



UNIVERSITY OF TRENTO

International PhD Program in Biomolecular Sciences

XXVI Cycle

**Cellular mimics that sense and respond
to external stimuli**

Tutor

Dr. Sheref S. Mansy

Armenise-Harvard Laboratory of synthetic and reconstructive biology

CIBIO (Centre for Integrative Biology)

Ph.D. Thesis of

Laura Martini

Armenise-Harvard Laboratory of synthetic and reconstructive biology

CIBIO (Centre for Integrative Biology)

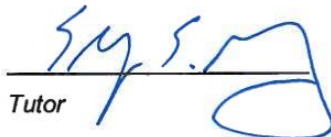
Academic Year 2013-2014

'Declaration

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



PhD candidate



Tutor

Abstract

To date little effort has been expended on the construction of cellular mimics from a minimum number of component parts. Such systems are desirable, because the cellular mimics could serve as useful tools to more deeply delve into the systems level reactions that sustain life and as a platform from which new types of technologies could be generated. Herein the building of cellular mimics that can sense and respond to external stimuli is presented. The majority of our efforts in building cellular mimics are directed towards the sensory element. Initially, previously characterized natural and artificial RNA sensors, i.e. a riboswitches, are exploited. Subsequently, the cellular mimics are implemented as chemical translators for natural bacterial cells.

To expand the capabilities of the engineered cellular mimics, we sought to develop a methodology for the selection of new RNA-based sensors capable of detecting new analytes. The tested methodologies were based on mRNA display and strand displacement reactions. The mRNA display selection did not lead to the identification of a sensor responsive to malachite green after eight cycles of selection. Conversely, via ligand induced triggering of a strand displacement reaction, new RNA sensors for thiamine pyrophosphate were selected from a small library. The sensors displayed translational control ability as is typical of certain classes of riboswitches. The strand displacement-based selection method represents a first step towards the *in vitro* evolution of sensing elements than can be exploited for new cellular mimics with programmable sensing capability.

Table of contents

Abstract	1
Abbreviation list	4
Chapter 1. Introducing cell-free synthetic biology	6
1.1 Compartment formation, growth and division	9
1.2 Genome replication	12
1.3 Protein Synthesis	14
1.4 Interaction with the environment: sensing and responding	15
Chapter 2. Cellular mimics sense and respond to external stimuli via riboswitches	25
2.1 Riboswitches as small ligand sensor devices	26
2.2 Artificial cells as chemical translators	55
Chapter 3. <i>In vitro</i> selection of new RNA- based sensors by mRNA display	72
3.1 <i>In vitro</i> and <i>in vivo</i> selection of functional nucleic acids	73
3.2 An <i>in vitro</i> riboswitch selection	75
3.3 Material and method	79
3.3.1 Library design	79
3.3.2 Reagents	81
3.3.3 mRNA display steps for riboswitch selection	82
3.4 Preliminary results for method development	84
3.4.1 Library assembly optimization	85
3.4.2 Transcription and RNA purification	87
3.4.3 Crosslinking reaction monitoring	88
3.4.4 Translation, fusion formation and isolation	91
3.4.5 Amplification testing at the selection cycle end	109
3.4.6 Control experiments concerning the MG ligand	112
3.5 mRNA display selection of a malachite green responsive riboswitch	113
3.6 Selection results and discussion	119

Chapter 4. <i>In vitro</i> selection of RNA- based sensors via ligand triggered strand displacement	126
4.1 A ligand induced conformational shift triggers strand displacement for RNA sensors selection	127
4.1.1 Methods	130
4.2 Results and Discussion	135
4.2.1 Small molecule binding to RNA can induce strand displacement	135
4.2.2 Strand displacement is an effective method of selecting RNA-based sensors	139
4.3 <i>In vitro</i> selection of TPP analogs- responsive RNA sensors	147
4.3.1 Material and method	148
4.3.2 Results and discussion	151
Chapter 5. Conclusions	161
5.1 Future perspective	163
Bibliography	165
Appendix	168
Acknowledgments	177

Abbreviation list

³⁵ S-MET	L- ³⁵ S Methionine
³² P-ATP	³² P- alpha adenosine triphosphate
αHL	α-hemolysin
AMP- PNP	adenosine 5'-(β,γ-imido)triphosphate
ATPS	aqueous two phase systems
A3PS	aqueous three phase systems
CPM	counts per minute
d(A)	deoxy-adenosine
d(T)	deoxy-thymine
FACS	fluorescence-activated cell sorting
FAKE – HIS	fake minus Histidine construct
GUV	giant unilamellar vesicles
HIS- TAG	six Histidine tag
HMP	4-amino-5-hydroxymethyl-2-methylpyrimidine
IVC	<i>in vitro</i> compartmentalization
IPTG	isopropyl β-D-1-thiogalactopyranoside
LUCA	last universal common ancestor
METADE	adenosine 5'-(α,β-methylene)diphosphate
MG	malachite green
NGS	next generation sequencing
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PCR	polymerase chain reaction
qPCR	quantitative (real time) PCR

RBS	ribosome binding site
RT- qPCR	reverse transcription quantitative PCR
RT- PCR	reverse transcription PCR
SDS	sodium dodecyl sulphate
TPP	thiamine pyrophosphate
SELEX	systematic evolution of ligands by exponential enrichment
UTR	untranslated region
YFP	yellow fluorescent protein

Chapter 1.

Introducing cell-free synthetic biology

This work has been adapted and improved from:

Piecing Together Cell-like Systems

Domenica Torino, **Laura Martini** and Sheref S. Mansy; *Current Organic Chemistry*, **2013**, 17, 1751-1757.

Attached after this introduction.

**Adapted and reprinted with the permission of the copyright holder*
© 2013 Bentham Science Publishers

Synthetic biology is the application of engineering studies to biological entities. In other words, the practical aspects of engineering disciplines are involved in understanding and unraveling biological complexity.¹ The gain in knowledge can also lead to a productive commercial output, e.g. engineered organisms for biofuel² or drug³ production, even if this is not the only focus. Artificial cell constructions are also investigated.⁴

Synthetic biology can be divided into two major subgroups of analysis and application: *in vivo* and *in vitro* (Fig.1.1). The former exploits the genetic engineering of bacteria, mammalian cells and yeast. The aim is to reprogram cellular behaviour. This can be achieved by the assembly of devices with standard and modular biological parts, e.g. transcriptional promoters and ribosome binding sites (RBS), to control a specific bacterial function or output. *In vitro* synthetic biologists still use and engineer biological devices, but these are employed to *mimic* cellular function, such as protein expression or sensing.⁵ In other words, *in vitro* synthetic biology focuses on the reconstitution of biological functions in a cell-free environment. The building blocks are the components a cell offers, biological macromolecules as lipids, nucleic acids, proteins, either natural or artificial.⁶

Mimicking cellular function means to reconstruct those properties that define matter as alive. A possible drawback in this definition is that life by itself has still to be defined. We know some of the properties that describe life, but we do not know the threshold for general identification. A so called top-down approach can be followed to determine a minimal set of functions a living system displays.⁷ In this method, depletion of unwanted functionalities and genes down to a minimal system starts from a natural cell. In other words, the complexity of a living system is step-by-step reduced, even if uncertainty is never removed completely. Conversely, a bottom-up approach tries to remove the mysteries behind life by reconstituting their properties from scratch.⁸ The assembly parts are defined, known, non-living components. The final result is a laboratory-made cell-like system. Although these cellular mimics are a simpler form of a cell, from their reconstitution we may gain insight into cellular function and evolution.

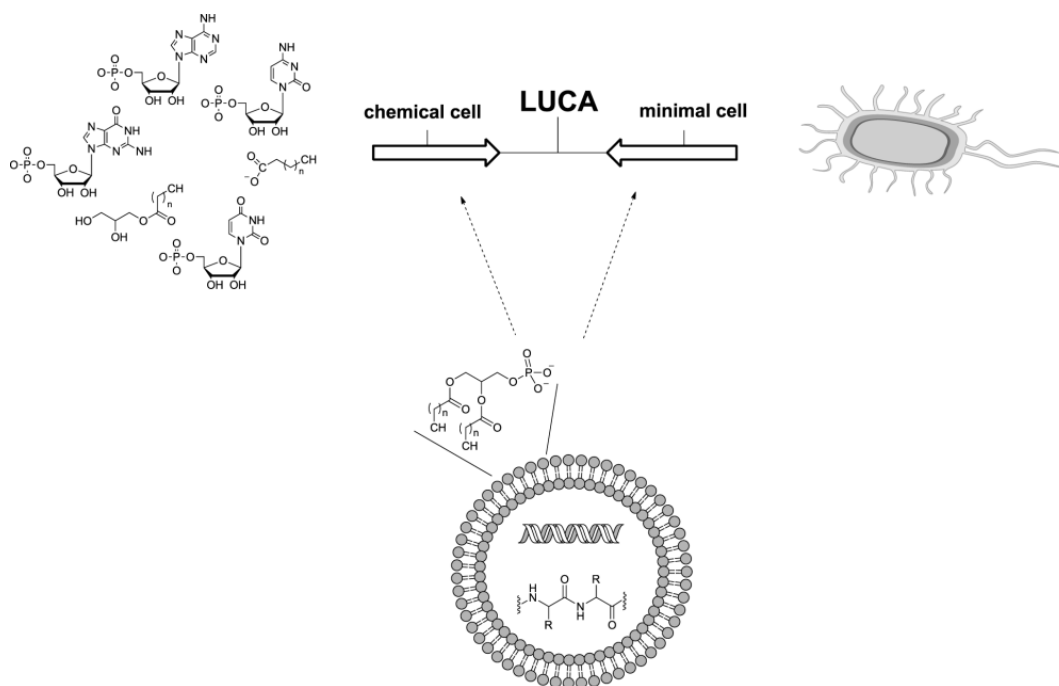


Fig.1.1 Different approaches to building new, artificial cells. Typically, laboratories begin with chemicals (left) or an existing cell (right). Presumably an approximation of the last universal common ancestor (LUCA) exists in between these two extremes. A third approach is to piece together cellular mimics from existing biological components (bottom). Reprinted with the copyright holder permission, © 2013 Bentham Science Publishers.

This thesis work exploits a bottom-up approach for the building, characterization and evolution of new cellular mimics with sensing ability. In this chapter all the necessary components that we can combine for building a system with life-like properties are considered. Generally, in order to build a cellular mimic, the necessary components are a compartment able to grow and divide, a genetic information polymer that can self-replicate, a system for protein synthesis and a capability to respond and adapt to the environment (Fig.1.2).

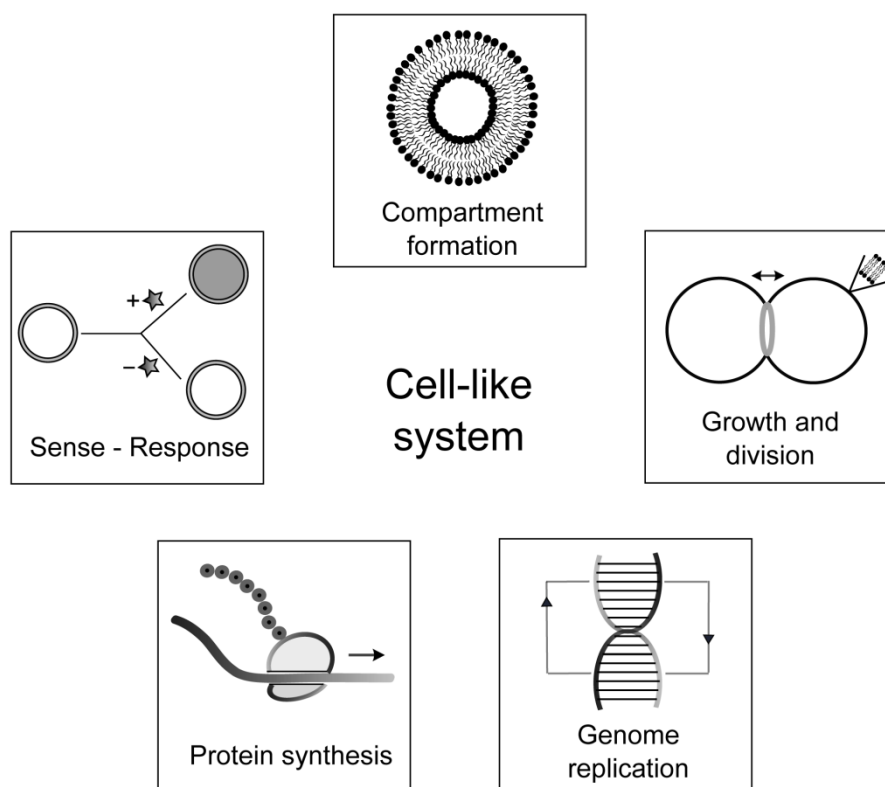


Fig.1.2. Life-like properties a cellular mimic can be implemented with.

1.1 Compartment formation, growth and division

The compartment function is to define an internal space from an outside environment, i.e. a barrier. In the inside, molecules can group, interact with each other and evolve.⁹ Energy can be stored and thus used to promote biological reactions.¹⁰ Compartments can be composed of lipids or proteins. Lipids have been widely used in laboratory constructions of cell-like systems. The ease of forming bilayer structures in water is the primary cause for their employment. Vesicles can be formed simply by vortexing a lipid film in an aqueous buffer. The lipid component can consist of fatty acids, mono- and diacylphospholipids, or other amphiphiles. The lipid composition affects the stability of the vesicles to solution conditions. While fatty acids demonstrate low stability and robustness,

phospholipids can stand wider ranges of conditions. However, fatty acid vesicles can grow, divide and uptake nutrients.^{9,11,12} On the contrary, phospholipid compartments are semi-permeable and nutrient uptake by simple diffusion is partially inhibited. Thus, phospholipid vesicles require an engineering approach to exchange small molecules with the external solution. Pore forming proteins, e.g. alpha-hemolysin,¹³ or physical mechanisms that induce packaging defects, such as those that arise at the phase transition temperature,¹⁴ can be used. Moreover, changes in the lipid composition can modulate vesicle permeability. For example, cholesterol concentration above 20 mol% can reduce the permeability of solutes across phospholipid vesicle membranes¹⁵.

Vesicles are not the only system exploited for building cellular mimic compartments. Water-in-oil emulsion droplets are easily formed by mechanical stirring a mineral oil lipid phase with an aqueous solution.¹⁶ The result is a dispersion of internally hydrophilic compartments in an oil interface. Such systems allow high encapsulation efficiency of water soluble molecules, but lack the possibility to exchange directly hydrophilic components with the external phase.¹⁷ Water-in-oil emulsions are better suited for protein and nucleic acid *in vitro* evolution, rather than for constructing cellular mimics.^{16,18}

Systems that mimic the cellular cytoplasm and micro-compartments have also been exploited, i.e. two (ATPS) and three (A3PS) phase systems. Their composition is based on the combination of a polyethylene glycol (PEG)-rich solution and a dextran aqueous phase within giant vesicles.¹⁹ Recently, an ATPS has also been engineered to act as a microscale bioreactor.²⁰ The PEG/dextran solution is mixed with PEGylated liposomes. The resulting bioreactor consists of an aqueous two-phase system with an interface of liposomes. This interesting system where aggregates of liposomes *en masse* provide barriers that define a larger aqueous droplet display increased permeability thereby facilitating the uptake of nutrients and larger macromolecules in comparison to traditional liposomes.²⁰ Moreover, the compartments are typically of similar sizes and are better suited for the study of cellular division (see below).

In the perspective of building a cellular mimic, the compartment should not be considered just as a physical static element. In fact, cell membranes are dynamic entities, able to grow and divide. Division has been investigated in cellular mimic constructions starting from different points of view using protein and non-protein based approaches. Protein-based efforts arise from studies on the divisome of *E. coli*.²¹ FtsZ is a monomeric protein that polymerizes to form a division ring that allows cell separation. The polymerization occurs in the middle of the dividing cell. Correct localization of this constriction Z-ring is due to a set of accessory proteins, belonging to the Min family. Although the expression of Fts and Min proteins together *in vitro* into a functioning divisome has still not been fully achieved, some attempts have demonstrated their feasibility in reconstructing division processes in vesicles. For example, FtsZ has been engineered to attach directly to tubular vesicles membrane and, upon GTP dependent polymerization, causes indentation to the membrane surface.²² Recently, the interaction of FtsZ polymers and the membrane anchor protein ZipA have been studied *in vitro* in giant unilamellar vesicles (GUV).²³ At a high density of ZipA molecules, shrinkage of the vesicles is observable. This effect resembles the division constriction forces obtained in living cells.

Non-protein based division processes are based on chemical-physical forces, such as osmotic pressures. For example, these forces cause budding and division in phospholipid vesicles composed of multiple types of lipids that can separate into different phase domains (ordered and unordered) resembling natural membrane rafts.²⁴ The limit of this approach is that only one single division is possible, since the daughter vesicles do not retain the same lipid composition as the mother vesicles. Conversely, vesicles containing an ATPS in a hypertonic solution are able to overcome the one-cycle division limitation and to proceed to a second generation of vesicles.²⁵ Notably, this system demonstrated how it is possible to reconstruct a simple form of division *in vitro*, even if it is not fully comparable to an extant living cell. In fact, a vesicle growth phase would be required for a complete reconstitution of an ongoing replication cycle of the compartment.

Growth of lipid compartments has been achieved through the simple addition of micelles to fatty acid vesicles.¹¹ Upon the addition of external forces, e.g. agitation or through a radical mediated oxidation of thiol group containing compounds that interacts with the membrane,²⁶ division cycles are also possible. Thus vesicle growth and division coupling is feasible. However, the system is based only on fatty acid vesicle, while phospholipids do not retain the same dynamics. Nevertheless, Luisi and coworkers reconstituted one step of the diacyl glycerophospholipid synthesis pathway within a vesicle via enzymatic synthesis.²⁷ The internal supply system allows the phospholipid vesicle to grow, since the newly formed lipids naturally insert into the bilayer.

1.2 Genome replication

One property of living systems is reproduction. In this process, the division to daughter cells is usually coupled to and preceded by replication of the genetic information. The genome can be either DNA or RNA. In the laboratory, DNA amplification reactions are routinely exploited. However, reconstruction of PCR reactions in vesicles, although possible,²⁸ is not feasible for the construction of a cellular mimic due to the need of temperature steps for thermocycling that the cellular mimic cannot regulate. Moreover, the reaction to proceed requires oligonucleotides to act as primers. Thus, different methods have to be explored. *In vitro* reconstruction of complete genome replication is achieved via the viral phi29 system. This bacteriophage in fact possess the ability to self-replicate a linear DNA genome using only four proteins, including a DNA polymerase, a single strand binding protein, a double strand binding protein and a terminal protein. The process is isothermal and is demonstrated to work *in vitro*.²⁹ Since the priming is protein based, there is no need to add oligonucleotide primers to initiate replication. Further, since replication in the phi29 system is end-to-end, the telomere problem is avoided. Non-phi29-based isothermal DNA replication systems require a DNA polymerase, a single strand binding

protein, and a helicase. To overcome external oligonucleotide addition, a primase can also be included.³⁰ However, the later removal of the RNA primers synthesized by the primase would cause a telomere problem in that the ends of the DNA would not be replicated. Therefore, the DNA would be shortened after each round of replication.

In the aim of overcoming DNA complexity of replication and of reducing the system to minimal terms, RNA has been considered by some to be an alternative genetic polymer to DNA. The usage of RNA genomes would notably minimize the complexity of information processing from genes to protein, as RNA would act both as information storage and translator. RNA polymerases also do not require primers thus the telomere problem would not exist. However, the coupling of transcription and translation opens the possibility of competition between replication and protein synthesis processes. The folding of the RNA may also influence the RNA polymerase or ribosome binding.³¹ The RNA genome would require more than 100 genes to sustain protein synthesis. Nevertheless, the information storage ability and catalytic activity of the RNA could potentially remove the need of protein synthesis.³²

In a recent publication,³³ Yomo and coworkers build an artificial system able to replicate and evolve an RNA genome. The mechanism involves the translation of a genome encoded RNA-dependent RNA polymerase, i.e. the Q β replicase, in a cell-free system. Through cycles of fusion and division of the compartment, manually performed, the RNA genome evolves towards a reinforcement of the interaction with the replicase. Even if the protein expression factors are externally supplied, this experiment shows how it is possible to build cellular mimics, able to self-replicate and evolve in a Darwinian manner. Nevertheless, the division and fusion cycles are still decoupled to the replication process. In this direction, a study from Sugarawa et al.³⁴ demonstrates how DNA replication can guide a vesicle to divide. However, the composition of the bilayer membrane is not retained after one division cycle.

1.3 Protein Synthesis

Naturally reduced genome studies show the evolutionary preservation of genes that are involved in protein synthesis.³⁵ A core set of almost 100 genes that encodes the protein expression factors has been established.³¹ Thus, achieving protein synthesis in cell-like systems is evolutionary relevant. Moreover, translation processes sustain many of the described features presented above. For these reasons, accomplishing protein synthesis *in vitro* in a cellular mimic is important. Transcription is usually carried out by bacteriophage DNA-dependent RNA polymerases, e.g. from T7 or Sp6. The translation machinery used is mostly based on living cells. Bacterial and eukaryotic components are exploited in cell extracts, i.e. S30 lysates and rabbit reticulocyte lysates, respectively. The undefined mixtures of lysates retain the complexity of cellular systems, hiding components and cofactors concentrations. Otherwise, defined, purified components that allow protein expression can be used. This system (PURE system), based on the technology developed by Ueda and colleagues,³⁶ is composed of singularly purified elements necessary for protein production. Thus all the components of the *in vitro* reaction are known. Moreover, the PURE system lacks protease and RNase activity.

The encapsulation of the PURE system has been exploited for protein synthesis in vesicles and in water-in-oil emulsions³⁷ and recently in APTS water-in oil droplets.³⁸ However, the full reconstitution of the transcription and translation machinery in cell-like systems has still to be performed. In other words, it is possible to encapsulate by external addition the components, but the elements cannot be regenerated in the compartment. The unfeasibility of component regeneration is a main limitation to the use of cellular mimics. In fact the system is not able to sustain long-term activity. From another point of view, the lack of internal regeneration is also an advantage, since there is no fear of losing control of the artificial cell.

1.4 Interaction with the environment: sensing and responding

The features of life described until now characterize a cellular mimic as a system able to replicate, grow and divide. However, the system would just be a closed compartment, excluding uptaking of nutrients, performing functions in a repetitive way. Nevertheless, living organisms interact and adapt to the environment, mostly regulating internal processes depending on external factors. The information flow is not only one-way directed to the cells. Living organism shape and modify their surroundings, establishing a communication with the environmental solution. Absence of the reciprocal communication leads to cell death. In order to survive, a cell needs to adapt to the external changes. To mimic this adaptation process, an artificial system should be able to sense and respond to stimuli. A sensor is an intracellular converter which allows for the decoding of a chemical message and to actuate a response reaction with an effector domain. The output is the result of signaling cascade that leads to gene expression control. The sensing mechanism is exploited via protein or other molecule-based approaches.

Prokaryotic two component signaling belongs to protein-based regulation class. These systems are generally composed of a membrane receptor molecule that self-activates upon ligand binding, starting a cascade response pathway. For example a ligand-bound receptor phosphorylates a transcription factor, which in turn regulates gene expression. Two component systems are an excellent toolbox for synthetic rewiring of natural circuits.⁶ In fact, through rational design it is possible to alter the regulation specificity towards new substrates.³⁹ Moreover, chimeric proteins can be designed to act as molecular switches in response to multiple or excluding signals.⁴⁰

Protein-based regulation also includes transcription factors responsive to small molecule ligands. Negative feedback loops have been characterized *in vitro* and in liposomes, e.g. an arabinose responsive circuit.⁴¹ Notably, the seven sigma factors of *E. coli* polymerase are exploited, adding control layers in the reaction.

However protein-based sensing relies mostly on already characterized and preexisting repressors and transcription factors. In other words, new combinations of behaviours are achieved, but the players are the same. Attempts to evolve new protein sensors include the selection of orthogonal systems, which increase the layers of specific control in reactions. For example, recently, the evolution of orthogonal T7 RNA polymerases-promoters pairs have been shown *in vitro* and *in vivo*, and up to six combinations have been used simultaneously without cross interference.⁴² Nevertheless, other approaches for adaptation mechanisms should be considered.

RNA offers many advantages in the construction of a sense-response mechanism.⁴³ RNA is a versatile tool from designing and engineering viewpoints. RNA sequences are easily manipulated since the building components and the interactions are known. Moreover, most of the properties of the system relies on RNA secondary structure, which can be readily modeled with computational software, based on the thermodynamics of structure prediction, e.g. Vienna RNA fold package, or on kinetic parameters for folding investigation.⁴³ These RNA-based devices should be preferred to protein regulatory systems since the actuation of the information processing is faster (not involving the protein synthesis apparatus) and less energy consuming for the system. In other words, synthetic circuits that rely only on RNA function can be built *in vitro* and compartmentalized.^{32,33}

The assembly of RNA-based sensor devices can be dependent on already existing natural RNA controllers or newly engineered regulators. Selections, both *in vivo* and *in vitro*, are also a possible way to build new RNA tools, expanding the possibility of the system towards sensing new ligands (see Chapter 3). The final results are molecules able to sense a specific molecule and cause a direct or non-direct effect. Direct control involves a conformational change caused by ligand binding correlated with an activation of the response actuator domain. For example, riboregulators called antiswitches⁴⁴ have been engineered in yeast. Antiswitches directly respond to small molecule binding to the aptameric regions through the exposure or sequestration of a mRNA antisense domain (ON

and OFF antiswitches, respectively). Conversely, the non-direct effect is related to a distinct transmitter domain, e.g. a bridge helix, which coordinates the ligand binding event to functional regulation.⁴⁵ Many systems are engineered by joining an aptamer domain with a ribozyme domain. The Hammerhead ribozyme has been coupled and engineered as transmitter of two functionalities: ligand binding and cleavage activity. These aptazymes (the name derives from the sum of the two single components) are both rationally designed⁴⁶ and selected.⁴⁷ The actuator domain, upon activation, regulates gene expression by several mechanisms, different for bacteria, yeast and mammalian applications. For example, translation can be modulated via RBS sequestration in *E. coli*,⁴⁸ or via alternative splicing in eukaryotes.⁴⁹

The majority of artificial devices are applied to *in vivo* studies for cellular behaviour reprogramming. However, the next chapter presents a mechanism of *in vitro* application and characterization of a synthetic RNA regulator, i.e. an engineered riboswitch.⁵⁰

Piecing Together Cell-like Systems

Domenica Torino, Laura Martini and Sheref S. Mansy*

CIBIO, University of Trento, via delle Regole 101, 38123 Mattarello (TN), Italy

Abstract: Several laboratories are pursuing the synthesis of cellular systems from different directions, including those that begin with simple chemicals to those that exploit existing cells. The methods that begin with nonliving components tend to focus on mimicking specific features of life, such as genomic replication, protein synthesis, sensory systems, and compartment formation, growth, and division. Conversely, the more prevalent synthetic biology approaches begin with something that is already alive and seek to impart new behavior on existing cells. Here we discuss advances in building cell-like systems that mimic key features of life with defined components.

Keywords: Cell-like, minimal cell, origin of life, protocell, riboswitch, synthetic biology.

BUILDING CELLULAR SYSTEMS FROM THEIR PARTS

Building life from scratch in the laboratory is an old dream with new tools at its disposal. We can now rapidly and affordably synthesize genes [1], assemble genomes [2], evolve new function [3] and make precise changes throughout an existing genome [4]. Most of these technological advances are applied to the modification of existing cells (Fig. 1), and so the resulting data do not directly address life's beginnings or clearly delineate the required components of cellular function. Any work that either exploits existing cells or cell lysates makes use of a complex, undefined mixture of reaction components that we do not have the tools to fully understand. Here we try to highlight steps forward in building fully defined life-like systems from a minimum number of components.

It is hoped that the process of building cell-like systems in the laboratory will give us insight into what is required to endow a system with the properties of life. There is no currently agreed upon threshold that must be crossed in order to label a chemical system living. However, as progress is made in recreating the functions of life in the laboratory, we may reach a point in which a threshold is crossed, even if it is not recognized until afterwards. Further, since the system would have been built with fully defined components, this systematic approach should give us a much better understanding of the necessary components of life. It is worth noting that this approach does not probe how the molecules of life were built, nor does it directly address the historical path taken between the prebiotic chemistry of Earth [5-11] and the emergence of protocellular structures [12-14].

The central dogma [15] of molecular biology offers one perspective on what is needed to build a cell. Typically, information stored in a replicating DNA genome flows through RNA and then finally to proteins. However, this simple description of cellular life is in fact not very simple, which is born out by analyses of microorganisms. *E. coli*, for example, has well over 4000 genes, and purposeful genetic reductions only reduces the genetic content by 15% [16]. On the smaller end of the spectrum, *Mycoplasma genitalium* has 482 genes, and genetic knockouts have shown that approximately 100 of these genes are not required under laboratory condi-

tions [17]. But even with this small genome, which is small enough to be synthesized [2] approximately one third of the genes provide unknown function to the cell. We have reached a point where our technological ability to synthesize genomes has outpaced our understanding of what we are synthesizing.

By comparing sequences of disparate microorganisms, Moya and others suggest that a minimal cell would contain on the order of 200 genes, of which more than half would be necessary for protein synthesis [18]. Interestingly, a natural symbiotic microorganism, *Carsonella ruddii*, has only 182 genes, a value similar to theoretical predictions of a minimum gene set [19]. Further, over half of the *C. ruddii* genome is dedicated to protein synthesis. In some ways, the impression is given that living systems are nothing more than just a bag of protein synthesizing machinery. Clearly life is more than just protein synthesis, but at least as far back as the last universal common ancestor, protein synthesis has been a crucial aspect of cellular function [20].

One conception of a simplified, laboratory-made cell consists of a vesicle compartment that contains a replicating DNA genome and transcription-translation machinery that responds to changing environmental conditions (Fig. 2). Much of the needed functions for such a cell-like system appears to depend on protein function. Nevertheless, origins of life research has shown that under specific chemical conditions, several features of life emerge without the participation of proteins. Perhaps future approaches that combine the lessons learned from origins research with those gained from attempts to exploit biological machinery will allow for the synthesis of a simplified cell.

COMPARTMENT TYPES

Compartmentalization is considered to be one of the key steps along the transition from simple chemistry to cellular life [21]. The enclosure of a chemical system within a semipermeable membrane causes several useful features to emerge. For example, encapsulation facilitates evolutionary processes [22, 23, 24], provides for an energy storage mechanism [21-25] and likely influenced accessible chemistry. Although it is possible that prebiotic boundary structures were defined by substances other than lipids, no living systems to date have been identified that are capable of surviving without lipid membranes. Further, several lines of evidence argue for the presence of lipids on prebiotic Earth, including simulated prebiotic syntheses of lipids [5-11] and the identification of lipid

*Address correspondence to this author at the CIBIO, University of Trento, via delle Regole 101, 38123 Mattarello (TN), Italy; Tel: +39 0461-883438; Fax: +39 0461-883937; E-mail: mansy@science.unim.it

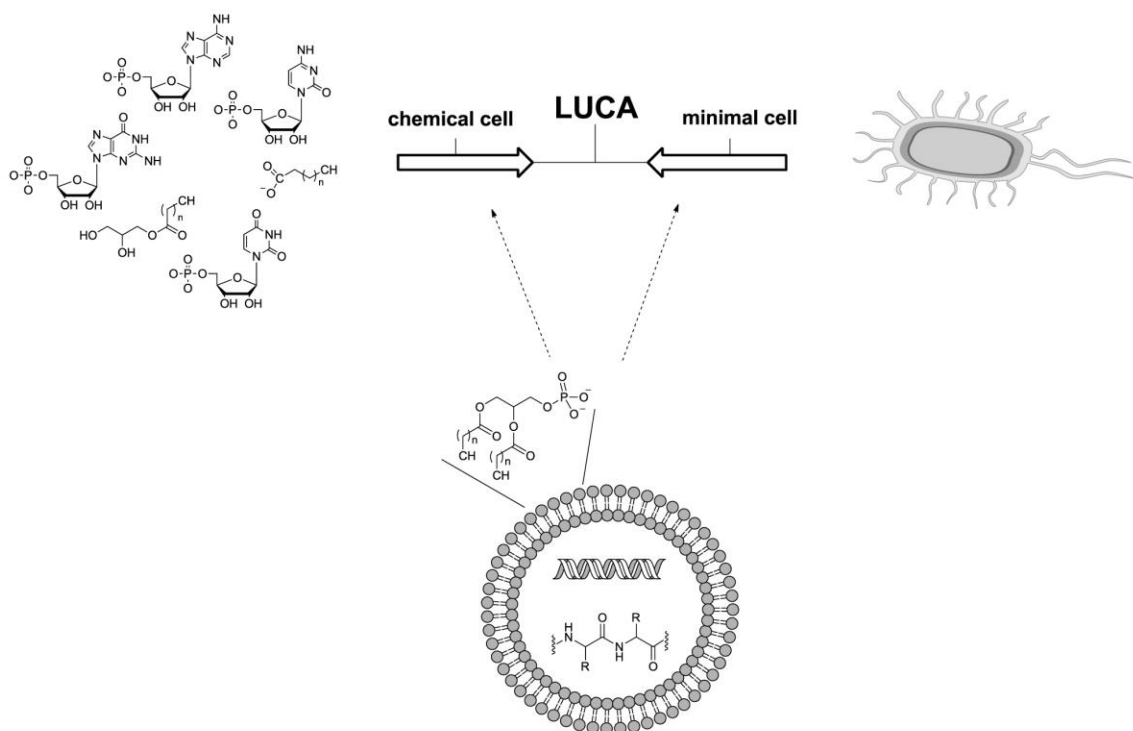


Fig. (1). Different approaches to building new, artificial cells. Typically, laboratories either begin with chemicals (left) or an existing cell (right). Presumably an approximation of the last universal common ancestor (LUCA) exists in between these two extremes. A third approach is to piece together cellular systems from existing biological components (bottom).

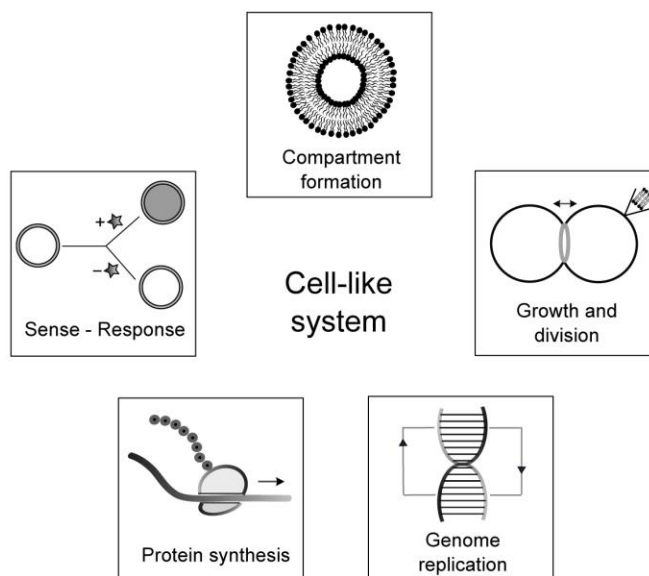


Fig. (2). Features of cellular life that are mimicked by cell-like systems.

molecules within carbonaceous meteorites [8]. Finally, vesicles form easily in aqueous solution, thus suggesting that there were vesicles on Earth even before there was life. The latter point has led

some to suggest that there once existed a lipid world in which hereditary was mediated by lipid composition rather than by specific nucleic acid sequences [26].

Prebiotically plausible lipids are generally thought to be saturated, single-chained amphiphiles. In the laboratory, fatty acids and fatty acid derivatives are often used as an approximation of what could have been present. Not only do such lipids form vesicles, they also exhibit many useful features not dependent upon protein function, including the ability to grow and divide, uptake nutrients, and retain macromolecules [23,25,27-34]. The main disadvantages of fatty acid based vesicles is the encountered difficulty in recovering encapsulated enzymatic activity from some enzymes and the vesicle's decreased stability in comparison with phospholipid vesicles. For example, fatty acid vesicles are stable over a narrow pH and salinity range [30,35,36] and difficulty has been encountered in reconstituting DNA polymerase activity within fatty acid vesicles [32-37]. Therefore, fatty acid vesicles are typically used for protocellular research rather than for attempts to build cell-like systems similar to life as we know it.

