Molecular Programming with a Transcription and Translation Cell-Free Toolbox: From Elementary Gene Circuits to Phage Synthesis

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jonghyeon G.Y.N Shin

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor Of Philosophy

August, 2012

Acknowledgements

I would like to thank my advisor, Vincent Noireaux, for his support and guidance through my graduate studies. I also thank to my collegues in Vincent's lab, Jerome Chalmeau, Nadezda Monina and Jonathan Gapp for making my lab life delightful. I expand my thanks to my collaborator, Paul Jardine for many helpful and valuable discussions. My thanks go even further to my friends, Jonathan Gapp, Matt Loth and Burak Himmetoglu for their editing help on this thesis.

Most of all, I thank my family, Sun, Min, Yoon, and my parents and my brother for their unlimited support and love.

Abstract

Cell-free synthetic/systems biology is an emerging field connecting biology, chemistry, physics, and engineering to understand biological systems and expand their capabilities. *In vitro* approaches compared to *in vivo* allow much better control of parameters and give much more freedom to program and study biological systems. Among the *in vitro* approaches, a transcription and translation (TX-TL) cell-free gene expression system mimicking a natural biological system offers the closest context to an intact cell. The conventional cell-free system as a playground to perform an experiment, however, has a couple of serious problems such as an insufficient sink system and the lack of transcriptional diversity. In this dissertation, I report the preparation of a custom-made *E. coli* cell-free system for the purpose of quantitative synthetic/systems biology, demonstrate synthetic gene circuits with cell-free toolbox, and show cell-free synthesis of a functional entity from genome-sized DNA. The custom-made cell-free system expresses genes with only endogenous TX-TL machinery and the sink systems for two biomolecules, mRNA and protein, can be applied in it. Moreover, mathematical models of gene expression including sink systems in this cell-free system are described. As a concept of cell-free toolbox, this cell-free system also makes it possible to use a variety of transcriptional activation and repression units to construct elementary circuit motifs. Furthermore, a bacteriophage as complex as T7 phage is synthesized from its genome-sized DNA with this cell-free system. This cell-free synthesis in a single test tube includes the central dogma of molecular biology including transcription, translation, and DNA replication as an internal process, and self-assembly and DNA packaging as a post-gene-expression process.

Contents

List of Tables

List of Figures

Chapter 1

Introduction[1]

1.1 An Overview of Cell-Free Synthetic Biology

Scholars describe the twentieth century as the era of physics because 1) quantum mechanics and relativity theories emerged and changed the way we think, 2) an enormous knowledge was developed from these new theories and 3) the resulting applications of the new physics had a considerable impact on our day-to-day life. We can ask what is going to be the next scientific revolution during the twenty-first century. Nowadays, people are more interested in the extension of life, in finding treatments against cancer and all sorts of viruses and diseases. It seems natural to label the twenty-first century as the era of biology, as well as its numerous related sub-fields such as biotechnology and bioengineering. Systems biology and synthetic biology are two other recent and rapidly growing multidisciplinary research areas dealing with biological systems [2] . The term 'synthetic biology' was first used in the early 1900's [1] and spread during the twentieth century including a couple of big steps; for example, the discovery and characterization of restriction enzymes for which the Nobel prize was awarded to Nathans, D., Arber,

[1] **I acknowledge that Vincent Noireaux's editing help was essential to achieve proper English usage and readability of this chapter.**

[2] Systems biology and synthetic biology are two recent and close scientific research fields. While systems biology focuses on understanding existing natural systems, synthetic biology consists of constructing artificial (non-existing) systems as a means to understand biology. Biological information, especially gene networks and regulation, is the main topic for both systems and synthetic biology. The ideas and techniques used in both fields overlap significantly. In this dissertation, I will use the term synthetic biology, which is more representative of my work.

W. and Smith, H. in 1978 [2]. Another major step, made during the late twentieth century, was the technical progress made in DNA sequencing and synthesis [3]. DNA became fast to sequence and cheap to synthesize, which promoted a very fast expansion of synthetic biology at the beginning of the twenty-first century [4]. Moreover, the development of modeling and simulation techniques to predict engineered biological systems and measurement using high technology make the quantitative research possible.

Interestingly, even though biology era is replacing the physics era, physicists and physics-based researcher such as Stanislas Leibler, Uri Alon, Michael Elowitz, and James Collins are in the forefront of modern synthetic biology era. In 2000, two innovative articles by these authors were published one after another in Science. Elowitz and Leibler engineered a synthetic oscillator composed of three transcriptional repressors, called a repressilator (one represses one of two other genes: TetR represses CI, CI represses LacI, and LacI represses TetR). The repressilator is a biological clock that does not exist in nature [5]. Gardner, Cantor and Collins built a synthetic bistable gene circuit. This network, composed of two repressors and promoters (promoter A is repressed by repressor A which is expressed by promoter B, and this repression is inhibited by inducer A^*) acts as a toggle switch, which can be flipped back and forth by adding chemical inducers to the medium [6]. Both papers are considered as turning points in modern synthetic biology; these works were reductionist, constructive (new DNA construction, not existing in nature) and quantitative (mathematical modeling and simulation compared to experimental data). Since then, Uri Alon has been one of the most active members in this area. He established that complex gene networks are composed of simple building blocks named network motifs. His papers and books have been cited more than 15,000 times [7, 8, 9, 10]. Many mathematicians, chemists, biologists, bioengineers, computer scientists, inspired by this work, are now working in this area. Most of the work produced is performed *in vivo* [5, 6, 11, 12, 13, 14, 15] or *in silico* [16, 17] because cells have already been studied a lot and biological systems are very complex objects.

In vitro synthetic biology is a much smaller field. Cell-free synthetic biology provides an experimental framework to understand and broaden our knowledge of biological systems in a constructive way by building biochemical networks out of a cell. This approach seeks to reconstruct biological systems, such as gene networks, from scratch

in a test tube. In most of the cell-free synthetic biology experiments, the reaction composition is entirely known as opposed to *in vivo* experiments, which use real living cells as factory. *In vitro*, small and simple building blocks, making complex systems, are studied in isolation. For this reason, cell-free approaches to biological processes are more quantitative [18] and the behavior of a complex system can be predicted from the description of simple building blocks [19]. Moreover, *in vitro*, artificial systems, that would be impossible to engineer *in vivo*, can be constructed [20, 21].

Various cell-free synthetic biology approaches have been investigated during the last decade. Enzyme-free biochemical circuits, using DNA only in buffered aqueous solutions, is the most active area [20, 22, 23]. The reactions are based on nucleotide sequence recognition, strand displacement and branch migration. Despite being relatively far from real biological systems, sophisticated biochemical reactions can be developed. Achievements include the construction of predictable large-scale biochemical system [24, 25, 26], the construction of complex dynamical behaviors from arbitrary systems of coupled chemical reactions [27] and the development of algorithmic self-assembly of nanostructure (called DNA origami) [28, 29, 30]. However, the construction of certain classes of biological behaviors, such as temporal and spatial pattern formation, seems far more complicated to construct with enzyme-free DNA system. To construct such systems, cell-free synthetic biology approaches using a reduced set of enzymes have been used [19, 31, 32]. Oscillators have been recently constructed with such systems [19, 32].

Among the other noticeable cell-free reductionist approaches, the construction of complex biochemical systems with a cell-free transcription and translation system (TX-TL) is the most complex but the closest to real biological systems. The TX-TL molecular machinery are extracted from cells and the endogenous information (genomic DNA and messenger RNAs) is removed to construct biochemical systems with only gene networks engineered in the laboratory. This approach allows performing complex information processes with the entire chain of information: DNA, messenger RNA and proteins [33]. This approach also escapes many constraints found in *in vivo* synthetic biology approaches. With such cell-free TX-TL system, one can construct, in principle, DNA programs to study gene regulation, to engineer network with predictable outputs, and to program the self-organization of complex biological systems outside real living organisms. In addition, the TX-TL machinery can manufacture all of their molecular

components. This allows envisioning the construction of self-reproducing systems from scratch. It is exciting to remind here the concepts of self-reproducing automata (artificial life) proposed by John von Neumann in the middle of 1900 [34, 35]. These concepts have been successful in computer architecture and computer science. The concepts developed by von Neumann can be applied to cell-free synthetic biology [36].

Although many types of cell-free TX-TL systems have been developed, no *in vitro* TX-TL systems have been optimized for synthetic biology purposes. During my PhD, I have been the first, with my collaborators, to develop a TX-TL cell-free toolbox specifically optimized for synthetic biology performed in a test tube. This system is becoming a biomolecular breadboards for prototyping and bebugging synthetic biocircuits.

1.2 Brief Results Inside This Dissertation

As of August 2012, my research has been reported in seven peer-reviewed publications [37, 38, 39, 40, 41, 42, 43].

1.2.1 A TX-TL Cell-Free Platform to Carry Out Quantitative Synthetic Biology in a Test Tube

Commercially available *E. coli* TX-TL cell-free gene expression systems have been optimized for applications in biotechnology rather than for quantitative synthetic biology [44, 45, 46]. These conventional systems use bacteriophage RNA polymerases (RNAP), for example T7 RNAP, one of the most powerful transcriptional machinery, added to a crude cell-free extract that provides the translation machinery. These systems do not offer any ways to control the dynamics of expression, such as mRNA and protein degradation. Therefore, such hybrid cell-free systems are absolutely not adapted for synthetic biology purposes.

The first goal of my research was to prepare and optimize an endogenous TX-TL cell-free gene expression system. The system is endogenous because, unlike the other systems described so far, we use the endogenous TX and TL machinery from the cells. The crude extract is prepared from *Escherichia coli* cells because it is a well-characterized organism, the preparation is cost-effective, and it requires low-cost energy sources [33]. Alongside the crude extract preparation, we developed two methods to adjust the mRNA

and protein degradation rates [37, 38]. The following points are covered in chapter 2 of my dissertation:

- Description of custom-made cell-free extract
- Components of cell-free reaction
- mRNA inactivation and rate of mRNA lifetime
- Protein degradation and combination with mRNA inactivation

1.2.2 A TX-TL Cell-Free Toolbox: elementary synthetic gene circuits

The other important bottleneck of conventional hybrid cell-free system is the lack of a repertoire of transcription units. These systems use only a few bacteriophage RNA polymerase. As a consequence, circuit connectivity is limited. Unlike an electric circuit, a gene circuit cannot repeatedly use the same motif because biological information is a bulk information. An electric circuit is well organized on a 2-dimensional circuit board and the current only flows along the 1-dimensional wire from one spot to another. In cells, transcription factors present in the cytoplasm have access to the entire DNA information. Therefore, having two or more copies of the same regulatory part does not make sense in biological systems. The second goal of my work was to develop the cell-free system as a toolbox with a library of transcription factors for both activation and repression of genes. I, then, showed how gene circuits were constructed with this toolbox [42]. In chapter 3 of my dissertation I discuss the following:

- List and characterization of transcriptional activation units
- Features of activation units in our custom-made cell-free system
- Elementary gene circuits
- Transcriptional repression units and synthetic gene circuits
- Toward artificial cell: elementary gene circuit in liposome

1.2.3 Mathematical Models of Gene Expression

In collaboration with the Bar-Ziv group at the Weizmann institute, we developed a coarse-grained model that describes the cell-free expression of one gene. This model is the first model that describes quantitatively protein synthesis *in vitro* [40]. To account for the coupling between TX and TL, a coupled coarse-grained model was developed. This coupled coarse-grained model was also developed to explain negative feedback process, for example negative feedback by auto regulation. In chapter 5 of my dissertation, I discuss two models:

- Coupled coarse-grained model
- Coupled coarse-grained model with negative feedback process

1.2.4 Genome-Sized DNA Programming

Cell-free expression with our custom-made system became very powerful, as we continuously optimized the system in the course of my PhD. Beyond the construction of cell-free gene circuits, we decided to test the capacity of our system to run large DNA programs that encode complex self-organized structure and function. Instead of constructing by hand a DNA program made of tens of genes and regulatory parts, we looked for natural DNA programs that could be expressed with our toolbox. The bacteriophage T7, composed of 57 genes, was selected as a natural DNA program compatible with our system. It is a good start with natural genome because the result immediately tells us whether the cell-free synthesis in the system works or not. In chapter 5, we show that large DNA programs, such as the T7 genome, can be entirely expressed into real functional phages. This work includes the expression, self-assembly, DNA replication, and packaging of the phage T7 [43]:

- The reason that bacteriophage T7 genome is selected
- Cell-free synthesis of bacteriophage T7 and its genome
- Extension of phage work; synthesis of bacteriophage ΦX174

Chapter 2

Cell-Free System[1]

2.1 Cell-Free Extract

The cell-free system consists of hardware, which is the transcriptional and translational machinery provided by the cytoplasmic extract from *E. coli*; software, which is the information on the nucleic acid coding. The schematic diagram of gene expression with a cell-free system in a test tube is shown in Figure 2.1. This *in vitro* approach is the only experimental platform on which to perform informational and metabolic processes using the central dogma of molecular biology in a single test tube and to construct a feeding exchange configuration for long-lived biochemical reactions, including those of an artificial cell. The PURE (Protein synthesis Using Recombinant Elements) system, another *in vitro* TX-TL-based biochemical system which is composed of minimal enzymes to synthesize protein from DNA, can be used to study informational and metabolic processes, yet it has not been reported as being used for long-lived biochemical reaction in a continuous feeding-exchange configuration [47]. The cell-free system is not identical but almost the same composition as an intact cell because cell-free extract is directly obtained from a cell by removing the cell wall and endogenous information (genomic DNA and messenger RNAs). Researchers using a cell-free system can work under circumstances similar to that of the intact cell but with much better control over parameters such as DNA concentration, ion concentration, half-life of messenger RNA, etc. Of

[1] **I acknowledge that Jonathan Gapp's editing help was essential to achieve proper English usage and readability of this chapter.**

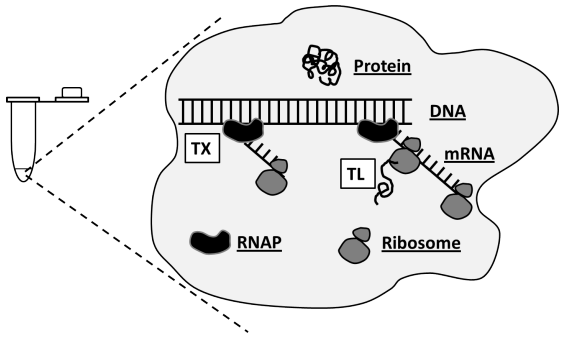

Figure 2.1: A gene encoded on DNA is expressed to protein through mRNA with the cell-free system providing transcriptional and translational machinery in a single test tube.

equal importance is the fact that they also have much more freedom in programing and studying these biological systems. Moreover, certain genes from other organisms, including even humans, can be expressed with the *E. coli*-based cell-free system [48]. In addition to these advantages, the cell-free system provides the ability to bypass toxicity problems so often met in *in vivo* experimentation [49]. For instance, studying a toxin protein *in vivo* is almost impossible because this protein disrupts certain processes in the cell once the toxin gene encoded on recombinant DNA has been inserted into the cell and is overexpressed. Therefore, a cell-free system will potentially become an alternative experimental platform to mimic cell circumstances with many advantages, the biggest of which is the ability to perform experiments that would be impossible in an *in vivo* environment.

Well-characterized and well-developed cell-free extract makes this idea possible. The question is, how well can cell-free extract be prepared so as to replicate an intact cell and, how well can a cell-free extract be engineered for a specific purpose if required? The preparation of a cell-free system has been developed and modified since it was used more than a hundred years ago. It is not clear when, where, how, why and by whom the first cell-free extract was used, but the first successful demonstration was shown by Eduard Buchner who won the Nobel Prize for chemistry in 1907. His research was converting sugar to ethanol using yeast cell-free extract [50]. In 1958, Paul Zamecnik and Mahlon Hoagland, without former knowledge of Crick's adaptor hypothesis, the central dogma of molecular biology, or the sequence hypothesis; identified and isolated

intermediate carrier molecules. These carrier molecules exactly matched Crick's adaptor hypothesis, now named tRNA. All of this was accomplished using cell-free extract from rat liver cells [51]. The cracking of the genetic code using *E. coli* cell-free extract by Marshall Nirenberg and Heinrich Mattaei in 1961 was not only the first step to solving genetic coding mechanisms in DNA for peptide synthesis, but also opened up the possibility of using cell-free extract for DNA-dependent biological mechanisms [52, 53]. After their notable experiment, many research groups in the 60's and 70's used *E. coli* cell-free extract to study the genetic coding mechanism such as regulation of gene expression [54, 55]. Meanwhile, the first landmark work for *E. coli* cell-free extract preparation was accomplished by Geoffrey Zubay in 1973 [56].

Modern streamlined protocol of *E. coli* cell-free extract was developed by Julie Pratt in early the 80's [57]. In the early 90's, she used the T7 bacteriophage transcription machinery in an *E. coli* cell-free system [58]. A plasmid encoding the T7 promoter was added into the *E. coli* cell-free extract containing T7 RNA polymerase (RNAP) and this hybrid system expressed much more protein than previous systems because the T7 bacteriophage's transcription set was very powerful and specific. The increased potency of the promoter is because naturally, a lytic phage or a phage in lytic cycle must express its genes in order to create its offspring and escape from a host cell before cell division occurs, causing dilution of materials. More delicate modifications to this cell-free system expanded the application area from proteomics to biotechnology and even to industry [33, 44, 45, 46, 59, 60, 61, 62]. Three main types of modification have occurred: economical and simple crude extract preparation, high-protein yield and synthesis rates, and development of an energy regeneration source. All modifications, however, have been slanted too much in one direction toward the application. When one looks back to the 60's, it is fully apparent that *E. coli* cell-free system had the merit to study DNA-dependent biological mechanisms. As such, a couple of researchers tried to use conventional cell-free systems developed for application to study synthetic gene circuits [63], pattern formation [64] and synthetic vesicles [65, 66], but these systems were not well received. For the aims of synthetic biology, the development of a cell-free system must value control of the system rather than pure protein production capability. That is, the cell-free system must have the potential for the use of a variety of transcriptional units and some control over inactivation/degradation processes coupled with synthesis

processes (sink-source pair mechanism), etc. rather than simply being concerned with large-scale protein production. Even though a cell-free system from a prokaryotic cell has a disadvantage in its limited post-translational modification, *E. coli* cell-free extract is the most reliable system and it has more advantages than disadvantages when compared with other organisms [59]. However, another modification of *E. coli* cell-free extract for the purpose of DNA-dependent information process was needed.

E. coli cell-free extract containing translational machinery includes core RNAP and housekeeping transcription initiation sigma factor σ^{70} as the transcriptional machinery. The complex of core RNAP and σ^{70}, called RNAP holoenzyme with σ^{70}, starts transcription and core RNAP finishes producing mRNA, similar to what happens in an *E. coli* cell [67]. Genes cloned under the σ^{70} specific promoter on a recombinant plasmid, however, were expressed with RNAP holoenzyme much less effectively than expression by the T7 bacteriophage transcription set (T7 RNAP and its specific promoter). After Julie Pratt's findings, most modifications of the cell-free system were done with the T7 bacteriophage set so as to make a conventional cell-free system using a cell-free extract with the purpose of massive protein production more applicative usage. In this study, however, the goal is to allow for information processing in a cell-free system, requiring the development of *E. coli* cell-free extract for this purpose. This unique cell-free system, later defined as a cell-free toolbox (chapter 3), is characterized by a couple of features:

1. all endogenous machinery (in this chapter) [37],
2. mRNA inactivation and protein degradation (section 2.3) [38],
3. a variety of transcriptional units (section 3.2) [42].

This system, which primarily uses RNAP holoenzyme with σ^{70} for transcription even though it was thought to be less efficient than conventional hybrid cell-free extract, has a couple of advantages: 1) similar to a cell and 2) variety of transcriptional regulatory units. *E. coli* has seven sigma factors, σ^{19}, σ^{24}, σ^{28}, σ^{32}, σ^{38}, σ^{54} and σ^{70}. All sigma factors except housekeeping sigma factor σ^{70}, which *E. coli* uses as the primary transcriptional initiation factor in order to stimulate most of the genes, are usually silent until a cell is exposed to a certain environment. This exposure causes expression of the

gene in order to cause the cell to adapt and sustain functionality [67]. In this way, our cell-free system works like *E. coli*, making it possible to study existing informational networks in a cell as well as to construct more complex synthetic circuits due to the larger number of regulatory units available in *E. coli* than in the bacteriophage set.

A big challenge yet remaining was overcoming the low capacity for gene expression. Conventional hybrid cell-free system could express as much as 1 mg/ml of protein, but the cell-free system using only endogenous machinery could do only 100-fold less than the conventional one [59]. When beginning to develop the *E. coli* cell-free extract in 2007, the capacity of protein synthesis was less than 500 nM (about 0.01 mg/ml of repoter protein, eGFP). Since then, the *E. coli* cell-free extract has been developed with the testing of different buffers and the engineering of a σ^{70} specific promoter, as well as streamlining the cell-free extract preparation protocol. Presently, the cell-free system can express reporter gene, deGFP at 0.65 mg/ml of functional protein in batch mode, 0.75 mg/ml of total protein in batch mode and more than 3 mg/ml of functional protein in feeding-exchange dialysis mode [37, 42]. This protocol is described below. In brief summary, it is a 3 day protocol consisting of bacteria plating (Day 1), cell growing (Day 2) and crude extract preparation (D-day). A crude extract preparation is streamlined; cell growth $\rightarrow$ cell lysis (by bead beating) $\rightarrow$ filtration (primitive crude extract) $\rightarrow$ pre-incubation $\rightarrow$ dialysis $\rightarrow$ cell-free crude extract. Cell-free crude extract can be stored at -80 $^{\circ}C$ for a few years while maintaining its capacity to efficiently synthesize protein.

Day 1 (D-2): bacteria plating

- Cell: BL21 Rosetta2 competent cell bearing pRARE2 plasmid on which seven rare tRNAs are encoded

- Prepare a 25 ml of 1.1 % agar plate in the morning: 0.775 g of 2xYT powder (31 g/l) + 0.275 g of bacto-agar powder(1.1 % (w/v)) + 1.55 ml of phosphate solution (645 mM of potassium phosphate dibasic solution and 355 mM of potassium phosphate monobasic solution) + 25 μl of chloramphenicol (34 μg/ml) $\rightarrow$ autoclave 2xYT + bacto-agar and phosphate separately, mix them all at 55 $^{\circ}C$, and pour 25 ml of mixture in a 100 mm diameter petri dish

- Add tiny amount of bacteria onto the surface of agar plate, spread it well, and incubate it in a 37 $^{\circ}C$ in the late afternoon

Day 2 (D-1): cell growing

- At noon, start a miniculture; 4 ml of 2xYT broth at 31 mg/ml + 0.27 ml of phosphate solution (645 mM of potassium phosphate dibasic solution and 355 mM of potassium phosphate monobasic solution) + 4 μl of chloramphenicol at 34 mg/ml, with one colony picked from an agar plate (from D-2)

- Prepare 2xYT medium, S12A and S12B buffers, and phosphate in order to use at D-day

- 2xYT medium: dissolve 62 g of 2xYT powder with 1.88 L of water in total volume $\rightarrow$ prepare 2 $\times$ 2 L bottles

- S12A buffer: dissolve 10.88 g of magnesium glutamate (MW=388.61) and 24.4 g of potassium glutamate (MW=203.23) with 1.9 L of water in 2 L beaker with magnetic stick on magnetic stirrer, and add water up to 2 L after adjusting pH at 7.7 with 50 ml of Tris at 2 M and some amount of acetic acid

- S12B buffer: dissolve 10.88 g of magnesium glutamate (MW=388.61) and 24.4 g of potassium glutamate (MW=203.23) with 1.9 L of water in 2 L beaker with magnetic stick on magnetic stirrer, and add water up to 2 L after adjusting pH at 8.2 with Tris at 2 M

- Phosphate: 170 ml of potassium phosphate dibasic solution (1 M) + 93.5 ml of potassium phosphate monobasic solution (1 M)

- After autoclaving them, place 2xTY medium and phosphate on bench (at room temperature), and S12A and S12B buffer in 4 $^{\circ}C$ fridge

- At 8:30 pm, start a midiculture, 50 ml of 2xYT broth at 31 mg/ml + 3.3 ml of phosphate solution (645 mM of potassium phosphate dibasic solution and 355 mM of potassium phosphate monobasic solution) + 50 μl of chloramphenicol at 34 mg/ml + 100 μl of the first miniculture

Day 3 (D-day): crude extract preparation

- All steps are done on ice (0∼4 $^{\circ}C$) except maxiculture and 80 min pre-incubation (both are at 37 $^{\circ}C$)
- At 5:00 am, start a maxiculture: add 124 ml of phosphate into 1.88 L of 2xTY medium, and adjust volume up to 2 L with water → aliquot it at 600 ml in 4 L erlenmeyer flask: 6 × 4 L erlenmeyer flask
- At 8:00 am, measure OD at 600 nm, and if needed, grow cells until OD at 600 nm becomes 1.5: normally it takes 20∼30 min more to reach OD600=1.5
- Meanwhile, measure weight of 50 ml centrifuge bottles, and add 4 ml of DTT at 1 M into a S12A buffer bottle
- Start to centrifuge (parameters: 5000 rpm for 12 min at 4 $^{\circ}C$) 500 ml centrifuge bottles with full of cell-medium in order to get pellet of cell → 4 centrifuge bottles
- Remove supernatant, and do one more run with same parameters as the previous step
- Remove supernatant, resuspend pellet with 220 ml of S12A buffer with 2 mM of DTT, and centrifuge with same parameters as the previous step
- Remove supernatant and do one more run with same parameters as the previous step
- Remove supernatant, resuspend pellet with 37.5 ml of S12A buffer with 2 mM of DTT, and transfer cell to 50 ml centrifuge bottles → 4 centrifuge bottles
- Centrifuge them at 4000 rpm for 8 min at 4 $^{\circ}C$
- Remove supernatant, and centrifuge them at same parameters as the previous step but 2 min
- Remove supernatant as much as possible, and measure weight of 50 ml centrifuge bottles containing cell pellets

- Calculate amount of S12A buffer and glass bead (a 100 μm diameter) that will be added in each individual 50 ml centrifuge bottle depending on weight of cell $\rightarrow$ S12A buffer in ml and bead in g are calculated as weight of cell pellet in g $\times$ 0.9 and $\times$ 5.0, respectively $\rightarrow$ i.e.) 4 g of cell means 3.6 ml of S12A buffer and 20 g of bead
- Add S12A buffer in a bottle, and dissolve pellet well by vortexing
- Mix sterile dry bead with cell pellet homogeneously: add one third of bead in a bottle and vortex $\rightarrow$ repeat two more times
- Transfer mixture into 2 ml bead beating tubes, and do bead beating with parameters: 46 rpm for 30 sec $\rightarrow$ twice (give 30 sec break between 2 runs)
- Make a structure for filtration: use micro bio-spin chromatography column
- Centrifuge them at 7,500 rpm for 10 min at 4 $^{\circ}C$
- Transfer supernatant (primitive crude extract) in 1.7 ml micro-centrifuge tube, and centrifuge them with the parameters: 12,000 rpm for 20 min at 4 $^{\circ}C$
- Transfer 500 μl of supernatant in new 2 ml bead beating tubes, and do 80 min pre-incubation at 37 $^{\circ}C$
- Transfer them into 1.7 ml micro-centrifuge tubes, and centrifuge them at 12,000 rpm for 10 min at 4 $^{\circ}C$
- Meanwhile, prepare dialysis setting: add 900 ml of S12B buffer with 1 mM of DTT and magnetic stick in 1 L beaker, and wash dialysis cassettes (10 kDa MWCO) applied on floaters into S12B buffer on magnetic stirrer (300 rpm) at 4 $^{\circ}C \rightarrow 2 \times$ 1 L beaker
- Once centrifuge is done, inject 2$\sim$2.5 ml of crude extract into dialysis cassette (maximum capacity at 3 ml), and dialyze sample for 3 hours
- Meanwhile, check the protein concentration by Bradford protein assay
- Extract crude lysate from dialysis cassettes $\rightarrow$ we call it *E. coli* crude extract

- aliquot it by 30 μl, and quickly freeze them with liquid nitrogen $\rightarrow$ $\sim$9.5 mg/ml is optimal of final protein concentration in 90 μl total volume of cell-free reaction $\leftarrow$ cell-free components are diluted 20 - 30 times compared with *in vivo*, 200$\sim$320 mg/ml [68, 69]

2.2 Cell-Free Reaction

2.2.1 Nutrients

Cell-free reaction (typically 90 μl) is composed of one third of crude extract and two thirds of nutrients and information source. As explained, crude extract whose raw protein concentration is 27$\sim$30 mg/ml (9$\sim$10 mg/ml of final concentration in a reaction) provides transcriptional and translational machinery. Crude extract is dialyzed for 3 hours as the last step of preparation, so small size molecules (under 10 kDa) including building blocks such as ATP, GTP, amino acids and supplemental components such as cAMP and folinic acid are additionally required for the biochemical reactions of gene expression. In order to allow for long-lasting biochemical reaction time and more efficient protein synthesis in a test tube, energy regenerating resources should be added as well. Nutrients are also modified in the cell-free system while adjusting the concentration of materials also boosts the quality of the cell-free system. Herein, three buffers are compared; PEP, CP and 3-PGA buffer, whose components are similar except for the energy regenerating resource. Also included is a short description of the importance of each compound.

