UNIVERSITY OF TRENTO

International PhD Program in Biomolecular Sciences

XXVII Cycle

"The relationship between genotype and phenotype in cell-free transcription-translation reactions."

Tutor

Prof. Sheref S. Mansy

University of Trento

Ph.D. Thesis of

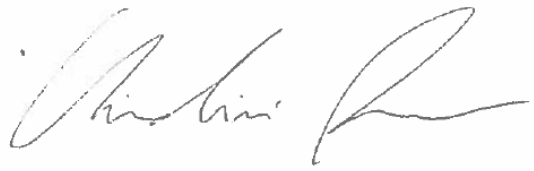
Fabio Chizzolini

University of Trento

Academic Year 2014-2015

Declaration of Authorship.

I, the undersigned Fabio Chizzolini, confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.



Abstract

Cell-free transcription and translation reactions lie at the heart of the rising field known as *in vitro* synthetic biology and their existence is fundamental for the reconstitution of artificial cells. While researchers are exploring different ways to create such reactions, the common feature that they share is the use of a template DNA to carry the information for the specific function that the reaction is required to perform. The scope of this thesis is to elucidate the relationship between the genotype and the phenotype in such reactions, investigating both transcription and translation using state of the art fluorescence spectroscopy.

Table of contents

Chapter 1: Introduction.

The concept at the heart of cell-free synthetic biology is that of the central dogma of molecular biology. One of the greatest achievements of life spanning 3.9 billion years of history is that the pathway described by the central dogma can be isolated from life itself and turned either into a technology or back into a new and different, artificial life form. While these applications are the product of recent technological advancements, the basic idea of disrupting a living cell in order to make its components interact with exogenous material dates back to the 1950s. *E. coli* cellular extracts were used to elucidate the role of mRNA in protein synthesis and, more importantly, in deciphering the triplets that compose the genetic code[1].

Cell-free systems were made not only from *E. coli* cells, but also from a wide range of different cells. Work done on cell-free systems made from rat liver investigated the relationship between metabolism and protein synthesis[2]. Cellular extracts made from two different strains of *Mycobacterium friburgensis*, either sensitive or insensitive to streptomycin, elucidated the antibiotic mode of action via the blockage of protein synthesis rather than ribonucleic acid synthesis[3]. Besides these examples, cellular extracts have contributed to the elucidation of other fundamental aspects of cellular metabolism, such as fatty acid[4] and carbohydrates[5] synthesis, DNA replication[6], nitrogen fixation[7], and many others.

Since then, many important steps have increased the activity and the stability of such systems. In its early years, *E. coli* cellular extracts were incubated directly with the desired mRNA. One of the first important steps forward for this technology was when DNA started to be used as the template for the cell-free reaction. The DNA would be transcribed into RNA which would then direct the synthesis of protein. Coupling transcription and translation in the same reaction removed the need to obtain purified mRNA, which was particularly difficult for prokaryotic mRNAs of specific genes. The efficiency of translation was also likely increased due to postulated coupling effects of transcription and translation similar to what is observed *in vivo*. As a result, full-length proteins became easier to synthesize[8]. In order to increase the efficiency of transcription and translation further, the native *E. coli* RNA polymerase was soon replaced with more processive bacteriophage RNA polymerases, such as those from T7 or SP6[9].

More recently, a deeper understanding of cellular metabolism led to the recognition of the elements that are crucial for the regeneration of molecules (mainly ATP and GTP) necessary to sustain both transcription and translation, and for the need to remove the accumulating byproducts of such reactions (mainly inorganic phosphate). Swartz[10,11] and Noireaux[12] independently showed how it is possible to use the endogenous metabolic processes of the cell to greatly increase the efficiency of *E. coli* cellular extracts by using relatively cheap molecules, such as maltose[12] or pyruvate[10,11]. For example, when maltose, a disaccharide, is added to the cell-free transcription-translation reaction, the sugar increases protein synthesis by stimulating both the recycling of inorganic phosphate and the regeneration of ATP molecules. This is possible, because enzymes that are endogenous to the cellular extract are able to hydrolyze maltose with a simultaneous addition of inorganic phosphate to form glucose-1-phosphate, which is then converted into glucose-6-phosphate, thus activating the glycolysis biochemical pathway[12]. The activation of the glycolysis pathway achieves the effect of both reducing the accumulation of inorganic phosphate and improving ATP regeneration, thereby lengthening the duration of a batch reaction up to 10 hours, while at the same time increasing the final protein yield from 0.5 mg/mL to 1.5 mg/mL[12].

In addition to the optimization of extract composition, the nature of the bacteria from which the extracts are made have been modified as well to increase transcription-translation efficiency. For example, it was shown that one of the factors that can halt protein synthesis is the depletion of four amino acids: arginine, tryptophan, cysteine and serine[13]. Their depletion is due to the presence of endogenous enzymes in the cellular extract that can degrade or modify the amino acids, such as arginine decarboxylase, tryptophanase and serine deaminase. Removing the genes coding for such enzymes by modifying the *E. coli* genome increased protein synthesis[13]. In another example, RNA stability was also increased by removing RNase E from the bacterial genome[14].

Because of the described technological advancements in efficiency of protein synthesis, cell-free transcription-translation reactions are now mature enough to show the first pharmaceutical applications. Different therapeutics, such as Tralokinumab (discovered using the ribosome display selection technique), Pegdinetanib (discovered using the mRNA display selection technique) and others, are in the preclinical or clinical stages[15]. Cell-free transcription-

translation reactions played different roles in the development and production of such therapeutics, either through *in vitro* high-throughput selection techniques followed by more conventional methods of synthesis to manufacture the selected molecule or by directly scaling up the cell-free reaction to manufacturing scales. Being able to jump from discovery to manufacturing using the same system required cell-free protein production to ramp up and reach the gram to kilogram scale. One example of the latter approach comes from Zawada et al.. The research group used an optimized extract to produce as much as 700 mg/L of a biologically active cytokine, human granulocyte-macrophage colony-stimulating factor, in up to a 100 L bioreactor. Scientists managed to do so without any appreciable loss in efficiency, thus demonstrating that cellular extracts can produce proteins of pharmaceutical grade quality at commercially relevant yields and scales[16].

Very detailed protocols for the production of *E. coli* cellular extracts are now available from different laboratories, both to make the production of cell-free transcription-translation reactions easier, and to minimize the batch-to-batch and the lab-to-lab variabilities. Surprisingly, it is still quite challenging for different laboratories to independently produce cellular extracts with the same activity, as this activity is affected by several factors, such as the efficiency of cellular lysis. One of the most popular protocols within the scientific community, published by Noireaux and colleagues, covers in detail every step of the cellular extract production and is even available in video format[17].

While *E. coli* cellular extracts are most commonly employed as batch reactions, cell-free transcription-translation reactions can also be used under different conditions. For example, it has been shown that cellular extracts can be adapted to be used in a continuous flow apparatus in which the buffer is continuously changed, thus greatly lengthening transcription-translation activity[18]. Another unusual example of the flexibility of cellular extracts that was shown to work is to spot the cellular extracts onto some simple filter paper. Upon rehydration the cellular extracts retained their full activity, raising some interesting implications for simple and cheap portable diagnostics[19].

While cellular extracts have historically been the only way to perform cell-free biology, today that is not the case anymore. In 2001, Ueda and colleagues from the University of Tokyo

illustrated a new, innovative system termed "protein synthesis using recombinant elements" (PURE) system[20]. In this new system, the transcription-translation reactions are not an endogenous activity retained after the disruption of the living cell but rather are the result of the concerted activity of a series of exogenous, purified components. Each one of the proteins required by *E. coli* for translation were over-expressed and purified with a histidine tag, then mixed together with T7 RNA polymerase, purified ribosomes, and all of the required tRNAs, amino acids, and nucleotide triphosphates. While the purification of such a relevant number of components (more than 30) makes it complex and labor expensive, the PURE system certainly provides an unprecedented level of freedom and control over the system itself. One of the advantages is that RNA and DNA molecules are more stable in the PURE system compared to cellular extracts. Both linear and circular DNA templates, for example, can be used without any significant difference in activity. In order to make it easier for other laboratories to reproduce, the Church lab has recently published an article in which genomic modifications led to the creation of seven different *E. coli* strains that comprised in total all the necessary histidine-tagged components of the PURE system[21]. While the activity was only 11% of the original PURE system, it is still an important step forward for the simplification of its constitution.

Another recent article shows how it is possible to apply the PURE system approach to different organisms. Scientists from the company New England Biolabs achieved efficient translation at high temperatures (up to 65 °C) by using purified components from the thermophilic model organism *Thermus thermophilus*[22], opening up new possibilities for engineering and testing thermostable proteins. Moreover, the availability of such reconstituted translation systems from two very distantly related organisms (the last shared common ancestor between *T. thermophilus* and *E. coli* has been estimated to be around 3.2 billion years ago) allowed the scientists to test whether or not single components from the two reconstituted systems showed functional compatibility[22]. Interestingly, there was quite a significant degree of compatibility between the protein components of the *E. coli* and *T. thermophilus* PURE systems. Finally, the ability of this new reconstituted system to sustain protein synthesis at high temperatures provided researchers with a unique biological assay to test the activity of resurrected ancient elongation factors, which were shown to be thermostable and have been predicted to come from ancient thermophilic species[23].

The PURE system has been used to express a wide range of different proteins, such as difficult to reconstitute membrane proteins[24], and even the very same aminoacyl-tRNA synthetases[25] that are used by the PURE system. Multi-subunit complexes have also been reported, such as the DNA replication system of *E. coli* which consists of 13 proteins[26]. Moreover, the total control over the reaction components makes the PURE system particularly suitable for the incorporation of non-standard amino acids, i.e. amino acids not naturally found in biology, thus conferring to the protein new chemical properties, structures and functions[27]. The ability to either increase, decrease, omit or replace one or more of the PURE system components gives to the system an unmatched flexibility compared to cellular extracts, which inevitably reflects the conditions of the living cell from which the extract has been made[28]. This flexibility is also important for selection techniques, such as mRNA display.

While having total control over the components of the PURE system has many advantages, a limited and defined number of components also carries with it some disadvantages. Two disadvantages are particularly relevant when it comes to defining and understanding the relationship that exists between genotype and phenotype. The first one is that the PURE system is lacking, because of the way this transcription-translation mix is manufactured, all the different co-factors that are required for most of the transcriptional activators of *E. coli* to work. Then there is the fact that, while with cellular extracts it is possible to choose whether to rely on the endogenous *E. coli* RNA polymerase for the transcription step or to opt for the more robust T7 RNA polymerase, with the PURE system, at present, the only available RNA polymerase for transcription is the T7 RNA polymerase. One of the main differences between the two polymerases is the mRNA synthesis rate, with T7 RNA polymerase being roughly eight times faster than the endogenous *E. coli* RNA polymerase[29]. Because of the strict coupling that exists in *E. coli* between transcription and translation (with the mRNA being translated while it is still being transcribed), it is obvious that the use of the T7 RNA polymerase alters in some way the behavior of the system, if compared to our knowledge of the genetic processes in *E. coli*. However, a recently published article showed that it is possible to use, with comparable efficiency, *E. coli* RNA polymerase holoenzyme (saturated with the sigma factor σ_{70}) with the PURE system by adding two transcription elongation factors[30]. Finally, the fact that the mRNA

produced via transcription is going to be the *only* mRNA present in the reaction is probably going to alter the way the ribosomes interact with it, compared to *in vivo*.

Different laboratories have shown that the PURE system can be used in a continuous flow apparatus that ensures a continuous supply of energy molecules to the reaction, while at the same time removing inhibitory byproducts, thus lengthening the reaction and increasing the overall protein synthesis yield[31]. Interestingly, these examples frequently exploit the use of microfluidic devices, following the same trend observed for experiments that require the encapsulation of such cell-free transcription-translation reactions into artificial vesicles[32]. Both the PURE system and *E. coli* cellular extracts have been shown to retain their activity while encapsulated in artificial vesicles[33,34].

These simple observations, the fact that cell-free systems can support transcription and translation starting from genetic material, and that they can be encapsulated into artificial vesicles while retaining their activity, spurred several laboratories in the world to try and create artificial cells, while at the same time trying understand what an artificial cell really is. At this stage, researchers are investigating, first, the molecular processes that underlie each of the properties of the living cell, while secondly, trying to reconstitute them. This process is still considering each property as an independent module to be reconstituted first *in vitro*, then inside artificial cells, while in a later phase these modules will be integrated together to increase the complexity of such artificial cells.

Several examples are now available to show the potential of this approach. The ability of natural cells to sense their environment, mainly through the recognition of different chemical stimuli, has been reconstituted *in vitro* and in artificial cells. Different articles showed how synthetic genetic circuits can detect the presence of small molecules, such as IPTG[35], tetracycline[36] and arabinose[37], both *in vitro*[35,36] and in artificial cells[37]. While these works mainly relied on transcriptional repressors to sense the small molecule of interest, other approaches are available. For example, the theophylline riboswitch has been shown to direct protein synthesis only in the presence of theophylline, working correctly both in vitro[33] and in artificial cells[33,38]. Having an artificial cell that can sense chemical stimuli from the environment is the first step for building artificial cell-cell communication, which is another

important property of the living cell. However, sensing is not enough, and in order to communicate an artificial cell would also need to be able to send a response of some sort. Interestingly, the theophylline riboswitch sensing module has been recently adapted to control the synthesis, in an artificial cell, of the pore forming protein α-hemolysin. This circuit, when theophylline is present, leads to the escape, outside of the artificial cell membrane, of the lactose analog IPTG, which finally acts as the message to communicate that the artificial cell sends to a natural cell, in response to the theophylline signal[38]. While this example clearly shows how different modules can be assembled together to create artificial cells with increasingly complex behaviors, it still relies on molecules that are quite orthogonal to the living bacterial cell, thus reducing the integration between artificial and natural cells. In an effort to overcome these limitations, researchers are now trying to hijack the natural quorum sensing system present in different bacterial species, thus leading to the creation of artificial cells that can communicate with natural cells in a chemical language that they understand.

The ability to divide itself is another important property of the living cell that several laboratories are trying to reconstitute. While no definitive system has yet been defined for artificial cells, several laboratories are interestingly exploring the use of the bacterial division pathway, mainly consisting of Fts and Min proteins[39]. FtsZ is a tubulin homolog that plays a major role in bacterial cell division, forming a constricting ring that is able to generate the force to divide the cell into two[39]. Interestingly, the expression in vesicles of a version of FtsZ modified with the addition of a C-terminal amphipathic helix (as *in vivo* FtsZ is cytosolic and interacts with the membrane through the other proteins), showed the formation of multiple Z rings in vesicles[40]. Moreover, the brighter Z rings also produced visible constrictions on the membrane of the vesicles, thus suggesting that FtsZ was generating a constricting force, unfortunately not strong enough to elicit cell division[40]. The role of the Min proteins, on the other hand, is to direct the location of cell division. They do so by creating concentration gradients within the cell, with lowest point of Min proteins concentration being in the middle of the cell, where the machinery for cell division will assemble[39]. Recent work showed how it is possible to reconstitute the correct Min protein gradient inside a cell-shaped compartment, and that this gradient is able to accumulate FtsZ in the middle of the compartment[41]. Still, the combination of the two systems within an artificial cell has not been reported yet, so the

problem of reconstituting cell division using the same protein machinery as the living cell is open and unresolved yet. Proteins that interact with the cellular membrane are also fundamental for another important feature of the living cell: the high degree of intracellular spatial order. Macromolecular complexes occupy a specific place within the cell, and this precise organization would be impossible without the presence of the cytoskeleton. Among the cytoskeletal proteins that have been recently described in *E. coli* we can find FtsZ, as well as MreB and MreC. MreB, the bacterial homologue to the eukaryotic actin, when expressed using encapsulated transcription-translation reactions, have been shown to form filamentous structures located near the inner membrane of artificial cells[42]. Filamentous structures near the membrane also play an important role in cell division. Therefore, elucidating the way to reconstruct the cytoskeleton will also be of primary importance in achieving artificial cell division.

The ability of the living cell to replicate its genome is another fundamental property that scientists are trying to reconstitute in artificial cells. While replicating DNA *in vitro* in the laboratory has become routine since the introduction of the PCR and affordable thermocyclers, *in vitro* isothermal DNA replication is still challenging. One promising DNA replication system is that of bacteriophage Φ29 which requires only four proteins to replicate DNA[43]. Moreover, Φ29 DNA polymerase shows a high processivity, allowing for reported replication of up to 70 kb[44] from a single binding event, with high fidelity, and with the possibility of obtaining microgram amounts of DNA starting from nanograms in a single-step isothermal reaction[45]. However, these experiments were performed by purifying and then using the four proteins required for DNA amplification, and they have yet to be tested inside an artificial cell, so some key steps are still missing before this system can be used. Nomura and colleagues followed a completely different approach when they showed that the replication process could be achieved by expressing the minimal set of genes required for DNA replication by *E. coli* using PURE system reactions[26]. However, the procedure that they had to use was quite convoluted and it involved incubating PURE system reactions at low temperatures, such as 27 °C, and expressing the different functional parts of the replication system (such as DNA polymerase III holoenzyme and the RNA priming) in different reactions, grouping together the products after

protein synthesis occurred[26]. All these factors could make it more challenging for the process to be reconstituted inside an artificial cell.

All these efforts are slowly changing our perception of life and of technology, blurring the line between the two. These studies not only have a direct impact on the development of artificial cells, but also an indirect and broader impact on several different fields of research. The quest for the artificial cell is deepening, for example, our understanding of what life is, from both a scientific and philosophical point of view. Moreover, the basic molecular understanding of the fundamental processes of the living cell is crucial for the definition of the "minimal" set of elements required to define a living cell.

In this thesis, we focus on something much simpler, yet of considerable importance. In order for an artificial cell to resemble a natural, living cell, the artificial cell needs to possess functional genomic material. It is very important, therefore, to know exactly what this piece of DNA should be like, and this has been the main area of interest for this thesis. The focus is primarily on the PURE system as the cell-free transcription-translation reaction system of choice. While there are several laboratories working on the same problem, this work is different and novel in some key aspects. First, all the work was done by having the creation of an artificial cell as the final goal. While other laboratories typically regard cell-free transcription-translation reactions as a quick benchmark to fine-tune genetic circuits to be employed *in vivo*, our aim is to define the rules governing gene expression in cell-free transcription-translation reactions as the first step in the development of a genome for an artificial cell. Second, when performing our experiments we tried to make good use of the controlled reaction environment that is the PURE system. For every PURE system reaction, we recorded not only the translational activity, but also the transcriptional activity, in an effort to gain a deeper understanding of the dynamics involved in cell-free gene expression. Moreover, we quantified, in terms of molar concentration, both the RNA and the protein produced in the experiments that will be presented in this thesis.

List of main goals addressed in this thesis:

(i) Investigate the relationship between the template DNA design and both transcription and translation in PURE system reactions, in an effort to gain a more precise control over cell-free gene expression with respect to the currently available methodologies;

(ii) Establish fluorescence spectroscopy methodologies to precisely quantify both RNA and protein production in real-time within PURE system reactions, in order to have a deeper understanding of the relationship that connects the template DNA, the transcribed RNA and the translated protein;

(iii) Investigate the use of a series of different T7 transcriptional promoters, both in the context of single-gene expression and in the context of genetic operons composed of either two or three genes, to control both transcription and translation in PURE system reactions;

(iv) Explore the use of a series of *E. coli* ribosome binding sites to control single-gene expression in PURE system reactions;

(v) Examine the role of the 5'-UTR and 3'-UTR, excluding the ribosome binding site, on both transcription and translation in PURE system reactions;

(vi) Combine the characterized T7 transcriptional promoters and *E. coli* ribosome binding sites in an effort to increase our ability to control single-gene expression in PURE system reactions;

(vii) Use the acquired data to train a computational model in order to predict gene expression in PURE system reactions when employing different genetic parts;

(viii) Apply our set of characterized genetic parts on two simple genetic circuits: a cascade circuit composed by another viral RNA polymerase, T3 RNA polymerase, and a repressor circuit, composed by the EsaR repressor;

(ix) Investigate the inherent variability of transcription and translation in PURE system reactions both in the context of single-gene expression and in the context of simple genetic circuits;

(x) Generate a series of genetically encoded fluorescent pictures with the scope of visually representing both our ability to control gene expression in PURE system reaction and the associated variability of gene expression.

Chapter 2: Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength.

2.1. Introduction to the article.

In the article that I am presenting in this chapter, we investigated the influence of T7 transcriptional promoter strength on both mRNA and protein production in PURE system reactions using different genetic architectures, such as one-gene, two-gene and three-gene operons. It was important to do so for different reasons. First, most of the papers showing expression of multi-subunit protein complexes with the PURE system relied only on changing the concentration of the different template DNAs, one template for each expressed protein. While such an approach showed promising results, it was not the best approach when the design of an artificial cell is the ultimate goal. For example, the encapsulation of multiple pieces of DNA inside of a single vesicle is difficult. The main issue is that the encapsulation of DNA template is inefficient and variable. The efficiency of encapsulation is influenced by the technique employed to generate the liposomes, as well as by the chemical composition of the liposomes and of the solution to be encapsulated. The efficiency can be as low as <1% and generally not more than 33% with traditional techniques[46]. Therefore, it is important to have all of the needed genetic elements on one piece of DNA so that the likelihood of having the whole "genome" inside of a vesicle compartment is increased. We decided to use genetic operons for two reasons. First, operons decrease the number of genetic parts required to express more than one gene, making the construction of genetic circuits and later of artificial cells easier. Second, we wanted to gain more insight into the dynamics of the transcription and translation reactions, especially in terms of what limits the amount of protein produced. While it is known that metabolic load plays an important role in protein synthesis, we wanted to understand if it was possible to increase protein production by reducing (using our T7 transcriptional promoters) transcription. On the other hand, prior work showed that increasing template concentration for monocistronic genetic constructs increases mRNA and protein production, therefore indicating that there were other effects, such as ribosome inactivation, that could cause protein synthesis to stop. We thought that by using two-gene and three-gene operons we could detect more easily signs of metabolic load on our system. Finally, reports indicate how, in cellular extract

reactions, the relationship between DNA template concentration and protein expression is directly linear only below a certain saturating concentration of the DNA template[35,47]. Above the saturating concentration, increasing the DNA template does not lead to any increase in protein expression. Therefore, in the article we will briefly investigate also the relationship between the DNA template concentration, and both transcription and protein expression in PURE system reactions.

Another interesting feature of the article is that our report was the first research article to show how the recently described Spinach aptamer can be used to measure transcription *in vitro*. The Spinach aptamer is an RNA aptamer that is able to bind specifically (K_d = 537 nM) the small molecule 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), which is structurally similar to the fluorophore that confers fluorescence to GFP. DFHBI alone has, similarly to the GFP fluorophore when the protein is denaturated, a low quantum yield (0.0007) and therefore gives a negligible fluorescence signal (0.04% of wild-type *Aequorea* GFP brightness). However, when the Spinach aptamer binds to DFHBI, the specific contacts between the aptamer and the fluorophore prevent intramolecular motions of DFHBI, thereby making fluorescence the major pathway available to dissipate the energy of the excited state fluorophore[48]. Therefore, the quantum yield of the fluorophore increases to 0.72 and the brightness goes up to 80% of wild-type *Aequorea* GFP. By encoding the Spinach aptamer in a transcribable unit of DNA it then becomes possible to monitor the synthesis of RNA in real-time by following the increase of fluorescence in the presence of DFHBI. With such Spinach encoding constructs we characterized a series of T7 transcriptional promoters not only in terms of protein production, but also by directly measuring RNA production with the combined use of Spinach and a fluorescent protein constructs.

First, we investigated whether Spinach could really be used to measure RNA production using transcription-only reactions with purified T7 RNA polymerase. Then we characterized what is the minimal number of bases that are required before the promoter for efficient transcription. After that, we modified the sequence composition of the bases before the promoter to check for any impact on the transcription rate. We also investigated the role of the six bases spanning from the +1 to the +6 of the nascent RNA transcript. We then characterized a library of 21 T7 transcriptional promoters, taking some of the candidate promoters directly from the genome of

the T7 phage, while we randomly created the others. The resulting promoters gave a good distribution of intensities, i.e. activities. Finally, we confirmed the results by both testing some of the promoters using a RT-qPCR assay and with ethidium bromide stained agarose gel electrophoresis.

After this part, we moved onto testing our promoters in PURE system reactions. First, we tested a few configurations of the aptamer to understand what would be the best for both the protein and the RNA fluorescence signal. We also compared the use of linear and circular template DNA. After that, we performed the experiment with the PURE system reactions, and we discovered that while RNA production was consistent with the previous transcription-only experiments, protein production seemed to cluster to "high" and "low" values, with very few values in-between. This is an effect not yet fully understood. One possibility is that the RNA that is transcribed is the only mRNA in the reaction. Thus even a very small concentration of RNA will be translated efficiently by the ribosomes.

Then we started to use the different T7 promoters with two-gene and three-gene operons. To our surprise, we discovered that the position of a gene within the operon had a bigger influence on expression compared to the transcriptional promoter. Specifically, the further the gene was from the first position, the lower the expression became, with the biggest gap being from the first to the second position. We hypothesize that the drop in expression could due to the folding of the mRNA which could mask or block the ribosome binding sites of the second and the third genes. It will be important to determine in the future if the same effect is observed with *E. coli* RNA polymerase. We also changed the order of the genes in the operons, but the result remained the same. Interestingly, at least on one occasion we observed that changing the order of the genes had an impact also on the transcription of the operon. This work elucidated some of the peculiarities of PURE system reactions and established the main assays and techniques that will then be used in the next chapter of this thesis.

Chapter 3: Combinations of different T7 transcriptional promoters and ribosome binding sites control gene expression in cell-free transcription-translation reactions.

3.1. Introduction.

After investigating the use of different T7 transcriptional promoters to modulate gene expression in PURE system reactions, we next turned to the characterization of the ribosome binding site. The ribosome binding site (or Shine-Dalgarno sequence) is located in the 5'-UTR (untranslated region) of the mRNA, usually 6-7 nucleotides upstream of the start codon. The ribosome binding site sequence is complementary to the 3' end of the rRNA and can therefore physically interact with the ribosome through base-pairing interactions thereby recruiting the ribosome to the mRNA. Therefore, it should be possible to influence gene expression by using different ribosome binding sites[49]. Moreover, recent analyses done *in vivo* have highlighted that, when mixing together different promoters, 5'-UTRs and coding sequences in a combinatorial manner, the 5'-UTR of the mRNA is responsible for 46% of the overall variability in gene expression, ranking as the most important element for modulating gene expression[50].