Contemporary cells exploit membranes of complex composition including monoacyl and diacyl lipids and proteins. Laboratory constructions tend to ignore this complexity and instead rely on the ease in which many lipids alone form vesicles. Of the commonly used vesicle systems, those built with diacylphospholipids are the most robust. However, this robustness comes at a cost. Diacylphospholipids are generally impermeable, thereby posing a difficulty in their use for building cell-like structures. One approach to overcome this limitation is to exploit membrane proteins, such as the bacterial toxin α -hemolysin. This protein expresses as a soluble monomer that then spontaneously oligomerizes into a pore in the presence of a membrane [38]. An alternative approach is to utilize the phase transition temperature of lipid membranes to create packing defects that can be exploited for permeation [39]. Both the α -hemolysin and packing defect mechanisms have been used to feed in substrates for encapsulated RNA [40] and protein synthesis reactions [38]. Interestingly, symbiotic organisms with small genomes, such as *C. ruddii* and *Buchnera aphidicola*, are thought to be heavily dependent on passive diffusion mechanisms for nutrient uptake [41]. In short, phospholipid membranes need not be viewed as an impenetrable barrier. Simple passive diffusion based mechanisms exist for nutrient uptake and waste release that may not be much different than what is used by some microorganisms with small genomes.

There is another approach to building compartments with dimensions similar to extant cells. Water-in-oil (w/o) emulsions are easy to make and are extremely efficient in encapsulating hydrophilic molecules. They have been used extensively for molecular evolution experiments [42, 43] and further developments in the technology allow for the delivery of reactants directly to the water droplets without breaking the emulsion [44]. However, w/o emulsions have not proved useful for constructing cell-like systems due the lack of solute exchange across the water-oil interface.

COMPARTMENT GROWTH & DIVISION

A compartment must grow and divide to allow for the cell-like system to replicate. For fatty acid vesicle systems, growth and division can be accomplished simply by adding lipids to preexisting vesicles [27-29,33,34]. Similar mechanisms for phospholipids are not available because of the decreased dynamics of diacyl phospholipids [45]. However, if phospholipid synthesis reactions were reconstituted within vesicles, then once the lipids were produced they would naturally partition into the membrane resulting in vesicle growth. The difficulty with reconstituting such enzymatic activities is that many of the lipid synthesis enzymes are membrane pro-

teins. Nevertheless, Luisi and colleagues have built an encapsulated enzyme system that can produce diacyl glycerophospholipids when provided with fatty acids and glycerol [46]. Although this is an important development, more work is needed to move beyond the use of simple lipids (fatty acids) to build more complex lipids (phospholipids). Also, additional enzymes may be required to flip a fraction of the newly synthesized lipids from the inner- to the outer-leaflet.

Phospholipid vesicles can be coerced into dividing through simple chemical-physical forces. For example, Baumgart *et al.* found that phospholipid vesicles containing liquid ordered (L_o) and liquid disordered (L_d) domains can bud and divide when placed under osmotic stress [47]. However, since the daughter vesicles do not retain the same membrane lipid composition as the parental vesicle, further rounds of division are not possible. Presumably if an additional mechanism were incorporated to restore the original membrane composition then further rounds of division could occur. More recently, Andes-Koback and Keating revealed that osmotic gradients can induce budding and division of phospholipid vesicles that contain an aqueous two-phase system, one enriched in dextran and the other aqueous phase enriched in polyethyleneglycol [11]. Since the resulting daughter vesicles retain the composition of the parental vesicle, they are also capable of dividing. If the Andes-Koback and Keating mechanism were combined with a vesicle growth system, then a growth – division cycle could be built. The system in its current state, however, does require external intervention to adjust the osmolality of the extravascular space. It is interesting to note that some biological evidence exists suggesting that life can persist in the absence of protein mediated division [48, 49].

Protein based systems have also shown promise in dividing vesicles. Although *in vivo* cell division mechanisms are complex [50], recent studies have shown that parts of the system can be reconstituted *in vitro* [51, 52]. For example, FtsZ is a highly conserved division protein that polymerizes into a ring that interacts with other members of the divisome to split the cell into two. Even if FtsZ is not thought to directly interact with the inner membrane in natural systems, FtsZ constructs can be engineered to anchor into the membrane of synthetic vesicles by the addition of a small helix to the C-terminus of FtsZ [52]. Further, in the presence of GTP this engineered version of FtsZ forms constricting rings that cause visible indentations within the membranes of tubular vesicles [52]. While these vesicles do not divide, the data suggest that a proper mix of Fts proteins could be sufficient to reconstitute functional cell division machinery within vesicles.

REPLICATING GENOMES

There are many ways to copy nucleic acids *in vitro*, but none are currently amenable to the construction of cell-like systems. The use of an RNA genome is attractive for at least two reasons. First, RNA genomes appear to simplify the system by removing the need of a class of molecules, i.e. DNA. Second, RNA polymerases do not require a primer. Although there are no known cells that use an RNA genome, there are viral systems that could be exploited. For example, the bacteriophage ϕ 10 uses a double stranded RNA genome that is fully replicated by a single RNA polymerase [53]. This may seem like a simple system to reconstitute in the laboratory for building cell-like systems, but significant complications may arise when protein synthesis machinery is incorporated. Ribosomes require single stranded RNA as a template. This means that in addition to the double stranded RNA genome, single stranded RNA also

must be present for protein synthesis. In other words, transcription of the genome, even an RNA genome, is always required. Although there are several classes of RNA viral systems with different genome organizations, e.g. viruses that use a circular RNA genome, they all share these same difficulties. An exciting alternative RNA system would exploit an RNA replicase, i.e. an RNA enzyme that copies an RNA genome. Despite the impressive advancements made in building better RNA replicases [54] much more progress is needed before they serve as a practical alternative.

The use of a DNA genome has the advantage of better separating replication and transcription, thus avoiding competition between the two processes [55]. Additionally, methods to replicate DNA *in vitro* have existed for several decades. Some attempts have been made to exploit these technologies for building cell-like systems. PCR based mechanisms for DNA copying inside of phospholipid vesicles have been demonstrated [56]. While useful as a first step, this approach is not practical, because it requires manipulation that is not regulated by the cell-like system (i.e. thermocycling), and it requires the addition of oligonucleotide primers. Another approach uses a more complex mixture of proteins, including a helicase and single strand binding proteins, in addition to a DNA polymerase to replicate DNA inside of vesicles. This system overcomes the thermocycling limitation of PCR, but still requires the addition of oligonucleotides, and it is only capable of replicating short (<100 bp) strands of DNA [37]. An extension of this technique would exploit a primase [57] to remove the need of adding oligonucleotide primers, but this would also add a significant problem. Genomic replication must be complete, which means that there must be a mechanism to ensure that there is no loss of the genomic termini, i.e. telomeres. Since primases add RNA primers that must later be removed, the simplified systems described so far are incapable of fully replicating a genome end-to-end.

The telomere problem is perhaps the biggest challenge to constructing a simple genomic replication mechanism. Viral DNA replication systems could potentially give a simpler solution to this problem than the isothermal bacterial systems described above. For example, the *Bacillus subtilis* bacteriophage phi29 uses only four proteins to replicate its entire genome, including a highly processive DNA polymerase that possesses strong strand displacement activity, a single strand DNA binding protein, a double strand DNA binding protein, and an additional protein that is required to initiate replication (terminal protein) [58]. The natural phi29 genome is linear, contains covalently attached terminal protein at the 5'-termini, and is replicated fully via a protein priming mechanism. Despite the peculiarities of the phi29 system, Salas and colleagues recently demonstrated that DNA can be fully replicated *in vitro* with this four protein component system if the template DNA contains appropriate nucleotide sequences at both termini to define the origins of replication [59]. Thus far the phi29 genomic replication system is the simplest and best characterized isothermal system available.

PROTEIN SYNTHESIS

Much of what has been described is dependent upon the activity of proteins. Cell-like systems that require protein function, therefore, require transcription and translation machinery. Transcription is a very straightforward process to reconstitute *in vitro*, particularly if the commonly used bacteriophage RNA polymerases are exploited. T7 and SP6 RNA polymerases consist of a single protein domain, do not require accessory factors for function, and provide robust activity. Conversely, the synthesis of proteins is a highly

complex process requiring over 100 genome encoded components [60]. Nevertheless, this complicated process is understood well enough to be reconstituted *in vitro* from purified, defined components [61]. Moreover, several laboratories have shown that this minimal transcription-translation system functions in water-in-oil emulsion droplets [62] and in vesicles [63, 64] despite the statistical difficulty associated with encapsulating multiple components within a single compartment [65, 66].

ADAPTING TO CHANGING ENVIRONMENTAL CONDITIONS

Not every process within a cell needs to be monitored and coordinated with other cellular functions. For example, mitochondrial genome replication is not thought to be coordinated with division. Nevertheless, no living system simply repeats the same functions over and over again without regard to its surroundings. Cells must constantly adapt to changing intracellular and extracellular conditions. This is largely due to the fact that life and the environment are intimately linked. Life both feeds off of and shapes the environment. Conversely, the environment dictates which living things can and cannot survive. Therefore, if a cell is to survive for an appreciable length of time, the cell must be able to continually adapt to changing environmental conditions, some of which are induced by the cell's own existence.

To adapt to changing environmental conditions, a cell needs to be capable of sensing and responding to stimuli. A straightforward solution to the problem would be to encode a sense-response system that exploits environmentally responsive transcription factors. For example, proteins such as FNR, IscR, and CooA control gene expression in response to oxygen, iron, and carbon monoxide levels, respectively [67,68]. Alternatively, bacterial two component systems [69] could be similarly used to alter gene expression profiles in response to environmental changes. Here a sensor-kinase would control the phosphorylation state and thus the activity of a response-regulator protein, which would then result in changes in gene expression (Fig. 3). Further layers of control could be built into sensory pathways by exploiting the sigma factors of bacterial polymerases. While bacterial RNA polymerases are more complicated than T7 or SP6 bacteriophage RNA polymerases, their use of initiation factors to guide promoter recognition provide for an opportunity to build in additional control elements. Although thoroughly studied *in vivo*, few attempts have been made to reconstitute any of these protein sensory systems *in vitro* [70].

A disadvantage of relying on existing proteins for sensing capabilities is that the use of existing proteins severely limits the types of cell-like systems achievable. In other words, we are forced to build cell-like systems that sense what natural cells are already capable of sensing if we use existing proteins. Although some examples of building new proteins with desired activities have been reported [71, 72], the present methods are not yet sufficiently developed to be widely applicable. Therefore, other sensing mechanisms that are easier to engineer are desirable.

Natural and synthetic RNA sensors that control protein production exist and are an attractive alternative to protein based sensing mechanisms. One class of RNA sensors is known as riboswitches and exist in the 5'-untranslated regions of bacterial genes. Riboswitches either turn on or off transcription or translation in response to ligand binding directly to the mRNA [73] (Fig. 3). The distinct advantage of RNA sensors over protein sensors is that we have years of RNA selection and engineering technologies [74] to build on, and so it is possible to build synthetic riboswitches capa-

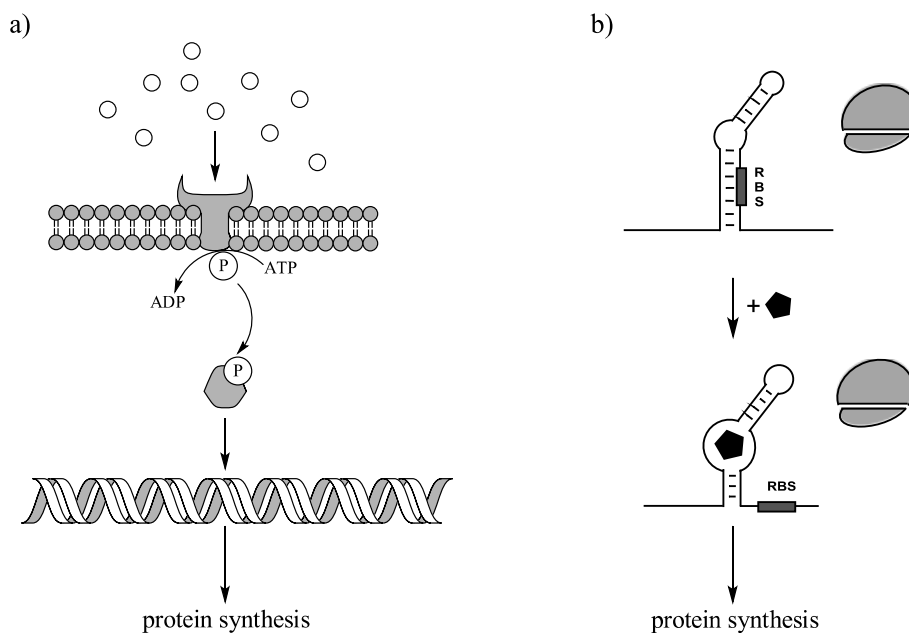


Fig. (3). Simple biological sensors could be built from protein or RNA components. (a) A two component protein system consisting of a sensor-kinase and a response regulator. In this example, a membrane bound sensor-kinase phosphorylates the response regulator only in the presence of ligand. The phosphorylated response regulator then activates gene expression. (b) An example of one type of riboswitch system that directly controls protein production in response to the presence or absence of ligand. Upon the binding of ligand, the ribosome binding site (RBS) is unmasked, thereby allowing for ribosome binding and thus protein synthesis. Both systems shown here illustrate an "on" switch; however, analogous "off" switches also exist.

ble of sensing and responding to molecules beyond those currently sensed by natural cells [75].

The majority of the riboswitch research has focused on measuring sensing activity inside of existing cells, as is the case for the protein based systems described above. However, recent work has shown that a previously selected theophylline riboswitch [76, 77] was capable of turning on gene expression *in vitro*, in water-in-oil emulsions, and in vesicles [78]. Further, this cell-like system could be built to sense molecules outside of the vesicle compartment, thereby demonstrating that artificial RNA sensors could serve as a foundation for the construction of cellular mimics that go beyond what natural biology gives us.

LIMITATIONS

Many challenges still remain. The progress made in mimicking some of the features of life is tempered by the fact that the majority of the experiments have been performed under different conditions. Therefore, it is unclear whether the integration of different cell-like functions in their present form into a single system is possible. There are also fundamental problems associated with the use of purified translation machinery. Although reconstituted transcription-translation systems are able to drive cascading, genetically encoded networks [61,79], none of the described systems are capable of supporting long-term activity. We are currently able to engineer *in vitro* systems that turn gene expression on and then off [38,80,81]. However, we are not able to begin another cycle, because we are not able to regenerate the translation machinery *in vitro*. Currently, the minimal translation system is composed of proteins and RNAs that are isolated from *E. coli*. Until we are able to produce functioning ribosomes from *in vitro* transcription-

translation reactions, there will be severe limits on what types of cell-like systems can be manufactured in the laboratory.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We thank the Armenise-Harvard foundation and CIBIO for financial support.

REFERENCES

- [1] Endy, D. Foundations for engineering biology. *Nature*, **2005**, 438, 449-453.
- [2] Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R. Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z. Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A., 3rd; Smith, H. O.; Venter, J. C. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **2010**, 329, 52-56.
- [3] Ellington, A. D.; Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **1990**, 346, 818-822.
- [4] Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C.J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science*, **2011**, 333, 348-353.
- [5] Deamer, D. W. Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature*, **1985**, 317, 792-794.
- [6] Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science*, **1953**, 117, 528-529.
- [7] Oro, J. Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. *Nature*, **1961**, 191, 1193-1194.
- [8] Pizzarello, S. The chemistry of life's origin: a carbonaceous meteorite perspective. *Acc. Chem. Res.*, **2006**, 39, 231-237.

- [9] Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **2009**, 459, 239-242.
- [10] Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. Borate minerals stabilize ribose. *Science*, **2004**, 303, 196.
- [11] Rushdi, A. I.; Simoneit, B. R. Lipid formation by aqueous Fischer-Tropsch synthesis over a temperature range of 100 to 400 degrees C. *Orig. Life Evol. Biosph.*, **2001**, 31, 103-118.
- [12] Deamer, D. W.; Dworkin, J. P. Chemistry and physics of primitive membranes. *Top. Curr. Chem.*, **2005**, 259, 1-27.
- [13] Mansy, S. S.; Szostak, J. W. Reconstructing the Emergence of Cellular Life through the Synthesis of Model Protocells. *Cold Spring Harb. Symp. Quant. Biol.*, **2009**, 74, 47-54.
- [14] Stano, P.; Luisi, P. L. Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chem. Commun. (Camb.)*, **2010**, 46, 3639-3653.
- [15] Crick, F. Central dogma of molecular biology. *Nature*, **1970**, 227, 561-563.
- [16] Posfai, G.; Plunkett, G., 3rd; Feher, T.; Frisch, D.; Keil, G. M.; Umenhoffer, K.; Kolisnychenko, V.; Stahl, B.; Sharma, S. S.; de Arruda, M.; Burland, V.; Harcum, S. W.; Blattner, F. R. Emergent properties of reduced-genome *Escherichia coli*. *Science*, **2006**, 312, 1044-1046.
- [17] Glass, J. I.; Assad-Garcia, N.; Alperovich, N.; Yooseph, S.; Lewis, M. R.; Maruf, M.; Hutchison, C. A., 3rd; Smith, H. O.; Venter, J. C. Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. USA*, **2006**, 103, 425-430.
- [18] Gil, R.; Silva, F. J.; Pereto, J.; Moya, A. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.*, **2004**, 68, 518-537.
- [19] Nakabachi, A.; Yamashita, A.; Toh, H.; Ishikawa, H.; Dunbar, H. E.; Moran, N. A.; Hattori, M. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, **2006**, 314, 267.
- [20] Woese, C. The universal ancestor. *Proc. Natl. Acad. Sci. USA*, **1998**, 95, 6854-6859.
- [21] Morowitz, H. J. Beginnings of cellular life: Metabolism recapitulates biogenesis. *Yale University Press: New Haven*, 1992.
- [22] Chen, I. A.; Roberts, R. W.; Szostak, J. W. The emergence of competition between model protocells. *Science*, **2004**, 305, 1474-1476.
- [23] Szostak, J. W.; Bartel, D. P.; Luisi, P. L. Synthesizing life. *Nature*, **2001**, 409, 387-390.
- [24] Deamer, D. W. The first living systems: a bioenergetic perspective. *Microbiol. Mol. Biol. Rev.*, **1997**, 61, 239-261.
- [25] Chen, I. A.; Szostak, J. W. Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles. *Proc. Natl. Acad. Sci. USA*, **2004**, 101, 7965-7970.
- [26] Segre, D.; Ben-Eli, D.; Deamer, D. W.; Lancet, D. The lipid world. *Orig. Life Evol. Biosph.*, **2001**, 31, 119-145.
- [27] Blochliger, E.; Blocher, M.; Walde, P.; Luisi, P. L. Matrix effect in the size distribution of fatty acid vesicles. *J. Phys. Chem. B*, **1998**, 102, 10383-10390.
- [28] Chen, I. A.; Szostak, J. W. A kinetic study of the growth of fatty acid vesicles. *Biophys. J.*, **2004**, 87, 988-998.
- [29] Hanczyc, M. M.; Fujikawa, S. M.; Szostak, J. W. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science*, **2003**, 302, 618-622.
- [30] Hargreaves, W. R.; Deamer, D. W. Liposomes from ionic, single-chain amphiphiles. *Biochemistry*, **1978**, 17, 3759-3768.
- [31] Mansy, S. S.; Schrum, J. P.; Krishnamurthy, M.; Tobe, S.; Treco, D. A.; Szostak, J. W. Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, **2008**, 454, 122-125.
- [32] Mansy, S. S.; Szostak, J. W. Thermotability of model protocell membranes. *Proc. Natl. Acad. Sci. USA*, **2008**, 105, 13351-13355.
- [33] Walde, P.; Wick, R.; Fresta, M.; Mangone, A.; Luisi, P. L. Autopoietic self-reproduction of fatty acid vesicles. *J. Am. Chem. Soc.*, **1994**, 116, 11649-11654.
- [34] Zhu, T. F.; Szostak, J. W. Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc.*, **2009**, 131, 5705-5713.
- [35] Chen, I. A.; Salehi-Ashtiani, K.; Szostak, J. W. RNA catalysis in model protocell vesicles. *J. Am. Chem. Soc.*, **2005**, 127, 13213-13219.
- [36] Monnard, P. A.; Apel, C. L.; Kanavarioti, A.; Deamer, D. W. Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium. *Astrobiology*, **2002**, 2, 139-152.
- [37] Torino, D.; Del Bianco, C.; Ross, L. A.; Ong, J. L.; Mansy, S. S. Intravesicular isothermal DNA replication. *BMC Res. Notes*, **2011**, 4, 128.
- [38] Noireaux, V.; Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. USA*, **2004**, 101, 17669-17674.
- [39] Monnard, P. A.; Deamer, D. W. Nutrient uptake by protocells: a liposome model system. *Orig. Life Evol. Biosph.*, **2001**, 31, 147-155.
- [40] Monnard, P. A.; Luptak, A.; Deamer, D. W. Models of primitive cellular life: polymerases and templates in liposomes. *Phil. Trans. R. Soc. B*, **2007**, 362, 1741-1750.
- [41] McCutcheon, J. P.; Moran, N. A. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci. USA*, **2007**, 104, 19392-19397.
- [42] Pietrini, A. V.; Luisi, P. L. Cell-free protein synthesis through solubilized exchange in water/oil emulsion compartments. *ChemBiochem*, **2004**, 5, 1055-1062.
- [43] Tawfik, D. S.; Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.*, **1998**, 16, 652-656.
- [44] Bernath, K.; Magdassi, S.; Tawfik, D. S. Directed evolution of protein inhibitors of DNA-nucleases by *in vitro* compartmentalization (IVC) and nanodroplet delivery. *J. Mol. Biol.*, **2005**, 345, 1015-1026.
- [45] Mansy, S. S. Membrane transport in primitive cells. In *The Origins of Life, Cold Spring Harb. Perspect. Biol.*, **2010**, 2, 193-206.
- [46] Kurama, H.; Stano, P.; Ueda, T.; Luisi, P. L. A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta*, **2009**, 1788, 567-574.
- [47] Baumgart, T.; Hess, S. T.; Webb, W. W. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature*, **2003**, 425, 821-824.
- [48] Chen, I. A. Cell division: breaking up is easy to do. *Curr. Biol.*, **2009**, 19, R237-R238.
- [49] Leaver, M.; Dominguez-Cuevas, P.; Coxhead, J. M.; Daniel, R. A.; Errington, J. Life without a wall or division machine in *Bacillus subtilis*. *Nature*, **2009**, 457, 849-853.
- [50] Margolin, W. Themes and variations in prokaryotic cell division. *FEMS Microbiol. Rev.*, **2000**, 24, 531-548.
- [51] Loose, M.; Fischer-Friedrich, E.; Ries, J.; Kruse, K.; Schwillke, P. Spatial regulators for bacterial cell division self-organize into surface waves *in vitro*. *Science*, **2008**, 320, 789-792.
- [52] Osawa, M.; Anderson, D. E.; Erickson, H. P. Reconstitution of contractile FtsZ rings in liposomes. *Science*, **2008**, 320, 792-794.
- [53] Makeyev, E. V.; Grimes, J. M. RNA-dependent RNA polymerases of dsRNA bacteriophages. *Virus Res.*, **2004**, 101, 45-55.
- [54] Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science*, **2011**, 332, 209-12.
- [55] Ichihashi, N.; Matsuura, T.; Kita, H.; Hosoda, K.; Sunami, T.; Tsukada, K.; Yomo, T. Importance of translation-replication balance for efficient replication by the self-encoded replicase. *ChemBiochem*, **2008**, 9, 3023-3028.
- [56] Oberholzer, T.; Albrizio, M.; Luisi, P. L. Polymerase chain reaction in liposomes. *Chem. Biol.*, **1995**, 2, 677-682.
- [57] Li, Y.; Kim, H. J.; Zheng, C.; Chow, W. H.; Lim, J.; Keenan, B.; Pan, X.; Lemieux, B.; Kong, H. Primase-based whole genome amplification. *Nucleic Acids Res.*, **2008**, 36 (13), e79.
- [58] Salas, M. 40 years with bacteriophage phi29. *Annu. Rev. Microbiol.*, **2007**, 61, 1-22.
- [59] Mencia, 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. USA*, **2011**, 108, 18655-18660.
- [60] Forster, A. C.; Church, G. M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2006**, 2, 45.
- [61] Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.*, **2001**, 19, 751-755.
- [62] Zheng, Y.; Roberts, R. J. Selection of restriction endonucleases using artificial cells. *Nucleic Acids Res.*, **2007**, 35 (11), e83.
- [63] Stano, P.; Kuruma, Y.; Souza, T. P.; Luisi, P. L. Biosynthesis of proteins inside liposomes. *Methods Mol. Biol.*, **2010**, 606, 127-145.
- [64] Sunami, T.; Matsuura, T.; Suzuki, H.; Yomo, T. Synthesis of functional proteins within liposomes. *Methods Mol. Biol.*, **2010**, 607, 243-256.
- [65] Pereira de Souza, T.; Stano, P.; Luisi, P. L. The minimal size of liposome based model cells brings about a remarkably enhanced entrapment and protein synthesis. *ChemBiochem*, **2009**, 10, 1056-1063.
- [66] Pereira de Souza, T.; Steiniger, F.; Stano, P.; Fabr, A.; Luisi, P. L. Spontaneous crowding of ribosomes and proteins inside vesicles: a possible mechanism for the origin of cell metabolism. *ChemBiochem*, **2011**, 12(15), 2325-2330.
- [67] Dufour, Y. S.; Kiley, P. J.; Donohue, T. J. Reconstruction of the core and extended regulons of global transcription factors. *PLoS Genet.*, **2010**, 6 (7), e1001027.
- [68] Fleischhacker, A. S.; Kiley, P. J. Iron-containing transcription factors and their roles as sensors. *Curr. Opin. Chem. Biol.*, **2011**, 15, 335-341.
- [69] Stock, A. M.; Robinson, V. L.; Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.*, **2000**, 69, 183-215.
- [70] Moker, N.; Kramer, J.; Uden, G.; Kramer, R.; Morbach, S. *In vitro* analysis of the two-component system MtrB-MtrA from *Corynebacterium glutamicum*. *J. Bacteriol.*, **2007**, 189, 3645-3649.
- [71] Seelig, B.; Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature*, **2007**, 448, 828-831.
- [72] Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*, **2010**, 329, 309-313.
- [73] Winkler, W. C.; Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.*, **2005**, 59, 487-517.
- [74] Lorsch, J. R.; Szostak, J. W. Chance and necessity in the selection of nucleic acid catalysts. *Acc. Chem. Res.*, **1996**, 29, 103-110.
- [75] Topp, S.; Gallivan, J. P. Emerging applications of riboswitches in chemical biology. *ACS Chem. Biol.*, **2010**, 5, 139-148.

- [77] Topp, S.; Gallivan, J. P. Guiding bacteria with small molecules and RNA. *J. Am. Chem. Soc.*, **2007**, 129, 6807-6811.
- [78] Martini, L.; Mansy, S. S. Cell-like systems with riboswitch controlled gene expression. *Chem. Commun. (Camb.)*, **2011**, 47, 10734-10736.
- [79] Ishikawa, K.; Sato, K.; Shima, Y.; Urabe, I.; Yomo, T. Expression of a cascading genetic network within liposomes. *FEBS Lett.*, **2004**, 576, 387-390.
- [80] Noireaux, V.; Bar-Ziv, R.; Libchaber, A. Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. USA*, **2003**, 100, 12672-12677.
- [81] Shin, J.; Noireaux, V. Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. *J. Biol. Eng.*, **2010**, 4, 9.

Received: September 29, 2011

Revised: May 07, 2013

Accepted: May 12, 2013

Chapter 2.

Cellular mimics sense and respond to external stimuli via riboswitches

This work has been adapted from:

1. *Cell-like systems with riboswitch controlled gene expression*

Laura Martini and Sheref S. Mansy; *Chemical Communications*, **2011**, 47, 10734–10736.

2. *Measuring Riboswitch Activity In vitro and in Artificial Cells with Purified Transcription–Translation Machinery*

Laura Martini and Sheref S. Mansy; in Atsushi Ogawa (ed.), *Artificial Riboswitches: Methods and Protocols, Methods in Molecular Biology*, **2014**, vol. 1111: 153-64.

3. *Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour*

Roberta Lentini, Silvia Perez Santero, Fabio Chizzolini, Dario Cecchi, Jason Fontana, Marta Marchioretto, Cristina Del Bianco, Jessica L. Terrell, Amy C. Spencer, **Laura Martini**, Michele Forlin, Michael Assfalg, Mauro Dalla Serra, William E. Bentley & Sheref S. Mansy; *Nature Communications*, **2014**, 5: 4012.

Attached the reprints of the papers*.

* Reprinted with the permission of the copyright holders

1. *Royal Society of Chemistry*

2. *Springer Science + Business Media (License Number 3487701102975)*

3. *Nature Publishing Group & Palgrave Macmillan*

In this chapter the engineering of cellular mimics to sense and respond to an external stimuli is presented. The need for sense-response activity has already been introduced. A riboswitch responsive to the small molecule theophylline is exploited as sensing element. The riboswitch controls translation of a reporter gene. In the presence of theophylline, gene expression is permitted and using a cell-free transcription and translation machinery a fluorescent protein is produced.

Initially, *in vitro* characterization of the riboswitch behaviour is performed monitoring in real-time the fluorescent protein production activated by the presence of theophylline. Next, the riboswitch system is encapsulated in water-in-oil emulsions and vesicles. When the ligand is present in the environment, the information is internally processed through the RNA sensor device. The fluorescent response demonstrates that the cell mimics are able to react to the signal arising from outside of the compartment. The last part of the chapter presents a recent work in which the riboswitch response is translated into a chemical message for *E. coli*.

2.1 Riboswitches as small ligand sensor devices

Riboswitches are genetically encoded riboregulators dependent on the direct binding of small molecules.^{43,50} Riboswitches are encoded naturally in the 5' UTR of mRNA of metabolite genes, thus controlling gene expression upon metabolite binding. Riboswitch regulation occurs at the transcriptional or translational level via different mechanisms. For example, riboswitches can modulate the activity of a transcription termination site by disrupting interactions between a terminator and an anti-terminator stem loop.⁵¹ As a result, a truncated transcript is produced that is incapable of supporting protein expression. Alternatively, riboswitches can regulate directly protein production by modulating the accessibility of the ribosome binding site via sequence sequestration or steric hindrance.⁵²

A third mechanistic variation is the modulation of RNA splicing.⁵³ Many natural riboswitches control transcription, while synthetic riboswitches generally control translation.

Riboswitch activity is due to the conformation that the RNA acquires after the binding of a specific ligand to the aptamer domain. Thus, every riboswitch is structurally predisposed to sense a specific molecule. This makes riboswitches very powerful and selective sensors, detecting the presence of target molecules with high affinities.⁴⁸ From their discovery in 2002,⁵⁴ many naturally occurring riboswitches have been identified and characterized.⁵³ However, even synthetic riboswitches have been developed.⁵⁵ Rational design and *in vivo* selections have expanded the possibility of sensing different ligands. The engineering and evolution of synthetic riboswitches will be investigated in the next chapter.

The theophylline riboswitch is a synthetic regulator developed via *in vivo* selection starting from the theophylline aptamer TCT8-4^{48,55} (Fig. 2.1). The dissociation constant (K_d) is 400 nM,⁵⁶ and the riboswitch demonstrates a very high ability to distinguish an analog of theophylline (caffeine). The proposed mechanism of action is based on RBS helix disruption. Conversely than before, a cotranscriptional kinetic trap mechanism has been recently proposed to mediate the switching.⁵⁶ In other words, the equilibrium between the bound and unbound states of the riboswitch is not thermodynamically regulated, but depends on the kinetics of folding while the RNA is transcribed. Nevertheless, due to the high specificity for the ligand and *in vivo* reported activity, the theophylline riboswitch is extensively used. The theophylline aptamer is exploited for aptazymes^{57,58} and antiswitches,⁴⁴ resulting in many literature reports. For this reason the theophylline responsive riboswitch is initially investigated and the activity *in vitro* is characterized

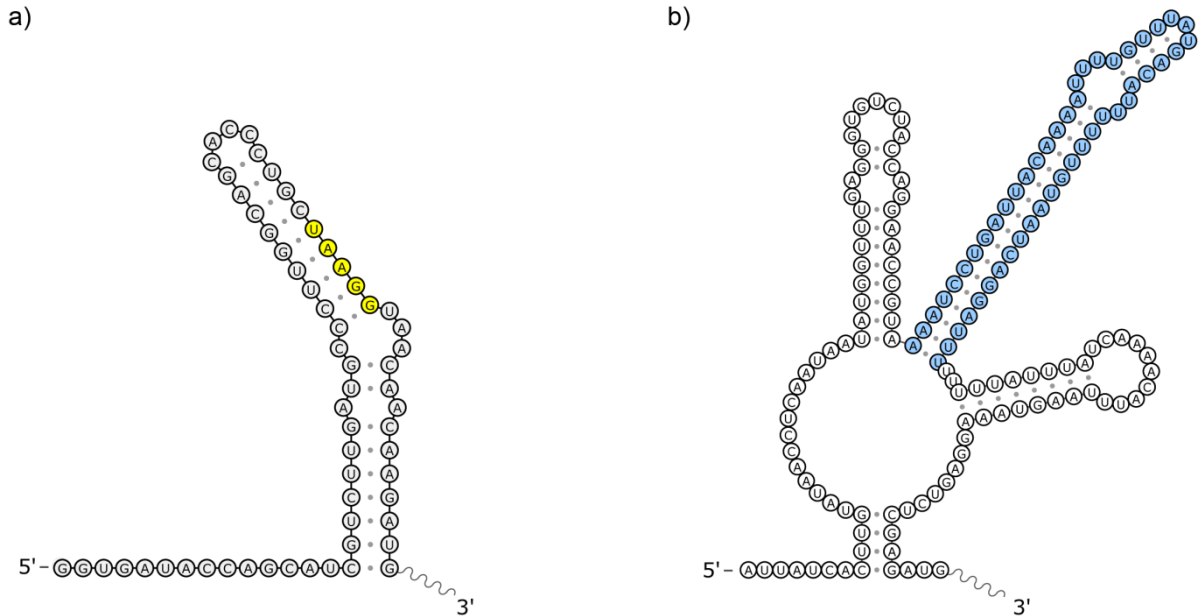


Fig. 2.1. Theophylline-responsive riboswitch (a) and adenine-responsive riboswitch (b) structures in the OFF state. Regions of the riboswitch thought to be functionally important are highlighted. The theophylline riboswitch regulates translation via modulation of accessibility of the RBS (yellow). The adenine riboswitch presents a terminator stem loop that regulates transcription (cyan). Structures are based on Vienna RNA fold webserver calculations.

For comparison with the synthetic riboswitch behaviour, the *in vitro* activity of a natural transcriptional control regulator is also assayed. The transcriptional control riboswitch is responsive to adenine and was identified in the 5' UTR of the *ydhL* mRNA coding for a purine efflux pump in *B. subtilis*⁵⁹ (Fig. 2.1). The adenine-riboswitch is one of the few examples of a natural ON regulator. In other words, when the ligand is present in solution, the conformational change induced in the RNA structure causes activation of gene expression. Conversely, in absence of the molecule, the riboswitch remains in an OFF state, inhibiting protein production. Considering the adenine riboswitch, the modulation is mediated through disruption of a terminator stem. Thus, the *ydhL* riboswitch is a transcriptional regulator. Nevertheless, *ydhL* (also called *pbuE*) adenine-responsive riboswitch is not the only RNA controller identified. The adenine sensor from the *add* gene of *V. vulnificus* is a translational control riboswitch.⁵¹ However, the investigation is focused

on the transcriptional control riboswitch in order to also explore natural riboswitches in a cell-free environment.

Riboswitch characterization is usually done *in vivo* by monitoring the expression of a reporter gene. β -galactosidase, fluorescent proteins⁶⁰ or cell motility⁶¹ genes have been exploited in bacteria. Another characterization approach is based on the detection of structure modifications induced by the ligand binding. The conformational shift is analyzed by investigation of altered RNA patterns after alkaline digestion.⁵⁹ However, these characterizations are all indirect means of monitoring riboswitch control. In other words, the ligand induced conformation change is not monitored in real time.

In the first two publications is presented an *in vitro* assay for riboswitch activity in a cell-free environment. Using the PURE system, theophylline and adenine-responsive riboswitches are directly monitored assessing protein production in the presence and absence of the ligand. The use of a yellow fluorescent protein (A206K YPet) as an *in vitro* reporter allows for real time detection. In the presence of the cognate ligand the riboswitch configuration is in an active state, allowing protein production, thus the fluorescence increases. Conversely in absence of the ligand, little or no protein is synthesized. As expected, the theophylline riboswitch is able to control YPet production in a PURE system reaction. On the other end, the activity of the adenine-responsive riboswitch reveals a dependence on the RNA polymerase (*Fig.2, Martini L., Mansy S. S, Chem. Comm., 2011, 47, 10734–10733*). Moreover, additional protein factors are likely required for transcriptional riboswitch regulation.⁵¹ Thus, in order to build a minimal, defined system, the focus is set only on the theophylline-responsive riboswitch.