List of nutrient with the main reasons in cell-free system

- HEPES: organic chemical buffering agent, better at maintaining physiological pH
- ATP (Adenosine TriPhosphate): building block of mRNA, main biochemical energy source and transport, and multifunctional coenzyme
- GTP (Guanosine TriPhosphate): building block of mRNA, biochemical energy source
- CTP (Cytidine TriPhosphate): building block of mRNA

- UTP (Uridine TriPhosphate): building block of mRNA
- tRNA (transfer RNA): bridge molecule connecting single amino acid to three letter codon
- Coenzyme A: coenzyme helping ATP regeneration
- NAD (Nicotinamide Adenine Dinucleotide): electrons carrier in redox (REDuction - OXidation) reaction
- cAMP (cyclic Adenosine MonoPhosphate): CRP (transcription factor cAMP receptor protein, also called CAP - catabolite gene activator protein) - cAMP complex which increases gene expression
- Folinic acid: helping DNA replication and RNA transcription
- Spermidine (Putrescine): binding and precipitating of DNA and stimulating T7 RNAP to assist *in vitro* transcription of mRNA
- PEP (PhosphoEnolPyruvic acid): substrate of ATP regeneration mechanism by enzyme, pyruvate kinase which exists in crude extract
- CK (Creatine Kinase): enzyme mediating ATP regeneration from creatine phosphate
- CP-K (Creatine Phosphate, dipotassium): substrate of ATP regeneration mechanism by enzyme, creatine kinase
- 3-PGA (3-PhosphoGlyceric Acid): substrate of ATP regeneration mechanism by enzyme, pyruvate kinase which exists in crude extract
- AA (Amino Acids): building block of protein
- DTT (DiThioThreitol): reducing the disulfide bonds of proteins (make proteins stable)
- Magnesium (Mg): essential ion to the basic nucleic acid chemistry of life (therefore it is important to all cells of all known living organism) because of the interaction between phosphate and magnesium ions; crucial ion to the protein structure and

accordingly the enzyme functionality for their catalytic action such as transcription and translation

- Potassium (K): serving as mediating biomolecular interactions
- PEG8000 (PolyEthylene Glycol, 8 kDa): flexible, water-soluble and biologically compatible polymer which provides molecular crowding and creates high osmotic pressures

Three energy regenerating resources, PEP, CP(CK + CP-K) and 3-PGA with HEPES, nucleoside triphosphate, tRNA, coenzyme A, NAD, cAMP, folinic acid and spermidine (or putrescine) are bufferized according to the literature and magnesium, potassium, amino acids, DTT and PEG8000 are separately added in the reaction [2] [62, 70, 71, 72]. Concentration of each element in each buffer is as follows,

i) PEP buffer

50 mM of HEPES pH 8, 1.5 mM of ATP and GTP, 0.9 mM of CTP and UTP, 0.2 mg/ml of tRNA, 0.26 mM of coenzyme A, 0.33 mM of NAD, 0.75 mM of cAMP, 0.068 mM of folinic acid, 1 mM of putrescine, and 30 mM of PEP

ii) CP buffer

50 mM of HEPES pH 8, 1.5 mM of ATP and GTP, 0.9 mM of CTP and UTP, 0.17 mg/ml of tRNA, 0.26 mM of coenzyme A, 0.33 mM of NAD, 0.75 mM of cAMP, 0.068 mM of folinic acid, 0.5 mM of spermidine, 0.0032 mg/ml of CK, and 30 mM of CP-K

iii) 3-PGA buffer

50 mM of HEPES pH 8, 1.5 mM of ATP and GTP, 0.9 mM of CTP and UTP, 0.2 mg/ml of tRNA, 0.26 mM of coenzyme A, 0.33 mM of NAD, 0.75 mM of cAMP, 0.068 mM of folinic acid, 1 mM of spermidine, and 30 mM of 3-PGA

The optimal condition of each buffer in terms of protein production is adjusted with magnesium glutamate (0∼10 mM additionally), potassium glutamate (0∼120 mM additionally), amino acids (0.5∼1.5 mM of each amino acid), DTT (0∼1.5 mM additionally) and PEG8000 (0.5∼4 %) [3] . Even if gene expression depends on all parameters added,

[2] Magnesium glutamate and potassium glutamate are used for the source of magnesium and potassium, respectively [72].

[3] Crude cell-free extract was dialyzed against S12B buffer containing 4.5 mM of magnesium glutamate, 20 mM of potassium glutamate and 1 mM of DTT. The quantity of magnesium glutamate, potassium glutamate and DTT shown in this thesis is additional in crude extract.

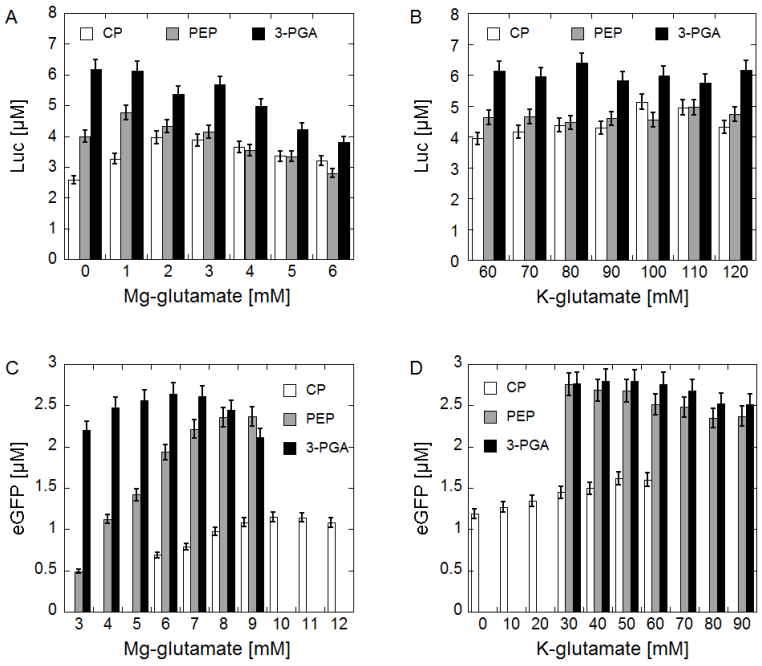

Figure 2.2: End-point protein expressions as a function of magnesium glutamate and potassium glutamate with three different buffers, CP, PEP and 3-PGA buffer are shown. **(A)** A reporter protein, Luc was synthesized as a function of magnesium glutamate with 5 nM of plasmid in which gene *luc* was cloned under σ^{70} specific promoter. The cell-free reaction (90 mM of potassium glutamate, 1.5 mM of each amino acid, and 2 % of PEG8000) was incubated at 22 $^{\circ}C$ for 8 hours. **(B)** A reporter protein, Luc was synthesized as a function of potassium glutamate with 5 nM of plasmid in which gene *luc* was cloned under σ^{70} specific promoter. The cell-free reaction (2, 1 and 0 mM of magnesium glutamate for CP, PEP and 3-PGA buffer, respectively, 1.5 mM of each amino acid, and 2 % of PEG8000) was incubated at 22 $^{\circ}C$ for 8 hours. **(C)** A reporter protein, eGFP was synthesized as a function of magnesium glutamate with 5 nM of plasmid in which gene *eGFP* was cloned under σ^{70} specific promoter. The cell-free reaction (40 mM of potassium glutamate, 1 mM of each amino acid, and 2 % of PEG8000) was incubated at 29 $^{\circ}C$ for 8 hours. **(D)** A reporter protein, eGFP was synthesized as a function of potassium glutamate with 5 nM of plasmid in which gene *eGFP* was cloned under σ^{70} specific promoter. The cell-free reaction (10, 8 and 6 mM of magnesium glutamate for CP, PEP and 3-PGA buffer, respectively, 1 mM of each amino acid, and 2 % of PEG8000) was incubated at 29 $^{\circ}C$ for 8 hours.

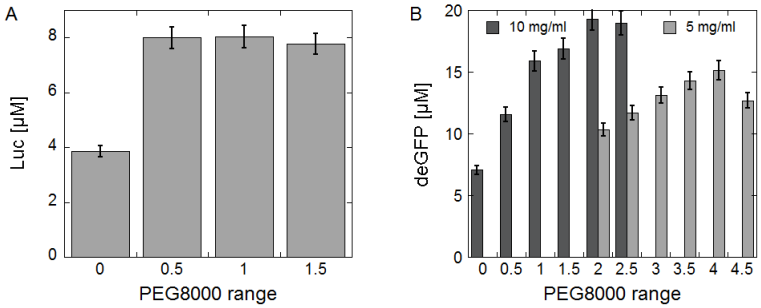

Figure 2.3: End-point protein expressions as a function of concentration of PEG8000 with 3-PGA buffer are shown. **(A)** A reporter protein, Luc was synthesized with 5 nM of plasmid in which gene *luc* was cloned under σ^{70} specific promoter. The cell-free reaction (0 mM of magnesium glutamate, 80 mM of potassium glutamate, and 1.5 mM of each amino acid) was incubated at 22 $^{\circ}C$ for 8 hours. **(B)** A reporter protein, deGFP was synthesized with 5 nM of plasmid in which gene *deGFP* was cloned under σ^{70} specific promoter. The cell-free reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, and 1.5 mM of each amino acid) was incubated at 29 $^{\circ}C$ for 8 hours. deGFP protein production is compared when protein quantity of crude extract (dark grey) was halved (light grey).

the main parameters that have the most effect are magnesium glutamate, potassium glutamate, PEG8000 and plasmid concentration (Figure 2.2, 2.3). The result of plasmid concentration and construction will be shown in the next chapter, so 5 nM of plasmids, pBEST-UTR1-Luc and pBEST-UTR1-eGFP or pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (how the plasmid was named is also shown in the next chapter), are used for standardizing the optimization in this chapter. As shown in the figure 2.2, magnesium glutamate concentration has to be precisely adjusted for both reporters in three different buffers whereas potassium glutamate is less sensitive than magnesium glutamate. Protein production at the optimal conditions is greater with 3-PGA buffer than PEP or CP buffer, so all experiments for this thesis were done with 3-PGA buffer.

Molecular crowding is another important parameter, having a large effect on protein production. The reconcentration of components in the cell-free reaction due to the physical space occupied by PEG8000 results in a decrease in the internal distance between the components. Two reporter proteins have different optimal concentrations because of dependence on the size of mRNA and protein as well as molecular crowding effects on

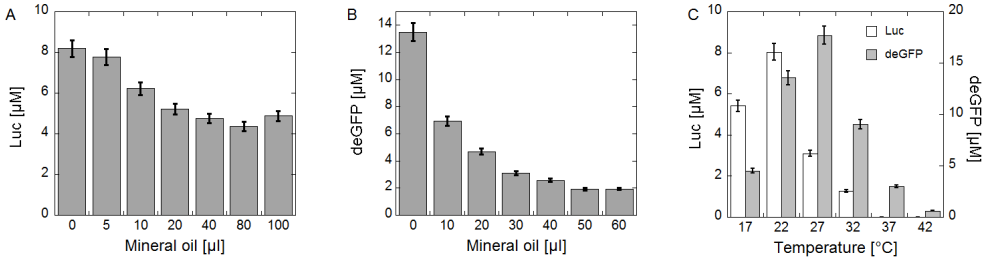

Figure 2.4: End-point protein expressions as a function of mineral oil added and the reaction incubation temperature are shown. **(A)** End-point measurement of Luc production is shown under oxygen cut circumstance by mineral oil. Different quantity of mineral oil was added on the top of 5 μl of the cell-free reaction in a 1.7 ml microcentrifuge tube and the reaction (0 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 1 % of PEG8000, and 3-PGA buffer) was incubated at 22 $^{\circ}C$ for 8 hours. **(B)** End-point measurement of deGFP production is shown under oxygen cut circumstance by mineral oil. Different quantity of mineral oil was added on the top of 5 μl of the cell-free reaction in a 1.7 ml microcentrifuge tube and the reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 2 % of PEG8000, and 3-PGA buffer) was incubated at 29 $^{\circ}C$ for 8 hours. **(C)** Two reporter proteins were synthesized at six different incubation temperatures for 8 hours. End-point protein expression was done with the same condition as in the panel (A) for Luc protein and as in the panel (B) for deGFP protein.

protein folding (Figure 2.3). deGFP gene expression with half the protein concentration of the normal cell-free crude extract stock solution was measured while adjusting PEG8000 concentration (Figure 2.3 (B)). 75 % of protein production with half the protein quantity of crude extract (5 mg/ml) was measured with 4 % of PEG8000. In the event that an overly dense extract creates a problem or light extract has an advantage when measuring the output, as is the case for experiments using QCM (Quartz crystal microbalance), half the quantity of crude extract can be used with compatible protein yield.

In addition, the surrounding environmental parameters, oxygen and temperature, were realized as critical ones (Figure 2.4). An experiment to study oxygen effect was carried out by blocking ambient oxygen with a layer of mineral oil on top of the cell-free reaction in a tube, so that only oxygen dissolved in the cell-free reaction was used. The result of Luc protein production shows that cell-free reaction in contact

with oxygen boosts the gene expression by two fold (Figure 2.4 (A)). deGFP protein expression is affected more than Luc because maturation of deGFP fluorophore requires oxygen [73]. A 6∼7-fold decrease in fluorescent deGFP protein was measured in the case where ambient oxygen was completely blocked from the cell-free reaction (Figure 2.4 (B)). Temperature dependence may be due to factors such as transcription, translation, sequence-dependent stability of mRNA and protein folding as well as maturation of the fluorophore (Figure 2.4 (C)). Luc protein production was optimal at 22 $^{\circ}C$ and this result is similar to the previous studies *in vivo* [74]. 29 $^{\circ}C$ and 32 $^{\circ}C$ are observed as optimal temperatures for deGFP and eGFP protein expression, respectively (eGFP data is not shown, see my publicatin [37]). This result is also similar to the previous *in vivo* study with wild type GFP [75].

2.2.2 Plasmid

Plasmid is another very important parameter for protein synthesis in a cell-free system. In this thesis, genes, *Luc*, *eGFP* or its variants are encoded on a plasmid to measure output signal by luminescence or fluorescence, respectively, and are used to characterize hardware (cell-free reaction, as described in the previous chapters) and software (engineered circuits, as will be discussed in the ensuing chapters). Unfortunately, the characterization of the cell-free system is very different according to the reporter protein used. As shown in the previous chapters, the optimal condition of ions, PEG8000 and incubation temperature for Luc and eGFP/deGFP were not identical and the same is true of plasmid concentration. There are, however, two common aspects related to plasmid concentration between the two reporters (Figure 2.5). 1) Protein production is maximal with 10∼20 nM plasmid concentration. At this concentration, ∼10 μM (∼0.6 mg/ml) of functional Luc protein was detected while ∼25 μM (∼0.65 mg/ml) of fluorescent deGFP protein was measured. 2) Plasmid concentration can be split by two regimes; a linear regime and a saturated regime. When a twofold input of plasmids yields twice more protein production, that input plasmid concentration is in the linear regime (regime I). Regime II shows that twice the input plasmid concentration does not produce twice as much protein. In the linear regime, transcription of input DNA and translation of synthesized mRNA occur without any restriction in terms of working machinery. At a later point, it will be shown that this separation becomes very important

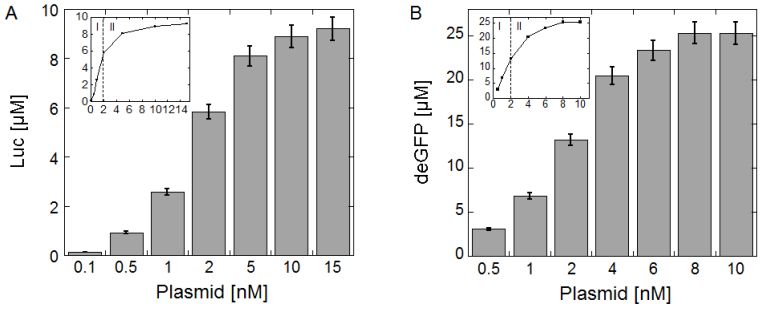

Figure 2.5: End-point protein expressions as a function of plasmid concentration are shown. **(A)** A reporter protein, Luc was synthesized as a function of plasmid, pBEST-UTR1-Luc, concentration. The cell-free reaction (0 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 1 % of PEG8000, and 3-PGA buffer) was incubated at 22 $^{\circ}C$ for 8 hours. **Inset**: It shows two regimes of plasmid concentration, a linear regime (I) and a saturated regime (II). **(B)** A reporter protein deGFP was synthesized as a function of plasmid, P_{70}–*deGFP*, concentration. The cell-free reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 1 % PEG8000, and 3-PGA buffer) was incubated at 29 $^{\circ}C$ for 8 hours. **Inset**: It shows two regimes of plasmid concentration, a linear regime (I) and a saturated regime (II).

when constructing synthetic circuits. It becomes evident that 2 nM of total input DNA is roughly observed as a barrier to distinguish the two regimes.

There is another important fact to remember, which is that gene expression is sequence dependent. How well a gene is expressed depends on how fast and how strong RNAP holoenzymes and ribosomes can recognize and bind to the promoter for transcription and to the RBS (Ribosome Binding Site) for translation, respectively. Conventional hybrid cell-free systems use T7 RNAP and its specific promoter to express genes because this combination is very powerful, yielding high protein in batch mode within a couple of hours. The endogenous *E. coli* transcription pair, RNAP holoenzyme and its specific promoter are not as efficient as the bacteriophage T7 transcription pair, therefore DNA used in our cell-free system must be designed well if it is to yield a quantity of protein comparable to the conventional system. Even though the aim of this research is not to produce large amounts of protein, a large capacity of protein production expands the potential of what can be done with our cell-free system. There are three parts that we can engineer to enhance production: promoter, UTR (UnTranslated Region including

Plasmid pBEST-	Protein production [μM]	(mg/ml)
PtacI-eGFP	0.02	(0.00067)
PtacI-UTR1-eGFP	1.05	(0.0284)
PtacI-UTR1-eGFP-T500	1.10	(0.0298)
PtacI-UTR1-eGFP-Del6	3.16	(0.0836)
PtacI-UTR1-eGFP-Del6-229	3.72	(0.0944)
PtacI-UTR1-eGFP-Del6-229-T500	4.65	(0.118)
OR2-OR1-Pr-UTR1-eGFP	1.40	(0.0375)
OR2-OR1-Pr-UTR1-eGFP-Del6	3.30	(0.0874)
OR2-OR1-Pr-UTR1-eGFP-Del6-229	4.24	(0.108)
OR2-OR1-Pr-UTR1-eGFP-Del6-229-T500	5.07	(0.129)

DNA PART	Sequences
PtacI	TTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGT GAGCGGATAACAATT
OR2-OR1-Pr	TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCG GTGATAATGGTTGCA
UTR1	AATAATTTTGTTTAACTTTAAGAAGGAGATATA
T500	CAAAGCCCGCCGAAAGGCGGGCTTTTCTGT

Table 2.1: Plasmid encoding green fluorescent protein is modified to maximize synthesis of gene. Modification of regulatory parts and gene are shown with the result of protein production. Protein production is compared with 1 nM of plasmid. PtacI, OR2-OR1-Pr, UTR1 and T500 sequences are shown.

with RBS) and the gene itself.

pBEST-Luc purchased from Promega was the root plasmid from which all other plasmids were constructed. This plasmid has a PtacI consensus promoter which is strong and specific to RNAP holoenzyme with σ^{70} and a modification was made to have the UTR site contain the T7 *g10* leader sequence for highly efficient translation initiation [76, 77]. The maximum protein production with a plasmid, pBEST-UTR1-Luc was ∼10 μM (∼0.6 mg/ml) at the range of 10∼20 nM plasmid concentration (Figure 2.5 (A)). No other modification yielded more Luc proteins, but the amount of protein synthesized is compatible to a conventional, hybrid, cell-free system whose maximum protein production is about 1 mg/ml. There was, however, a sizeable drop in protein production with *eGFP* gene. Even if it is shown in the Table 2.1 that pBEST-UTR1-eGFP boosted the protein expression level 50 times compared to pBEST-eGFP, the maximum eGFP protein production from pBEST-UTR1-eGFP was ∼4 μM (∼0.1 mg/ml)

(data not shown, see my publication [37]). In order to test whether this is due to the transcription process or not, pure *E. coli* RNAP saturated with σ^{70} was added to the cell-free reaction, however, the protein production did not increase (data not shown, see my publication [37]). It was realized that the decrease was due to the poor translation of the *eGFP* gene even though a strong UTR1 was used. The decrease in translation was because the *eGFP* gene has ribosome binding site-like (RBS-like) sequences at the beginning of the gene so the N-terminal sequences of eGFP were being truncated. According to a previous work, a new *eGFP* gene was formed by modifying the N-terminus via a silent mutation to remove the same potential problem [37]. A new gene called *eGFP-Del6* cloned with PtacI-UTR1 was synthesized at 3 times the amount of *eGFP* cloned with PtacI-UTR1 (Table 2.1). Another modification on the gene was done in the C-terminal according to the previous work [37]. This new gene called *eGFP-Del6-229* [4] boosted the gene expression up another 20 % (Table 2.1).

Amplification of the plasmid containing the *deGFP* gene poses its own set of difficulties. Heightened over-expression of deGFP causes toxicity, inhibiting the healthy growth of *E. coli*. Even though the ptacI consensus promoter is fused with the LacO operator and no toxicity problem has been observed so far, this regulatory part may leak as the result of a mutation, causing the toxicity problem. Therefore, a new regulation of the gene is needed to solve this problem though two of the most modified plasmids, pBEST-UTR1-Luc and pBEST-UTR1-deGFP appear to resist this toxicity problem. As a solution, the CI repressor from bacteriophage λ-switch was employed to regulate gene expression during the amplification. This is a very good regulator for two reasons. 1) The regulatory part consists of Pr, a strong σ^{70} specific promoter and OR2-OR1, strong operator sites located near the CI protein gene. Pr promoter flanked by OR2 and OR1 is even stronger than PtacI with LacO (Table 2.1). Indeed, the mutation was observed with the pBEST-OR2-OR1-Pr-UTR1-deGFP during the plasmid amplification *in vivo*. 2. *E. coli* strain KL740, in which the temperature sensitive lambda repressor CI857 which is active below 30 $^{\circ}C$ is over-expressed, can be used to amplify a plasmid in a stable manner [78]. The final modification is done with a transcription terminator. Because a RNAP can be re-used once the transcription is completed, a transcription terminator provides a workable RNAP more efficiently. Indeed, with

[4] This will be a common reporter gene in this thesis, so it is now named as deGFP.

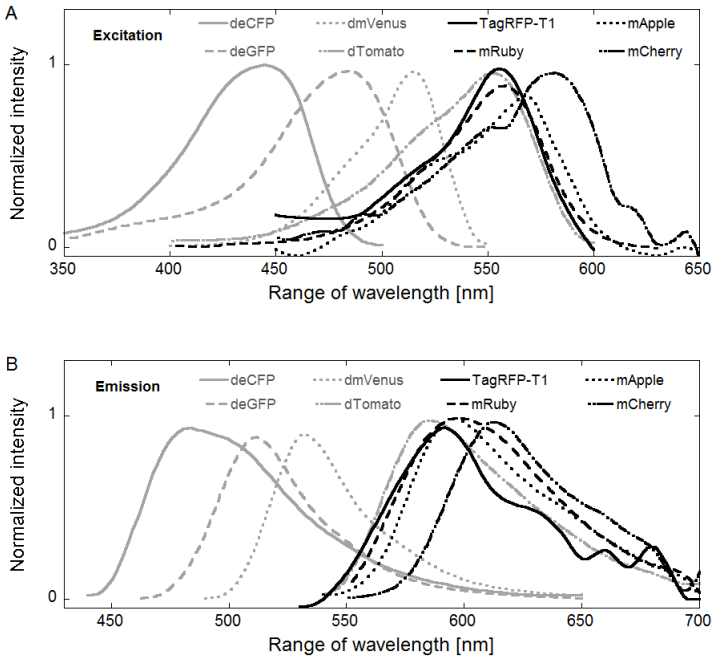

Figure 2.6: Excitation and emission spectrum of eight fluorescent reporters expressed with the cell-free system are shown.

a strong transcription terminator, T500, 20 % more protein production was observed (Table 2.1) [79]. In the end, fluorescent reporter protein production with the newly constructed plasmid, pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (P_{70}–*deGFP*, short notation in this thesis) became 200-fold higher than had been the case with the original plasmid, pBEST-eGFP (Table 2.1) and the maximum protein yield of deGFP came to ∼25 μM (∼0.65 mg/ml) at a 10∼20 nM plasmid concentration range (Figure 2.5 (B)). In addition, the toxicity can be solved with another treatment, origin of replication. All plasmids described so far have ColE1 as an origin of replication whose copy number is a few hundred. Another origin of replication used in this thesis is p15A whose copy number is about 10, therefore decreasing toxicity as required for certain genes (Chapter 3).

Despite the sole focus of this study on Luc and eGFP/deGFP to this point, our cell-free system is not specific to reporter proteins and other fluorescent reporter proteins can be synthesized in the system. Spectra of eight fluorescent reporters expressed in

the cell-free system are shown in Figure 2.6. All reporter genes, deGFP, deCFP (from eCFP with the same modification as deGFP), dmVenus (from mVenus with the same modification as deGFP), dTomato, TagRFP-T1, mRuby, mApple and mmCherry (from mCherry with the similar modification as deGFP) are cloned under the plasmid, pBEST-OR2-OR1-Pr-UTR1-(*gene*)-T500 (P_{70}–*gene*). The protein production of each reporter was different even if the same concentration of plasmid was used, so the intensity was normalized to 1. Therefore all reporters can be potentially used in the cell-free system.

2.3 Quantitative Synthetic Biology

2.3.1 Gene Expression

As pointed out in the introduction, the quantitative study of biological networks was one of the main goals here, and it was the impetus for creating this particular *E. coli* cell-free system. For the quantitative study of both *in vivo* and *in vitro*, the sink mechanism is as important as the synthesis [80]. For example, if toxin is over-expressed, this makes the gene network malfunction. In an *in vitro* system, even if there is no issue with toxicity, the presence of sink processes becomes more important because there is no dilution effect by cell division. In a sense, cell division is sort of sink system in that the number of materials are halved after the process unless it is in the stationary state. As mentioned, there are two sink processes, one for mRNA and one for protein. Synthesis and degradation of two biopolymers have to be balanced in order for the system to sustain a functional genetic network. Unlike a conventional system whose purpose is removing a sink from the system so as to produce more protein, cell-free systems for quantitative biology have to involve a sink mechanism for both biopolymers. In this chapter, mRNA inactivation and protein degradation are covered by introducing two different mechanisms. Before going forward, quantitative kinetics of gene expression will be introduced.

A fluorescent reporter has an advantage when measuring the kinetics of gene expression. Output signal measured by luminescence cannot trace the gene expression instantaneously. Luminescent measurement of Luc requires time consuming, manual effort and is done with custom-made PMT (PhotoMultiplier Tubes) with a LABVIEW program. Data is collected at 1 minute increments necessitating that to measure 12

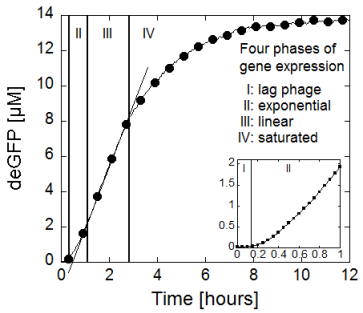

Figure 2.7: Typical kinetics of single gene expression is shown. It consists of 5-15 min lag phase, ∼1 hour exponential synthesis phase (acceleration of gene expression), ∼2 hours linear synthesis phase (gene expression reaches terminal velocity), and saturated synthesis phase. P_{70}–*deGFP* was incubated with the cell-free reaction at 29 $^{\circ}C$ (2 mM magnesium glutamate, 80 mM potassium glutamate, 1.5 mM of each amino acid, 2 % PEG8000, and 3-PGA buffer). **Inset**: a blow-up of the first hour of gene expression.

hours of kinetics, one must stay in front of the PMT with more than 200 samples (assumed every 3 min measurement) in a dark room for the full duration of the kinetic measurement. Practically, this makes it a logistical impossibility to do the kinetics measurement. Present technology and instrumentation, however, do allow for facile measurement of fluorescent signals. In this work, measurements were done with an instrument called a plate reader. The software of the instrument facilitates measurement of output signals in a large variety of configurations with only a few clicks. For instance, one can program the software such that the instrument measures the output signal while maintaining 30 seconds of shaking at 29 $^{\circ}C$ incubation temperature at 3 minute increments. Also, the output signal of synthetic circuit in a liposome can be measured with fluorescent protein on the microscope, but it wouldn't be possible with luminescence. Therefore, a fluorescent protein was used as the main reporter and Luc was used as a counterpart of the fluorescent reporter to confirm the results.

Using a fluorescent reporter with plate reader technology allows for quantitative measurement of expression. Here is presented the typical kinetics of gene expression as shown in the Figure 2.7. The plate reader (Wallac Victor III from PerkinElmer or Synergy H1 from Biotek) was programmed with the following parameters; 3 min

incremental measurement, no shaking, 29 $^{\circ}C$ incubation temperature and 12 hours total time. 10 μl of cell-free reaction sample was transferred into the well of a 384-well plate and was sealed with sticky tape to avoid evaporation. Typically, four expression phases are observed: a lag phase, an exponential protein synthesis phase, a linear protein synthesis phase and a saturation phase. Gene expression from a single gene in a single plasmid starts with a 5$\sim$15 min lag phase and includes transcription initiation, mRNA synthesis, translation initiation, protein synthesis and fluorophore maturation (phase I). Then, one hour of an exponential growth curve is observed (phase II). This is where data has to be fit with a mathematical model where the assumption is that there is no limitation of energy resources allowing for prediction or estimation of the internal mechanism of the genetic circuit and its resulting output. After this phase, two phases (phase III and IV) display saturation of working machinery, limitation of energy resources and accumulation of waste. Ultimately it was decided that the two phases would have to be distinguished in the future with more studies. Overall protein synthesis in a test tube lasts 4$\sim$6 hours and reached the maximum at around 8 hours.