The role of the 5'-UTR does not solely arise from the ribosome binding site sequence but also the structure formed by the 5'-UTR. The folding of the 5'-UTR can influence the availability of the ribosome binding site to the ribosomes[51]. Similarly, the sequence of the 5'-UTR can interact with the coding sequence of the downstream gene[50,52]. Building evidence is linking all these effects to the secondary and tertiary structures in which the mRNA folds[50,52]. While the connection between gene expression and mRNA structure is currently under investigation by several research groups, we are still far from a complete understanding of how gene expression and mRNA structure interact with each other. Even though many algorithms to predict RNA structure exists, the correlation between mRNA structure and gene expression has not been perfectly elucidated yet. The biggest experimental challenge is the total conformational space of the mRNA, which is enormous if we consider the different 5'-UTRs, ribosome binding sites, coding sequences, and their possible interactions.

While different groups have proposed different strategies to control the influence of the mRNA structure on protein expression, the two main approaches that arose are to either remove or predict the influence of RNA structure. The first one exploits the fact that while a ribosome is sliding along a piece of RNA translating it, the ribosome is also unwinding the secondary structures of that RNA[53]. Therefore, an additional ribosome binding site upstream of the target ribosome binding site has been reported to diminish the structures formed by the mRNA that can influence the expression from the downstream ribosome binding site[53]. The second approach is completely different, and relies on computational models to predict gene expression from a ribosome binding site, the most famous being the Ribosome binding sites Calculator[54]. Algorithms that can predict the structure of the mRNA structure are used for such computational models. These algorithms can be used either to estimate the gene expression arising from a given ribosome binding site and 5'-UTR sequence, or to generate novel ribosome binding sites and 5'-UTR sequences designed to achieve a desired level of gene expression[54].

While many articles investigated the role of the ribosome binding site and of the mRNA secondary structure *in vivo* using *E. coli*, not much has been done in cell-free systems. Only a few papers report the use of different ribosome binding sites to modulate gene expression in *in vitro* systems[36,55], but still no studies were specifically designed to explore the use of different ribosome binding sites to control protein expression in cell-free transcription-translation reactions. Therefore, we sought to investigate the influence of different ribosome binding sites upon protein expression in PURE system reactions.

3.2. Results and discussion.

3.2.1. The expression of the second gene influences the expression of the first gene in synthetic two-genes operons when using PURE system reactions.

First, in order to generate the different ribosome binding sites we applied a "Design of Experiments" approach. We did so because we wanted to explore as much as possible of the available sequence space, without escalating the number of experiments. Using this approach, we divided the sequence composition into three different variables: the number of base pairing interactions between the ribosome and the ribosome binding site, the position of such paired bases within the ribosome binding site, and the nucleotide composition of the non-pairing bases

within the ribosome binding site. Using a D-criterion optimal statistical design of experiments[56], Michele Forlin designed 16 different ribosome binding sites to possibly describe the expression efficiency with respect to these three variables (Table 1).

Table 1. List of designed ribosome binding sites.

name	sequence
RL055A	TAAGGAGAA taatct ATG
CD_104	**CGGAA**AG**GT** taatct ATG
CD_105	**CGGA**GAG**GT** taatct ATG
CD_106	**CGG**GGAG**GT** taatct ATG
CD_107	TA**GAAGA**A**C** taatct ATG
CD_108	TAAGGA**A**A**C** taatct ATG
CD_109	**CG**AGGA**A**A**C** taatct ATG
CD_110	**C**AAGGAGAC taatct ATG
CD_111	**GCCTTCTTT** taatct ATG
CD_112	**GCCTTC**G**GT** taatct ATG
CD_113	**GCCTT**AG**GT** taatct ATG
CD_114	TAA**TTCTTG** taatct ATG
CD_115	TAAG**TCTTG** taatct ATG
CD_116	TAAGGA**TTG** taatct ATG
CD_117	**GCCT**G**CTTG** taatct ATG
CD_118	TAAGGAG**GG** taatct ATG
CD_119	TAAGGAG**GT** taatct ATG
CD_120	**CGGA**GAG**GC** taatct ATG
CD_121	TAAGGAGAA taatct ATG
CD_122	TAAGGAGAA taatct **TAA**
CD_123	TAAGGAGAA taatct **TTG**
CD_124	TAAGGAGAA taatct **GTG**
CD_125	TAAGGAG**G**A taatct ATG
CD_126	TAAGGAGA**T** taatct ATG
CD_127	**G**AAGGAGA**T** **at**a**ca**t ATG

The designed ribosome binding sites to be tested. RL055A is the reference sequence previously used in our group to characterized different fluorescent proteins[57]. CD_121 has, instead of the start codon of the first gene, the stop codon "TAA". CD_122 has, instead of the start codon of the second gene, the stop codon "TAA". CD_123 and CD_124 have, instead of the more common start codon "ATG", the rarer start codons "TTG" and "CTG" respectively. CD_127 is the standard pET21b ribosome binding site.

We decided to include those sequences in a two-gene operon, consisting of two fluorescent proteins, mRFP1 followed by GFPmut3b (Figure 1a). This operon was already used with success in our laboratory to characterize the expression of different fluorescent proteins in

PURE system reactions[57]. In our experimental set up, the first protein, mRFP1, would always be under the control of the same, strong, ribosome binding site, therefore acting as an internal control to normalize the data. The second gene, GFPmut3b, would be the one under the expression of the variable ribosome binding site. In this way, the ratio between the two proteins would be a precise indication of the influence of the different ribosome binding sites on protein expression in PURE system reactions.

However, when we performed the experiment the result was rather surprising. Changing the expression of the second gene (because of the different ribosome binding sites) also changed the expression of the first gene (Figure 1b). There seemed to be a recognizable trend, in which the level of expression of the first gene directly correlates with the expression of the second (Figure 1b). This challenges the intuitive view that the expression of genes from an operon depends only on the available resources in the PURE system reaction, while at the same time highlighting the centrality of the mRNA structure for gene expression. Moreover, this is an indication that the translational process dynamically modifies the structure of the mRNA. In this regard, it is easier to understand the role that the expression of the second gene exerts on the expression of the first gene. Probably the rate of translation of the second gene produces a big effect on the overall structure of the mRNA, which in turn affects the expression of the first gene.

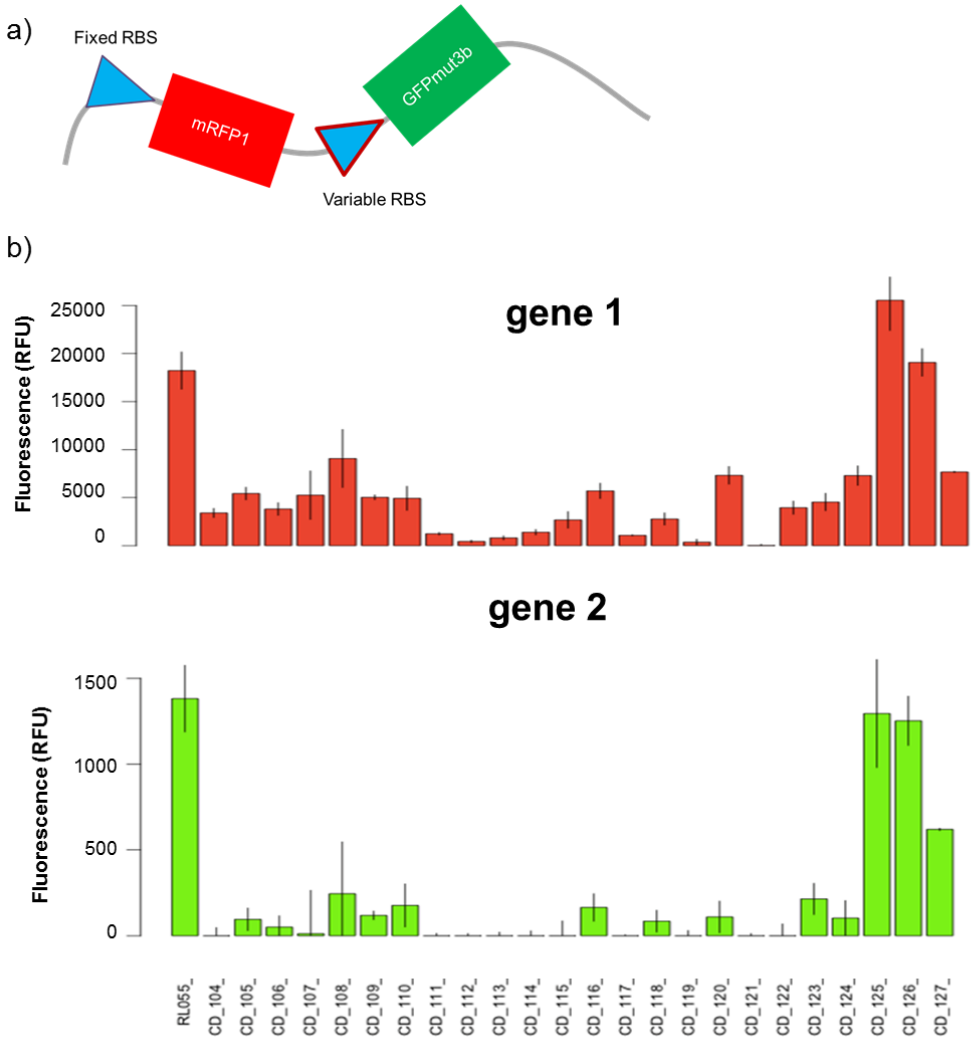


Figure 1. Ribosome binding site characterization. (a) The two-gene operon employed to characterize the ribosome binding sites. A fixed ribosome binding sites controls the expression of the first gene, the fluorescent protein mRFP1. The second gene, the fluorescent protein GFPmut3b, is under the control of a variable ribosome binding site. (b) Expression of the first gene is not constant, but is highly influenced by the expression of the second gene.

In order to explore more in detail this effect, we focused particularly on two different samples. We compared our "positive control" operon (RL055A), which expresses both genes strongly, and the "negative control" operon (CD_122), in which a stop codon (TAA) substitutes the start codon (ATG) of the second gene. Comparing these two systems is thus the best strategy to investigate the role of the expression of the second gene on the expression of the first gene in a two-gene operon. First, we wanted to understand if this effect could be observed regardless of the nature of the template DNA employed. To do so, we compared the expression of our two

operons in the form of either plasmid DNA or PCR generated DNA. To our surprise, we observed that the feedback effect of the expression of the second gene on the expression of the first was present only when using plasmid DNA and not DNA generated with PCR (Figure 2). To understand if this was due to the supercoiling of the plasmid DNA, we digested the plasmid using a restriction enzyme with a single restriction site in the backbone of the plasmid to linearize the template. Then, we compared expression from this template with the one arising from the exact same template, only generated by PCR, but having exactly the same length and base composition. Again, we could detect the feedback effect only in the template created using plasmid DNA, and not the DNA generated from PCR (Figure 2). Next, we tried to recover plasmid DNA from an *E. coli* strain that lacks the genes required for DNA methylation, as DNA methylation is the main difference between plasmid DNA and DNA generated from PCR. Interestingly, this approach partially removed the feedback effect, reducing it (Figure 2). Finally, we tried to incubate the plasmid DNA with two different RNases, a mixture of RNase A/T1, or RNase H. After the incubation, the plasmid DNA would be purified again, and used as the template for a PURE system reaction. This treatment led to a complete disappearance of the feedback effect (Figure 2). These results are quite puzzling and are still under investigation.

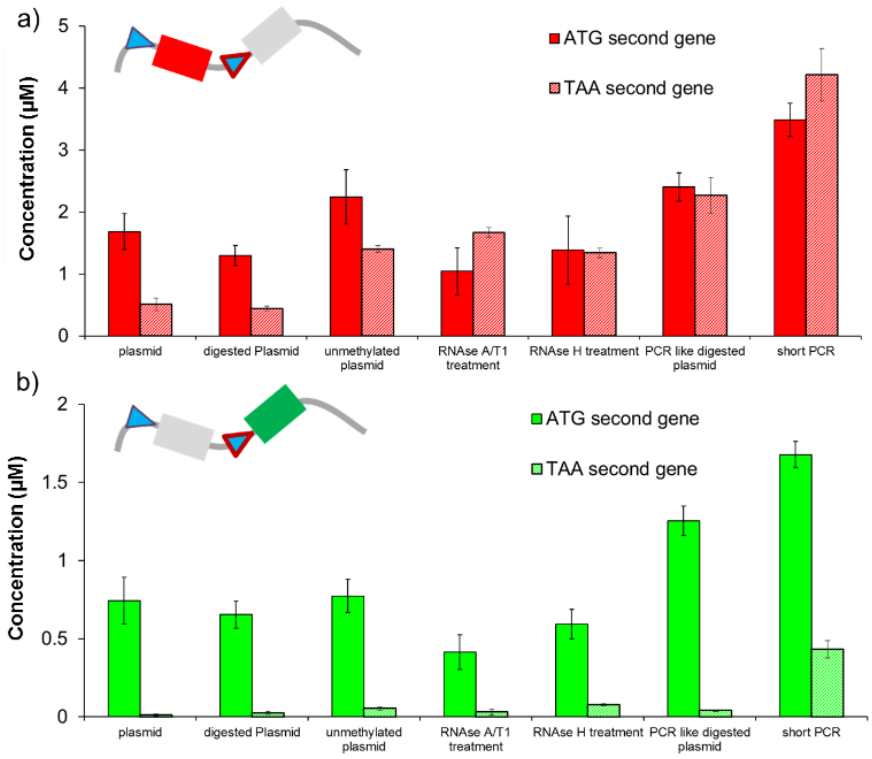


Figure 2. Different templates were employed to test the influence on the expression of the first gene by the expression of the second gene in a two-genes operon. Two constructs were employed, the standard RL055A and the negative control CD_122, in which the start codon of the second gene is replaced by a stop codon. (a) The expression of the first gene is more influenced by the expression of the second gene when using a plasmid template. Removing methylation from the plasmid partly removes the influence of the expression of the second gene. An RNase treatment completely removes the influence of the expression of the second gene. Linear templates assembled via PCR do not show any kind of relationship between the expression of the two genes. (b) The expression of the second gene is abolished when changing the start codon of the gene for a stop codon in all the templates except for a PCR product that only includes the insert of the construct.

Because in some of the tested samples we could detect the production of the second gene even if the start codon was replaced by a stop codon, we tried to investigate what was the role of the expression of the first gene in this observed behavior. We compared the expression of GFPmut3b, so the second gene of the operon, from two different templates, the one harboring the full two-gene operon (with the stop codon instead of the start codon of the second gene) and one in which only the second gene was present as a single-gene template with a stop codon instead of the start codon. To our surprise, the expression of GFPmut3b increased when the first gene was not present and only the second gene was (Figure 3). This effect is probably similar to what we observed in the previous chapter, in which genes that are further away from the 5' end of the transcribed mRNA show a reduced expression. Interestingly, it shows that even if the

start codon is not present the ribosomes are still able in some way to produce the fluorescent protein. Moreover, the first start codon available that is in frame would result in the loss of the first 77 aminoacids of the fluorescent protein, which is 33% of the total length of the protein, thus it also seems unlikely that this is what is happening.

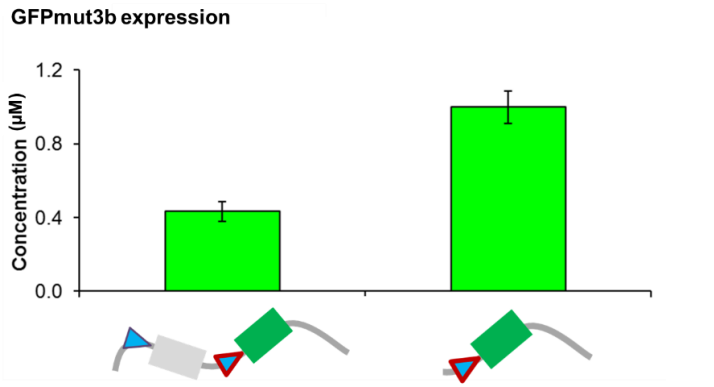


Figure 3. Gene expression arising genes in which the start codon have been replaced by a stop codon. The first sample is the second gene in a two-genes operon, while the second sample is a single gene. Comparing these two samples we can conclude that the observed expression is not due to the expression of the first gene in the two-genes operon, but can be observed also independently.

3.2.2. The "Bicistronic Design" does not work in PURE system reactions.

In order to understand if we could overcome the issue of the RNA structure, especially when using genetic operons in PURE system reactions, we tested the "Bicistronic Design" (BCD), recently described by Endy and colleagues[53]. The idea is to exploit the fact that while the ribosomes slide on the mRNA during translation, the ribosomes disrupt the mRNA structures. Therefore, in order to insulate the accessibility of a given ribosome binding site from the perturbations of the mRNA structure, it will simply be sufficient to place the ribosome binding site downstream of another ribosome binding site. This upstream ribosome biding site will drive the expression of a small peptide, and the desired ribosome binding site will be placed within the coding sequence of the small peptide. The continuous flow of ribosomes will keep the second ribosome binding site available regardless of the mRNA structure. Even though the published work was done *in vivo* using *E. coli*[53], we wanted to investigate if this approach could be applied also *in vitro*.

In order to properly test if this approach could remove the influence of mRNA secondary structures from gene expression in our system we designed both a positive and a negative control. The positive control would be a single gene construct using our standard promoter, leader sequence and ribosome binding site, while the negative control would be the same construct but modified in such a way that the beginning of the coding sequence would, when transcribed, create a hairpin with the ribosome binding site, thus masking it from the ribosomes. We created two different versions of both the positive and the negative controls: one following our usual genetic design, and one identical to the published one (Table 2). Correctly, for both the designs the hairpin from the negative control would completely abolish gene expression. However, when the upstream ribosome binding sites were inserted, only a modest recovery in gene expression was achieved (Figure 4a). We then decided to perform the same experiment, only this time using our two-gene operon. Again, we generated a negative control by modifying the coding sequence of the second gene to form a hairpin with the ribosome binding site when transcribed. We employed the very same designs reported in Table 5, applying them to the second gene of the operon: GFPmut3b. We then observed a similar drop in expression of the second gene (the one with the hairpin), even though it was not completely abolished, as it was in the single gene experiment, but a minimal expression was retained. Upon insertion of the second ribosome binding site upstream of the one blocked by the hairpin we could only detect a very modest recovery in the expression of the second gene (Figure 4b). Because of the poor results, we then abandoned the use of this approach.

Table 2. Tested Bicistronic designs.

Sample name	Sequence
MCD (our design)	CCGGTTAATACGACTCACTATAGGGAGATCTTAATCATGCTAAGGAGGTTTTCTAATGGACTCCCCC GATAAAAGTAAAGTGATTAACAGC – GFPmut3b CDS
HAIRPIN (our design)	CCGGTTAATACGACTCACTATAGGGAGATCTTAATCATGCTAAGGAGGTTTTCTAATGACCTCCTTA GATAAAAGTAAAGTGATTAACAGC – GFPmut3b CDS
BCD (our design)	CCGGTTAATACGACTCACTATAGGGAGAGGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAA GCAATTTTCGTACTGAAACATCTTAATCATGCTAAGGAGGTTTTCTAATGACCTCCTTAGATAAAAG TAAAGTGATTAACAGC- GFPmut3b CDS
MCD (published design)	GCGGATCCGAATTCAATTAGTTTGAACTTATAAGGAGAATAATTAATGCGC – GFPmut3b CDS
HAIRPIN (published design)	GCGGATCCGAATTCAATTAGTTTGAACTTATAAGGAGAATAATTAATGTTCTCCTTACGC – GFPmut3b CDS
BCD (published design)	GAAGGTCGTCACTCCACCGGTGCTTAATAAGGAGGGAATTCAAATGGTTTCAACTTATAAGGAGAAT AATTAATGTTCTCCTTACGC – GFPmut3b CDS

Modifications of the leader sequence and of the beginning of the fluorescent protein GFPmut3b coding sequence. Marked in red we can see the standard ribosome binding sites, while marked in green we have start codons. With red letters we can see the hairpin

introduced to base pair and therefore block the ribosome binding site. Underlined in black we have the coding sequence of the additional peptide required for the Bicistronic design. Marked in cyan we have the stop codon of the additional peptide required by the Bicistronic design.

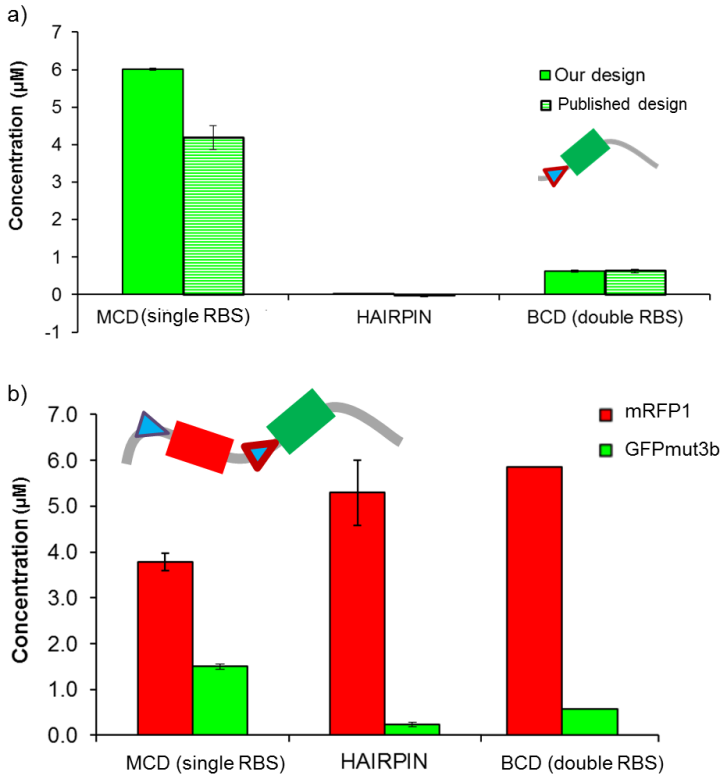


Figure 4. Bicistronic design (BCD) experiment in PURE system reactions. (a) A sequence placed so to form a hairpin with the ribosome binding site in the mRNA is able to shut down expression from a single gene construct. Inserting a second ribosome binding site controlling the expression of a coding sequence going through the hairpin does not relieve the hairpin blockage of translation. (b) The same can be observed when the template is a two-gene operon. Both the hairpin, and subsequently the additional ribosome binding site were applied to the second gene of the operon, the fluorescent protein GFPmut3b.

3.2.3. *The 3'-UTR is required for optimal expression in PURE system reactions, while the 5'-UTR can influence gene expression regardless of the ribosome binding site.*

We then sought to continue the project by characterizing different ribosome binding sites using single gene constructs, but before doing so we briefly explored the role of the 5'-UTR (without modifying the ribosome binding site) and of the 3'-UTR when using linear DNA templates in PURE system reactions. First, we tested linear DNA templates with different 3'-UTRs. The templates would end either directly with the stop codon of the coding sequence (TAA), or with parts of increasing length or complexity, going from just 56 bases of low folding stability, to the Spinach aptamer followed by T-ϕ terminator (Table 3, Figure 5a). Interestingly, this experiment

clearly showed that a linear DNA template ending with the stop codon of the coding sequence is extremely inefficient in driving gene expression in PURE system reactions. Moreover, there seem to be a mixed dependency from both the length and the folding stability of the included DNA section after the stop codon, with the best result obtained by including the T-ϕ terminator (Figure 5a). Finally, we confirmed that this effect is present in two-gene operons, but only affects the second gene of the operon. It seems like the presence of the coding sequence of the second gene stabilizes the expression of the first gene, while for the correct expression of the second the presence of the T-ϕ terminator is required (Figure 5b).

Table 3. 3'-end of the templates employed for the experiment on the role of the 3'-UTR.

Sample name	Ending of the template
TAA	TAA
TAA+56bp	TAATCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAA
TAA+Spinach	TAAGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGGACGCAACTGAATGAAATGGTGA AGGACGGGTCCAGGTGTGGCTGCTTCGGCAGTGCAGCTTGTTGAGTAGAGTGTGAGCTC CGTAACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA
TAA+91bp	TAACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGA AGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAAC
TAA+longTerm	TAACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGA AGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTA AACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGG A
TAA+Sp+Term	TAAGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGGACGCAACTGAATGAAATGGTGA AGGACGGGTCCAGGTGTGGCTGCTTCGGCAGTGCAGCTTGTTGAGTAGAGTGTGAGCTC CGTAACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCATAG CATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTCGAGCACCACCAC CACCACCACTGAGATCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACC GCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTT GCTGAAAGGAGGAACT
TAA+Term	TAACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGA AGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTA AACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACT

These are all the 3'-UTRs that were tested in cell-free transcription-translation reactions. From the negative control (template ending right at the TAA of the coding sequence), we increase in length and complexity of the 3'-UTR.

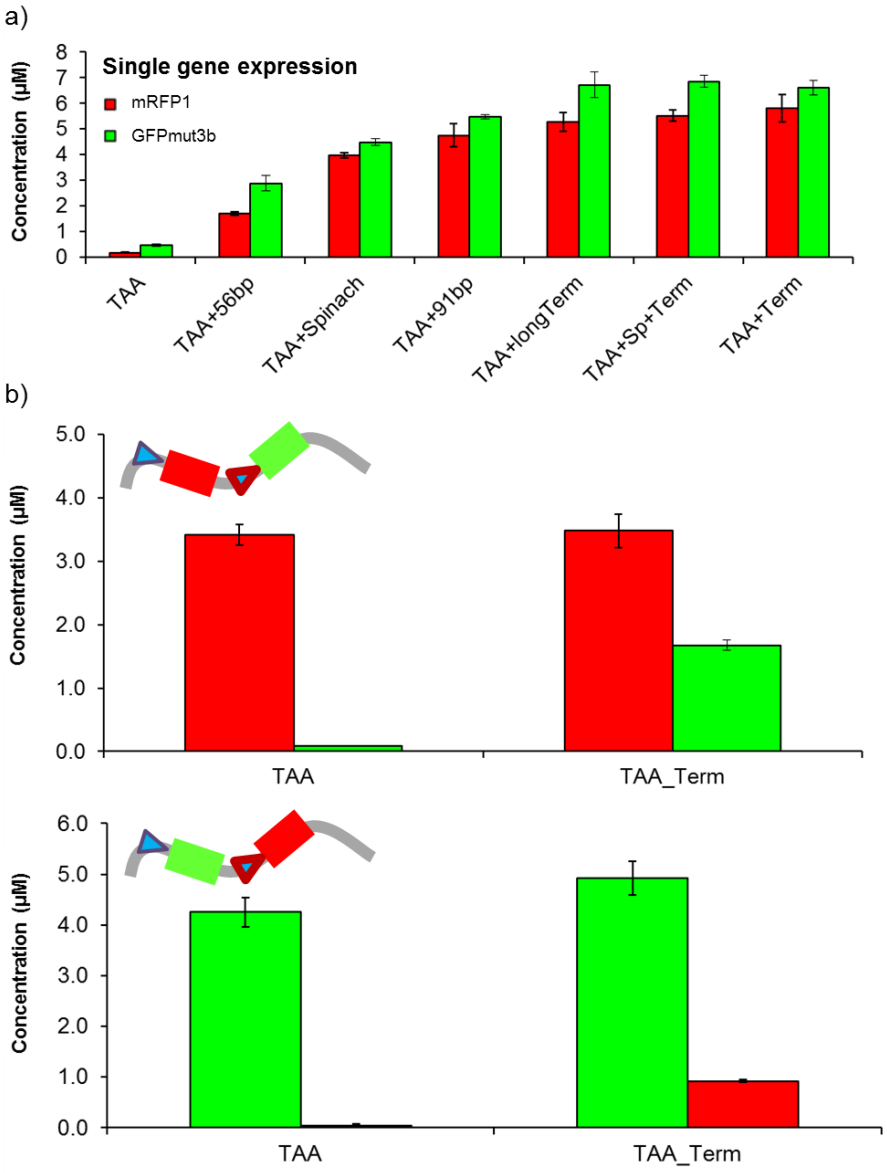


Figure 5. Role of the 3'-UTR of the mRNA in PURE system reactions. (a) Role of the 3'-UTR when using single-gene templates encoding either the fluorescent protein mRFP1 or GFPmut3b. When the mRNA ends right at the stop codon gene expression is severely impaired. An additional 56 unstructured bases can partially rescue expression. However, at least 91 unstructured bases, or structured ones such as a transcriptional terminator, or the Spinach aptamer are able to fully recover gene expression. (b) The second gene in a two-genes operon can act similarly to a 3'-UTR in regard to the expression of the first gene. The expression of the second gene still requires the presence of a structured region such as transcriptional terminator.