The previously characterized design is exploited for the construction of cellular mimics responsive to theophylline. The PURE system reaction and the DNA encoding the theophylline riboswitch are encapsulated in water-in-oil emulsions and in vesicles. Fluorescence is achieved in the presence of the small molecule ligand in both the

compartments. Moreover, when the ligand is added to the external solution, the cellular mimic is able to sense from the environment the ligand, and through activation of the riboswitch, the fluorescent output is produced (*Fig.3, Martini L., Mansy S. S, Chem. Comm., 2011, 47, 10734–10733*). Results obtained can facilitate the characterization of riboswitch behaviour in real time. Also, the study can be a model for the engineering of new cell-like systems able to sense and respond to different small ligands present in the environment.

My contribution to these works concerns the design of the genetic construct, the setting and collecting of all the experimental measurements, manuscript writing and editing. The materials and methods, the full results and a detailed discussion are described in the published method chapter *Measuring Riboswitch Activity In vitro and in Artificial Cells with Purified Transcription–Translation Machinery* (Martini L., Mansy S. S, *Methods Mol. Biol.*, 2014, 1111, 153) and the paper *Cell-like system with riboswitch controlled gene expression* (Martini L., Mansy S. S, *Chem. Comm.*, 2011, 47, 10734–10733).

Cite this: *Chem. Commun.*, 2011, **47**, 10734–10736

www.rsc.org/chemcomm

COMMUNICATION

Cell-like systems with riboswitch controlled gene expression†

Laura Martini and Sheref S. Mansy*

Received 1st July 2011, Accepted 10th August 2011

DOI: 10.1039/c1cc13930d

Synthetic riboswitches can be used to control protein expression under fully defined conditions *in vitro*, in water-in-oil emulsions, and in vesicles. The developed system could serve as a foundation for the construction of cellular mimics that respond to molecules of our choosing.

There are many possible approaches towards synthesizing cells, including the modification of existing microorganisms¹ and the bottom-up assembly of cell-like systems from minimal components.² Engineering within existing organisms has the benefit of providing all of the needed ingredients of life without requiring extensive knowledge of the component parts. However, this approach thus far has failed to yield a fully describable biological system. Further, the use of the existing cell architecture may severely limit the types of life-like systems achievable. Although building a cell from component parts is challenging, success would give a much simplified, potentially fully understood life-like system that is easier to model and to engineer. The resulting system would not need to mimic every aspect of contemporary cells.³ Bottom-up constructions that have identity and sense and respond to their environment would possess several features of life without having the ability to reproduce (Fig. 1). Such a system would be attractive for basic research and technological applications, since the constructed system would not be capable of evolving, *i.e.* the fear of losing control over the system would be removed.

The biological parts needed to assemble a cell-like system that responds to external stimuli include a genome, transcription-translation machinery, a sensor, and a compartment. The use of DNA and lipids to direct transcription-translation and to define the compartments, respectively, is well established. Similarly, there are several simple options for transcription, including phage, such as T7, and bacterial, such as *E. coli*, RNA polymerases. Of these two, T7 RNA polymerase is simpler since it consists of a single protein domain and does not necessitate a sigma factor. Despite the complexity of translation, 31 *E. coli* proteins can be individually purified and assembled with RNA to provide an active, minimal

protein synthesis system (PURE system) that functions *in vitro*,⁴ in water-in-oil (w/o) emulsions,⁵ and in vesicles.⁶

The challenging component of the described bottom-up cell-like system is the sensor. Gene expression can be controlled by naturally occurring proteins that directly or indirectly influence the activity of transcription factors and translational attenuators. However, the use of such protein-based mechanisms generally requires the exploitation of natural proteins, thus limiting the types of molecules that can be sensed. Since protein engineering and design remains a difficult task, other methods are desirable.

RNA molecules provide an attractive alternative, since RNAs can be evolved and engineered in the laboratory to perform a range of functions, including ligand binding, catalysis, and the control of gene expression.⁷ One class of regulatory RNAs, known as riboswitches, is located within the untranslated regions of mRNA. In bacteria riboswitches can up or down regulate gene expression in response to ligand binding directly to the mRNA.⁸ Control can either be at the transcriptional or translational level, either switching between truncated and full-length transcripts (transcriptional control) or between different mRNA conformations that control the accessibility of a ribosome binding site (translational control).

To establish if riboswitches can function as robust on-off switches *in vitro*, in w/o emulsions, and in vesicles with fully defined components, we tested two different previously characterized riboswitches that up-regulate gene expression in response to ligand binding. The two riboswitches were a synthetic theophylline responsive riboswitch selected by Lynch and Gallivan⁹ that functions through a translational control

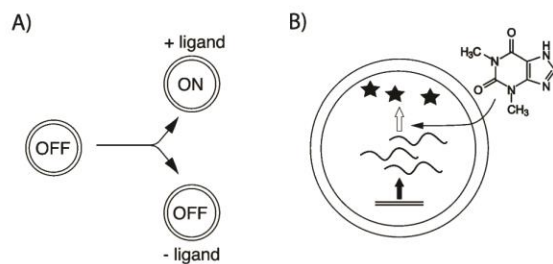


Fig. 1 Compartmentalized, cell-like systems that sense and respond to their environments through riboswitch activity. (A) The presence of an extravesicular ligand converts the cell-like system from the OFF-state to the ON-state. (B) RNA (squiggly line) is transcribed from DNA (double line). RNA is only translated into protein (star) in the presence of the activator ligand, which in this case is theophylline.

CIBIO, University of Trento, via delle Regole 101, 38100 Mattarello (TN), Italy. E-mail: mansy@science.unitn.it;

Fax: +39 0461-883091; Tel: +39 0461-883438

† Electronic supplementary information (ESI) available: Control reactions of theophylline riboswitch activity in w/o emulsions and vesicles, a control reaction showing YPet w/o emulsion expression, and experimental details. See DOI: 10.1039/c1cc13930d

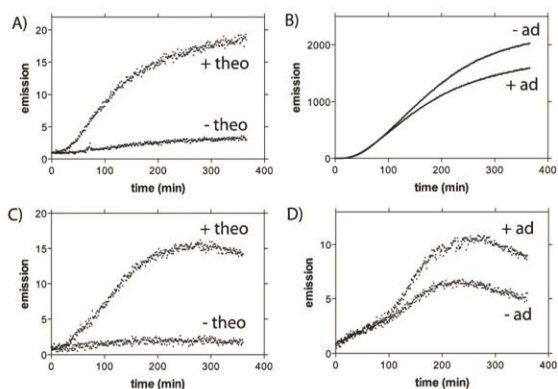


Fig. 2 *In vitro* riboswitch activity. Reactions were either carried out at 37 °C with the PURE system (A, B) or with an *E. coli* S30 system (C, D). (A, C) Theophylline riboswitch activity in the presence (+ theo) and absence (– theo) of theophylline. (B, D) Adenine riboswitch activity in the presence (+ ad) and absence (– ad) of adenine. The production of protein was monitored by fluorescence spectroscopy.

mechanism and a natural *B. subtilis* adenine-sensing riboswitch¹⁰ that functions through a transcriptional control mechanism.

Both riboswitches were inserted into the 5'-untranslated region of sequences encoding the yellow fluorescent protein YPet.¹¹ The genetic constructs also included a T7 transcriptional promoter and an *E. coli* ribosome binding site (see ESI† for more details). Riboswitch activity with minimal transcription and translation components was monitored in real time by fluorescence spectroscopy. As shown in Fig. 2A, little protein expression occurred in the absence of theophylline for the synthetic riboswitch system, whereas protein levels were 6-fold higher after 6 h in the presence of theophylline. Conversely, the adenine-sensing riboswitch did not adequately control gene expression. Significant expression occurred in the absence of adenine, and the system failed to show induction upon adenine addition (Fig. 2B). Since the ribosome binding site within the theophylline riboswitch was weaker than the ribosome binding site used with the adenine riboswitch, the overall YPet yield was significantly lower for the theophylline riboswitch system.

To determine if the adenine-sensing riboswitch failed to show switching activity due to features of the T7 RNA polymerase, we repeated the *in vitro* assay using an *E. coli* *tac* promoter and an *E. coli* cell lysate that contained *E. coli* RNA polymerase. Under these conditions, both the theophylline and the adenine riboswitches showed activity, exhibiting 8-fold and 1.7-fold increases in YPet protein levels, respectively (Fig. 2C and D). However, the adenine riboswitch continued to show significant gene expression in the absence of a ligand. For riboswitches that exploit terminator–antiterminator RNA structures to modulate transcription elongation, some specificity for the structural features and the catalytic rate of the RNA polymerase is expected.¹² For example, transcription of the adenine riboswitch sequence gives different mRNA product ratios depending on whether transcription is mediated by *B. subtilis*, *E. coli*, or T7 RNA polymerase, with the use of T7 RNA polymerase resulting in no detectable riboswitch

activity.^{12a} Conversely, riboswitches that control gene expression at the translational level produce full-length transcripts regardless of the presence or absence of ligands and thus would not be expected to show a dependence on the RNA polymerase. Further, transcriptional control riboswitches often require additional accessory protein factors for efficient signalling.^{12a,13} Since we endeavoured to build the simplest system possible, we focused our remaining efforts on the theophylline riboswitch.

Having established that the theophylline riboswitch displays robust activity in a fully defined, minimal *in vitro* system, we next sought to test its activity within w/o emulsion compartments. The water droplets of a w/o emulsion mimic several features of cellular compartments, providing a space similar to the dimensions of bacterial cells that is capable of coupling genotype with phenotype.¹⁴ We encapsulated the minimal theophylline riboswitch system within w/o emulsion droplets stabilized with a mixture of non-ionic surfactants (span-80, Tween-80, and triton X-100). The average water droplet diameter was 10 μm. Robust theophylline riboswitch activity was observed within the confines of the w/o emulsion droplets. No fluorescence was detected by microscopy in the absence of theophylline (Fig. S1, ESI†), while fluorescence was detected in the presence of theophylline (Fig. 3A and B). However, fluorescence in the presence of theophylline was generally low due to the weak ribosome binding site within the theophylline riboswitch. Since larger droplets can hold greater quantities of reactants, large water droplets showed detectable fluorescence. Conversely, reactions using a strong ribosome binding site

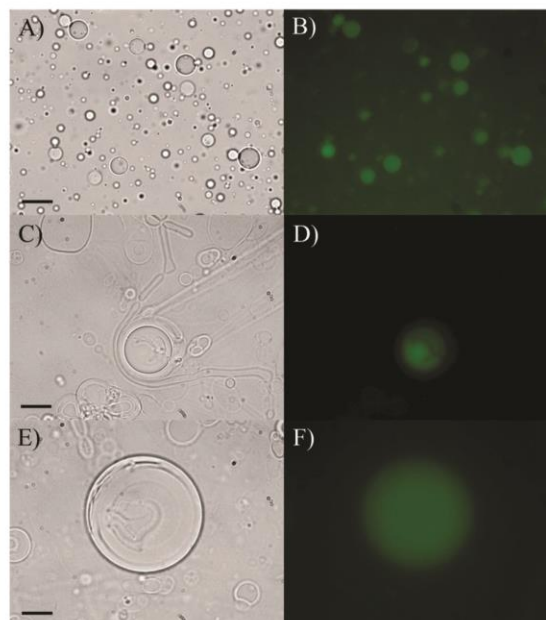


Fig. 3 Compartmentalized theophylline riboswitch activity after 6 hours at 37 °C. Reactions were carried out in w/o emulsions (A, B) or vesicles (C–F). Compartments shown in panels (A–D) contained theophylline from the beginning of the reaction. In panels E, F theophylline was added to the exterior of the vesicles. Panels A, C, E are bright field and panels B, D, F are epifluorescence images. The black bar in the lower left corner of panels A, C, E represents 10 μm.

showed a less pronounced dependence on the compartment size (Fig. S2, ESI†).

Next, we reconstituted the minimal theophylline riboswitch system inside of phospholipid vesicles. Vesicles are better able to mimic cellular conditions by allowing for the selective exchange of chemicals between the intra- and extra-vesicular environments.¹⁵ To ensure that the vesicle encapsulation process did not degrade riboswitch activity, we tested the activity of the theophylline riboswitch within vesicles composed of a mixture of diacyl phospholipids and cholesterol. Encapsulation was achieved by hydrating lyophilized vesicles⁶ with the minimal theophylline riboswitch system in the presence and absence of theophylline. The encapsulation procedure was conducted on ice in order to inhibit premature protein synthesis. Extravesicular reactions were suppressed by proteolytic digestion. After incubation at 37 °C to allow for transcription and translation, YPet production was observed by fluorescence microscopy in vesicles containing theophylline (Fig. 3C and D) but not in vesicles lacking theophylline (Fig. S3, ESI†). Therefore, the minimal transcription–translation system was capable of supporting riboswitch activity within phospholipid vesicles.

Living cells must be capable of responding to their surroundings in order to adapt to changing environmental conditions. To test whether a riboswitch could impart environmental response activity on a vesicle that contained minimal transcription–translation machinery, we assembled vesicles with the minimal riboswitch system as described above. However, theophylline was not included. Subsequently, theophylline was added extravesicularly, and the response of the system was monitored by fluorescence microscopy. As shown in Fig. 3E and F, fluorescence was only observed within the vesicles after the addition of theophylline. Since theophylline is membrane permeable, the generation of fluorescence was consistent with theophylline permeating across the membrane into the internal space of the vesicle and activating gene expression by directly binding to the theophylline riboswitch. In other words, this fully defined system was capable of sensing and responding to its surroundings.

Here we described a cell-like system built from known parts and with minimal components. The incorporation of a genetic construct that encoded a synthetic riboswitch was capable of converting an encapsulated protein synthesis system into a cellular mimic capable of sensing and responding to its environment. Considering the ability of RNA selection technologies to generate aptamers capable of sensing molecules of our choosing, we feel that cell-like systems similar to what was described herein could be built to sense and respond to a wide variety of biological and abiological molecules.

In addition to bottom-up synthetic biology applications, we believe that our *in vitro* fluorescence assay will facilitate riboswitch research. Most riboswitch studies exploit *in vivo* activity assays or use *in vitro* assays to probe RNA conformational changes or riboswitch activity at the RNA transcript level. While *in vivo* experiments are important, it can be difficult to decipher which cellular factors are required for

riboswitch activity from *in vivo* data alone. The *in vitro* mRNA transcript level assays do not share this complication; however, since the ultimate function of a riboswitch is to control protein levels, it is important to have an assay that measures protein production. A few laboratories have monitored riboswitch activity through *in vitro* translation assays. A limitation of these current approaches is the use of ill-defined cell extracts to mediate translation, thus introducing some of the same problems associated with *in vivo* assays, *i.e.* the use of ill-defined mixtures does not allow for the identification of the required components for activity. Further, protein level riboswitch activity assays to date exploit radioactive labelling^{12a} thereby making the collection of high resolution kinetic data difficult.

The use of fully defined transcription–translation machinery allows for the systematic investigation of the required cellular factors for riboswitch function. Additionally, the exploitation of YPet, one of the brightest fluorescent proteins,¹¹ as a signal output allows for facile, direct, real-time measurements that should permit deeper kinetic and mechanistic investigation. Our data from this assay show that a synthetic theophylline riboswitch requires no additional components for activity beyond a T7 RNA polymerase and the minimal *E. coli* translation machinery.

This work was supported by the Armenise-Harvard foundation and CIBIO.

Notes and references

- 1 D. Endy, *Nature*, 2005, **438**, 449–453.
- 2 J. W. Szostak, D. P. Bartel and P. L. Luisi, *Nature*, 2001, **409**, 387–390.
- 3 L. Cronin, N. Krasnogor, B. G. Davis, C. Alexander, N. Robertson, J. H. Steinke, S. L. Schroeder, A. N. Khlobystov, G. Cooper, P. M. Gardner, P. Siepmann, B. J. Whitaker and D. Marsh, *Nat. Biotechnol.*, 2006, **24**, 1203–1206.
- 4 Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa and T. Ueda, *Nat. Biotechnol.*, 2001, **19**, 751–755.
- 5 Y. Zheng and R. J. Roberts, *Nucleic Acids Res.*, 2007, **35**, e83.
- 6 T. Sunami, T. Matsuura, H. Suzuki and T. Yomo, *Methods Mol. Biol.*, 2010, **607**, 243–256.
- 7 (a) J. E. Weigand and B. Suess, *Appl. Microbiol. Biotechnol.*, 2009, **85**, 229–236; (b) M. Wieland and J. S. Hartig, *ChemBioChem*, 2008, **9**, 1873–1878.
- 8 (a) F. J. Grundy and T. M. Henkin, *Crit. Rev. Biochem. Mol. Biol.*, 2006, **41**, 329–338; (b) A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov and E. Nudler, *Cell*, 2002, **111**, 747–756; (c) W. C. Winkler and R. R. Breaker, *Annu. Rev. Microbiol.*, 2005, **59**, 487–517.
- 9 S. A. Lynch and J. P. Gallivan, *Nucleic Acids Res.*, 2009, **37**, 184–192.
- 10 M. Mandal and R. R. Breaker, *Nat. Struct. Mol. Biol.*, 2004, **11**, 29–35.
- 11 A. W. Nguyen and P. S. Daugherty, *Nat. Biotechnol.*, 2005, **23**, 355–360.
- 12 (a) J. F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Masse and D. A. Lafontaine, *PLoS Genet.*, 2011, **7**, e1001278; (b) J. K. Wickiser, W. C. Winkler, R. R. Breaker and D. M. Crothers, *Mol. Cell*, 2005, **18**, 49–60.
- 13 I. Irnov and W. C. Winkler, *Mol. Microbiol.*, 2010, **76**, 559–575.
- 14 D. S. Tawfik and A. D. Griffiths, *Nat. Biotechnol.*, 1998, **16**, 652–656.
- 15 S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobe, D. A. Treco and J. W. Szostak, *Nature*, 2008, **454**, 122–125.

Supporting Information Available

Cell-like Systems with Riboswitch Controlled Gene Expression

Laura Martini and Sheref S. Mansy*

CIBIO, University of Trento, via delle Regole 101, 38100 Mattarello (TN), Italy

CONTENTS:

Figure S1. Control emulsion reaction.	3
Figure S2. YPet emulsion expression.	4
Figure S3. Control vesicle reaction.	5
Methods.	6

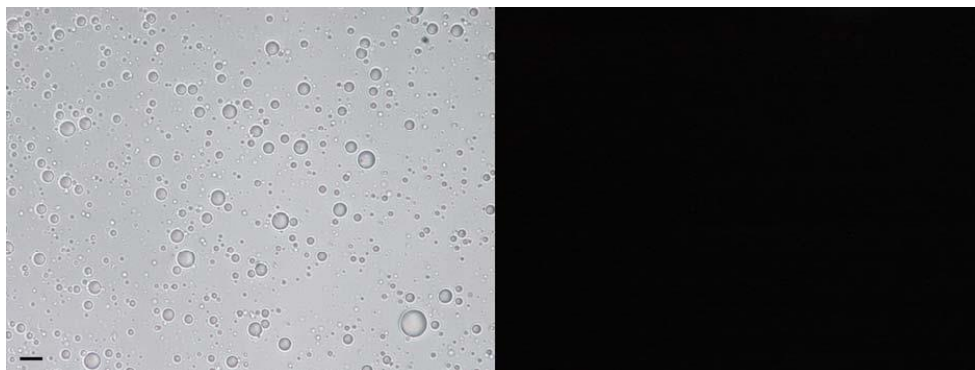


Figure S1. Control emulsion reaction. The theophylline riboswitch was encapsulated in w/o emulsion droplets with transcription-translation machinery provided by the PURE system. No theophylline was present. Bright field (left) and epifluorescence (right) images are shown. The black bar in the lower left corner of the bright field image (left) represents 10 μm .

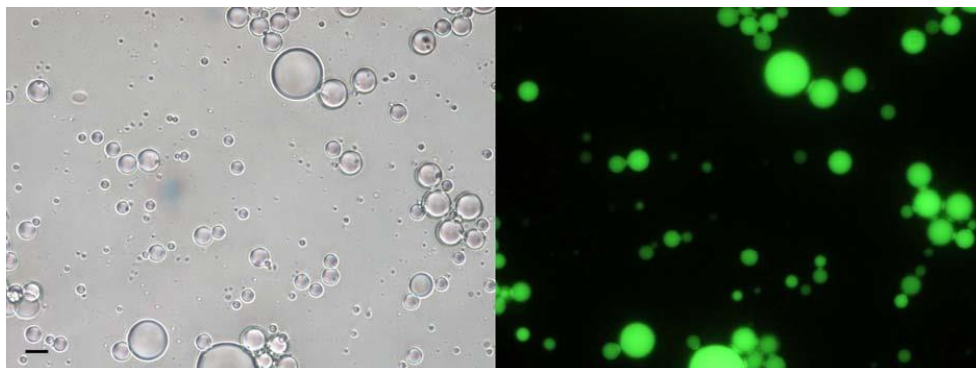


Figure S2. YPet emulsion expression. Robust expression of YPet with the PURE system inside of w/o emulsion droplets was achieved with a strong ribosome binding site. Here the strong ribosome binding site was provided by the adenine riboswitch construct. Bright field (left) and epifluorescence (right) images are shown. The black bar in the lower left corner of the bright field image (left) represents 10 μm .



Figure S3. Control vesicle reaction. The theophylline riboswitch was encapsulated within phospholipid vesicles with transcription-translation machinery provided by the PURE system. No theophylline was present. Bright field (left) and epifluorescence (right) images are shown.

Methods

Materials. All genetic constructs were synthesized by Genscript. Unless otherwise indicated, all chemicals and plasmid purification kits were from Sigma-Aldrich. The *E. coli* cell lysate transcription-translation system (S30 T7 High Yield Protein Expression System) was from Promega, and the PURE system (PURExpress *in vitro* protein synthesis kit) was purchased from New England BioLabs. Lipids were either from Genzyme (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC)), Sigma-Aldrich (cholesterol), or NOF-Europe (N-(carbonyl-methoxypolyethyleneglycol 5000)-1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine (DSPE-PEG 5000)).

***In vitro* transcription-translation with the PURE system.** The reactions were assembled on ice. All DNA templates were phenol-chloroform extracted and ethanol precipitated prior to use. 17.5 μL of the kit components were mixed with 250 ng DNA, 20 units RiboLock RNase Inhibitor (Fermentas), and RNase free Milli-Q water in a final volume of 25.5 μL . If adenine or theophylline was included, they were added to 1 μM and 500 μM , respectively. The entire reaction volume was then placed in a quartz micro cell cuvette (Hellma). Fluorescence measurements were with a PTI QuantaMaster 40 UV Vis Spectrofluorometer and the temperature was held at 37 $^{\circ}\text{C}$ with a Peltier temperature controlled single sample holder. Excitation and emission were at 517 nm and 530 nm, respectively. Emission was recorded every minute for 6 hours. Data were subsequently normalized by setting the initial emission to one. At least three sets of data were acquired for each riboswitch construct and averaged. The standard deviation for each data point was less than 40%.

***In vitro* transcription-translation with an *E. coli* S30 lysate.** The reactions were assembled directly in Nunc 96 well plates on ice. All DNA templates were phenol-chloroform extracted and ethanol precipitated prior to use. 38 μL of Promega S30 kit components (20 μL of premix factors plus 18 μL of *E. coli* S30 lysate), 40 units RiboLock RNase Inhibitor (Fermentas), 10 nM plasmid DNA, 8 mM magnesium acetate were mixed together, and RNase free Milli-Q water was added to bring to a final volume of 50 μL . If adenine or theophylline was included, they were added to 1 μM and 500 μM , respectively. Fluorescence emission was collected with a Tecan infinite 200 multiplate reader at 37 $^{\circ}\text{C}$. The excitation wavelength was 510 nm, and the emission wavelength was 540 nm. Data were acquired every minute for 6 hours. The data were normalized as described above for "*In vitro* transcription-translation with the PURE system."

Water-in-oil emulsion reactions. Water-in-oil emulsions were prepared essentially as described by Davidson et al.¹ Briefly, 475 μL mineral oil, 22.5 μL Span-80, 2.5 μL Tween-80, and 0.25 μL Triton X-100 were stirred with a 9 mm Teflon stir bar in a 50 mL Falcon conical tube. 500 μL of this oil solution was then cooled on ice. Subsequently, 25.5 μL of chilled, aqueous solution containing all of the needed components for transcription-translation was added to 500 μL of the chilled oil phase and stirred for 3 min in an ice water bath. The aqueous phase contained all of the PURE system components plus 500 ng DNA template. If adenine or theophylline was included, the ligand was present at 1 μM and 5 mM, respectively. This concentration of theophylline was necessary due to the partitioning of theophylline to the oil phase. Reactions were initiated by incubating aliquots in

Theophylline riboswitch 12.1³ with a *tac* promoter and A206K YPet

GTTGACAATTAATCATCGGCTCGTATAATGTGTGGCCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCC
TGCTAAGGTAACAACAAGATGGTGTCCAAAGGCGAAGAAGTGTTCACCGGTGTGGTCCGATTCTGGTGGAACTGG
ATGGCGACGTTAACGGTCATAAAATTTAGTGTGTCCGGCGAAGGTGAAGGCGATGCGACCTATGGCAAACTGACGCT
GAAACTGCTGTGCACCACCGGTAAGTGTCCGGTCCCGTGGCCGACCTGGTGACCACGCTGGGTTATGGCGTGCAG
TGTTTCGGCGCTACCCGGACCACATGAAACAACACGATTTCTTTAAAGTGCCATGCCGGAAGGCTATGTTCCAGG
AACGTACCATCTTTTCAAAGATGACGGTAACTACAAAACCCGCGGGAAGTTAAATTTGAAGCGGATACGCTGGT
CAACCGTATGAACTGAAAGGTATCGACTTCAAAGAAGATGGCAATATTTCTGGGTCATAAACTGGAATATACTAC
AATAGCCACAACGTGTATATTTACCGCGGATAAACAGAAAAACGGCATCAAAGCCAACCTTCAAAATCCGCCATAACA
TCGAAGACGGCGGTGTCAACTGGCCGATCACTACCAGCAAAACACCCCGATTGGTGTATGGTCCCGTCCGTGCTGCC
GGATAATCATTATCTGTACATACCAGTCGAACTGTTTAAAGACCCGAATGAAAAACGTGATCACATGGTGTCTGTG
GAATTTCTGACCGCGCCCGCATTACGGAGGGTATGAACGAACTGTATAAATGATAA**TTAGTTAGTTAG**CAGATCC
GGCTGCTAACAAAGCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGTAGCAATAA**AGAGAATATAAAAAGCCAGA**
TTATTAATCCGGCTTTTTTGTATT

B. subtilis adenine riboswitch⁴ with a T7 promoter and A206K YPet

ATTTAATACGACTCACTATAGATTATCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGT
AAAATCCGTGATTACAAAATTTGGTTATGACATTTTGTAAATCAGGATTTTTTTTATTATCAAAAACATTTAAGT
AAAGGAGTCTCGAGATGGTGTCCAAAGGCGAAGAAGTGTTCACCGGTGTGGTCCGATTCTGGTGGAACTGGATGG
CGAGCTTAACGGTCATAAAATTTAGTGTGTCCGGCGAAGGTGAAGGCGATGCGACCTATGGCAAACTGACGCTGAAA
CTGTGTGCACCACCGGTAAGTGTCCGGTCCCGTGGCCGACCTGGTGACCACGCTGGGTTATGGCGTGCAGTGT
TCGGCGCTACCCGGACCACATGAAAACAACGATTTCTTTAAAGTGCCATGCCGGAAGGCTATGTTTCAGGAACG
TACCATCTTTTCAAAGATGACGGTAACTACAAAACCCGCGGGAAGTTAAATTTGAAGGCGATACGCTGGTCAAC
CGTATTTGAACTGAAAGGTATCGACTTCAAAGAAGATGGCAATATTTCTGGGTCATAAACTGGAATATACTACATA
GCCACAACGTGTATATTTACCGCGGATAAACAGAAAAACGGCATCAAAGCCAACCTTCAAAATCCGCCATAACATCGA
AGACGGCGGTGTCAACTGGCCGATCACTACCAGCAAAAACCCCGATTGGTGTATGGTCCCGTCCGTGCTGCCGGAT
AATCATTATCTGTACATACCAGTCGAACTGTTTAAAGACCCGAATGAAAAACGTGATCACATGGTGTCTGCTGGAAT
TTCTGACCGCGCGCCGATACGGAGGGTATGAACGAACTGTATAAATGATAA**CGGCCGCTTAGTTAGTTAG**CAG
ATCCGGCTGCTAACAAAGCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGTAGCAATAA**AGAGAATATAAAAAGC**
CAGATTATTAATCCGGCTTTTTTGTATT

B. subtilis adenine riboswitch⁴ with a *tac* promoter and A206K YPet

GCTAGCGTTGACAATTAATCATCGGCTCGTATAATGTGTGGCCATTATCACTTGTATAACCTCAATAATATGGTTT
GAGGGTGTCTACCAGGAACCGTAAAATCCCTGATTACAAAATTTGTATTGACATTTTTTGTAAATCAGGATTTTTT
TTATTTATCAAAAACATTTAAGTAAAGGAGTCTCGAGATGGTGTCCAAAGGCGAAGAAGTGTTCACCGGTGTGGTCC
CGATTCTGGTGGAACTGGATGGCGACGTTAACGGTCATAAAATTTAGTGTGTCCGGCGAAGGTGAAGGCGATGCGAC
CTATGGCAAACTGACGCTGAACTGTGTGCACCACCGGTAAGTGTCCGGTCCCGTGGCCGACCTGGTGAACACG
CTGGGTTATGGCGTGCAGTGTTCGCGCGCTACCCGGACCACATGAAACAACGATTTCTTTAAAGTGCCATGC
CGGAAGGCTATGTTTCAGGAACGTACCATTTTTTCAAAGATGACGGTAACTACAAAACCCGCGGGAAGTTAAAT
TGAAGGCGATACGCTGGTCAACCGTATTGAACTGAAAGGTATCGACTTCAAAGAAGATGGCAATATTTCTGGTCA
AACTGGAATATACTACAATAGCCACAACGTGTATATTTACCGCGGATAAACAGAAAAACGGCATCAAAGCCAAC
TCAAAATCCGCCATAACATCGAAGACGGCGGTGTTCAACTGGCCGATCACTACCAGCAAAAACCCCGATTGGTGA
TGGTCCGGTCCGTGCTGCCGATAATCATTATCTGTACATACCAGTCGAACTGTTTAAAGACCCGAATGAAAAACGT
GATCACATGGTGTCTGCTGGAATTTCTGACCGCGCGCCGATACGGAGGGTATGAACGAACTGTATAAATGATAAG
CGGCCGCT**TTAGTTAGTTAG**CAGATCCGGCTGCTAACAAAGCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGTAG
CAATAA**AGAGAATATAAAAAGCCAGATTATTAATCCGGCTTTTTTGTATT**

Supporting References

- 1 E. A. Davidson, P. J. Dlugosz, M. Levy and A. D. Ellington, *Curr Protoc Mol Biol*, 2009, **Chapter 24**, Unit 24 26.
- 2 T. Sunami, T. Matsuura, H. Suzuki and T. Yomo, *Methods Mol Biol*, 2010, **607**, 243-256.
- 3 S. A. Lynch and J. P. Gallivan, *Nucleic Acids Res*, 2009, **37**, 184-192.
- 4 M. Mandal and R. R. Breaker, *Nat Struct Mol Biol*, 2004, **11**, 29-35.

Chapter 11

Measuring Riboswitch Activity In Vitro and in Artificial Cells with Purified Transcription–Translation Machinery

Laura Martini and Sheref S. Mansy

Abstract

We present a simple method to measure the real-time activity of riboswitches with purified components in vitro and inside of artificial cells. Typically, riboswitch activity is measured in vivo by exploiting β -galactosidase encoding constructs with a putative riboswitch sequence in the untranslated region. Additional in vitro characterization often makes use of in-line probing to explore conformational changes induced by ligand binding to the mRNA or analyses of transcript lengths in the presence and absence of ligand. However, riboswitches ultimately control protein levels and often times require accessory factors. Therefore, an in vitro system capable of monitoring protein production with fully defined components that can be supplemented with accessory factors would greatly aid riboswitch studies. Herein we present a system that is amenable to such analyses. Further, since the described system can be easily reconstituted within compartments to build artificial, cellular mimics with sensing capability, protocols are provided for building sense-response systems within water-in-oil emulsion compartments and lipid vesicles. Only standard laboratory equipment and commercially available material are exploited for the described assays, including DNA, purified transcription–translation machinery, i.e., the PURE system, and a spectrofluorometer.

Key words Riboswitch, Transcription–translation, In vitro compartmentalization, Liposome, Emulsion, Cell-free synthetic biology

1 Introduction

Riboswitches are genetically encoded control elements that respond to small molecules through direct binding. Sensing is mediated by an aptamer [1–4] sequence within the mRNA that controls the conformation of the expression platform. Usually ligand binding turns off gene expression; however, natural on-riboswitches exist [5]. The induced conformational changes either regulate transcription through terminator–anti-terminator activity, translation by modulating the accessibility of the ribosome binding

site, mRNA processing, or splicing [6]. In addition to natural riboswitches, many riboswitches have been engineered by modifying previously selected aptamer sequences [7] or by mutating natural riboswitches to display new functionality [8]. Most of the characterized natural riboswitches control transcription, whereas synthetic riboswitches typically control translation.

Monitoring transcription *in vitro* is straight forward [9, 10] thereby allowing for the characterization of riboswitches that alter transcript length in a manner dependent upon the presence or absence of ligand. However, riboswitches that control ribosome binding site accessibility produce transcripts of the same length regardless of the presence or absence of ligand, making methods that quantify differences at the RNA level less insightful. Moreover, riboswitches ultimately control protein synthesis, regardless of the specific mechanism exploited. Therefore, more direct methods that probe the influence of riboswitch activity on protein synthesis are desirable. This is most often achieved by placing the riboswitch in question within a genetic construct that encodes β -galactosidase, a fluorescent protein [7, 11] or more recently, a protein involved in motility [12] and monitoring the activity of the reporter protein in *Escherichia coli*. In other words, the assay is carried out within the cell and absorbance or fluorescence are quantified.

The advantage of such methods is that the activity of the riboswitch within a living cell is monitored, meaning that the measured activity is not a result of imperfect *in vitro* approximations of *in vivo* conditions. However, there are several limitations of such *in-cell* assays. First, the influence of accessory proteins could easily be missed, since their participation in sensing or transducing chemical messages is largely uncontrolled in such experiments. Second, the putative ligand either must be capable of crossing the membrane (to allow for exogenous delivery) or easy to manipulate in terms of concentration. For example, the activity of the flavin mononucleotide (FMN) riboswitch was characterized at the transcriptional level *in vitro* [13, 14], but the influence of FMN on protein synthesis was not investigated, presumably due to the difficulty in quantifying and modulating intracellular FMN concentrations.

Herein we present a simple method to characterize the influence of riboswitch activity on protein synthesis *in vitro*. Guidelines for the design and assembly of the genetic construct, and the evaluation of *in vitro* riboswitch activity by monitoring the synthesis of fluorescent protein with fully defined components are described (Fig. 1). Importantly, this real-time fluorescence assay is amenable to the screening of protein accessory factors and ligands, including ligands that are metabolites, impermeable, or toxic. It should be noted that the described protein synthesis assay does not replace current methods that characterize transcriptional activity. The investigation of both transcription and translation is needed in order to fully define the mechanistic details of riboswitch activity.

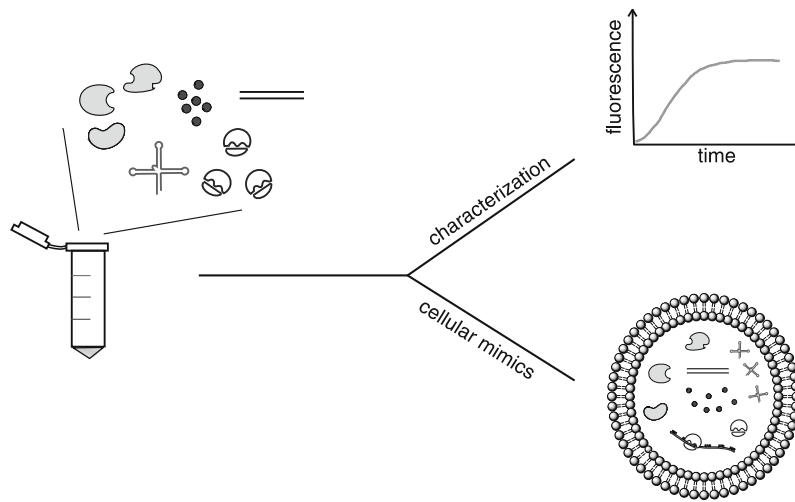


Fig. 1 Cell-free systems for in vitro riboswitch characterization and the construction of artificial, cellular mimics. The PURE system is used to characterize in real-time riboswitch activity through the expression of a reporter protein either in vitro or inside of a compartment with dimensions similar to living cells

We additionally describe how this riboswitch controlled in vitro transcription–translation system can be encapsulated within compartments to build cellular mimics (Fig. 1). As opposed to the majority of artificial cell studies that focus on self-replication, riboswitch sensing-based cellular mimics integrate more fully with the environment and thus could potentially serve as a platform for future technologies. The example described here uses water-in-oil (w/o) emulsion droplets [15], vesicles [16], and a previously reported theophylline riboswitch [7, 17, 18].