2.3.2 mRNA Inactivation

mRNA is an unstable biopolymer. This instability is due to complex and different enzymatic mechanisms as well as structural loss of transcription resulting from hairpin looping, which is sequence dependent. This mRNA instability involving either process leads to the inactive form of mRNA and is therefore considered broadly as mRNA inactivation rather than simply mRNA degradation [38]. In this thesis, toxin MazF is introduced in the system to control protein synthesis by mRNA inactivation. The *E. coli* interferase MazF cleaves ACA sequences on only single strand messenger RNA but not on other types of RNAs, such as ribosomal RNA and transfer RNA [81, 82]. This feature allows for specific targeting of mRNA, providing a method to control a genetic network with no effect on other molecules or polymers. Another good reason to use MazF is its ability to interact with anti-toxin MazE. MazF-MazE coupling follows the stoichiometric ratio MazF:MazE = 2:1. Once this pair forms, MazF loses its cleaving function [81, 82]. The precise control of gene expression by MazF coupled with the sequence dependence of mRNA inactivation is shown in the Figure 2.8. mRNA inactivation by MazF is very efficient, so just a small amount of MazF can precisely control final protein production

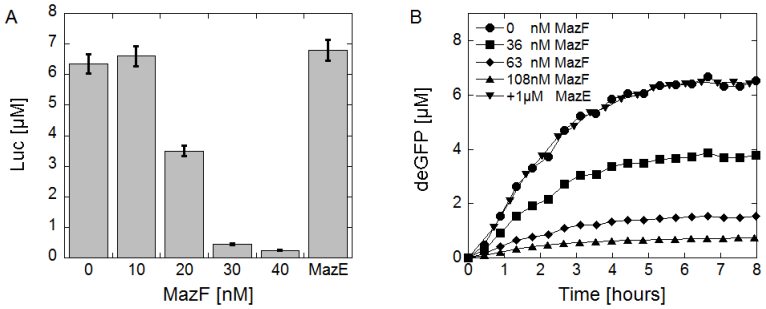

Figure 2.8: **(A)** A reporter protein, Luc was synthesized as a function of MazF concentrations added in the cell-free reaction. 5 nM of pBEST-UTR1-Luc in the cell-free reaction (0 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 1 % of PEG8000, and 3-PGA buffer) supplemented with MazF was incubated at 22 $^{\circ}C$ for 8 hours. Protein production was recovered with 1 μM of MazE in the cell-free reaction supplemented with 140 nM of MazF. **(B)** Kinetics of deGFP expression with 1 nM of plasmid, P_{70}–*deGFP*, in the cell-free reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 2 % of PEG8000, 3-PGA buffer, and 29 $^{\circ}C$ of incubation temperature) supplemented with different concentrations of MazF are shown. The kinetics and the protien production were recovered with 1 μM of MazE in the cell-free reaction supplemented with 90 nM of MazF.

that is anywhere from zero to the maximum as well as its kinetics. The MazF quantity added to control Luc protein production is different from that of deGFP. 20 nM of MazF decreases the Luc protein production by half while 36 nM of MazF is required to get half the quantity of deGFP protein production. Also, the Toxin-anti-Toxin complex allows complete recovery of maximum protein production as well as restoration of protein synthesis kinetics. In the end, mRNA inactivation by MazF can precisely control protein synthesis. With the efficiency of mRNA inactivation being sequence dependent, MazF-MazE pairing makes sink system even more subject to control.

2.3.3 Protein Degradation

Protein is another biopolymer which has to be controlled to sustain homeostasis and gene regulation in the organism. In bacteria, protein degradation is processed mainly with AAA+ proteases [83, 84]. Protein degradation tags (also named degrons) are peptides added to the C or N-terminus of a protein either by being encoded in DNA

or by being added externally due to misfolding, denaturing or incomplete polypeptide synthesis. Proteins labeled with these degradation tags are directly degraded by AAA+ proteases. Unlike mRNA inactivation controlled by adding a toxin enzyme, protein degradation control is accomplished by the fusion of degrons to the N- or C-terminus of a target protein which will then cause interaction with the ClpXP complex, an *E. coli* endogenous protease left in the cell-free extract [38]. In this thesis, seven degrons, one N-terminal tag and six C-terminal tags are fused to deGFP: OmpA (N-terminal tag), SsrA, SsrA/D (SsrA with one mutation peptide), SsrA/DD (SsrA with two mutation peptide), Crl, YbaQ and YdaM (Table 2.2) [85]. As shown in Figure 2.9 (A, B), deGFP protein with different degrons leads different production and kinetics. The end-point protein production as it is related to a specific degron is ordered as follows: no tag > SsrA/DD > SsrA/D > YdaM > Crl > OmpA > YbaQ > SsrA at 0.1 nM of plasmid and no tag > SsrA/DD > SsrA/D > YdaM > SsrA > Crl > YbaQ > OmpA at 1 nM of plasmid. The dependence of plasmid concentration on the degradation rate is also shown. For instance, the 60 % drop of deGFP expression with SsrA degron was measured with 0.1 nM of plasmid while the 40 % drop was measured with 1 nM of plasmid. Purified His-eGFP-SsrA protein was tested to estimate the degradation rate and approximately 10 nM/min of constant rate is observed up to 1 μM (Figure 2.9 (C)). The degradation rate was, however, decreased down to 4 nM/min once 10 μM His-eGFP-SsrA was tested (data not shown, see my publication [38]).

At this point, one simple question arises: Can we quantitatively predict the result with the quantified data obtained above for mRNA inactivation and protein degradation

Degrons	Sequences
SsrA	GCAGCAAACGACGAAAACTACGCTTTAGCTGCT
SsrA/D	GCAGCAAACGACGAAAACTACGCTTTAGATGCT
SsrA/DD	GCAGCAAACGACGAAAACTACGCTTTAGATGAC
Crl	TTTCGTGATGAACCTGTTAAACTTACCGCC
YbaQ	AGAAGGGAAGAAAGAGCAAAGAAGGTAGCA
YdaM	TGCAAGAATGATGGAAGAAATAGGGTACTAGCAGCA
OmpA	AAAAAAACTGCTGCTATCGCGATCGCGGTCATG

Table 2.2: Sequence of seven degrons are shown.

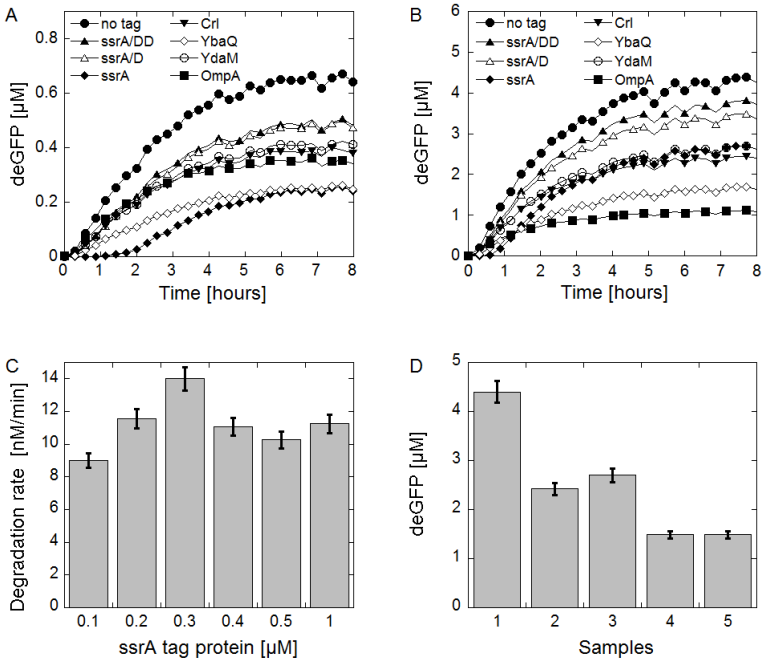

Figure 2.9: **(A)** Kinetics of deGFP expression with 0.1 nM of plasmids, P_{70}–*deGFP-Tag*, in the cell-free reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 2 % of PEG8000, 3-PGA buffer, and 29 $^{\circ}C$ of incubation temperature) are shown. The reporter protein, deGFPs were fused with seven degrons recognized by AAA+ proteases. **(B)** Kinetics of deGFP expression with 1 nM of plasmid, P_{70}–*deGFP-Tag*, in the cell-free reaction (same conditions as in the panel (A)) are shown. **(C)** Degradation rate of purified His-eGFP-ssrA protein at different concentration in cell-free reaction (same condition as in the panel (A)) are shown. ∼10 nM/min of constant degradation rate up to 1 μM protein was observed. **(D)** End-point measurement of deGFP expression is shown. The concurrent processes with two sink mechanisms for mRNA and protein were tested and predicted. 1 nM of plasmid, P_{70}–*deGFP* or P_{70}–*deGFP-ssrA*, were incubated in the cell-free reaction (2 mM of magnesium glutamate, 80 mM potassium glutamate, 1.5 mM of each amino acid, 2 % of PEG8000 and 3-PGA buffer) at 29 $^{\circ}C$ for 8 hours. 1) no degradation (0 nM of MazF, deGFP), 2) mRNA inactivation (36 nM of MazF, deGFP), 3) protein degradation (0 nM of MazF, deGFP-ssrA), 4) expected protein production with mRNA inactivation (36 nM of MazF) and protein degradation (deGFP-ssrA), 5) protein production measured with mRNA inactivation (36 nM of MazF) and protein degradation (deGFP-ssrA).

once both sink mechanisms concurrently exist in cell-free system? This answer will be the first step in telling us whether this cell-free system can be employed in quantitative research. The prediction of deGFP expression under co-existing sink processes behaves as multiplication of the two effects because these two different mechanisms are assumed to be independent; in the case that 80 % of deGFP gene expression with MazF added and 60 % of deGFP gene expression with degradation tags was separately observed, deGFP protein production is supposed to be 48 % (= 80 % ×60 %) compared to deGFP gene expression without any sink mechanism. In separate tests, it was observed that 36 nM of MazF gives 50 % of deGFP expression while the SsrA tag shows 60 % of deGFP expression, with each case having 1 nM of plasmid (Figure 2.8 (B), 2.9 (B)). Therefore the prediction of deGFP production fused to SsrA tag in the cell-free reaction supplemented with 36 nM of MazF is estimated to 30 % (= 50 % ×60 %). The prediction was in 100 % agreement to the measurement within the error bar (Figure 2.9 (D)). This result promises the capacity of our cell-free system toward use in quantitative research.

2.4 Coarse-Grained Model of Protein Synthesis

2.4.1 Single Gene Expression

It has just been shown that the cell-free system can be the experimental platform to carry out the quantitative study of protein synthesis containing sink mechanisms (mRNA inactivation and protein degradation) as well as source processes (transcription and translation). To understand the physics of gene expression, it is necessary to describe the information flow and its structure quantitatively. The source-sink nonlinear dynamics with feedback processes explaining reaction mechanisms, i.e. pattern formation, is however as complex as containing more than 100 biomolecules in the simplest case. Moreover, even though some delicate methods have been developed, *in vivo* measurement of gene expression involving many unknown parameters makes the quantitative approach harder because other indirect reaction mechanisms, i.e. cell division, can provide even more parameters. The advantage of an *in vitro* approach for quantitative study is that it enables the research to focus on a single event or single reaction line.

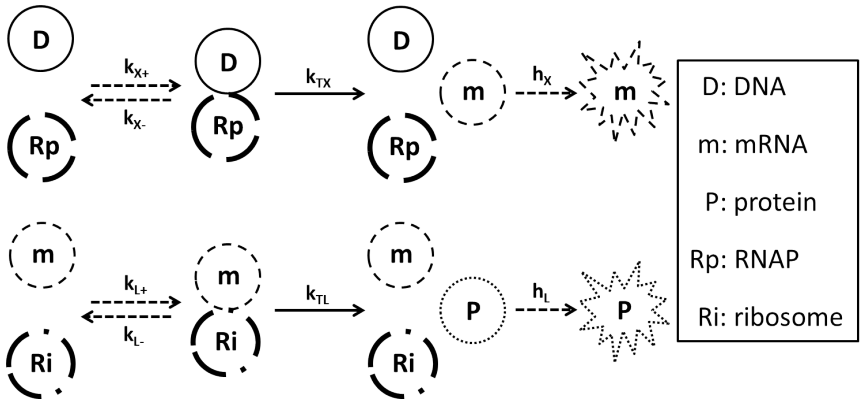

Figure 2.10: Biomolecules and their enzymatic reactions are represented as a ball and particle collision with the retarded time, respectively. k, h, X, L, +, -, TX and TL stand for the reaction rate, degradation rate, related to transcription, related to translation, association, dissociation, transcription and translation, respectively.

Despite the advantage of an *in vitro* approach, there are still more than 100 biomolecules and their biochemical reactions. The coarse-grained model uses far less parameters to describe the whole process of gene expression by means of defining one kinetic parameter implying a couple of or even tens of parameters of internal biochemical reactions (Figure 2.10). For instance, a single parameter k_{X+} which is defined as association rate of RNAP to DNA contains

- the interaction of core RNAP and transcriptional initiation sigma factor to form RNAP holoenzyme,
- the interaction of RNAP holoenzyme and DNA,
- the search of promoter on DNA by RNAP holoenzyme by sliding, hopping, and jumping on DNA,
- making readable single strand DNA by opening a piece of double strand DNA,
- detaching sigma factor from core RNAP for mRNA synthesis,

just to name a few main steps. By this way, the coarse-grained model mainly includes six biomelecules with eight reaction parameters. The thing that makes this model even better is that it can be adopted to the mathematical model of gene expression *in*

vivo [5, 6, 14]. According to the enzyme kinetics; that is, the chemical reaction catalyzed by enzymes, the following equations can be constructed for the simplest coarse-grained model.

$$\frac{d[mRNA]}{dt} = A - \beta[mRNA], \tag{2.1}$$

$$\frac{d[deGFP_d]}{dt} = \alpha[mRNA] - \kappa[deGFP_d] - \gamma[deGFP_d], \tag{2.2}$$

$$\frac{d[deGFP_f]}{dt} = \kappa[deGFP_d] - \gamma[deGFP_f], \tag{2.3}$$

where A is for mRNA synthesis from input DNA corresponding to input DNA concentration, RNAP concentration and k_{X+}, k_{X-}, k_{TX}; β is for mRNA inactivation rate corresponding to h_X; α is for protein synthesis rate corresponding to ribosome concentration and k_{L+}, k_{L-}, k_{TL}; γ is for protein degradation rate corresponding to h_L; κ^{-1} is for fluorescent protein maturation rate; $[mRNA]$ is for mRNA concentration; $[deGFP_d]$ is for deGFP protein which is folded but not maturated; and $[deGFP_f]$ is for fluorescent deGFP protein. It is assumed that energy resources and working machinery are not limited in this model. This simple form of the equation set can be modified according to the conditions. For instance, Michaelis-Menten kinetics will be explained more in the chapter 4. Therefore this equation set is the basic format to explain gene expression by the coarse-grained model. Two examples of quantitative studies will be shown in the following sections, 2.4.2 and 2.4.3.

2.4.2 deGFP Maturation

During translation, the amino acids join together successively to form a polypeptide chain. These macromolecules have to be folded to form a functional three-dimensional structure, which depends on many parameters such as the amino acids sequence, the solvent, the concentration of ions, the temperature and the molecular crowding [67, 86, 87, 88]. Some proteins even need a chaperone protein to be present at the location where the protein is supposed to be folded [89]. This protein folding process, occurring either during or after translation, physically changes its conformation. Some proteins, however, need a chemical structure change after folding process. The fluorescent reporter protein, GFP family is one of this kind. Even if the random coil of a fluorescent protein is folded correctly, it doesn't have functional form yet. Three more chemical reaction

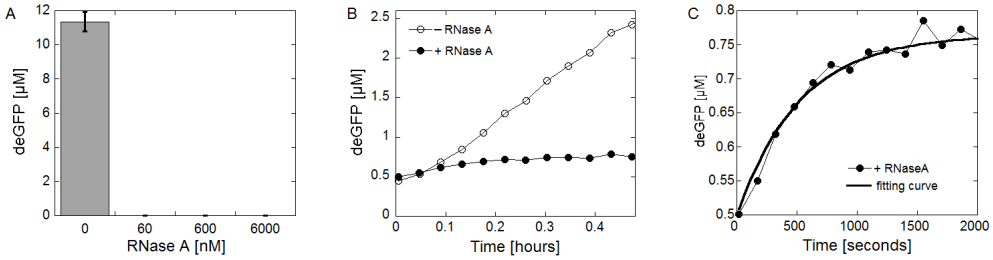

Figure 2.11: **(A)** End-point deGFP expressions with RNase A concentrations are shown. 5 nM of P_{70}–*deGFP* was incubated in the cell-free reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 2 % of PEG8000, 3-PGA buffer, and 29 $°C$ of incubation temperature) supplemented with RNase A are shown. **(B)** The kinetics of deGFP expression in the cell-free reaction (same condition as in the panel (A)) after the addition of RNase A are shown. The cell-free reaction without RNase A was incubated at room temerature for 30 min, and either 0 nM or 600 nM of RNase A was added in the cell-free reaction, and then the kinetics of gene expression was observed at 29 $°C$ of incubation temperature. **(C)** A blow-up of the kinetics for the first 30 min in the panel (B) is shown. The coarse-grained model was used to estimate the maturation time of deGFP protein.

steps called cyclisation, oxidation and dehydration are required for fluorophore structure and this mechanism takes time referred to as maturation time [73, 90]. In this chapter, maturation time of deGFP is estimated by fitting data using a coarse-grained model.

In experiment, deGFP protein is expressed in the cell-free reaction at room temperature for 30 min, and ribonuclease A, RNase A is added, then the fluorescent signal is measured at 29 $°C$ (Figure 2.11). This powerful enzyme degrades all single strand RNA, including mRNA and rRNA very quickly. So, once this enzyme is added into the system, all TX-TL processes are quickly halted: 600 nM of RNase A stops gene expression completely (Figure 2.11 (A)). For example, once Luc was used for the maturation test, data with RNase A showed no signal increasing due to protein maturation even 30 sec later which is minimum time lapse of luminescent measurement (data not shown, see my publication [38]). For the mathematical model, an even simpler equation than equation set (Equation 2.1, 2.2 and 2.3) is constructed because no equation is needed for the synthesis and the degradation of mRNA and protein. So only the last equation

without a degradation term is used to fit the data,

$$\frac{d[deGFP_f]}{dt} = \kappa[deGFP_d]. \tag{2.4}$$

This equation yields

$$[deGFP_f](t) = [deGFP_0](1 - e^{-\kappa t}) \tag{2.5}$$

where $[deGFP_0] = [deGFP_f] + [deGFP_d]$ is total initial protein. The parameter $1/\kappa$ extracted by fitting data with the equation 2.5 directly gives the maturation time of deGFP protein. From the Figure 2.11 (B, C), deGFP maturation time in cell-free reaction was estimated to be 8∼8.5 min. The same experiment with 1 hour incubation before fluorescent signal measurement yielded the same maturation time (data not shown). This result is compatible with the results of other groups [91, 92].

2.4.3 mRNA Lifetime

Among major biomacromolecules, the most unstable one is mRNA, the information carrier existing between the information source and the product of the information. The single stranded genetic material is inactivated by either enzymatic mechanism or structural loss of functionality. Certain enzymes cleave specific sequences, such as the toxin MazF which cleaves at only the ACA site of mRNA [81, 82, 93, 94]. The 3-dimensional conformation depends on the molecule's tendency to position itself to have minimal free energy, an aspect which is sequence dependent. Thus the inactivation of mRNA depends on the internal sequence and the size of mRNA, and the external concentration of ions, the temperature and the molecular crowding, to name a few of the conditions. *E. coli* whose genome consists of more than 4,000 protein-coding genes includes a lot of different mRNA in terms of size and sequence even in the healthy steady state. Accordingly, mRNA half-life, also called lifetime, is variable. The average global lifetime of total mRNA in *E. coli* was estimated to be ∼6 min [95]. Even though *E. coli* cell-free extract is supposed to include almost all cytoplasmic materials, global mRNA lifetime in the cell-free system is expected to be longer than ∼6 min since protein concentration in the cell-free system is approximately 20∼30-fold diluted. In order to estimate the global mRNA lifetime in cell-free reaction, deGFP gene whose size is 678 nt and 25.39 kDa is used and data is fit using the equation set from the coarse-grained model.

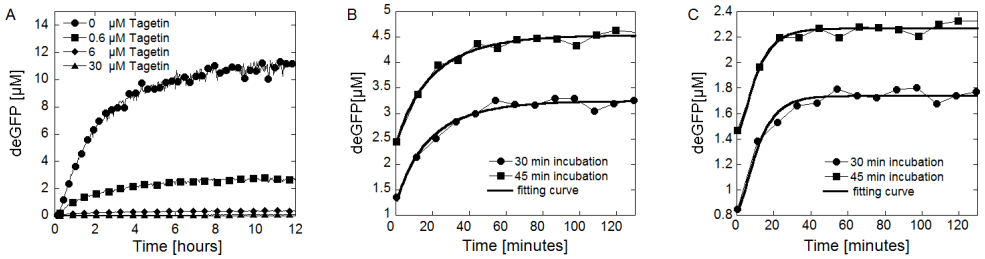

Figure 2.12: **(A)** Kinetics of deGFP expressoin with 5nM of plasmid, P_{70}–*deGFP* incubated in the cell-free reaction (2 mM of magnesium glutamate, 80 mM of potassium glutamate, 1.5 mM of each amino acid, 2 % of PEG8000, 3-PGA buffer, and 29 $°C$ of incubation temperature) supplemented with the different concentrations of Tagetin are shown. **(B)** Kinetics of deGFP expression with 5nM of plasmid, P_{70}–*deGFP* after addition of Tagetin are shown. 30 μM of Tagetin was added to the cell-free reaction (same condition as in the panel (A)) after 30 or 45 min of incubation at room temperature. The coarse-grained model was used to estimate the endogenous mRNA lifetime. **(C)** Kinetics of deGFP expression with 5 nM of plasmid, P_{70}–*deGFP* in cell-free reaction supplemented with 36 nM of MazF after addition of Tagetin are shown. 30 μM of Tagetin was added to the cell-free reaction (same condition as in the panel (A) but supplemented with 36 nM of MazF) after 30 or 45 min of incubation at room temperature. The coarse-grained model was used to estimate the mRNA lifetime in the cell-free reaction with the acceleration process of mRNA turnover.

For this test, core *E. coli* RNAP inhibitor, tagetin was used to stop transcription (Figure 2.12 (A)) [96]. The cell-free reaction is incubated at room temperature for 30 min and 45 min, and the kinetics are measured at 29 $°C$ after adding 30 μM of tagetin. Once tagetin is added in the cell-free system, there is no more mRNA synthesis, so the output signal is only from mRNA present at the moment the tagetin was added and from maturation of deGFP protein folding. The equation set is therefore simplified;

$$\frac{d[mRNA]}{dt} = -\beta[mRNA], \tag{2.6}$$

$$\frac{d[deGFP_d]}{dt} = \alpha[mRNA] - \kappa[deGFP_d], \tag{2.7}$$

$$\frac{d[deGFP_f]}{dt} = \kappa[deGFP_d]. \tag{2.8}$$

This equation set yields

$$[deGFP_f](t) = \frac{\alpha m_0}{\beta(\beta-\kappa)}(\kappa(e^{-\beta t}-1) + \beta(1-e^{-\kappa t})), \tag{2.9}$$

where m_0 is the concentration of active mRNAs at the moment of tagetin being added. The parameter $1/\beta$ extracted by fitting data with the equation 2.9 directly gives the lifetime of messenger RNA from the *deGFP* gene. Parameter α independence was observed while κ was manually given at 8 min, then $\sim$13 min endogenous global mRNA lifetime was estimated (Figure 2.12 (B)) [97]. This is two times the lifetime than found *in vivo*. To control mRNA lifetime down to the *in vivo* level, the MazF enzyme is added. Due to the doubled mRNA lifetime, it was rationalized that 36 nM of MazF, which leads to half the protein production, would work for decreasing mRNA lifetime down by one half. MazF was added to the cell-free reaction with the plasmid at the beginning and the cell-free reaction was incubated at room temperature for 30 min and 45 min, then 30 μM of tagetin was added to stop transcription. The output data was fit by equation 2.9 with same maturation time of deGFP and it gave a 6 minute global mRNA lifetime which is comparable to one *in vivo* (Figure 2.12 (C)) [95].

Chapter 3

Cell-Free Toolbox[1]

3.1 Introduction

3.1.1 Limitations of Conventional Hybrid Cell-Free Expression System

TX-TL cell-free system has been considered a black box because there were many uncharacterized properties of the system. The system uses the crude extract obtained from an engineered *E. coli* strain, and the hybrid cell-free system contains the powerful transcriptional machinery such as T7 RNAP manually added in crude extract and its specific promoter on recombinant plasmid. Now, conventional cell-free system is widely used for many applications such as proteomics and biotechnology [46]. Even though the 'black box' concept was chipped away by many efforts to improve the performance, a further improvement of cell-free system was required for synthetic gene circuit construction; especially development of sink processes for mRNA and protein and expanding transcriptional units. In the previous chapter, it was shown that sink mechanisms involving acceleration of mRNA turnover and degradation of protein, which are necessary for quantitative study, could be introduced into a cell-free system, as developed in this thesis. In addition, the protein production with this cell-free system is fortunately comparable with the conventional one even though the transcription with endogenous machinery is not as efficient and powerful as one with T7 RNAP [37].

[1] **I acknowledge that Jonathan Gapp's editing help was essential to achieve proper English usage and readability of this chapter.**

In order to construct a synthetic gene circuit and have it perform in a cell-free system, one more problem is still left: the lack of transcriptional repertoire. Unlike an electric circuit, biological gene circuit is not allowed to use same motif multiple times in one reaction. It was shown that synthetic gene circuit, pattern formation and artificial cell system was restricted when using the conventional hybrid cell-free system [63, 64, 65, 66]. Even though the cell-free system described in this thesis, which is using endogenous TX-TL machinery containing σ^{70} as a primary transcription initiation sigma factor, expanded the list of transcription units; still only σ^{70} and the regulatory part with σ^{70} specific promoter were used. *E. coli* has a total of seven sigma factors: σ^{19}, σ^{24}, σ^{28}, σ^{32}, σ^{38}, σ^{54} and σ^{70} whose number represents the size of each sigma factor in the unit of kDa. Except for the housekeeping sigma factor σ^{70} always existing in the cell, other sigma factors are expressed in certain specific condition. For example, once *E. coli* faces starvation, RNA holoenzyme with σ^{70} transcribes gene σ^{28} and RNAP holoenzyme with σ^{28} expressed, transcribes all genes which are required to build flagella in order to move toward a rich nutrient area. By mimicking this approach such as an activating cascade, all sigma factors can be used in the cell-free system. Therefore here is offered a new concept of cell-free system as not a 'blackbox' but a 'toolbox' because now we have tools to construct and perform synthetic gene circuit with cell-free system.

3.1.2 Bottom-Up Approach to Artificial Cell

Protein synthesis with cell-free system lasts 4∼6 hours in a batch mode (Figure 2.7). The cell-free system as a reservoir of non-limiting energy resources is preserved for the first 1∼1.5 hours and it was reported that the machinery from cell-free crude extract was able to work for days [98]. These results indicate that the gene expression from DNA within the cell-free system can be in the exponential and linear synthesis phase for hours if infinite energy resources are available. Furthermore, these phases can be extended even longer if the byproduct of the biochemical reactions are diluted. A couple of large-scale bioreactors for the long-lived biochemical reaction were built for high-throughput of proteins [99]. The long-lived biochemical reaction is important for study and for performing synthetic gene circuits, both quantitatively and qualitatively, because more complex circuits require more energy resources and produce more byproduct.

Cell-free reaction can also be encapsulated in synthetic phospholipid vesicle (liposome) [65, 66]. This plasma membrane not only separates the interior of liposome from outside but also acts as an intermediary to exchange materials from one side to the other. In this sense, gene expression with cell-free system in a liposome can last longer and this bioreactor can be considered as a microscopic long-lived bioreactor whose size is in the range of a few μm to 50 μm in diameter. The cell-sized, long-lived bioreactor adopts the concept of a primitive artificial cell [100]. This artificial cell system becomes a platform for a bottom-up approach to research informational and self-organizational properties of living matter from scratch [65, 100, 101]. Instead of the top-down approach for redesigning an organism by re-engineering its genome, the bottom-up approach starts with completely synthetic information and materials and can tell more about synthetic life even if it is not the expected results [102].

3.2 Genetic Circuits

3.2.1 Repertoire of Transcriptional Regulatory Elements

The translation process in bacteria is universal which means that only one translational machine synthesizes protein from any kind of messenger RNA in its host. *E. coli* TX-TL cell-free crude extract is directly obtained from *E. coli*, so only one translation process exists in cell-free system. Transcription, however, is more variable than translation because synthesizing messenger RNA of operons on *E. coli* genome is distinguished by seven transcription initiation factors called sigma factors. Except housekeeping sigma factor 70, σ^{70} which always exists in a cell, all other sigma factors are mostly silent until the cell faces certain conditions. There are a couple of regulations of sigma factors' activation such as anti-sigma factors, but the main explanation is as follows. Once the cell meets specific circumstance such as starvation and heat shock, certain processes make the operon encoding the specific sigma factor gene open, then that sigma factor is expressed, after a housekeeping sigma factor transcribes a sigma factor gene. It, in turn, starts to transcribe genes which are necessary to sustain life against the environment. The following is the main reasons of each sigma factors expressed:

- σ^{19} (*fecI*): the ferric citrate sigma factor, regulates the fec gene for iron transport,

- σ^{24} (*rpoE*): the extracytoplasmic/extreme heat stress sigma factor,
- σ^{28} (*rpoF*): the flagellar sigma factor,
- σ^{32} (*rpoH*): the heat shock sigma factor,
- σ^{38} (*rpoS*): the starvation/stationary phase sigma factor,
- σ^{54} (*rpoN*): the nitrogen-limitation sigma factor.
  (from wikipedia)

Despite this variety of transcription units, the construction of transcription/translation circuits is limited by the transcription repertoire of modern cell-free systems, which are composed of only a few bacteriaphage RNA polymerases and promoters. *E. coli* endogenous cell-free expression system, based on the housekeeping transcription machinery, described in the previous chapter, is used to develop a repertoire of alternative transcriptional activation units. We tested the main transcription space of *E. coli*, composed of six sigma factors in addition to the primary σ^{70}. Here, a transcriptional activation unit as a transcriptional activation protein and its specific promoter is defined. A transcriptional activation cascade is the serial assembly of two or more transcriptional activation units.