We then analyzed the role of the 5'-UTR (excluded the ribosome binding site) on single-gene expression in PURE system reactions. We tested two random 5'-UTRs without any particularly strong secondary structure (Table 4). We applied those to several different templates, some of which also included testing the role of the 3'-UTR as we have done previously, while at the

same time changing the 5'-UTR (Figure 6a). Interestingly, for the fluorescent protein mRFP1, using a different leader sequence would result in a different gene expression (Figure 6a). We then used the Spinach aptamer to investigate if this difference was due to a difference in transcription or translation, and we found out that it is only a translational difference (Figure 6b).

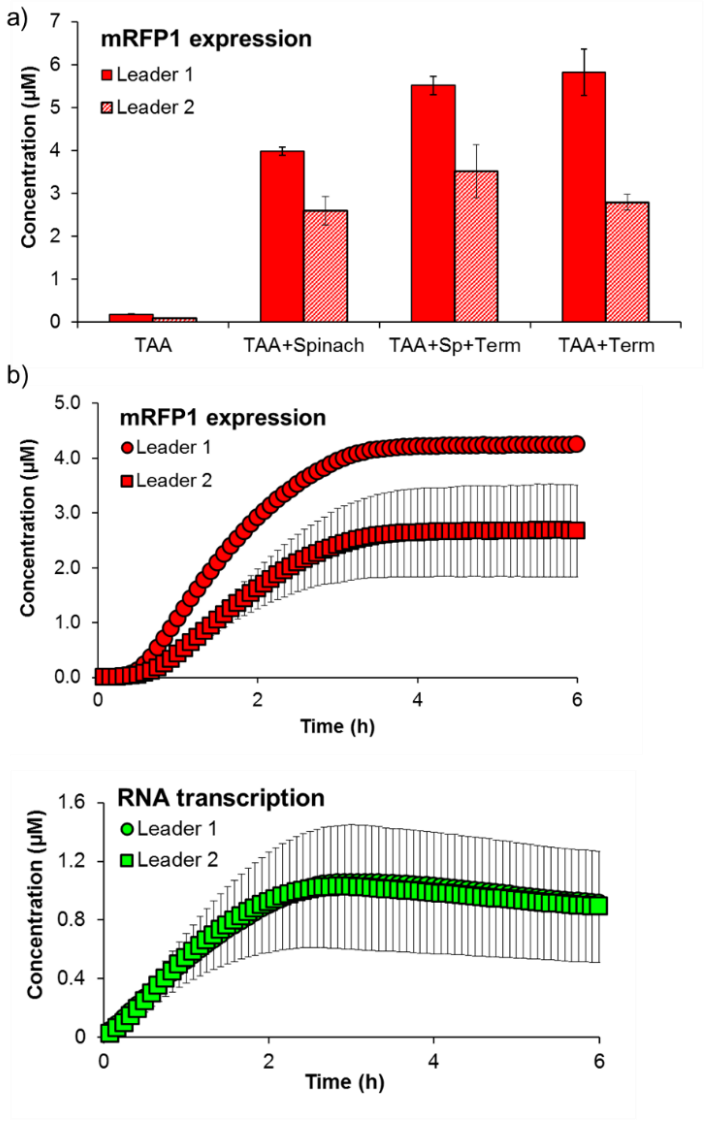


Figure 6. Expression of the fluorescent protein mRFP1 can be modulated by a leader sequence without changing the ribosome binding site in PURE system reactions. (a) Single gene expression is modulated by the leader sequence regardless of the 3'-UTR. (b) The leader sequence modulates gene expression by influencing the translation of the gene. Transcription is not influenced by the different leader sequence, indicating a possible modulation of the interaction between the mRNA and the ribosomes.

Finally, we generated three more leader sequences (Table 4), in order to find some that would not influence gene expression, to be used for experiments in which the expression of more than one gene at the same time is required. Two of the new leader sequences were randomly designed, again with no particularly stable secondary structure and by keeping fixed the ribosome binding site, while the third was generated using the RBS calculator, with a different ribosome binding site as well (Table 4). Interestingly, the two randomly designed sequences showed optimal gene expression, while the sequence generated with the RBS calculator had the weakest gene expression (Figure 7a). This time, using Spinach, we also observed a difference in transcription while using different leader sequences, and interestingly the sequence generated with the RBS calculator had the better transcription rate (Figure 7b).

Table 4. Leader sequences tested with cell-free transcription-translation reactions.

Leader sequence code	Sequence
Leader 1	TTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAG GAGATATACATATG
Leader 2	GCGGATCCGAATTCAATTAGTTTGAACTTATAAGGAGAATAATCTATG
Leader 3.1	ATAATCATATTAGAATGCTTTAAGAAGGAGATATACATATG
Leader 3.2	TCTAAGTTTTTCCACTTGGTTTAAGAAGGAGATATACATATG
Leader 3.RBS calc	GGTATAAAAAGCAAATACTAGGGGGGTAGAGAATG

Different leader sequences were tested for their influence on gene expression. Highlighted in red the ribosome binding site. In green the start codon of the red fluorescent protein mRFP1.

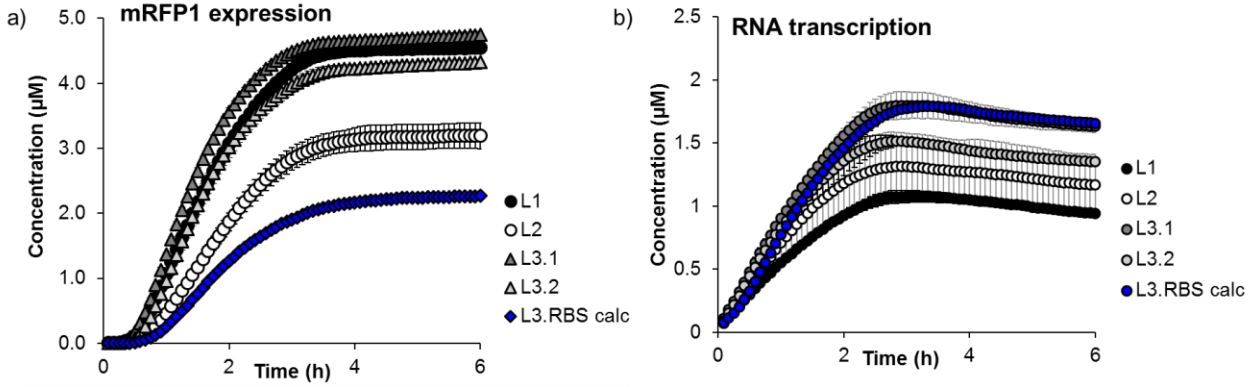


Figure 7. Expression of the fluorescent protein mRFP1 with different leader sequences in PURE system reactions. (a) Translation can be influenced by different leader sequences. (b) Transcription can be influenced as well by different leader sequences.

Finally, we tested three leader sequences using three different fluorescent proteins to test for their consistency. To our surprise, we found a good degree of consistency, while the only significant difference was that the leader 2 was decreasing the expression of only one of the three coding sequences that we tested (Figure 8).

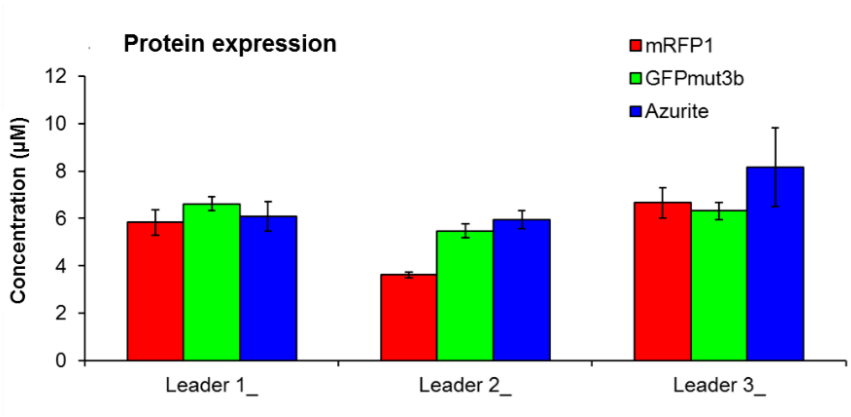


Figure 8. Different leader sequences can influence differently the expression of three distinct fluorescent proteins. The protein mRFP1 is the only protein that shows a reduced expression when using leader 2. Leader 3 here is the same sequence as Leader 3.1 in Table 5 and in Figure 7.

3.2.4. Single-gene characterization of different ribosome binding sites reveals a better control over protein expression compared to the use of T7 transcriptional promoters.

After having briefly investigated the role of the 5'-UTR and of the 3'-UTR on gene expression in PURE system reactions, we proceeded to characterize our set of different ribosome binding sites. We characterized them independently using two different fluorescent proteins: mRFP1 and GFPmut3b (Figure 9a). The characterization resulted in some unexpected data. First, for at least three of the ribosome binding sites there is a significant discrepancy in gene expression between the two fluorescent proteins, while for the majority of the ribosome binding sites the trend is consistent. It is not clear what caused such a divergent behavior for those ribosome binding sites. Then, if we only focus on mRFP1 data the distribution is definitely a step forward compared to the one obtained with the different T7 transcriptional promoters, with more intermediate values of expression. The overall distribution of expression intensities is definitely a step forward, if we compare it with the one obtained using the different T7 transcriptional promoters. On the other hand, however, if we also consider the distribution of the GFPmut3b data, it seems to be once again clustered against a "high" and "low" expression value (Figure

9a). The reasons for the discrepancy in gene expression between the two fluorescent proteins is not clear, however the variability of gene expression when using different ribosome binding sites could play a role. Finally, we tried to correlate the observed expression levels with the parameters that we used to generate the ribosome binding sites. As we can see from Figure 9b, the correlation between the number of bases pairing with the rRNA, the position of the pairing bases within the ribosome binding site, and gene expression is clearer for GFPmut3b than for mRFP1. Moreover, the observed correlation for GFPmut3b also accounts for the "high" and "low" expression values, even though it is definitely an improvement compared to the distribution of protein expression obtained with the T7 transcriptional promoters. On the other hand, the observed correlation for the mRFP1 data seems to be somewhat disturbed and not as clear as the GFPmut3b one. While it is not clear why this is happening, the differences in the coding sequences of the two fluorescent proteins might play an important role. Therefore, while using different ribosome binding sites to control gene expression in PURE system reactions seems to be a promising approach, it might also carry some drawbacks not associated with the use of different T7 transcriptional promoters.

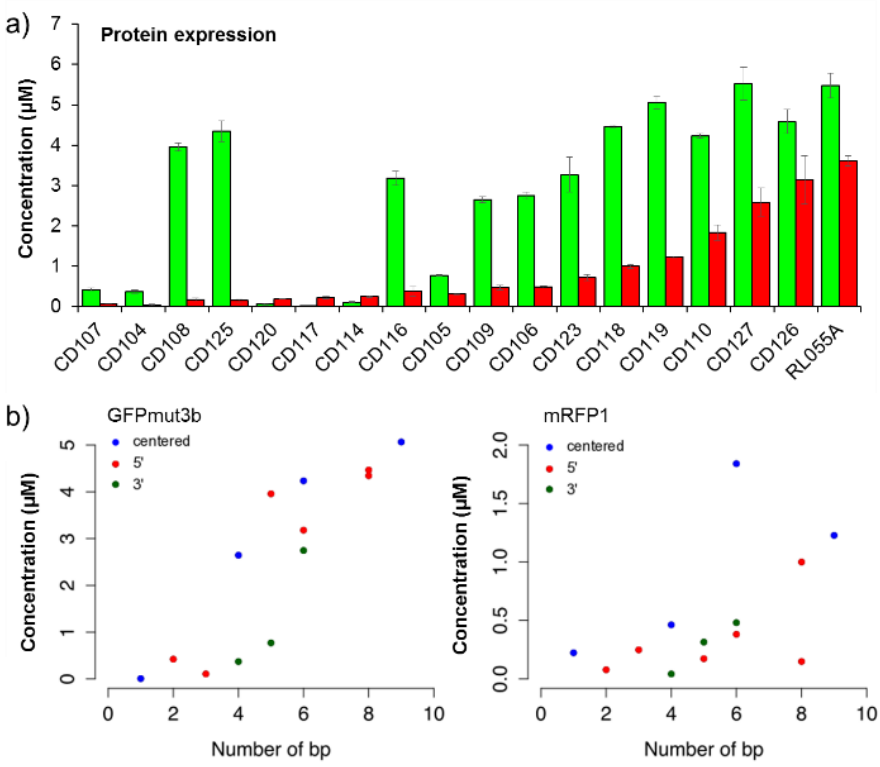


Figure 9. Single gene characterization of the different ribosome binding sites. (a) Different ribosome binding sites were characterized using PURE system reactions as linear, single-gene, templates. The characterization was done expressing both the fluorescent protein mRFP1 and GFPmut3b. Some of the samples differ quite significantly between the two proteins, possibly due to interactions between the ribosome binding site and the mRNA. (b) Relationship between the number of bases that show a perfect match with the ribosome binding site consensus sequence and gene expression. The position of the matching bases is also taken into account. The relationship shows, as expected, that if more bases from the ribosome binding site are able to interact with the ribosomes than a higher level of gene expression is achieved.

3.2.5. A model for genetic expression in PURE system reactions.

After the characterization of T7 transcriptional promoters and ribosome binding sites Michele Forlin built a biological model that could be used to predict the behavior of genetic circuits using different parts. The model describes both transcription and translation in PURE system reactions using 7 species, plus some intermediate ones, and 5 reactions, plus degradations (Figure 10). The model is a simplification of the PURE system components driving the expression, and has been developed following what have been previously done by Stogbauer and colleagues[58]. Resources necessary for transcription and translation were modeled as a single species (R1 and R2, respectively) subject to degradation. Here resources refers not only to the molecular machinery required to sustain transcription and translation, such as the T7 RNA

polymerase, the ribosomes and all the accessory protein factors, but also to the small molecules required to sustain such reactions. Following the PURE system composition[20] we could only define the initial concentration of T7 RNA polymerase and ribosomes. For all the rest of species and for the reaction rates we had to perform parameter estimation based on experimental data.

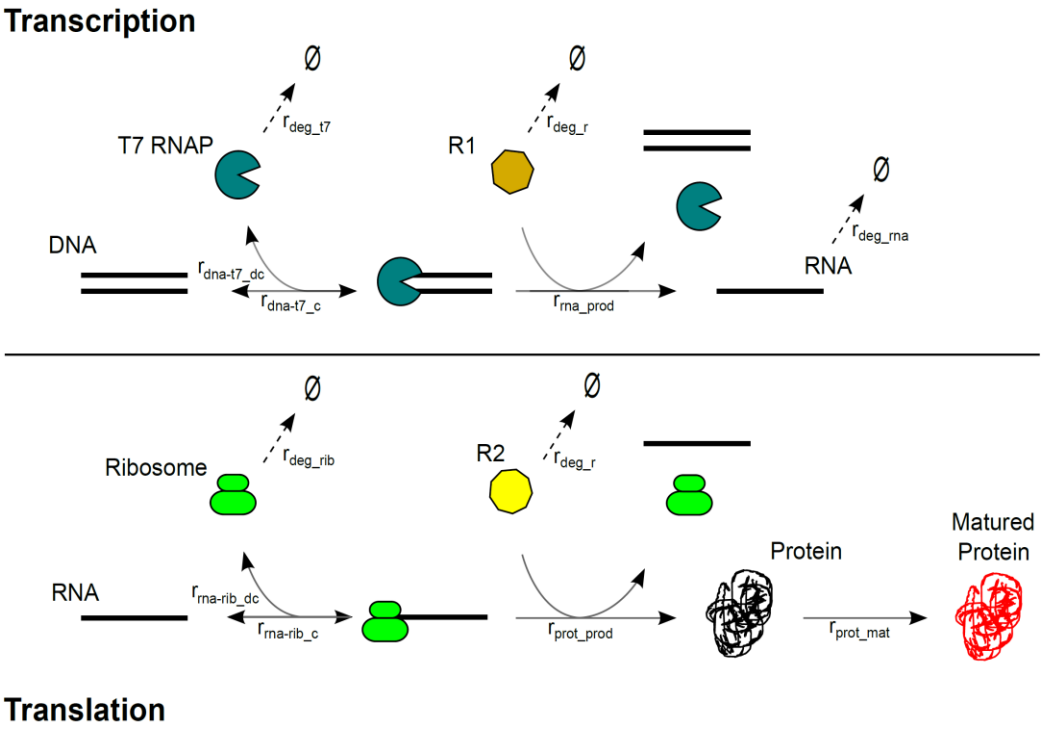


Figure 10. Graphical representation of the biological model used to predict gene expression in PURE system reactions. Upon interacting with the template DNA, the T7 RNA polymerase can either transcribe RNA or detach from the DNA. Moreover, upon transcribing RNA resources R1 will be consumed until transcription halts. The ribosomes will then interact with the transcribed RNA, either by translating it into a protein, or by detaching from the RNA. During translation resources R2 will be consumed until translation is not sustainable anymore. All these steps have been kinetically modeled based on the experimental data acquired.

We implemented the model in COPASI[59], a freely available software. Once implemented we could use the parameter estimation tool within COPASI to infer reaction rates and initial concentration of unknown species (R1 and R2) from the experimental results on transcription and translation kinetic profiles with different T7 transcriptional promoters and ribosome binding sites. Every different T7 transcriptional promoter affected the reaction rates driving the binding and unbinding of T7 RNA polymerase with the DNA template (rdna-t7_c and rdna-t7_dc) while a different RBS affected the binding and unbinding of ribosomes with the mRNA template (rrna-rib_c and rrna-rib_dc). After several rounds of parameter estimation using

different optimization algorithms we were able to identify a set of reaction rates and initial concentrations for R1 and R2 producing a reasonable fit with the experimental evidence. While for some peculiar cases the fitting still requires some adjustment, the resulting model can be used to predict the behavior of desired circuits using different combinations of T7 transcriptional promoter and ribosome binding sites.

3.2.6. Combining different T7 transcriptional promoters and ribosome binding sites improves control over gene expression in PURE system reactions.

We initially used the model to predict gene expression when using one of the possible combinations of promoters and ribosome binding sites. We applied the model to predict the expression of 12 combinations of promoters and ribosome binding sites with the fluorescent protein Azurite (Table 5). We then performed the experiment to test if the data would match the predictions, but unfortunately, that was not the case, with only a few of the samples actually matching the predictions (Figure 11). In order to ensure that the observed result was real we then proceeded to repeat the experiment once more using the same combinations, and again the result was not matching the predictions (Figure 11). This time, however, the data from the second run of the experiment was poorly matching even the data from the first run of the experiment (Figure 11). Several factors might explain this observed inconsistency. First, Azurite, belonging to the class of the blue fluorescent proteins, is one of the least bright fluorescent proteins available. Therefore, the lower level of signal required that the gain of the instrument to be higher compared to the other fluorescent proteins, thus increasing the error in measurement. Then, the coding sequence of Azurite could, for reasons related to the structure of the mRNA, lead to a higher degree of variability, compared to the other fluorescent proteins. Finally, this variability could be an intrinsic part of PURE system reactions and would then be visible for all our fluorescent proteins, regardless of the relative coding sequence.

Table 5. Combinations of promoters and ribosome binding sites.

Sample number	Promoter	RBS	Prediction (μM)
(1)	FC115	CD105	0.45
(2)	FC074	CD105	0.74
(3)	FC095	CD109	1.38
(4)	FC095	CD110	1.8
(5)	FC094	CD109	2.29
(6)	FC115	CD106	2.61
(7)	FC107	CD110	3.03
(8)	FC094	CD118	3.55
(9)	FC089	CD118	4.13
(10)	FC089	CD119	4.57
(11)	FC107	RL055	5.06
(12)	FC090	RL055	5.33

The combination of T7 transcriptional promoters and ribosome binding sites are shown, along with the predicted protein output.

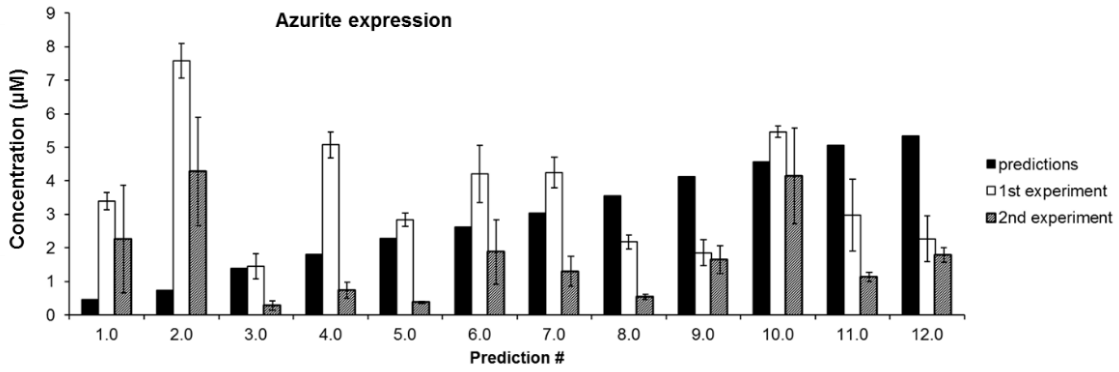


Figure 11. Azurite expression does not match the predicted values. Experiment was repeated twice, highlighting the strong variability associated with the expression of the fluorescent protein Azurite.

In order to have a better understanding of the weight of these different hypotheses in the observed behavior, we decided to repeat the combinations of promoters and ribosome binding sites for the other two fluorescent proteins, mRFP1 and GFPmut3b. For mRFP1 we would also acquire data from transcription using the Spinach aptamer. We would also repeat the data with Azurite, but this time including the Spinach aptamer as well. Finally, we would repeat these

experiments three different times independently, thus allowing us to measure intra and inter-experiment variability, both for transcription and translation.

First, we modified our combinations of promoters and ribosome binding sites, selecting four T7 transcriptional promoters and four ribosome binding sites out of the ones that we characterized. We did not employ our model to choose the promoters and ribosome binding sites, but rather we relied on the characterization data to select for different transcriptional and translational intensities (Table 6).

Table 6. Combinations of promoters and ribosome binding sites.

Promoter code	Strength	RBS code	Strength
FC074	Strong	CD127	Strong
FC115	Mid	CD110	Mid
FC108	Weak	CD109	Weak
FC109	Very weak	CD105	Very weak

T7 transcriptional promoters and ribosome binding sites associated with different degrees of transcriptional and translational intensities are reported.

First, we tested the sixteen combinations (four ribosome binding sites by four promoters) reported in Table 6, monitoring the expression of mRFP1 and Spinach. Data from the transcription matched our expectations, and confirmed what was observed in the previous chapter about the T7 transcriptional promoters (Figure 12a). Looking at translation data, however, again showed the problem of the two clusters of expression, which was either high or low (Figure 12a). Repeating the samples using Azurite and Spinach led again to some unexpected results: regarding transcription, the data was very similar to the RNA levels obtained with mRFP1, thus confirming the high reliability of transcription levels from different T7 transcriptional promoters (Figure 12b). However, translational data are again hard to interpret. Overall, the expression was significantly lower compared to what was observed with the protein mRFP1. Finally, we repeated the same samples with the protein GFPmut3b, this time without the Spinach aptamer due to the overlapping spectral properties of the two systems. Again, the expression was somewhat reduced compared to mRFP1 (Figure 13).