2 Materials

All solutions should be prepared using diethyl pyrocarbonate (DEPC) treated water. All reagents are nuclease-free, molecular biology grade. The theophylline riboswitch sequence used here is available from the Registry of Standard Biological Parts (BBa_J89000).

2.1 Template Preparation

1. *E. coli* DH5 α or similar laboratory, cloning strain.
2. Commercial plasmid miniprep kit.
3. 25:24:1 Phenol–chloroform–isoamyl alcohol mixture (*see Note 1*).

2.2 In Vitro Transcription and Translation

1. PURExpress in vitro protein synthesis kit (New England Biolabs).
2. Riboswitch ligand molecule (e.g., theophylline).

3. RNase Inhibitor (RiboLock RNase Inhibitor, Fermentas).
4. Quartz ultra-micro cell cuvette (105.252-QS, Hellma).
5. QuantaMaster 40 UV-Vis Spectrofluorometer with a Peltier temperature controlled single sample holder (Photon Technology International) or a similar spectrofluorometer.

2.3 Emulsion Preparation

1. Mineral Oil.
2. Span 80.
3. Tween 80.
4. Triton X-100.
5. 9 mm Teflon stir bar.
6. Magnetic stir plate.
7. Gilson Microman Positive displacement pipettes.

2.4 Liposome Preparation

1. 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC).
2. Cholesterol.
3. *N*-(carbonyl-methoxypolyethyleneglycol5000)-1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine (DSPE-PEG 5000, NOF-Europe).
4. Rotary evaporator, e.g., Rotavapor R-210 with Vacuum Pump V-700 (Buchi).
5. IKA T 10 basic ULTRA-TURRAX disperser with a 5 mm diameter dispersing tool.
6. Mini-Extruder (Avanti Polar Lipids, Inc.).
7. Nucleopore Track-Etch Membrane 0.4 μm (Whatman).
8. Centrifugal evaporator, e.g., CentriVap Centrifugal Vacuum Concentrator (Labconco).
9. Tris saline buffer; 50 mM Tris-HCl, 50 mM NaCl, pH 7.4 supplemented with 10 mg/mL Proteinase K (Fermentas).

2.5 Imaging

1. Zeiss Observer Z1 microscope (Carl Zeiss S.p.A.) or similar fluorescence microscope.

3 Methods

3.1 DNA Template Preparation

The DNA template can be either a circular plasmid or a linear PCR product that contains a series of modular elements to allow for ribo-switch controlled protein synthesis (Fig. 2). The sequence should contain a transcriptional promoter, a sequence encoding the ribo-switch that contains a ribosome binding site (preferably the natural ribosome binding site sequence, if possible), a gene coding for a fluorescent protein to act as a reporter, and a transcriptional terminator. The transcriptional promoters for T7 and *E. coli* RNA polymerases are typically used. However, since the activity of a riboswitch can depend

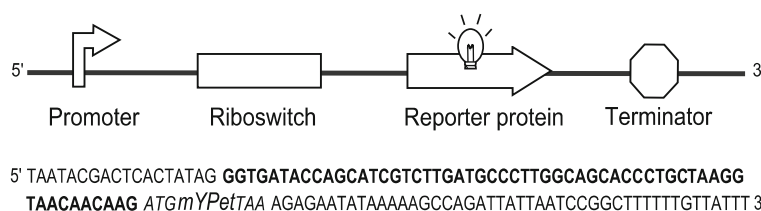


Fig. 2 The composition of a theophylline riboswitch for the real-time observation of riboswitch activity. The construct shown here contains a standard T7 transcriptional promoter, a theophylline riboswitch that contains a ribosome binding site, a gene that codes for A206K YPet (mYPet), and a T7 transcriptional terminator. Only the start and stop codons of mYPet are shown. The riboswitch sequence is shown in *bold*

on the RNA polymerase [19], particularly for riboswitches that use a terminator–anti-terminator mechanism, the choice of which promoter to use can significantly impact *in vitro* riboswitch activity. Similarly, it is advisable that the sequence chosen for the riboswitch portion of the construct contains the ribosome binding site. Riboswitches are often associated with inefficient ribosome binding sites, either due to a lack of potential base-pairing interactions with the ribosome or because of structural features of the riboswitch that obstruct ribosome binding site—ribosome interaction. Further, since riboswitches typically do not fully block protein synthesis in the off-state nor mediate robust expression in the on-state, *i.e.*, riboswitch control is leaky and generally mediates more subtle changes in expression [7], a ribosome binding site not tuned to the activity of the riboswitch could complicate analyses. The reporter should be a protein that expresses well *in vitro* and is easily detectable. We find the green fluorescent proteins super folder GFP (sfGFP) and GFPmut3b and the yellow fluorescent proteins YPet and Venus to be particularly good choices [20]. Finally, incorporating a hairpin transcriptional terminator is advisable, even if not absolutely required when using a linear PCR product as a template, because structured RNA termini increase RNA stability and thus protein yield [21].

1. The DNA template should contain from 5' to 3' a T7 promoter followed by two GG nucleotides to enhance transcription, a sequence encoding the riboswitch [17] and the RBS, and the gene coding for the reporter protein followed by a transcriptional terminator (Fig. 2).
2. The template should be amplified either by PCR or by transforming a typical laboratory cloning strain of *E. coli*, such as DH5 α , and purified with a commercially available kit according to the manufacturer's instructions.
3. Subsequently, the DNA is phenol–chloroform extracted [22] with an equal volume of Tris-buffered Phenol–Chloroform (*see Note 1*).
4. The DNA is ethanol precipitated [23], resuspended in sterile water, and stored at -20 °C.

3.2 *In Vitro* Transcription and Translation Reaction

The major advantage of working *in vitro*, compared with *in vivo*, is that the system operates only with what is provided. In other words, *in vitro* activity cannot depend on unidentified cellular components, because they are not present. To date, only *E. coli* [24] and *Thermus thermophilus* [25] translation machinery have been reconstituted *in vitro* from purified components. Of these two, only the *E. coli* system, i.e., the PURE system, is commercially available. It should be noted that in contrast to *in vivo* or cell-extract conditions, reactions with purified transcription–translation machinery do not contain nucleases. The lack of nucleases decreases the amount of DNA template needed. Whether protein production is more or less efficient with the PURE system in comparison with cell-extract based systems depends on the specific folding properties of the expressed protein.

1. The PURE system components should be aliquoted on ice and stored in 0.2 mL microcentrifuge tubes. Convenient volumes are 10 μ L aliquots of solution A and 7.5 μ L aliquots of solution B. Aliquots are stored at -80 °C.
2. Assemble the reaction components, except for the DNA template, on ice following the manufacturer's instructions. Supplementary reagents can also be added, such as RNase inhibitor (e.g., 20 U RiboLock RNase Inhibitor) or the ribo-switch ligand (e.g., 0.5 mM theophylline).
3. The assembled reaction is transferred to a quartz cuvette and incubated at 37 °C.
4. The reaction is initiated by the addition of the DNA template and monitored by fluorescence spectroscopy for 6 h (Fig. 3) (see Note 2). The DNA template concentration should be screened. We used 250 ng of plasmid template in 25.5 μ L total reaction volume. If YPet is used as the reporter protein, the excitation and emission wavelengths are 517 nm and 530 nm, respectively.

3.3 *In Vitro* Compartmentalization

Since cellular life is chemically distinct from the environment, efforts to mimic cells in the laboratory often times exploit w/o emulsion droplets or vesicles to approximate the compartment of the living, chemical system [26, 27]. However, even if cellular life is distinct from the environment, life cannot exist in isolation and must in some manner interface with the environment to survive [28]. Since lipid vesicles are semipermeable and more similar to the types of barriers found in biology, vesicles are better suited than w/o emulsion droplets for the construction of cellular mimics. Nevertheless, the encapsulation efficiency of w/o emulsions is nearly 100 %, whereas encapsulation efficiency in vesicles is at best 30 % [29]. It is for this reason that the screening of compartmentalized reactions is carried out with w/o emulsions. Once optimal conditions are identified, similar vesicle systems are setup.

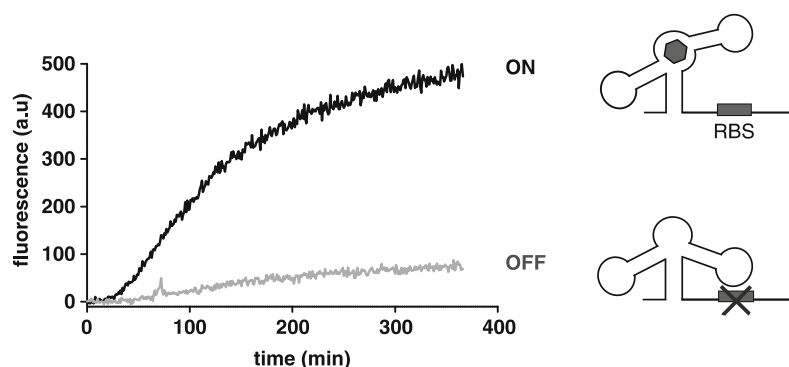


Fig. 3 In vitro theophylline riboswitch activity observed by measuring the expression of the reporter protein mYPet. The presence of the ligand activates protein expression. A schematic representation of the mRNA in the ligand-bound ON state and the uncomplexed OFF state are shown. Protein production is inhibited in this case in the OFF state because the RBS is not available for base-pairing with the ribosome. The data were taken from a previous in vitro theophylline riboswitch study [18]

3.3.1 Cell-Free Expression in w/o Emulsion

The method described here is based on that of Davidson et al. [30] but has been scaled down to be compatible with small volume PURE system reactions. Positive displacement pipettes are used to handle oil samples.

1. The oil phase that will be used for the w/o emulsion is assembled in a 15 mL Falcon tube (*see Note 3*), in the following order: 474.75 μL of mineral oil, 22.5 μL of span 80, 2.5 μL of tween 80, and 0.25 μL of triton X-100.
2. The aqueous phase is first assembled in a microcentrifuge tube by mixing the PURE system components on ice as described above for the in vitro reactions. Note that the optimal DNA template concentration may be different when transcription-translation is performed in a compartment versus in vitro. We used 500 ng of plasmid DNA in 25.5 μL of total aqueous volume (*see Note 4*) for expression in w/o emulsion droplets. The theophylline concentration was also increased to 5 mM to compensate for partitioning into the oil phase (*see Note 5*).
3. The 15 mL tube containing the oil phase is placed in a 250 mL beaker filled with ice water on a magnetic stir plate. A teflon stir bar is inserted in the oil phase and the oil is mixed by stirring at maximum speed for 1 min.
4. The emulsion is formed by the drop-wise addition of the aqueous phase containing the PURE system reaction to the oil phase over 1 min with continuous stirring. The emulsion is then stirred for an additional 3 min (*see Note 6*).
5. Finally, the emulsion is transferred to a 2 mL microcentrifuge tube and incubated at 37 $^{\circ}\text{C}$ for 6 h. 5 μL aliquots are removed every hour for observation by fluorescence microscopy.

3.3.2 Cell-Free Expression in Vesicles

The freeze-dried empty liposome (FDEL) method, as described by Yomo and colleagues [16], is used with slight modification to build the vesicles that house the transcription–translation reaction. FDEL vesicles encapsulate macromolecular, hydrophilic components relatively efficiently. A variety of lipid compositions can be exploited. Here 12 μmol of 58:39:3 POPC–cholesterol–DSPE-PEG 5000 is used.

1. Each lipid is dissolved in chloroform and mixed in a 5 mL round-bottom flask.
2. The solution is subjected to rotary evaporation for 1 h. The resulting thin lipid film is then hydrated with 1 mL of DEPC-treated water and vortexed for 20 s or until a homogeneous opaque solution is formed.
3. The vesicle solution is then transferred to a 2 mL microcentrifuge tube and disrupted with an IKA T 10 basic homogenizer at high speed (level 4 setting) for 1 min (*see Note 7*).
4. Samples are extruded through 400 nm polycarbonate filters 11 times with an Avanti mini-extruder. 40 μL aliquots of the vesicles are placed in 1.5 mL microcentrifuge tubes, frozen in liquid nitrogen (*see Note 8*), and lyophilized with a centrifugal evaporator overnight at 30 °C. A thin opaque lipid layer can be observed at the bottom of the microcentrifuge tube after lyophilization. At this stage the samples can be stored at –20 °C.
5. A PURE system reaction is assembled on ice in a total volume of 22.1 μL , including 500 ng of the template plasmid (*see Note 4*). 20 U of RiboLock RNase inhibitor is added to the solution.
6. 10 μL of the assembled PURE system reaction is added to an aliquot of FDEL vesicles on ice and incubated without agitation for 2.5 h (*see Note 9*). The unused portion of the PURE system reaction can be stored at –80 °C.
7. The hydrated liposomes are then diluted 20-fold in Tris saline buffer supplemented with proteinase K in a 0.2 mL microcentrifuge tube and incubated at 37 °C. The inclusion of proteinase K is to degrade extravesicular proteins.
8. At this point the ligand to be sensed, e.g., 5 mM theophylline (*see Note 5*) is added. Theophylline is capable of diffusing across the membrane, binding directly to the mRNA, and activating translation thereby resulting in fluorescence. Control reactions in the absence of ligand should result in no or significantly reduced fluorescence.
9. The reactions are incubated at 37 °C for 6 h. 5 μL aliquots are removed from the reaction every 1.5 h and visualized by fluorescence microscopy.

3.4 Microscope Sample Preparation

Detecting activity inside of vesicles by fluorescence microscopy is more difficult than *in vitro* measurements with a spectrofluorometer. First, encapsulation efficiency is low, particularly when over 80 different components need to be encapsulated within one vesicle in order for protein synthesis to proceed [31] (*see* **Notes 4** and **9**). Second, riboswitches typically have weak ribosome binding sites and thus produce less protein than constructs typically exploited for recombinant expression. Finally, fluorescent proteins photobleach and require time to mature. We use a monomeric version of the yellow fluorescent protein YPet as a reporter because the characteristics of YPet are more amenable to investigation by microscopy. YPet is one of the brightest fluorescent proteins and is more photostable than the majority of available fluorescent proteins [32]. The expression of monomeric YPet with the PURE system requires approximately 2 h to reach half maximal fluorescence [20].

1. 5 μ L aliquots are removed from the reaction and spotted on a clean glass slide. A cover slip is added.
2. The slide is then left for 2 min to rest on the bench. This step helps decrease the number of rapidly moving vesicles.
3. The sample is then observed by bright field and epifluorescence with 100 \times magnification (Fig. 4). Care should be taken to decrease photobleaching by decreasing exposure time.

4 Notes

1. It is preferable to avoid Phenol–Chloroform solutions that contain ethylenediaminetetraacetic acid (EDTA), since the chelation of metals by EDTA can interfere with enzyme activity.
2. A plate reader or a real-time PCR machine can be used in place of a spectrofluorometer.
3. We also made w/o emulsions in 13 mL Sarstedt tubes with stirring with x-shaped spinplus stir bars, as described by Davison et al. [30]. The resulting emulsions were more homogeneous and more stable than the emulsions we obtained with Falcon tubes and linear stir bars. However, we found the Davidson et al. emulsion droplets to be smaller and thus more difficult to observe by microscopy.
4. The PURE system instruction manual suggests screening between 25 to 250 ng of template DNA for 25 μ L reactions. For the *in vitro* characterization of the theophylline riboswitch, 250 ng of plasmid template was found to be optimal. However, better results were obtained with 500 ng DNA template for compartmentalized reactions.

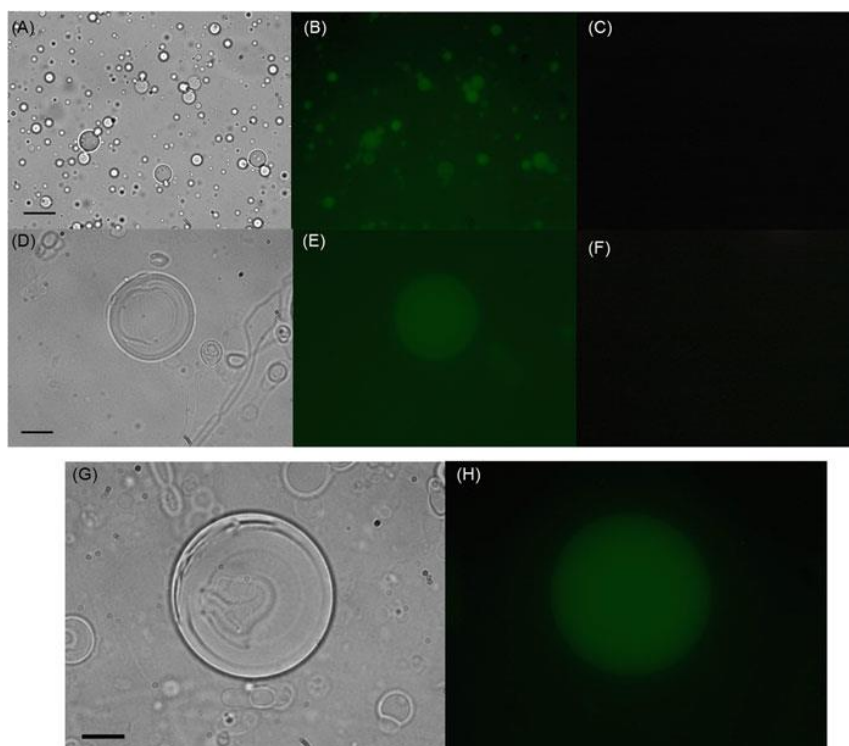


Fig. 4 Fluorescence microscopy images of riboswitch activity in compartments. Upper panels (a–c): activity recorded in emulsion compartments. Middle panels (d–f): activity recorded in liposomes. Lower panels (g, h): activity due to riboswitch sensing of the environment. Images are all epifluorescence, except for panels a, d, g, which are bright-field. Panel b, e show the fluorescence resulting in compartments when the ligand is present. The corresponding controls (i.e., in the absence of ligand) are also shown (c, f). In panel h, the activity of a riboswitch as a sensor element is shown. The data are from the characterization of a cell-free theophylline riboswitch system and panels g and h are reproduced with permission from the Royal Society of Chemistry [18]

5. Since the oil–water partition coefficient of theophylline is low [33], much more theophylline is required to activate the riboswitch in the presence of oil than in aqueous solution.
6. Stirring is an important parameter to consider when generating an emulsion. A constant stir force should be used and the stir bar must be compatible with the tube holding the aqueous-oil mixture. For example, the conical shape of a Falcon tube is not compatible with x-shaped stir bars. The efficiency of mixing can be qualitatively assessed by eye by including in the aqueous phase 1 mM HPTS (8-hydroxypyrene-1,3,6-trisulfonic acid) and observing the distribution of green color throughout the tube during stirring.
7. We also tested sonication as a possible disruption method. Sonication at 70 A of amplitude for 5 min resulted in higher dispersion. However, the number of liposomes observed by

microscopy was lower than that observed with the IKA T homogenizer.

8. Freezing can also be performed with dry ice without any appreciable difference in liposomes formation.
9. The slow vesicle formation process mediated by the natural swelling method described herein results in fewer, but larger vesicles that are easier to observe by microscopy than by other methods that exploit vortexing. Additionally, compartment size impacts protein synthesis efficiency with larger vesicles being more compatible with protein synthesis [34].

References

1. Soukup GA, Breaker RR (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* 5:1308–1325
2. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Braker RR (2002) Genetic control by a metabolite-binding mRNA. *Chem Biol* 9:1043
3. Wakeman CA, Winkler WC (2009) Analysis of the RNA backbone: structural analysis of riboswitches by in-line probing and selective 2'-hydroxyl acylation and primer extension. In: Serganov A (ed) *Riboswitches*. *Methods Mol Biol*, vol 540, pp 173–191
4. Ellington AD, Szostak JW (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 346:818–822
5. Serganov A, Yuan Y-R, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* 11:1729–1741
6. Serganov A, Nudler E (2013) A decade of riboswitches. *Cell* 152:17–24
7. Desai SK, Gallivan JP (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* 126:13247–13254
8. Nomura Y, Yokobayashi Y (2007) Reengineering a natural riboswitch by dual genetic selection. *J Am Chem Soc* 129:13814–13815
9. Artsimovitch I, Henkin TM (2009) In vitro approaches to analysis of transcription termination. *Methods* 47:37–43
10. Mironov A, Epshtein V, Nudler E (2009) Transcriptional approaches to riboswitch studies. In: Serganov A (Ed) *Riboswitches*. *Methods Mol Biol*, vol 540: 39–51, Humana Press, NY, USA
11. Muranaka N, Yokobayashi Y (2010) Posttranscriptional signal integration of engineered riboswitches yields band-pass output. *Angew Chem Int Ed* 49:4653–4655
12. Topp S, Gallivan JP (2007) Guiding bacteria with small molecules and RNA. *J Am Chem Soc* 129:6807–6811
13. Soukup GA, Breaker RR (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A* 96:3584–3589
14. Wickiser JK, Winkler WC, Breaker RR, Crothers DM (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* 18:49–60
15. Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 16:652–656
16. Sunami T, Matsuura T, Suzuki H, Yomo T (2010) Synthesis of functional proteins within liposomes. *Methods Mol Biol* 607:243–256
17. Lynch SA, Gallivan JP (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res* 37:184–192
18. Martini L, Mansy SS (2011) Cell-like systems with riboswitch controlled gene expression. *Chem Commun* 47:10734–10736
19. Lemay J-F, Desnoyers G, Blouin S, Heppell B, Bastet L, St-Pierre P, Massé E, Lafontaine DA (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet* 7:e1001278
20. Lentini R, Forlin M, Martini L, Del Bianco C, Spencer AC, Torino D, Mansy SS (2013) Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. *ACS Synth Biol* 2:482–489
21. Aiba H, Hanamura A, Yamano H (1991) Transcriptional terminator is a positive regulatory element in the expression of the *Escherichia coli crp* gene. *J Biol Chem* 266:1721–1727
22. Sambrook J, Russel DW (2001) Commonly used techniques in molecular cloning. In: *Molecular cloning*. (Ed) Jan Argentine Cold Spring Harbor Laboratory Press, NY, USA

23. Sambrook J, Russell DW (2006) Standard ethanol precipitation of DNA in microcentrifuge tubes. *Cold Spring Harb Protoc.* doi:[10.1101/pdb.prot4456](https://doi.org/10.1101/pdb.prot4456)
24. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19:752–755
25. Zhou Y, Asahara H, Gaucher EA, Chong S (2012) 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* 40:7932–7945
26. Forlin M, Lentini R, Mansy SS (2012) Cellular imitations. *Curr Opin Chem Biol* 16:586–592
27. Ichihashi N, Matsuura T, Kita H, Sunami T, Suzuki H, Yomo T (2010) Constructing partial models of cells. *Cold Spring Harb Perspect Biol.* doi:[10.1101/cshperspect.a004945](https://doi.org/10.1101/cshperspect.a004945)
28. Bianco CD, Mansy SS (2012) Non replicating protocells. *Acc Chem Res* 45:2125–2130
29. Colletier J-P, Chaize B, Winterhalter M, Fournier D (2002) Protein encapsulation in liposomes: efficiency depends on interaction between protein and phospholipid bilayer. *BMC Biotechnol* 2:9
30. Davidson EA, Dlugosz PJ, Levy M, Ellington AD (2009) Directed evolution of proteins in vitro using compartmentalization in emulsions. *Curr Protoc Mol Biol.* doi:[10.1002/0471142727.mb2406s87](https://doi.org/10.1002/0471142727.mb2406s87)
31. Lazzerini-Ospri L, Stano P, Luisi P, Marangoni R (2012) Characterization of the emergent properties of a synthetic quasi-cellular system. *BMC Bioinform* 13(Suppl 4):S9
32. Nguyen AW, Daugherty PS (2005) Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* 23:355–360
33. Donoso P, O'Neill SC, Dilly KW, Negretti N, Eisner DA (1994) Comparison of the effects of caffeine and other methylxanthines on $[Ca^{2+}]_i$ in rat ventricular myocytes. *Br J Pharmacol* 111:455–458
34. Pereira de Souza T, Stano P, Luisi PL (2009) The minimal size of liposome-based model cells brings about a remarkably enhanced entrapment and protein synthesis. *Chembiochem* 10:1056–1063

2.2. Artificial cells as chemical translators

Cellular mimics have been engineered to sense and respond to theophylline from the environment with a small molecule-dependent RNA regulator. However, this system gives a response in the internal compartment, displaying a fluorescent output. It is possible to further engineer the cell mimic in order to achieve a response activity which is not limited to the compartment. Thus, the information sent by the artificial cell can be processed by another receiving system. In other words, the sensing and responding functions can be mediated by the cell mimic, causing a behavioural effect as final response on living cells, e.g. bacteria.

In contrast to the typical synthetic biology approach of controlling biological functions via engineering of living systems, the described artificial cell does not require direct genetic modification of the receiving bacteria. Moreover, cellular mimics can potentially be exploited for therapeutic function, e.g. mediating the disruption of *Pseudomonas aeruginosa* biofilms in cystic fibrosis patients. As another advantage, artificial cells degrade rapidly thereby avoiding the fear of losing control of the system over time.

In this context, the publication *Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour* (Lentini et al., *Nat. Commun.*, 2014, 5, 4012) presents an artificial cell that operates as a chemical translator, converting an input signal, to which the bacteria is naturally unable to respond, into a chemical message that the cell can sense and respond to (Fig. 1, Lentini et al., *Nat. Commun.*, 2014, 5, 4012). A cellular mimic responsive to theophylline is engineered that allows for the release of isopropyl β -D-1-thiogalactopyranoside (IPTG). The DNA of the artificial cell codes for the theophylline riboswitch, which controls the translation of the pore-forming protein, i.e. α -hemolysin (α HL). IPTG is entrapped in the cellular mimic and the phospholipid bilayer is impermeable to the small molecule. Thus, only in presence of theophylline, the cell mimic activates and allows the expression of α HL, releasing IPTG. The detection by the bacteria

of the IPTG chemical signal is assessed firstly by transforming *E. coli* with a plasmid carrying a fluorescent protein under the control of an IPTG responsive *lac* operon. Fluorescence is directly monitored by flow cytometry. Secondly, the variation of gene expression of untransformed *E. coli* is measured by reverse transcription quantitative PCR (RT-qPCR).

Initial optimization of the system is performed through the testing of the components of the system *in vitro*. The riboswitch activity is assessed through the investigation by fluorescence kinetic measurements of the theophylline dependent expression of a fusion construct consisting of α HL and super folder GFP. However, the lack of difference in protein expression in presence and absence of theophylline suggests the existence of internal RBS sequences that interfere with the determination of riboswitch activity by monitoring fluorescence. In fact, a variant with a mutated version of α HL allows for the reduction of the background fluorescent signal, thereby revealing riboswitch activity (*Fig. 2, Lentini et al., Nat. Commun., 2014, 5, 4012*). Moreover, hemolysis tests on red blood cells show that the α HL produced *in vitro* is functional, and pores can be effectively formed.

In order to demonstrate that *E. coli* can sense and respond to IPTG released from the cellular mimic, bacteria are transformed with a plasmid encoding the fluorescent protein GFPmut3b under the control of an IPTG-inducible operon. After incubation of the cellular mimics and bacteria, in the presence and absence of theophylline, flow cytometry results clearly demonstrate fluorescence of the *E. coli* cells (*Fig. 3, Lentini et al., Nat. Commun., 2014, 5, 4012*). Thus, the artificial cells are able to communicate with *E. coli*, translating an unrecognized chemical signal (theophylline) into a signal (IPTG) that bacteria can respond to.

The expression of a fluorescent reporter may be considered as an engineered response in bacteria since the cells were transformed with an exogenous plasmid coding for the fluorescent protein. For this reason, investigation by RT-qPCR of the variation of

expression of *lac* operon genes *LacZ*, *LacY* and *LacA* is carried out for untransformed bacteria in presence and absence of theophylline. The results confirm the ability of the artificial cell to translate the theophylline message into information contained within IPTG, processed by the natural cells. These results highlight the possibility to interface cellular mimics with living system, leading to regulation and reprogramming. Moreover, the sensing abilities of bacteria can further be expanded by the use of a chemical translator.

Detailed designs, materials and methods and discussion are attached in the paper. My involvement in this publication relates to the setting up of the theophylline riboswitch experiments. Concerning the construction of the α HL-super folder GFP reporter template, I analyzed the RNA folding with the secondary structure prediction software *mfold*, leading to the identification of internal RBS sequences. These RBSs were then successfully mutated to optimize the characterization of the *in vitro* activity of the riboswitch. Finally, I was involved in editing the manuscript before publication.

ARTICLE

Received 4 Jan 2014 | Accepted 30 Apr 2014 | Published 30 May 2014

DOI: 10.1038/ncomms5012

OPEN

Integrating artificial with natural cells to translate chemical messages that direct *E. coli* behaviour

Roberta Lentini¹, Silvia Perez Santero^{1,2}, Fabio Chizzolini¹, Dario Cecchi¹, Jason Fontana¹, Marta Marchioretto³, Cristina Del Bianco¹, Jessica L. Terrell^{4,5}, Amy C. Spencer¹, Laura Martini¹, Michele Forlin¹, Michael Assfalg², Mauro Dalla Serra³, William E. Bentley^{4,5} & Sheref S. Mansy¹

Previous efforts to control cellular behaviour have largely relied upon various forms of genetic engineering. Once the genetic content of a living cell is modified, the behaviour of that cell typically changes as well. However, other methods of cellular control are possible. All cells sense and respond to their environment. Therefore, artificial, non-living cellular mimics could be engineered to activate or repress already existing natural sensory pathways of living cells through chemical communication. Here we describe the construction of such a system. The artificial cells expand the senses of *Escherichia coli* by translating a chemical message that *E. coli* cannot sense on its own to a molecule that activates a natural cellular response. This methodology could open new opportunities in engineering cellular behaviour without exploiting genetically modified organisms.

¹CIBIO, University of Trento, via delle Regole 101, 38123 Mattarello (TN), Italy. ²Department of Biotechnology, University of Verona, 37134 Verona, Italy. ³National Research Council—Institute of Biophysics & Bruno Kessler Foundation, Via alla Cascata 56/C, 38123 Trento, Italy. ⁴Fischell Department of Bioengineering, University of Maryland, College Park, Maryland 20742, USA. ⁵Institute for Bioscience and Biotechnology Research, University of Maryland, College Park, Maryland 20742, USA. Correspondence and requests for materials should be addressed to S.S.M. (email: mansy@science.unitn.it).

Synthetic biology thus far has relied upon the engineering of new cellular function through the insertion and deletion of genetic information in living cells. This genetic engineering based approach has progressed rapidly. There is now available a set of well-characterized biological parts^{1–3} that can be used to build complex genetic circuitry within and between the living cells^{4–6}. Further, entire genomes can be edited⁷ and synthesized⁸, suggesting that fully designed organisms with heretofore unseen capabilities are likely in the future.

Despite the wide range of technologies and target pathways exploited, the desire to control microorganisms to date has always employed direct genetic intervention. The limitations of these prevalent methods are due to the difficulties of engineering living systems, including evolutionary pressures that may alter engineered pathways over time and the potential long-term consequences of altering ecosystems with engineered organisms. However, it may not be necessary to genetically modify living cells. Extant life is already extremely complex, endowed with numerous sensory and metabolic pathways tuned by billions of years of evolution to be efficiently responsive to changing intracellular and extracellular conditions. A simple change in pH, for example, results in the up and downregulation of nearly 1,000 genes in *Escherichia coli*⁹. In other words, cells are already capable of sensing many different stimuli and capable of performing many tasks. Therefore, it should be possible to exploit these existing cellular pathways to control cellular behaviour without changing the genetic makeup of the cells.

Here we explore this idea of engineering *E. coli* through alternative means by targeting the sensory pathways of *E. coli*. To do so without altering the genetic content of the bacterium, we instead construct artificial cells that could interact with natural cells in order to evoke a behavioural response. The artificial cells in this system function as chemical translators that sense molecules that *E. coli* alone cannot sense. In response, the artificial cells release a molecule that *E. coli* can naturally respond to, thereby translating an unrecognized chemical message into a recognized chemical message. In this way, the sensory capabilities of *E. coli* are expanded without altering the genetic content of the bacterium. The artificial cell is built with a phospholipid vesicle containing isopropyl β -D-1-thiogalactopyranoside (IPTG), DNA, and transcription–translation machinery. The DNA template codes for a previously selected riboswitch that activates translation in response to the presence of theophylline¹⁰. The theophylline riboswitch controls the synthesis of the pore forming protein α -hemolysin (α HL). Therefore, in the presence, but not the absence, of theophylline a pore forms that releases entrapped IPTG. *E. coli* alone does not respond to theophylline,

and IPTG does not cross the vesicle membrane of the artificial cell in the absence of the pore. The ability of *E. coli* to receive the chemical message sent by the artificial cells is assessed in two ways. First, the fluorescence of *E. coli* carrying a plasmid encoding a fluorescent protein behind an IPTG-responsive, *lac* operator sequence is evaluated. Second, the gene expression of untransformed *E. coli* is monitored by reverse transcription quantitative PCR (RT–qPCR). To our knowledge, this is the first artificial, cell-like system capable of translating unrecognized signals into a chemical language that natural cells can recognize. The integration of artificial translator cells with natural cells represents a new strategy to introduce synthetic features to a biological system while circumventing the need for direct genetic manipulation.

Results

The theophylline-sensing device is functional *in vitro*. To build artificial cells that sense theophylline and in response release IPTG (Fig. 1), a theophylline-sensing genetic device was built with a T7 transcriptional promoter, a theophylline riboswitch and a gene encoding a fusion between α HL and super folder GFP at the carboxy terminus. If functioning properly, this arrangement should result in the expression of protein and thus green fluorescence only in the presence of theophylline. However, cell-free expression in the presence and absence of theophylline showed similar levels of fluorescence (Fig. 2a). Since this same riboswitch was previously shown to function *in vitro*¹¹, the sequence of the α HL–GFP gene was more closely examined. Multiple pairs of potential ribosome binding sites (RBS) and start codons were identified within the α HL portion of the gene that were in-frame with the GFP-encoding region. The theophylline riboswitch controls translation, meaning that sequences behind the theophylline riboswitch are always transcribed. Translation from the RBS within the riboswitch is activated by direct binding of theophylline to the messenger RNA. Therefore, if additional sequences outside of the riboswitch but within the α HL portion of the gene were recognized by the ribosome, then regardless of the theophylline concentration, the expression of truncated peptide products with fluorescently active GFP would have been possible. To test if such internal RBSs were present, the theophylline riboswitch and thus the RBS preceding the α HL–GFP sequence was deleted. *In vitro* transcription–translation of this construct showed the accumulation of fluorescence over time similar to the riboswitch containing construct (Fig. 2b). Sequence analysis revealed three potential RBS–start codon pairs within the α HL coding portion of the gene. Of these, a putative RBS of

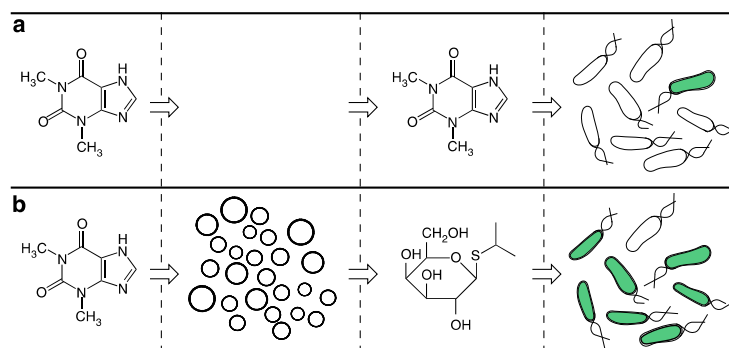


Figure 1 | Artificial cells translate chemical signals for *E. coli*. (a) In the absence of artificial cells (circles), *E. coli* (oblong) cannot sense theophylline. (b) Artificial cells can be engineered to detect theophylline and in response release IPTG, a chemical signal that induces a response in *E. coli*.