A two-stage transcriptional activation cascade was constructed for each of the *E. coli* σ^{19}, σ^{24}, σ^{28}, σ^{32}, and σ^{38} transcription factors (Figure 3.1 (A - G)). The σ^{54} unit, which requires the gene *ntrC* for the transcriptional co-activation of genes, was studied separately (Chapter 3.2.4). Each transcription factor was cloned under the same promoter P_{70}. To create a set of degradable factors, the five sigma factors were also tagged with AAA+ specific proteolytic degrons. σ^{19}, σ^{28}, and σ^{32} were tagged with the C-terminal SsrA degron. σ^{24} and σ^{38}, not active with the SsrA degron (data not shown), were tagged with the N-terminal OmpA degron. The reporter gene *deGFP* was cloned under a promoter specific to each sigma factor. We chose strong promoters previously described in the literature [103, 104, 105, 106, 107, 108]; P_{19}: promoter of the gene *fec*, specific to σ^{19}, P_{24}: promoter of the gene *htrA*, specific to σ^{24}, P_{28}: promoter of the gene *tar*, specific to σ^{28}, P_{32}: promoter of the gene *groE*, specific to σ^{32} and P_{38}: promoter of the gene *osmY*, specific to σ^{38}. For comparison, we also constructed a T7 and a T3 transcriptional activation units (Figure 3.1 (H, I)). Degradable versions of these two

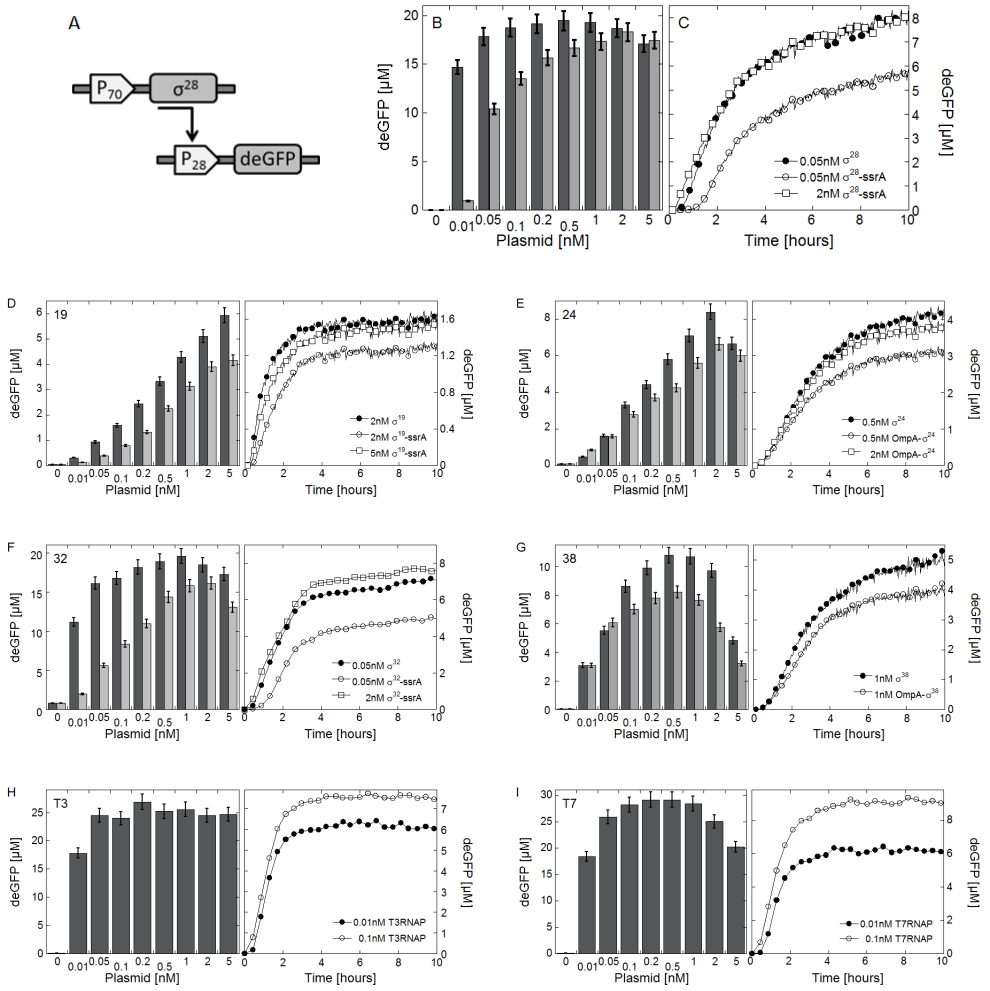

Figure 3.1: Seven two-stage transcriptional activation cascades are shown. **(A)** Schematic diagram of the two-stage transcriptional activation cascade using σ^{28} is shown. σ^{28} cloned under a promoter P_{70}, specific to σ^{70} (P_{70}–σ^{28}), activates the expression of gene *deGFP* cloned under a promoter P_{28}, specific to the σ^{28} (P_{28}–*deGFP*). **(B)** End-point deGFP expression as a function of the concentration of plasmid, P_{70}–σ^{28} are shown (dark gray bar: σ^{28}, light gray bar: $\sigma^{28} - SsrA$). The concentration of plasmid, P_{28}–*deGFP* was fixed to 5 nM and the cell-free reaction was incubated under the σ^{28} condition in TABLE 3.1. **(C)** Kinetics of deGFP expression with 5 nM of plasmid, P_{28}–*deGFP* are shown. The experiments were done under the σ^{28} condition in TABLE 3.1. **(D - I)** The same characterizations were performed for σ^{19}, σ^{24}, σ^{32}, σ^{38}, T3 RNAP and T7 RNAP.

units with AAA+ specific degrons were not tested. The transcriptional activation units were first validated as usable units on the basis of the maximum deGFP production in batch mode reaction, also defined as the output signal of the cascade. The requirement for the magnitude of the output signal was set to 500 nM, the average cytoplasmic protein abundance in *E. coli* cells [109]. This concentration was used as the relevant threshold output signal for all of the circuits constructed in this thesis.

The magnesium and the potassium glutamate concentrations to get the maximum deGFP production for each cascade were determined, before determining the optimum plasmid concentrations in each stage (Table 3.1). The magnesium and the potassium glutamate concentrations span a range of 4 mM and 80 mM, respectively, σ^{19} and σ^{38} being the most distant. These rather small differences, within what has been measured in *in vitro* transcription assays, are expected since promoter selectivity is based on cellular conditions [110]. We found that all of the activation units have an output signal largely above the reference concentration of 500 nM.

As mentioned before, plasmid concentration is crucial parameter affecting the gene expression in cell-free system. In the two-stage cascade, the two parameters are coupled; transcriptional activator plasmid (a plasmid encoding the transcriptional activator protein such as sigma factors and bacteriophage RNA polymerases) and reporter plasmid (a plasmid encoding the reporter protein such as deGFP). So, the experiments were done the following way: a reporter plasmid was fixed at 5 nM and the concentration range of a transcriptional activator plasmid was performed from 0 nM to 5 nM (Figure 3.1). Then the concentration of the transcriptional activator plasmid was fixed at that concentration which gave the maximal protein production and the concentration range of the reporter plasmid was performed from 0 nM to 20 nM. Finally, the concentration of reporter plasmid was fixed where it gave the maximal protein production and the concentration range of the transcriptional activator plasmid was performed from 0 nM to 5 nM. Table 3.1 shows the different optimal concentrations of transcriptional activator plasmid and reporter plasmid for each sigma factor. Degradable sigma factors were also compared with non-degradable sigma factors. To observe the degradation efficiency of each sigma factor with either SsrA or OmpA, the same conditions for magnesium glutamate and potassium glutamate were used.

transcription factor	Mg-glutamate [mM]	K-glutamate [mM]	plasmid A [nM]	plasmid B [nM]	deGFP [μM]	deGFP (mg/ml)
σ^{70}	3	60	NA	10	25	(0.63)
σ^{19}	3	20	2	10	7	(0.18)
σ^{19}-*SsrA*	3	20	5	10	7	(0.18)
σ^{24}	6	30	2	10	11	(0.28)
OmpA-σ^{24}	6	30	2	10	8	(0.20)
σ^{28}	3	60	0.2	5	21	(0.53)
σ^{28}-*SsrA*	3	60	2	5	18	(0.46)
σ^{32}	5	70	0.5	5	19	(0.48)
σ^{32}-*SsrA*	5	70	2	5	14	(0.36)
σ^{38}	5	100	0.5	15	13	(0.33)
OmpA-σ^{38}	5	100	0.5	15	10	(0.25)
σ^{54}/*NtrC*	5	30	1 (each)	10	5	(0.13)
T3 RNAP	2	80	0.2	5	27	(0.69)
T7 RNAP	2	80	0.2	5	29	(0.74)

plasmid A: plasmid encoding the transcription factor gene
plasmid B: plasmid encoding the reporter gene (reporter plasmid)

Table 3.1: Optimal conditions of magnesium glutamate, potassium glutamate, and plasmids concentration for the maximal protein production are shown. Except for endogenous σ^{70}, deGFP was synthesized through a two-stage cascade for other transcriptional activation units (except σ^{54}) shown in Figure 3.1 and for σ^{54} and NtrC shown in Figure 3.3.

In *E. coli*, the sigma factor family competes for the same core RNA polymerase. In our experiments, the change in the magnitude of the output signal for each unit with respect to the concentration of plasmid encoding the alternative sigma factor is an indirect measurement of the competition between each alternative sigma factor and the primary one. The change in the magnitude of the output signal depends on the relative binding affinity for each sigma factor for the core RNA polymerase. The magnitude of the output signal depends also on the strength of the promoter. Two trends are observed. For the σ^{28} and σ^{32}, the output signal is high even at low plasmid concentration. In the case of σ^{28}, the output signal only increases by 25 % when the plasmid concentration is increased by a factor of 100 in the linear regime of plasmid concentrations (0.01 nM to 1 nM). For the σ^{19}, σ^{24}, and σ^{38} units, a continuous increase of the output signal is observed when the plasmid concentration is increased in the linear regime of plasmid concentration. These observations globally agree with the relative binding affinities of

the seven *E. coli* sigma factors for the core RNA polymerase [111]. As expected, the strong σ^{28} and σ^{32} take over the core RNA polymerase even at low plasmid concentration. The transmission of information through these two units is highly efficient compared to the σ^{19}, σ^{24}, and σ^{38} units. The switch-like input-output function of the σ^{28} and σ^{32} units may be a problem for circuit constructions as one may want a linear easy-adjustable output signal. The input-output function of these two units could be linearized by adjusting the strength of the different regulatory parts (promoters and untranslated regions) and/or by using degradable sigma factors. A net decrease of the output signal is observed for the σ^{38} and T7 units at high plasmid concentration, for which no explanation can be provided. The sharp response of the T7 and the T3 units to the change of concentration of the plasmid encoding the RNA polymerase is attributed to the high efficiency and specificity of bacteriophage transcriptions.

As expected, the reporter protein production at low concentration of plasmid encoding σ^{28}, a regime where the activity of the AAA+ proteases is not saturated, is much higher than the degradable version $\sigma^{28} - SsrA$. This trend, also observed for $\sigma^{19} - SsrA$ and $\sigma^{38} - SsrA$, is not observed for $OmpA - \sigma^{24}$ and $OmpA - \sigma^{38}$, presumably due to the weakness or the inaccessibility of the OmpA tag. As opposed to the OmpA tag, the degradation with the SsrA degron is enhanced by the specific SspB factor present in the extract, which may also explain the difference observed between the two degrons [112]. The Michaelis constant of the SspB-SsrA degradation pathway is inferior to 10 nM in our cell-free system. At high concentration of plasmid encoding sigma factors, the production of reporter protein with the non-degradable and the degradable versions of sigma factors is similar. In this regime of plasmid concentrations, the rate of protein degradation by the AAA+ proteases is negligible compared to the rate of protein synthesis. This trend is observed for all the sigma factors.

The kinetics of deGFP expression for each unit is similar to the typical protein synthesis dynamics in cell-free systems. An accumulation of reporter protein is observed for a few hours before the reaction stops. A 5 to 10 min delay of expression is observed for each transcriptional activation cascade compared to the expression of deGFP from a promoter P_{70}, a typical time for such two-stage cascades. This delay is much larger for the degradable version of the σ^{28} and σ^{32} transcriptional activation units at low plasmid concentration due to the high efficiency of proteolysis with the SsrA tag. The

kinetics of deGFP expression through the σ^{19} unit is two times shorter, which could be due to an increased instability of this sigma factor [113].

3.2.2 Crosstalk among Activation Units

A quantitative estimation of non-specific signals generated in elementary gene circuits is essential to construct complex informational systems with predictable behaviors. We next studied the crosstalks between the elements (sigma factors and promoters) of all the units. To determine the complete crosstalk space of the activation unit set (σ^{19}, σ^{24}, σ^{28}, σ^{32}, σ^{38}, σ^{70}, T3 and T7), we measured in individual assays the leak of deGFP expression through each promoter against the other sigma factors and the two bacteriophage RNA polymerases. The specific and non-specific end-point production of active deGFP and the maximum rate of deGFP synthesis (data not shown, see my publication [42]) were first measured and scaled in the linear regime of plasmid concentrations (Table 3.2). The generation of undesirable signals due to the presence of alternative endogenous sigma factors in the extract was also examined. All supplementary data can be found in my publication [42].

The non-specific gene expression generated through the promoter P_{70} by the alternative sigma factors could not be measured since the σ^{70} is present in the reaction. The decrease of gene expression from the promoter P_{70} when one alternative sigma factor is expressed in the reaction agrees with the relative binding affinities of each sigma factor for the core enzyme measured *in vitro* [111], P_{28} and P_{32} being the most competitive (Table 3.2). Protein synthesis through the promoter P_{70} is not sensitive to the expression of the T3 or the T7 RNA polymerases, which do not compete for the core RNA polymerase.

Several features emerge from the crosstalk table (Table 3.2). First, σ^{32} is the most leaky unit. The non-specific expression induced by σ^{32} through the other promoters (P_{70} excluded) is rather small, less than 1 % in the worst case. The non-specific expression through the promoter P_{32}, however, is large even in the presence of the primary sigma factor only. This could be due either to the leak of the primary sigma factor on the promoter P_{32} or to the presence of endogenous σ^{32} in the extract. Based on the intracellular concentration of σ^{32} (< 10 nM) and taking into consideration the dilution factor during extract preparation (20 to 30 times dilution), we hypothesize that the

deGFP[μM]		Transcription factor expressed							
		$(\sigma^{70})^*$	σ^{19}	σ^{24}	σ^{28}	σ^{32}	σ^{38}	*T3 RNAP*	*T7 RNAP*
Promoter	P_{70}	1	0.92	0.85	0.62	0.74	0.94	0.95	0.90
	P_{19}	<0.01	1	0.01	0.02	<0.01	0.01	<0.01	<0.01
	P_{24}	0.01	0.04	1	<0.01	0.01	0.04	0.01	0.01
	P_{28}	<0.01	<0.01	<0.01	1	<0.01	<0.01	<0.01	<0.01
	P_{32}	0.04	0.04	0.04	0.02	1	0.04	0.04	0.04
	P_{38}	<0.01	0.16	0.04	0.01	<0.01	1	0.01	0.01
	P_{T3}	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1	<0.01
	P_{T7}	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1

deGFP[μM]		Transcription factor expressed							
		$(\sigma^{70})^*$	σ^{19}	σ^{24}	σ^{28}	σ^{32}	σ^{38}	*T3 RNAP*	*T7 RNAP*
Promoter	P_{70}	1	0.94	0.83	0.41	0.43	0.89	1	0.87
	P_{19}	0.07	1	0.1	0.07	0.08	0.09	0.07	0.08
	P_{24}	0.05	0.06	1	0.05	0.05	0.06	0.05	0.06
	P_{28}	0.01	0.02	0.02	1	0.01	0.02	0.01	0.01
	P_{32}	0.05	0.04	0.05	0.03	1	0.05	0.06	0.05
	P_{38}	0.06	0.15	0.1	0.06	0.07	1	0.07	0.07
	P_{T3}	0.02	0.02	0.03	0.02	0.03	0.03	1	0.02
	P_{T7}	0.01	0.02	0.03	0.02	0.02	0.02	0.02	1

Table 3.2: Crosstalk of a sigma factor to non-specific promoters was measured and scaled in the linear regime (0.1 nM of plasmid encoding the sigma factor and 1 nM of reporter plasmid, top table) and the saturated regime (0.5 nM of plasmid encoding the sigma factor and 5 nM of reporter plasmid, bottom table) of plasmid concentration. Scaled values of the end-point deGFP productions for six *E. coli* sigma factors, and for the T3 and T7 RNAP, were measured with respect to each of the specific promoters (shaded in gray) and against each of the other non-specific promoter. These tables were obtained by dividing each row by the specific value of each unit. Except endogenous σ^{70} present in cell-free crude extract (* in the table), transcription factors were expressed as shown in the Figure 3.1. The cell-free reaction was incubated at σ^{70} condition (Table 3.1) for 8 hours. Expression through the promoter P_{70} is high in all the cases because the endogenous σ^{70} present in cell-free crude extract. Supplementary data of this table and more data for the maximum deGFP synthesis rate can be found in my publication [42].

leak observed on the promoter P_{32} is due to the primary sigma factor. Other well-characterized σ^{32} dependent promoters could be tested to determine whether a more specific unit can be constructed with this sigma factor.

Another important feature is the high specificity of σ^{28} and its promoter. The leak through the promoter P_{28} is almost systematically below the detection limit. The non-specific expression generated by σ^{28} through the other promoters (P_{70} excluded), less than 0.2 % in the worst case, is smaller than the leak generated by the other sigma factors. The σ^{28} unit is the most efficient and the most specific sigma factor unit tested in this work. The low level of crosstalk for the two bacteriophage units, well-known for their high specificity, was expected. Overall, we can also conclude that all of the alternative sigma factors are either not present in the extract or are present at insufficient concentrations to generate undesirable signals.

The last feature in the crosstalk table is the high level of non-specific expression induced by the σ^{19} through the promoter P_{38}. Whether this crosstalk can be decreased with a more specific σ^{38}-dependent promoter has yet to be tested. The entire crosstalk table was also determined for the saturation regime of plasmid concentration (Table 3.2). As anticipated, the crosstalk level in this regime is much higher, but no new features are observed.

3.2.3 Passive Transcriptional Regulation by Competition

E. coli has seven transcription initiation factors, sigma factors and one core transcription machinery, core RNAP. Core RNAP is charge of synthesizing messenger RNA, but promoter selectivity depends on a sigma factor bound on a core RNAP (RNAP holoenzyme). RNAP holoenzyme formation relies on the affinity of sigma factor to core RNAP, leading to a competition between sigma factors to occupy core RNAP which exist in a cell [111, 114]. Therefore, the competition of sigma factors is responsible for passive gene regulation [115, 116]. To determine the passive transcription regulations that result from such enzymatic contests, a competition assay was performed as follows: the expression of deGFP through each of the σ^{19}, σ^{24}, σ^{28}, σ^{32}, σ^{54} and σ^{70} single transcriptional activation units was carried out with and without all the other sigma factors in the reaction. The assay was performed in the linear and the saturation regime of plasmid concentrations (Figure 3.2 (A, B)),

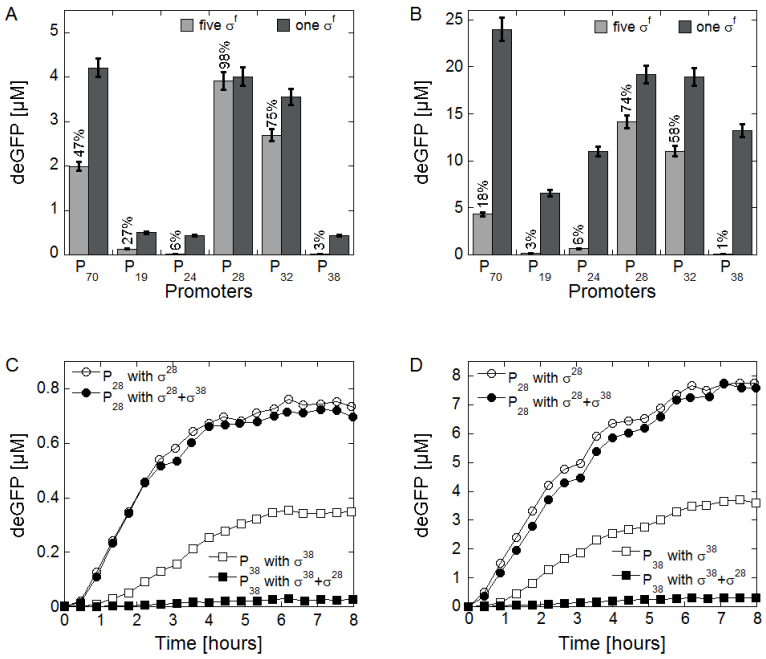

Figure 3.2: **(A)** End-point deGFP expression with one or five sigma factors expressed in the cell-free reaction in a single test tube are shown. The cell-free reactions with 0.1 nM of plasmid encoding each of the sigma factors and 1 nM of reporter plasmid (linear regime of plasmid concentration, σ^{70} condition in TABLE 3.1) were incubated at 29 $^{\circ}C$ for 8 hours. In the first set of reaction (one σ^{f}), deGFP was synthesized through each individual sigma transcriptional activation cascade. In the second set of reaction (five σ^{f}), σ^{19}, σ^{24}, σ^{28}, σ^{32}, and σ^{38} cloned under the promoter P_{70} were expressed simultaneously, whereas only one reporter plasmid, deGFP cloned under the promoter P_{19}, P_{24}, P_{28}, P_{32}, or P_{38} was added to the reaction. The percentage indicates the level of expression compared to the one σ^{f} case. **(B)** The same experiments as in the panel (A) in the saturation regime (0.5 nM of plasmid encoding each of the sigma factor and 5 nM of reporter plasmid) were done. **(C)** Kinetics of passive transcription regulation by competition with σ^{28} and σ^{38} are shown. The cell-free reactions with 0.2 nM of each plasmid (linear regime of plasmid concentration, σ^{70} condition in TABLE 3.1) were incubated at 29 $^{\circ}C$. The expressoin of deGFP through the σ^{28} is not sensitive to the co-expression of σ^{38} in the reaction. The expression of deGFP through the σ^{38} decreases by a factor of 20 when σ^{28} is also expressed in the same reaction. **(D)** The same experiments as in the panel (C) in the saturation regime (1 nM of plasmid encoding each of the sigma factors and 4 nM of reporter plasmid) were done.

The σ^{28}, σ^{32} and σ^{70} units resist competition the most, a result in agreement with *in vitro* measurements of binding affinity between individual sigma factors and the core RNA polymerase [111]. The σ^{28} unit is mostly insensitive to competition, even at high plasmid concentration. A 10-fold increase of sensitivity is observed for σ^{19} between the two regimes of plasmid concentrations. Cell-free expression with σ^{24} or σ^{38} is strongly inhibited by the presence of the other sigma factors in both regimes.

To better characterize passive transcription regulation by competition, we tested two subsets of sigma factors. First, we studied gene expression with σ^{28} and σ^{38}, the most distant sigma subunits in competition sensitivity. Whereas deGFP synthesis through the promoter P_{28} is insensitive to the presence of σ^{38} at low and high plasmid concentrations, gene expression through the promoter P_{38} is decreased by a factor of 20 when σ^{28} is expressed (Figure 3.2 (C, D)). The passive transcriptional repression created by σ^{28} illustrates the additional gene regulations that emerge when transcriptional activation units are used in the same circuit. In this particular case, σ^{28} can be used to down regulate the upstream part of a circuit, which increases the efficiency of the information flow and reduces the crosstalks. We shall see later that the sigma factor competition does not prevent the construction of parallel circuit.

3.2.4 AND Gate

We have been investigating and characterizing six sigma factors out of seven sigma factors, but one sigma factor, σ^{54} is still left as excluded sigma factor. In *E. coli*, the transcription through σ^{54}-dependent promoters is co-activated by the enhancer protein NtrC (Nitrogen regulatory protein C). The *E. coli* glnAp2 regulatory DNA part, a well-characterized σ^{54} specific promoter, contains a set of operator sites specific to NtrC located upstream of the promoter region [117]. Transcription through the glnAp2 promoter works literally as an AND gate. Phosphorylation of the co-activator NtrC, also required for activation, is carried out either by the kinase/phosphatase NtrB (Nitrogen regulatory protein B) or by autophosphorylation with specific chemical substrates [118]. The reaction conditions and the buffer 3-PGA are sufficient for the phosphorylation of NtrC. The output signal is twice as great when the high-energy substrate carbamyl phosphate, specific for the phosphorylation of NtrC, is added to the reaction. No increase of deGFP synthesis is observed when NtrB is also expressed in the reaction (data

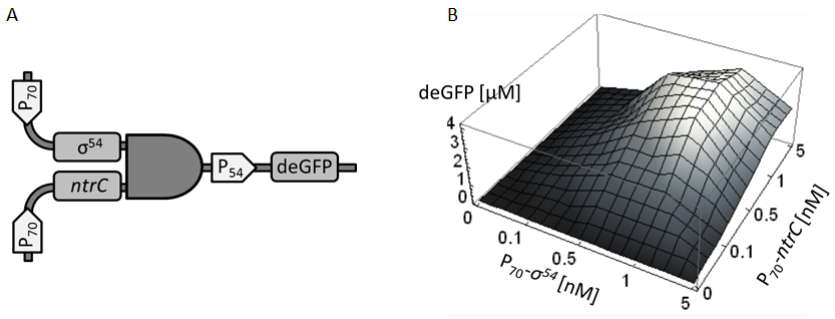

Figure 3.3: A two-stage transcriptional activation cascades for σ^{54} and NtrC are shown. **(A)** Schematic diagram of AND gate circuit using σ^{54} and co-activator, NtrC is shown. Two genes, σ^{54} and *ntrC* cloned under σ^{70} specific promoter, P_{70} are required to activate gene *deGFP* cloned under σ^{54} specific promoter, P_{54}. **(B)** 3D plot of end-point deGFP expression as a function of the concentration of two plasmids, P_{70}–σ^{54} and P_{70}–*ntrC* is shown. 5 nM of plasmid, P_{54}–*deGFP* was fixed in the cell-free reaction (σ^{54} condition in TABLE 3.1). 10 mM carbamyl phosphate was supplemented to increase gene expression.

not shown).

To construct a synthetic AND gate, we cloned the σ^{54} and the *ntrC* genes under the same promoter P_{70} in separate plasmids and the *deGFP* gene under the glnAp2 promoter (Figure 3.3 (A)). Tagging NtrC and σ^{54} with AAA+ specific degrons was not tested. The AND gate was studied with a slightly modified extract. For this transcriptional activation unit only, we noticed that the magnitude of the output signal was much higher when the crude extract was prepared with the S30 buffer A adjusted to pH 8.2 rather than 7.7. The activity of the other transcriptional activation units is decreased by only 20 % with this extract. In the optimal conditions, the maximun deGFP production is largely above the reference concentration of 500 nM (Figure 3.3 (B), Table 3.1). The orthogonality of the glnAp2 promoter with respect to the primary σ^{70} was determined by measuring the synthesis of deGFP in the presence of one of the inputs. The output signal as a function of each input is similar. The protein production (0.04 μM) and the rate of protein synthesis (0.25 nM/min) decrease by a factor of 100 when only one input is used.

3.2.5 Series Circuits: Multiple-Stage Cascade

Using the repertoire of transcriptional activation units and the preliminary circuit design rules, we next constructed a five-stage transcriptional activation cascade that creates a one-hour delay with a specific and biologically relevant end-point output signal (i.e. with a magnitude of at least 500 nM). Our objective was to understand how the crosstalks alter the signals specificity for a circuit composed of a large number of activation units placed in series. The optimization of the output signal was also used to test the adjustable parameters of the toolbox.

We chose to order the units by strength from the weakest to the strongest to obtain a favorable cascading of the amplification factors (Figure 3.4 (A)). The position of σ^{19} in the second stage just after σ^{38} takes advantage of the leak of σ^{19} through P_{38} to create a positive feedback loop. The position of σ^{28} in the third position creates a negative feedback loop, by competition-induced passive regulation, that turns off the second stage of the cascade (P_{38}) and partially inhibits transcription of the first and third stages (promoter P_{70} and P_{19}). This creates an auto-regulation of the first part of the cascade, which improves both the transmission of the information along the series circuit and the bookkeeping of resources. σ^{28} turns off the first stages of the cascade, which decreases unnecessary protein synthesis along the cascade. Furthermore, because it is the most orthogonal and the most competitive sigma factor, σ^{28} attenuates the leaks from the other sigma factors. The T7 unit, used as the last amplification stage, is not sensitive to competition and it is the strongest amplifier. The reporter gene is placed in the last stage of the circuit to monitor the output signal. The T3, σ^{24}, σ^{32} and AND gate units were not used in this example of circuit construction.