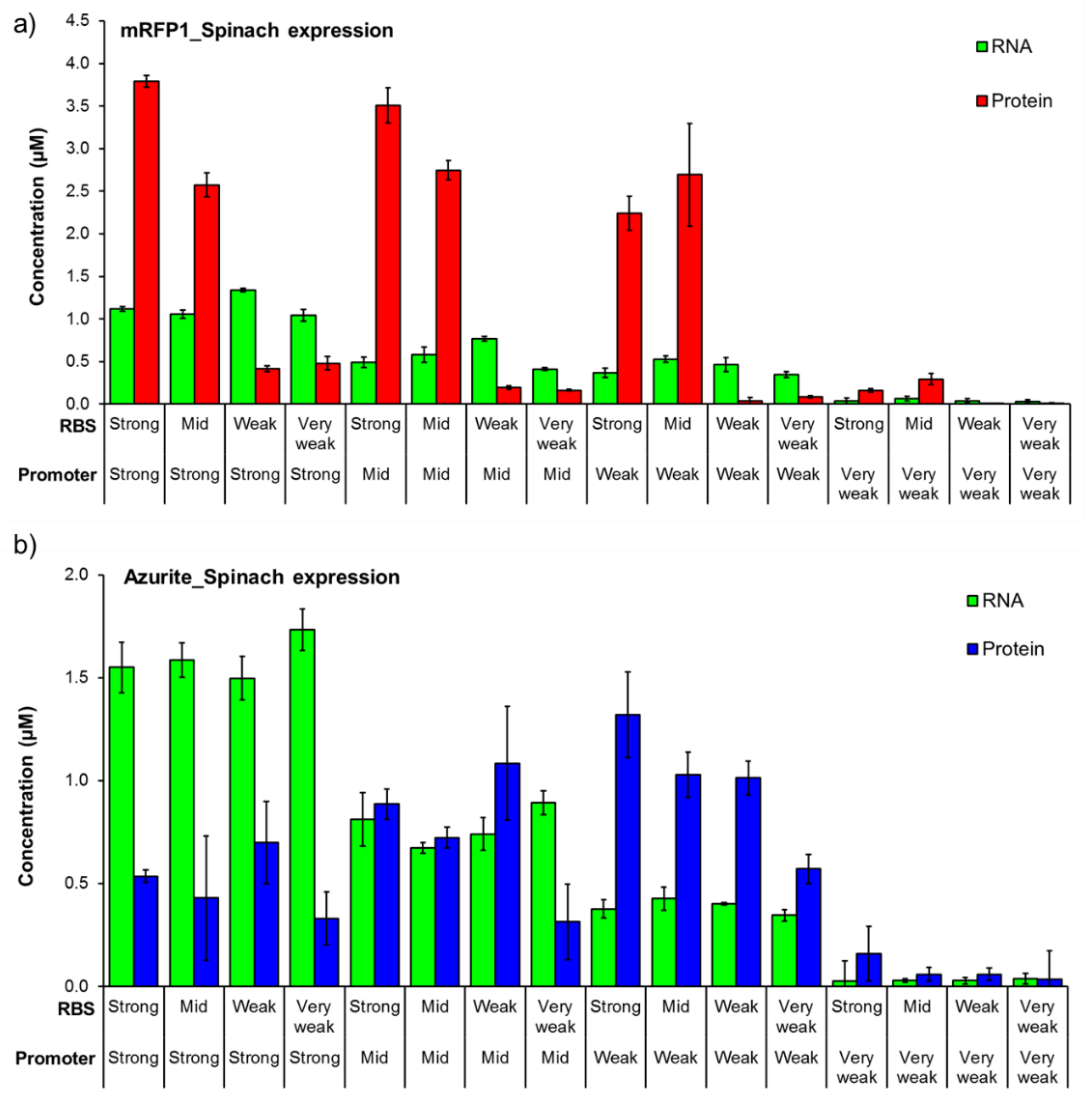


Figure 12. Single gene expression of the fluorescent proteins mRFP1 and Azurite employing different combinations of T7 transcriptional promoters and ribosome binding sites in PURE system reactions. (a) Using different T7 transcriptional promoters results in a difference in RNA levels matching our previous data. The different ribosome binding sites, on the other hand, are not as predictable in their behavior. (b) Again, using different T7 transcriptional promoters results in a difference in RNA levels matching our previous data, thereby highlighting the reliability of T7 transcriptional promoters regardless of the coding sequence to be expressed. Translational data, however, clearly shows a less then optimal expression of the fluorescent protein Azurite, thereby highlighting the more important role of the coding sequence in translation compared to transcription.

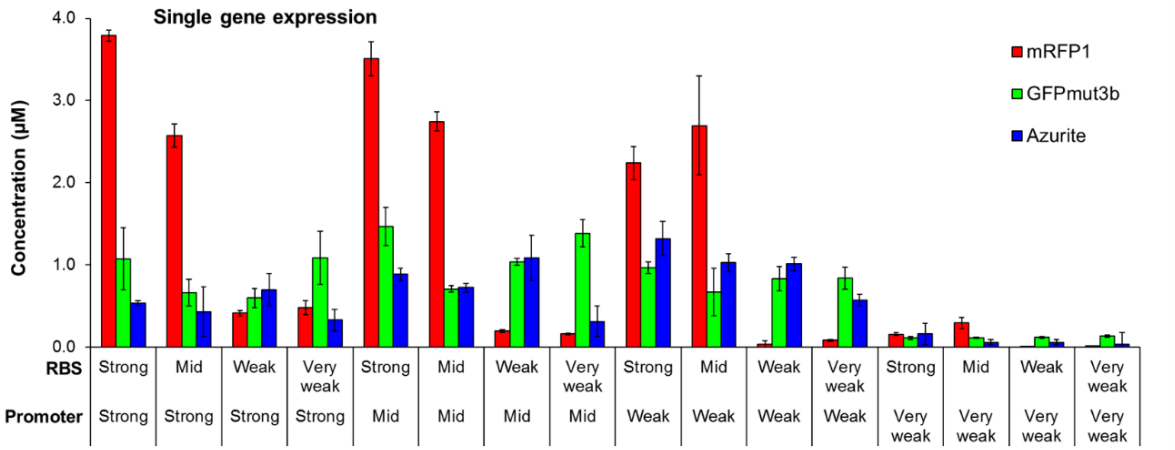


Figure 13. Combinations of T7 transcriptional promoters and ribosome binding sites for single-gene expression of different fluorescent proteins.

Next, because all the tested combinations with the two weakest promoters resulted in an extremely weak output, we chose to discard the samples with the two weakest promoters, and focus only on the remaining eight samples. Then, we used the eight samples to run two additional replicates of the experiment in order to probe the variability arising from PURE system reactions. The numerical indicator that we used to describe the measured variability is the coefficient of variation. The coefficient of variation is defined as the ratio of the standard deviation to the mean. We report two different coefficients of variation: the "intraday" and the "extraday". The "intraday" coefficient refers to the variability that is observed within the single triplicate, while the "extraday" coefficient describes the variability observed among the different, independent, triplicates. Therefore, in the "intraday" coefficient we will mostly find the variability coming from the physical process of assembling each reaction by mixing the different components. In the "extraday" coefficient, on the other hand, we will find an indication of the variability coming from different batches of PURE system. The result is quite interesting (Figure 14). First, we can see that generally transcription shows a lower variability, compared to translation (Figure 15). This is probably because the transcriptional process is a 1-step process that involves only a single protein, T7 RNA polymerase. Then, we can see that the variability for Azurite expression is significantly higher than mRFP1 (Figure 15). Again, this is an indication that a different coding sequence can lead to unpredictable and possibly disruptive effects.

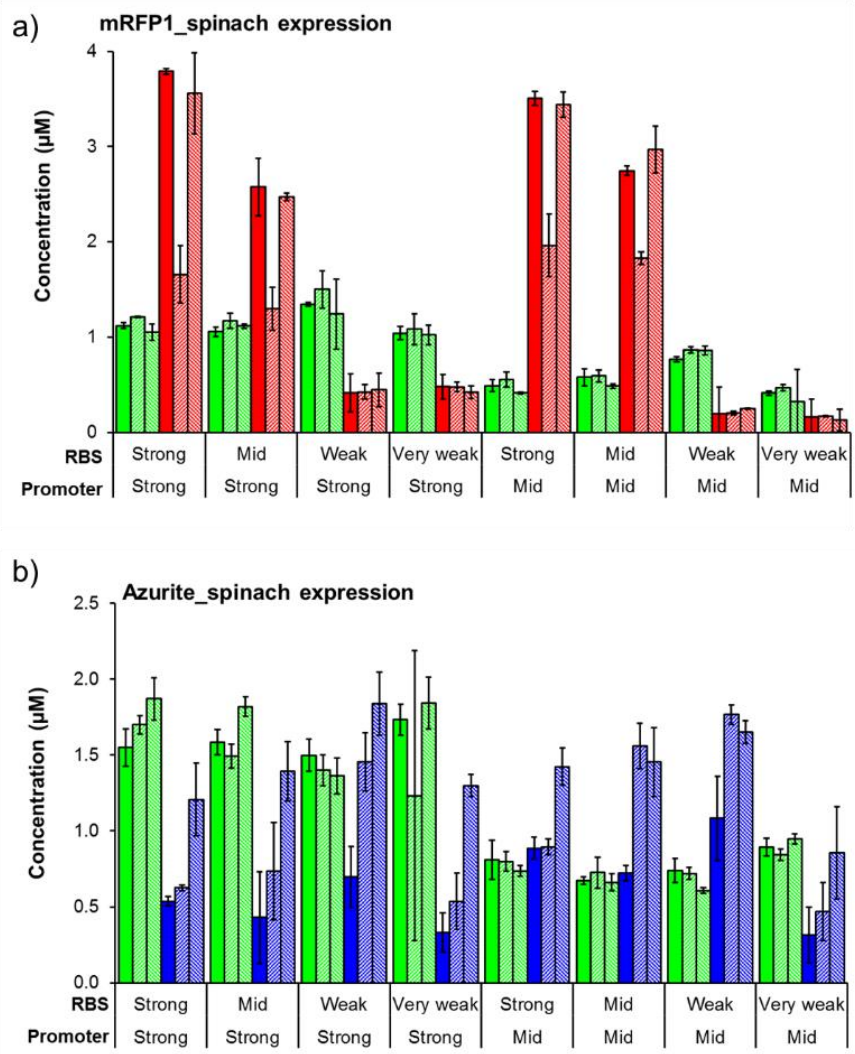


Figure 14. Transcription is more consistent across different batches of PURE system reactions, compared to translation. (a) Three independent experiments of single gene expression of the reporter mRFP1_spinach clearly shows that transcription is more consistent than translation. (b) Similarly, three independent experiments of Azurite_spinach expression again show how transcription is more consistent than translation in PURE system reaction.

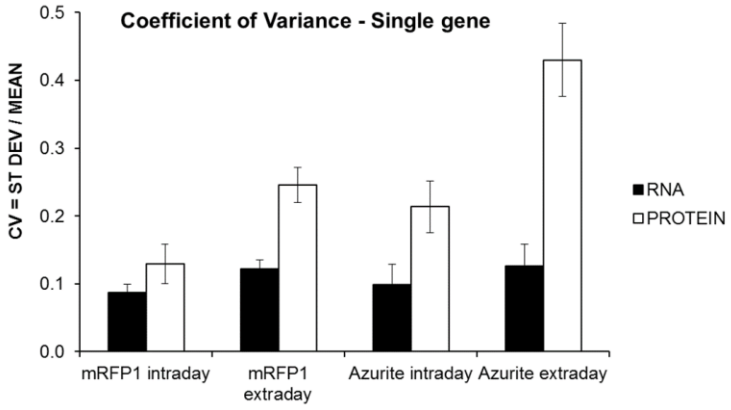


Figure 15. Coefficients of variation (CV) for the expression of the two fluorescent proteins mRFP1 and Azurite. CVs are shown both for transcription and translation. We can see how transcription is more consistent between different batches of PURE system in respect to translation by comparing the "intraday" and "extraday" RNA and PROTEIN CVs, for both fluorescent proteins. Moreover, the CV of Azurite translation is significantly higher than the CV of mRFP1 translation, thereby suggesting a role of the coding sequence in gene expression variability.

3.2.7. Removing predicted internal ribosome binding sites from Azurite coding sequence does not improve expression in PURE system reactions.

In order to see if we could overcome the problem of Azurite low expression and high variability we tried to look for internal ribosome binding sites (iRBSs) in the coding sequence of the protein. These internal ribosome binding sites are parts of the coding sequence of a protein that, because of their sequence composition, and because of the structure of the mRNA, can recruit ribosomes and drive translation. These sequences were recently described in a paper that computationally analyzed several coding sequences for internal ribosome binding sites[60]. Their role, however, has not been fully elucidated yet, but internal ribosome binding sites are suspected to reduce the expression of a gene, mainly due to the translational pausing[61]. Moreover, to date no study investigated the influence of internal ribosome binding sites on cell-free transcription and translation reactions. Therefore, we decided to use the dedicated tool that can be found online at the website of the RBS calculator, to check for internal ribosome binding sites in both mRFP1 and Azurite coding sequences. The computational tool indeed reported a difference in the number and the intensity of the internal ribosome binding sites between the two coding sequences, with the strongest iRBSs being present in the Azurite coding sequence. Next, we selected the four strongest internal ribosome binding sites, and we modified the coding sequence in order to remove the iRBSs, while at the same time retaining the same amino

acid composition of the fluorescent protein. Unfortunately, when we did a comparison by expressing the two different versions of Azurite in PURE system reactions, there was virtually no difference between the two coding sequences (Figure 16).

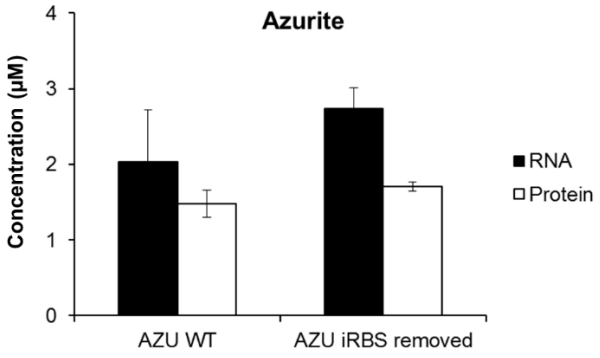


Figure 16. Removing predicted internal ribosome binding site in Azurite coding sequence did not improve the expression of the fluorescent protein in PURE system reactions. Final concentration of both mRNA and protein are very similar between the two different templates.

3.2.8. Visualizing the achieved control of protein expression and the relative variation by composing a picture with PURE system reactions.

Then, we sought to find a way to display both the different gene expression intensities that we achieved, and the associated variability. As we are expressing fluorescent proteins, we decided to exploit the expression of fluorescent proteins in different concentrations to compose a simple picture. The final picture would be composed by a series of "pixels." Each pixel is physically a well in a 1536-well plate filled with 2 μL of a PURE system reaction incubated with a different linear DNA template. Each DNA template was chosen among four different templates that elicit a different transcriptional and translational output. The template DNA chosen for each pixel reflected the "intensity" of the color required to compose the picture. By using both mRFP1 and the Spinach aptamer, we managed to successfully compose two pictures of the "Yin-yang" symbol. We created the pictures by using the same set of reactions, only that for the first picture we exploited the different RNA levels arising from transcription and detected using the Spinach aptamer. For the second picture, we used the different protein levels arising from translation and detected using the mRFP1 fluorescent protein. We employed a fluorescence scanner to record fluorescence signals. When assembling the reactions for the picture we also made sure to use

the highest number of different batches of PURE system reaction compatible with the workflow. By looking at the result (Figure 17), we can see the difference in variability by comparing the RNA and the protein pictures. Moreover, we can also compare the two pictures with two pictures created by a computer program applying both the "intensity" of each pixels color and the variability of such "intensity" (Figure 16). It interesting to compare the predicted pictures with the actual pictures. It is also interesting to note the partial control that we reached over protein expression, with the "strongest" color, the full black, being easily identifiable, while the two weaker shades of black are not as distinguishable. This quite correctly graphically represents what we have seen in the previously described experiments.

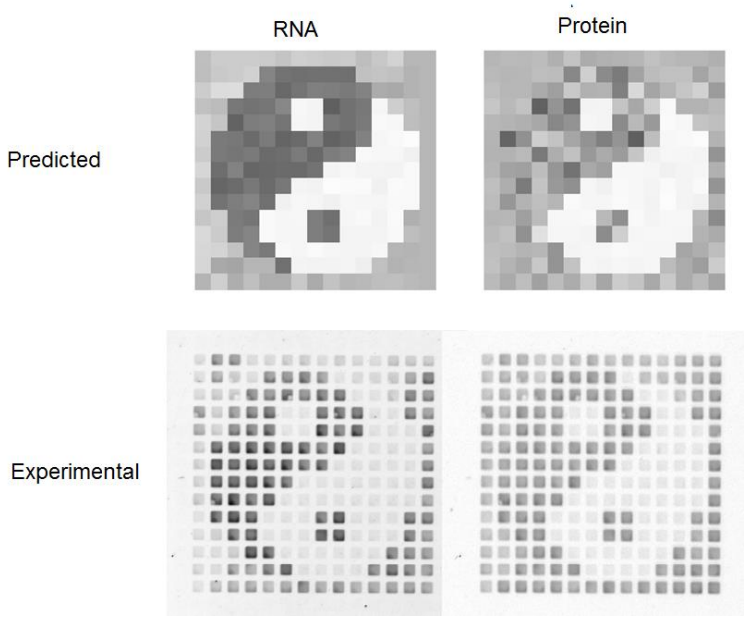


Figure 17. "Yin-Yang" pictures reconstituted using PURE system reactions placed in a 1538 wells plate. The pictures were reconstituted both using RNA concentration (via the Spinach aptamer) and protein concentration (via the fluorescent protein mRFP1). The final pictures can be compared with the pictures generated taking into consideration the noise associated with transcription and translation in PURE system reactions.

3.2.9. A simple cascade circuit in PURE system reactions successfully achieves intermediate levels of protein expression with minimal increase in variability.

Then, we decided to use our set of characterized parts in order to explore how their use would influence the behavior of two simple genetic circuits: a cascade and a repressor. Understanding how these circuits work in PURE system reaction is important, as genetic circuitry will be

necessary to increase the complexity of artificial cells. Moreover, the number of articles dealing with this kind of optimization, in PURE system reactions, is very limited. To set up our first circuit, the cascade, we relied on the protein T3 RNA polymerase to propagate the signal. Briefly, a T7 promoter (the first point of regulation) controls the expression of the T3 RNA polymerase, along with a ribosome binding site (the second point of regulation). Then, a T3 promoter controls the expression of our reporter composed of the fluorescent protein mRFP1 and the Spinach aptamer, again along with a ribosome binding site (the third point of regulation).

We decided to use, as variable parts, three different T7 transcriptional promoters and three different ribosome binding sites, out of the four that we tested in the previous experiments, ranking them as "strong", "medium" and "weak" (Table 7).

Table 7. Promoters and ribosome binding sites combinations for the genetic circuits.

Promoter code	Strength	RBS code	Strength
FC074	Strong	CD127	Strong
FC108	Mid	CD110	Mid
FC109	Weak	CD109	Weak

T7 transcriptional promoters and ribosome binding sites associated with different degrees of transcriptional and translational intensities are reported.

However, the total number of possible arrangements, considering that we have three different positions in which there is a variable part (as we decided not to modify the T3 transcriptional promoter), is 27. Therefore, we decided to apply our model in order to decrease the number of samples that we had to test. Using this method, we selected 13 samples to test. We also decided to compare two different genetic architectures of our circuit: the "divided" architecture in which the two genes are placed onto two different pieces of DNA, and thus added independently to the reaction, and "united" architecture in which the two genes are placed on the same piece of DNA, and thus added together to the reaction (Figure 18). As the majority of articles reporting the expression of multiple proteins in PURE system reaction relied on the use of multiple pieces of DNA (one for each expressed components), we wanted to understand if this approach could have any significant impact on both expression and variability of expression of our reporters.

Therefore, we performed the 13 samples, using the two different genetic architectures, and we repeated every experiment three times independently. The goal is to use the resulting data to assess the variability of the outcome, and to compare it with the variability measured for the single-gene constructs.

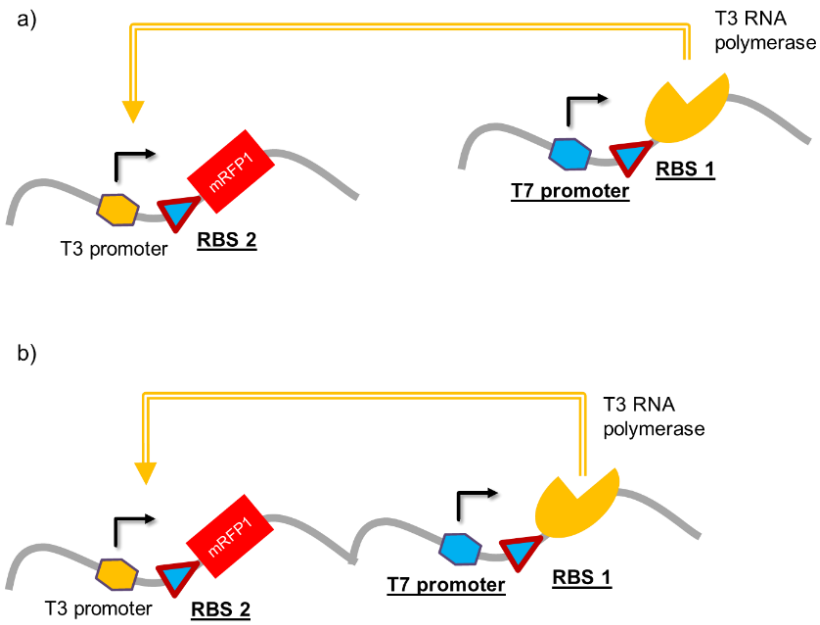


Figure 18. The cascade genetic circuit. Briefly, the expression of T3 RNA polymerase by a variable T7 transcriptional promoter and a variable ribosome binding sites leads to the expression of the mRFP1_spinach reporter through a T3 transcriptional promoter and a variable ribosome binding site. (a) The two genes are placed on two different constructs. (b) The two genes are placed on the same construct.

After performing all these experiments, the results were quite interesting. First, it is surprisingly clear, even without performing the coefficient of variation analysis, that the genetic architecture in which the two genes are on the same piece of DNA is superior to the one in which the two genes are on two different pieces of DNA (Figure 19). This effect is striking both in terms of expression and of stability of expression, and is observed both for transcription and for translation (Figure 19). Perhaps this effect is due to the increased variability in template concentration when using two different pieces of DNA. It should be noted, however, that, at least for one sample, also using the one piece of DNA leads to a high variability in terms of protein expression (Figure 19).

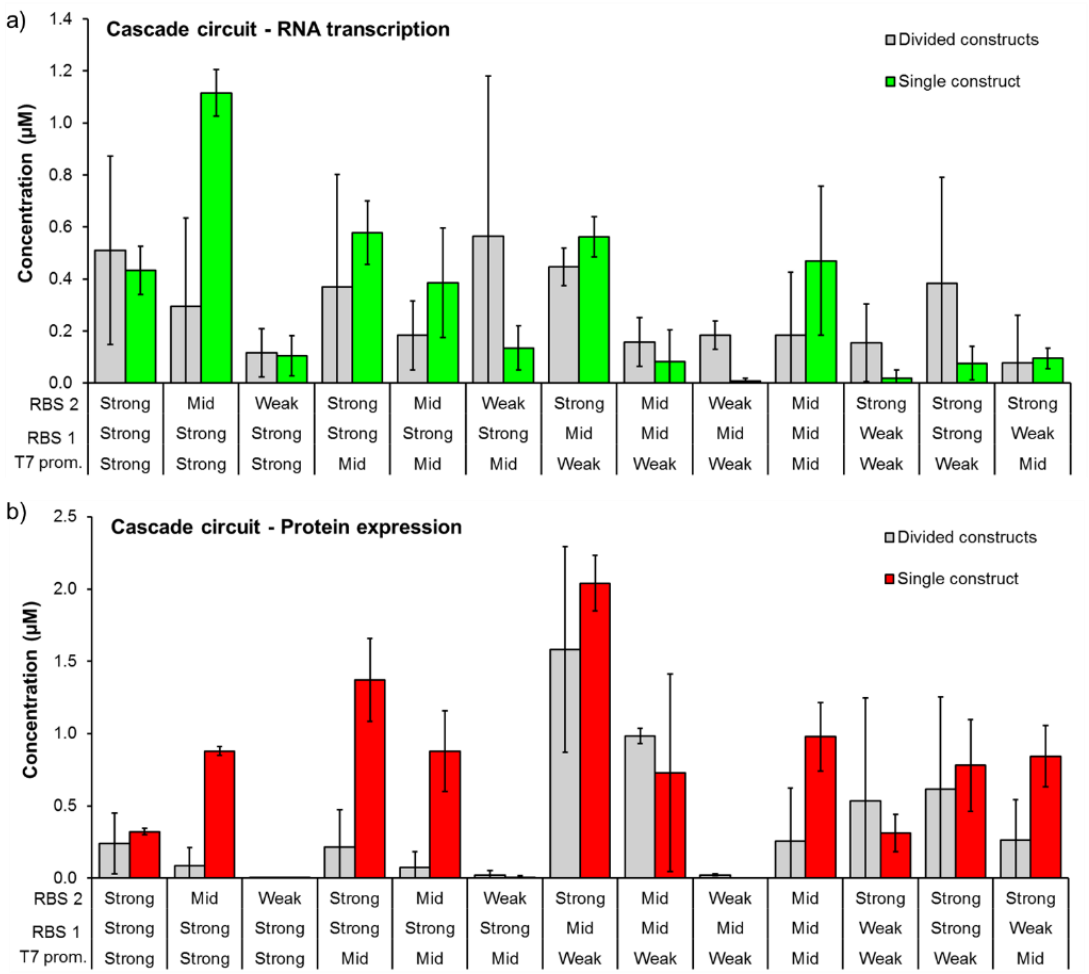


Figure 19. Cascade circuit performance is significantly influenced by the genetic architecture employed. Data from three independent experiments is aggregated and plotted for transcription and translation of the reporter gene mRFP1_spinach (a) Transcription is affected by the employed genetic architecture. Specifically, having the two genes on the same construct reduced transcriptional variability across three independent experiments. (b) Translation is greatly affected by the employed genetic architecture. Having the two genes on the same construct greatly improved protein expression, while at the same time reducing protein expression variability.

It is not clear whether this variability is dependent upon the specific parts (promoters and ribosome binding sites) that were employed in that sample. That sample, also, illustrates how there seems to be a threshold concentration of mRNA, below which there is barely detectable protein production, and above which there is good protein synthesis (Figure 19). However, in some other samples, a higher concentration of mRNA is not directly correlated with a higher concentration of protein, but quite the contrary, on a lower protein concentration (Figure 19). This could be an indication of the limited resources available in the PURE system. Finally, it is

interesting to note that the “Strong, Strong, Strong” configuration, in which all the possible points of regulation show the highest possible over expression, is actually one of the worst samples, both in terms of transcription and translation of the reporter gene (Figure 19). This is one of the first indications that, as much as *in vivo*, also *in vitro* it is not always the best option to over express everything, when trying to implement a genetic circuit. Specifically, in this circuit it seems that just a small concentration of T3 RNA polymerase is enough for an optimal performance, while increasing the expression of the T3 RNA polymerase has a negative impact on the expression of the reporter. Another interesting thing that we observed is that, when applying this genetic circuit, we achieved different intermediates levels of gene expression, which is something that we were struggling to achieve when using our single-gene constructs. Our impression is that the appearance of such intermediates levels of protein expression is due to the modulation of expression of the T3 RNA polymerase. We achieved different concentrations of T3 RNA polymerase by applying our different genetic parts. Because the commercial PURE system employs a defined concentration of T7 RNA polymerase, it was not possible to test lower concentrations of the T7 RNA polymerase. However, the results obtained with the cascade circuit clearly indicates that modulating the concentration of T7 RNA polymerase in the PURE system could be a good strategy for achieving a better control over protein expression. Finally, we did not drop considerably in terms of protein expression levels, therefore suggesting that this sort of circuit could help in regulating protein expression in PURE system reactions. Finally, we performed both the “intraday” and “extraday” coefficient of variation analysis. Interestingly, the variability slightly increases, for both transcription and translation, and especially the “extraday”. This is an indication that as the complexity of the expressed genetic circuits will increase also their variability might increase (Figure 20).