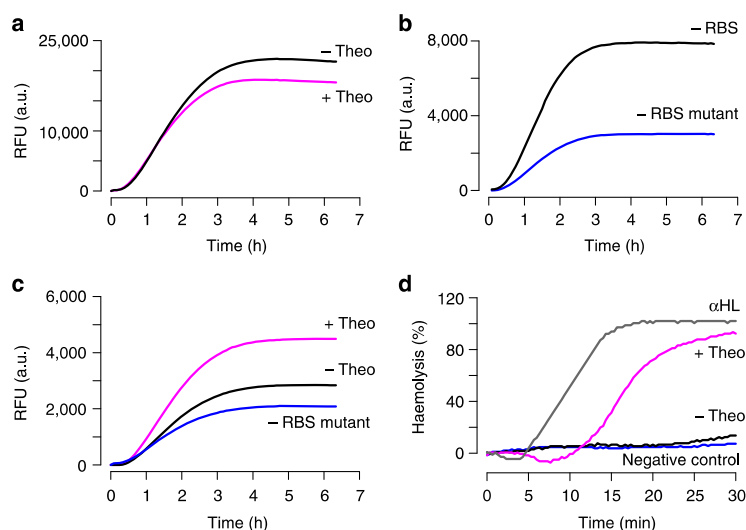


Figure 2 | *In vitro* characterization of the theophylline-sensing device and α HL. (a) The cell-free expression of α HL-GFP behind a theophylline riboswitch gives rise to similar levels of fluorescence both in the presence (+ theo) and absence (- theo) of theophylline at 37 °C. (b) The removal of the theophylline riboswitch and thus, the RBS preceding the start codon of α HL-GFP shows production of a fluorescent protein product when incubated with transcription-translation machinery (- RBS). The removal of a putative internal RBS within the α HL coding portion of the fusion construct significantly decreases the production of the fluorescent protein product (- RBS mutant). (c) The activity of the theophylline-sensing device is observable by fluorescence when an internal RBS is removed. The top and middle curves are the *in vitro* expression of α HL-GFP behind the theophylline riboswitch in the presence (+ theo) and absence of theophylline (- theo), respectively. Background fluorescent protein production is shown with the same construct lacking the theophylline riboswitch (- RBS mutant) used in b. (d) The cell-free expression of theophylline riboswitch-controlled α HL-degraded red blood cells (RBCs) in the presence (+ theo) but not the absence of theophylline (- theo). Control reactions include the expression of an α HL construct lacking the theophylline riboswitch (α HL) and RBCs alone (negative control). RBC degradation was monitored by attenuation at 22 °C. The exploited constructs were SP011A for panel A, SP002A and AS014A for panel B, RLO69A and AS014A for c, and RLO67A and JF001A for d (Supplementary Table 1). Data are averages of three independent reactions.

AAAGAA was selected as the most likely candidate for giving fluorescent protein expression based on sequence composition and spacing¹². The putative internal RBS was removed by mutation to TCTACC, resulting in a carboxy-terminal GFP tagged K30S E31T α HL construct. Fluorescence from this mutated construct was reduced threefold, consistent with the removal of an internal RBS (Fig. 2b). Finally, K30S E31T α HL-GFP was placed behind the theophylline riboswitch to test the activity of the cell-free sensing device. A clear difference was observed between protein expression in the presence and absence of theophylline (Fig. 2c), and the fluorescence arising in the absence of theophylline was within 20% of the construct lacking an RBS upstream of the full gene. The data were consistent with a functioning riboswitch sensor with background fluorescent protein expression arising from internal RBS within α HL. Therefore, the final artificial cellular mimic described below was built with α HL lacking a GFP-tag to avoid complications arising from the expression of truncated fluorescent protein product.

Active α HL is produced in response to theophylline *in vitro*. To ensure that the cell-free expressed α HL was active as a pore, the ability of α HL to degrade rabbit red blood cells was assessed through a standard haemolysis assay¹³. Each construct was expressed *in vitro* at 37 °C for 6 h after which, an aliquot was removed and added to red blood cells. Haemolysis was quantified by measuring attenuation at 650 nm. In the presence of theophylline, 90% haemolysis was observed when the genetic construct containing a riboswitch-controlled α HL was expressed. The cell-free expression of the same construct in the absence of

theophylline gave haemolysis levels similar to the negative control reactions (Fig. 2d), as was expected for a functioning theophylline riboswitch that controls the production of α HL. Control reactions with commercial α HL-purified protein and *in vitro*-expressed α HL and α HL-GFP all were fully active (Fig. 2d, Supplementary Table 2), whereas aliquots from *in vitro*-expressed GFP alone and α HL with a carboxy-terminal His-tag were inactive (Supplementary Table 2). α HL with a carboxy-terminal His-tag was previously shown to have reduced activity¹⁴. Also, comparison of the riboswitch activity fluorescence data with the haemolysis assay data was consistent with the production of GFP containing protein fragments from an internal RBS without an active α HL domain. For example, the α HL-GFP construct lacking one of the putative internal RBSs failed to produce protein with haemolysis activity (Supplementary Table 2), despite giving rise to fluorescence during *in vitro* transcription-translation (Fig. 2b).

Artificial cells can translate chemical messages for *E. coli*. After demonstrating that the riboswitch was able to control the *in vitro* expression of α HL in response to theophylline and that the expressed α HL molecules formed functional pores, the component parts were next assembled inside of phospholipid vesicles to build artificial cells. Theophylline is capable of passing through the membrane of vesicles¹¹. Phospholipid vesicles were generated in the presence of IPTG, transcription-translation machinery and DNA encoding α HL under the control of the theophylline riboswitch. The vesicles were then purified by dialysis at 4 °C to remove unencapsulated molecules. The receiver bacterial cells were mid-exponential phase *E. coli* BL21(DE3) pLysS carrying a

plasmid encoding GFP behind a T7 promoter and a *lac* operator sequence. In this commonly exploited system, IPTG induces the expression of a chromosomal copy of T7 RNA polymerase in *E. coli* BL21(DE3) and derepresses the expression of GFP from the plasmid. Background expression is typically low with such an arrangement because of the presence of constitutively expressed lysozyme from pLysS, a natural inhibitor of T7 RNA polymerase.

To test if the artificial cells could function as chemical translators for *E. coli*, the artificial cells were incubated with *E. coli* BL21(DE3) pLysS carrying the GFP-encoding plasmid at 37 °C, and the fluorescence of *E. coli* was evaluated by flow cytometry. A control reaction in which theophylline was directly added to *E. coli* in the absence of artificial cells failed to show green fluorescence after 3 h (Fig. 3a). Similarly, IPTG loaded vesicles that did not contain the machinery necessary to form pores did not induce fluorescence in *E. coli*. Therefore, theophylline was not able to induce a detectable response in *E. coli*, and IPTG could not cross the vesicle membrane in the absence of α HL, which was consistent with permeability measurements (Supplementary Fig. 1). However, when *E. coli* was incubated with artificial cells and theophylline, 17 \pm 10% and 69 \pm 3% of the bacteria fluoresced green after 0.5 and 3 h, respectively. When the same experiment was repeated in the absence of theophylline, 3 \pm 1% and 24 \pm 5% of the bacteria were fluorescent after 0.5 and 3 h, respectively (Fig. 3a,b). Longer incubations resulted in diminishing differences between the two samples suggesting the presence of low levels of α HL expression in the absence of theophylline. Also, the GFP response was encoded within a medium copy number plasmid. Therefore, higher background levels of GFP were to be expected in comparison with gene expression from the chromosome. The flow cytometry experiments were consistent with the ability of artificial cells to translate an unrecognized chemical signal (theophylline) into a signal (IPTG) that *E. coli* could respond to.

Although the artificial cells were capable of communicating with *E. coli*, the induction of GFP synthesis, as observed above, exploited an engineered response. To assess whether artificial cells could elicit a natural, chromosomally encoded response,

RT-qPCR was used to measure gene expression from the *lac* operon of *E. coli*. The *lac* operon is one of the most thoroughly characterized sensory pathways¹⁵. The presence of allolactose (or the non-hydrolyzable analogue IPTG) induces the expression of *lacZ*, *lacY* and *lacA*. To facilitate detection of *E. coli* responding to the chemical message sent from the artificial cells, *E. coli* BL21 (DE3) pLysS were grown in LB supplemented with glucose to decrease the background expression of the *lac* operon and then transferred to M9 minimal media prior to incubation with artificial cells. The artificial cells were prepared as described for the GFP induction experiments above. After incubating together artificial cells with *E. coli* in the presence and absence of theophylline for 4 h, aliquots were collected for RNA isolation. The RNA was then reverse transcribed and *lacZ*, *lacY*, and *lacA* expression quantified by qPCR. The RNA isolated from bacteria incubated with artificial cells plus theophylline showed on average over 20-fold higher *lacZYA* expression than samples incubated with artificial cells alone (calculated from AC/(AC + theo) as shown in Fig. 3c). Taken together, the data are consistent with the ability of artificial cells to translate chemical messages and induce both engineered and natural pathways in *E. coli*.

Discussion

Direct genetic engineering of living cells is not needed to control cellular behaviour. It is possible, instead, to coerce desired activity through communication with artificial cells. The foundation for such technologies has already been laid by both cell-free and *in vivo* studies. Engineered communication paths between living cells have been constructed to coordinate cellular activities in response to external stimuli^{6,16} and are being developed for therapeutic purposes¹⁷. In these systems, sender cells often can process information and in response release molecules that affect other cells. What has been shown herein builds on these past efforts but does so by integrating reconstituted, non-living systems with living cells. This allows for the genetic engineering component of the system to be moved from the living, evolving, replicating cells to the more controllable, ephemeral artificial

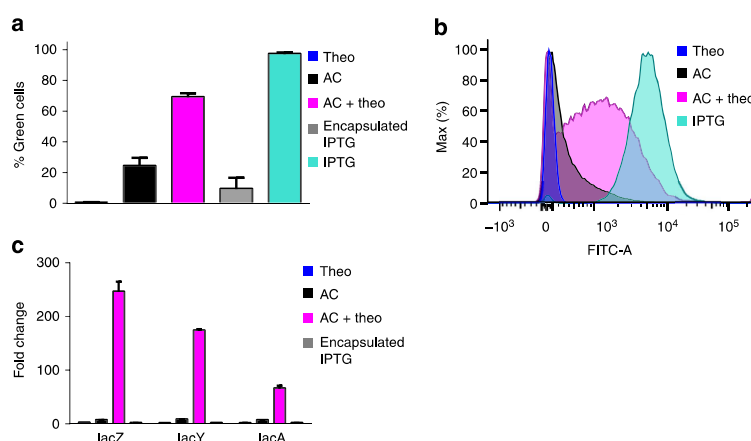


Figure 3 | The artificial translator cells are functional. (a) Artificial cells can induce the expression of a plasmid encoded gene within *E. coli* in response to a molecule that *E. coli* cannot naturally sense. BL21(DE3) pLysS carrying a plasmid encoding GFP behind a *lac* operator sequence was incubated with the following components at 37 °C for 3 h: theophylline (theo), artificial cells (AC), artificial cells plus theophylline (AC + theo), IPTG encapsulated inside of vesicles (encapsulated IPTG), and unencapsulated IPTG (IPTG). *E. coli* fluorescence was quantified by flow cytometry. The reported averages and s.e.m. were calculated from three separate reactions run on three different days from independently assembled artificial cells. (b) A histogram of a subset of the FACS data used in panel a shows a clear shift in the *E. coli* population in the presence of artificial cells plus theophylline. (c) Artificial cells can induce the expression of chromosomally encoded genes of *E. coli*. After 4 h of incubation of artificial cells with *E. coli* at 37 °C, the messenger RNA encoding *lacZ*, *lacY* and *lacA* was quantified by RT-qPCR. Data are reported as averages of three measurements and error bars represent s.e.m.

cells. When the artificial cells degrade, the natural cells go back to their original state, thereby diminishing the possibility of unintended long-term consequences. For example, rather than engineering bacteria to search for and clean up environmental contaminants, artificial cells could be built to sense the contaminant molecules and in response release chemoattractants that bring natural bacteria capable of feeding on the contaminants¹⁸ to the affected site.

Several recent reports have described the engineering of seek-and-destroy bacteria for the eradication of tumours or bacterial infections^{19–22}. However, these methods ultimately rely on administering living bacteria to the patient. Artificial cells could be built to carry out similar tasks if the sensor module of the artificial cell was designed to detect the chemical conditions associated with the ailment. For instance, rather than spraying engineered bacteria into the lungs of cystic fibrosis patients, artificial cells could be built to detect the presence of *Pseudomonas aeruginosa* biofilms through the quorum signalling molecules that are naturally secreted by the organism, such as *N*-(3-oxododecanoyl)-L-homoserine lactone, a molecule capable of crossing membranes without the aid of transporters. Subsequently, the artificial cells could release small molecules, for example, D-amino acids²³, to disperse the biofilm and thus clear the infection. Moreover, the use of dispersion rather than killing would decrease the probability of the bacteria developing resistance. Similar strategies with artificial cells could be developed to substitute for engineered probiotics that integrate with gut microbiota²⁴ and prevent disease^{25,26}.

There are several limitations to these first generation artificial cells. First, heterogeneity in membrane lamellarity and in encapsulation efficiency²⁷ results in a mixture of artificial cells with varying degrees of activity. Microfluidic-based methods for compartment formation and solute encapsulation would likely alleviate many of the complications associated with vesicle-to-vesicle and batch-to-batch variability. Also, a system fully dependent upon the permeability properties of the membrane limits the types of molecules that can be sensed and released. The development of specific membrane-associated sensors and transporters will likely be necessary as the complexity of artificial cells increase. Finally, the simple release of encapsulated molecules means that release could result from compartment degradation as opposed to an engineered response to the detection of a specific molecule. It is, therefore, important to develop an output that is mediated by synthesis so that compartment degradation would only result in the release of inactive starting molecules. An example of such a system is the biological nanofactory described by Fernandes *et al.*²⁸ that synthesizes a signalling molecule from *S*-(5′-deoxyadenosin-5′)-L-homocysteine via two enzymatic steps.

The absence of a living chassis opens up greater opportunities to assemble or biofabricate various mechanisms or functions that would be difficult to implement with living cells. For example, chemical systems housed within inorganic and peptide-based compartments are capable of sensing the environment through, in part, the gating behaviour of the non-lipid compartment^{29,30}. Further, artificial cells can synthesize and release signalling molecules sensed by living cells without exploiting genetically encoded parts^{31,32}. The possibility of merging advances with non-genetically encoded and genetically encoded parts may lead to the construction of artificial cells that are better able to imitate natural cellular life^{33,34}.

Methods

Genetic constructs. The gene encoding *Staphylococcus aureus* α HIL was synthesized by Genscript. Super folder GFP (BBa_I746916) was from the registry of standard biological parts (<http://parts.igem.org>). The theophylline riboswitch

sequence was from Lynch and Gallivan¹⁰ and was amplified from a previously described construct¹¹. All genes were subcloned into pET21b (Novagen) with NdeI and XhoI restriction sites. Mutagenesis was performed by Phusion site-directed mutagenesis (Thermo Scientific). All constructs were confirmed by sequencing at Genechiron or Eurofins MWG Operon. Sequences of all the exploited constructs are listed in Supplementary Table 1. All experiments were repeated at least three times. Data are reported as averages with standard error, or representative runs are shown.

In vitro characterization of the riboswitch. Plasmids were amplified in *E. coli* Novablue (Novagen) and purified with Wizard Plus SV Minipreps DNA Purification System (Promega). Plasmid DNA was phenol–chloroform extracted, ethanol precipitated and resuspended in deionized and diethyl pyrocarbonate-treated water. PCR products were purified with Wizard Plus SV Gel and PCR Clean-Up Systems (Promega). Transcription–translation reactions used the PURExpress *In Vitro* Protein Synthesis Kit (New England Biolabs) supplemented with 20 units of Human Placenta RNase Inhibitor (New England Biolabs). Reactions were monitored by fluorescence with a CFX96 Touch real-time PCR (Bio-Rad) using the SYBR green filter set.

α -hemolysin activity. Each construct was expressed with the PURExpress *In Vitro* Protein Synthesis Kit at 37 °C in a final volume of 25 μ l either in the presence or absence of 1.5 mM theophylline for 6 h. Rabbit red blood cell (RBC) suspensions (adjusted to $D = 0.1$ at 650 nm) were added to a microplate where the reaction mixtures were serially diluted. Changes in attenuation of the RBC suspension were measured at 650 nm with a microplate reader (UVmax, Molecular Devices) for 30 min at 22 °C as reported in Laventie *et al.*³⁵ The results are reported as percentage of haemolysis or as the time necessary to reach 50% of haemolysis.

Preparation of *E. coli* receiver cells. Mid-exponential *E. coli* BL21(DE3) pLysS transformed with a plasmid encoding super folder GFP behind a T7 promoter and a *lac* operator sequence (CD101A¹²) were grown in LB supplemented with 100 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol to an optical density of 0.5 at 600 nm. A quantity of 200 μ l aliquots in 10% (vol/vol) glycerol were flash frozen with liquid nitrogen and stored at –80 °C for later use. Aliquots were rapidly thawed and mixed with 2 ml LB supplemented with 100 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol and incubated for 2 h at 37 °C with 220 r.p.m. shaking. Finally, the cells were gently pelleted and resuspended in 1 ml M9 minimal media.

Preparation of artificial cells. Vesicles were prepared as previously described^{36,37}. Briefly, 12.5 mg 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and 12.5 mg cholesterol (Avanti Polar Lipids) in chloroform were mixed in a round bottom flask. A thin lipid film was made through rotary evaporation with a Buchi Rotovapor R-210 equipped with a Buchi Vacuum Pump V-700 for 5 h. A quantity of 2 ml DEPC-treated deionized water was then added to the thin lipid film and vigorously vortexed. The resulting liposome dispersion was homogenized with an IKA T10 basic homogenizer at a power setting of 4 for 1 min. A quantity of 100 μ l aliquots were frozen in liquid nitrogen or dry ice and lyophilized overnight in a vacuum concentrator (Centriprep DNA concentrator, Labconco) at 40 °C. The lyophilized empty liposomes were stored at –20 °C. A quantity of 100 μ l aliquots of freeze-dried liposomes were hydrated with 25 μ l of 100 mM IPTG (Sigma) dissolved in 50 mM HEPES pH7.6, 25 μ l of the PURE system, 500 ng DNA and 20 units of human placenta RNase inhibitor (final volume of 50 μ l), unless otherwise noted. Solutions were gently mixed for 30 s.

To remove extravesicular material, the vesicles were dialyzed following a method previously described by Zhu and Szostak³⁸. The original membranes of 500 μ l Slide-a-Lyzer dialysis cassettes (Pierce) were exchanged with 25 mm diameter polycarbonate track-etched membranes with a 1 μ m pore size (Whatman). A quantity of 50 μ l of unpurified vesicles were loaded onto the center of the dialysis system with a 100 μ l Hamilton syringe and dialyzed against 250 ml of buffer A (50 mM HEPES, 10 mM MgCl₂, 100 mM KCl, pH 7.6) by stirring. The first four rounds of dialysis were for 10 min each. Two more rounds of dialysis in which the buffer was changed after 30 min incubations were further performed. All dialysis steps were carried out at 4 °C.

Artificial-natural cell communication. Purified vesicles containing DNA, the PURE system, and IPTG were incubated with *E. coli* BL21(DE3) pLysS transformed with CD101A in M9 minimal media supplemented with 1 mg ml⁻¹ of Proteinase K and 5 mM theophylline at 37 °C in a final volume of 40 μ l. Control reactions did not contain theophylline. At different time points, 1 μ l was removed and diluted 1:100 in PBS. The sample was then analysed by flow cytometry with a FACSCanto A (BD Biosciences). The FITC filter was used for the detection of positive cells. The incident light was at 488 nm for forward scatter (FSC), side scatter (SSC) and fluorescence. Detection for SSC and fluorescence was at 488 \pm 10 nm and 530 \pm 30 nm, respectively. The threshold parameters were 200 for both FSC and SSC. The PMT voltage settings were 525 (FSC), 403 (SSC) and 600 (FITC). The flow rate was set to 'low'. For each sample 30,000 events were

collected. Reactions were repeated three times on three separate days. Data were analysed using FlowJo software (TreeStar, USA).

Samples were also evaluated by RT-qPCR. Here, the dialyzed vesicles and *E. coli* were incubated as described above for 4 h at 37 °C. Subsequently, the total RNA was extracted with the RNeasy Mini kit (Qiagen). A quantity of 10 µl of 500 ng of RNA was reverse transcribed using RevertAid Reverse Transcriptase (Thermo Scientific). cDNA was quantified with a CFX96 Touch real-time PCR (Bio-Rad) with SYBR green detection. Each sample was diluted to 5 ng and measured in triplicate in a 96 wells plate (Bio-Rad) in a reaction mixture containing SoAdvanced SYBR green supermix (Bio-Rad) and 180 nM of each primer in a 10 µl finale volume. The primers used to quantify *lacZ*, *lacY* and *lacA* expression were *lacZ* FW: 5'-TACGATGCGCCCATCTACAC-3', *lacZ* REV: 5'-AACAAACC GTCGGATTCTCC-3', *lacY* FW: 5'-GGTTTCCAGGGCGCTTATCT-3', *lacY* REV: 5'-TTCATTCACTGACGACGCA-3', *lacA* FW: 5'-GCGTCACCATC GGGGATAAT-3', *lacA* REV: 5'-CCACGACGTTTGGTGAATG-3'. Gene expression was normalized to the expression of *idnT*³⁹ with the following primers: 5'-CTGCCGTTGCGTGTATTATT-3' and 5'-GATTGCTCGATGGTGCCTC-3'.

References

- Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
- Chen, Y. J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).
- Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **110**, 14024–14029 (2013).
- Smolke, C. D. & Silver, P. A. Informing biological design by integration of systems and synthetic biology. *Cell* **144**, 855–859 (2011).
- Albert, R., Collins, J. J. & Glass, L. Introduction to focus issue: quantitative approaches to genetic networks. *Chaos* **23**, 025001 (2013).
- Regot, S. *et al.* Distributed biological computation with multicellular engineered networks. *Nature* **469**, 207–211 (2011).
- Isaacs, F. J. *et al.* Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
- Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
- Maurer, L. M., Yohannes, E., Bondurant, S. S., Radmacher, M. & Slonczewski, J. L. pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *J. Bacteriol.* **187**, 304–319 (2005).
- Lynch, S. A. & Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.* **37**, 184–192 (2009).
- Martini, L. & Mansy, S. S. Cell-like systems with riboswitch controlled gene expression. *Chem. Commun.* **47**, 10734–10736 (2011).
- Lentini, R. *et al.* Fluorescent proteins and *in vitro* genetic organization for cell-free synthetic biology. *ACS Synth. Biol.* **2**, 482–489 (2013).
- Cassidy, P. & Harshman, S. Studies on the binding of staphylococcal 125I-labeled alpha-toxin to rabbit erythrocytes. *Biochemistry* **15**, 2348–2355 (1976).
- Mantri, S., Sapra, K. T., Cheley, S., Sharp, T. H. & Bayley, H. An engineered dimeric protein pore that spans adjacent lipid bilayers. *Nat. Commun.* **4**, 1725 (2013).
- Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
- Prindle, A. *et al.* A sensing array of radically coupled genetic 'biopixels'. *Nature* **481**, 39–44 (2012).
- Weber, W. & Fussenegger, M. Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.* **13**, 21–35 (2012).
- Kostka, J. E. *et al.* Hydrocarbon-degrading bacteria and the bacterial community response in gulf of Mexico beach sands impacted by the deepwater horizon oil spill. *Appl. Environ. Microbiol.* **77**, 7962–7974 (2011).
- Xiang, S., Fruehauf, J. & Li, C. J. Short hairpin RNA-expressing bacteria elicit RNA interference in mammals. *Nat. Biotechnol.* **24**, 697–702 (2006).
- Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.* **355**, 619–627 (2006).
- Gupta, S., Bram, E. E. & Weiss, R. Genetically programmable pathogen sense and destroy. *ACS Synth. Biol.* **12**, 715–723 (2013).
- Saeidi, N. *et al.* Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol. Syst. Biol.* **7**, 521 (2011).
- Kolodkin-Gal, I. *et al.* D-amino acids trigger biofilm disassembly. *Science* **328**, 627–629 (2010).
- Kotula, J. W. *et al.* Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl Acad. Sci. USA* **111**, 4838–4843 (2014).
- Duan, F. & March, J. C. Engineered bacterial communication prevents *Vibrio cholerae* virulence in an infant mouse model. *Proc. Natl Acad. Sci. USA* **107**, 11260–11264 (2010).
- Goh, Y. L., He, H. & March, J. C. Engineering commensal bacteria for prophylaxis against infection. *Curr. Opin. Biotechnol.* **23**, 924–930 (2012).
- Weitz, M. *et al.* Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator. *Nat. Chem.* **6**, 295–302 (2014).
- Fernandes, R., Roy, V., Wu, H. C. & Bentley, W. E. Engineered biological nanofactories trigger quorum sensing response in targeted bacteria. *Nat. Nanotechnol.* **5**, 213–217 (2010).
- Li, M., Harbron, R. L., Weaver, J. V., Binks, B. P. & Mann, S. Electrostatically gated membrane permeability in inorganic protocells. *Nat. Chem.* **5**, 529–536 (2013).
- Huang, X. *et al.* Interfacial assembly of protein-polymer nano-conjugates into stimulus-responsive biomimetic protocells. *Nat. Commun.* **4**, 2239 (2013).
- Gardner, P. M., Winzer, K. & Davis, B. G. Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria. *Nat. Chem.* **1**, 377–383 (2009).
- Gupta, A. *et al.* Encapsulated fusion protein confers 'sense and respond' activity to chitosan-alginate capsules to manipulate bacterial quorum sensing. *Biotechnol. Bioeng.* **110**, 552–562 (2013).
- Cronin, L. *et al.* The imitation game—a computational chemical approach to recognizing life. *Nat. Biotechnol.* **24**, 1203–1206 (2006).
- Forlin, M., Lentini, R. & Mansy, S. S. Cellular imitations. *Curr. Opin. Chem. Biol.* **16**, 586–592 (2012).
- Laventie, B. J. *et al.* p-Sulfonato-calix[n]arenes inhibit staphylococcal bicomponent leukotoxins by supramolecular interactions. *Biochem. J.* **450**, 559–571 (2013).
- Yu, W. *et al.* Synthesis of functional protein in liposome. *J. Biosci. Bioeng.* **92**, 590–593 (2001).
- Spencer, A. C., Torre, P. & Mansy, S. S. The encapsulation of cell-free transcription and translation machinery in vesicles for the construction of cellular mimics. *J. Vis. Exp.* **80**, e51304 (2013).
- Zhu, T. F. & Szostak, J. W. Preparation of large monodisperse vesicles. *PLoS ONE* **4**, e5009 (2009).
- Zhou, K. *et al.* Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Mol. Biol.* **12**, 18 (2011).

Acknowledgements

We thank the Armenise-Harvard foundation, the Autonomous Province of Trento (Ecomm), CIBIO, NSF (CBET#1160005, CBET#1264509), and DTRA (HDTRA1-13-0037) for financial support. A.C.S. was supported by the Marie-Curie Trentino COFUND from the Autonomous Province of Trento. We also thank M. Pizzato, P. Torre and F. Ausubel for helpful discussions, and H. Bayley for α HL construct suggestions.

Author contributions

Design, cloning and mutagenesis of genetic constructs were done by R.L., A.C.S., J.F., S.P.S., M.F., and C.D.B. *In vitro* riboswitch activity was investigated by R.L., S.P.S., C.D.B., L.M., M.F. and A.C.S. α HL activity was measured by R.L., S.P.S., M.M., and M.D.S. R.L., J.L.T., D.C., F.C. and S.P.S. ran the cell flow cytometry experiments, and RT-qPCR was performed by R.L. and J.F. S.S.M. supervised the project. All authors analysed and interpreted the data and contributed to the writing of the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

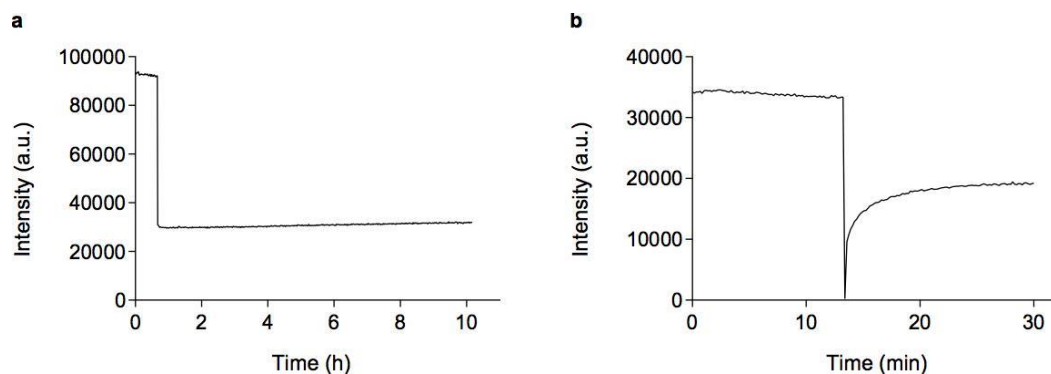
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Lentini, R. *et al.* Integrating artificial with natural cells to translate chemical messages that direct *E. coli* behaviour. *Nat. Commun.* **5**:4012 doi: 10.1038/ncomms5012 (2014).



This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>



Supplementary Figure 1. Vesicle permeability to IPTG. (A) A shrink-swell²⁻⁴ assay was used to assess whether IPTG was capable of crossing vesicle membranes. POPC:cholesterol vesicles with entrapped calcein were prepared in 10 mM MgCl₂, 100 mM KCl, 50 mM HEPES, pH 7.6 as described in the methods and subsequently purified by gel filtration chromatography with sepharose-4b. An aliquot of the vesicle sample was diluted two-fold with 1.0 M IPTG (final concentration = 0.5 M) at 37 °C. The reaction was monitored by fluorescence with excitation and emission wavelengths of 495 nm and 515 nm, respectively. The rapid decrease in fluorescence was due to both dilution with the solute solution and calcein self-quenching. If IPTG were capable of crossing the membrane, a recovery of fluorescence would have been observed. (B) The permeability of POPC vesicles to ribose was observable with the shrink-swell assay. The recovery in fluorescence after two-fold dilution with 1.0 M ribose (final concentration = 0.5 M) was due to equilibration of ribose and water across the membrane.

Supplementary Table 1. DNA sequences used in this study

NAME	NOTE	SEQUENCE
AS014A	T7 promoter, K30S E31T αHL-sfGFP	ATTTAATACGACTCACTATAG ATG GATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTC CAATACCACCGTTAAAACCGGTGATCTGGTGACCTATGATTCTACCAACGGTATGCATAAAAAAGTGT TTTACTCGTTTATTGACGATAAAAACCATAACAAAAAAGTCTGGTTCATCCGCACCAAAAGGCACCATG CGGGTCAATACCGTGTGACTCCGAAGAAGGTGCGAACAAAAGCGGTCTGGCTTGGCCGTCTGCCTT TAAAGTGCAGCTGCAACTGCCGATAATGAAGTGGCGCAGATTTCAGATTATTATCCGCGTAATAGCA TCGATACCAAAGAATATATGAGTACCCTGACCTATGGTTTTAATGGCAATGTTACCGGTGATGATACG GGTAAAATTGGCGGTCTGATTGGCGCAATGTGCCATTGGTCATACGCTGAAATACGTGCAACCCG ATTTCAAACCATTCTGAAAAGTCCGACCGATAAAAAAGTGGTTGGAAAAGTTATCTTCAACAACATG GTGAATCAGAACTGGGGTCCGTACGATCGCGATTCTGGAATCCGGTTTATGGCAATCAGCTGTTTAT GAAAACCGCAACGGTAGTATGAAAGCGCGGATAATTTCTGGACCCGAACAAAGCCTCAAGCCTG CTGTCCAGCGGTTTTAGCCCGGATTTGCCACGGTATTACCATTGGATCGAAAGCCAGCAAAACAGCA GACCAACATTGATGTGATACGAACTGTCGCGTGTGATGATTATCAACTGCATTGGACCTCAACCAATT GGAAAGGCACCAATACCAAAGATAAATGGACGGATCGCAGTTCAGAACGCTACAAAATTGATTGGGA AAAAGAAGAATGACCAACGGATCCGGCAGCGTTCT ATG CGTAAAGCGAAGAGCTGTTCACTGG TGTCGTCCTATTCTGGTGGAACTGGATGGTGTGCAACGGTCATAAGTTTTCCGTGCGTGGCGAGG GTGAAGGTGACGCAACTAATGGTAACTGACGCTGAAGTTCATCTGTACTACTGGTAACTGCCGGTA CCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTCTTGTCTGTTATCCGGACCATATG AAGCAGCATGACTTCTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACGCACGATTTCTTTAA GGATGACGGCACGTACAAAACCGTGCAGGAAAGTAAATTTGAAGGCGATACCTGGTAAACCGCATT GAGCTGAAAGGCATTGACTTTAAGAAGACGGCAATATCTGGGCCATAAGCTGGAAATCAATTTTA ACAGCCACAATGTTTACATCACCGCGATAAACAAAAAAATGGCATTAAAGCGAATTTTAAATTCGC CACACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATG GTCCTGTTCTGTCCAGACAATCACTATCTGAGCACGAAAGCGTCTGTCTAAAGATCCGAACGAG AAACGCGATCATATGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGATGGTATGATGAAC TGTAACA TAA CTCGAGCACCACCACCACCACCTGAGATCCGGCTGCTAACAAAGCCGAAAGGA AGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATAACCCCTTGGGGCTCTAACCGGGTCT TGAGGGGTTTTTTG
DT101A	T7 promoter, αHL- His tag	TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCTCTAGAAATAATTTGTTAACT TTAAGAAGGAGATACAT ATG GATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTCC AATACCACCGTTAAAACCGGTGATCTGGTGACCTATGATAAAGAAAACGGTATGCATAAAAAAGTGT TTACTCGTTTATTGACGATAAAAACCATAACAAAAAAGTCTGGTTCATCCGCACCAAAAGGCACCATG GGGTCAATACCGTGTGACTCCGAAGAAGGTGCGAACAAAAGCGGTCTGGCTTGGCCGTCTGCCTT AAAGTGCAGCTGCAACTGCCGATAATGAAGTGGCGCAGATTTCAGATTATTATCCGCGTAATAGCAT CGATACCAAAGAATATATGAGTACCCTGACCTATGGTTTTAATGGCAATGTTACCGGTGATGATACGG GTAAAATTGGCGGTCTGATTGGCGCAATGTGTCCATTGGTCATACGCTGAAATACGTGCAACCGGAT TTCAAACCACTTCTGAAAAGTCCGACCGATAAAAAAGTGGTTGGAAAGTTATCTTCAACAACATGGT GAATCAGAACTGGGGTCCGTACGATCGGATTCTGGAATCCGGTTTATGGCAATCAGCTGTTTATGA AAACCGCAACGGTAGTATGAAAGCGGCGATAATTTCTGGACCGAACAAGCCTCAAGCCTGCT GTCCAGCGTTTTAGCCCGGATTTTCCACGGTATTACCATGGATCGCAAAGCCAGCAAAACAGCAGA CCAACATTGATGTGATCTACGAACGTGTGCGTGTGATTATCAACTGCATTGGACCTCAACCAATTGG AAAGGCACCAATACCAAAGATAAATGGACGGATCGCAGTTCAGAACGCTACAAAATTGATTGGGAAA AAGAAGAAATGACCAACCTCGAGCACCACCACCACCACCT TGA GATCCGGCTGCTAACAAAGCCCG AAAGGAAGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATAACCCCTTGGGGCTCTAAAC GGGTCTTGGGGTTTTTTG
JF001A	T7 promoter, <u>theophylline</u> <u>riboswitch</u> αHL	AATTAATACGACTCACTATAG GGT GATACCAGCATCGTCTT GATGCCCTTGGCAGCACCTGCTAAGG TAACAACAAGAT GATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTCCAATACCACCG TTAAAACCGGTGATCTGGTGACCTATGATAAAGAAAACGGTATGCATAAA AAAGTGTTTACTCGTTT ATTGACGATAAAAACCATAACAAAAAAGTCTGGTTCATCCGCACCAAAAGGCACCAATTGCGGGTCAATA CCGTGTGACTCCGAAGAAGGTGCGAACAAAAGCGGTCTGGCTTGGCCGTCTGCCTTTAAAGTGACG CTGCAACTGCCGATAATGAAGTGGCGCAGATTTCAGATTATTATCCGCGTAATAGCATCGATACCAA