The concentration of each gene was first fixed to 0.2 nM to study the circuit in the linear regime of plasmid concentration (1 nM of total plasmid concentration, σ^{70} condition in Table 3.1). In these conditions, the output signal barely reaches the 500 nM reference level (Figure 3.4 (B)) with a relatively low rate of deGFP production ($\sim$1.5 nM/min). The output signal, however, is specific with a signal to leak ratio of 10. To get a greater amplification and faster response, the gain of each stage was increased by adjusting the concentration of each plasmid to 1 nM (5 nM of total plasmid concentration, saturation regime of plasmid concentration), which corresponds to the concentration of genes in *E. coli*. Whereas a much larger output signal and a faster

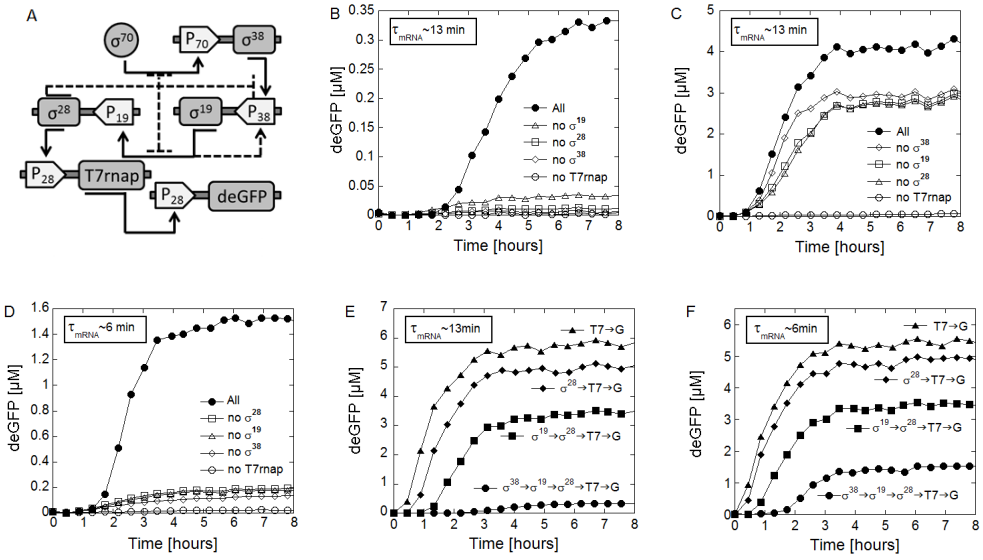

Figure 3.4: **(A)** Schematic diagram of a five-stage transcriptional activation cascade is shown. Nonspecific interactions are shown as dotted lines (inhibition of transcription due to the competition of σ^{28} with the other sigma factors, crosstalk of σ^{19} on P_{38}). **(B)** Kinetics of deGFP expression in the linear regime of plasmid concentration (0.2 nM of each plasmid) with an endogenous global mRNA mean lifetime of 13 min is shown. The specificity of the output signal is confirmed by the four negative controls, which consist of removing one of the four sigma factor stages (no σ^{19} means no P_{38}–σ^{19}, signal to leak ratio > 10). The output signal, however, barely reaches the biologically relevant concentration, fixed to 500 nM. **(C)** Kinetics of deGFP expression in the saturated regime of plasmid concentration (1 nM of each plasmid) with an endogenous global mRNA mean lifetime of 13 min is shown. The negative controls indicate that the output signal of the cascade is not specific. **(D)** Kinetics of deGFP expression in the saturated regime of plasmid concentration (1 nM of each plasmid) with a global mRNA mean lifetime of 6 min is shown. The specificity of the output signal is confirmed by the four negative controls (signal to leak ratio > 10), and the output signal is largely above the biologically relevant concentration of 500 nM. **(E)** Kinetics of deGFP expression through two-, three-, four- and five-stage transcriptional activation cascades with a global mRNA mean lifetime of 13min in the linear regime of plasmid concentration (0.2 nM of each plasmid) are shown. **(F)** Kinetics of deGFP expression through two-, three-, four- and five-stage transcriptional activation cascades with a global mRNA mean lifetime of 6 min in the saturated regime of plasmid concentration (1 nM of each plasmid) are shown.

response are observed, the circuit loses its specificity (Figure 3.4 (C)). Except for the T7 stage, large output signals are observed when one of the stages is removed from the series circuit. The crosstalks between units lead to a large amplification of non-specific signals when the concentration of each plasmid is increased from 0.2 nM (linear regime of plasmid concentration) to 1 nM (saturation regime of plasmid concentration). The different steps of construction of the circuit at 1 nM of each plasmid show that the amplification of the leak is already significant when the circuit is composed of only two stages (data not shown, see my publication [42]).

Rather than looking for a better range of plasmid concentration to optimize the output signal, we tested the behavior of the circuit as a function of the global messenger RNA turnover. The concentration of each plasmid was held at 1 nM. With 36 nM of MazF, the global mRNA mean lifetime was decreased from 13 min to 6 min, a lifetime in the range of the mRNA half-life measured in *E. coli*. In this case, the magnitude of the output signal (1.5 μM) is largely above the reference mark (Figure 3.4 (D)). The output signal is specific, with a signal to leak ratio of 10. Furthermore, the rate of deGFP synthesis is increased by a factor of 10 ($\sim$15 nM/min), and the onset of reporter synthesis is observed after one hour of incubation, rather than two hours in the case of low plasmid concentration (0.2 nM) and slow mRNA turnover (13 min). The acceleration of the mRNA inactivation rate becomes critical at the five-stage level. With a smaller number of stages, both systems (1 nM plasmid and 6 min mean lifetime, 0.2 nM plasmid and 13 min mean lifetime) have specific output signals with similar magnitudes (Figure 3.4 (E, F)). This circuit demonstrates how critical the mRNA turnover is for the specificity of the signal. Non-specific signals are filtered out only by a 2-fold acceleration of the global mRNA turnover.

3.2.6 Parallel Circuits

Gene circuits usually consist of series and parallel connections. The transcription regulation induced by competition between sigma factors, although useful for some circuit configurations (Figure 3.4 (A)), can be also as a serious limitation to design series or parallel circuits. To show that strong and weak sigma factors can be used simultaneously, multiple transcriptional activation units were carried out in parallel. First, the σ^{28} and the σ^{38} cascades were performed simultaneously (Figure 3.5 (A)). σ^{28} was tagged with

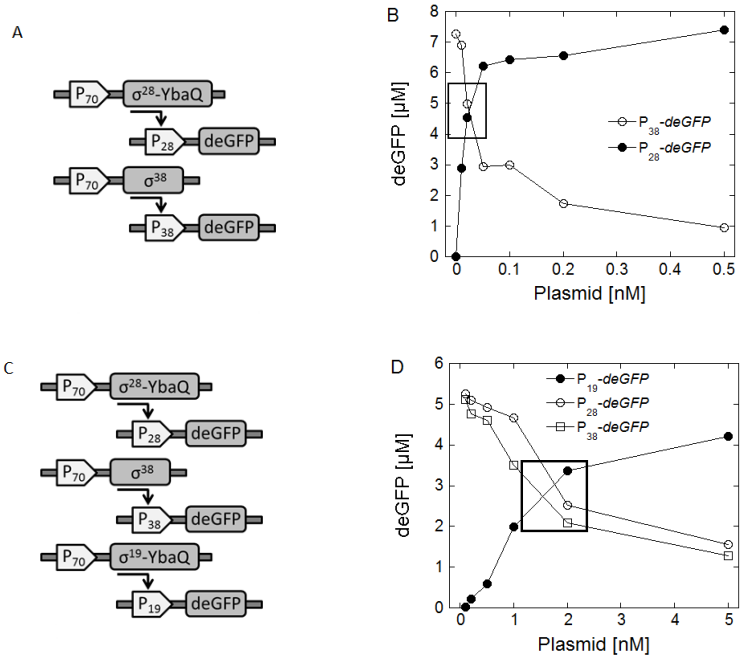

Figure 3.5: **(A)** Schematic diagram of two two-stage transcriptional activation units carried out in parallel is shown. σ^{28} is tagged with the degron, YbaQ. **(B)** End-point deGFP expressions as a function of the concentration of plasmid, P_{70}–$\sigma^{28} - YbaQ$ are shown. The cell-free reaction with 0.5 nM of P_{70}–σ^{38}, 10 nM of P_{38}–$deGFP$, and 2 nM of P_{28}–$deGFP$ (σ^{70} condition in TABLE 3.1) was incubated at 29 $°C$ for 8 hours. The rectangular box indicates that the output signals of each cascade have the same magnitude for a P_{70}–$\sigma^{28} - YbaQ$ plasmid concentration of about 0.025 nM. **(C)** Schematic diagram of three two-stage transcriptional activation units carried out in parallel is shown. σ^{19} and σ^{28} are tagged with the degron, YbaQ. **(D)** End-point deGFP expressions as a function of the concentration of plasmid, P_{70}–$\sigma^{19} - YbaQ$ are shown. The cell-free reaction with 0.5 nM of P_{70}–σ^{38}, 10 nM of P_{38}–$deGFP$, 0.025 nM of P_{70}–$\sigma^{28} - YbaQ$, 2 nM of P_{28}–$deGFP$, and 10 nM of P_{19}–$deGFP$ (σ^{70} condition in TABLE 3.1) was incubated at 29 $°C$ for 8 hours. The rectangular box indicates that the output signals of each cascade have the same magnitude for a P_{70}–$\sigma^{19} - YbaQ$ plasmid concentration of about 2 nM.

the YbaQ AAA+ specific degron to attenuate its domination. By simply adjusting the gene concentrations, the endpoint output signal of each cascade can be chosen over a wide range of magnitudes all above the 500 nM reference concentration (Figure 3.5 (B)). The same observations were made with three transcriptional activation units placed in parallel (Figure 3.5 (C)). When the σ^{19} cascade is added to the previous circuit, the output signals of each unit can be adjusted over a wide range of magnitudes (Figure 3.5 (D)). Parameters can be chosen so as to get output signals with same magnitude in each branch of the parallel circuits. The freedom to adjust the gene concentrations, the global mRNA turnover, and the protein degradation rate allows constructing various series and parallel circuits without even tuning the strength of the other regulatory parts (promoter, untranslated region, ribosome binding site).

3.2.7 Transcriptional Repression Units

In gene regulation, active transcription repression is an essential counterpart to transcription activation. To show that transcriptional repression could be effectively implemented with our cell-free system, four σ^{70} transcriptional repression units were studied (Figure 3.6 (A)).

First, the lactose system with the synthetic regulatory element $P_{LlacO-1}$, composed of a strong promoter specific to σ^{70} and two lac operators was tested [119]. The *lacI* repressor gene was cloned under the $P_{LlacO-1}$ element to make a negative feedback loop. The *deGFP* reporter gene was cloned under the same element in a separate plasmid. The repression of deGFP expression is observed after 30 min of incubation (Figure 3.6 (B)). The reporter protein is fully expressed when 0.5 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to the reaction. The range of IPTG concentration required to inhibit the repression is similar to the amount of IPTG used for *in vivo* induction (data not shown, see my publication [42]). A much higher concentration of lactose, also comparable to *in vivo* experiments, is required to induce the expression of deGFP (data not shown, see my publication [42]). At low plasmid concentration, however, these observations are slightly biased by the presence of endogenous lac repressor in the extract. A 50 % repression is observed when only the reporter plasmid is used at a concentration of 0.5 nM (data not shown, see my publication [42]). At higher plasmid concentration, the repression due to the presence of endogenous lac repressor

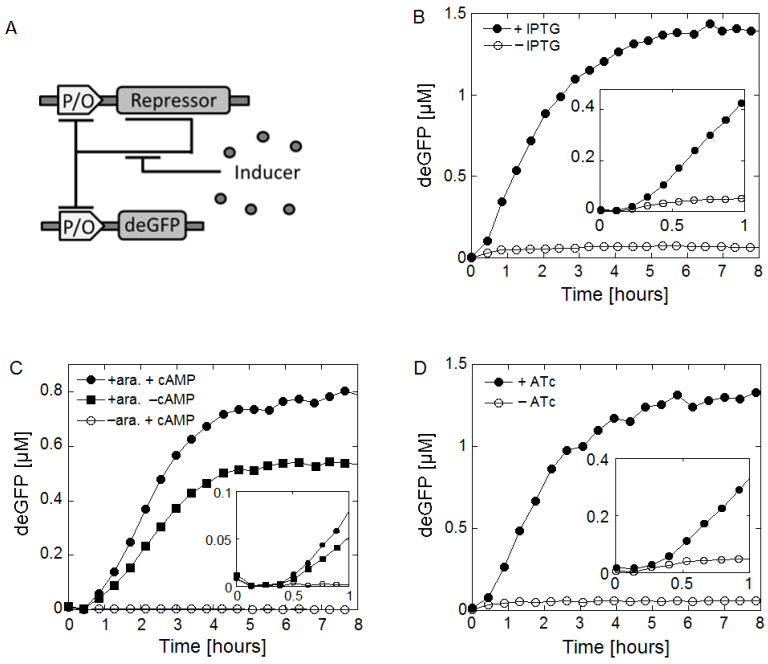

Figure 3.6: Lactose, arabinose, and tetracycline inducible transcriptional repression units are shown. **(A)** Schematic diagram of the circuit is shown. A promoter/operator (P/O) represents a σ^{70} specific promoter and a certain repressor specific operator. Repression is inhibited by its specific inducer. **(B)** Kinetics of the lactose system with or without inducer, IPTG, are shown. The *E. coli lacI* repressor gene and the *deGFP* gene were cloned under the $P_{LlacO-1}$ regulatory element. The cell-free reaction with 1 nM of $P_{LlacO-1}$–*LacI* and 0.5 nM of $P_{LlacO-1}$–*deGFP* (σ^{70} condition in TABLE 3.1) was incubated at 29 °*C*. The expression of deGFP was fully derepressed when 0.5 mM of IPTG was added to the reaction. **Inset**: It shows a blow-up of the first hour of expression. **(C)** Kinetics of the arabinose system with or without inducer, arabinose, are shown. The *deGFP* gene was cloned under the araBAD promoter/operator into the plasmid pBAD bearing the arabinose repressor gene *araC*. The cell-free reaction with 5 nM of pBAD–deGFP (σ^{70} condition in TABLE 3.1) was incubated at 29 °*C*. The expression of deGFP was derepressed when 10 mM of arabinose was used in the reaction. A slight increase of deGFP synthesis was observed when cAMP was added to the reaction. **Inset**: It shows a blow-up of the first hour of expression. **(D)** Kinetics of the tetracycline system with or without inducer, anhydrotetracycline, are shown. The *E. coli tetR* repressor gene and the *deGFP* gene were cloned under the $P_{LtetO-1}$ regulatory element. The cell-free reaction with 2 nM of $P_{LtetO-1}$–*TetR* and 1 nM of $P_{LtetO-1}$–*deGFP* (σ^{70} condition in TABLE 3.1) was incubated at 29 °*C*. The expression of deGFP was fully derepressed when 10 μM of anhydrotetracycline (ATc) was added to the reaction. **Inset**: It shows a blow-up of the first hour of expression.

protein in the system is negligible.

Next, the arabinose inducible system with the plasmid pBAD, which contains the araBAD regulatory element and the *araC* repressor gene was tested. Transcription through the araBAD element, a σ^{70} specific promoter with two operator sites, is repressed by the protein AraC in the absence of arabinose and activated by AraC in the presence of arabinose [120, 121]. Transcription through the araBAD promoter is also stimulated by the cAMP catabolic activator protein Crp via a CAP operator site [121, 122]. The synthesis of deGFP, cloned under the araBAD promoter, is repressed in the absence of arabinose, whether cAMP is present in the solution or not (Figure 3.6). When a concentration of 10 mM of arabinose is used in the reaction (0.2 % w/v), a typical concentration used for induction in *E. coli* cells, deGFP is fully expressed. A 2-fold maximum increase of gene expression is observed upon the addition of 0.75 mM of cAMP to the reaction (data not shown, see my publication [42]), a stimulation smaller than the 5- to 6-fold increase observed *in vivo* [122]. The expression of the *crp* gene in the reaction (cloned under a promoter P_{70}) does not have any effect on the expression of deGFP (data not shown, see my publication [42]), presumably because the concentration of endogenous CRP protein present in the extract is high enough to coactivate the expression. Compared to the reference plasmid P_{70}-*deGFP*, the deGFP production in the open state ($\sim 1\mu$M) and the synthesis rate ($\sim$5 nM/min) are smaller. This is due to the promoter and to the native untranslated region, not as strong as the promoter P_{70} and the UTR1 untranslated region used in this study. The synthesis of deGFP in the repressed state is at or below the background level ($0.1 <$ nM/min). Consequently, the efficiency of repression, determined by the ratio of the rate of protein synthesis in the open state and the repressed state, is larger than 1000 (detection limit). These observations are not biased by the presence of endogenous AraC repressor in the extract. The expression of deGFP is identical in the presence and in the absence of arabinose when the gene araC is knocked out of the plasmid construction (data not shown, see my publication [42]).

Although not systematic, leftovers of repressor proteins in the extract can slightly bias the function of synthetic circuits at low plasmid concentrations. An approach to get around this limitation is to use repressors not present in the *E. coli* extract. The tetracycline system, constructed as a negative feedback loop, was tested with the

synthetic regulatory element $P_{LtetO-1}$, composed of a strong promoter specific to σ^{70} and two tet operators [119]. The tetracycline repressor gene *tetR* and the *deGFP* gene were cloned under $P_{LtetO-1}$ in two separate plasmids. As expected, no repression is observed when only the reporter gene is used (data not shown, see my publication [42]). The repression of deGFP expression is observed after 30 min of incubation (Figure 3.6). The reporter protein is fully expressed when a concentration of 10 μM (5 μg/ml) of anhydrotetracycline (ATc) is used in the reaction. The high concentration range of ATc required for full expression (data not shown, see my publication [42]), fifty times greater than the amount used for *in vivo* induction, is due to the high level of TetR repressor expressed in the open state, on the order of a few micromolars. The rate of deGFP synthesis decreases 200-fold between the open and closed states. The concentration of TetR repressor monomers required for repression, estimated from the kinetics of deGFP expression to be between 10 nM and 50 nM, agrees with previous characterizations of the tetracycline operon (dimer dissociation constant $K_d \sim 10^{-7}$ to 10^{-8} M and repressor dimer-operator dissociation constant $K_d \sim 10^{-12}$ to 10^{-13} M [123, 124]).

Toggle switch was constructed with two inducible repression units, lactose system and tetracycline system consisting of two outputs and four possible states (Figure 3.7 (A)). Each branch of the circuit was first tested individually (data not shown, see my publication [42]). The concentration of the regulatory part $P_{LlacO-1}$ was kept high so that the action of endogenous lac repressor is negligible. Without any inducer in the reaction both sides of the circuit are closed because both promoters have comparable strength and both repressions occur at comparable concentrations of repressor dimers (Figure 3.7 (B, C)). As expected, addition of one inducer opens only one side of the circuit, both sides are open in the presence of IPTG (50 μM) and ATc (10 μM).

Non-inducible system was studied as well. Coliphages also provide a large number of repression units. The lambda repressor, for example, is fully active in our cell-free system, thus gene *cI* was cloned under the regulatory part, OR2-OR1-Pr (P_{70}). This part consists of a σ^{70} specific promoter between two CI binding operator sites, OR1 and OR2. This repression unit was controlled by mutation of operators (P_{70-mut}). Reporter gene *eGFP* was cloned under either P_{70} which is supposed to be repressed by CI or under P_{70-mut} which is not repressed (Figure 3.8 (A)). A 100-fold repression is observed between the closed and the open states (Figure 3.8 (B)). As expected, no

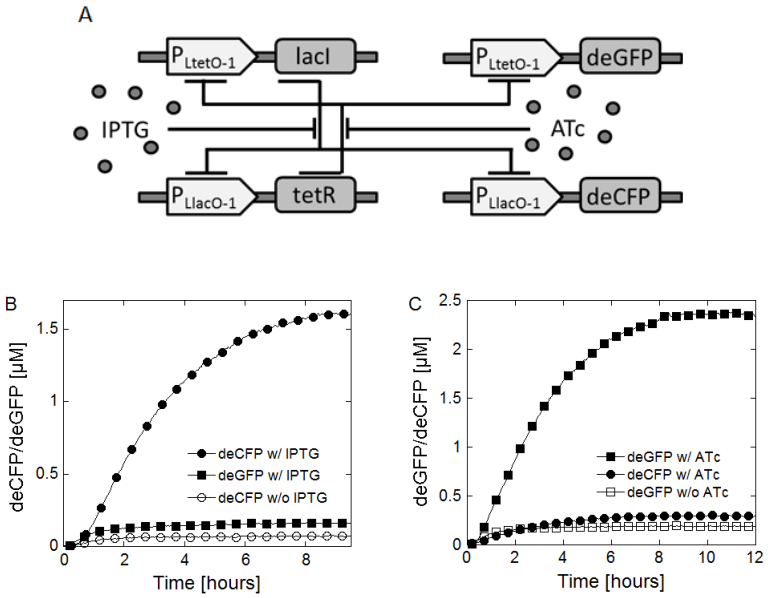

Figure 3.7: An inducible circuit constructed with two transcriptional repression cascades is shown. **(A)** Schematic diagram of the circuit composed of the lactose and tetracycline inducible repressions and the reporter genes *deCFP* and *deGFP* is shown. The concentrations of the four plasmids, 1 nM of $P_{LtetO-1}$–*lacI*, 2 nM of $P_{LtetO-1}$–*deGFP*, 2 nM of $P_{LlacO-1}$–*tetR*, and 4 nM of $P_{LlacO-1}$–*deCFP* were fixed in the cell-free reaction with σ^{70} condition in TABLE 3.1 (deCFP channel: 434 nm excitation and 485 nm emission; deGFP channel: 485 nm excitation and 528 nm emission). **(B)** Kinetics of deCFP and deGFP expression in the closed state (0 μM of IPTG) and the open state (50 μM of IPTG) are shown. deCFP was expressed in the presence of IPTG and repressed without IPTG. deGFP was repressed by the TetR repressor. **(C)** Kinetics of deCFP and deGFP expression in the closed state (0 μM of ATc) and the open state (10 μM of ATc) are shown. deGFP was expressed in the presence of ATc and repressed without ATc. deCFP was repressed by the LacI repressor.

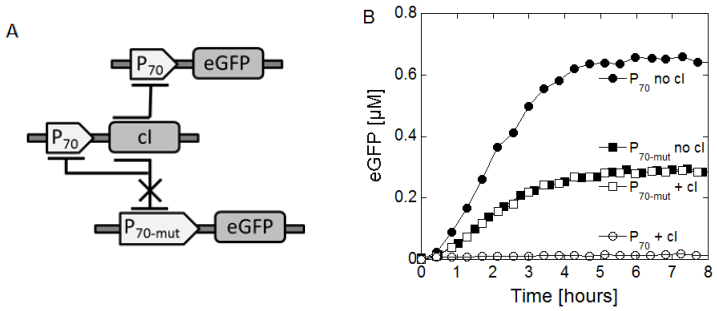

Figure 3.8: A non-inducible repression circuit is shown. **(A)** Schematic diagram of a non-inducible circuit using lambda repressor CI is shown. The lambda repressor gene *cI* and the *eGFP* gene were cloned under the P_{70} promoter. The same constructions were made with a mutated promoter, P_{70-mut} that cannot be repressed by CI. **(B)** Kinetics of eGFP expression with two sets of regulatory parts are shown. The cell-free reaction with 0.5 nM of reporter plasmid, either P_{70}–*eGFP* or P_{70-mut}–*eGFP*, (σ^{70} condition in the TABLE 3.1) was incubated with and without 1 nM of the plasmid encoding the repressor CI, either P_{70}–*cI* or P_{70-mut}–*cI* respectively, at 29 °*C*. The strength of the promoter P_{70-mut} is slightly inferior to the promoter P_{70}.

change is observed between the closed and the open states with P_{70-mut}.

3.3 Artificial Cell Approach

3.3.1 Long-lived Gene Expression in Macroscopic Continuous System

The availability of resources is an issue for batch mode cell-free reactions, especially for relatively slow processes requiring a significant amount of energy and building blocks like transcription and translation. Typically, batch mode cell-free expression reactions last 4 to 5 hours in the best conditions. In batch mode, gene expression is independent from the resources only for a short period of time (∼1 to 2 hours for conventional systems and our system). The decrease of the energy charge, the degradation of some amino acids and the change of pH during protein synthesis rapidly alter the kinetics of gene expression [45]. The accumulation of waste products is also a concern for pure synthetic biochemical systems. The development of cell-free reactions stable over long periods of time is necessary to test and to model larger circuits.

Long-lived cell-free gene expression reactions that produce proteins in high yield have

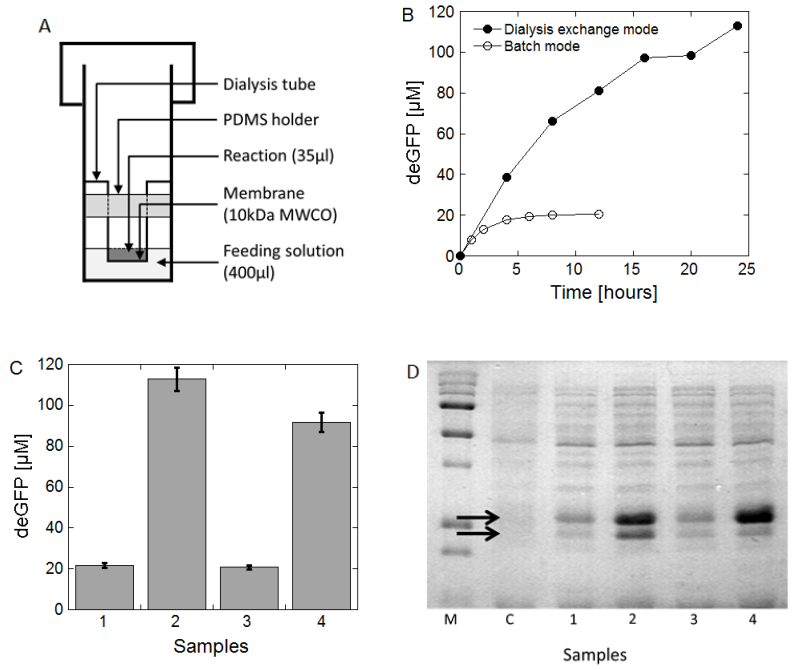

Figure 3.9: Macroscopic structure for long-lived biochemical reaction is shown. **(A)**Schematic diagram of the custom-made, dialysis exchange system is shown. A dialysis tube (MWCO 10 kDa, 35 μl of cell-free reaction) is held in a large tube (7 ml) containing 400 μl of feeding solution. **(B)** Kinetics of deGFP expression in batch mode and dialysis mode are shown. The cell-free reactions with 5 nM of plasmid, P_{70}–*deGFP* (σ^{70} condition in the TABLE 3.1) were incubated at 29 $^{\circ}C$. **(C)** End-point deGFP expression in batch mode (columns 1 and 3) and dialysis mode (columnes 2 and 4) are shown. The expression of deGFP was carried out with 5 nM of the plasmid, P_{70}–*deGFP* (columns 1 and 2; σ^{70} condition in the TABLE 3.1) and through the two-stage cascade (columns 3 and 4; 0.2 nM of P_{70}–σ^{28} and 5 nM of P_{28}–*deGFP* with σ^{28} condition in the TABLE 3.1. **(D)** SDS-PAGE 13 % of cell-free reaction is shown. M: marker, C: cell-free reaction with no plasmid, 1 -4: same sample as in the panel (C). The top arrow indicates the deGFP band, and the bottom arrow indicates the ampicillin resistance protein band.

been already engineered with conventional hybrid systems [98, 99]. Typical continuous-exchange bacteriophage systems are prepared for 1 ml of reaction and 10 ml of feeding solution. To test our endogenous system in continuous-exchange mode, we devised a dialysis reactor that works with 35 μl of reaction and 400 μl feeding solution (Figure 3.9 (A)). Two different reactions were tested. In the first reaction, deGFP is expressed through the promoter P_{70} (plasmid P_{70}–*deGFP*). We observe a time extension of cell-free expression by a factor of 4, from 4 hours in batch mode to 16 hours in dialysis mode (Figure 3.9 (B)). In the second reaction, deGFP is expressed through the transcriptional activation cascade P_{70}–$\sigma^{28} \rightarrow P_{28}$–*deGFP*. A time extension of cell-free expression by a factor of 4, from 4 hours in batch mode to 16 hours in dialysis mode was observed (data not shown). In both cases, the amount of active deGFP produced is also increased by a factor of 5 to 6 (Figure 3.9 (C), 0.55 mg/ml in batch mode, 2.5 to 3 mg/ml in dialysis mode). In batch mode, a total concentration of 0.7 mg/ml deGFP (0.55 mg/ml active deGFP based on fluorescence) is measured on SDS PAGE for both reactions (Figure 3.9 (D)). In dialysis mode, a total concentration of 4 mg/ml deGFP is measured for both reactions. Our endogenous cell-free toolbox works as well as conventional hybrid systems in both batch and dialysis modes.

In the first reaction (plasmid P_{70}–*deGFP*), the ampicillin antibiotic resistance gene, present in the plasmid under a σ^{70} specific promoter, is also significantly expressed in dialysis mode (Figure 3.9 (D), sample 2). In the second reaction, the antibiotic resistance gene is almost not expressed due to the competition of σ^{28} with the housekeeping σ^{70}. σ^{28} regulates its own expression by passively repressing the promoter P_{70} through which it is transcribed. We could not confirm this auto-regulation mechanism from the SDS PAGE because the molecular mass of σ^{28} is comparable to the molecular mass of deGFP (25 kDa). We performed the same experiment with the gene *ntrB* encoding the protein NtrB of mass 32 kDa. As expected, the quantity of σ^{28} measured on gel is barely visible (data not shown, see my publication [42]).