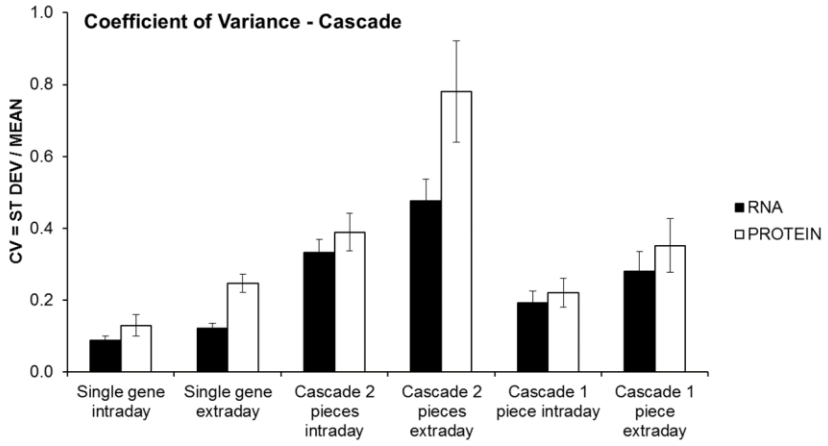


Figure 20. Comparison of the transcription and translation CV of single gene expression with the cascade circuit in the two tested genetic architectures. The architecture with the two genes on two separate constructs is the one with the highest CV, therefore with the highest variability, both of transcription and translation. The architecture with the two genes on the same construct shows a modest increase in all the CVs, compared to the single gene expression CVs, probably because of the intermediate step of T3 RNA polymerase expression.

Finally, we decided to apply the same methodology we applied before and select four samples to make another two "Yin-Yang" pictures, one for each genetic architecture. The result again shows quite clearly the difference between the two genetic architectures (Figure 21). It is interesting to note how different is the variability between the different pixels of the pictures with the two architectures. Moreover, it is also interesting to note that the pictures done using the architecture with the two genes on the same piece of DNA shows intermediate levels that are even better than the single gene picture, highlighting again how important probably is to accurately control the concentration of the RNA polymerase present in PURE system reactions.

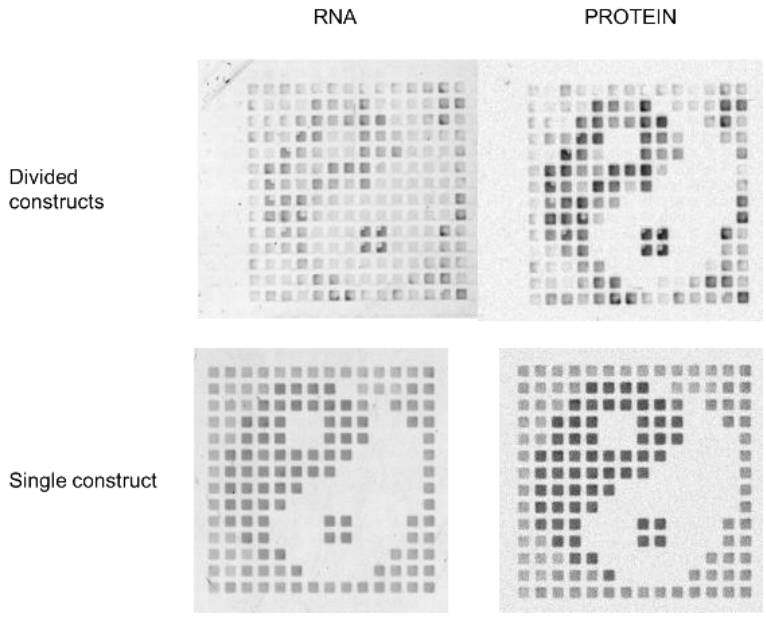


Figure 21. Comparison of four "Yin-Yang" pictures reconstituted using the two different genetic architectures of the cascade circuit. Consistent with previous data, we can immediately see that the genetic architecture with the two genes on the same construct is associated with a lower variability both in transcription and translation compared to the genetic architecture with the two genes on two different constructs.

3.2.10. Using characterized parts to implement a repressible circuit in PURE system reactions: the problem of the "off" state.

Next, we decided to investigate the use of our different genetic parts (promoters and ribosome binding sites) on a repressible circuit. The repressor that we decided to use is EsaR, a LuxR homologue from the organism *Pantoea stewartii* subsp. s*tewartii*. EsaR is part of the quorum-sensing regulatory pathway used by bacteria to communicate. The protein binds a specific DNA sequence (EsaR operator), and releases the DNA when a signal molecule, part of the quorum-sensing system, 3-oxohexanoyl-homoserine lactone (3OC6HSL), binds the repressor.
Therefore, when the repressor is present without the 3OC6HSL it can block transcription from a promoter carrying its cognate operator, while when the 3OC6HSL is present the repression is relieved and transcription can take place. For the experiments described here, we did not use the wild-type version of EsaR, but we used a modified version, EsaR-D91G, which was selected for an increased sensitivity to the 3OC6HSL[62].

First, we had to test if the EsaR repressor could be expressed and be functional in PURE system reactions, and if the EsaR operator could be applied to a T7 transcriptional promoter. To do so, we inserted the operator at increasing distances from the +1 of the T7 transcriptional promoter, therefore identifying the best positioning of the operator (Table 8, Figure 22).

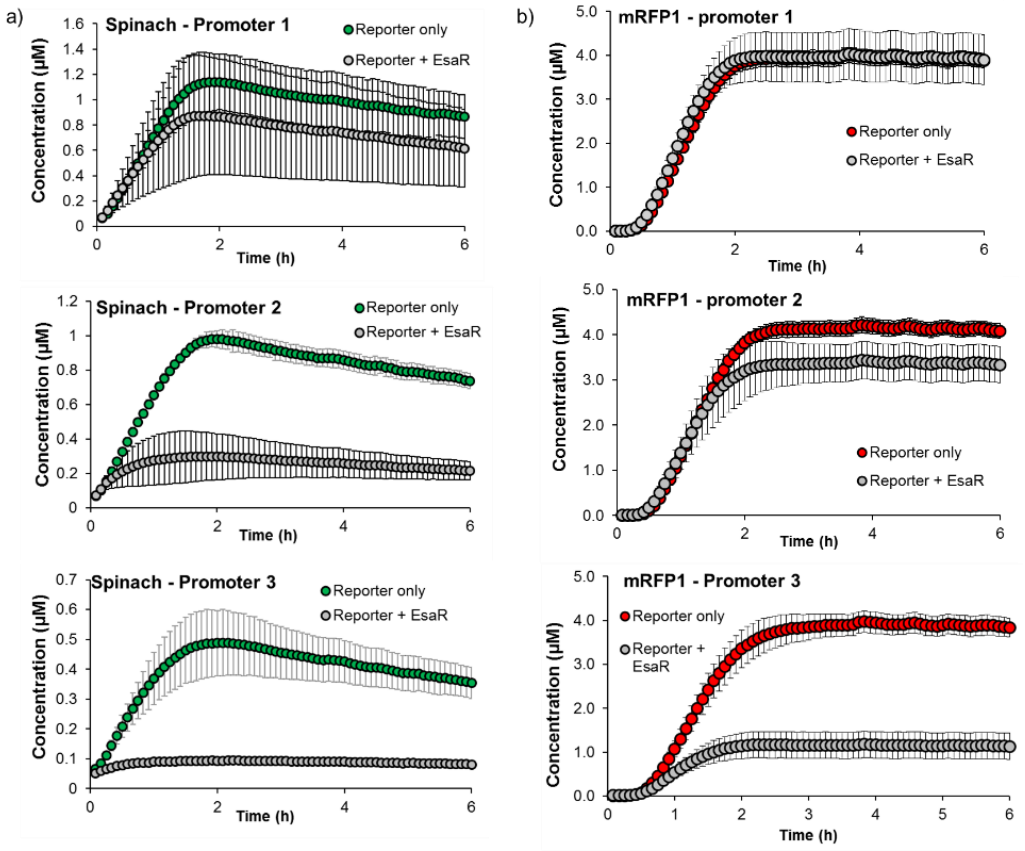


Figure 22. Repression of gene expression increases as the EsaR operator is closer to the transcriptional starting site. The EsaR operator is places in three different positions (Table 8), then both transcription and translation are measured using the reporter mRFP1_spinach, with or without co-expressing the transcriptional repressor EsaR. (a) Transcription data shows that the closest the operator is to the transcriptional starting site, the more effective the repression will be. (b) Translational data confirms the previous observation.

Table 8. T7 promoters with EsaR operator.

Promoter code	T7 promoter	+1	Spacer	EsaR operator
Promoter 1	CCGGTTAATACGACTCACTATA	G	GGAGAttgtgagcg	GCCTGTACTATAGTGCAGGT
Promoter 2	CCGGTTAATACGACTCACTATA	G	GGA	GCCTGTACTATAGTGCAGGT
Promoter 3	CCGGTTAATACGACTCACTATA	G		CCTGTACTATAGTGCAGGT

The EsaR operator is placed at different distances from the transcriptional starting site.

Then, we screened different concentrations of the 3OC6HSL that again resulted in the best behavior of the circuit. After that, we assembled our circuit. Briefly, one T7 transcriptional promoter controls, along with a ribosome binding site, the expression of the EsaR repressor, while another T7 transcriptional promoter, this time harboring also the EsaR operator, along with another ribosome binding site, drives the expression of our reporter, again the fluorescent protein mRFP1 and the Spinach aptamer (Figure 23). This time, having four different nodes in which we can insert our three parts of different intensity, the total number of possible combinations increased to 81. Therefore, we employed our model again to select eight samples with the best ON/OFF ratio and overall expression. This time we used only one genetic architecture, with the two genes on the same piece of linear DNA, as we shown with the previous circuit that this is the best configuration. Again, we performed the experiment three different times independently, to assess the variability of our results.

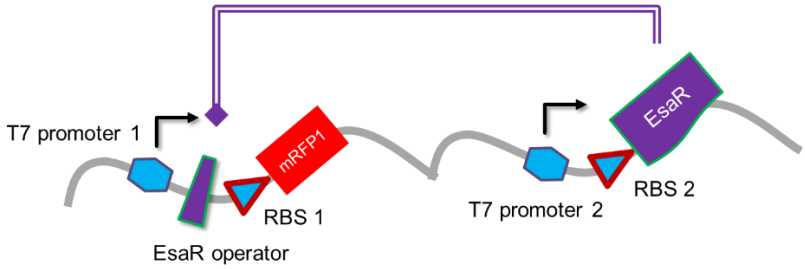


Figure 23. The repressor genetic circuit. Briefly, the transcriptional repressor EsaR is under the control of a variable T7 transcriptional promoter and of a variable ribosome binding site. When expressed, the repressor is able to bind to the EsaR operator, thereby blocking transcription of the reporter gene mRFP1_spinach.

The system is working great if we look only at the transcription levels (Figure 24). Five samples out of the eight we tested showed a good ON/OFF ratio, while at the same time reaching a good level of transcription. On the other hand, when we look at protein levels, background expression from the OFF state increases, probably because, as we have seen before, just a tiny amount of mRNA is enough to translate a significant amount of protein (Figure 24). While the two samples with the best overall expression still show a 3-fold and 4-fold difference between the OFF and the ON state, it is only for the samples with a significantly lower expression range that the ON/OFF ratio improves. Again, this is probably because the expression of both proteins is under the control of T7 promoters. While expression from the reporter is probably shut down as

soon as EsaR is functional, still there is a period in which it is not, and that is why we see the generally high OFF state in the tested samples. Another interesting thing that we can note in this circuit is that, as we have previously seen for the cascade circuit, over-expressing everything by using a "strong" part for each regulatory node does not result in the best circuit performance (Figure 24). Finally, our coefficient of variation analysis shows values that are similar to the coefficients of variation obtained for the cascade circuit, albeit slightly lower for translation of the reporter gene (Figure 25). This is probably because, since there is not really a cascade here, the variability does not propagate as much as it did in the cascade circuit.

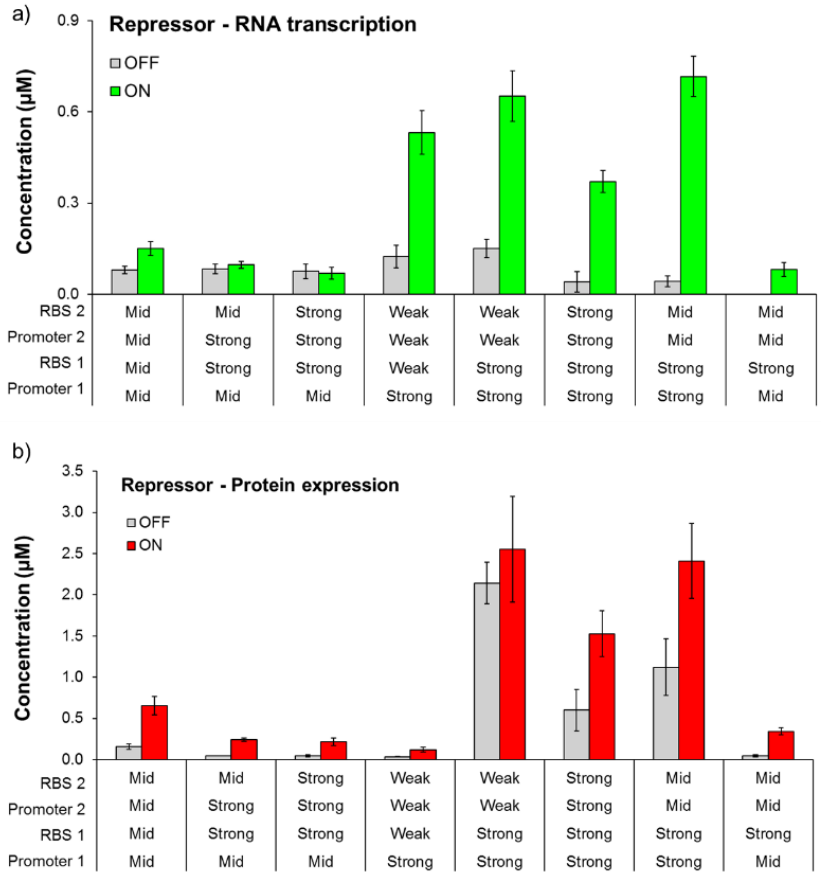


Figure 24. The repressor circuit in PURE system reactions. (a) When looking at the transcription data the repressor circuit works perfectly, with many samples showing good expression in the "on" state, and very low expression in the "off" state. (b) Translation data highlights how the main issue associated with this genetic circuit is the high "off" state.

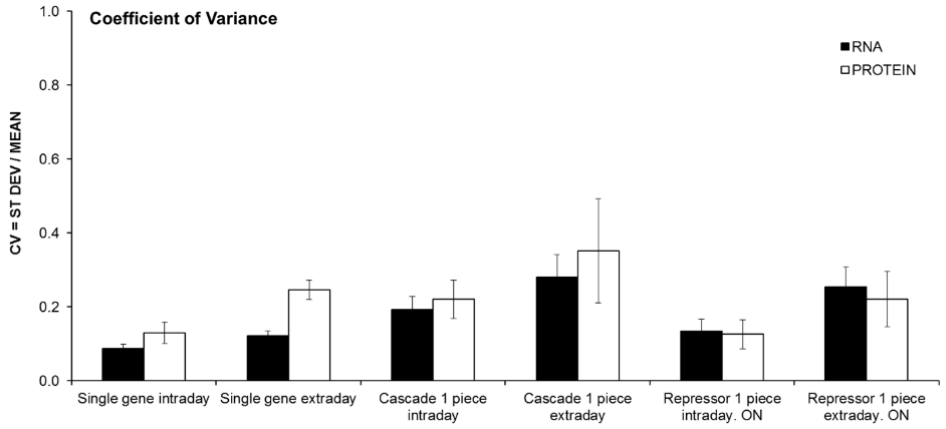


Figure 25. Both the transcriptional and translational CVs in the repressor circuit are comparable to the transcriptional and translational CVs of the cascade circuit.

Finally, we decided to compare our experimental results with some predictions that we generated using our model. We generated predictions of the final protein concentration both for the "off" and the "on" state. As we can see from Figure 26, the experiment and the predicted results match quite nicely, even though the experimental results seem to be slightly scaled down compared to the predicted ones. Therefore, in this case the model could be used to reduce the number of required experiments, predicting the outcome of different combinations of T7 transcriptional promoters and ribosome binding sites.

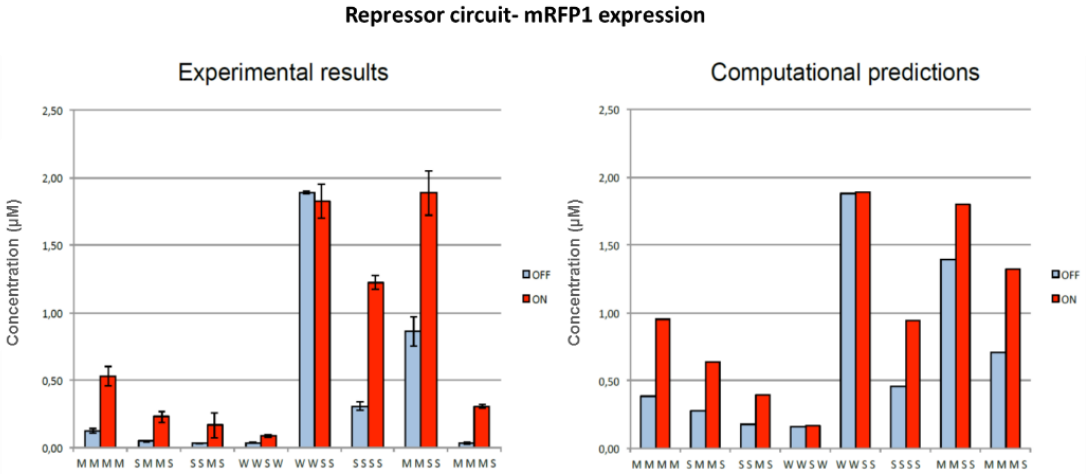


Figure 26. Experimental values from the repressor circuit in PURE system reactions match closely computationally predicted values, albeit being slightly scaled down.

3.2.11. E. coli transcriptional promoters readily allow for intermediate levels of protein expression in cellular extract reactions.

Several of the effects that we observed could derive from some specific features of the PURE system, such as the use of the T7 RNA polymerase, or the lack of other factors that could, on the other hand, be present in the cellular extract. Therefore, we decided to begin exploring the same questions, but using cellular extract reactions, instead of PURE system reactions. Particularly, we would like to explore the differences in regulation and control of protein production when using the endogenous *E. coli* RNA polymerase, in contrast to the viral T7 RNA polymerase found in the PURE system.

First, we tried to set up a reporter to detect both transcription and translation in cellular extracts reaction, similarly to what we have shown with the PURE system. For translation, we would use again the red fluorescent protein used so far in PURE system reactions, mRFP1. For transcription, we tried to use the Spinach aptamer, but the signal was extremely weak, and barely distinguishable from the base line, so we discarded it (Figure 27).

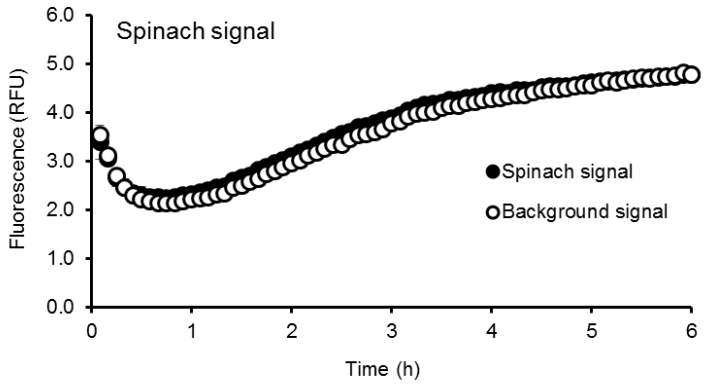


Figure 27. The fluorescence signal from the Spinach aptamer is negligible in cellular extract reactions. Here it is plotted against the background signal from a cellular extract reaction with no DNA template.

Next, we tried to use a molecular beacon with a 2'-O-methylribonucleotide backbone, as described in an article by Marras and colleagues[63]. However, when trying the molecular beacon in cellular extract reactions, we soon realized that something is creating an unspecific signal that is present even if there is no template DNA present in the reaction, coming from the

molecular beacon (Figure 28). It seems as if something is interacting with the beacon leading to it to open up and therefore spiking the fluorescence signal. We, therefore, abandoned the use of the fluorescent probe.

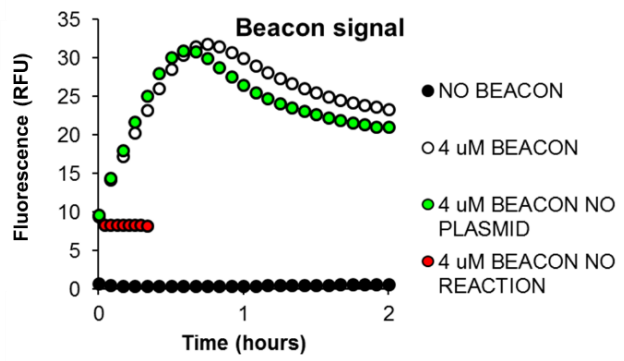


Figure 28. Modified 2'-O-methylribonucleotide beacon shows an unspecific fluorescent signal in cellular extract reactions. Fluorescent background is negligible in reactions without the molecular beacon. The background is higher, but stable, when the beacon is present without the cellular extract reaction. Finally, the beacon shows the same fluorescence kinetic profile in a cellular extract reaction both with and without the template DNA encoding the RNA sequence required to open the molecular beacon.

Next, we decided to start characterizing different *E. coli* transcriptional promoters, and we considered enough for this preliminary stage consider only the final protein output. First, we tested five of the *E. coli* σ_{70} promoters from the Registry of Standard Parts, comparing their performance with the T5 *E. coli* σ_{70} promoter (Table 9, Figure 29).

Table 9. E. coli promoters.

Promoter code	-35 box	spacer	-10 box	+1
T5 promoter	TTGCTT	TGTGAGCGGATAACAAT	TATAAT	AGATTCA
J23100	TTGACG	GCTAGCTCAGTCCTAGG	TACAGT	GCTAGCA
J23104	TTGACA	GCTAGCTCAGTCCTAGG	TATTGT	GCTAGCA
J23116	TTGACA	GCTAGCTCAGTCCTAGG	GACTAT	GCTAGCA
J23117	TTGACA	GCTAGCTCAGTCCTAGG	GATTGT	GCTAGCA
J23118	TTGACG	GCTAGCTCAGTCCTAGG	TATTGT	GCTAGCA

Series of E. coli transcriptional promoters tested with cellular extract reactions.

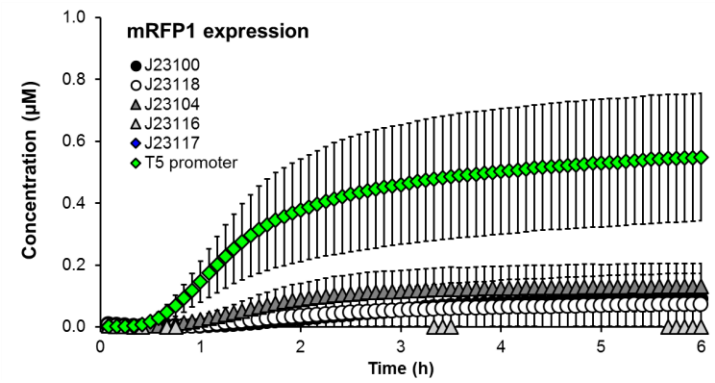


Figure 29. Low expression of the fluorescent protein mRFP1 when using the reported E. coli transcriptional promoters.

However, because of their relatively poor performance, we decided to try another approach. We kept the T5 *E. coli* promoter sequence, and we replaced either the -35 or the -10 box with sequences from other *E. coli* promoters, such as the *tac* promoter (Table 10). This approach proved to be better (Figure 30), and we readily identified five novel *E. coli* promoters, which interestingly produced also intermediate levels of protein expression, which is something that we struggled to achieve with PURE system reactions.

Table 10. E. coli transcriptional promoters.

Promoter code	-35 box	spacer	-10 box	+1
T5 promoter	TTGCTT	TGTGAGCGGATAACAAT	TATAAT	AGATTCA
T5.1	TTGCTA	TGTGAGCGGATAACAAT	TATAAT	AGATTCA
T5.2	TTGACA	TGTGAGCGGATAACAAT	TATAAT	AGATTCA
T5.3	TTGACG	TGTGAGCGGATAACAAT	TATAAT	AGATTCA
T5.4	TTGCTT	TGTGAGCGGATAACAAT	TATTAT	AGATTCA
T5.5	TTGCTT	TGTGAGCGGATAACAAT	TACTAT	AGATTCA

Series of E. coli transcriptional promoters tested with cellular extract reactions.

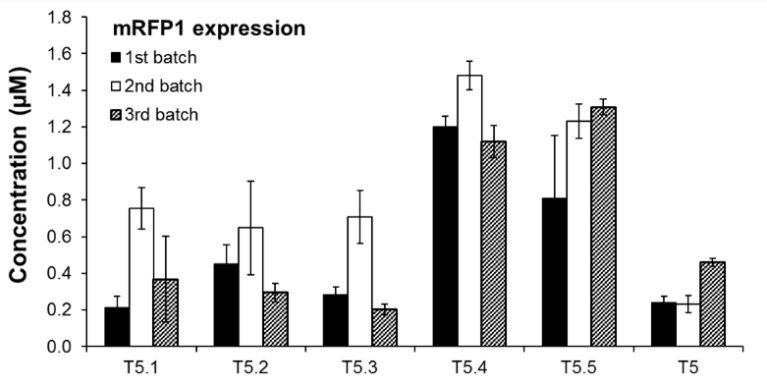


Figure 30. E. coli transcriptional promoters readily show many intermediated degrees of protein expression. Three independent experiments were performed in order to assess variability of gene expression in cellular extracts.

Therefore, it seems that employing the wild-type *E. coli* RNA polymerase is another successful strategy to control protein production in cell-free transcription-translation reactions. Obviously, further experiments will be required to elucidate if these promoters can be associated with different ribosome binding sites. Finally, it will be interesting to see if all these genetic elements can be incorporated into more complex genetic circuitry to improve their performance in cellular extract reactions. Finally, we repeated the experiments using our second set of *E. coli* transcriptional promoters for three different triplicates to assess the variability of protein synthesis in cellular extract reactions. The result was surprisingly similar to what previously observed using PURE system reactions (Figure 31). The "intraday" noise is more similar between the PURE system and the cellular extract reactions, while the "extraday" noise of the cellular extract reaction is higher than the one of the PURE system reaction. It has to be noted, however, that also the error of the coefficient of variation measurement is higher in cellular extract reactions, compared to PURE system reactions. The reason is that we observe a more significant sample-to-sample variability in the coefficient of variation, therefore increasing the error in the final coefficient of variation. Further investigation will be required to understand the many differences between PURE system and cellular extract reactions.