		<p>AGAATATATGAGTACCTGACCTATGGTTTTAATGGCAATGTTACCGGTGATGATACGGGTAAAATTG GCGGTCTGATTGGCGCAATGTGCCATTGGTCATACGCTGAAATACGTGCAACCGGATTTCAAACCC ATTCTGGAAGTCCGACCGATAAAAAAGTGGGTTGAAAGTTATCTTCAACAACATGGTGAATCAGA ACTGGGGTCCGTACGATCGCGATTCTGGAATCCGGTTTATGGCAATCAGCTGTTTATGAAAACCCGC AACGGTAGTATGAAAGCGCGGATAATTTCTGGACCCGAACAAGCCTCAAGCTGCTGTCAGCG GTTTTAGCCCGATTTTCCACGGTTATTACCATGGATCGCAAAGCCAGCAAACAGCAGACCAACATT GATGTGATCTACGAACGTGTGCGTGATGATTATCAACTGCATTGGACCTCAACCAATTGAAAAGGCAC CAATACCAAAGATAAATGGACGGATCGCAGTTCAGAACGCTACAAAATTGATTGGGAAAAAGAA ATGACCAACTAACTCGAGCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCGAAAGG AAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCTCTAACGGGTC TTGAGGGGTTTTTG</p>
RL067A	T7 promoter, α HL	<p>TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATCCCTCTAGAATAATTTGTTAACT TTAAGAAGGAGATATACATATGATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTCC AATACCACGGTAAAACCGGTGATCTGGTGACCTATGATAAAGAAAACGGTATGCATAAAAAAGTGT TTACTCGTTTATTGACGATAAAAAACATAACAAAAACTGCTGGTCATCCGCACCAAAGGCACATTGC GGGTCAATACCGTGTACTCCGAAGAAGTGCGAACAAAAGCGGTCTGGCTTGGCCGTCTGCCTTT AAAGTGCAGCTGCAACTGCCGATAATGAAGTGGCGCAGATTTCAGATTATTATCCGCGTAATAGCAT CGATACCAAAGAATATATGAGTACCCCTGACCTATGGTTTTAATGGCAATGTTACCGGTGATGATACGG GTAAAATTGGCGGTCTGATTGGCGCAATGTGTCCATTGGTCATACGTGAAATACGTGCAACCGGAT TTCAAACCACTTCTGAAAAGTCCGACCGATAAAAAAGTGGGTTGAAAAGTTATCTTCAACAACATGGT GAATCAGAACTGGGGTCCGTACGATCGCGATTCTGGAATCCGGTTTATGGCAATCAGCTGTTTATGA AAACCCGCAACGGTAGTATGAAAAGCGCGGATAATTTCTGGACCCGAACAAGCCTCAAGCCTGCT GTCCAGCGTTTTAGCCCGGATTTTCCACGGTTATTACCATGGATCGCAAAGCCAGCAAACAGCAGA CCAACATTGATGTGATCTACGAACGTGTGCGTGATGATTATCAACTGCATTGGACCTCAACCAATTGG AAAGGCACCAATACCAAAGATAAATGGACGGATCGCAGTTCAGAACGCTACAAAATTGATTGGGAAA AAGAAGAAATGACCAACTAACTCGAGCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGC CCGAAAAGGAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCTCT AAACGGGTCTTGGGGTTTTTG</p>
RL069A	T7 promoter, <u>theophylline</u> <u>riboswitch</u> , K30S E31T α HL	<p>AATTAATACGACTCACTATAGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTGCTAAGG <u>TAACAACAAGATG</u>ATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTCAATACCACC GTTAAAACCGGTGATCTGGTGACCTATGATTCTACCAACGGTATGCATAAAAAAGTGTTTACTCGTTT ATTGACGATAAAAAACCATACAAAAAAGTCTGGTTCATCCGCACCAAAGGCACCAATTCGGGTCAATA CCGTGTGACTCCGAAGAAGGTGCGAACAAGCGGTCTGGCTTGGCCGTCTGCCTTTAAAGTGAG CTGCAACTGCCGATAATGAAGTGGCGCAGATTTCAGATTATTATCCGCGTAATAGCATCGATACCAA AGAATATATGAGTACCCTGACCTATGGTTTTAATGGCAATGTTACCGGTGATGATACGGGTAATAATTG GCGGTCTGATTGGCGCAATGTGTCCATTGGTCATACGCTGAAATACGTGCAACCGGATTTCAAACCC ATTCTGAAAAGTCCGACCGATAAAAAAGTGGGTTGAAAAGTTATCTTCAACAACATGGTGAATCAGA ACTGGGTCCGTACGATCGCGATTCTGGAATCCGGTTTATGGCAATCAGCTGTTTATGAAAACCCGC AACGGTAGTATGAAAGCGCGGATAATTTCTGGACCCGAACAAGCCTCAAGCCTGCTGTCCAGCG GTTTTAGCCCGGATTTTCCACGGTTATTACCATGGATCGCAAAGCCAGCAAACAGCAGACCAACATT GATGTGATCTACGAACGTGTGCGTGATGATTATCAACTGCATTGGACCTCAACCAATTGAAAAGGCAC CAATACCAAAGATAAATGGACGGATCGCAGTTCAGAACGCTACAAAATTGATTGGGAAAAAGAA ATGACCAACGGATCCGGCAGCGTCTATGCGTAAAGGCGAAGAGCTGTTCACTGGTGTCTCCCTA TTCTGGTGAACCTGGATGGTGTGATGCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGA CGCAACTAATGGTAACTGACGCTGAAGTTCATCTGTACTACTGGTAACTGCCGGTACCTTGGCCGA CTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTGTTATCCGGACCATATGAAGCAGCAT GACTTCTCAAGTCCGCATGCCGAAGGCTATGTGCAGGAACGCACGATTTCTTTAAGGATGACGG CACGTACAAAACGCGTGCAGAAAGTGAATTTGAAGGCGATACCTGGTAAACCGCATTGAGCTGAAA GGCATTGACTTTAAAGAAGACGGCAATATCCTGGGCATAAGCTGGAATACAATTTAACAGCCACAA TGTTTACATCACCGCGATAAACAAAAAATGGCATTAAAGCGAATTTAAAATTCGCCACAACGTGG AGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACCTCAATCCGGTGTGTCCTGTTCTG CTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAACGCGATCA TATGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAAGTGTACAAAATAA</p>

		CTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACCGGGTCTTGAGGGGTTTT TTG
SP002A	T7 promoter, <u>αHL-sfGFP</u>	ATTTAATACGACTCACTATAG ATG GATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTC CAATACCACCGTTAAACCGGTGATCTGGTGACCTATGATAAAGAAAACGGTATGCATAAAAAAGTGT TTTACTCGTTTATTGACGATAAAAAACATAACAAAAAAGCTGGTTCATCCGCACCAAAGCACCATTG CGGGTCAATACCGTGTGACTCCGAAGAAGGTGCGAACAAAAGCGGTCTGGCTTGGCCGTCTGCCTT TAAAGTGAGCTGCAACTGCCGATAATGAAGTGCGCAGATTTAGATTATTATCCGCGTAATAGCA TCGATACCAAAGAATATAGTACCCTGACCTATGTTTTAATGGCAATGTTACCGGTGATGATACG GGTAAAATTGGCGGTCTGATTGGCGCAATGTGTCCATTGGTCATACGCTGAAATACGTGCAACCGG ATTTCAAACCATTTCTGAAAAGTCCGACCGATAAAAAAGTGGGTTGAAAAGTTATCTTCAACAACATG GTGAATCAGAACTGGGGTCCGTACGATCGCGATTCTGGAATCCGGTTTATGGCAATCAGCTGTTAT GAAAACCGCAACGGTAGTATGAAAGCGCGGATAATTTCTGGACCCGAACAAAGCCTCAAGCCTG CTGTCCAGCGGTTTTAGCCCGGATTTGCCACGGTATTACCATGGATCGAAAGCCAGCAACAGCA GACCAACATTGATGTGATCTACGAACGTGTGCGTGATGATTCAACTGCATTGGACCTCAACCAATT GAAAAGGCACCAATACCAAAGATAAATGGACGGATCGCAGTTGAGAACGCTACAAAAGTACTGGGA AAAAGAAGAAATGACCAACGGATCCGGCAGCGGTTCT ATG CGTAAAGGCGAAGAGCTGTTCACTGG TGTCGTCCTATTCTGGTGAAGTGGATGGTGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGCGAGG GTGAAGGTGACGCAACTAATGGTAACTGACGCTGAAGTTCATCTGACTACTGGTAACTGCCGGTA CCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTGTCTGTTATCCGGACCATATG AAGCAGCATGACTTCTCAAGTCCGCCATGCCGGAAGGCTATGTGAGGAAACGACGATTCTCTTTAA GGATGACGGCAGCTACAAAACCGTGCAGGAAAGTAAATTTGAAGGCGATACCTGGTAAACCGCATT GAGCTGAAAGGCATTGACTTTAAGAAGACGGCAATATCCTGGGCATAAGCTGGAATACAATTTTA ACAGCCACAATGTTTACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAATTCGC CACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATG GTCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGAAAGCGTTCTGTCTAAAGATCCGAACGAG AAACCGCATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGCATCAGCATGGTATGGATGAAC TGTACAAA TAA CTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCGAAAGGA AGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACCGGGTCT TGAGGGGTTTTTTG
SP011A	T7 promoter, <u>theophylline</u> <u>riboswitch</u> , <u>αHL-sfGFP</u>	AATTAATACGACTCACTATAGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTGCTAAGG <u>TAACAACAAGATG</u> GATTCTGATATCAATATCAAACCGGCACCACCGATATCGGCTCAATACCAC GTTAAACCGGTGATCTGGTGACCTATGATAAAGAAAACGGTATGCATAAAAAAGTGTTTTACTCGTT TATTGACGATAAAAAACATAACAAAAAAGCTGGTTCATCCGCACCAAAGCACCATTGCGGGTCAAT ACCGTGTGACTCCGAAGAAGGTGCGAACAAAAGCGGTCTGGCTTGGCCGTCTGCCTTTAAAGTGCA GCTGCAACTGCCGATAATGAAGTGCGCAGATTTAGATTATTATCCGCGTAATAGCATCGATACCA AAGAATATAGTACCCTGACCTATGTTTTAATGGCAATGTTACCGGTGATGATACGGGTAATAAT GGCGTCTGATTGGGCCAATGTGCCATTGGTCATACGCTGAAATACGTGCAACCGGATTTCAAAC CATTCTGAAAGTCCGACCGATAAAAAAGTGGGTTGAAAAGTTATCTTCAAACAACGGTGAATCAG AACTGGGGTCCGTACGATCGCGATTCTGGAATCCGGTTTTATGGCAATCAGCTGTTTATGAAAACCG CAACGGTAGTATGAAAGCGCGGATAATTTCTGGACCCGAACAAAGCCTCAAGCTGCTGTCCAGC GGTTTTAGCCCGGATTTGCCACGGTATTACCATGGATCGAAAGCCAGCAACAGCAGACCAACAT TGATGTGATCTACGAACGTGTGCGTGATGATTATCAACTGCATTGGACCTCAACCAATTGAAAGGCA CCAATACAAAGATAAATGGACGGATCGCAGTTGAGAACGCTACAAAATTGATTGGGAAAAAGGA AATGACCAACCGGATCCGGCAGCGGTTCT ATG CGTAAAGGCGAAGAGCTGTTCACTGGTGTGCTCCCT ATTCTGGTGAAGTGGTGGTGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGCCGAGGGTGAAGGTG ACGCAACTAATGGTAACTGACGCTGAAGTTTACTGTACTACTGGTAACTGCCGTACCTTGGCCG ACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTGTTATCCGGACCATATGAAGCAGCAT GACTTCTCAAGTCCGCCATGCCGGAAGGCTATGTGAGGAACGACGATTTCTTTAAGGATGACGG CAGTACAAAACGCGTGCAGGAAAGTAAATTTGAAGGCGATACCTGGTAAACCGCATTGAGCTGAAA GGCATTGACTTTAAAGAAGACGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCCACA TGTTTACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAATTCGCCACAACGTGG AGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATGGTCTGTTCTG

		CTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAACGCGATCA TATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAACTGTACAA TAA CTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTGGGGCCTCTAACGGGTCTTGAGGGGTTTT TTG
--	--	---

Start and stop codons are in bold and the theophylline riboswitch is underlined. The T7 promoter and T7 terminator sequences are TAATACGACTCACTATA and CTAGCATAACCCCTGGGGCCTCTAACGGGTCTTGAGGGGTTTTTTG, respectively.

Supplementary Table 2. The activity of cell-free expressed α HL

construct name	$t_{1/2}$ (min)	Comments
JF001A	> 30	α HL behind theophylline riboswitch in the absence of theophylline
JF001A	16.5	α HL behind theophylline riboswitch in the presence of theophylline
DT101A	> 30	α HL-His tagged
SP002A	4.5	α HL-GFP
RL067A	10.0	α HL
Sigma-Aldrich α HL	9.5	commercial α HL
CD101A ¹	> 30	GFP
AS014A	> 30	K30S E31T α HL-GFP lacking an internal RBS

Each construct was expressed *in vitro* with the PURE system and subsequently added to rabbit red blood cells. Hemolysis was measured by attenuation as described in the methods. When indicated, the theophylline concentration was 1.5 mM. Sigma-Aldrich α HL indicates purchased purified protein and was not *in vitro* expressed.

Supplementary References

1. Lentini R., *et al.* Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. *ACS Synth. Biol.* **2**, 482-489 (2013).
2. Bittman R. & Blau L. Permeability behavior of liposomes prepared from fatty acids and fatty acid methyl esters. *Biochim. Biophys. Acta* **863**, 115-120 (1986).
3. Chen P. Y., Pearce D. & Verkman A. S. Membrane water and solute permeability determined quantitatively by self-quenching of an entrapped fluorophore. *Biochemistry* **27**, 5713-5718 (1988).
4. Sacerdote M. G. & Szostak J.W. Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proc. Natl Acad. Sci. USA.* **102**, 6004-6008 (2005).

Chapter 3.

***In vitro* selection of new RNA- based sensors by mRNA display**

In this chapter expanding the sensing ability of riboswitches towards new ligand-responsive devices is presented. Initially the main methodologies that are used to develop and engineer riboregulators are introduced. Both *in vivo* and *in vitro* selection approaches have been carried out in the last twenty years, for functional nucleic acids and novel protein evolution. In particular, in the chapter the focus is set on a protein selection technique, i.e. mRNA display. The efforts for the *in vitro* selection via mRNA display of a translational control riboswitch responsive to the small molecule malachite green are presented. Preliminary experiments, technique development for RNA application and the *in vitro* selection performed are described and discussed in detail. At the end of the selection, next generation sequencing (NGS) data do not converge on a specific sequence. Thus a different approach can be considered to achieve the *in vitro* selection of new riboswitches (see Chapter 4).

3.1 *In vitro* and *in vivo* selection of functional nucleic acids

The developing of functional nucleic acids, as aptamers and riboregulators, has been well established by twenty years of studies.⁶²⁻⁶⁴ The polymerase chain reaction and discovery of reverse transcriptase made possible the first aptamer selection in the 1990s.⁶⁵ From that time on, RNA and DNA aptamers, ribozymes, and riboswitches have been selected by *in vivo* or *in vitro* methods, allowing the emergence of new functionalities as specific ligand binding or catalysis. Generally, a selection scheme is based on linking genotypic information with a desired phenotype. In other words, in a selection process, starting from large libraries of sequences, only those molecules that encode a specific function are enriched through the cycles. For example, a RNA able to bind ATP can be selected from a pool of sequences upon iterative cycles of incubation with the ligand.⁶⁵ Separation of functional molecules from un-functional molecules allows the enrichment.

In vivo selection methods rely on cell population transformation. Clones are tested for a specific phenotype. For example, cell motility and fluorescence-activated cell sorting (FACS)⁵⁵ can be used to distinguish functional sequences. This approach was mainly involved in riboswitch and aptazymes selections.^{58,66} The theophylline riboswitch was *in vivo* selected assessing β -galactosidase production in *E. coli*.⁶⁶ Moreover, a natural riboswitch repressing translation in presence of thiamine pyrophosphate (TPP) was reengineered to activate protein expression when binding the ligand.⁶⁷ The conversion was achieved by a dual selection based on the *tetA* gene.

If the construction of functional devices are to be exploited for the modification of cell behavior, then RNA sequences derived from *in vivo* selections are preferred, since *in vitro* selected RNA sequences may not be functional under *in vivo* conditions.⁶⁸ However, the maximum complexity of the pool capable of being screened *in vivo*, i.e. the available sequence space, is limited by transformation efficiency. Only 10^8 - 10^9 sequences can be screened by *in vivo* selection methods. To overcome this limitation, *in vitro* selections are applied for new aptamers and synthetic riboregulators.

In vitro selections, also known by the acronym SELEX (Systematic Evolution of Ligands by Exponential Enrichment) allows to assess in theory 4^N variants. N is the number of randomized nucleotides in the initial starting pool. However, due to technical limitations of the procedure, the sequence space is reduced to 10^{14} - 10^{15} . In an *in vitro* selection procedure, the starting DNA library can be commercially synthesized or generated by error prone PCR. A randomized region is flanked by two constant parts that allow transcription and amplification at the end of the cycle. The length of the randomized region ranges from 30 to 70 nucleotides.⁴³ The isolation of functional sequences is generally done by affinity chromatography. The ligand is bound to a support phase and the RNA pool is initially passed through the column. After several washings, only molecules with binding properties are retained on the resin and finally eluted. Reverse transcription PCR (RT-PCR) is then performed to obtain the new DNA pool. This procedure is for aptamer selections, where

RNAs are isolated for binding to a specific ligand. However, the general scheme described can be adapted to better suit the sought after RNA activity. Multiple cycles, between 10-20, are repeated in order to achieve enrichment.

Common parameters that influence the selection, for both *in vivo* and *in vitro* methods, are stringency and counter-selection. Increasing the stringency of a procedure allows for the isolation of molecules with high affinity for the recognized ligand. For example, decreasing the available concentration of the ligand during the selection favors those sequences that bind with higher affinity. Counter-selection permits to increase the selectivity of the process. Counter-selection is performed by incubating the library with an analog of the ligand used for the selection. Molecules that possess activity are discarded, in order to enhance substrate specificity.

3.2 An *in vitro* riboswitch selection

In the previous chapter the *in vitro* characterization of a riboswitch and how the riboregulator is exploited for the construction of cellular mimics able to communicate through sensing and responding with the environment are presented. The theophylline responsive riboswitch is assessed. The cellular mimic relies on the specific recognition of theophylline for starting the communication process. The dependence towards this molecule severely limits the possible applications the system presents. It would be advisable to engineer cellular mimics able to respond to different molecules than theophylline. Advantages for example would be 1. to gain the ability to seek and destroy a soil contaminant or 2. to target specific cells, e.g. tumor cells. The cellular mimic can be engineered to sense different molecules through the identification of new riboswitches. Until now, the potential of riboswitches as sensor tools has been little explored. It is possible to expand the capabilities of riboswitches as tools in synthetic biology by selecting new riboswitches that sense new target molecules in a custom-made, specific fashion.

As previously described, few attempts were made in converting existing natural riboswitches by *in vivo* selection, e.g. the TPP riboswitch,⁶⁷ or in designing new aptazymes based sensors.⁴⁷ However, to date, a full *in vitro* selection for new riboswitches has not been achieved. The development of an *in vitro* selection technique are presented in this chapter. Again, the RNA regulators taken into account are those switches able to control protein expression at the translational level. Thus, the relevant element to consider during the selection design is the modulation of the RBS upon ligand binding. The RNA is in an “OFF” conformation when the ligand is not bound and the RBS is sequestered. Vice versa, the “ON” conformation is structured with a freely accessible RBS.

Since a translational control riboswitch is sought, a protein-based selection strategy is initially considered. Many *in vitro* techniques for the evolution of new enzymes have been set in the previous years, e.g. *in vitro* compartmentalization (IVC)^{16,18} and mRNA display.^{69,70} These selections share the same aim, i.e. to evolve proteins with new functionality from a random pool of sequences. Similarly to functional nucleic acid selections, a linkage between the phenotype (protein in this case) and RNA (or DNA) is necessary. However, the methods differ by how the linkage is achieved and where the selection is carried out. For example, the selection step takes place inside designed compartments for IVC or in solution for mRNA display. IVC exploits the compartment encapsulation of components not only for isolation from the external environment, but also for direct coupling of the nucleic acid with the encoded protein variant. Thus the linkage is obtained inside an emulsion droplet. In principle, only one genotype information molecule is encapsulated inside the compartment and protein production is performed via cell-free transcription and translation. For this reason, the number of possible variants is reduced to the available emulsion droplets. 1 mL of emulsion solution can generally provide 10^9 compartments, thus libraries with higher complexity cannot be screened. The binding between the nucleic acid and the encoded protein is accomplished through coupling with tags, e.g. a streptavidin tag fused to the protein binds the biotinylated corresponding DNA

molecule.¹⁶ Upon breaking the emulsion after each selection round, functional proteins are isolated and the linked genetic information is processed through RT-PCR for amplification before entering the following cycle.

Conversely, mRNA display does not exploit compartmentalization and tags for coupling genotype and phenotype. This selection methodology is designed starting from ribosome display,⁷¹ where a nascent peptide is displayed on the ribosome at the end of its translation. In this case, the interaction between the mRNA, the encoded protein and the ribosome is stabilized by non-covalent bonds, which should be preserved during the selection process.⁷¹ Instead, mRNA display is based on a direct covalent link between the mRNA and the protein that each RNA encodes. The formation of a stable bond is due to the coupling of a linker containing a puromycin molecule at the 3'-end of the RNA. Puromycin is an antibiotic able to enter the A site of the ribosome while translation is occurring, mimicking a tyrosyl-tRNA. Once in the ribosome site, the puromycin promotes the formation of an amide bond with the C-terminal of the nascent protein. The fusion molecule that is produced is then isolated via a separation method, e.g. a binding column, and finally the genetic information is reverse transcribed and amplified.

mRNA display is a protein selection technique. In order to select for a riboswitch, the mRNA display method is adapted. In fact, if RNA sequences are in an ON conformation, they would expose the RBS, thus allowing for protein production, and the RNA-protein fusions form. Conversely, RNA sequences in an OFF conformation do not allow for protein production and thus no RNA-protein fusions are formed. Via protein-affinity chromatography, only those RNA molecules showing switching activity in the presence of the ligand are isolated (Fig. 3.1).

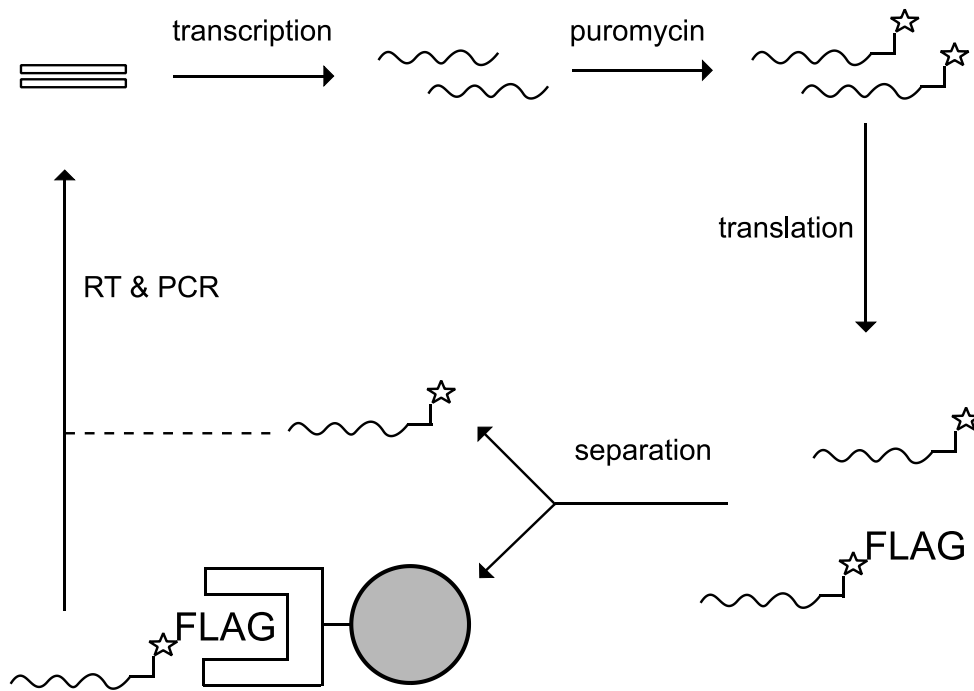


Fig 3.1 mRNA display cartoon showing a selection round. A double strand DNA library (top left) is initially transcribed. The RNA is coupled with a linker containing a molecule of puromycin (star). Upon translation, functional RNAs are fused with the peptide encoded via a covalent bond mediated by puromycin. Next the fusion complexes are isolated by FLAG affinity resin and sequences are reverse transcribed (RT) and amplified by PCR.

In the general selection scheme for mRNA display, the DNA library is initially transcribed into RNA via a T7 RNA promoter presented in the 5'-end constant region. The pool is then coupled to a DNA linker containing a deoxy-adenosine (dA) stretch and a puromycin molecule. The coupling is achieved through two different methods: via splint mediated ligation⁷² or a photo-crosslinking reaction.⁷³ The former exploits a DNA splint to anchor by base pair complementarity the 3'-end of the mRNA with the 5'-end of the linker and to allow enzymatic ligation. The latter involves functionalization of the 5'-end of the linker with a psoralen molecule, which by UV irradiation forms a stable covalent bond between the RNA and the puromycin-containing linker. After the crosslinking, the RNA pool is translated with cell-free transcription and translation machinery. Only functional molecules form the RNA-protein fusions. Isolation via a protein moiety, e.g. a FLAG peptide, allows recovery of functional sequences. Through RT-PCR the resulting

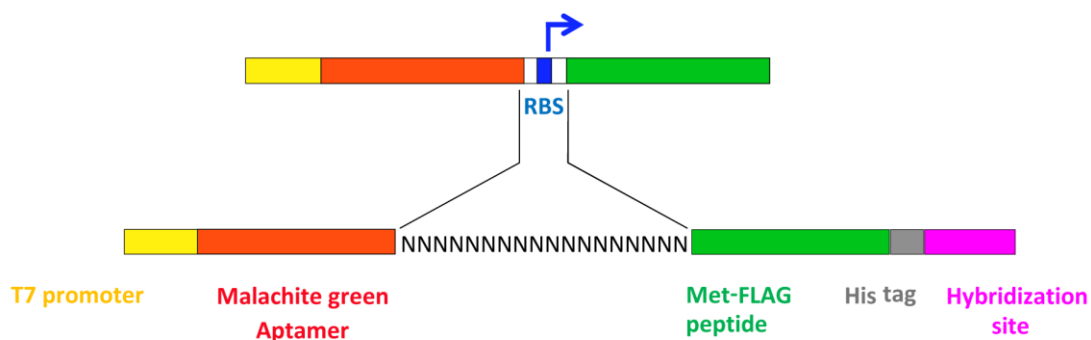
sequences are amplified. Together these steps represent one round of selection. Generally, between 8 and 15 cycles are carried out. Moreover, additional purification steps are introduced. For example, the (dA)₁₅ stretch in the linker allows for the purification of fusions from the translation reaction components. The order of the selection steps can also be modified according to the selection strategy, e.g. reverse transcription is performed before the affinity-column isolation of functional molecules in order to limit interactions between RNA and the resin. For riboswitch selections, the ligand is added to the translation reaction, inducing a switch in the RNA conformation that allows for translation. The library design and the selection steps performed are described and discussed in the following section.

3.3. Material and method

3.3.1 Library design

The library was rationally designed by inserting a randomized nucleotide stretch between an already characterized aptamer and a coding sequence (Fig. 3.2). This choice was based on previous aptazymes selections and on *in vivo* riboswitch examples.^{58,66} In riboswitches selections, the randomization did not involve the full sequence of the riboregulator. Only a specific linker or communication module was selected. Since the key element for the translational control riboswitch selection was the RBS, selection via mRNA display for an RBS modulated by ligand binding was initially planned. The library was built introducing a known aptamer region followed by eighteen random nucleotides. The RNA aptamer changed conformation upon ligand binding. The aim of the selection was to translate the conformational shift in the RNA structure induced by the aptamer-ligand binding to the following random stretch, specifically to a RBS sequence that would possibly arise during the selection. RBS accessibility would then be regulated by the ligand binding in the aptamer region. The number of eighteen random residues inserted before the start codon of a reporter gene was determined mainly for two reasons. First, an optimal RBS

sequence, e.g. AGGAG, was generally composed of five base pairs.⁷⁴ Optimal spacing between the ATG of the protein and the +1 position after the RBS could affect translation, and 6 base pairs were found to be optimal.⁷⁵ Moreover, starting from the example of the theophylline riboswitch,⁶⁶ a region of eighteen nucleotides between the aptamer and the start codon, including the RBS, was suggested for synthetic riboswitch screening and discovery.⁷⁶ These considerations did not take into account the position where a putative RBS could develop in the randomized part. Nevertheless, having eighteen random base pairs the library complexity was 10^{10} (4^{18} sequences). This value was higher than the number of possible sequences screened by *in vivo* selections and lower than the technical limit of mRNA display selections (10^{12}).



5' – GC**TAATACGACTCACTATAGGATCGCGACTGGCGAGAGCCAGGTAACGAATCGATCC** – 18 N –
 – **ATGGACTACAAGGACGACGACACAAG**GGCAGCCATCATCACCATCACCATATGGGCC**ACCCGGCTATTAA** – 3'

Fig. 3.2 Overview of the library exploited for *in vitro* selection of a malachite green responsive riboswitch. The sequence included a T7 promoter (yellow) and the malachite green aptamer (red) at the 5'-end. Eighteen randomized nucleotides were inserted downstream of the aptamer in order to select for a RBS element modulated by the ligand-induced conformational change. The coding sequence contained a Methionine-FLAG peptide (green), a His- tag (grey) and an hybridization site for crosslinking (purple). The DNA sequence is reported (148 bp). In the sequence, the start codon (ATG) is highlighted in bold. The nucleotides color follows the figure code.

The malachite green (MG) aptamer was used as the sensor for the riboswitch selection.⁷⁷ Malachite green binds with high affinity to the RNA aptamer (the dissociation

constant is 117 nM) and upon binding the complex becomes fluorescent.⁷⁸ The transfer of this property to a riboswitch represented an additional value to the selection, opening possible applications and characterization methods of the resulting sequences.

The library presented at the 5'-end a T7 RNA promoter for efficient transcription. Downstream of the randomized nucleotides, a start codon was inserted. The RNA sequence coded for the FLAG peptide. This small peptide was exploited in order to avoid possible truncation of a longer protein and to maximize the fusion production. Moreover, a FLAG tag also allowed easy separation of functional sequences via column purification on anti-FLAG resin beads. Nevertheless another tag was also added, i.e. six Histidine tag (His-tag), to increase the length of the coded region and to avoid interference of the fusion complex in bead-FLAG peptide binding. Later it was discovered that the His-tag caused unwanted secondary structure formation in the mRNA, according to thermodynamic prediction.

An additional feature was necessary in the library design for performing an mRNA display based *in vitro* selection. In order to crosslink the puromycin-containing linker with the mRNAs, an hybridization site should be present at the 3'-end of the sequence. This site was designed complementary to the linker strand. The overall design followed the assembly presented by Lohse and coworkers.⁷³

3.3.2 Reagents

All the reagents were purchased from Sigma Aldrich unless specifically indicated. Oligonucleotides were bought from Eurofins Genomics (GmbH), with some exceptions. The puromycin-containing linker was ordered from the W.M. Keck facility at Yale University. The sequence specifications have already been reported.⁷⁹ The oligonucleotides for the library assembly were from Tema Ricerca (Integrated DNA Technologies, IDT). DreamTaq DNA

polymerase, RNase A, Proteinase K, RiboRuler Low Range RNA Ladder were bought from Carlo Erba Reagents (Fermentas). Reverse transcriptase Superscript II was purchased from Life Technologies. Oligo (dT) Cellulose Type 7 and Ultrafree-MC centrifugal filter devices were purchased from Millipore Italia (GE-Healthcare and Millipore respectively). Oligo (dT) magnetic beads, T7 RNA polymerase, RNase inhibitor from human placenta, Yeast Inorganic Pyrophosphatase, the PURE system (PUREXpress), DeepVent-R polymerase, Theromopol buffer, Phusion polymerase, DNase I, dNTPs and NTPs were purchased from New England Biolabs (NEB, Gmbh). Low melting Nusieve GTG agarose was purchased from Lonza Italia. L-³⁵S Methionine (³⁵S-Met), ³²P- alpha adenosine triphosphate and scintillation liquid were obtained from Perkin Elmer. RTS 100 E.coli HY Kit for cell-free translation was from Roche Italia (5 Prime). DNA Gel extraction and PCR purification kit was purchased from Promega Italia. qPCR SYBR[®] Green Supermix was from Bio-Rad. All the solutions were prepared with DEPC-treated water and filtered through 0.22 µm filters in order to be RNase free.

3.3.3 mRNA display for riboswitch selection

The mRNA display protocol was extensively described by B. Seelig.⁷⁹ However, few changes were made to adapt the procedure to the selection of RNA rather than protein. An overview of the developed selection protocol is highlighted in Fig.3.3.

Briefly, the MG library DNA pool was assembled and amplified. Next, transcription and RNA purification via PAGE were performed. Recovered RNA was photo-crosslinked to a puromycin-containing linker. After ethanol precipitation, the RNA was added to an *in vitro* transcription and translation system for translation reaction. Translation was performed in the presence or absence of MG depending on the cycle. The sample was then purified by oligo (dT) affinity chromatography. The resin efficiently bound the crosslinked RNA via recognition of a (dA)₁₅ stretch of the linker. mRNA-FLAG peptide fusions and unreacted

crosslinked RNAs were isolated. Next, reverse transcription was carried out. Ethanol precipitated samples were then subjected to the FLAG-binding selection column. This selection allowed for the retrieving of only the fusion complexes, i.e. the functional riboswitch sequences. Finally, the selected sequences were amplified and another round of selection could start.

The selection steps were specifically designed and tested with different approaches. The setting up of the selection is described and discussed in the next section.

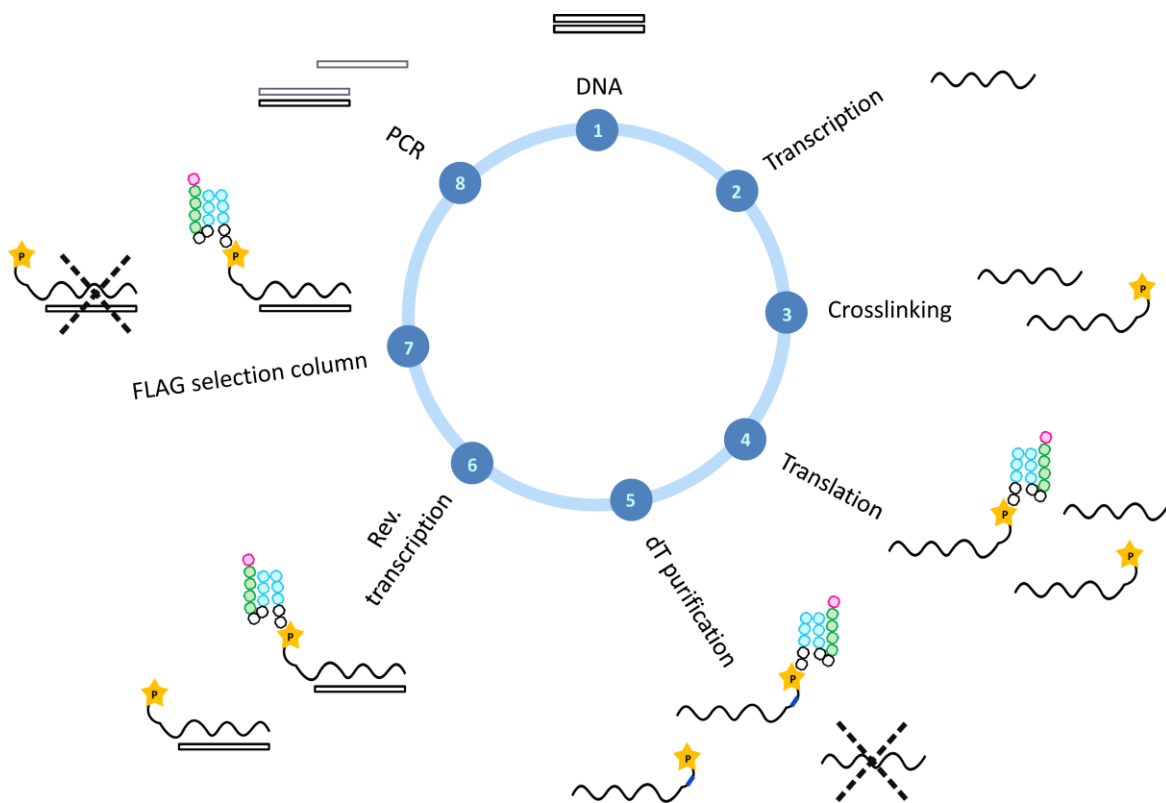


Fig. 3.3 mRNA display scheme of the steps necessary for one round of selection. Eight principal steps were performed during each cycle of the selection. Purification steps are not included in the cartoon. RNA is represented as a wavy line, DNA strands are linear. In the crosslinking step, not all the RNA was crosslinked to the puromycin-containing linker. In the translation reaction, only those RNAs that were functional produced the fusions. Step 5 allowed for the removal of uncrosslinked RNAs. Reverse transcription was performed on the crosslinked RNAs, obtaining a RNS-DNA duplex. The FLAG column permitted to separate the fusion complexes from non-functional RNAs. The cDNAs of the eluted sequences were finally amplified.