3.3.2 Liposome: Microscopic Continuous System

The cell-free toolbox presented in this work is also devised to program synthetic phospholipid vesicles. Cell-free gene expression carried out in cell-sized liposomes is a bottom-up approach to constructing an artificial cell that focuses on the informational

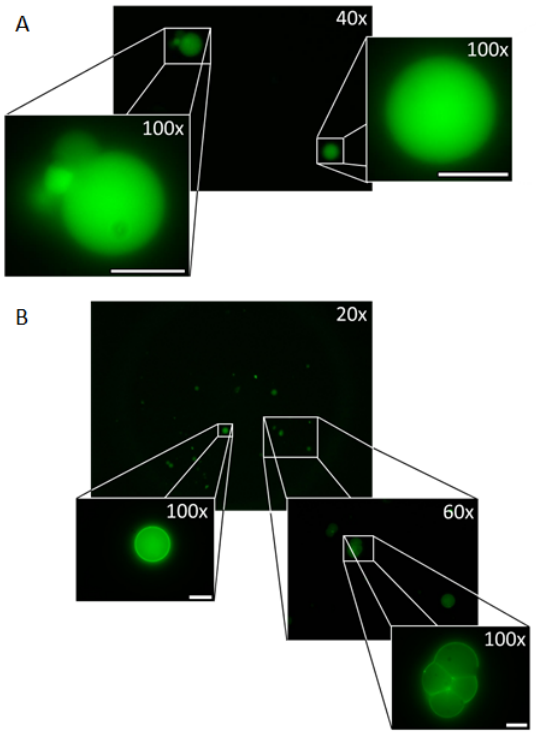

Figure 3.10: Fluorescence images of cell-free expression inside phospholipid vesicles after 10 hours of incubation at room temperature are shown. **(A)** The cell-free reaction with 5 nM of plasmid, P_{70}–*deGFP* (σ^{70} condition in the TABLE 3.1) was incubated in vesicles at room temperature. Scale bar with 20x objective: 50 μm, and scale bar with 100x objective: 10 μm. **(B)** The cell-free reaction with 0.2 nM of P_{70}–σ^{28} and 5 nM of P_{28}–α*H-eGFP* (σ^{28} condition in the TABLE 3.1) was incubated in vesicles at room temperature. Scale bar with 20x objective: 50 μm, scale bar with 60x objective: 25 μm, and scale bar with 100x objective: 25 μm.

and self-organization properties of living matters. The idea of an artificial cell with cell-free TX-TL system was already demonstrated with conventional cell-free system and has been discussed recently [21, 36, 65, 66, 125]. Cell-sized liposomes were produced as follows,

1. 1 μl of cell-free reaction was added in 200 μl of mineral oil,
2. Vortex the mixture,
3. Pipet 100 μl of emulsion and add it on top of 200 μl of feeding solution, and
4. Centrifuge it at 11000 rpm for 11 sec.

A heterogeneous solution of single and aggregates of liposomes with a diameter of one micrometer to a few tens of micrometers is directly produced in a feeding solution (Figure 3.10). The expression of deGFP inside the liposomes is either as high as in a test tube or slightly greater due to the permeability of the phospholipid membrane [65]. The expression of the fusion protein αHemolysin-eGFP allows extending gene expression inside the vesicles by creating a selective permeability of the membrane through the channel of molecular mass cutoff 2-3 kDa. In addition, an artificial phospholipid vesicle as an experimental device to perform the functionality of proteins in cell-like structure was studied. When *E. coli* structure protein MreB, one of which are in charge of cylinder shape, was expressed in cell-free system, the stick-like structure was formed. When this protein, however, expressed together with membrane protein MreC in the liposome, a ring structure was observed [41]. This result shows that the protein expressed in the cell-free system is not only an expressed gene but also a functional protein which has a memory of its own function.

3.3.3 On/Off Switch in Liposome: Toward Artificial Cell

In addition, we tested the arabinose system, the tightest repression tested, in liposome. A cell-free reaction containing the plasmid pBAD-*deGFP* was encapsulated inside vesicles. With no arabinose added to the feeding solution or to the reaction, the repression of deGFP expression is as efficient as when the reaction is carried out in a test tube (Figure 3.11 (A, B)). The expression of deGFP inside the liposomes is identical whether 10 mM of arabinose is added to the feeding solution or to the reaction (Figure 3.11 (C)).

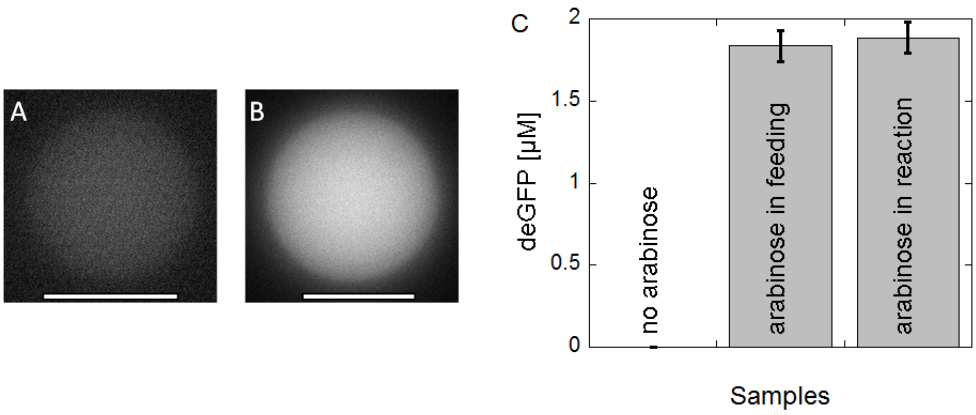

Figure 3.11: Arabinose inducible transcriptional repression system encapsulated in liposome is shown. **(A)** Fluorescence image of deGFP expression with no arabinose added to the cell-free reaction or to the feeding solution is shown. The cell-free reaction with 5 nM of plasmid, pBAD–*deGFP* (σ^{70} condition in the TABLE 3.1), in liposome was incubated at room temperature for 10 hours. Scale bar: 10 μm. **(B)** Same as in panel (A) with 10mM arabinose added to the feeding solution. Scale bar: 10 μm. **(C)** End-point measurements of deGFP expression inside phospholipid vesicles (average fluorescence intensity of 20 large phospholipid vesicles) are shown. In the absence of arabinose, no deGFP can be measured, whereas 1.5 - 2 μM deGFP was produced inside the vesicles when 10mm arabinose was added either to the reaction or to the feeding solution.

Expression of deGFP is a factor of two higher than in the test tube due to the feeding of the reaction through the membrane. Adding arabinose only to the feeding solution is enough to completely open the system. The vesicle system adds to the range of applications of the toolbox by providing a cell-like environment to study self-organization processes via the expression of synthetic gene circuits.

Chapter 4

Mathematical models[1]

4.1 Coupled Coarse-Grained Model

The simplest mathematical model with regard to gene expression from DNA to protein through mRNA is the coarse-grained model. As explained, this model simplifies all detailed biochemical complexity into just a couple of parameters (Figure 2.10). In the previous chapter, this mathematical model was constructed and used to extract two parameters: mRNA lifetime and deGFP maturation time. Despite this model's simplicity, it also was successfully used to fit data from *in vivo* experiments [5, 14]. However, this model has excluded two important parameters. Recalling that the coarse-grained model was constructed like a particle-particle interaction, an inquiry must be made concurring transcription and translation time. mRNA is not synthesized instantly but instead is elongated by successive addition of nucleic acids via transcriptional machinery at reported rates of ∼50 nt/s *in vivo* [69]. Protein is elongated by successive addition of amino acids via translational machinery at reported rates of ∼18 AA/s *in vivo* [69]. This time discrepancy can be solved by accounting for these terms with the appropriate parameters.

The second difficulty posed to this simplest form of the coarse-grained model for the analysis of gene expression in a bacterial system is due to the lack of a cell nucleus

[1] **I acknowledge that Matt Loth's and Jonathan Gapp's editing help was essential to achieve proper English usage and readability of this chapter.**

in prokaryotes. This specific organization allows for the coupled processes of transcription and translation. That is, while mRNA is synthesized by transcriptional machinery, translational machinery can begin protein synthesis once the ribosome binding site (RBS) is transcribed and available for the machinery to bind to it. It was reported that transcription and translation is not coupled with a conventional *E. coli* cell-free system due to use of the high-affinity T7 RNAP as the primary transcription machinery. Our cell-free system may have a coupled TX-TL processes because *E. coli* core RNAP and transcription initiation factor, σ^{70} were used as the primary transcription machinery. Thus, modification of coarse-grained model is required. Recently, German researcher, Kremling, developed the mathematical model for coupled TX-TL processes in bacterial system [126]. However, this model introducing nascent mRNA is not as simple as the coarse-grained model. The coupled coarse-grained model constructed in this work was formulated from the simple coarse-grained model with the introduction of three simple conditions.

mRNA synthesis The schematic diagram of the coarse-grained model showed an one-to-one interaction between DNA and RNAP for transcription (Figure 2.10). From this diagram, the mRNA synthesis is given by

$$D_f + R_{pf} \underset{k_{X-}}{\overset{k_{X+}}{\rightleftharpoons}} DR_p \xrightarrow{k_{TX}} D_f + R_{pf} + m, \qquad \text{: mRNA synthesis} \tag{4.1}$$

where D_f, R_{pf}, DR_p and m represent the concentration of free DNA, free RNAP, DNA-RNAP complex and mRNA. Two equations of enzymatic kinetics can then be constructed for DR_p and m,

$$\dot{DR_p} = k_{X+} \cdot D_f \cdot R_{pf} - (k_{X-} + k_{TX}) \cdot DR_p, \tag{4.2}$$

$$\dot{m} = \frac{k_{TX}}{l} \cdot DR_p, \tag{4.3}$$

where l is the length of mRNA. The quasi-steady-state for intermediate stage DR_p as well as the stable DNA and RNAP were assumed,

$$\begin{cases} \dot{DR_p} = 0, & \text{: steady state assumption} \\ \left.\begin{array}{l} D_o = D_f + DR_p, \\ R_{po} = R_{pf} + l^* \cdot DR_p, \end{array}\right\} & \text{: mass conservation} \\ M_{TX} \equiv \dfrac{k_{X-} + k_{TX}}{k_{X+}}. & \text{: Michaelis-Menten constant} \end{cases} \tag{4.4}$$

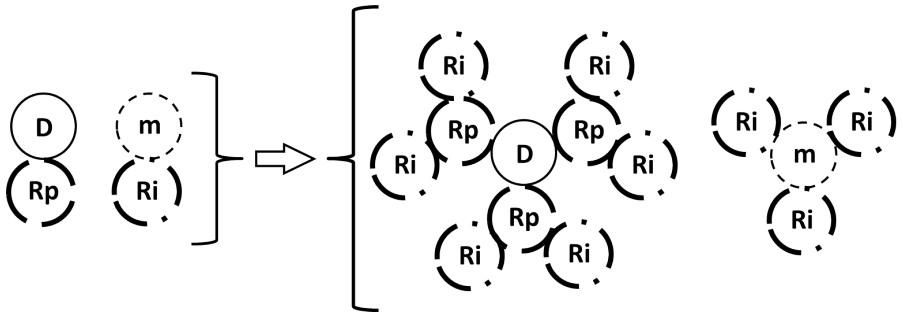

Figure 4.1: The coarse-grained model whose schematic diagram was shown in the Figure 2.10 did not show the coupled system of transcription and translation. In order to introduce a coupled coarse-grained model, a new schematic diagram is suggested. All steps will be the same except the catalysis process shown in this figure. The same symbols as Figure 2.10 are used.

Here, the effective length of mRNA, l^* is introduced because RNAP binding on DNA is not one-to-one but one-to-many depending on the length of operons (l) and the effective distance ($\bar{l}$) between RNAPs on DNA. The effective length l^* is defined as the length of mRNA divided by effective distance ($l/\bar{l}$). For example, when l is 100 bp and RNAP can bind on DNA every 20 bp apart ($\bar{l}$), the effective length of mRNA (l^*) is 5. This number describes how many RNAPs are on the DNA (Figure 4.1). It is important to calculate the number of free RNAP which could potentially bind to DNA. When all conditions are applied to equation 4.2 and the quadratic equation is solved, the equation

$$DR_p = \frac{1}{2 \cdot l^*} \left\{ l^* \cdot D_o + R_{po} + M_{TX} \pm \sqrt{(l^* \cdot D_o + R_{po} + M_{TX})^2 - 4 \cdot l^* \cdot D_o \cdot R_{po}} \right\}. \tag{4.5}$$

is obtained. Only a minus sign is selected because the positive sign gives a non-sense result, $DR_p > R_{po}$. Therefore, the mRNA synthesis term will be

$$\dot{m} = \frac{k_{TX}}{2 \cdot l \cdot l^*} \left\{ l^* \cdot D_o + R_{po} + M_{TX} - \sqrt{(l^* \cdot D_o + R_{po} + M_{TX})^2 - 4 \cdot l^* \cdot D_o \cdot R_{po}} \right\}. \tag{4.6}$$

mRNA inactivation The schematic diagram of mRNA inactivation is not precisely shown in Figure 2.10. The mRNA inactivation includes mRNA degradation by RNases as well as mRNA structural loss of functionality such as stem-loop formation. Here, the structural loss is treated as an enzymatic mechanism. So the schematic diagram of

mRNA inactivation is constructed as following,

$$m_f + E_{1f} \underset{h_{X-}}{\overset{h_{X+}}{\rightleftharpoons}} mE_1 \xrightarrow{h_{TX}} \not{m}_f + E_{1f}, \qquad : \text{mRNA inactivation} \tag{4.7}$$

where h_{X+}, h_{X-} and h_{TX} represent association, dissociation and catalysis reaction rate, respectively. E_1 is the concentration of the provisional alternative enzyme which is responsible for mRNA inactivation, and m_f and mE_1 represent the concentration of free mRNA and mRNA-enzyme complex. $\not{m}_f$ represents the degradation of mRNA. Two equations of enzymatic kinetics for mE_1 and $\not{m}_f$ are

$$\dot{mE_1} = h_{X+} \cdot m_f \cdot E_{1f} - (h_{X-} + h_{TX}) \cdot mE_1, \tag{4.8}$$

$$\dot{\not{m}}_f = h_{TX} \cdot mE_1. \tag{4.9}$$

mRNA length is not accounted for in this equation set as mRNA is no longer functional once even one part of the mRNA is damaged by an enzymatic reaction. In the reaction diagram, mRNA serves as the substrate so the condition sets are written by

$$\begin{cases} \dot{mE_1} = 0, & : \text{steady state assumption} \\ E_{1o} = E_{1f} + mE_1, & : \text{mass conservation} \\ H_{TX} \equiv \dfrac{h_{X-} + h_{TX}}{h_{X+}}. & : \text{Michaelis-Menten constant} \end{cases} \tag{4.10}$$

When applying these conditions into equation 4.8, we will get

$$mE_1 = \frac{m_o \cdot E_{1o}}{H_{TX} \cdot m_o}. \tag{4.11}$$

Therefore mRNA inactivation term will be

$$\dot{\not{m}}_f = \frac{h_{TX} \cdot m_o \cdot E_{1o}}{H_{TX} + m_o}. \tag{4.12}$$

If the fast cutting and $m_o \ll E_{1o}$ are assumed, equation 4.12 can be simplifed

$$\dot{\not{m}}_f = h_{TX} \cdot m_o. \tag{4.13}$$

protein synthesis According to the schematic diagram of protein synthesis shown in Figure 2.10, the schematic equation is given by

$$m_f + R_{if} \underset{k_{L-}}{\overset{k_{L+}}{\rightleftharpoons}} mR_i \xrightarrow{k_{TL}} m_f + R_{if} + p, \qquad : \text{protein synthesis} \tag{4.14}$$

where m_f, R_{if}, mR_i and p represent the concentration of free mRNA, free ribosome, mRNA-ribosome complex and protein, respectively. Enzymatic equations for mR_i and p are written by

$$\dot{mR_i} = k_{L+} \cdot m_f \cdot R_{if} - (k_{L-} + k_{TL}) \cdot mR_i, \tag{4.15}$$

$$\dot{p} = \frac{k_{TL}}{n} \cdot mR_i, \tag{4.16}$$

where n is the length of protein. Under the same assumption as mRNA synthesis, this equation set will meet the following conditions,

$$\begin{cases} \dot{mR_i} = 0, & : \text{steady state assumption} \\ \left.\begin{array}{l} m_o = m_f + mR_i, \\ R_{io} = R_{if} + n_1^* \cdot mR_i + n_2^* \cdot DR_p, \end{array}\right\} & : \text{mass conservation} \\ M_{TL} \equiv \dfrac{k_{L-} + k_{TL}}{k_{L+}}, & : \text{Michaelis-Menten constant} \end{cases} \tag{4.17}$$

where n_1^* is effective length of protein. Ribosome binding on mRNA is not one-to-one, so n_1^* is defined as $n/\overline{n1}$ where $\overline{n1}$ represents the effective distance between ribosomes on mRNA. This is the same concept as RNAPs on DNA (l^*). For the coupled coarse-grained model, one more parameter n_2^* is introduced in order to include the information of coupled process of transcription and translation. Ribosomes can be on nascent mRNA to start translation once RBS on mRNA is synthesized. Nascent mRNA is defined as mRNA in the middle of transcription process, so nascent mRNA is still on DNA yet completely free. Figure 4.1 shows the modification of DNA-RNAP complex and mRNA-ribosome complex to explain coupled system. Therefore, n_2^* can be defined as $l^* \cdot \overline{n2}$ where $\overline{n2}$ represents the average number of ribosomes on DNA. Applying the condition set 4.17 into equation 4.15 gives

$$\begin{aligned} mR_i = \frac{1}{2 \cdot n_1^*} \Bigg\{ & n_1^* \cdot m_o + R_{io} - n_2^* \cdot DR_p + M_{TL} \pm \\ & \pm \sqrt{(n_1^* \cdot m_o + R_{io} - n_2^* \cdot DR_p + M_{TL})^2 - 4 \cdot n_1^* \cdot m_o \cdot (R_{io} - n_2^* \cdot DR_p)} \Bigg\}, \end{aligned} \tag{4.18}$$

where DR_p is from equation 4.5. Only a minus sign is selected because the positive sign

gives non-sense result which is $mR_i > R_{i\,o}$. Therefore, protein synthesis term will be

$$\dot{p} = \frac{k_{TL}}{2 \cdot n \cdot n_1^*} \Big\{ n_1^* \cdot m_{\,o} + R_{i\,o} - n_2^* \cdot DR_p + M_{TL} - \\ - \sqrt{(n_1^* \cdot m_{\,o} + R_{i\,o} - n_2^* \cdot DR_p + M_{TL})^2 - 4 \cdot n_1^* \cdot m_{\,o} \cdot (R_{i\,o} - n_2^* \cdot DR_p)} \Big\}. \tag{4.19}$$

protein degradation Protein degradation in our cell-free system is mainly due to protease ClpXP complex, therefore only protein tagged with degrons is degraded. If the concentration of ClpXP is written as E_2, and $E_{2\,f}$ and pE_2 represent the concentration of free ClpXP and protein-ClpXP complex, then the schematic equation is given by

$$p_{\,f} + E_{2\,f} \underset{h_{L-}}{\overset{h_{L+}}{\rightleftharpoons}} pE_2 \xrightarrow{h_{TL}} \not{p}_{\,f} + E_{2\,f} \qquad : \text{protein degradation} \tag{4.20}$$

where h_{L+}, h_{L-} and h_{TL} represent association, dissociation and catalysis reaction rate, respectively. $\not{p}_{\,f}$ represents the degradation of protein. The same shape of equations 4.8, 4.9 and conditions 4.10 were used, and the equation of protein degradation,

$$\dot{\not{p}}_{\,f} = \frac{h_{TL} \cdot p_o \cdot E_{2\,o}}{H_{TL} + p_o}, \tag{4.21}$$

where H_{TL} is defined as $\frac{h_{L-} + h_{TL}}{h_{L+}}$.

$$G_f(t) = \frac{\alpha \cdot m}{\beta \cdot (\beta - \kappa)} \cdot (\kappa \cdot (e^{-\beta \cdot t} - 1) + \beta \cdot (1 - e^{-\kappa \cdot t})) \tag{4.22}$$

4.2 Coupled Coarse-Grained model with Negative Feedback Process

Pattern formations involving nonlinear dynamics together with feedback processes can be found in many processes, such as morphogenesis, cell growth, and cell differentiation. The quantitative study of temporal and spatial patterns is therefore crucial for understanding the life of a single cell. These patterns include self-organization of biomolecules (in both temporal and spatial dimensions) as well as cell to cell communication within a multicellular system. Studying these patterns in nature is important, however, this thesis focuses on understanding patterns in a synthetic gene circuit; understanding synthetic gene circuits will allow for the reconstitute of life using a bottom-up approach.

To this end, a couple of landmark works concerning both temporal and spatial patterns were published recently [5, 11, 14, 127]. In these works, researchers engineered a synthetic gene circuit to study patterns, including a synthetic biological clock and the localization of reaction induced by a stimulating condition. These experiments were all carried out *in vivo*. Of interest to this chapter, is the fact that these studies also attempted to use a coarse-grained model to predict the patterns that would be observed. This chapter develops the coupled coarse-grained model to describe temporal oscillation which could be achieved with a negative feedback loop. Unfortunately, oscillatory behavior could not be obtained with the engineered gene circuit in our cell-free toolbox. An explanation of why such difficulties remain is given at the end of this section.

Reliability of model To proceed to a temporal oscillatory pattern, it is necessary to determine whether the coupled coarse-grained model is reliable. To check the reliability of this model, saturation of protein synthesis in terms of plasmid concentration was compared to simulation (Figure 4.2). As previously mentioned, plasmid concentration could be split into two regimes: a linear regime and a saturation regime. Protein production reached the maximum at the optimal concentration of plasmid (Figure 2.5 (B)) and a linear regime could be extended depending on several different parameters such as concentration of RNAP, concentration of ribosome, the strength of promoter, the strength of RBS and so on. Figure 4.2 (A, D) shows the extension of the linear regime depending on the strength of the promoter. For the test, a plasmid, P'_{70}–*deGFP* was constructed wherein a single base pair in the promoter region -35 was mutated (TTGACA $\rightarrow$ TAGACA) from the original plasmid, P_{70}–*deGFP*. Protein production decreased 6-fold at 2 nM of plasmid (from 12 μM to 2 μM) but the linear regime was extended 3-fold (from 2 nM to 6 nM). Another plasmid, P''_{70}–*deGFP*, was mutated by one base pair each at both promoter region -35 and -10 (TTGACA $\rightarrow$ TAGACA at -35 and GATAAT $\rightarrow$ GCTAAT at -10) from the original P_{70}–*deGFP*. Subsequent testing measured a 60-fold decrease of protein production at 2 nM of plasmid (from 12 μM to 0.2 μM) and 12-fold extension of linear regime (from 2 nM to 25 nM) (data not shown).

In order to simulate this phenomenon observed in the cell-free system, the equations of the coupled coarse-grained model for a single gene *deGFP* expression were

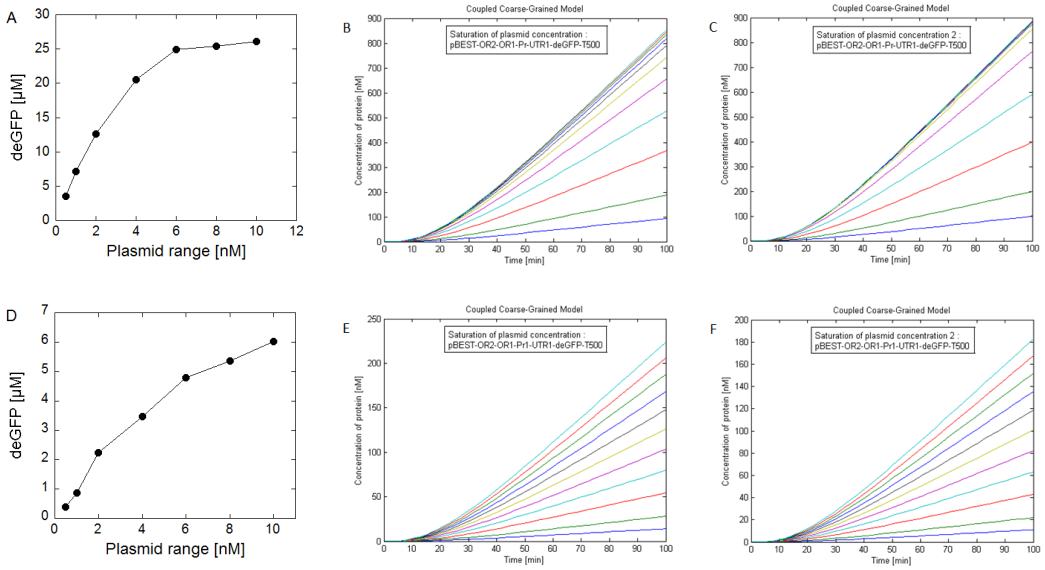

Figure 4.2: Saturation of deGFP production as a function of plasmid concentration and simulation with coupled coarse-grained model are shown. **(A)** End-point protein production as a function of the concentration of plasmid, P_{70}–*deGFP* is shown. Plasmid concentration at the range of 0.5, 1, 2, 4, 6, 8 and 10 nM were tested at the σ^{70} condition in Table 3.1. **(B)** The coupled coarse-grained model was used to simulate the saturation of protein synthesis as a function of plasmid concentration with parameter set 1 in Table 4.1. The plasmid concentration range was at 0.5 nM and from 1 nM to 10 nM at every 1 nM. The simulated kinetics are shown such that the bottom corresponds to the low concentration of plasmid (0.5 nM) and the high concentration (10 nM) corresponds to the top. **(C)** The same simulation as in the panel (B) was done with parameter set 2 in Table 4.1. **(D)** End-point protein production as a function of the concentration of plasmid, P'_{70}–*deGFP* is shown. To decrease the strength of the promoter, one base pair in promoter region "-35" was mutated. Plasmid concentration was the same as in the panel (A). **(E)** The coupled coarse-grained model was used to simulate the saturation of protein synthesis as a function of plasmid concentration with parameter set 1 in Table 4.1. To simulate weak promoter, only k_{X+} was changed and given as multiplying by 10^{-2} from Table 4.1. The plasmid concentration range and the display of the simulated kinetics were the same as in the panel (B). **(F)** The same simulation as in the panel (E) except k_{X+} given as multiplying by 10^{-3} was done with parameter set 2 in Table 4.1.

parameters	unit	set 1	set 2	parameters	unit	set 1	set 2
k_{X+}	$[min^{-1} \cdot nM^{-1}]$	120	1000	k_m	$[min^{-1}]$	1/8	1/8
k_{X-}	$[min^{-1}]$	60	100	k_{T+}	$[min^{-1} \cdot nM^{-4}]$	- (0.01)	- (0.1)
k_{TX}	$[min^{-1}]$	300	300	k_{T-}	$[min^{-1}]$	- (0.01)	- (0.1)
h_{TX}	$[min^{-1}]$	1/13 (1/5)	1/13 (1/5)	k_r	$[min^{1/c}]$	- (10)	- (1000)
				c	[#]	- (2)	- (2)
k_{L+}	$[min^{-1} \cdot nM^{-1}]$	120	1000				
k_{L-}	$[min^{-1}]$	60	100	$\bar{l}$	[bp]	678	678
k_{TL}	$[min^{-1}]$	100	100	$\bar{\bar{l}}$	[bp]	60	60
h_{L+}	$[min{-}1 \cdot nM^{-1}]$	- (10)	- (10)	n	[$a.a.$]	225	225
h_{L-}	$[min{-}1]$	- (0.1)	- (0.1)	$\overline{n1}$	[$a.a.$]	60	60
h_{Tl}	$[min{-}1]$	- (1/8)	- (1/8)	$\overline{n2}$	[#]	1.8	1.8
R_{p0}	$[nM]$	50	50	R_{i0}	$[nM]$	1000	1000
E_{20}	$[nM]$	- (1)	- (1)				

Table 4.1: Two parameter sets are shown. Both sets are used for the saturation of protein synthesis (Figure 4.2) and genetic oscillation (Figure 4.3). Parameters for oscillation are in parenthesis and '−' means no use. The explanation of parameters, k_{T+}, k_{T-}, k_r and c is in the section, temporal oscillation.

constructed,

$$\dot{m} = \frac{k_{TX}}{\bar{l}} \cdot DR_p - h_{TX} \cdot m, \tag{4.23}$$

$$\dot{p_d} = \frac{k_{TL}}{n} \cdot mR_i - k_m \cdot p_d, \tag{4.24}$$

$$\dot{p_f} = k_m \cdot p_d, \tag{4.25}$$

where p_d, p_f and k_m^{-1} represent the concentration of dark deGFP, fluorescent deGFP and maturation time, respectively. Table 4.1 shows the parameters used for simulation in Figure 4.2 (B, C, E, F). mRNA inactivation rate and deGFP maturation time which were estimated in section 2.4 were fixed and the size of gene *deGFP* was fixed as well. The three new parameters introduced in the coupled coarse-grained model, $\bar{l}$, $\overline{n1}$ and $\overline{n2}$ were assumed and the same number was used in both sets. Parameter set 1 was estimated based on the literature value and 20∼30-fold dilution factor of cell-free system compared to a living cell [42, 69, 128]. Parameter set 2 was estimated to have a similar order, 10^{-8}∼10^{-9}, as parameter set 1 for the dissociation constant of RNAP-DNA and ribosome-mRNA. In order to simulate the weak promoter, the association constant between RNAP and DNA was changed to 1.2 (from 120) for set 1 and to 1.0 (from 1000)

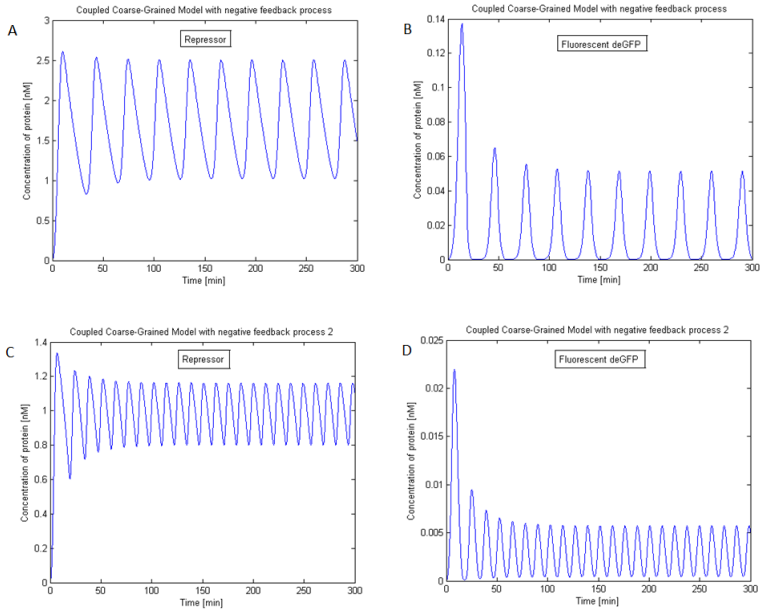

Figure 4.3: The tetracycline negative feedback system described in section 3.2.7 was assumed. 1 nM of the repressor plasmid, $P_{LtetO-1}$–*TetR-degron* and 0.5 nM of the reporter plasmid, $P_{LtetO-1}$–*deGFP-degron* were tested. **(A)** The kinetics of monomeric repressor protein, TetR presented in the system is shown. The ocsillatory behavior was obtained with parameter set 1 in Table 4.1. **(B)** The kinetics of fluorescent reporter protein presented in the system is shown. The oscillatory behavior was obtained with parameter set 1 in Table 4.1. **(C)** The oscillatory behavior of monomeric TetR presented in the system was obteined with parameter set 2 in Table 4.1. **(D)** The oscillatory behavior of fluorescent deGFP presented in the system was obtained with parameter set 2 in Table 4.1.

for set 2. Two parameter sets with a strong promoter show protein saturation (Figure 4.2 (A)) whereas the extension of the linear regime is observed with weak promoter in both parameter sets (Figure 4.2 (D)). Therefore, the protein synthesis saturation phenomenon could be simulated by the coupled coarse-grained model and the model could be extended to explain and pre-estimate the gene expression by a synthetic gene circuit within a cell-free system (Figure 4.2).