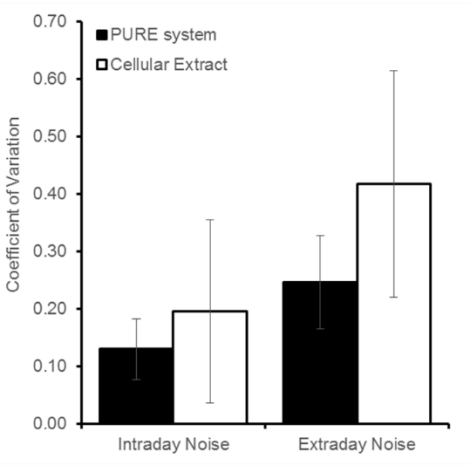


Figure 31. Coefficient of variation for gene expression is slightly higher in cellular extract reactions, compared to PURE system reactions.

3.3. Matherials and methods

3.3.1. Genetic Constructs

Genes encoding the proteins GFPmut3b (BBa_E0040), mRFP1 (BBa_E1010) and T3 RNA polymerase (BBa_K346000) were from the registry of standard biological parts (http://partsregistry.org). Genes encoding the proteins Azurite and EsaR were from Addgene (Plasmid #14034 and Plasmid #47646). Spinach surrounded by the tRNA scaffolding sequence was synthetized by Genscript following Jaffrey et al.[48] All genes were subcloned into pET21b by isothermal Gibson assembly.[64] All constructs were confirmed by sequencing by GATC Biotech. More details on the genetic constructs can be found in the following table.

3.3.2. Cell-Free Transcription–Translation (PURE system reaction)

Unless otherwise indicated, 9 µL transcription–translation reactions with the PURExpress in vitro protein synthesis kit (New England Biolabs) contained 12.6 nM of linear DNA template and 4 units of human placenta RNase inhibitor (New England Biolabs). When needed, DFHBI (Lucerna) was added to a final concentration of 60 µM. The reaction components were assembled in an ice-cold metal plate, and then the reaction was initiated by incubation at 37 °C. Reactions were monitored for 6 h with a Rotor-Gene Q 6plex system (Qiagen). Channel blue was used to detect Azurite (excitation, 365 ± 20 nm; emission, 460 ± 20 nm), channel green was

used to detect GFPmut3b and Spinach (excitation, 470 ± 10 nm; emission, 510 ± 5 nm), channel orange was used to detect mRFP1 (excitation, 585 ± 5 nm; emission, 610 ± 5 nm). Each reaction was repeated at least three times. The template DNA concentration was modified for the following experiments: the first two-gene operons experiment employed a circular DNA template at a concentration of 0.5 nM (Figure 1), while the following experiment employed either circular or linear DNA template at a concentration of 0.5 nM (Figure 2). Then, the cascade genetic circuit with the two genes on two separate pieces of DNA employed the following template concentrations: 12.6 nM (reporter gene encoding for mRFP1 and Spinach) and 4.2 nM (gene encoding for T3 RNA polymerase). The cascade genetic circuit with two genes on the same piece of DNA experiments employed a template DNA concentration of 5 nM. Finally, the experiments performed with the repressible genetic circuit employed a template DNA concentration of 5 nM. Moreover, in the "ON"-state of the repressible genetic circuit experiments, a final concentration of 5 μM of the molecule 3-oxohexanoyl-homoserine lactone.

3.3.3. Genetically encoded picture using PURE system reactions

The picture of the size of 14x14 pixels is generated using four different DNA templates. The following table was used to calculate the number of different 9 μL PURE system reactions needed to generate the picture.

sample	n. of pixels	pixel volume (μL)	final volume required	number of 10 μL reactions
Strong	58	2	116	13
Intermediate Strong	40	2	80	9
Intermediate Weak	40	2	80	9
Weak	58	2	116	13

Each reaction was assembled as previously described, ensuring that each sample had at least one of the solutions that compose the reaction (solution A or solution B) coming from a different batch compared to the other samples. After 6 h of incubation at 37 °C, each reaction was randomly pipetted into the wells (of a 1538 wells plate) representing the pixels of the

corresponding intensity. After filling all the required wells, the plate was centrifuged at 4000 rpm for 1 min at 4 °C with a Thermo Scientific Legend X1R centrifuge with a T20 microplate rotor. A Typhoon Trio from GE Healthcare was then used to visualize the picture. For the RNA picture, a blue laser was used (488 nm) in combination with the "526 SP" filter (short-pass filter transmitting light below 526 nm). For the protein picture, a green laser was used (532 nm) in combination with the "610 BP 30" filter (transmitting light between 595 nm and 625 nm). For both the pictures, the setting of the gain was 1000 V.

3.3.4. Cell-Free Transcription–Translation (cellular extract reactions)

Unless otherwise indicated, a standard *E. coli* S30 Extract System for Circular DNA (Promega) reaction of 50 μL is divided into five separate 10 μL reactions, each incubated with 3.5 nM of plasmid DNA. The plasmid DNA was isolated with a Wizard SV Mini Prep kit from Promega, and further purified by a phenol:chloroform extraction followed by an ethanol precipitation. When needed, DFHBI (Lucerna) was added to a final concentration of 60 μM. For the molecular beacon experiments, a 2-o-methyl RNA probe was synthetized from Eurofins with the following sequence: 5'-FAM-[CGCUUUUUUUUUUUGCG]-DAB-3', with each base having the modified 2-o-methyl ribonucleic backbone. The plasmid was modified accordingly to display the antiparallel sequence of the beacon in the 3'-UTR of the mRNA right after the end of the Spinach sequence. In the experiments the beacon was present at a final concentration of 4 μM. The reaction components were assembled in an ice-cold metal plate, and then the reaction was initiated by incubation at 37 °C. Reactions were monitored for 6 h with a Rotor-Gene Q 6plex system (Qiagen). Channel green was used to detect Spinach and the molecular beacon (excitation, 470 ± 10 nm; emission, 510 ± 5 nm), channel orange was used to detect mRFP1 (excitation, 585 ± 5 nm; emission, 610 ± 5 nm). Each reaction was repeated at least three times.

3.3.5. Proteins and RNA Standard Curves

Standard curves to translate fluorescence readouts into molar concentrations were created as follows. His-Tagged versions of GFPmut3b, Azurite, and mRFP1 were generated by mutating the stop codon of the constructs CD100A, FB009A, RL008A[57] by Phusion site-directed mutagenesis (Finnzymes). The resulting plasmids encoded GFPmut3b, Azurite, or mRFP1 with

24 additional residues including a carboxy-terminal hexahistidine-tag. Each His-tagged construct was used to transform *E. coli* BL21(DE3) pLysS (Promega) cells to be grown in LB supplemented with 50 μg/mL ampicillin and 100 μg/mL chloramphenicol at 37 °C to an optical density of 0.5 at 600 nm before induction with 0.4 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). The cells were harvested 4 h after the addition of IPTG by centrifugation at 5000 rpm for 10 min with a Beckman Coulter Avanti J-E centrifuge with a JLA 9100 rotor. Cell pellets were then resuspended in 40 mL buffer R (50 mM NaH2PO4, 300 mM NaCl, 5 mM imidazole, pH 8), supplemented with 100 μL protease inhibitor cocktail (Sigma), and sonicated on ice (4 cycles, 10 s each cycle with 1 min cooling on ice between cycles) with a Branson Sonifier 450. Lysates were centrifuged at 15 000 rpm for 30 min at 4 °C with a Thermo Scientific Legend X1R centrifuge with a Fiberlite F15-8 × 50 cy rotor. The cleared lysates were loaded on Ni-NTA columns (Qiagen) and successively washed with buffer R and buffer R supplemented with 20 mM imidazole. Bound proteins were eluted with buffer R plus 250 mM imidazole. Eluted proteins were dialyzed against 20 mM Tris-HCl, 150 mM NaCl, 5 mM 2-mercaptoethanol, pH 8. Protein concentrations were determined from the extinction coefficients of GFPmut3b ($\varepsilon^{280\ nm}$ = 21 890 M–1 cm–1)[65], Azurite ($\varepsilon^{383\ nm}$ = 26 200 M–1 cm–1)[66], and mRFP1 ($\varepsilon^{584\ nm}$ = 44 000 M–1 cm–1)[67] with an Agilent 8453 UV–vis.

Spinach mRNA was purified by using acidified phenol extraction followed by an ethanol precipitation. Transcription reactions were assembled with a 60 nM final concentration of the template DNA of the different constructs of which the mRNA we wanted to purify. The transcription reactions were performed with 150 units of T7 RNA polymerase (New England Biolabs), 5 mM of each nucleotide (New England Biolabs), 20 units of human placenta RNase Inhibitor (New England Biolabs) and 0.1 units of inorganic pyrophosphatase (New England Biolabs). The reactions were assembled on ice and then incubated overnight at 37 °C. The following day the transcription reactions were supplemented with 40 mM of ethylenediaminetetraacetic acid (EDTA) pH 8.0, 0.5 mM of CaCl2 and 2.5 units of DNase I (RNase-free) (New England Biolabs), and incubated at 37 °C for an additional 1.5 h. The samples were then extracted with 5:1 phenol:chloroform (Sigma). The upper aqueous phase was subsequently extracted with 24:1 chloroform:isoamyl alcohol (Sigma) and ethanol precipitated[68]. RNA samples were resuspended in 0.1 mM EDTA and mixed with 1 volume of

2× RNA Loading Dye (Thermo Scientific). Samples were loaded on a 1% agarose gel and compared against a lane containing RiboRuler High Range RNA Ladder (Thermo Scientific) for quantification with the software ImageJ[69]. Varying amounts of the resulting mRNAs were then incubated with 60 μM DFHBI to construct a standard curve that relates fluorescence to RNA concentration.

3.3.6. Statistical analysis

The coefficient of variation was calculated as the ratio between the standard deviation and the mean value of each experiment:

$$C_v = \frac{\sigma}{\mu}$$

For the "intraday" coefficient of variation, we calculated the coefficient of variation for each sample in each experiment, but without mixing together the same samples from different experiments, and then we would calculate the mean value and the standard deviation of all the coefficient of variations obtained in that manner. For the "extraday" coefficient of variation, we calculated the coefficient of variation of the different samples by pooling together the different values of the same samples from the different experiments, after which we would calculate the mean value and the standard deviation of all the coefficient of variations obtained in that manner.

Chapter 4: Using RNA molecules to coordinate proteins in cell-free transcription-translation reactions.

4.1. Introduction.

While being able to control protein expression in cell-free transcription-translation reactions is crucial for the creation of artificial cells, that is not the sole feature required in order to increase the complexity of artificial cells. For example, the catalytic efficiency of enzymes can be significantly lower *in vitro* compared to *in vivo* for a variety of reasons. One of the reasons for the lower catalytic activity, especially when reconstituting metabolic pathways that involve more than a single enzyme, is that the endogenous spatial organization of the enzymes is lost in the transition from *in vivo* to *in vitro* conditions. Moreover, several articles showed how placing in close proximity different enzymes that catalyze subsequent steps of a metabolic pathway increases the overall catalytic activity of the pathway[70,71]. While we were aware of the work done using protein scaffolds to join together different enzymes, we decided to use another kind of molecule: RNA. Several RNA aptamers available are able to interact specifically with different peptide domains. Moreover, it has been shown that if these peptide domains are included in the coding sequence of different enzymes, then the relative RNA aptamers can bind the enzymes through the peptide domains[70]. Finally, as the interaction between the RNA aptamer and the relative peptide domain relies on the correct folding of the RNA aptamer, it is possible to influence the binding between the RNA and the enzyme by interfering with the folding of the RNA aptamer. Therefore, opening up the possibility of influencing the catalytic rate of a process via interaction with an RNA molecule. To investigate if spatial proximity between proteins can be obtained by using an RNA molecule as the scaffold we decided to rely on the Förster resonance energy transfer (FRET). FRET is a mechanism by which an energy transfer can happen between two fluorophores, if the emission spectra of one of the two fluorophores partially overlaps the excitation spectra of the other fluorophore. A donor fluorophore in its electronic excited state may transfer energy to an acceptor fluorophore through nonradiative dipole–dipole coupling. As the efficiency of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor, FRET is

extremely sensitive to small changes in distance. Therefore, we chose to use this effect to understand if we could place two proteins in close proximity using an RNA molecule.

4.2. Results and discussion.

4.2.1. FRET is detectable when using the aptamers PP7 and Biv-TAT, but not PP7 and MS2.

The first step was to understand if we could place two proteins in close physical proximity in a cell-free transcription-translation reaction. Therefore, we designed an RNA molecule displaying two RNA aptamers, the MS2 aptamer and the PP7 aptamer, separated by a short hairpin. The two aptamers would then be able to recruit two fluorescent proteins modified to include the peptide sequences recognized by the two aptamers. The RNA molecule was designed in such a way that the hairpin would also include a loop: by opening the hairpin by base pairing the loop with another RNA molecule we would increase the distance between the two proteins, thus directly influencing the catalytic flux through the enzymes. Our idea was that before directly measuring the catalytic efficiency of a pair of enzymes, we would use another methodology to confirm the physical proximity of the two proteins, i.e. spatial proximity was probed by FRET between two fluorescence proteins. We included the peptide domain recognized by the aptamer on two different fluorescent proteins (mVenus and mCerulean). We decided to use the two fluorescent proteins mCerulean and mVenus because the emission spectra of mCerulean partially overlaps the excitation spectra of mVenus. Therefore, mCerulean functions as the donor and mVenus as the FRET acceptor.

Preliminary experiments, however, showed that the FRET signal could not be detected. It seemed that the main issue was the interaction between the MS2 aptamer and the relative peptide domain. Since it was not clear whether the problem was the folding of the MS2 aptamer, or if the problem was the interaction with the fluorescent protein, we decided to change two things. First, we substituted the MS2 aptamer with the Biv-TAT aptamer, consequently changing also the peptide domain in the relative fluorescent protein. Second, we completely modified the RNA molecule carrying the aptamers, and we decided to use the pRNA 3WJ as the structural scaffold. That is an RNA composed of three different interacting domains that can fold in a three way junction (a structure with three arms)[72]. This RNA is an extremely stable structure[72]. Moreover, each arm can be modified to contain different RNA

aptamers, and the excellent folding capability of the overall structure ensures the correct folding of the different RNA aptamers that are included[72]. We decided to include aptamer domains in each of the three arms of pRNA 3WJ. One arm contains the malachite green aptamer for detection of the correct folding of the pRNA[73]. Another arm houses a streptavidin aptamer. The third arm has two aptamer domains that are used to localize the fluorescent proteins fused to the aptamer ligands. These two aptamers are PP7[74] and the Biv-TAT[75].

First, in order to check if the system is working, we only used purified components. We purified the two fluorescent proteins modified to include the peptides recognized by the RNA aptamers, mVenus::Biv-TAT and mCerulean::PP7, and we purified the pRNA. Then, we mixed them together. In this first round of experiments, we recorded the fluorescence by using a fluorimeter with a slit aperture of one nm to reduce interference between the different fluorescence channels. This is particularly important for the FRET signal channel, which uses the excitation wavelength of the donor protein (mCerulean) for the excitation, and the emission wavelength of the acceptor protein (mVenus) for the emission. The result of the experiments are reported using the ratio of the FRET signal channel over the donor protein signal channel (mCerulean). This is because FRET results in an increase in the emission of the acceptor fluorophore and a decrease in the emission of the donor fluorophore. Adding the purified pRNA to an equimolar concentration of the two proteins unfortunately did not lead to any significant change FRET. Therefore, something in the system was not working properly (Figure 1).

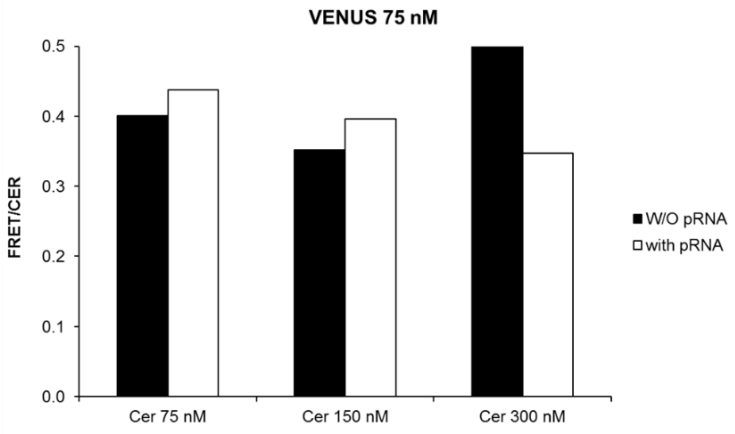


Figure 32. FRET is not observed when the pRNA is present. Three different concentrations of the donor protein have been tested, and none of them led to any detectable FRET between the two fluorescent proteins.

One potential problem was that the K_d of the two aptamers was different. Specifically, the K_d of the Biv-TAT aptamer was 60 nM[76], while the K_d of the PP7 aptamer was 1 nM[74]. However, given the concentration of the modified fluorescent proteins that we used, we should still have saturated all the binding sites in the aptamers, even if considering the higher K_d. Therefore, something must be present in our system that reduces the affinity between the aptamers and the peptide tags. However, we still tried to fix the concentration of the donor protein (mCerulean) and of the pRNA in the system and titrate the acceptor protein (mVenus). Using this approach, we identified the optimal concentration for the acceptor protein (mVenus), which was found to be 4 μM (Figure 2). Again, this is much higher than should be, and it is not clear why. After this, we then tried to change the concentration of the donor protein (mCerulean) after fixing the concentration of the other two components of the system, the pRNA and the acceptor protein (mVenus). Again, we used the data from the different fluorescence channels to identify the optimal concentration of the donor protein (Figure 3).

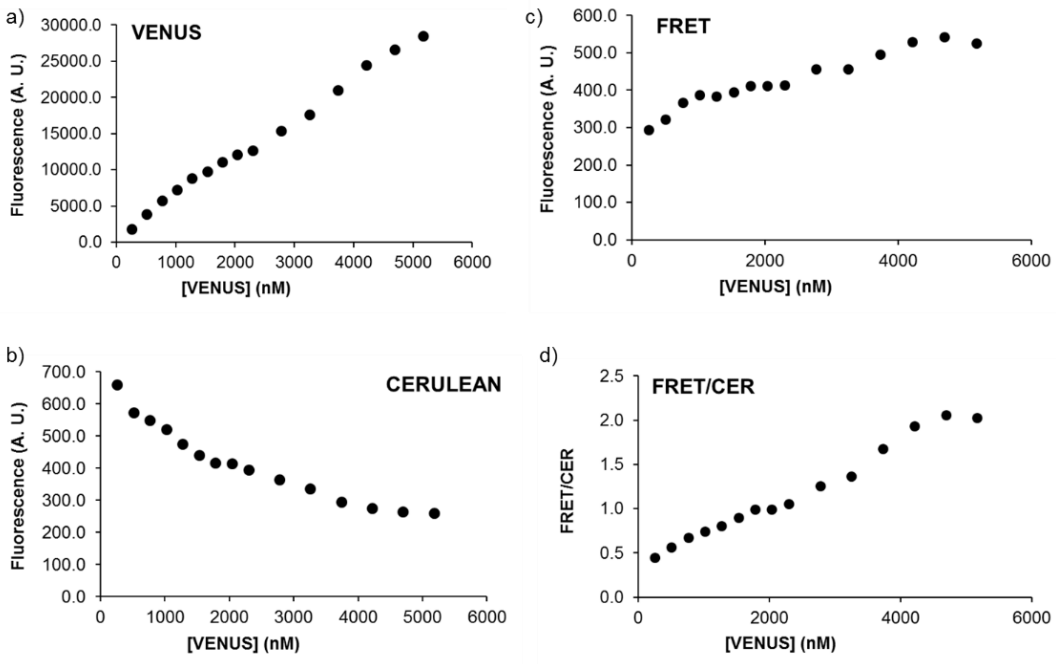


Figure 33. FRET requires a high concentration of the acceptor protein to be detected. (a) Direct relationship of the signal recorded from the fluorescent protein mVenus from the concentration of mVenus present in the system. (b) Decrease of the signal recorded from the fluorescent protein mCerulean, as the concentration of mVenus increases, possibly due to an increase in the FRET between the two fluorescent proteins. (c) Increase in the FRET signal is directly correlated with an increase in mVenus concentration. (d) Ratio of the FRET over the mCerulean signal shows a plateau around 4 μ M of mVenus.

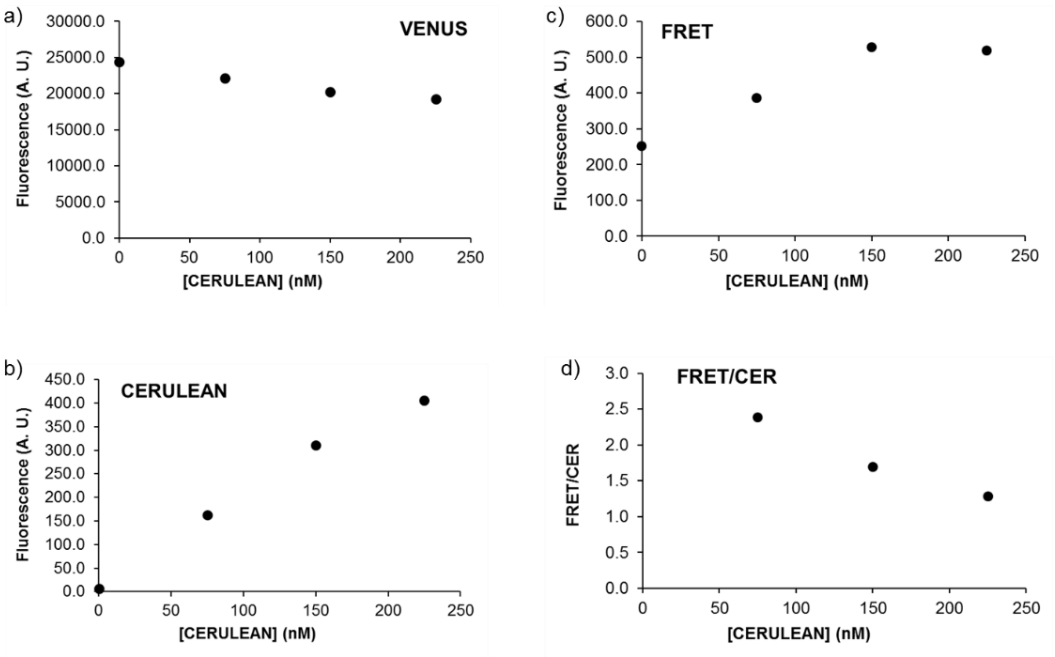


Figure 34. Increasing mCerulean concentration above 100 nM does not increase FRET signal. (a) mVenus signal is stable as mCerulean concentration increases. (b) Direct relationship of the signal recorded from the fluorescent protein mCerulean from the concentration of mCerulean present in the system. (c) FRET signal does not increase as mCerulean concentration increases. (d) Ratio of the FRET over the mCerulean signal decreases as the concentration of mCerulean increases.

Finally, after fixing the concentration of the two fluorescent proteins we could start the final experiment. Here, we started with the right concentration of the two proteins in the absence of the pRNA. Subsequently pRNA was added to the system. As can be seen in Figure 4, the ratio of the FRET channel fluorescence signal over the donor protein fluorescence signal showed a two-fold increase after the addiction of the pRNA to the system. To confirm that the change in the ratio was really due to the presence of the pRNA, after waiting for the fluorescence signals to stabilize, we added a mixture of RNase A and RNaseT1 to the system to degrade the pRNA. As expected, the ratio of the FRET channel fluorescence signal over the donor protein fluorescence signal goes back to the value it had before the addition of the pRNA to the system (Figure 4). This is a good indication that the pRNA construct can indeed recruit two proteins and that the two proteins are physically close together. The main drawback of this system is in the use of an aptamer (Biv-TAT) with a weak affinity for its ligand.

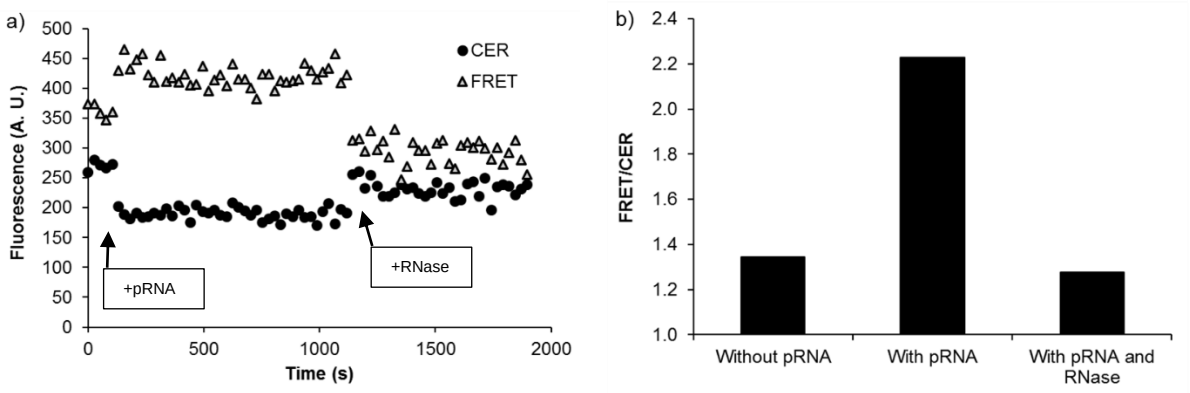


Figure 35. FRET is detected in the system and is dependent on the presence and integrity of the pRNA. (a) Fluorescent signal recorded from the FRET channel increases upon addition of the pRNA to the system, and then decreases upon the addition of RNases to the system. Conversely, signal recorded from the mCerulean channel decreases upon addition of the pRNA to the system, and increases upon addition of RNases to the system. (b) The two signals are reported together as the ratio between the FRET signal over the mCerulean signal.

4.2.2. FRET is not detectable when using the MS2 aptamer with the pRNA

Since the MS2 aptamer displays a significantly lower K_d (5 nM[74]) compared to the Biv-TAT aptamer, we tested a pRNA construct containing the MS2 aptamer. We decided to do so because when we tried the MS2 aptamer previously we were not using the pRNA. Therefore, we did not have any definitive indication that the problem lied in the MS2 aptamer. We modified the pRNA to remove the Biv-TAT aptamer and to include the MS2 aptamer, while at the same time modifying the acceptor protein (mVenus) to include the peptide tag recognized by the MS2 aptamer. Then, we tried different concentrations of the two fluorescent proteins to see if you could use the pRNA to detect FRET. Unfortunately, we could not find any condition at which the pRNA would elicit a difference in the ratio of the FRET channel fluorescence signal over the donor protein fluorescence signal (Figure 5). We therefore concluded that the MS2 aptamer was not working.