3.4. Preliminary results for method development

The optimal conditions for the selection were tested with control constructs, which were designed according to the riboswitch library. The constructs were named “Fake” and “Fake-minus-His” (Table 3.1). The name was related to the construct function of mimicking the library behaviour. In fact, the constructs would be used as test sequences for assessing the cycle steps. Thus the sequence composition was similar to the library sequence, excluding the randomized part. The Fake-minus-His construct differs from the Fake only by the removal of the His- tag in the coding region. The developing of the Fake-minus-His construct was due to a thermodynamic analysis via the *mfold* web prediction tool. In fact, the Fake sequence was initially designed to have an open RBS in order to act as a positive control. However, analysis of the RNA secondary structure demonstrated how the RBS was blocked by base pair annealing at 37 °C. Thus, the Fake-minus-His sequence was designed. The removal of the His- tag before the hybridization site allowed to free the RBS according to structure prediction. For this reason both constructs were exploited in the setting up of the technique.

Table 3.1 DNA sequences of construct used for mRNA display optimization.

Constructs	Lenght (bp)	
Library (18N)	148	5'- GCTAATACGACTCACTATAGGATCGCGACTGGCGAGAGCCAGGTAAC GAATCGATCCNNNNNNNNNNNNNNNNNNNN NATGG ACTACAAGGACG ACGACGACAAGGGCAGCCATCATCACCATCACCATATGGGCCACCGG CTATTAA-3'
Fake	120	5'- GCGCGCGGATTTAATACGACTCACTATAGGTTAAG AAGGAG ATATA CAT ATGG ACTACAAGGACGACGACGACAAGGGCAGC CATCATCACCA TCACCAT ATGGGCCACCGGCTATTAA-3'
Fake-minus-His	104	5'- GCGCGCGGATTTAATACGACTCACTATAGGTTAAG AAGGAG ATATA CAT ATGG ACTACAAGGACGACGACGACAAGGGCAGCATGGGCCACC GGCTATTAA-3'

The Fake and Fake-minus-His constructs were used for mRNA display technique setting up. Both the constructs presented a RBS (cyan). The start codon is highlighted in bold. Fake and Fake-minus His differed for a His- tag in the coding sequence (orange).

3.4.1 Library assembly optimization

The library assembly was performed by overlapping extension of oligonucleotides. The forward oligonucleotide contained the randomized part followed by 20 bp, which constituted the overlapping region. The reaction (50 µL) was carried out with different Taq polymerases. DreamTaq, Trustart Taq, One Taq Hot start, Taq DNA pol (standard) and Homemade Taq were compared (Fig. 3.4). 2.5 pmol of each oligonucleotide were used. The buffers used for DreamTaq, Trustart, Taq standard were specifically supplied by the manufacturer, the One Taq Hot start and Homemade Taq polymerases exploited the Thermopol buffer. The protocol was constituted by just one cycle of 10 min at 94 °C, 15 min at 55 °C, 45 min at 72 °C. Addition of the polymerase followed the initial denaturation step.

Clearly the Homemade Taq polymerase did not work (sample 4, Fig. 3.4). Conversely, a band was observable for the DreamTaq product (samples 1, 2). Increase in the overall volume of the reaction, and consequent dilution of the oligonucleotides concentration did not improve the result (sample 3). DreamTaq polymerase result was reproducible and final product evaluation on agarose gel (2%) showed that DreamTaq processivity was higher than all the other polymerase tested.

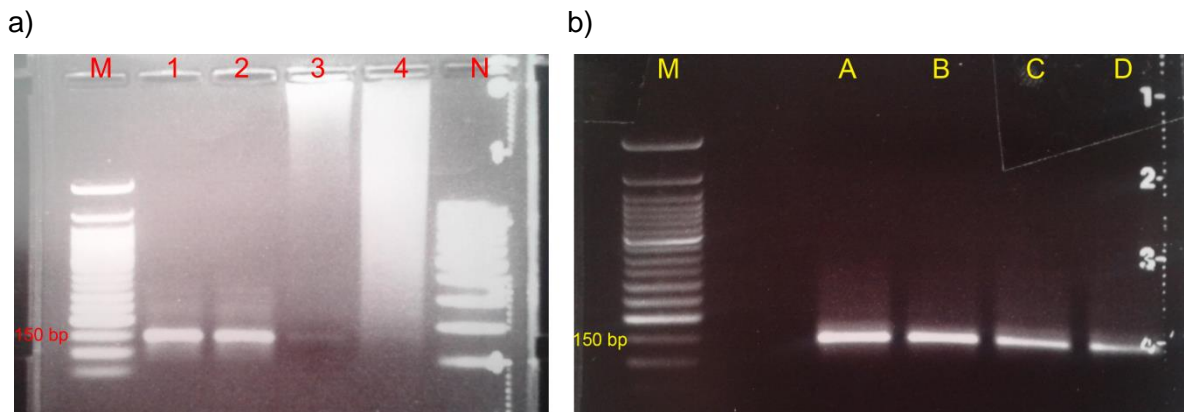


Fig 3.4. Assembly of the double strand DNA library by overlapping oligonucleotide extension. The products were analyzed by 2% agarose gels. Panel a) DreamTaq and Homemade Taq polymerases were exploited in a 50 μ L reaction. Homemade Taq failed to show activity (sample 4). Conversely, DreamTaq result was reproducible (samples 1 and 2). Increasing the reaction volume to 200 μ L did not improve the result (sample 3). Panel b) DreamTaq, TruStart, One Taq Hot start and Taq standard were assessed (samples A, B, C, D respectively) in a 50 μ L assembly reaction. Marker (M) is a 50 bp ladder (NEB), Marker N is a 100 bp ladder (Fermentas). Note that the agarose gel in panel a) was exposed to the UV light with the gel support.

For the production of the Fake and Fake-minus-His construct, the agarose gel extracted double strand product (10 ng) was amplified by PCR with Homemade Taq polymerase. Next, PCR purification was performed with the column-based clean up commercial kit before transcription, since the Fake construct agarose gel run showed the absence of additional unspecific bands in the PCR product. Library and test constructs were stored at -20°C.

3.4.2 Transcription and RNA purification

Transcription reactions were performed as previously described⁷⁹ in a 50 μ L reaction. Conditions were screened by radioactively labeling with 32 P-alpha-ATP the RNA of the Fake construct. The radioactive nucleotide was added in the transcription mixture (20 μ Ci). Incubation time at 37 °C and starting DNA template amount were assessed (Fig. 3.5). 10 μ L aliquots were collected at specific time points of 2 h, 4 h, and 6 h. The evaluation was done by 8 M Urea- 4% PAGE. Samples were denaturated by addition of 2X loading dye (8 M urea, 20% wt/vol sucrose, 0.1% SDS, 0.05% xylene cyanol, 0.09 M Tris, 0.09 M borate, 10 mM EDTA) and incubated at 94 °C for 5 min. The gel was run in TBE buffer for 5 h at 140 V. Phosphorimaging was performed with the Typhoon Trio Variable Mode Imager (Perkin Elmer) after exposure of the radioactive gel in a screen cassette (GE Healthcare) for 40 min.

The best condition seemed to incubate the reaction for 4 h at 37 °C. However, longer incubation did not improve the overall result. As template, 1 pmol and 10 pmol of double strand linear DNA were tested. Resulting RNA was correlated with the template amount.

Time and DNA template setting were not the only tests performed. Other small improvements in RNA purification and isolation were achieved. In the selection protocol, after the transcription reaction, the RNA was processed with the addition of 37 mM EDTA, 0.5 mM CaCl_2 and DNase I and incubated at 37 °C for 1 h. Next, ethanol precipitation was performed prior to loading on a 8M Urea- 4% PAGE denaturing gel. While setting up the technique, before loading on the gel, the RNA pellet, even if transcribed by 1 pmol of DNA as starting material, was difficult to resuspend in 20 μ L of DEPC-water. Elimination of the salt addition in the ethanol precipitation protocol helped solve the problem. 8 M Urea- 4% PAGE was used in the selection process for RNA purification. Recovery was done by UV

shadowing after the gel run. An important step in the optimization of the UV shadowing was removal of bromophenol blue from the sample loading dye.

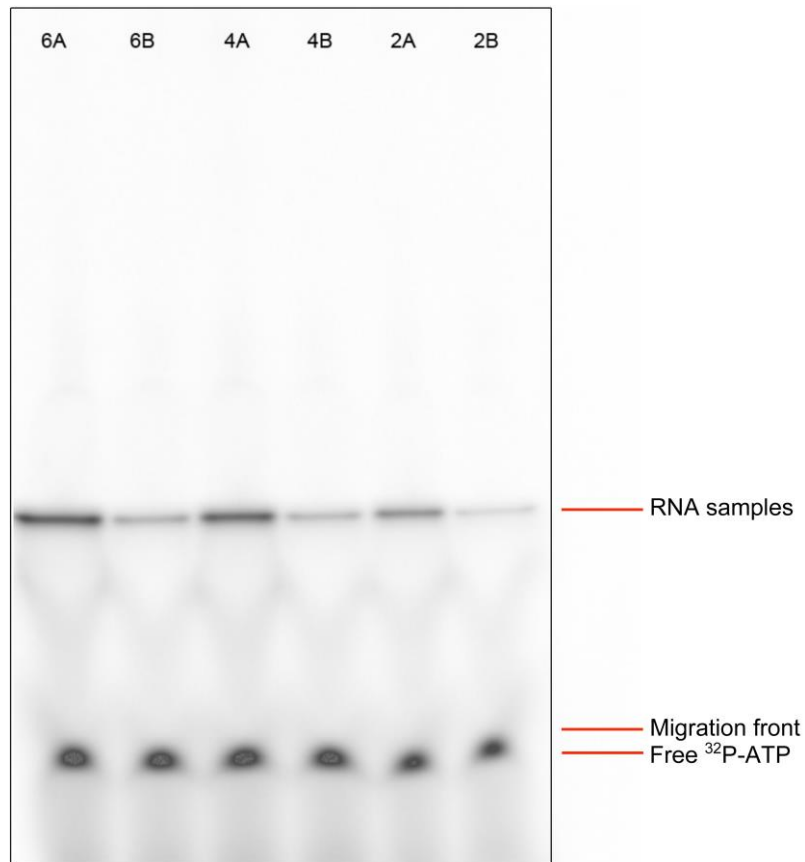


Fig. 3.5 ^{32}P radioactive RNA production evaluation on a 8M Urea- 4% PAGE. Transcription was performed from 10 pmol and 1 pmol of starting double strand DNA template (samples A and B, respectively). The reactions were incubated at 37 °C, aliquots were taken after 2 h, 4 h, 6 h (samples 2, 4, 6) and run on the gel. The image was generated by 40 min exposure on a phosphoimaging screen and analyzed by phosphoimager.

3.4.3 Crosslinking reaction monitoring

Crosslinking reaction was performed according to the Seelig B. protocol.⁷⁹ The volume of the reaction was scaled down to 50 μL . The concentration of puromycin-containing linker was 7.5 μM . Initially the reaction was tested using radioactively labeled RNA. 1 and 10 pmol of Fake construct PCR product were transcribed in the presence of 20

$\mu\text{Ci } ^{32}\text{P}$ -alpha-ATP. Then, the obtained radioactive RNAs were Urea-PAGE purified as previously described (see page 87). Recovery from the gel was performed by UV shadowing and crush and soak method.⁷⁹ The resulting samples were dissolved in 20 μL DEPC-water. Crosslinking reactions were set up using the whole amount of obtained RNA in XL buffer (100 mM KCl, 1 mM spermidine, 1 mM EDTA, 20 mM HEPES-KOH, pH 7.5) in the presence of the puromycin containing-linker. Reactions were initially incubated at 70 °C for 10 min, cooled to 25 °C in 5 min and finally transferred to a transparent 96-well plate. Next, samples were irradiated on ice for 20 min at 365 nm with a handheld UV lamp (UVP). Resulting RNAs were ethanol precipitated prior to being load on a 8 M Urea- 4% PAGE gel. Control samples, i.e. uncrosslinked RNAs produced respectively from 1 pmol or 10 pmol DNA, were included in the analysis. Finally, exposure of the acrylamide gel was carried out for 45 min. Results showed how the crosslinking process worked, even if the gel was exposed too much (Fig. 3.6). However, it was possible to see a higher band in all the crosslinked samples.

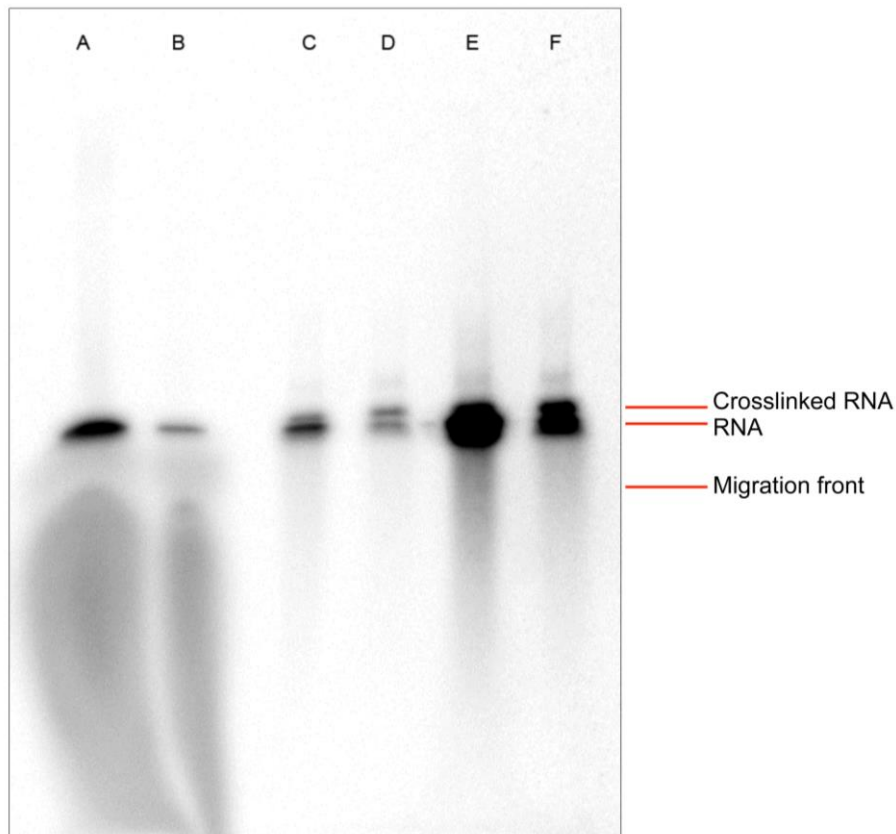


Fig. 3.6 Crosslinking on ^{32}P radioactive RNA. Labeled RNAs were photo-crosslinked with the puromycin-containing linker. Results were evaluated with 8 M Urea-4% PAGE. Samples A and B were uncrosslinked control RNAs, respectively transcribed from 10 pmol and 1 pmol DNA. Samples C and E were obtained from 10 pmol DNA as transcription template, samples D and F from 1 pmol DNA. All the RNA was crosslinked in samples E and F, while half of the obtained RNA was exploited for crosslinking reaction of samples C and D. Crosslinked samples presented an higher band compared to the single control RNAs.

This experiment confirmed that the crosslinking reaction was working. During the selection procedure, crosslinking would be a good reference point for monitoring the RNA status in each cycle. However, performing the control PAGE gel after each round would have been difficult and time consuming. For this reason, an agarose gel was developed as checkpoint for the selection round before entering the translation step. This gel allowed for the verification of RNA integrity and crosslinking efficiency. An aliquot of the RNA (1 μL of 10 μL resulting after the ethanol precipitation) was checked with a native 2% agarose gel in TAE buffer. The RNA was denatured by the formaldehyde-containing loading dye and an

incubation step at 70 °C for 10 min. The RNA ladder (commercial) was prepared accordingly. Clearly, the crosslinked band was visible (Fig. 3.7). This experiment was repeated with the Fake-minus-His construct. The result was the same. Comparison between crosslinked and uncrosslinked bands of the same RNA confirmed the expected efficiency of the reaction, i.e. almost 50%.⁷⁹

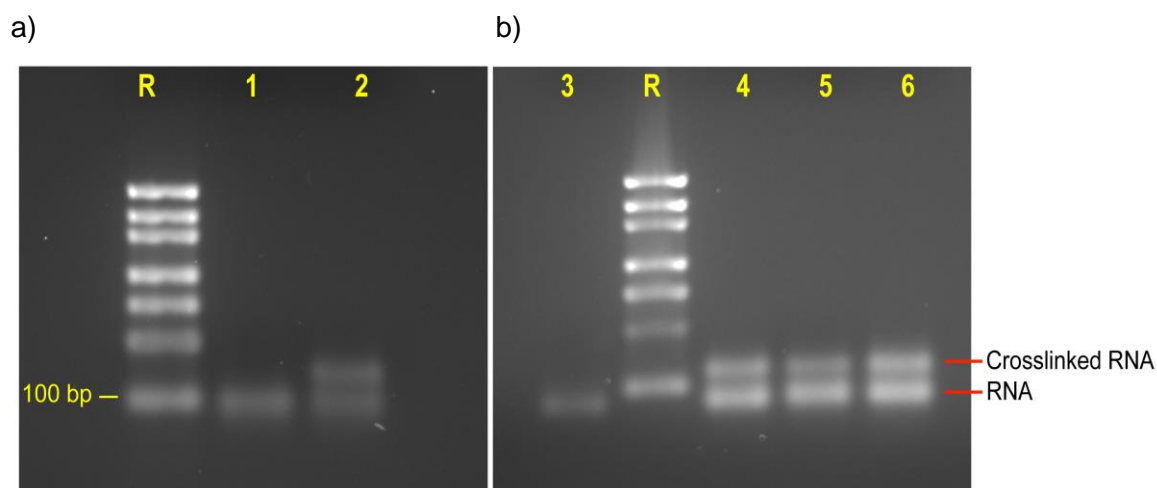


Fig. 3.7 Agarose gel evaluation of crosslinking reaction. An aliquot of crosslinked RNA was run after the reaction on a 2% agarose gel to verify the coupling with the puromycin-containing linker. Samples were denaturated prior to loading. Samples 1 and 3 are uncrosslinked and included as reference control. The efficiency of the process is ~50% as demonstrated by the two bands presence in crosslinked samples. R is the 100 bp RNA ladder (Fermentas). Panel a) Fake construct samples; panel b) Fake-minus-His samples.

3.4.4 Translation, fusion formation and isolation

In this section the development of the main part of the selection technique is presented. Many experiments were carried out with different approaches. Initially mRNA-peptide fusion formation was monitored by PAGE. Next, Real time PCR (qPCR) was exploited, and finally ³⁵S-Methionine scintillation counting was performed. Exploiting different techniques was possible since the fusion complex consisted of both an RNA and a peptide species.

Initially, the translation reaction was performed with an *E. coli* cell lysate (RTS 100 *E. coli* HY Kit). Many of the mRNA display protocols were based on rabbit reticulocyte lysate as *in vitro* transcription and translation system.⁷⁹ However, the aim of the project was to select for a riboswitch that is functional in prokaryotic cells. Thus, the bacterial extract seemed like the optimal system. The reaction (50 μ L) included the lysate, a reaction mix solution, amino acids, reconstitution buffer, 10 U RNase inhibitor and the crosslinked RNA. Methionine was provided separately from the cell lysate components, allowing for the possible addition of the radioactive labelled amino acid. Translation was generally performed at 30 °C, as suggested by the kit manufacturer, for 4 h. Moreover, the lysate solution was supplemented with 100 mM KCl and 0.9 mM Magnesium acetate.⁷⁹ Immediately after translation, for optimal fusion production, additional KCl and MgCl₂ were added to the reaction to achieve a final concentration of 531 mM and 50 mM, respectively. The solution was then incubated at room temperature for 5 min.

1. PAGE gel analysis of translation products

The preliminary results of transcription and crosslinking reactions were obtained by PAGE gel visualization as described above. Thus, analysis of the migration on a polyacrylamide gel was exploited to check also translation and fusion formation for Fake construct samples. Initially, a 8 M Urea-4% PAGE gel was run on the translation product from radioactive labelled RNA. In this experiment, the Fake construct crosslinked RNA was subjected to translation (10 μ L). The reaction was performed as described above and the products were ethanol precipitated. However, the gel failed to show bands.

A different approach was to radioactive label the peptide of the fusion via ³⁵S-Methionine inclusion in the translation reaction. In this experiment, the RNA was not labelled with ³²P. In the translation reactions the amount of un-labelled Methionine was reduced according to the manufacturer protocol of the cell lysate kit and 20.4 μ Ci of L-³⁵S

Methionine was supplemented. The amount of the initial RNA was not measured, but the same aliquot of the crosslinked Fake RNA was added for each of the samples. Translation and salt addition was performed as stated above, even if the reaction volume was scaled down to 26.3 μ L. Two of the samples were treated with RNase A and Proteinase K, respectively, for 30 min at 37 °C. These samples were controls for comparison with the RNA-peptide fusions. Again, the translation products were run on a 8 M Urea-4% PAGE gel. Purification before loading on the gel was done by ethanol precipitation (without salt addition). Visualization of the gel was performed by overnight gel exposure to the phosphorimager screen, since the ^{35}S signal is less intense than ^{32}P signal and requires more time for visualization. Unfortunately, changing the labelling method gave negative results (Fig. 3.8). Spots were visible for all the samples, including the reactions that were degraded by the protease and RNase. The gel was exposed too much and did not provide useful information. However, it seemed clear that RNA denaturing gels were not the optimal method to assess fusion formation.

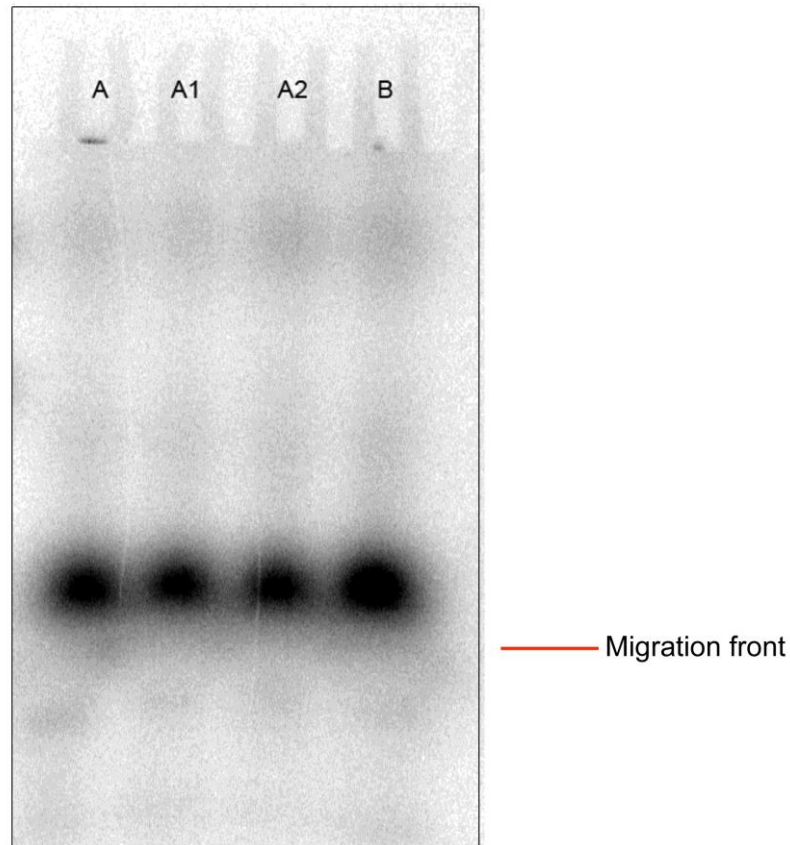


Fig. 3.8 ^{35}S radioactive samples run after translation on a 8 M Urea- 4% PAGE. 10 μL of sample A were loaded on the gel, while A1 and A2 aliquots were treated with RNase A and Proteinase K, respectively, for 30 min at 37 $^{\circ}\text{C}$ prior to loading. Conversely, the whole translation product was loaded for sample B. Signals were visible for all the samples; however, no useful information about fusion formation was obtained.

From the gel result it was not possible to determine if the translation reaction and fusion formation were accomplished. Generally, cell-free transcription and translation systems have been optimized for protein production and the buffer composition heavily influences the performance. For example, the PURE system cannot withstand high concentrations of NaCl without affecting protein yields. For this reason, a control experiment was performed in order to assess if the addition of salt to the translation reaction could possibly cause inhibition. The RTS 100 HY *E. coli* kit includes also a GFP encoding control plasmid for testing reaction functionality. Two separate reactions were monitored over time at a QuantaMaster 40 UV VIS spectrofluorometer (PTI) with excitation

and emission wavelengths of 395 and 504 nm, respectively. The first reaction was carried out exactly as specified in the manufacturer protocol. Briefly, 0.5 µg of control GFP vector were exploited in a 25 µl reaction and incubated for 6 h at 30 °C. To the second reaction 100 mM KCl and 0.9 mM Magnesium acetate were added, as previously done. The kinetics clearly demonstrated that the presence of additional salt could inhibit the translation reaction (Fig.3.9).

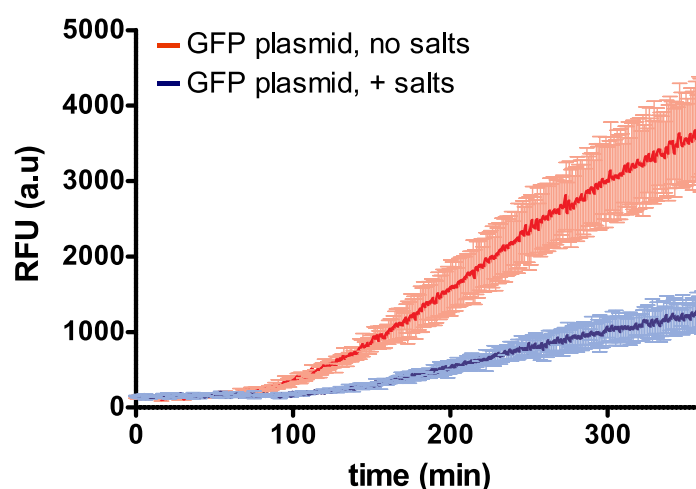


Fig. 3.9 Fluorescence kinetics for testing salt effects on an *in vitro* and transcription cell lysate. The production of GFP was assessed with the RTS 100 HY *E. coli* kit at 30 °C for 6 h. KCl and Magnesium acetate addition in the mixture affected the reaction causing efficiency decrease.

Considering salt inhibition in the translation reaction, a different method for incubation with salt after the translation was tested. In particular, an mRNA display selection that exploited a S30 *E. coli* cell lysate⁸⁰ was taken into account. Sisido and coworkers⁸⁰ at the end of translation added only KCl (500 mM) to the reaction, excluding MgCl₂, and incubated the solution for 1 h at 10 °C. Comparison between this procedure and the previous incubation method was carried out with ³²P labelled RNA. Again, radioactive crosslinked RNA was obtained. Conversely than before, translations reactions were not supplemented with salt in the solution. The crosslinked RNA was aliquoted in three samples for translation, thus each reaction contained the same amount of starting material (sample

A, B, C). Moreover, a control translation reaction with uncrosslinked RNA was prepared (sample D). After 4 h at 30 °C, sample B was incubated after addition of 531 mM KCl and 50 mM MgCl₂ for 5 min at 25 °C as before (thereafter called Seelig procedure). Sample C was processed following the Sisido procedure. On the other hand, any salt was supplemented to sample A, which was immediately stored at 4 °C. Results were evaluated with a 4% PAGE gel, containing not only urea for RNA denaturation, but also 0.1% of sodium dodecyl sulfate (SDS) for protein denaturation. Prior to loading, a 5 µL aliquot of the samples was acetone precipitated as suggested by the kit manufacturer. This purification step was performed adding initially 50 µL of -20 °C cold acetone. Samples were then incubated on ice for 5 min and centrifuged at 10,000 rpm for 5 min. The pellet was evaporated for 10 min at 70 °C after removal of the supernatant. Finally, samples were resuspended in the loading dye (10 µL of 8 M urea, 20% wt/vol sucrose, 0.1% SDS, 0.05% xylene cyanol, 0.09 M Tris, 0.09 M borate, 10 mM EDTA). Moreover, to evaluate the run, a prestained PAGE protein ladder was added in a well of the gel. Before exposure to the screen, some of the ladder bands were labelled by hand by painting a small amount of ³²P-ATP. The Urea-SDS PAGE result was again a failure (Fig. 3.10). In fact, none of the treated sample bands was visible on the gel. Nevertheless, P-32- labelled crosslinked and uncrosslinked Fake RNAs were clearly visible on the gels.

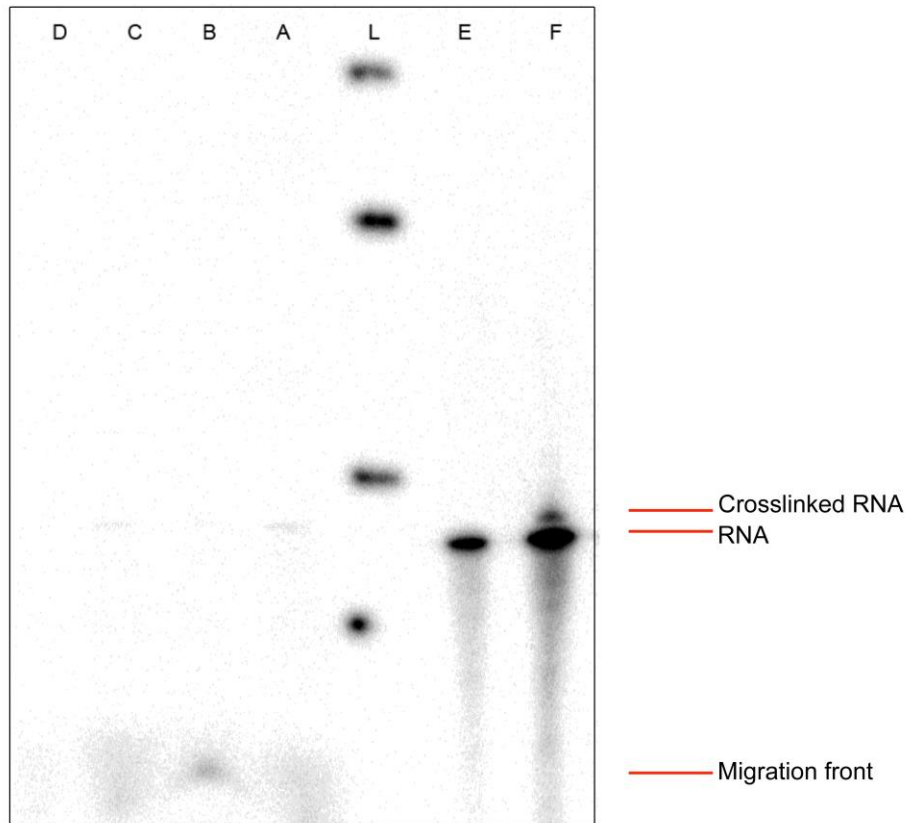


Fig. 3.10 8 M Urea-0.1% SDS- 4% PAGE of translation products. The RNA was ^{32}P labelled prior to crosslinking. After translation, sample A was not treated, while sample B was supplemented with 531 mM KCl and 50 mM MgCl_2 and incubated at 25 °C for 5 min. To sample C, 500 mM KCl was added after translation and the fusion formation was performed at 10 °C for 1 h. Uncrosslinked RNA was used as template for translation in sample D (control sample). A, B, C, D were acetone purified prior to gel loading. Sample E and F were controls and contained 1 μg uncrosslinked RNA and 1 μg crosslinked RNA. Acetone treated samples were not visible on the gel. L is a prestained protein ladder where specific bands were ^{32}P -ATP signed in order to serve as a reference in the final image. Bands highlighted were 10 kDa, 25 kDa, 70 kDa, 170 kDa (from bottom to top). However, since the gel composition was not specifically indicated for the ladder run, results were only qualitative considered.

From this result, repeating the experiments avoiding the acetone precipitation prior to loading were considered. Six reactions were planned, all derived from just a unique ^{32}P crosslinked RNA, preventing even small variations in the starting translation material to affect the reaction. Translations were performed in 50 μL of volume. Again, the different protocols at the end of the translation for fusion formation were assessed. Samples C and F were not treated after translation; samples B and E were supplemented with KCl and MgCl_2 at the end of the reaction and incubated 5 min at 25 °C, while samples D and A were

processed according to Sisido's procedure (Fig. 3.11). Of these samples, three were simply purified by ethanol precipitation and loaded on a 15% Urea- SDS PAGE (sample D, E, F). Conversely, the other samples were subjected to an anti-FLAG M2-agarose affinity column. This resin binds and allows the recovery of FLAG peptide-tagged species. 50 μ L of the resin were washed and resuspended.⁷⁹ 5 μ L of 10x FLAG-binding buffer (1.5 M KCl, 0.1% (wt/vol) Triton X-100, 50 mM β -mercaptoethanol, 500 mM HEPES-KOH, pH 7.4) were supplemented to each sample. These were then added to the resin and incubated for 4 h at 4 °C. A small empty column was used as chromatography support (Bio-Rad). Elution was performed twice with 100 μ L of 1x FLAG-binding buffer and two equivalents of FLAG peptide (the powder was solubilized in FLAG-binding buffer and stored at -20 °C). On the purified samples an ethanol precipitation was performed and finally products were loaded on the PAGE gel. A labelled crosslinked RNA was also included in a gel well as control (sample R). Since the concentration of acrylamide was higher than usual, the run was performed for 10 h at 140 V. Again, hand labelling with a small amount of ³²P was done on some bands of a prestained protein ladder (M). Exposure to the screen was carried out for 40 min. The resultant gel did not show any band for FLAG resin purified samples (Fig. 3.11). Conversely, un-purified translation reaction products were visible. Unfortunately, the bands seemed stuck in the wells. Comparison with the crosslinked RNA did not provide enough information to determine if the fusions were formed.

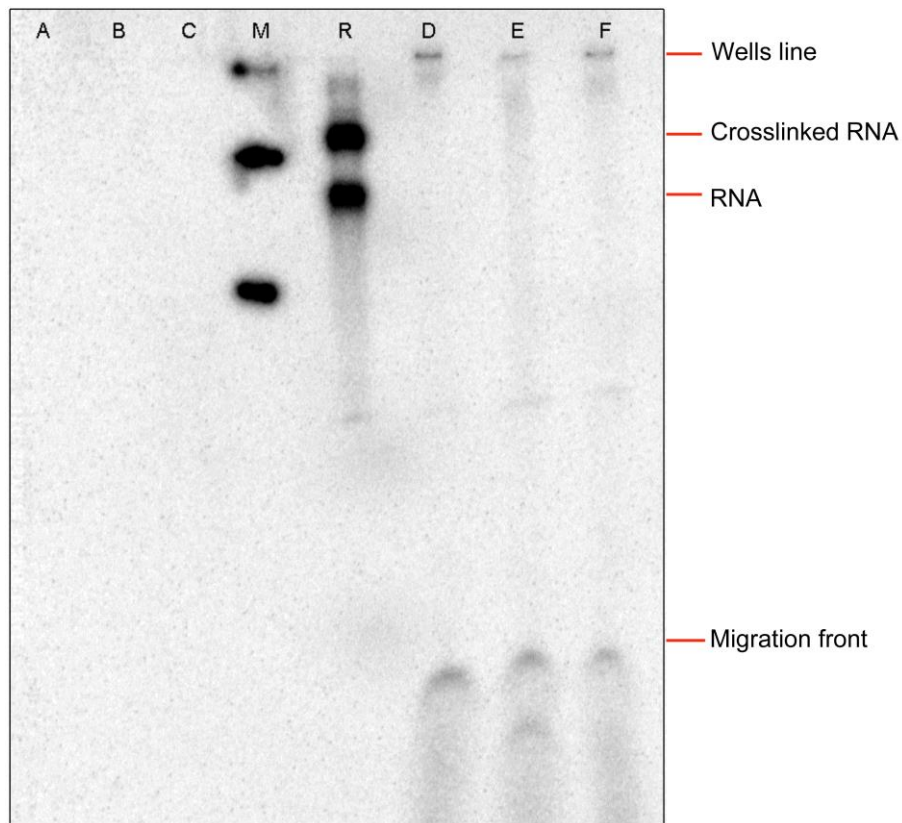


Fig. 3.11 8 M Urea- 0.1% SDS- 15% PAGE for fusion formation assessment. The RNA was radioactively transcribed and crosslinked. After translation, sample B and E were processed by addition of 531 mM KCl and 50 mM MgCl₂ and incubated 5 min at 25 °C. Sample A and D were supplemented with 500 mM KCl and incubated at 10 °C for 1 h. Sample C and F were not treated. Next, FLAG column purification was done on samples A, B, C. Results were compared with unpurified reactions (D, E, F respectively). Sample R was a crosslinked RNA control. M ladder was ³²P hand signed as qualitative reference. A weak signal, maybe corresponding to the fusions, is present for samples unpurified, even if stuck in the well of the gel.