Temporal Oscillation There are a couple of ways to involve temporal oscillation including the application of the michaelis-menten term or the application of phase-delay (time retarded) terms [129]. It is also possible to get the oscillatory patterns with

stochastic reaction kinetics [130]. To formulate the oscillation, a simple deterministic approach was used wherein the coupled coarse-grained model was fused with the Goodwin model [131]. Here, the two plasmids negative feedback system using Tet repressor studied in section 3.2.7 was assumed. Thus six equation sets, three equations per each protein synthesis are constructed,

$$\dot{m_t} = \frac{k_{TX}}{l} \cdot DR_p \cdot \frac{1}{1 + k_r \cdot p_{r,t}^c} - h_{TX} \cdot m_t, \tag{4.26}$$

$$\dot{p_r} = \frac{k_{TL}}{n} \cdot mR_i - \frac{h_{TL} \cdot p_r \cdot E_{2\,o}}{H_{TL} + p_r} - k_{T+} \cdot p_r + k_{T-} \cdot p_{r,t}, \tag{4.27}$$

$$\dot{p_{r,t}} = k_{T+} \cdot p_r - k_{T-} \cdot p_{r,t} - \frac{h_{TL} \cdot p_{r,t} \cdot E_{2\,o}}{H_{TL} + p_{r,t}}, \tag{4.28}$$

for the repressor gene and

$$\dot{m} = \frac{k_{TX}}{l} \cdot DR_p^{'} \cdot \frac{1}{1 + k_r \cdot p_{r,t}^c} - h_{TX} \cdot m, \tag{4.29}$$

$$\dot{p_d} = \frac{k_{TL}}{n} \cdot mR_i^{'} - k_m \cdot p_d - \frac{h_{TL} \cdot p_d \cdot E_{2\,o}}{H_{TL} + p_d}, \tag{4.30}$$

$$\dot{p_f} = k_m \cdot p_d - \frac{h_{TL} \cdot p_f \cdot E_{2\,o}}{H_{TL} + p_f}, \tag{4.31}$$

for the reporter gene, where m_t, p_r, $p_{r,t}$, k_r, k_{T+}, k_{T-} and c are the concentration of mRNA of gene *tetR*, the concentration of monomeric TetR, the concentration of tetrameric TetR, the association reaction rate of repressor to operator on DNA, the association rate of tetramerization of TetR, the dissociation rate of monomerization of TetR and the hill coefficient which represents cooperativity between repressors, respectively. $\mathrm{DR}_p^{(')}$ (equation 4.2) and $\mathrm{mR}_i^{(')}$ (equation 4.15) were adjusted by changing the DNA concentration used: 1 nM of repressor plasmid and 0.5 nM of reporter plasmid were tested. It was found that it is necessary to have a sink process for the protein in the system, therefore both proteins, TetR and deGFP, were tagged with the same degron and protein degradation terms were added in the equations. Protein degradation in our cell-free system is mainly due to the protease ClpXP, which has a slower effect than that seen in mRNA inactivation. Protein degradation will depend on the concentration of enzyme, therefore the Michaelis-Menten formation is used. The two parameter sets in Table 4.1 were tested (Figure 4.3). The same parameters used for protein synthesis saturation (except mRNA inactivation rate) simultaneously satisfied

the independent phenomena, oscillatory behavior of gene expression. The parameters used for sink mechanisms were 1/5 [1/min] for mRNA inactivation and 1/8 [1/min] for protein degradation in order to accelerate the biomolecules' turnover. Figure 4.3 shows the oscillatory behavior of two biomolecules, monomeric TetR (A and C) and fluorescent deGFP (B and D). As a result, the simulation with the coupled coarse-grained model promises a suitable mathematical model for gene expression of bacterial systems and shows the feasibility of the temporal pattern which can be executed in a cell-free system.

Why Is It Hard to Perform Oscillation Studies in Cell-Free Context? Investigation suggests that pattern formation involving nonlinear dynamics is obtained by good systematic balance of source and sink processes. It means that the sink process is as important as source process. The cell-free toolbox, which was developed in this thesis, provided such tools for mRNA inactivation and protein degradation. mRNA inactivation is able to be finely tuned depending on the amount of MazF added in the system within the entire DNA concentration range. The protein degradation ratio, however, is not as tunable as the mRNA inactivation. It depends on the number of degrons that one can choose. Even worse, we showed that the degradation of deGFP tagging with SsrA degron was not a linear function of p, the concentration of protein, nor is it an exponential function of p. It appeared as though the degradation of deGFP was independent of p, and was constant [40]. Thus there exists a limitation for the precise tuning of protein degradation.

Indeed, once two plasmids, $\mathrm{P}_{LtetO-1}$–$deGFP$-$YbaQ$ and $\mathrm{P}_{LtetO-1}$–$TetR$ were used in the cell-free system, it was confirmed that the protein degradation process did not behave as well as expected (data not shown). This two plasmids system is auto-regulated, so when gene $TetR$ is expressed, protein production of deGFP-SsrA is repressed. While repression by TetR was shown to be mostly effective at concentration lower than 100 nM, deGFP-SsrA protein, which was supposed to be degraded, was not decreased or was decreased very slowly. This result indicated that building pattern formations would require the new tools for effective protein degradation be added to the present cell-free toolbox, such as new degrons.

Even if the protein degradation process is refined or a library of many degrons is built, pattern formation will still be a challenge. When the parameters in Table 4.1 were adjusted, it became evident that all parameters were just as sensitive as the protein

degradation rate and the combination of all parameters could only produce oscillations which were varied. Therefore, each parameter has to be characterized carefully, so that parameters like maturation time of deGFP protein and adjustable parameters such as protein degradation and mRNA inactivation can be tightly controlled. In other words, the system must have a well-balanced source-sink mechanism because these processes depend not only on synthesis and degradation of mRNA and protein but also, other parameters in the equations such as strength of repressor on operator, association and dissociation reaction rate between monomeric repressor and functional repressor, for example tetrameric TetR. As a conclusion, the results of our modeling indicate that the cell-free toolbox is suitable to study pattern formation, once every parameter is well characterized and a better option for protein degradation is developed.

Chapter 5

Cell-Free synthesis of living entities[1]

5.1 Introduction

It has been shown that a cell-free system as a cell-free toolbox was developed for use in quantitative and synthetic biology. Development of the cell-free toolbox solves two problems: it addresses the lack of both sink processes and transcription units, and it characterizes both transcriptional activation and repression units. Furthermore, elementary synthetic gene circuits were analyzed in batch mode, in a feeding-exchange mode, and with synthetic phospholipid vesicles. Thus the application of a cell-free toolbox has been expanded from the characterization of simple gene circuits and to the pursuit of quantitative analysis using mathematical models, achieving a bottom-up approach for the development of the artificial cell. The next step is to address the following questions: how many genes in a complex network can be expressed in the cell-free toolbox? The *E. coli* genome contains more than 4,000 protein-coding genes (more than 2,500 operons) [132, 133]. Ideally, the *E. coli* cell-free system, derived directly from *E. coli* after removing cell membranes and the native genome has the potential of expressing 4,000 genes involving complex connections between them.

According to the von Neumann's automata scheme, which has been well accepted

[1] **I acknowledge that Paul Jardine's editing help was essential to achieve proper English usage and readability of this chapter.**

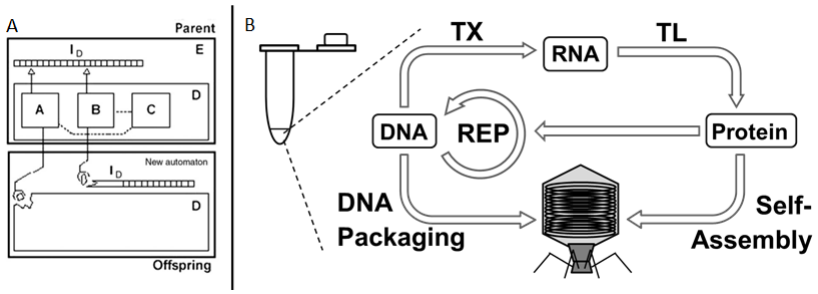

Figure 5.1: **(A)** von Neumann's automata scheme is shown. The parts of this scheme can be interpreted in molecular biology as following: Instructor, I_D is the DNA genome, part A is the TX/TL machinery, part B is the DNA replication machinery, and part C is the regulation. **(B)** Cell-free synthesis in the cell-free toolbox contains transcription and translation as well as DNA replication. In addition to the central dogma of molecular biology, self-organization processes including self-assembly and DNA packaging are involved in the cell-free synthesis of bacteriophage as a self-replicator.

in the field of computer architecture, a body containing well-designed instructions and well-defined machinery can be replicated, producing offspring (Figure 5.1) [34, 35]. The central dogma of molecular biology, confirmed within a cell membrane, is well matched to this concept [134]. This concept, with metabolism and self-sustaining processes in biological systems, can be defined as a primitive life form [135]. A bacterium is one of the simplest organisms that conforms to this definition of life. Based on the hardware developed in this thesis, one can ask if certain types of life forms can be replicated using their DNA genome as instructions. If the absence of a cell membrane in batch mode is problematic, what about performing the reaction in the presence of a liposome? It is apparent that this would likely not be possible using genomes of bacterial size given that no metabolism and no processes that allow the system to be self-sustaining against environment are included.

In contrast to bacteria, viruses are considered not a type of living organism but are rather used to define the edges of life [136]. They are parasitic to their host, and do not metabolize by themselves. They do not have self-sustaining processes but can still evolve by mutation [137]. However, viruses are self-replicate inside a host cell: genes in a genome of a virus are expressed in the host cell, and proteins and copies of replicated genome are self-organized to produce progeny, that are released from the host. As

discussed above, synthesizing living entities such as bacteria in cell-free toolbox would be very difficult, even if it was possible to express more than 4,000 protein-coding genes encoded within more than 2,500 operons. In contrast systems like bacteria, viruses as a self-replicator require only components of the cell in order to produce progeny. Therefore, the cell-free toolbox can be used an experimental platform to study the cell-free synthesis of viruses from their genome.

In addition, synthesis of viruses in the TX-TL cell-free toolbox can address a central question: "how many genes and how complex a network can be expressed in the cell-free toolbox?" The synthesis of bacterial viruses, or bacteriophage, that uses *E. coli* components for self-replication can reveal the capacity for cell-free synthesis in the custom-made, cell-free toolbox developed in this thesis. This research, using genomes that exist in nature, will be a first step toward developing complex synthetic gene networks because the genome selected is a complete gene network. Thus, phage synthesis in cell-free toolbox serve as a standard or toy model of expressing synthetic gene networks in cell-free toolbox. With further development of the cell-free toolbox and DNA programming, the capacity for a cell-free synthesis of a synthetic gene network will mature. Therefore, the cell-free toolbox will be a springboard for well-designed molecular programs.

5.2 Bacteriophage T7 Synthesis

5.2.1 Phage Synthesis

The choice of phages to carry out cell-free synthesis varied from a coliphage to a phage from another bacterium. A test-phage genome was sought that satisfied several conditions. A coliphage was determined to be a good starting point because of experimental platform. Given that the cell-free system hardware was prepared from *E. coli*, so the software originally programmed for that hardware would run better. In addition, a coliphage that had restricted dependence on host would be preferred. As a self-replicator inside a cell, some phages depend on their host significantly, whereas some require just a few components. The cell-free extract protocol was optimized to extract as much of cell cytoplasm as possible, but the extract, a 20 to 30 fold dilution of cytoplasm, likely does not contain all materials present in the cytoplasm. Therefore a phage that

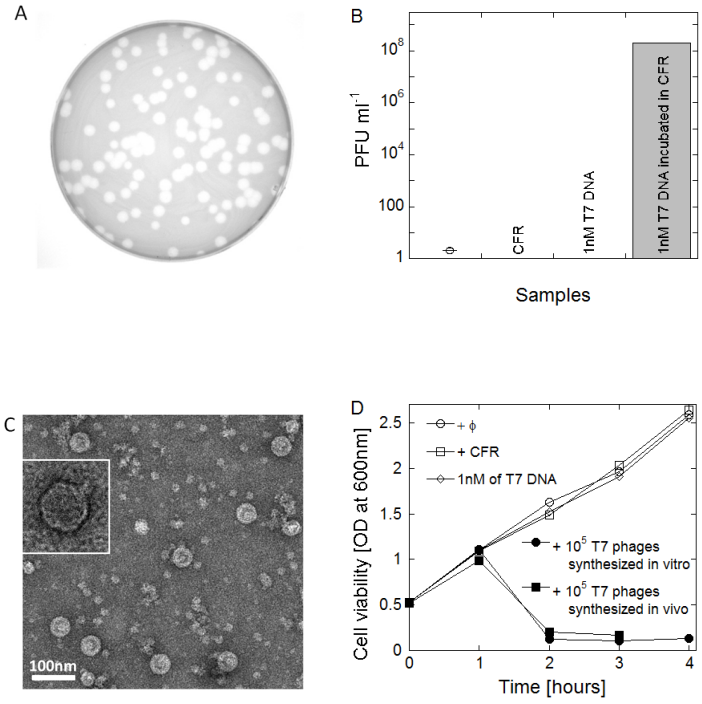

Figure 5.2: **(A)** A plate showing plaques produced by T7 phage synthesized in the cell-free reaction is shown. *E. coli* strain B was used as the host for T7 phage. The cell-free reaction sample (incubated with σ^{70} condition in Table 3.1 at 29 $^{\circ}C$ for 12 hours), was added to a population of *E. coli* strain B and incubated at 37 $^{\circ}C$ for 12 hours. **(B)** Plaque-forming units of samples were counted on plate. A 1 μl sample added to 200 μl overnight cell-culture and 3 ml of 0.55 % LB-agar was mixed at 50 $^{\circ}C$ and was spread on 1.1 % of LB-agar plate: nothing (ϕ), cell-free reaction with 5 nM of P_{70}–*deGFP* (incubation; CFR), cell-free reaction with 1 nM of T7 genome (no incubation; 1 nM T7 DNA) and cell-free reaction with 1 nM of T7 genome (12 hours incubation; 1 nM T7 DNA incubated in CFR). **(C)** T7 phage synthesized in the cell-free reaction with 1 nM of T7 genome observed by transmission electron microscopy (magnification: 59000X) is shown. **Inset**: A blow-up of single T7 phage particle, the capsid (55 nm diameter) and the tail (22 nm length) were visible. **(D)** Cell viability by absorbance at 600 nm is shown. 1 μl samples were inoculated in 5 ml Luria-Bertani medium with 0.5×10^9 of *E. coli* strain B at 37 $^{\circ}C$. The samples of ϕ (circle), +CFR (square), 1nM of T7 DNA (diamond), and 10^5 of T7 phages synthesized *in vitro* (filled circle) were the same samples as in the panel (B). The same quantity of T7 phage synthesized *in vivo* (*E. coli* strain B infection, filled square) were compared.

depends heavily on host factors might perform poorly. Moreover, the capacity of the present custom-made cell-free system must be considered. Based on the single-stage gene expression experiments, the cell-free toolbox yielded 25-30 μM of deGFP in batch mode. If this minimum protein-production threshold inside a cell were considered, a phage genome containing about 50-60 genes would be in a size range that could be supported. Guided by these three conditions, the coliphage T7 genome was selected. Unlike the T7 genome, lambda phage and M13, for example, depend heavily on the host, whereas T4 phage, whose genome size is $\sim$160 kbp with over 200 genes is likely too large [138, 139, 140]. Conversely, bacteriophage T7 is one of the most host-independent phage; only two molecules from the host, thioredoxin and CMP kinase, would be necessary in the cell-free crude extract to produce progeny [141]. The 40 kbp linear T7 dsDNA genome consists of 57 known protein-encoding genes. The T7 genome also encodes its own RNA and DNA polymerases.

To begin, 1nM of linear T7 DNA genome was incubated in the cell-free toolbox with the σ^{70} condition shown in Table 3.1 at 29 $^{\circ}C$ for 12 hours. After incubation, a 1 μl of sample was mixed with 200 μl of an *E. coli* B cell stock and 3 ml of 0.55 % LB-agar at 50 $^{\circ}C$. This mixture was evenly spread on 1.1 % LB-agar plate and incubated at 37 $^{\circ}C$. As shown in Figure 5.2 (A, B), the cell-free toolbox with 1 nM of T7 DNA genome, corresponding to roughly one genome in one *E. coli* cell, after 12 hours incubation, produced 0.1 to 1 billion plaques per 1 ml ($\sim 10^8$ PFU(Plaque Forming Units)/ml). The number of T7 phage synthesized was calculated according to the assumption that one phage forms one plaque, therefore 0.1 to 1 million phages per 1 μl were synthesized from 1 nM of T7 genome with cell-free toolbox. No plaques were observed with control samples. $\sim 10^9$ PFU/ml, the maximum synthetic particle yield, was produced using 2 nM of T7 genome in the reaction, whereas the maximum capacity to synthesize phage from 2 nM of DNA was estimated at $\sim 10^{11}$ PFU/ml. In addition, the activity of genetic information encoded for gene product 5.9 (gp5.9) is implied due to the deficiency of protein GamS, an inhibitor of the RecBCD exonuclease activity that, when added in the cell-free toolbox, protects linearized DNA [142]. The efficiency of phage synthesis from 2 nM of T7 genome was only 10-fold lower without GamS in the cell-free system, conditions that yield no signal from a linear PCR product encoding deGFP, because T7 DNA encodes its own RecBCD inhibitor enzyme, gp5.9

that replaces GamS (data not shown).

The physical shape of synthesized T7 phage from the cell-free toolbox was examined by transmission electron microscopy (TEM) (Figure 5.2 (C)). The typical shape of T7 phage, including head and tail but no visible fibers, is shown in a blow-up of the microscopy image. It was also confirmed that the biological capacity to infect a host cell with synthesized T7 phage within a cell-free system was similar to *in vivo* produced phage (Figure 5.2 (D)). Optical density (OD) of cell cultures at 600 nm wavelength was measured at 0.5 OD, corresponding to 5×10^8 cells/ml, at time=0, and the OD was measured every 30 min, which corresponds to an average generation time for *E. coli* as well as the time required for T7 phage to produce around 100 progeny, escape from the host cell, and infect neighboring cells [141]. Therefore the number of *E. coli* cells and T7 phages increases 4-fold and 10^4-fold every hour, respectively. No interruption to cell growth in the negative controls was observed, but cells inoculated with either T7 DNA seeded cell-free system materials or phage produced from cells cleared between 1 and 2 hours. Consequently, synthesis of viable T7 phage involving transcription and translation, as well as self-assembly and DNA packaging, occurs in the cell-free toolbox.

5.2.2 Phage DNA Replication

Next, characterizations of the DNA replication process was conducted. Experiments were performed where the cell-free reaction was supplemented with 0.5 mM of dNTPs (0.5 mM each of four deoxyribonucleotide-triphosphate bases)(Figure 5.3). First, phage synthesis as a function of the input genome concentration in the range of 0.1 nM and 1 nM was determined and compared with and without dNTPs (Figure 5.3 (A)). A nonlinear relationship between the numbers of T7 phage synthesized and the amount of T7 DNA added was observed within the entire input concentration range. T7 phage synthesis with dNTPs added was one to two orders of magnitude higher than without dNTPs. A higher efficiency of phage synthesis relative to input DNA was obtained at lower input concentration: a 10-fold lower input DNA, reduced from 1 nM to 0.1 nM yielded $10^7\sim10^8$-fold lower phage. The kinetics of phage synthesis was also compared with and without 0.5 mM of dNTPs addition (Figure 5.3 (B)). Overall, more phage synthesis was still observed with 0.5 mM of dNTPs than without dNTPs at any time point.

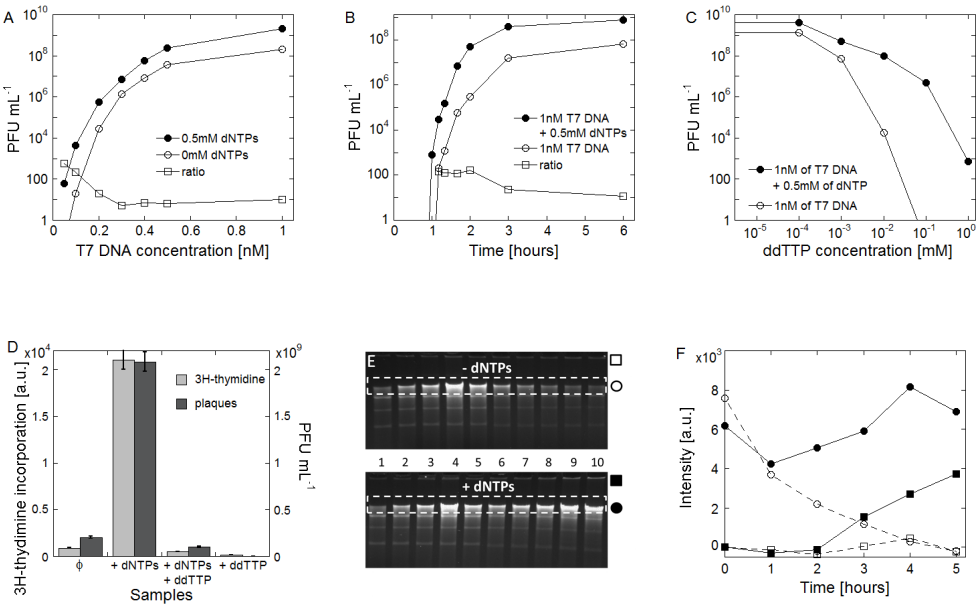

Figure 5.3: Characterization of the T7 phage cell-free synthesis and of the DNA replication are shown. **(A)** Number of T7 phage synthesized in the cell-free reaction (incubated for 12 hours), with or without the addition of 0.5mM of each dNTP (each of the four DNA bases), as a function of the T7 DNA genome concentration was measured by plaque-forming units. **(B)** Number of T7 phage synthesized in the cell-free reaction (with 1 nM of T7 DNA genome), with or without the addition of 0.5mM of each dNTP, as a function of time was measured by plaque-forming units. **(C)** Number of T7 phage synthesized in the cell-free reaction (incubated with 1 nM of T7 DNA genome for 12 hours), with or without the addition of 0.5mM of each dNTP, as a function of the concentration of the DNA replication inhibitor (ddTTP) was measured by plaque-forming units. **(D)** 3H-thymidine (2.5 μM) incorporation and number of plaques were measured (the cell-free reaction was incubated with 1 nM of T7 DNA genome for 12 hours) for four different samples. ϕ: nothing added, dNTPs: addition of 0.5 mM of each dNTP, dNTPs + ddTTP: addition of 0.5 mM of each dNTP and 25 μM of ddTTP, and ddTTP: addition of 0.5 mM of ddTTP. **(E)** Gel electrophoresis of the T7 DNA present in the cell-free reaction, with (bottom gel) or without (top gel) the additon of 0.5 mM of each dNTP, as a function of time is shown. Lanes 1 to 4: 0, 0.5, 1, and 2 nM of T7 DNA added to the cell-free reaction (no incubation). Lanes 5 to 10: T7 DNA incubated in the cell-free reaction measured every hour (1 nM of T7 DNA used; lane 5 to 10: 0 to 5 hours). The white dotted frame indicates the T7 DNA band. Symbols in right are used in the panel (F). **(F)** Kinetics of T7 DNA genome present in the cell-free reaction, with or without 0.5 mM of each dNTP, were constructed from the gels shown in the panel (E). The intensity of the T7 DNA band (monomeric genomes) and the intensity of the DNA present into the wells (concatemers) were measured.

Further analyses of whether T7 genome replication occurs simultaneously with phage synthesis were done using dideoxythymidine triphosphate (ddTTP), a chain-terminating, DNA replication inhibitor, and 3H-thymidine, radioactive material that can be incorporated into newly synthesized. Once ddTTP is incorporated into replicating DNA, DNA synthesis ceases due to chain termination. Phages cannot be synthesized from such incomplete genomes. Addition of ddTTP resulted in the most drastic differences between samples prepared with and without dNTP (Figure 5.3 (C)). For example, 0.1 mM of ddTTP completely inhibited phage synthesis without dNTP whereas $\sim 10^7$ phages were observed with cell-free toolbox supplemented with 0.5 mM of dNTP and 0.1 mM of ddTTP. In parallel, the amount of newly replicated DNA that incorporated 3H-thymidines was detected quantified by scintillation counting following TCA precipitation. The amount of radionucleotide incorporation into DNA was in relative agreement with the amount of phage synthesized (Figure 5.3 (D)).

Direct measurement of the amount of T7 DNA replication was determined using gel-electrophoresis. The T7 DNA synthesis rate without and with dNTP addition was measured after gel purification (Figure 5.3 (E, F)). The first 4 lanes contain DNA quantification standard (0, 0.5, 1 and 2 nM, respectively). Additional lanes include the cell-free toolbox before incubation and after 1, 2, 3, 4 and 5 hours incubation (lanes 5-10). T7 genome-sized DNA was degraded over time without dNTPs in the reaction, reaching background levels after 5 hours even though 3.3 μM of GamS was added to reactions. With dNTP supplementation, however, DNA was degraded during the first hour but then increased, suggesting that DNA synthesis rebounds and overcomes the degradation reaction. Beginning with 1 nM of DNA with and without added dNTP yielded 2.2 nM and 0.1nM after 15 hours incubation, respectively. Moreover, longer DNA that cannot enter the 0.8 % agarose gel was observed in dNTP sample. Such long DNA may be concatamers of T7 DNA that form during DNA synthesis [143]. As a result, DNA synthesis, including repair and replication, as well as phage synthesis involving transcription, translation, self-assembly, and the DNA packaging process, occurred concurrently within cell-free toolbox.

Host dependence of T7 phage synthesis, especially genome replication, was also tested. The *E. coli* component thioredoxin is used for T7 DNA genome replication *in vivo*. To test its role in the cell-free system, cell-free systems were prepared from two

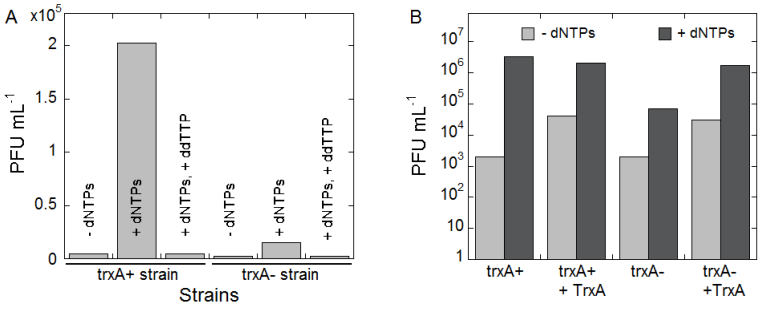

Figure 5.4: The effect of thioredoxin on phage synthesis is shown. **(A)** Two *E. coli* strains with and without thioredoxin, *E. coli BL21 trxA+* and *E. coli BL21 trxA-*, were used to prepare cell-free extract for the study of thioredoxin dependency of DNA replication. Samples with 1 nM of T7 genome were incubated in each cell-free reaction without dNTPs, with 0.5 mM of each dNTP, or with 0.5 mM of each dNTP and 25 μM of ddTTP at 29 $^{\circ}C$ for 12 hours. **(B)** T7 phage in cell-free reactions with or without 3 μM TrxA (synthesized in cell-free reaction separately) was synthesized and compared with or without 0.5 mM of each dNTP. All experiments were done with σ^{70} condition in Table 3.1.

E. coli strains: BL21 notated as *trxA+* and BL21(-) notated as *trxA-*. The *trxA-* strain was engineered to replace thioredoxin with kanamycin. The cell-free systems prepared from these two stains were not as efficient as regular cell-free toolbox in terms of phage synthesis overall, but phage yields were high enough to test and compare phage synthesis between them. The large difference between phage synthesis comparing reactions without and with dNTPs in *trxA+* BL21 cell-free system was observed, however only a small increase with dNTPs was seen with the *trxA-* BL21 cell-free system (Figure 5.4 (A)). Overall, 10-fold more phages were produced with the *trxA+* BL21 cell-free system compared to the *trxA-* BL21 cell-free system. Further study is required to explain the increase in *trxA-* BL21 cell-free system with dNTPs should follow given that it was supposed to be the same as without dNTPs. When the protein TrxA expressed with P_{70}–*trxA* in the regular cell-free toolbox, was added either the *trxA+* or *trxA-* BL21 cell-free system, phage synthesis became the same both with and without dNTPs addition (Figure 5.4 (B)).