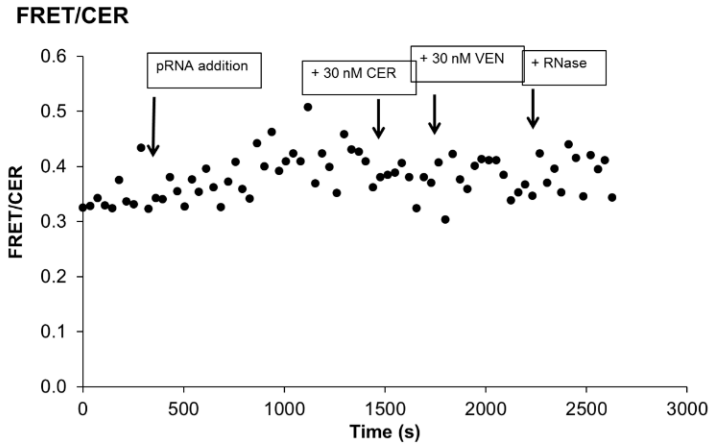


Figure 36. FRET is not detect when using the MS2 aptamer. Different conditions were tested, but none of them resulted in any appreciable difference in the ratio of the FRET over the mCerulean signal, indicating the absence of FRET between the two fluorescent proteins.

4.2.3. In vitro transcription can be used to produce the pRNA and detect FRET

Next, we tried to see if we could detect FRET by performing an *in vitro* transcription reaction in the presence of the two fluorescent proteins and of a template DNA with the pRNA under the control of a T7 transcriptional promoter. Interestingly, we could detect an increase in the usual ratio of the FRET over donor protein fluorescence signals, with a kinetic profile similar to *in vitro* transcription reactions from T7 transcriptional promoters recorded using the Spinach aptamer (Figure 6). However, after peaking, the ratio started to diminish, finally reaching a value similar to the one at the beginning of the reaction (Figure 6). Another experiment performed recording only the signal from the malachite green aptamer that is included in the pRNA indicated that the amount of template DNA was too high, probably leading to an excess of pRNA, thus negatively impacting FRET (Figure 7).

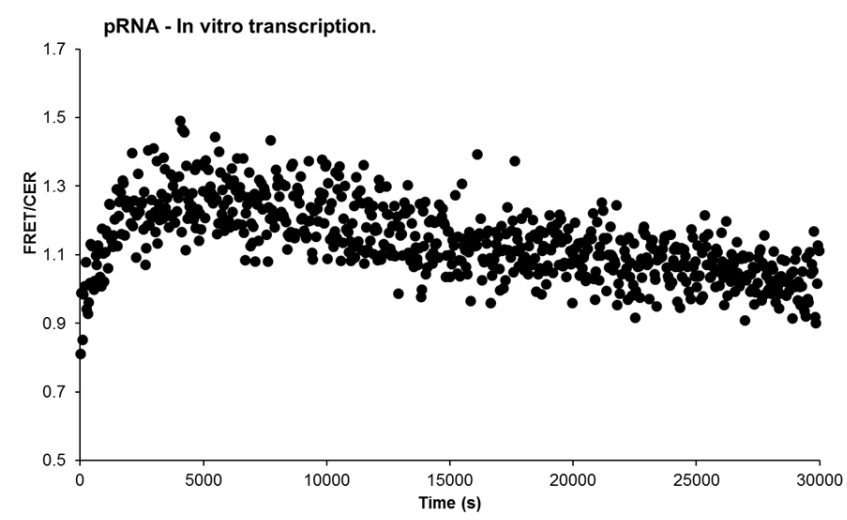


Figure 37. Transcribing the pRNA in the presence of the two fluorescent proteins results in a detectable FRET signal. The kinetic profile of the experiment resembles the transcription kinetic profiles obtained with the Spinach aptamer, particularly the first two hours. When the ratio start to decrease there might be too much pRNA in the system, thereby diminishing the FRET between the two fluorescent proteins.

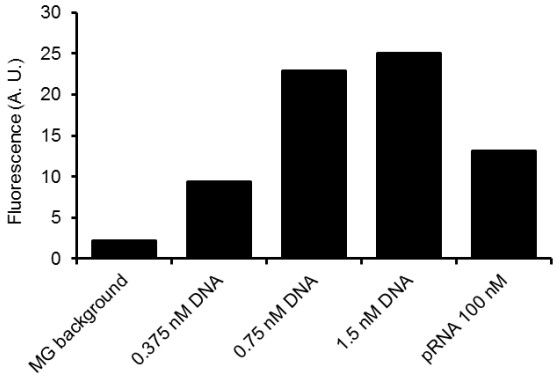


Figure 38. The amount of template DNA for the experiment reported in Figure 6 (4 nM) was probably leading to an excess of the pRNA in the system. Even 1.5 nM of template DNA in an in vitro transcription reactions lead to a fluorescent signal from the malachite green aptamer that is higher to the signal obtained from 100 nM of purified pRNA.

4.2.4. In vitro transcription-translation reactions can be used to synthesize both the pRNA and the fluorescent proteins to detect FRET

Finally, we tried to use PURE system reactions in order to make all the three components of our system, including the two fluorescent proteins and the pRNA. We would add to the PURE system reaction three different pieces of DNA, each carrying the information for one of the

components of the system. When the first experiment was performed, again we could detect an increase in the FRET over donor ratio followed by a decrease (Figure 8). Upon reducing the concentration of the template DNA for the pRNA the ratio would increase, slowly, but steadily and without the decrease that was observed in the first experiment (Figure 9). However, because we are also producing the fluorescent proteins within the reaction, the result is not as clear as the result obtained with the purified components. When we are using purified components the change in fluorescence can be safely assigned as a result to the presence of the pRNA. On the other hand, when fluorescent proteins are synthesized during the transcription-translation reaction, then an increase in fluorescence will occur regardless of the existence of FRET or not. Therefore, a clear assignment of the source of each signal is much more difficult to decipher.

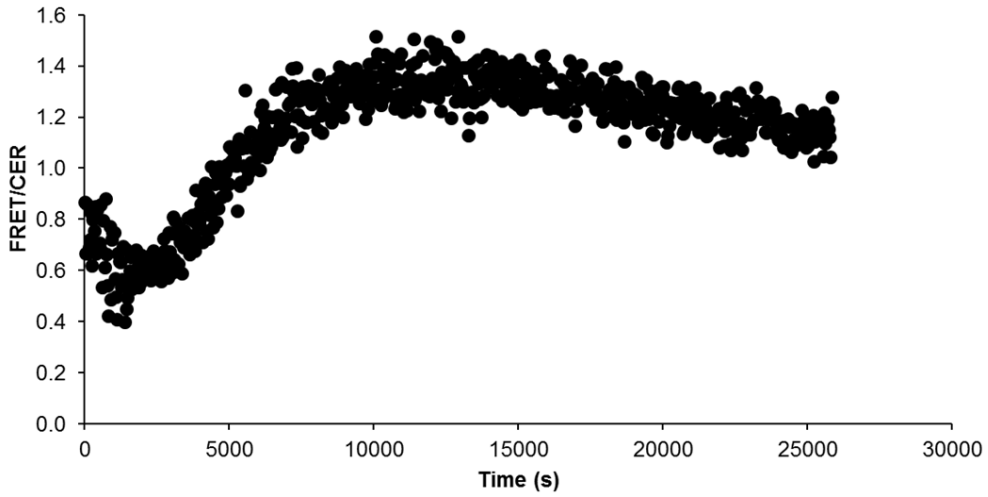


Figure 39. FRET is detectable when synthetizing the two fluorescent proteins and the pRNA in a PURE system reaction. The ratio of the FRET over the mCerulean signal decreases probably due to an excess of the template DNA encoding the pRNA.

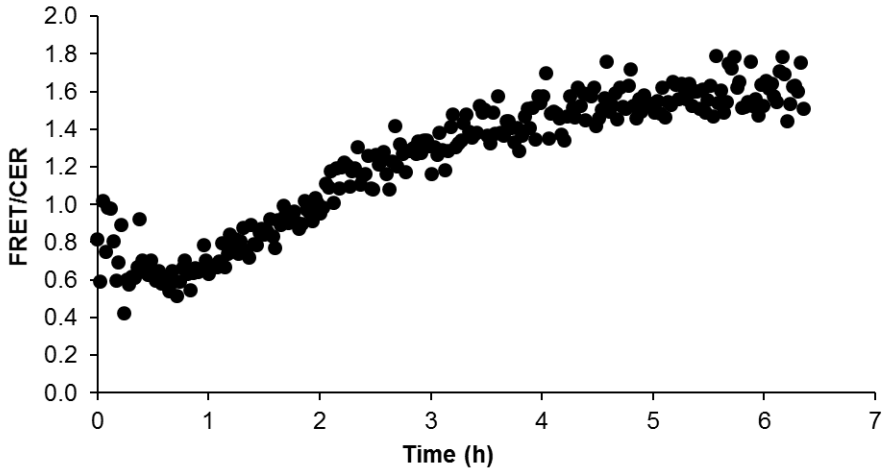


Figure 40. FRET is detectable when synthetizing the two fluorescent proteins and the pRNA in a PURE system reaction. The ratio of the FRET over the mCerulean signal slowly increases probably due to the low concentration of the template DNA encoding the pRNA employed.

4.3. Materials and Methods

4.3.1. Gene constructs

The Silver lab kindly provided the plasmids encoding the peptides recognized by the PP7 and the MS2 aptamers, particularly plasmids pCJDFA and pCJDFB[70]. Genes encoding the proteins mCerulean and mVenus were taken from plasmids RL005A and RL009A[57].

4.3.2. Proteins and pRNA purification

The fluorescent proteins were purified using a carboxy-terminal hexahistidine-tag. *E. coli* BL21(DE3) pLysS (Promega) were transformed with each His-tagged construct and grown in LB supplemented with 50 µg/mL ampicillin and 100 µg/mL chloramphenicol at 37 °C to an optical density of 0.5 at 600 nm before induction with 0.4 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). The cells were harvested 4 h after the induction by centrifugation at 5000 rpm for 10 min with a Beckman Coulter Avanti J-E centrifuge with a JLA 9100 rotor. A resuspension buffer (500 mM NaCl, 10 mM imidazole, 50 mM Tris HCl pH 8), supplemented with 200 uL protease inhibitor cocktail (Sigma), was used to resuspend the cell pellets. The cells where then sonicated on ice (4 cycles, 10 s each cycle with 1 min cooling on ice between cycles) with a Branson Sonifier 450. Lysed cells were centrifuged at 15 000 rpm

for 30 min at 4 °C with a Thermo Scientific Legend X1R centrifuge with a Fiberlite F15-8 × 50 cy rotor. The cleared lysate was loaded on a Ni-NTA column (Qiagen) and successively washed with the resuspension buffer. Bound protein was then eluted with resuspension buffer plus 250 mM imidazole. Eluted protein was dialyzed against 500 mM NaCl, 50 mM Tris- HCl pH 7.5. Protein concentrations were determined from the extinction coefficients of mVenus ($\varepsilon^{515\ nm}$ = 92 200 M–1 cm–1)[65] and mCerulean ($\varepsilon^{433\ nm}$ = 43 000 M–1 cm–1)[65], with an Agilent 8453 UV–vis.

To purify the pRNA a transcription reaction was assembled with a 60 nM final concentration of a linear template DNA composed of a T7 transcriptional promoter followed by the pRNA sequence. The transcription reaction was performed with 150 units of T7 RNA polymerase (New England Biolabs), 5 mM of each nucleotide (New England Biolabs), 20 units of human placenta RNase Inhibitor (New England Biolabs) and 0.1 units of inorganic pyrophosphatase (New England Biolabs). The reaction was assembled on ice and then incubated overnight at 37 °C. The following day the transcription reaction was supplemented with 40 mM of ethylenediaminetetraacetic acid (EDTA) pH 8.0, 0.5 mM of $CaCl_2$ and 2.5 units of DNase I (RNase-free) (New England Biolabs), and incubated at 37 °C for an additional 1.5 h. After the DNase treatment, the pRNA was purified using an RNA purification kit from E.Z.N.A. following manufacturer's protocol. The pRNA was then analyzed with an agarose gel for structural integrity and quantified with a NanoDrop 1000 spectrophotometer.

4.3.3. FRET detection with purified components

FRET detection reactions were assembled in a buffer with 200 mM TrisHCl pH 7.5, 1.4 M NaCl and 50 mM KCl. Unless otherwise indicated, when present, the pRNA was present at a final concentration of 100 nM. First, we tested an mVenus concentration of 75 nM with increasing concentrations of mCerulean (75 nM, 150 nM and 300 nM), with and without the presence of the pRNA. After adjusting the concentration of the fluorescent proteins as indicated in the "Results and Discussion" section the final concentration of the fluorescent proteins was for mVenus of 4 μM and for mCerulean of 120 nM. When required, 5 units of a mixture RNases A and RNase T1 (Thermo Scientific) were added to the reaction. Fluorescence was monitored at 37 °C with a Photon Technology International (PTI) QuantaMaster 40 UV–Vis

spectrofluorometer equipped with two detectors. The excitation and emission wavelengths were 434 and 475 nm, 515 and 528 nm, and 434 and 528 nm for mCerulean, mVenus, and FRET, respectively.

4.3.4. In vitro transcription reaction

A transcription reaction was assembled with a 4 nM final concentration of a linear template DNA composed of a T7 transcriptional promoter followed by the pRNA sequence. The transcription reaction was performed with 50 units of T7 RNA polymerase (New England Biolabs), 2 mM of each nucleotide (New England Biolabs), 20 units of human placenta RNase Inhibitor (New England Biolabs), 4 μM of mVenus and 120 nM of mCerulean. Fluorescence was monitored at 37 °C for 6 hours with a Photon Technology International (PTI) QuantaMaster 40 UV–Vis spectrofluorometer equipped with two detectors. The excitation and emission wavelengths were 434 and 475 nm, 515 and 528 nm, and 434 and 528 nm for mCerulean, mVenus, and FRET, respectively.

To monitor malachite green fluorescence signal arising from *in vitro* transcription reactions with different DNA template concentrations a Rotor-Gene Q 6plex system (Qiagen) was used, particularly the red channel (excitation, 625 ± 5 nm; emission, 660 ± 10 nm). Three different concentrations of DNA template were tested: 1.5 nM, 0.75 nM and 0.375 nM. The resulting malachite green signal was then compared to the signal from 100 nM of purified pRNA. The malachite green molecule was present at a final concentration of 10 μM in all the reactions.

4.3.5. In vitro transcription-translation reaction

Three different DNA templates were added to a PURE system reaction. The DNA template encoding the pRNA was present at a final concentration of either 1.5 nM (first experiment) or 0.375 nM (second experiment). The DNA template encoding mCerulean was present at a final concentration of either 2 nM (first experiment) or 1.75 nM (second experiment). The DNA template coding for mVenus was present at a final concentration of either 10 nM (first experiment) or 11 nM (second experiment). Fluorescence was then monitored at 37 °C for 6 h with a Photon Technology International (PTI) QuantaMaster 40 UV–Vis spectrofluorometer

equipped with two detectors. The excitation and emission wavelengths were 434 and 475 nm, 515 and 528 nm, and 434 and 528 nm for mCerulean, mVenus, and FRET, respectively.

Chapter 5: Discussion and conclusions.

In this thesis, we showed how it is possible to track both transcription and translation in real-time using fluorescence spectroscopy in PURE system reactions. To do so we employed the recently characterized Spinach aptamer and confirmed that this aptamer is functional in PURE system reactions, alongside a red fluorescent protein. Then we used this approach to characterize a series of T7 transcriptional promoter of different intensities. We identified the different T7 promoters mainly from the genome of the T7 bacteriophage, and we showed that using different T7 transcriptional promoters correctly results in a balanced distribution of transcriptional levels, meaning that T7 promoters are very good at controlling transcription. Moreover, the T7 transcriptional promoters retain the ability of precisely controlling transcription very consistently even between different genetic contexts, and are associated with a very limited variability. The variability is indeed inferior to the one associated with protein synthesis, probably because transcription requires the participation of fewer components.

Unfortunately, the same degree of control does not apply to translation: even if the transcriptional distribution arising from the different T7 promoters is balanced, the same is not true for the distribution of protein synthesis arising from the same promoters. Protein expression clustered against a "low" and a "high" value, therefore reducing the possibility of using different T7 transcriptional promoters to control precisely gene expression in PURE system reactions.

Our experiments seem to indicate that the main culprit behind the differences in the transcriptional and translational distributions is the T7 RNA polymerase. Because the T7 RNA polymerase is a very highly processive polymerase, employing T7 RNA polymerase for transcription results in a de-coupling between transcription and translation: the RNA is produced much faster than the ribosomes can translate it. On the other hand, in *E. coli* mRNAs are translated while they are still transcribed. Together the data indicate that the de-coupling due to the use of the T7 RNA polymerase for transcription is responsible for what we observe. First is that the ratio of protein to RNA is generally quite low (usually around 1), much lower than *in vivo*, meaning that the process is not optimal, and that de-coupling has a negative effect on translation. Moreover, it seems that because of the de-coupling, and because there are no

other mRNAs able to compete for the ribosomes, even a very modest amount of RNA is able to drive translation from the ribosomes, thereby leading quite quickly to the "high" state of protein expression. The final experiments of chapter 3, employing the cellular extract to characterize different *E. coli* promoters seem to confirm this behavior. In the extract, the use of different *E. coli* promoters results in a balanced distribution of protein expression levels.

The de-coupling could also help to explain the results of the experiments done using two and three-gene operons. We observed that the position of the gene within an operon influences gene expression more than the promoter, but this is true only for the second and the third position within the operon. As the gene moves further away from the 5' of the RNA, expression is greatly diminished. We speculated that this again is partly due to the de-coupling between transcription and translation. Because the RNA is not being actively translated while it is transcribed, it has time to fold into a complex three- dimensional structure. As the ribosome binding site that drives the expression of the first gene is located at the 5' end of the RNA molecule, it is going to be more accessible for the ribosomes than the ribosome binding sites driving the expression of the second and the third gene, as they will be buried within the structure. Metabolic load was also considered as one of the reasons of the unbalanced expression of the different genes within the operon. However, we were not able to find any indicative evidence of such an effect. By reducing the concentration of the DNA template we would reduce the expression of all the genes within the operon, therefore it seems that the accessibility of the ribosome binding site due to the folding of the RNA is the driving factor behind the imbalance of expression.

In order to see if we could remove the problem of the RNA structure we tried to apply the recently described Bicistronic design. Briefly, the presence of an additional ribosome binding site placed upstream of the desired ribosome binding site removes RNA structures, therefore allowing the ribosomes to start translation. This system was effective *in vivo* in *E. coli*[53], but unfortunately did not work in PURE system reactions.

Next, we decided to characterize a series of *E. coli* ribosome binding sites using single-gene constructs in order to understand if this genetic element would perform better with respect to the promoters. While it is true that we observed a better distribution of protein expression when

using ribosome binding sites instead of promoters, at the same time we also reported an increase in the variability associated with the expression. We also observed that the ribosome binding sites seem to be more dependent on the genetic context, especially of the 5'-UTR, with respect to the promoters.

Then, we moved on to combine our sets of T7 promoters and *E. coli* ribosome binding sites, again using single-gene constructs, to see if we could increase further our control of protein expression in PURE system reactions. We observe again that the data from the transcribed RNA is very consistent with the behavior of promoters as characterized before, highlighting again the reliability and the stability of the T7 promoters for transcription. We also report a decent increase in the distribution of protein expression. Moreover, we confirm the difference in variability between transcription and translation, again with the transcription being less variable than the translation. After training a computational model with the data sets acquired using a green fluorescent protein and a red fluorescent protein, we test the predictive power of the model by generating a series of prediction for a third protein, a blue fluorescent protein. The experimental values did not match the predictions. It seems that the employed blue fluorescent (Azurite) is difficult to express in PURE system reactions. Not only the experimental values did not match the predictions, the associated variability was also significantly higher, compared to the variability associated with the expression of the red fluorescent protein mRFP1. Because we suspected a role of the RNA structure in this behavior, we computationally analyzed the coding sequence of the fluorescent protein Azurite, looking for internal ribosome binding sites. Upon finding a significant difference in the predicted number and relative strength of the internal ribosome binding sites in the coding sequence of Azurite, compared to the ones found in the coding sequence of the red fluorescent protein mRFP1, we decided to modify Azurite coding sequence so as to exclude the internal ribosome binding sites. However, expression of the fluorescent protein did not increase, while at the same time the variability of expression did not decrease. Therefore, the internal ribosome binding sites were not responsible for the observed behavior.

Next, we moved on to investigate the performance of a simple genetic circuit in PURE system reactions: a cascade circuit, in which the canonical T7 RNA polymerase drives the expression of another viral polymerase, T3 RNA polymerase, which in turn leads to the expression of the

Spinach aptamer and a red fluorescent protein. First, we compared the performance of the circuit using two different genetic architectures, with the two genes either placed on the same piece of DNA, or with two genes separated on two different DNA fragments. Interestingly, having both genes on the same template DNA increased the performance of the circuit, while at the same time lowering the variability of expression. Therefore, it seems that, as the number of the genes to be expressed in PURE system reactions increases, it is important to keep all the genes on the same template DNA. With the cascade circuit we also showed how important it is to test genetic parts of different intensities, because of all our tested combinations, only one resulted in the highest expression of the reporter gene. Moreover, this one sample did not have the strongest parts to drive the expression of the T3 RNA polymerase, meaning that the approach of over-expressing all the genes in a genetic circuit in PURE system reactions decreases the performance of the genetic circuit. Finally, we measured the variability associated with the expression of the reporter gene, and we compared the data with the variability measured from single-gene expression. As expected, variability of expression from the cascade circuit is higher than variability of expression from single-gene, probably because of the additional step of expressing T3 RNA polymerase.

We then tested another simple genetic circuit, in which the repressor EsaR controls the expression of the reporter gene, composed as usual by the Spinach aptamer and the red fluorescent protein mRFP1, by interacting with the EsaR operator placed downstream of the promoter of the reporter gene. While the performance of the circuit is good, when considering only transcription data, there are some issues, when considering the translation data. The main problem is that the "OFF" state, in which there is supposed to be no expression of the reporter gene, shows a significant degree of expression nonetheless. While for most of the samples the expression is lower than the corresponding "ON" state, unfortunately having a residual expression from the "OFF" state will probably prevent the genetic circuit from working correctly. Again, we suspect that this problem stems once more from the use of the T7 RNA polymerase from transcription. Before the repressor is able to be expressed and bind the operator, the T7 RNA polymerase drives some transcription from the reporter gene, and even if the amount of transcribed RNA is quite low, it is still enough to drive translation from the ribosomes. This could be connected to the issue that we have observed previously, in which

even when using different T7 transcriptional promoters the expression was clustering against an "high" and a "low" value.

Finally, we tried to use different transcriptional promoters in a completely different system, an *E. coli* cellular extract. Because in this system it is possible to use the endogenous *E. coli* RNA polymerase with σ_{70} promoters, our goal was to compare differences in protein expression with the one obtained using T7 promoters in PURE system reactions. Even though we were not able to monitor transcription with the cellular extract, data from translation clearly showed a very well balanced distribution in protein expression. Therefore, this is another indication that the main issue of controlling gene expression in PURE system reactions is the use of the T7 RNA polymerase for transcription. It will be interesting to thoroughly test a modified PURE system, in which instead of T7 RNA polymerase for transcription, the endogenous *E. coli* RNA polymerase is used.

Appendix: genetic constructs sequences.

Construct name	Sequence (insert only)
RL055A (two-gene operon with mRFP1 and GFPmut3b)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGGCTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTTT CAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGT CCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGG ACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGA CTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGT GTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTG GTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGA ACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGT GGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTT ACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGA ACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAGCGGATCCGAATTCAATTAGTTTGAACTTATAA GGAGAATAATCTATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAG ATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAA ACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACT TTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGA GTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGAC ACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTT AAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCA TGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAG CGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGAC AACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCC TTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAACTCGAGCA CCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCC ACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGA AAGGAGGAACT
FC013A (mRFP1_Spinach reporter)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGGCTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTTT CAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGT CCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGG ACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGA CTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGT GTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTG GTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGA ACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGT GGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTT ACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGA ACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG GACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGGTGTGGCTGCTTCGGCAGTGCAGCTTGTTGA GTAGAGTGTGAGCTCCGTAACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGC GCCATAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTCGAGCACCACCACC ACCACCACTGAGATCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAA TAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACT
CD100A (GFPmu3b reporter)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGGCTAGCATGCGTAAAGGAGAAGAACTTTTCACTGGAGT TGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGT GAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTC CATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGATCATAT GAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTC AAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAA TCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTA TAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATT AGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCG ATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAA CGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGAT

	GAACTATACAAATAAGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACC ACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCAC CGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAA GGAGGAACT
DC051A (Azurite_Spinach reporter)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGGCTAGCATGTCTAAAGGTGAAGAATTATTCACTGGTGT TGTCCCAATTTTGGTTGAATTAGATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGT GAAGGTGATGCTACGTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTC CATGGCCAACCTTAGTAACTACTTTGAGCCATGGTGTTCAATGTTTTTCTAGATACCCAGATCATAT GAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTC AAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAA TCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGAATACAACTT CAACTCTCACAATATATACATCATGGCTGACAAACAAAAGAATGGTATCAAAGTGAACTTCAAAATT AGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATTGGTG ATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACCCAATCAGCCTTATCCAAAGATCCAAA CGAAAAGAGAGACCACATGGTCCTGTTAGAATTTAGGACTGCTGCTGGTATTACCCATGGTATGGAT GAATTGTACAAATAAGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGGACGCAACTGAATGAAATG GTGAAGGACGGGTCCAGGTGTGGCTGCTTCGGCAGTGCAGCTTGTTGAGTAGAGTGTGAGCTCCGTA ACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCATAGCATAACCCCTTG GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTCGAGCACCACCACCACCACCACTGAGATCTGCT AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTG GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACT
FC046A (Azurite iRBS removed)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGGCTAGCATGTCTAAAGGTGAAGAATTATTCACTGGTGT TGTCCCAATTTTGGTTGAATTAGATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGT GAAGGTGATGCTACGTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTC CATGGCCAACCTTAGTAACTACTTTGAGCCATGGTGTTCAATGTTTTAGCAGATACCCAGATCATAT GAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTC AAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAA TCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGCCACAAATTGGAGTACAACTT CAACTCTCACAATATATACATTATGGCTGACAAACAGAAGAATGGCATTAAAGTGAACTTCAAAATC AGACACAACATCGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATTGGTG ATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACCCAATCAGCCTTATCCAAAGATCCAAA CGAAAAGAGAGACCACATGGTCCTGTTAGAATTTAGGACTGCTGCTGGTATTACCCACGGTATGGAT GAATTGTACAAATAAGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGGACGCAACTGAATGAAATG GTGAAGGACGGGTCCAGGTGTGGCTGCTTCGGCAGTGCAGCTTGTTGAGTAGAGTGTGAGCTCCGTA ACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCATAGCATAACCCCTTG GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTCGAGCACCACCACCACCACCACTGAGATCTGCT AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTG GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACT
Cascade circuit (single construct)	CCGGTAATTAACCCTCACTAAAGGGAGATTGTGAGCGGATAACCAATTCCCCTCTAGAAATAATTTT GTTTAACTTTAAGAAGGAGATATACATATGGCTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTT TCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCG TCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGG GACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGG ACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGG TGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGT GGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCG AACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGG TGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCT TACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACG AACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCC GGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGGTGTGGCTGCTTCGGCAGTGCAGCTTGTTG AGTAGAGTGTGAGCTCCGTAACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGG CGCCACCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGT TTGTCGGTGAACGCTCTATCTCTCCGAGGAGGGAGTGCTGCAGTAATACGACTCACTATAGCCTGTA CTATAGTGCAGGTGGGAGAATAATCATATTAGAATGCTTTAAGAAGGAGATATACATATGATGAACA TCATCGAAAACATCGAAAAGAATGACTTCTCAGAAATCGAACTGGCTGCTATCCCGTTCAACACACT GGCTGACCACTACGGAAGCGCCTTGGCTAAAGAGCAGTTGGCTTTAGAACATGAGTCTTATGAGCTA GGCGAGCGCCGCTTCCTCAAGATGCTTGAGCGTCAAGCGAAAGCTGGTGAGATTGCAGACAACGCAG CCGCTAAGCCGTTACTCGCTACGCTTCTCCCTAAGTTAACCACACGTATCGTCGAGTGGCTCGAAGA GTACGCATCGAAGAAAGGCCGCAAGCCTAGCGCATACGCACCGCTCCAGTTACTCAAGCCGGAGGCC

	TCCGCGTTTATCACCCTGAAAGTTATCCTTGCGTCACTAACCAGTACGAACATGACAACCATTCAGG CCGCTGCTGGTATGCTGGGGAAAGCCATTGAGGACGAGGCACGATTTGGGCGCATCCGTGACCTAGA AGCGAAGCACTTCAAGAAGCACGTTGAGGAACAGCTTAACAAGCGCCACGGGCAAGTCTACAAGAAA GCATTTATGCAGGTGGTCGAGGCCGATATGATTGGTCGAGGTCTGCTTGGTGGCGAGGCGTGGTCTA GCTGGGATAAAGAAACCACGATGCACGTAGGGATTCGCCTGATTGAAATGCTGATTGAATCCACGGG TCTGGTGGAATTACAGCGCCACAACGCAGGTAACGCAGGCTCTGACCATGAGGCACTGCAACTGGCC CAAGAGTACGTGGACGTATTAGCGAAGCGTGCAGGCGCTCTGGCGGGTATCTCTCCGATGTTCCAGC CGTGTGTCGTACCGCCGAAACCTTGGGTAGCAATCACAGGGGGCGGCTATTGGGCTAACGGTCGCAG ACCTTTGGCACTCGTTCGCACTCACTCTAAGAAGGGCTTGATGCGCTACGAAGACGTTTACATGCCA GAAGTCTACAAGGCTGTGAACCTCGCGCAAAACACCGCATGGAAAATCAACAAGAAAGTTCTTGCTG TTGTCAATGAGATTGTTAACTGGAAGAATTGCCCGGTAGCAGACATTCCATCGCTGGAGCGCCAAGA GTTACCGCCTAAGCCTGACGACATTGACACCAACGAGGCAGCGCTCAAGGAGTGGAAGAAAGCCGCT GCTGGTATCTATCGCTTGGACAAGGCACGAGTGTCTCGCCGTATCAGCTTAGAGTTCATGCTGGAGC AGGCCAACAAGTTCGCAAGTAAGAAAGCAATCTGGTTCCCTTACAACATGGACTGGCGCGGTCGTGT GTACGCTGTGCCGATGTTCAACCCGCAAGGCAACGACATGACGAAAGGTCTGCTGACCCTTGCTAAA GGCAAGCCAATCGGTGAGGAAGGTTTCTACTGGCTGAAAATCCACGGTGCGAACTGTGCGGGTGTTG ATAAGGTTCCATTCCCGGAGCGCATCGCGTTCATTGAGAAGCACGTAGACGACATTCTGGCTTGCGC TAAAGACCCAATCAATAACACTTGGTGGGCTGAGCAGGATTCACCGTTCTGTTTCCTCGCGTTTTGC TTCGAGTATGCAGGCGTTACGCACCACGGTCTGAGCTACAATTGCTCTCTGCCGCTGGCGTTCGACG GGTCTTGCTCTGGTATCCAGCACTTCTCCGCGATGCTCCGCGATGAGGTAGGCGGTCGTGCGGTTAA CCTGCTGCCAAGCGAAACCGTGCAGGACATTTACGGCATCGTTGCACAGAAAGTAAACGAGATTCTC AAACAGGATGCAATCAACGGCACGCCTAACGAGATGATTACCGTGACCGACAAGGACACCGGGGAAA TCTCAGAGAAGCTCAAACTTGGAACCTCAACGCTGGCGCAACAGTGGCTGGCATATGGTGTAACCCG TAGCGTAACTAAACGTTCGGTCATGACGCTGGCTTACGGTTCCAAGGAGTTCGGCTTTCGTCAACAG GTATTGGATGACACCATTCAGCCTGCAATTGACAGCGGTAAGGGCTTGATGTTCACCCAACCGAACC AAGCGGCTGGCTATATGGCTAAGCTGATTTGGGATGCGGTAAGCGTGACCGTAGTTGCAGCGGTTGA GGCGATGAACTGGCTCAAATCTGCCGCTAAGCTGCTGGCTGCTGAGGTCAAGGACAAGAAGACCAAG GAGATTCTGCGCCACCGTTGCGCGGTTCACTGGACTACGCCGGACGGCTTCCCGGTCTGGCAGGAAT ACCGCAAGCCACTCCAGAAGCGTCTCGATATGATTTTCTTAGGGCAATTCCGTCTGCAACCGACGAT TAATACCCTCAAGGATTCAGGCATTGACGCACACAAGCAGGAGTCTGGCATCGCTCCTAACTTTGTT CACTCACAGGACGGTAGCCACCTCCGCATGACAGTCGTTTATGCTCACGAGAAGTATGGCATTGAGT CCTTTGCGCTCATCCATGACAGCTTTGGGACTATCCCGGCAGACGCTGGTAAGCTCTTTAAGGCTGT GCGTGAAACGATGGTTATCACCTATGAGAACAACGATGTGCTGGCAGACTTCTACTCTCAGTTTGCC GACCAGCTACACGAGACCCAACTGGACAAGATGCCTCCGCTTCCGAAGAAAGGAAACCTGAACCTGC AAGACATTCTCAAGTCTGACTTTGCCTTTGCATAAACTGAGATCCGGCTGCTAACAAAGCCCGAAAG GAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG TCTTGAGGGGTTTTTTG
Repressor circuit	TAATACGACTCACTATAGCCTGTACTATAGTGCAGGTGGGAGATTGTGAGCGGATAACCAATTCCCC TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTTCCTCCGAAGACGTTATCA AAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGG TGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCG CTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACC CGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAA CTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTAC AAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTT GGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCT GAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTT CAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCA TCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAGCCCGGATAGCTCAGTC GGTAGAGCAGCGGCCGGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGGTGTGGCTGCTTCGG CAGTGCAGCTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTCGCGTCCGGCCGCGGGTCCAGGGTTC AAGTCCCTGTTCGGGCGCCACCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTT CGTTTTATCTGTTGTTTGTCGGTGAACGCTCTATCTCTCCGAGGAGGGAGTGCTGCAGATCTCTCCG AGGAGGGAGTGCCGGTTAATACGACTCACTATAGGGAGAATAATCATATTAGAATGCTTTAAGAAGG AGATATACATATGTTCTCTTTCTTCCTTGAAAACCAAACAATAACGGATACGCTTCAGACTTACATA CAGAGAAAGTTATCTCCGCTGGGTAGTCCGGATTACGCTTACACTGTTGTGAGCAAAAAAAATCCTT CAAATGTTCTGATTATTTCCAGTTATCCTGACGAATGGATTAGGTTATACCGCGCTAACAACTTTCA GCTGACCGATCCCGTTATTCTCACGGCCTTTAAACGCACCTCGCCGTTTGCCTGGGATGAGAATATT ACGCTGATGTCCGGCCTGCGGTTCACCAAAATTTTCTCTTTATCCAAGCAATACAACATCGTTAACG GCTTTACCTATGTCCTGCATGACCACATGAACAACCTTGCTCTGTTGTCCGTGATCATTAAAGGCAA CGATCAGACTGCGCTGGAGCAACGCCTTGCTGCCGAACAGGGCACGATGCAGATGCTGCTGATTGAT TTTAACGAGCAGATGTACCGACTGGCAGGCACCGAAGGCGAGCGAGCCCCGGCGTTAAATCAGAGCG

	CGGACAAAACGATATTTTCCTCGCGTGAAAATGAGGTGTTGTACTGGGCGAGTATGGGCAAAACCTA TGCTGAGATTGCCGCTATTACGGGCATTTCTGTGAGTACCGTGAAGTTTCACATCAAGAATGTGGTC GTGAAACTGGGCGTCAGTAACGCCCGACAGGCTATCAGACTGGGTGTAGAACTGGATCTTATCAGAC CGGCAGCGTCAGCAGCAAGGTAAACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTG GCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT TTTTG
FC044A (mVenus::MS2::His_tag)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGAGCAAAGGCGAAGAACTGTTCACGGGTGTGGTTCCGAT CCTGGTTGAACTGGATGGCGATGTGAACGGTCATAAATTTAGCGTGTCTGGTGAAGGCGAAGGTGAT GCGACCTACGGCAAACTGACGCTGAAACTGATTTGCACCACGGGTAAACTGCCGGTTCCGTGGCCGA CCCTGGTGACCACGCTGGGTTATGGTCTGATGTGTTTCGCACGTTACCCGGATCACATGAAACGCCA TGATTTCTTTAAATCTGCGATGCCGGAAGGCTATGTGCAGGAACGTACCATCTTTTTCAAAGATGAT GGTAACTACAAAACCCGCGCGGAAGTTAAATTTGAAGGCGATACGCTGGTGAACCGTATTGAACTGA AAGGTATCGATTTCAAAGAAGATGGCAATATTCTGGGTCACAAACTGGAATACAACTACAACAGTCA TAACGTGTACATTACCGCCGATAAACAGAAAAACGGTATCAAAGCAAACTTCAAAATCCGTCACAAC ATCGAAGATGGCGGTGTTCAGCTGGCCGATCATTACCAGCAGAACACCCCGATTGGCGATGGTCCGG TGCTGCTGCCGGATAATCATTATCTGAGTTACCAGAGCAAACTGTCTAAAGATCCGAATGAAAAACG CGATCACATGGTTCTGCTGGAATTTGTGACCGCGGCCGGCATTACGCATGGTATGGATGAACTGTAT AAAACTAGAGGAGGAGGAGGATCAGGAGGAGGAGGATCAACTAGAATGGCTTCTAACTTTACTCAGT TCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGGT CGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCAGAGC TCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTG GTGTAGAGCTTCCTGTAGCCGCATGGCGTTCGTACTTAAATATGGAACTAACCATTCCAATTTTCGC CACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATT CCCTCAGCAATCGCAGCAAACTCCGGCATCTACAAAGCGGATCCGAATTCGAGCTCCGTCGACAAGC TTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGA AGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTC TTGAGGGGTTTTTTGCTGAAAGGAGGAACT
FC040A (mCer::PP7::His_tag)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGGTGAGTAAAGGCGAAGAGCTGTTCACAGGGGTTGTTCC GATTCTGGTCGAACTGGACGGGGACGTTAATGGTCACAAATTCAGCGTTAGCGGTGAGGGCGAGGGT GATGCCACTTATGGTAAACTGACCCTGAAATTCATCTGTACCACCGGCAAACTGCCTGTTCCTTGGC CTACACTGGTTACAACACTGACTTGGGGTGTTCAATGTTTTGCTCGCTATCCGGATCACATGAAACA GCACGATTTCTTCAAAAGCGCCATGCCTGAAGGTTATGTCCAAGAGCGTACGATCTTCTTTAAAGAC GACGGCAACTATAAAACCCGTGCCGAGGTGAAATTCGAAGGTGATACCCTGGTAAACCGTATCGAAC TGAAAGGGATCGACTTCAAAGAGGACGGGAACATTCTGGGCCATAAACTGGAGTATAACGCCATCAG CGATAATGTGTATATTACCGCCGACAAACAGAAAAACGGGATCAAAGCCAACTTCAAAATCCGCCAC AACATCGAGGATGGTAGCGTTCAACTGGCCGATCACTATCAACAGAATACCCCGATTGGTGATGGTC CTGTTCTGCTGCCTGATAACCACTATCTGAGCACCCAGTCTAAACTGTCCAAAGACCCGAACGAGAA ACGTGATCACATGGTTCTGCTGGAGTTTGTTACCGCTGCCGGCATTACTCTGGGTATGGATGAACTG TATAAAACTAGAGGAGGAGGAGGATCAGGAGGAGGAGGATCAACTAGAATGTCCAAAACCATCGTTC TTTCGGTCGGCGAGGCTACTCGCACTCTGACTGAGATCCAGTCCACCGCAGACCGTCAGATCTTCGA AGAGAAGGTCGGGCCTCTGGTGGGTCGGCTGCGCCTCACGGCTTCGCTCCGTCAAAACGGAGCCAAG ACCGCGTATCGCGTCAACCTAAAACTGGATCAGGCGGACGTCGTTGATTCCGGACTTCCGAAAGTGC GCTACACTCAGGTATGGTCGCACGACGTGACAATCGTTGCGAATAGCACCGAGGCCTCGCGCAAATC GTTGTACGATTTGACCAAGTCCCTCGTCGCGACCTCGCAGGTCGAAGATCTTGTCGTCAACCTTGTG CCGCTGGGCCGTAAAAAAGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGC ACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGC CACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTG AAAGGAGGAACT
FC041A (mVenus::Biv-TAT::His_Tag)	GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG TTTAACTTTAAGAAGGAGATATACATATGAGCAAAGGCGAAGAACTGTTCACGGGTGTGGTTCCGAT CCTGGTTGAACTGGATGGCGATGTGAACGGTCATAAATTTAGCGTGTCTGGTGAAGGCGAAGGTGAT GCGACCTACGGCAAACTGACGCTGAAACTGATTTGCACCACGGGTAAACTGCCGGTTCCGTGGCCGA CCCTGGTGACCACGCTGGGTTATGGTCTGATGTGTTTCGCACGTTACCCGGATCACATGAAACGCCA TGATTTCTTTAAATCTGCGATGCCGGAAGGCTATGTGCAGGAACGTACCATCTTTTTCAAAGATGAT GGTAACTACAAAACCCGCGCGGAAGTTAAATTTGAAGGCGATACGCTGGTGAACCGTATTGAACTGA AAGGTATCGATTTCAAAGAAGATGGCAATATTCTGGGTCACAAACTGGAATACAACTACAACAGTCA TAACGTGTACATTACCGCCGATAAACAGAAAAACGGTATCAAAGCAAACTTCAAAATCCGTCACAAC ATCGAAGATGGCGGTGTTCAGCTGGCCGATCATTACCAGCAGAACACCCCGATTGGCGATGGTCCGG TGCTGCTGCCGGATAATCATTATCTGAGTTACCAGAGCAAACTGTCTAAAGATCCGAATGAAAAACG CGATCACATGGTTCTGCTGGAATTTGTGACCGCGGCCGGCATTACGCATGGTATGGATGAACTGTAT

	AAAACTAGAGGAGGAGGAGGATCAGGAGGAGGAGGATCAACTAGAATGTCCGGCCCGCGTCCTCGTG GTACCCGTGGCAAAGGTCGCCGTATTCGCCGTGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGC GGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCT GAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA GGGGTTTTTTGCTGAAAGGAGGAACT
FC042A (pRNA)	GAAATTAATACGACTCACTATAGGGAGAATGCGGCCGCCGACCAGAATCATGCAAGTGCGTAAGATA GTCGCGGGTCGGCGGCCGCATAAAAATTGTCATGTGTATGTTGGGCGCAGGACTCGGCTCGTGTAGC TCATTAGCTCCGAGCCGAGTCCTCGAATACGAGCTGGGCACAGAAGATATGGCTTCGTGCCCAGGAA GTGTTCGCACTTCTCTCGTATTCGATTGCGCCCACATACTTTGTTGAGGATCCCGACTGGCGAGAGC CAGGTAACGAATGGATCCTCAATCATGGCAACTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG CTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTT GCTGAAAGGAGGAACT

References.

1. Nirenberg, M. Historical review: Deciphering the genetic code--a personal account. *Trends Biochem. Sci.* **29,** 46–54 (2004).
2. KORNER, A. The effect of hypophysectomy of the rat and of treatment with growth hormone on the incorporation of amino acids into liver proteins in a cell-free system. *Biochem. J.* **73,** 61–71 (1959).
3. ERDŐS, T. & ULLMANN, A. Effect of Streptomycin on the Incorporation of Amino-Acids labeled with Carbon-14 into Ribonucleic Acid and Protein in a Cell-free System of a Mycobacterium. *Nature* **183,** 618–619 (1959).
4. TIETZ, A. Fat synthesis in cell-free preparations of the locust fat-body. *J. Lipid Res.* **2,** 182–7 (1961).
5. RACKER, E. Synthesis of Carbohydrates from Carbon Dioxide and Hydrogen in a Cell-Free System. *Nature* **175,** 249–251 (1955).
6. Blow, J. J. & Laskey, R. A. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. *Cell* **47,** 577–87 (1986).
7. Smith, R. V & Evans, M. C. Nitrogenase activity in cell-free extracts of the blue-green alga, Anabaena cylindrica. *J. Bacteriol.* **105,** 913–7 (1971).
8. Zubay, G. In vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.* **7,** 267–87 (1973).
9. Krieg, P. A. & Melton, D. A. *Recombinant DNA Part F. Methods in Enzymology* **155,** (Elsevier, 1987).
10. Jewett, M. C. & Swartz, J. R. Substrate replenishment extends protein synthesis with an in vitro translation system designed to mimic the cytoplasm. *Biotechnol. Bioeng.* **87,** 465–72 (2004).
11. Jewett, M. C. & Swartz, J. R. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* **86,** 19–26 (2004).
12. Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* **99,** 162–8 (2014).
13. Michel-Reydellet, N., Calhoun, K. & Swartz, J. Amino acid stabilization for cell-free protein synthesis by modification of the Escherichia coli genome. *Metab. Eng.* **6,** 197–203 (2004).
14. Ahn, J.-H. *et al.* Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochem. Biophys. Res. Commun.* **338,** 1346–52 (2005).
15. Murray, C. J. & Baliga, R. Cell-free translation of peptides and proteins: from high throughput screening to clinical production. *Curr. Opin. Chem. Biol.* **17,** 420–6 (2013).
16. Zawada, J. F. *et al.* Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnol. Bioeng.* **108,** 1570–8 (2011).
17. Sun, Z. Z. *et al.* Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* e50762 (2013). doi:10.3791/50762

18. Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. & Alakhov, Y. B. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242,** 1162–4 (1988).

19. Pardee, K. *et al.* Paper-Based Synthetic Gene Networks. *Cell* **159,** 940–954 (2014).

20. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19,** 751–5 (2001).

21. Wang, H. H. *et al.* Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. *ACS Synth. Biol.* **1,** 43–52 (2012).

22. Zhou, Y., Asahara, H., Gaucher, E. A. & Chong, S. Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. *Nucleic Acids Res.* **40,** 7932–45 (2012).

23. Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451,** 704–7 (2008).

24. Kuruma, Y. & Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* **10,** 1328–44 (2015).

25. Awai, T., Ichihashi, N. & Yomo, T. Activities of 20 aminoacyl-tRNA synthetases expressed in a reconstituted translation system in Escherichia coli. *Biochem. Biophys. Reports* BBREPD1500219 (2015). doi:10.1016/j.bbrep.2015.08.006

26. Fujiwara, K., Katayama, T. & Nomura, S.-I. M. Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system. *Nucleic Acids Res.* **41,** 7176–83 (2013).

27. Hong, S. H., Kwon, Y.-C. & Jewett, M. C. Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. *Front. Chem.* **2,** 34 (2014).

28. White, E. R., Reed, T. M., Ma, Z. & Hartman, M. C. T. Replacing amino acids in translation: expanding chemical diversity with non-natural variants. *Methods* **60,** 70–4 (2013).

29. Makarova, O. V, Makarov, E. M., Sousat, R. U. I. & Dreyfus, M. Transcribing of Escherichia coli genes with mutant T7 RNA polymerases : Stability of lacZ mRNA inversely correlates with polymerase speed Microbiology : **92,** 12250–12254 (1995).

30. De Maddalena, L. L. *et al. GreA and GreB enhance Escherichia coli RNA polymerase transcription rate in a reconstituted transcription-translation system*. *bioRxiv* (Cold Spring Harbor Labs Journals, 2015). doi:10.1101/024604

31. Jackson, K., Kanamori, T., Ueda, T. & Fan, Z. H. Protein synthesis yield increased 72 times in the cell-free PURE system. *Integr. Biol. (Camb).* **6,** 781–8 (2014).

32. Ota, S., Yoshizawa, S. & Takeuchi, S. Microfluidic formation of monodisperse, cell-sized, and unilamellar vesicles. *Angew. Chem. Int. Ed. Engl.* **48,** 6533–7 (2009).

33. Martini, L. & Mansy, S. S. Cell-like systems with riboswitch controlled gene expression. *Chem. Commun. (Camb).* **47,** 10734–6 (2011).

34. Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H. & Libchaber, A. Toward an artificial cell based on gene expression in vesicles. *Phys. Biol.* **2,** P1–8 (2005).

35. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12672–7 (2003).

36. Karig, D. K., Iyer, S., Simpson, M. L. & Doktycz, M. J. Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res.* **40,** 3763–74 (2012).

37. Shin, J. & Noireaux, V. An E. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* **1,** 29–41 (2012).

38. Lentini, R. *et al.* Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour. *Nat. Commun.* **5,** 4012 (2014).

39. Martos, A., Jiménez, M., Rivas, G. & Schwille, P. Towards a bottom-up reconstitution of bacterial cell division. *Trends Cell Biol.* **22,** 634–643 (2012).

40. Osawa, M., Anderson, D. E. & Erickson, H. P. Reconstitution of contractile FtsZ rings in liposomes. *Science* **320,** 792–4 (2008).

41. Zieske, K. & Schwille, P. Reconstitution of self-organizing protein gradients as spatial cues in cell-free systems. *Elife* **3,** (2014).

42. Maeda, Y. T. *et al.* Assembly of MreB Filaments on Liposome Membranes: A Synthetic Biology Approach. *ACS Synth. Biol.* **1,** 53–59 (2012).

43. Mencía, M., Gella, P., Camacho, A., de Vega, M. & Salas, M. Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage Phi29. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 18655–60 (2011).

44. Blanco, L. *et al.* Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* **264,** 8935–8940 (1989).

45. Blanco, L., Lázaro, J. M., de Vega, M., Bonnin, A. & Salas, M. Terminal protein-primed DNA amplification. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12198–202 (1994).

46. Colletier, J.-P., Chaize, B., Winterhalter, M. & Fournier, D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol.* **2,** 1–8 (2002).

47. Shin, J. & Noireaux, V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *J. Biol. Eng.* **4,** 1–9 (2010).

48. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333,** 642–6 (2011).

49. Zelcbuch, L. *et al.* Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **41,** e98 (2013).

50. Mutalik, V. K. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* **10,** 347–53 (2013).

51. Osterman, I. A., Evfratov, S. A., Sergiev, P. V & Dontsova, O. A. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* **41 ,** 474–486 (2013).

52. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342,** 475–9 (2013).

53. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and

translation initiation elements. *Nat. Methods* **10,** 354–60 (2013).

54. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42,** 2646–59 (2014).

55. Chappell, J., Jensen, K. & Freemont, P. S. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* **41 ,** 3471–3481 (2013).

56. Pukelsheim, F. *Optimal design of experiments*. **50,** (siam, 1993).

57. Lentini, R. *et al.* Fluorescent Proteins and in Vitro Genetic Organization for Cell-Free Synthetic Biology. *ACS Synth. Biol.* **2,** 482–489 (2013).

58. Stogbauer, T., Windhager, L., Zimmer, R. & Radler, J. O. Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integr. Biol.* **4,** 494–501 (2012).

59. Hoops, S. *et al.* COPASI—a COmplex PAthway SImulator. *Bioinforma.* **22 ,** 3067–3074 (2006).

60. Whitaker, W. R., Lee, H., Arkin, A. P. & Dueber, J. E. Avoidance of Truncated Proteins from Unintended Ribosome Binding Sites within Heterologous Protein Coding Sequences. *ACS Synth. Biol.* **4,** 249–257 (2015).

61. Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484,** 538–541 (2012).

62. Shong, J., Huang, Y.-M., Bystroff, C. & Collins, C. H. Directed Evolution of the Quorum-Sensing Regulator EsaR for Increased Signal Sensitivity. *ACS Chem. Biol.* **8,** 789–795 (2013).

63. Marras, S. A. E., Gold, B., Kramer, F. R., Smith, I. & Tyagi, S. Real-time measurement of in vitro transcription. *Nucleic Acids Res.* **32,** e72 (2004).

64. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–5 (2009).

65. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Meth* **2,** 905–909 (2005).

66. Mena, M. A., Treynor, T. P., Mayo, S. L. & Daugherty, P. S. Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat Biotech* **24,** 1569–1571 (2006).

67. Vrzheshch, E. P. *et al.* Optical properties of the monomeric red fluorescent protein mRFP1. *Moscow Univ. Biol. Sci. Bull.* **63,** 109–112 (2008).

68. Sambrook, J. & Russell, D. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 2001). at <citeulike-article-id:144016>

69. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* **9,** 671–675 (2012).

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

71. Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat.*

Biotechnol. **27,** 753–9 (2009).

72. Shu, D., Khisamutdinov, E. F., Zhang, L. & Guo, P. Programmable folding of fusion RNA in vivo and in vitro driven by pRNA 3WJ motif of phi29 DNA packaging motor. *Nucleic Acids Res.* **42,** e10 (2014).
73. Baugh, C., Grate, D. & Wilson, C. 2.8 A crystal structure of the malachite green aptamer. *J. Mol. Biol.* **301,** 117–28 (2000).
74. Schifferer, M. & Griesbeck, O. Application of aptamers and autofluorescent proteins for RNA visualization. *Integr. Biol. (Camb).* **1,** 499–505 (2009).
75. Sachdeva, G., Garg, A., Godding, D., Way, J. C. & Silver, P. A. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Res.* **42 ,** 9493–9503 (2014).
76. BAYER, T. S., BOOTH, L. N., KNUDSEN, S. M. & ELLINGTON, A. D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* **11 ,** 1848–1857 (2005).