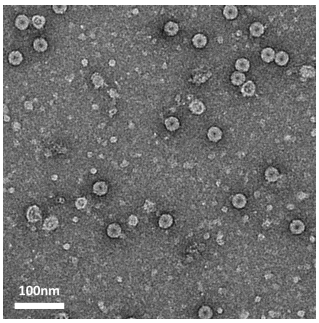

Figure 5.5: An additional molecular program was executed with the cell-free reaction. 30 nM of double stranded $\Phi X174$ genome was added in the cell-free reaction with the σ^{70} condition in Table 3.1. Synthesized $\Phi X174$ phages were observed by transmission electron microscopy.

5.2.3 Beyond T7 DNA Programming

As shown with T7 phage synthesis and T7 genome replication, the T7 genome can serve as an excellent toy model to perform cell-free synthesis in custom-made TX-TL cell-free toolbox. Even though the cell-free system is diluted 20∼30-fold relative to *in vivo*, the genetic program of the T7 genome runs well; It is not only synthesizing T7 phage, whose genome has 57 protein-encoding genes, but also targeting of performing its molecular program. In this sense, a simple question arose: is the T7 genome and T7 phage synthesis specifically designed for its success? To extend the concept of molecular programming, an additional genome found in nature was tested. The ssDNA phage ΦX174 genome size is ∼5.4 kbp and contains 11 known protein-encoding genes. *In vitro* ΦX174 phage synthesis has already done using a difference approach with purified and preformed components [144]. However, the main idea of our approach is to run a software program in the TX-TL cell-free toolbox hardware. Figure 5.5 taken by TEM showed that ΦX174 phage particles were synthesized in the cell-free toolbox upon addition of ΦX174 DNA.

Chapter 6

Conclusion[1]

The cell-free system is a well-developed experimental platform. It has been mostly used for large-scale recombinant protein production such as biotechnology and proteomic applications. In this present biological era, synthetic and systems biology connecting biology, chemistry, physics, engineering and other research fields, has come to the forefront of research in the 21st century. Thus there were several efforts to use the cell-free system as the experimental platform to carry out synthetic gene circuits similar to using a breadboard for custom-built electric circuits. However, those attempts ended up facing the limitations of a conventional cell-free system. Due to the main aim of cell-free system to date, which has been developed in the direction of larger protein production, the most powerful transcriptional machinery was sought and the sink mechanisms were eliminated. However, a large variety of transcriptional machinery and easily tunable sink mechanisms are required to reconstitute a biological network. Thus it is necessary to overcome the insufficient sink processes and expand the tools to build a successful synthetic circuit. This was the inspiration for this thesis.

In this thesis, a custom-made, cell-free system was developed for the research of synthetic biology. Accelerating mRNA turnover with precise control could be carried out by using the toxin MazF, while protein degradation was also able to be introduced by tagging degrons at N- or C- terminal of a protein. Moreover, all seven *E. coli* sigma factors containing 'AND' logic gate expanded the library of available transcriptional machinery.

[1] **I acknowledge that Jonathan Gapp's editing help was essential to achieve proper English usage and readability of this chapter.**

Despite these advantages, protein production with our cell-free system was comparable to a conventional cell-free system whose purpose was the production of large amounts of protein. Our cell-free system was characterized with a couple of phenomenological aspects which play an important role in constructing a gene circuit. These phenomenological aspects included orthogonality of transcriptional motif, passive regulation by competition and importance of mRNA inactivation and protein degradation. This thesis also covered the transcriptional repression units which make regulation of a genetic network possible in addition to studying the transcriptional activation units which allow information transmission to proceed. A well-characterized cell-free system, referred to as a cell-free toolbox, is now available to anybody who is interested in constructing a synthetic gene circuit *in vitro*.

This unique cell-free system also opens the possibility of developing an experimental platform. It was shown that a cell-free toolbox could perform the experiment in feeding-exchange mode as well as an artificial cell system. Feeding-exchange mode is important both in large-scale structure and in synthetic phospholipid vesicle because the kinetics of gene expression lasts longer and final output production is higher. This is especially crucial for the creation of a gene circuit contained in a cell-sized liposome; this approach to gene circuit development is the bottom-up approach to understanding life. Through this approach we can better understand life that exists in nature as well as create new life that might be useful in curing diseases and addressing other engineering problems in a biological context. Moreover, the cell-free synthesis shown in this thesis shows the widened scope of use for a cell-free system, both in application and for fundamental research. It is demonstrated that at least tens of genes were functional within complex networks. Self-assembly of procapsid, DNA packaging in procapsid and even DNA replication were observed in addition to the central dogma of molecular biology. The result of cell-free synthesis include many complex mechanisms and demonstrate the vast potential of what a cell-free system can accomplish. Thus, when complete molecular programing is ready to be tested, the cell-free toolbox is here and ready to be used!

In addition to the experimental developments, formulating the mathematical model for anything from a simple gene circuit to complex genetic network is of interest. A simple coupled coarse-grained model was developed in order to explain the coupled processes between transcription and translation in a bacterial system. It seems that

this model is promising in the explanation and description of a gene circuit in an *E. coli*-based cell-free system. As mentioned, each parameter in the equation has to be characterized accurately in order to predict the behavior of the gene circuit constructed. Thus, development of a mathematical model would do well to keep pace with experimental progress in determining the important parameters.

Finally, it is important to emphasize that the results reported here are not the final product. The cell-free toolbox is an open breadboard to be developed and modified for the special purposes of experimentalists and theorists alike. Any kind of development can improve the quality of the cell-free toolbox, from expanding the regulatory aspects to developing a well-designed molecular program. New technique to produce active phospholipid membrane, various and fine-tunable sink processes for protein, more diverse transcription and translation machinery, etc. will be just a few of the aspects of the present cell-free toolbox that we can try to enhance.

References

[1] Leduc, S. (1910) Thĕorie physico-chimique de la vie et gĕnĕrations spontanĕes. *Paris, Poinat, A. Ĕditeur*; Leduc, S. (1912) La biologie synthĕtique, ĕtude de biophysique. *Paris, Poinat, A. Ĕditeur*

[2] Nathans, D., Arber, W. and Smith, H. (1978) For the discovery of restircition enzymes and their application to problems of molecular genetics. The Nobel Prize in physiology or Medicine, The Nobel Foundation

[3] Bio FAB group (2006) Engineering life: building a fab for biology. *Science American* **294** (6): 44-51

[4] Prunick, P. and Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10** (6): 410-422

[5] Elowitz, M. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403** (6767): 335-338

[6] Gardner, T., Cantor, C. and Collins, J. (2000) Construction of a genetic toggle switch in *Escherichia coli. Nature* **403** (6767): 339-342

[7] Alon, U., Barkai, N. and Notterman, D. A. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Nat. Aca. Sci. USA* **96** (12): 6745-6750

[8] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002) Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298** (5594): 824-827

[9] Shen-Orr, S., Milo, R., Mangan, S. and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature genetics* **31** (1): 64-68

[10] Alon, U. (2006) An Introduction to Systems Biology - Design Principles of Biological Circuits (1st edition). *Chapman and Hall/CRC*

[11] Basu, S., Gerchman, Y., Collins, C., Arnold, F. and Weiss, R. (2005) A synthetic multicellular system for programmed pattern formation. *Nature* **434** (7037): 1130-1134

[12] Levskaya, A., Chevalier, A., Tabor, J., Simpson, Z., Lavery, L., Levy, M., Davidson, E., Scouras, A., Ellington, A., Marcotte, E. and Voigt, C. (2005) Synthetic biology: Engineering *Escherichia coli* to see light. *Nature* **438** (7067): 441-442

[13] Guido, N., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C., Elston, T. and Collins, J. (2006) A bottom-up approach to gene regulation. *Nature* **439** (7078): 856-860

[14] Stricker, J., Cookson, S., Bennett, M., Mather, W., Tsimring, L. and Hasty, J. (2008) A fest, robust and tunable synthetic gene oscillator. *Nature* **456** (7221): 516-519

[15] Tamsir, A., Tabor, J. and Voigt, C. (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469** (7329): 212-215

[16] Winfree, E., Liu, F., Wenzler, L. and Seeman, N. (1998) Design and self-assembly of two-dimensional DNA crystals. *Nature* **394** (6693): 539-544

[17] Hasty, J., Pradines, J., Dolnik, M. and Collins, J. (2000) Noise-based switches and amplifiers for gene expression. *Proc. Nat. Aca. Sci. USA* **97** (5): 2075-2080

[18] Forster, A. and Church, G. (2006) Synthetic biology projects *in vitro*. *Genome research* **17** (1): 1-6

[19] Montagne, K., Plasson, R., Sakai, Y., Fujii, T. and Rondelez, Y. (2011) Programming an *in vitro* DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**: 466

[20] Zhang, D., Turberfield, A., Yurke, B. and Winfree, E. (2007) Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science* **318** (5853): 1121-1125

[21] Jewett, M. and Forster, A. (2010) Update on designing and building minimal cells. *Curr. Opin. Biotechnol.* **21** (5): 697-703

[22] Isaacs, F., Dwyer, D. and Collins, J. (2006) RNA synthetic biology *Nat. Biotechnol.* **24** (5): 545-554

[23] Seelig, G., Soloveichik, D. Zhang, D and Winfree, E. (2006) Enzyme-free nucleic acid logic circuits. *Science* **314** (5805): 1585-1588

[24] Macdonald, J., Li, Y., Sutovic, M., Lederman, H., Pendri, K., Lu, W., Andrews, B., Stefanovic, D. and Stojanovic, M. (2006) Medium scale integration of molecular logic gates in an automation. *Nano Lett.* **6** (11): 2598-2603

[25] Qian, L., Winfree, E. (2011) A simple DNA gate motif for synthesizing large-scale circuits. *J. R. Soc., Interface* **8** (62): 1281-1297

[26] Qian, L., Winfree, E. and Bruck, J. (2011) Neural network computation with DNA strand displacement cascades. *Nature* **475** (7356): 368-372

[27] Soloveichik, D., Seelig, G. and Winfree, E. (2010) DNA as a universal substrate for chemical kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **107** (12): 5393-5398

[28] Rothemund, P. (2006) Folding DNA to create nanoscale shapes adn patterns. *Nature* **440** (7082): 297-302

[29] Venkataraman, S., Dirks, R., Rothemund, R., Winfree, E. and Pierce, N. (2007) An autonomous polymerization motor powered by DNA hybridization. *Nat. Nanotechnol.* **2** (8): 490-494

[30] Barish, R., Schulman, R., Rothemund, P. and Winfree, E. (2009) An information-bearing seed for nucleating algorithmic self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **106** (15): 6054-6059

[31] Kim, J., White, K. and Winfree, E. (2006) Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2**: 68

[32] Kim, J. and Winfree, E. (2011) Synthetic *in vitro* transcriptional oscillators. *Mol. Syst. Biol.* **7**: 465

[33] Hodgman, C. and Jewett, M. (2011) Cell-free synthetic biology: Thinking outside the cell. *Metab. Eng.* DOI: 10.1016/j.ymben.2011.09.002

[34] Von Neumann, J. (1951) The general and logical theory of automata. *Cerebral Mechanisms in Behavior: The Hixon Symposium, ed LA Jeffress, Wiley, New York*

[35] Von Neumann, J. and Burks, A. W. (1966) Theory of self-reproducing automata. *University of Illinois Press*

[36] Noireaux, V., Maeda, Y. and Libchaber, A. (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl. Acad. Sci. U.S.A.* **108** (9):3473-3480

[37] Shin, J. and Noireaux, V. (2010) Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* **4** (8)

[38] Shin, J. and Noireaux, V. (2010) Study of messenger RNA inactivation and protein degradatino in an *Escherichia coli* cell-free expression system. *J. Biol. Eng.* **4** (9)

[39] Chalmeau, J., Monina, N., Shin, J., Vieu, C. and Noireaux, V. (2011) α-Hemolysin pore formation into a supported phospholipid bilayer using cell free expression. *Biochim. Biophys. Acta* **1808** (1): 271-278

[40] Karzbrun, E., Shin, J., Bar-Ziv, R. and Noireaux, V. (2011) Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System *Phys. Rev. Lett.* **106**, 048104

[41] Maeda, Y., Nakadai, T., Shin, J., Uryu, K., Noireaux, V. and Libchaber, A. (2012) Assembly of MreB Filaments on Liposome Membranes: A Synthetic Biology Approach. *ACS Synth. Biol.* **1** (2): 53-59

[42] Shin, J. and Noireaux, V. (2012) An *E. coli* Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* **1** (1): 29-41

[43] Shin, J., Jardine, P. and Noireaux, V. (2012) Genome Replication, Synthesis, and Assembly of the Bacteriophage T7 in a Single Cell-Free Reaction. *ACS Synth. Biol.* (): - just published online (As Soon As Publishable)

[44] Katzen, F., Chang, G. and Kudlicki, W. (2005) The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23** (3): 150-156

[45] Swartz, J. (2006) Developing cell-free biology for industrial applications. *J. Ind. Microbiol. Biotechnol.* **33** (7): 476-485

[46] He, M. (2008) Cell-free protein synthesis: applications in proteomics and biotechnology. *N. Biotechnol.* **25** (2-3): 126-132

[47] Shimizu, Y., Kanamori, T. and Ueda, T. (2005) Protein synthesis by pure translation systems. *Methods* **36** (3): 299-304

[48] Patnaik, R. and Swartz, J. (1998) *E. coli*-Based *In Vitro* Trnascription/Translation: *In Vivo*-Specific Synthesis Rates and High Yields in a Batch System *Bio Techniques* **24** (5):862-868

[49] Sitaraman, K. and Chatterjee, D. (2009) High-Throughput Protein Expression Using Cell-Free System. *High Throughput Protein Expression and Purification* **498** :229-244

[50] Robert Kohler (1972) The background to Eduard Buchner's discovery of cell-free fermentation. *Journal of the History of Biology* **4** (1):35-61

[51] Hoagland, M., Stephenson, M., Scott, J., Hecht, L. and Zamecnik, P. (1958) A Soluble Ribonucleic Acid Intermediate in Protein Synthesis. *J. Biol. Chem.* **231** (1):241-257

[52] Nirenberg, M. and Matthaei, J. (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **47** (10):1588-1602

[53] Nirenberg, M. (2004) Historical review: Deciphering the genetic code - a personal account. *Trends in Biochemical Sciences* **29** (1):46-54

[54] Chambers, D. and Zubay, G. (1969) The stimulatory effect of cyclic adenosine 3'5'-monophosphate on DNA-directed synthesis of beta-galactosidase in a cell-free system. *Proc. Natl. Acad. Sci. U.S.A.* **63** (1):118-122

[55] Zalkin, H., Yanofsky, C. and Squires, C. (1974) Regulated *in vitro* synthesis of *Escherichia coli* tryptophan operon messenger ribonucleic acid and enzymes. *J. Biol. Chem.* **249** (2):465-475

[56] Zubay, G. (1973) *In vitro* synthesis of protein in microbial systems. *Annu. Rev. Genet.* **7** :267-287

[57] Pratt, J. (1984) Coupled transcription-translation in prokaryotic cell-free systems. *In Transcription and Translation: A Practical Approach* Edited by: Hames, B. and Higgins, J.; IRL Press: Oxford U.K.:179-209

[58] Nevin, D. and Pratt, J. (1991) A coupled *in vitro* transcription-translation system for the exclusive synthesis of polypeptides expressed from the T7 promoter. *FEBS Letters* **291** (2):259-263

[59] Carlson, E., Gan, R., Hodgman, C. and Jewett, M. (2011) Cell-free proteoin synthesis: Applications come of age. *Biot. Adv.* DOI: 10.1016/j.biotechadv.2011.09.016

[60] Liu, D., Zawada, J. and Swartz, J. (2005) Streamlining *Escherichia coli* S30 extract prepaartion for economical cell-free protein synthesis. *Biotechnol. Prog.* **21** (2):460-465

[61] Kigawa, T., Yabuki, T., Matsuada, N., Matsuda, T., Nakajima, R., Tanaka, A. and Yokoyama, S. (2004) Preparation of *Escherichia coli* cell extract for highly productive cell-free protoein expression. *J. Struct. Funct. Genomics* **5** (1-2):63-68

[62] Kim, T., Keum, J., Oh, I., Choi, C., Park, C. and Kim, D. (2006) Simple procedures for the construction of a robust and cost-effective cell-free protein system. *J. Biotechnol.* **126** (4):554-561

[63] Noireaux, V., Bar-Ziv, R. and Libchaber, A. (2003) Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. USA* **100** (22):12672-12677

[64] Isalan, M., Lemerle, C. and Serrano, L. (2005) Engineering gene networks to emulate Drosophila embryonic pattern formation. *PLoS Biol.* **3** (3):e64

[65] Noireaux, V. and Libchaber, A. (2004) A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. USA* **101** (51):17669-17674

[66] Ishikawam, K., Sato, K., Shima, Y, Urabe, I. and Yomo, T. (2004) Expression of a cascading genetic network within liposomes. *FEBS Lett.* **576** (3):387-390

[67] Alberts, B., Johnson, A., Lewis, J., Raff, M. and Roberts, K. (2002) Molecular Biology of the cell, 4th edition *Garland Science*

[68] Elewitz, M., Surette, M., Wolf, P., Stock, J. and Leibler, S. (1999) Protein mobility in the cytoplasm of *Escherichia coli*. *Journal of bacteriology* **181** (1):197-203

[69] Bremer, H. and Dennis, P.(1987) Modulation of chemical composition and other parameters of the cell by growth rate in *Escherichia and Salmonella: Cellular and Molecular Biology ASM Press* Washington D.C.:1527-1542

[70] Kim, D. and Swartz, J. (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng.* **74** (4):309-316

[71] Sitaraman, K., Esposito, D., Klarmann, G., Le Grice, S., Hartley, J. and Chatterjee, D. (2004) A novel cell-free protein synthesis system. *J. Biotechnol.* **110** (3):257-263

[72] Matsuda, T., Koshiba, S., Tochio, N., Seki, E., Iwasaki, N., Yabuki, T., Inoue, M., Yokoyama, S. and Kigawa, T. (2007) Improving cell-free protein synthesis for stable-isotope labelling. *J. Biomol. NMR* **37** (3):225-229

[73] Craggs, T. (2009) Green Fluorescent protein: structure, folding and chromophore maturation. *Chemical Society reviews* **38** (10):2865-2875

[74] Ataei, F., Hosseinkhani, S. and Khajeh, K. (2009) Limited proteolysis of luciferase as a reporter in nanosystem biology: a comparative study. *Photochem. Photobiol.* **85** (5):1162-1167

[75] Iskakova, M., Szaflarski, W., Dreyfus, M., Remme, J. and Nierhaus, K. (2006) Troubleshooting coupled *in vitro* transcription-translation system derived from *Escherichia coli* cells: synthesis of high-yield fully active proteins. *Nucleic Acids Res.* **34** (19):e135

[76] de Boer, H., Comstock, L. and Vasser, M. (1983) The tac promoter: a fucntional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. USA* **80** (1):21-25

[77] Olins, P., Devine, C., Rangwala, S. and Kavka, K. (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene* **73** (1):227-235

[78] KL740 was purchased from The Coli Genetic Stock Center at Yale

[79] Larson, M., Greenleaf, W., Landick, R. and Block, S. (2008) Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132** (6):971-982

[80] Rondelez, Y. (2011) Breaking down complexity. *Physics* **4**

[81] Zhang, J., Zhang, Y. and Inouye, M. (2003) Characterization of the interactions within the mazEF addiction module of *Escherichia coli*. *J. Biol. Chem.* **278** (34):32300-32306

[82] Kamada, K., Hanaoka, F. and Burley, S. (2003) Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. *Mol. Cell* **11** (4):875-884

[83] Dougan, D., Mogk, A. and Bukau, B. (2002) Protein folding and degradation in bacteria: To degrade or not to degrade? That is the question. *Cellular and Molecular Life Sciences* **59** (10):1607-1616

[84] Schmidt, R., Bukau, B. and Mogk, A. (2009) Principles of general and regulatory proteolysis by AAA+ proteases in *Escherichia coli*. *Res. Microbiol.* **160** (9):629-636

[85] Flynn, J., Neher, S., Kim, Y., Sauer, R. and Baker, T. (2003) Proteomic discovery of cellular substrates of teh ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11** (3):671-683

[86] Anfinsen, C. (1972) The formation and stabilization of prtoein structure. *Biochem. J.* **128** (4):737-749

[87] Anfinsen, C. (1973) Principles that Govern the Folding of Protein Chains. *Science* **181** (4096):223-230

[88] van den Berg, B., Ellis, R. and Dobson C. (1999) Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* **18** (24):6927-6933

[89] Ellis, R. (2006) Molecular chaperones: assisting assembly in addition to folding *Trends in Biochemical Sciences* **31** (7):395-401

[90] Raid, B. and Flynn, G. (1997) Chromophore formation in green fluorescent protein. *Biochemistry* **36** (22):6786-6791

[91] Megerle, J., Fritz, G., Gerland, U., Jung, K. and Radler, J. (2008) Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.* **95** (4):2103-2115

[92] Macdonald, P., Chen, Y. and Mueller, J. (2011) Chromophore maturation and fluorescence fluctuation spectroscopy of fluorescent proteins in a cell-free expression system. *Analytical Biochemistry* **421** (1):291-298

[93] Knapinska, A., Irizarry-Barreto, P., Adusumalli, S., Androulakis, I. and Brewer, G. (2005) Molecular Mechanisms Regulating mRNA Stability: Physiological and Pathological Significance. *Current Genomics* **6** (6):471-486

[94] Anderson, K. and Dunman, P. (2009) Messenger RNA Turnover Processes in *Escherichia coli*, Bacillus subtillis, and Emerging Studies in Staphylococcus aureus. *Int. J. Microbiol.* **2009** :1-15

[95] Selinger, D., Saxena, R., Cheung, K., Church, G. and Rosenow, C.(2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* **13** (2):216-223

[96] Mathews, D. and Durbin, R. (1990) Tagetitoxin inhibits RNA synthesis directed by RNA polymerases from chloroplasts and *Escherichia coli. J. Biol. Chem.* **265** (1):493-498

[97] Buxboim, A., Daube, S. and Bar-Ziv, R. (2008) Synthetic gene brushes: a structure-function relationship. *Mol. Syst. Biol.* **4** :181

[98] Spirin, A., Baranov, V., Ryabova, L., Ovodov, S. and Alakhov, Y. (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242** (4882):1162-1164

[99] Spirin, A. (2004) High-throughput cell-free systems for synthesis of functionally active proteins. *Trends in Biotechnol.* **22** (10):538-545

[100] Deamer, D. (2005) A giant step towards artificial life? *Trends in Biotechnol.* **23** (7):336-338

[101] Kuruma, Y., Stano, P., Ueda, T. and Luisi, P. (2009) A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *BBA-Biomembranes* **1788** (2):567-574

[102] Bedau, M., Church, G., Rasmussen, S., Caplan, A., Benner, S., Fussenegger, M., Collins, J. and Deamer, D. (2010) Life after the synthetic cell. *Nature* **465** (7297):422-424

[103] Enz, S., Braun, V. and Crosa, J. (1995) Transcription of the region encoding the ferric dicitrate-transport system in *Escherichia coli*: similarity between promoters for fecA and for extracytoplasmic function sigma factors. *Gene* **163** (1):13-18

[104] Lipinska, B., Sharma, S. and Georgopoulos, C. (1988) Sequence analysis and regulation of the htrA gene of *Escherichia coli*: a sigma 32-independent mechanism of heat-inducible transcription. *Nucleic Acids Res.* **16** (21):10053-10067

[105] Arnosti, D. and Chamberlin, M. (1989) Secondary Sigma-Factor Controls Transcription of Flagellar and Chemotaxis Genes in Escherichia-Coli. *Proc. Natl. Acad. Sci. USA* **86** (3):830-834

[106] Wang, Y. and deHaseth, P. (2003) Sigma 32-dependent promoter activity *in vivo*:Sequence determinants of the groE promoter. *Journal of Bacteriology* **185** (19):5800-5806

[107] Yim, H., Brems, R. and Villarejo, M. (1994) Molecular Characterization of the Promoter of Osmy, an rpoS-Dependent Gene. *Journal of Bacteriology* **176** (1):100-107

[108] Reitzer, L. and Magasanik, B. (1985) Expression of glna in *Escherichia coli* is regulated at tandem promoters. *Proc. Natl. Acad. Sci. USA* **82** (7):1979-1983

[109] Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F., Kerner, M. and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* sytosol. *BMC Genomics* **9** :102

[110] Glaser, B., Bergendahl, V., Anthony, L., Olson, B. and Burgess, R. (2009) Studying the salt dependence of the binding of sigma70 and sigma32 to core RNA polymerase using luminescence resonance energy transfer. *PLoS ONE* **4** (8):e6490

[111] Maeda, H., Fujita, N. and Ishihama, A. (2000) Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* **28** (18):3497-3503

[112] Levchenko, L., Seidel, M., Sauer, R. and Baker, T. (2000) A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289** (5488):2354-2356

[113] Ochs, M., Angerer, A., Enz, S. and Braun, V. (1996) Surface signaling in transcriptional regulation of the ferric citrate transport system of *Escherichia coli*: mutational analysis of the alternative sigma factor FecI supports its essential role in fec transport gene transcription. *Mol. Gen. Genet.* **250** (4):455-465

[114] Gruber, T. and Gross, C. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57** (1):441-466

[115] Farewell, A., Kvint, K. and Nystrom, T. (1998) Negative regulation by RpoS: a case of sigma factor competition. *Mol. Microbiol.* **29** (4):1039-1051

[116] Ishihama, A. (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Annu. Rev. Microbiol.* **54** (1):499-518

[117] Atkinson, M., Pattaramanon, N. and Ninfa, A. (2002) Governor of the glnAp2 promoter of *Escherichia coli*. *Mol. Microbiol.* **46** (5):1247-1257

[118] Feng, J., Atkinson, M., McCleary, W., Stock, J., Wanner, B. and Ninfa, A. (1992) Role of phosphorylated metabolic intermediates in the regulation of glutamine synthetase synthesis in *Escherichia coli*. *J. Bacteriol.* **174** (19):6061-6070

[119] Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25** (6):1203-1210

[120] Lobell, R. and Schleif, R. (1990) DNA looping and unlooping by AraC protein. *Science* **250** (4980):528-532

[121] Schleif, R. (2010) AraC protein, regulation of the I-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34** (5):779-796

[122] Lichenstein, H., Hamilton, E. and Lee, N. (1987) Repression and catabolite gene activation in the arabad operon. *J. Bacteriol.* **169** (2):811-823

[123] Hillen, W., Gatz, C., Altschmied, L., Schollmeier, K. and Meier, I. (1983) Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetics investigation of the regulatory reactions. *J. Mol. Biol.* **169** (3):707-721

[124] Biliouris, K., Daoutidis, P. and Kaznessis, Y. (2011) Stochastic simulations of the tetracycline operon. *BMC Sys. Biol.* **5** (9)

[125] Stano, P. and Luisi, P. (2010) Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chem. Commun. (Cambridge)* **46** (21):3639-3653

[126] Kremling, A. (2007) Comment on Mathematical Models Which Describe Transcription and Calculate the Relationship Between mRNA and protein Expression Ratio. *Biotechnol. Bioeng.* **96** (4): 815-819

[127] Tabor, J., Salis, H., Simpson, Z., Chevalier, A., Levskaya, A., Marcotte, E., Voigt, C. and Ellington, A. (2010) A synthetic genetic edge detection program. *Cell* **137** (7):1272-1281

[128] Tang, G., Bandwar, R. and Patel, S. (2005) Extended Upstream A-T Sequence Increase T7 Promoter Strength. *J. Biol. Chem.* **280** (49):40707-40713

[129] Monk, N. (2003) Oscillatory expression of HseI, p53, and NF-KB driven by transcriptional time delay. *Current Biology* **13** (16):1409-1413

[130] Mckane, A., Nagy, J., Newman, T. and Stefanini, M. (2007) Amplified Biochemical Oscillations in Cellular Systems. *2007* **128** (1):165-191

[131] Goodwin, B. (1965) Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation* **3** :425-437

[132] Blattner, F., Plunkett, C., Block, C., Perna, N., Burland, V., Riley, M., Collado-Vides, J., Glasner, J., Rode, C., Mayhew, G., Gregor, J., Davis, N., Kirkpatrick, H., Goedem, M., Rose, D., Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277** (5331): 1453-62

[133] Zhaxybayeva, O. and Doolittle, W. (2011) Lateral gene transfer. *Current Biology* **21** (7):R242-R246

[134] Crick, F. (1970) Central Dogma of Molecular Biology. *Nature* **227** (5258):561-563

[135] Koshland, D. (2002) The Seven Pillars of Life. *Science* **295** (5563):2215-2216

[136] Rybicki, E. (1990) The classification of organisms at the edge of life, or problems with virus synstematics. *S. Aft. J. Sci.* **86** :182-186

[137] Koonin, E., Senkevich, T. and Dolja, V. (2006) The ancient Virus World and evolution of cells. *Biology Direct* **1** :29

[138] Friedman, D. (1992) Interaction between bacteriophage labmda and its *Escherichia coli* host. *Curr. Opin. Genet. Dev.* **2** (5):727-38

[139] Russel, M. (1991) Filamentous phage assembly. *Mol. Microbiol.* **5** (7):1607-1613

[140] Miller, E., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. and Rueger, W. (2003) Bacteriophage T4 Genome. *Microbiol. Mol. Biol. Rev.* **67** (1):86-156

[141] Qimron, U., Marintcheva, B., Tabor, S. and Richardson, C. (2006) Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* **103** (50):19039-19044

[142] Karu, A., Sakaki, Y., Echols, H. and Linn, S. (1975) The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of *Escherichia coli*. *J. Biol. chem.* **250** (18):7377-7387

[143] Kelly, T. and Thomas, C. (1969) An intermediate in the replication of bacteriophage T7 DNA molecules. *J. Mol. Biol.* **44** (3):459-475

[144] Aoyama, A., Hamatake, R. and Hayashi, M. (1981) Morphogenesis of phi X174: *in vitro* synthesis of infectious phage from purified viral components. *Proc. Natl. Acad. Sci. U.S.A.* **78** (12):7285-7289