2. qPCR analysis for fusion formation assessing

The second approach for evaluation of fusion production was based on real time PCR measurements. This method did not directly probe fusion formation after translation, like a PAGE gel, but considered the product after the FLAG selection column. In fact qPCR assessed the number of molecules present in the elution of the FLAG affinity column in order to know exactly how many sequences were retrieved at the end of the selection. Intermediate steps such as oligo (dT) column purification, reverse transcription, ethanol precipitations were not directly monitored. The fraction of retrieved molecules in the FLAG

column elution was compared to the number of molecules that entered into the column, i.e. the reverse transcription output. Or the ratio could also be evaluated towards the number of molecules in the first FLAG column wash, which contains unbound sequences. If a Fake construct was used, the expected result was to have a higher number of molecules in the elution than in the wash, since the RNA-FLAG fusion bound to the resin.

The number of molecules could be calculated by comparing the results to a standard curve of dilutions of Fake construct DNA in the qPCR reaction. Each standard curve was freshly prepared and run in the same experiment of the considered samples. Standard curve samples range covered from 10^9 to 10 molecules, with an order of magnitude as dilution scale (10^9 , 10^8 , 10^7 and so on). Each value was assessed in triplicates. Generally, qPCR reactions were performed in 10 μ L in 96-well back plates (Bio-Rad) and monitored with a CFX96 Real time PCR Detection System (Bio-Rad). Reaction components included a commercial master mix (Bio-Rad), containing dNTPs, optimized buffer with syber green, and DNA polymerase, the primer couple specific for the analyzed construct and an aliquot of the investigated sample. 10^4 and 10^2 molecule standard dilutions were also exploited for a gradient run in order to establish the optimal annealing temperature. For Fake and Fake-minus-His constructs specific primers were designed and the gradient temperatures were 62.5 °C and 52.6 °C, respectively.

Initially this technique was exploited for the evaluation of the Seelig and Sisido procedures after the translation reaction. Un-labelled samples were tested. Translation reactions were carried out with the RTS 100 *E. coli* HY kit (50 μ L volume). The RNA concentration was measured before crosslinking. Thus, 1.5 μ g of crosslinked RNA was estimated to be present in each sample. Next, salt incubation after translation followed the Sisido and Seelig procedures described above. Successive steps were firstly an oligo (dT) column purification, reverse transcription, and finally the FLAG column. The oligo (dT) purification allowed for the isolation of the crosslinked RNA from the translation reaction. In fact, a (dA) stretch was included in the puromycin-containing linker. 6 mg of resin was

exploited for each translated sample.⁷⁹ The cellulose was dissolved in 540 μL of oligo (dT) cellulose-binding buffer (10 mM EDTA, 1 M NaCl, 0.2% wt/vol Triton X-100, 10 mM β -mercaptoethanol, 20 mM Tris-HCl pH 8.0) and incubated with tumbling at 4 $^{\circ}\text{C}$ for 15 min. The resin was washed three times with 1 mL binding buffer and six times with 1 mL oligo (dT) cellulose wash buffer (300 mM KCl, 5 mM β -mercaptoethanol, 20 mM Tris-HCl pH 8.0). Elution was performed three times with 400 μL of oligo (dT) elution buffer (5 mM β -mercaptoethanol, 2 mM Tris-HCl pH 8.0). To reduce the volume, ethanol precipitation was performed and the pellet resuspended in 20 μL DEPC-water. Next, reverse transcription (40 μL) was set up in 50 mM Tris pH 8.3, 3 mM MgCl_2 , 0.5 mM dNTPs, 0.1 U/ μL RNase inhibitor, 10 mM β -mercaptoethanol and 4 U of Superscript II reverse transcriptase.⁷⁹ The reaction included a specific reverse primer and all the oligo (dT) purified RNA-protein fusions. Incubation was at 42 $^{\circ}\text{C}$ for 30 min. An aliquot of 6 μL of reactions was then exploited for qPCR determination of number of molecules that were entering in the FLAG affinity column. The remaining reverse transcribed product (34 μL) was processed by Anti-Flag M2-agarose chromatography as described above. The final elution volume was 200 μL , while the first column wash was 1 mL. Both were ethanol precipitated and resuspended in 20 μL DEPC-water. 1.25 μL of retrieved DNA was used for qPCR. Triplicates were run for reverse transcription samples, first wash and elution products, each for both Sisido and Seelig treated reactions. The results demonstrated that a very low amount of the Fake molecules were producing fusions. Both the Sisido and Seelig procedure gave as final output 10^3 molecules (Table 3.2). 10^4 molecules entering into the FLAG selection column were few if considered that the estimated amount of RNA added in translation reaction was about 10^{13} sequences. Moreover, all these sequences should produce the RNA-FLAG peptide fusion, since the Fake construct was designed to be a positive control.

Table 3.2 Number of molecules determined by qPCR

Fusion incubation procedure	n. molecules	
	Rev. Trascr.	Elution
Seelig	2.1×10^4	1.2×10^3
Sisido	2.7×10^4	2×10^3

The number of molecules reported is the average of triplicates. Experiments were performed on aliquots of reverse transcribed samples and on aliquots of elution sample. Two samples were analyzed: the first one was processed after translation according to the procedure indicated by Seelig,⁷⁹ i.e. addition of 531 mM KCl and of 50 mM MgCl₂ then incubation for 5 min at 25 °C. The second sample was supplemented with 500 mM KCl and incubated at 10 °C for 1 h, according to what exploited by Sisido and coworkers.⁸⁰ Data were corrected for the original sample volume. The retrieved number of molecules is low if considered that 10^{13} molecules were added into the translation step.

After these results, some improvements to the procedure for fusion production were considered. The first element to be addressed was the translation step. Until now a cell lysate had been exploited. However, the PURE system could be used for efficient protein production, since no nucleases are present. In a preliminary test, an estimated amount of 3 µg total RNA, i.e. uncrosslinked and crosslinked RNA, was translated for 4 h at 37 °C in a PURE system reaction (25.5 µL). The solution included also 20 U of RNase inhibitor. Then, salt addition (531 mM KCl and 50 mM MgCl₂) as Seelig protocol⁷⁹ was performed. After incubation at 25 °C for 5 min, the sample was oligo (dT) purified. An improvement was introduced at the end of the purification step. The eluted samples were filter-centrifuged for 3 min at 12,000 g in order to completely remove the oligo (dT) resin particles. The successive ethanol precipitation, reverse transcription and FLAG column were performed exactly as described above. Only the FLAG column buffers were slightly modified, since no β-mercaptoethanol was added. qPCR analysis gave as result 10^5 molecules in the eluted samples, considering 10^{13} sequences introduced in the translation. This result was next confirmed by analysis with the Fake and Fake-minus-His constructs. Translation was carried out with the cell lysate and PURE system and the number of retrieved molecules compared. In this experiment, conversely than before, new magnetic FLAG beads were

exploited. The main difference was due to the easier separation via the magnetic support. 50 μL of beads were used and were washed for three times with 1 mL of FLAG clean buffer (100 mM glycine pH 3.5 and 0.25% wt/vol Triton X-100) and with 3 x 1 mL FLAG binding buffer (50 mM HEPES-KOH pH 7.4, 150 mM KCl, 0.01% wt/vol Triton X-100, and 5 mM β -mercaptoethanol). Next, the beads were resuspended with 500 μL of FLAG binding buffer and with the translation reaction. After incubation of 1 h at 4 $^{\circ}\text{C}$ (tumbling), the supernatant was removed. Beads were washed six times with 1 mL FLAG binding buffer and finally, two elutions of 100 μL were performed accordingly to the FLAG purification protocol.⁷⁹ After the FLAG column step, ethanol precipitation and qPCR were done. Comparison between the retrieved molecules clearly indicated how the PURE system was a better system for translation in the mRNA display selection protocol (Fig.3.12).

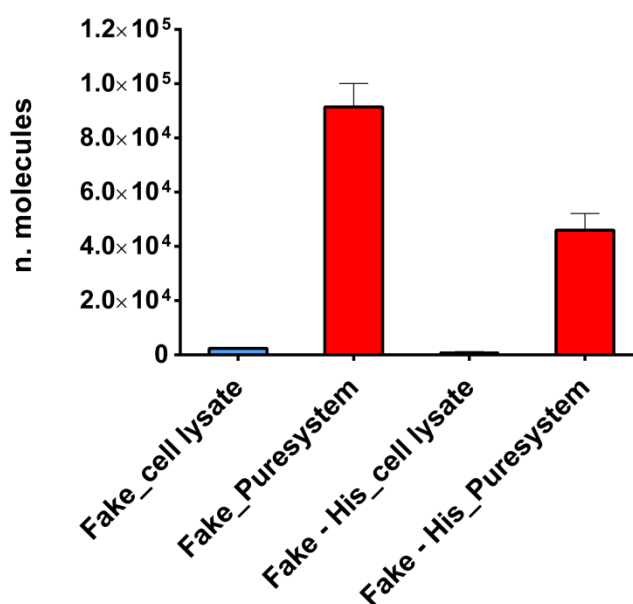


Fig. 3.12 qPCR analysis on retrieved number of molecules at the end of FLAG purification after exploiting different translation systems. Fake and Fake-minus-His (abbrev. Fake – His in the figure) constructs were translated exploiting a cell lysate for 4 h at 30 $^{\circ}\text{C}$ or the PURE system, for 4 h at 37 $^{\circ}\text{C}$. The qPCR measurements were done at the end of the FLAG column purification on eluted fractions. Samples were analyzed in triplicates. The PURE system showed an improvement in the retrieved molecules. Unexpectedly, Fake-His elution was lower than the Fake one.

Noticeably, these results also suggested a small difference between the Fake and Fake-minus-His construct. Introduction of the Fake-minus-His construct had been done for optimization of expression, since thermodynamic structure prediction showed how the RBS in the Fake construct RNA was closed by base pair annealing. Elution recovery after the mRNA display selection steps demonstrated that the possible RBS annealing did not affect translation for the Fake construct. Nevertheless, the high number of steps present between translation and qPCR analysis could also cause variations in the elution results.

3. ³⁵S- Methionine scintillation counting for fusion assessment

qPCR measurements gave insight into which *in vitro* transcription and translation system was better performing. However, the technique presented the limitation of monitoring only the final output of mRNA display, while intermediate steps were not considered. For this reason, a method that allowed assessment in every phase from translation to final amplification was preferred. Scintillation counting of labelled species permitted to directly visualize the fusion formation immediately after the translation. ³⁵S-Methionine was added to the PURE system reaction with Fake-minus-His RNA. Again, almost 3 µg of total RNA and 20.4 µCi of ³⁵S-Met were used. Incubation with salt was performed at the end of a 4 h translation reaction as Seelig procedure. New magnetic oligo (dT) beads were exploited for the purification of the RNA. Buffers were the same as the cellulose resin ones, without the addition of β-mercaptoethanol. 100 µL of beads were washed with oligo (dT) binding buffer for three times (1 mL each) and resuspended in 540 µL. After incubating 10 min in rotation at 4 °C, the translation reaction was added. Again, the mixture was incubated at 4 °C (20 min, still rotating). Finally, the supernatant was removed and the beads were washed initially with 6 x 1 mL oligo (dT) binding and then 6 x 1 mL oligo (dT) washing buffer. Fractions were collected. Next, four elutions of 400 µL each were carried out. Each of the fractions collected was then assessed by ³⁵S scintillation

counting. 1 μL sample aliquots were directly inserted in 1 mL of scintillation liquid. The oligo (dT) matrix specifically binds a linker region, allowing isolation of crosslinked RNA. Since ^{35}S -Methionine labels the Met-FLAG peptide, molecules that are visualized via scintillation counting in the oligo (dT) column elution are the RNA-peptide fusion. Results clearly demonstrated for the first time that fusion molecules were produced via translation and KCl and MgCl_2 salt incubation (Fig. 3.13).

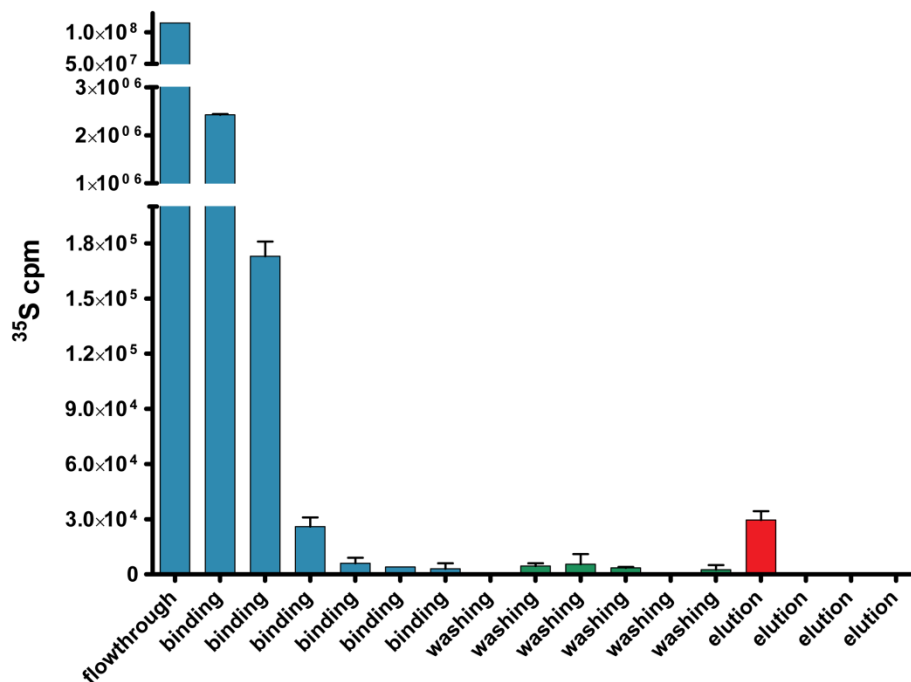


Fig. 3.13 ^{35}S scintillation counting analysis of fractions collected in the oligo (dT) column purification. The Fake-minus-His construct was translated in presence of ^{35}S -Met for 4 h at 37 °C. Fusion formation was done by incubation with KCl and MgCl_2 as previously described. Thereafter the reaction was oligo (dT) column purified. Aliquots were assessed for each fraction. Samples were read twice at the scintillator counter. In the flowthrough un-incorporated methionine was discarded. Fusion formation was visualized in the first column elution.

Next, the oligo (dT) elution was ethanol precipitated, thereafter reverse transcribed and finally purified by FLAG magnetic beads. Procedures were performed as described before (see page 101 for reverse transcription; see page 103 for magnetic beads FLAG purification). Fraction of the FLAG column were collected and analyzed by scintillation

counting. Here, 10 μL were exploited for the readings. Results showed that initially some of the fusions were lost since not binding to the resin. Nevertheless, signal was weakly detected, thus only one reading was done for each sample (Fig 3.14).

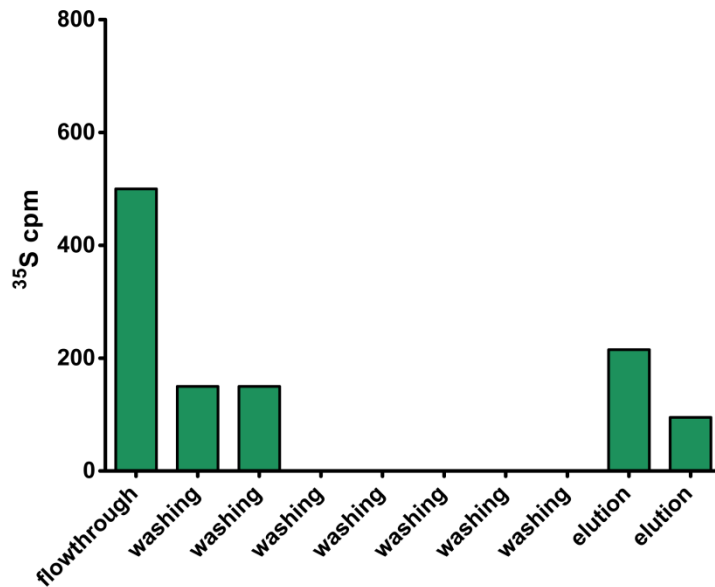


Fig. 3.14 ^{35}S scintillation counting data on FLAG column fractions. The experiment was performed with FLAG magnetic beads. At the end of the affinity column, 10 μL aliquots were measured. The obtained cpm values were low, thus samples were not read again. However, the signal recorded showed how molecules were retrieved in elution fractions from the column.

Direct monitoring of RNA-peptide fusion formation in the oligo (dT) purification allowed to optimize also the time required for translation. Until now, 4 h of PURE system reaction at 37 $^{\circ}\text{C}$ were generally performed. A ^{35}S -Met labelling experiment as just described was set up. Aliquots of translation reaction (6 μL) were taken at specific time points of 10 min, 30 min, 2 h and 4 h. Incubation with salt was done as previously indicated. In the oligo (dT) purification only 25 μL of magnetic beads were used for each sample, binding and washing steps were done with 500 μL of specific buffer, final elution volume was reduced to 100 μL . 10 μL were exploited for scintillation counting. Analysis of the first eluted fractions suggested that 30 min were optimal for translation to occur (Fig. 3.15).

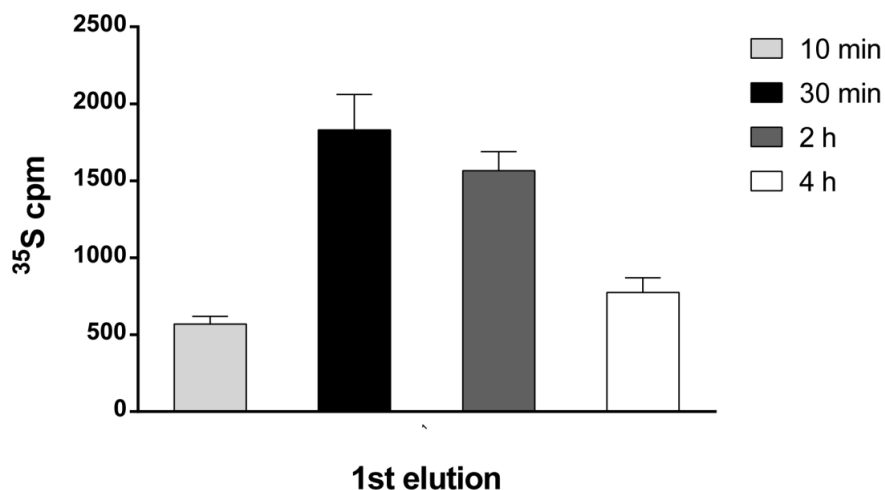


Fig. 3.15 ³⁵S scintillation counting for evaluation of optimal translation time with PURE system. 10 μ L aliquots of translation reaction were taken at time points of 10 min, 30 min, 4 h and 6 h. Samples were oligo (dT) purified and fractions were analyzed by scintillation counting. Here is presented the recorded radioactivity for the first oligo (dT) column elution of each time point. Incubation of 30 min at 37 °C was the optimal condition.

Considering the positive ³⁵S data obtained, a last experiment was set up to evaluate if Fake and Fake-minus-His constructs were retrieved from the mRNA display tested steps. Both samples were carried out as previously indicated. Translation reactions were performed incubating 30 min at 37 °C. Processing of the samples followed the usual flow, i.e. oligo (dT) purification, reverse transcription and finally FLAG selection column. For scintillation counting analysis, 1 μ L aliquots were taken for oligo (dT) purification monitoring, 10 μ L out of the reverse transcription reaction, 10 μ L of FLAG column fractions. Fake and Fake-minus-His sequences were successfully retrieved from the column (Fig. 3.16).

The results are reported as number of molecules. These values were calculated starting from ³⁵S-Methionine standard curve. In fact, scintillation counting was performed on dilutions of the free labelled amino acid. The total radioactivity and the methionine pmol used in a PURE system reaction were calculated according to the supplier protocol. Briefly, the total methionine pmol used were the sum of labelled (20.4 pmol) and un-labelled methionine (7500 pmol). The total counts in the reaction were calculated from the equation:

total counts per minute (cpm) registered for 5 μ L control dye multiplied for the reaction volume and divided by 5. For each elution fraction in the oligo (dT) column the correspondent pmol of incorporated methionine were calculated and the value divided for the number of methionine present in the constructs. Finally, the sum of eluted pmol was converted in the number of molecules. For reverse transcription samples and FLAG column fractions, a proportion with the oligo (dT) values of cpm and the corresponding number of molecules was calculated.

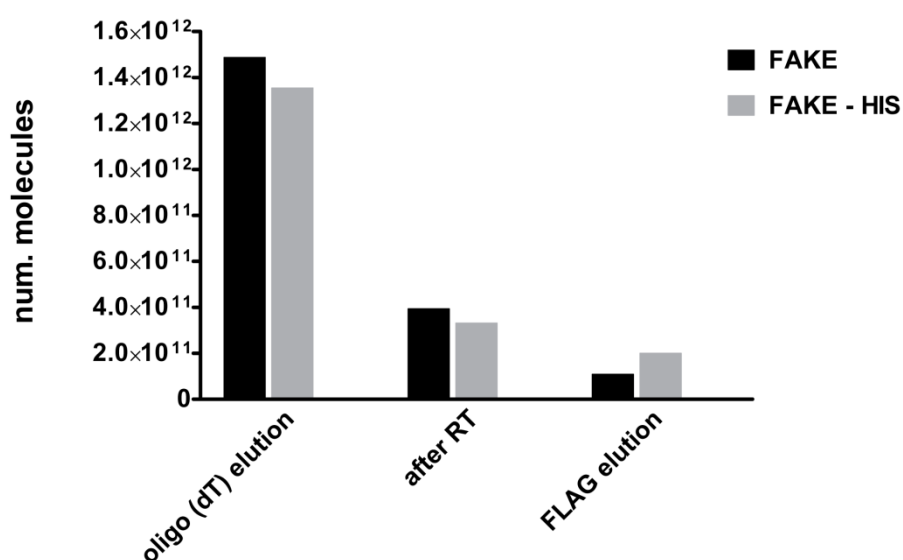


Fig. 3.16 Comparison of the number of molecules calculated for elution fractions of Fake and Fake-minus-His constructs. 35 S scintillation measurements were carried out on aliquots for each step of mRNA display, i.e. oligo (dT) column, reverse transcription (RT) and FLAG purification. From elution cpm values, the number of molecules was calculated. Average of two reading of each sample was considered. Final result after FLAG column purification demonstrated how the selection was working for both constructs.

From the results both constructs seemed to work. These data were similar to the qPCR results from a qualitative point of view. In fact, Fake and Fake-minus-His were both allowing for protein expression, and the thermodynamic prediction about Fake RNA structure to present the RBS annealed was discarded (for qPCR data, see page 103). Conversely, from a quantitative point of view, the number of molecules obtained at the end

of the cycle were different from the qPCR values. However, qPCR results were obtained without exploiting magnetic beads for both the purifications steps, therefore it was less easy to carry on. Moreover, by scintillation counting a direct monitoring of the intermediate passages was possible.

3.4.5 Amplification testing at the selection cycle end

At the end of the FLAG column, the product in the eluted fraction was a duplex constituted of a cDNA and RNA strand. The duplex was generated by reverse transcription before entering the FLAG purification. Thus, it did not present a T7 RNA promoter necessary for the new selection cycle. This addition and amplification of resulting sequences was performed via PCR on eluted sequences. An ethanol precipitation was required for reducing the sample volume before amplification.

cDNA formation by reverse transcription was initially evaluated, thus excluding amplification failure due to the template. Reverse transcription was carried out on total RNA (crosslinked and uncrosslinked) of a MG library. Three samples were prepared (A, B, C) in 40 μ L reverse transcription reactions. After the incubation for 30 min at 42 °C, samples were evaporated with a speedvac. Resuspension was in 10 μ L DEPC-water. Sample B was treated with RNase A for 30 min at 37 °C. Since in this experiment, total RNA, i.e. crosslinked and uncrosslinked, was used, RNase treatment of sample B removed all the RNA from the reaction. Thus, only the cDNA would be present. Sample C was a negative control, since no reverse transcriptase enzyme was present. Products were loaded on a 2% agarose gel and results evaluation confirmed the presence of the cDNA band in sample B (Fig. 3.17). Double bands were observed for sample A and C, indicating the presence of the RNA in the crosslinked and uncrosslinked forms. The experiment demonstrated that the reverse transcription reaction was functional and cDNA was produced.

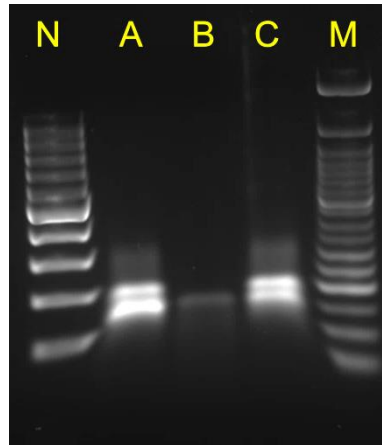


Fig. 3.17 Reverse transcription assessment. A MG DNA library was transcribed and the RNA crosslinked. From this template three reverse transcription reactions were performed. Results were run on a 2% agarose gel. Sample B was treated for 20 min at 37 °C with RNase A. The resulting band confirmed cDNA production. Sample C was a negative control since no reverse transcriptase was added. The displayed bands correspond to the total RNA template, i.e. crosslinked and uncrosslinked RNA. The enzyme catalyzed the cDNA production (sample B). Moreover, sample A was not treated for RNA digestion, production of cDNA is indicated by the higher intensity of the lower band, compared to sample C. Marker (M) is a 50 bp ladder (NEB), Marker N is a 100 bp ladder (Fermentas).

For amplification at the end of the cycle were exploited two different DNA polymerases, i.e. DeepVent-R and Phusion. Taq polymerase was not considered for the high rate of mutation insertion presented. The forward primer contained an overlapping region on the 5'-end of the duplex and a T7 promoter for allowing transcription of the amplified molecules. The PCR test was directly carried out at the end of the first cycle with the MG library. Roughly 10 ng of the elution product was amplified in 50 μ L reactions, exploiting 0.2 mM dNTPs, 0.5 μ M primer couple. 0.5 U of DeepVent-R and 1 U Phusion DNA polymerase were added to the reactions. Buffer Thermopol and High Fidelity were used for DeepVent-R and Phusion reactions respectively. 20 cycles were carried out. For DeepVent-R polymerase, the annealing temperature exploited was 54 °C, for Phusion reaction a 3 step protocol was followed, with an annealing temperature of 71 °C. The difference in temperature between the two PCRs was mainly due to previous experiments performed on the initial library amplification. At the end all the reactions were run on a 2%

agarose gel. DeepVent-R sample did not present any amplification product (Fig. 3.18). Conversely, the Phusion sample showed two similar bands close one to the other. The lower band was at the right height of the full initial library. From a preliminary amplification test carried out on a double strand assembled MG library, the amount of template was correlated with the increase in the second band production (Fig. 3.18, panel b). Possibly, the higher band could be constituted by two DNA strands forming an hetero-duplex. Since the starting material for the PCR was a mix of sequences that could base pair at the similar termini, although the sequences were different in the middle region, the hetero-duplex could form when the PCR reaction would not perform efficiently. For these reasons, after gel extraction the selection proceeded only with the lower specie. Moreover, the amount of template was reduced for the amplification in the selection.

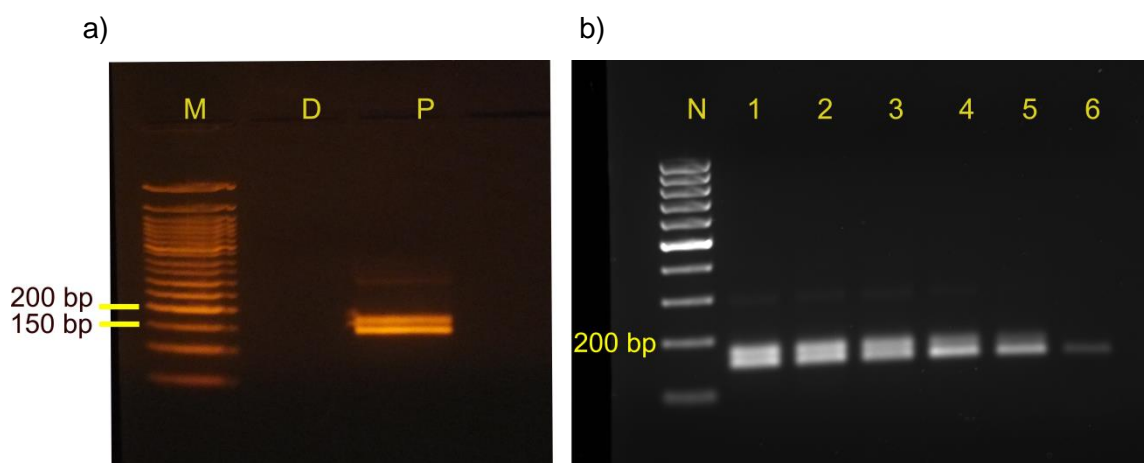


Fig. 3.18 Amplification reaction optimization. Panel a) PCR was performed with Deepvent R and Phusion polymerase (samples D and P, respectively) on the duplex, after the first selection cycle with MG library. All the amplification reaction was loaded on the gel for gel extraction. Deepvent-R polymerase did not work, while Phusion gave two close band as result. The upper band is probably caused by the template amount. Panel b) PCR amplification with Phusion polymerase on the assembled library. Samples contained decreasing amount of double strand template. From sample 1 to 6, 50 ng, 10 ng, 1 ng, 10 pg, 1 pg, 10 fg respectively were used. Marker (M) is a 50 bp ladder (NEB), Marker N is a 100 bp ladder (Fermentas).

3.4.6 Control experiments concerning the MG ligand

The preliminary experiments for the technique setting have been mostly performed with the Fake and the Fake-minus-His constructs. However, investigation if MG could interfere in the riboswitch selection was carried out. MG was tested for possible inhibition effect on the transcription and translation system. Exploiting the PURE system, production of a cyan fluorescent protein (mCerulean) in the presence of increasing concentrations of MG was assessed. The experiment was performed via the qPCR instrument (CFX96 Real time PCR Detection System, Bio-Rad), monitoring in FAM channel for 6 h at 37 °C. 10 μ L reactions were prepared and 10 ng of plasmid template were exploited. MG concentrations were 0.1 μ M, 0.32 μ M, 1 μ M, 10 μ M, respectively. Controls reactions with MG and water were also measured for comparison. The fluorescence signal of mCerulean was only slightly affected by MG concentration (Fig. 3.19). However, the final concentration exploited in the selection was 100 μ M. This higher concentration was not considered when performing *in vitro* transcription and translation mostly because the color of the solution was green. This could affect the fluorescence measurement, thus it was not tested directly.

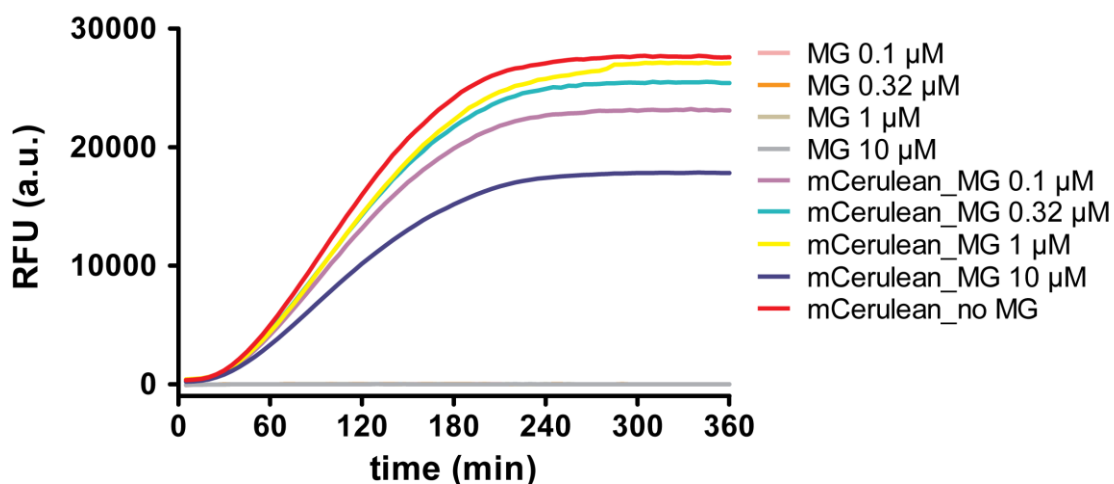


Fig. 3.19 Fluorescence measurements in time of mCerulean production in PURE system in presence of MG. Several concentrations of MG were tested. Data were collected in a qPCR instrument, recording with the FAM filter. MG alone was not fluorescent. Samples MG 0.1 μ M, MG 0.32 μ M, MG 1 μ M fluorescence measurements are not visible in the graph since the emission was low. PURE system protein expression was little affected by MG addition.

3.5 mRNA display selection of a malachite green responsive riboswitch

The selection for a MG responsive riboswitch was performed considering the experiment tests previously shown. In order to select for a riboregulator presenting a modulated RBS depending upon the presence or absence of the ligand, cycles were carried out with and without MG in the translation reaction. In particular, the selection strategy was designed considering initial presence of the ligand (ON cycles). However, conditions were modified through the selection according to the number of molecules that were retrieved after each cycle (see Fig. 3.20). OFF cycles were also performed, i.e. cycles 3 and 4. In an OFF cycle, MG was not added to the translation reaction. This strategy was defined in order to decrease the number of possible molecules that presented a conformation with an open RBS, which would allow the expression of the FLAG peptide and fusion formation without being regulated by the ligand presence. The OFF cycles acted as counter selection on the molecules pool. In absence of MG, the RNAs to be retrieved in the (counter) selection were those molecules that did not produce fusions. Thus the flowthrough of the FLAG column, containing the unbound sequences, was collected. Nevertheless, a drawback of performing OFF cycles was that RNAs with open RBS that were not able to form fusion efficiently were also collected. In fact, only a small percentage of fusions is produced at the end of the translation reaction. After two OFF cycles, an ON cycle was performed (cycle 5). From cycle 6 on the strategy of the selection was modified, although all the following cycles were performed including MG in the translation reaction. In particular, the reaction conditions became more stringent, in order to enrich for functional riboregulators. For example in cycle 6 the number of DNA molecules added in the transcription reaction was reduced. From cycle 7, the translation reaction was incubated for only 5 min instead than 30 min as previously done.

Details on conditions applied and the selection methods are reported in the following sub-sections.

1. *Library assembly*

At the beginning of the selection, the DNA pool was assembled by overlapping extension of oligonucleotides. The reaction (50 μ L) was performed with 2.5 pmol of each oligonucleotide. DreamTaq buffer, 0.2 mM dNTPs and 2.5 unit of DreamTaq polymerase were used. Initial denaturation was carried out for 10 min at 94 °C without the polymerase. Then, after polymerase addition, the reaction was incubated at 55 °C for 15 min and at 72 °C for 45 min. Just one cycle was performed, in order to avoid over-amplification of the pool. Next, the overlapped double strand product was extracted from a 2% Nusieve GTG agarose gel in TAE buffer, and column purified. The final volume of column elution was 50 μ L in sterile water. The obtained double strand product was then amplified by PCR with DeepVent polymerase (3 U) in Thermopol buffer in a 300 μ L reaction. The sample was aliquoted in six sample prior thermocycling. Primers bound specifically the 5' and the 3' of the construct. 20 cycles of amplification were carried out. Bands were gel extracted on 2% low melting agarose and finally combined.

2. *Transcription and DNase treatment*

10 pmol of DNA library was transcribed with 150 U of T7 RNA polymerase in T7 buffer (35 mM $MgCl_2$, 2 mM spermidine, 200 mM Hepes-KOH, pH 7.5). The reaction (50 μ L) included also 40 mM freshly prepared DTT, 5 μ g BSA, 5 mM each NTP, 0.05 U Inorganic Pyrophosphatase, 20U RNase inhibitor. The reaction was incubated at 37 °C for 4 h. From the second cycle till the fifth included all the PCR product obtained from the previous round was transcribed. For rounds six, seven and eight only 1.5 ng of the obtained PCR product (corresponding to 10^{10} molecules) were transcribed in a 50 μ L reaction. How this could affect the selection progress is discussed in the Result section.

At the end of transcription, 37 mM EDTA and 0.5 mM $CaCl_2$ were added to the reaction and the sample was processed with a RNase-free DNase I treatment to remove template DNA. The degradation was performed at 37 °C for at least one hour.

3. PAGE gel for RNA purification

Thereafter, the RNA samples were first subjected to ethanol precipitation. This step allowed the removal of residual components of the transcription reaction and permitted to reduce the solution volume. The ethanol precipitation was performed without the addition of salt. A wash with 100% ethanol was done and 20 min centrifugation at 4 °C allowed the formation of the RNA pellet. Residual ethanol was eliminated by incubation for 5 min at 70 °C of the RNA. Finally, the RNA was resuspended in 25 µL DEPC-water. 8 M Urea, 4% PAGE was exploited for RNA purification. The 2X loading dye contained 8 M urea, 20% wt/vol sucrose, 0.1% SDS, 0.05% xylene cyanol, 0.09 M Tris, 0.09 M borate, 10 mM EDTA. Samples were denatured for 5 min at 94 °C before loading on the gel. The gel was prepared and run in TBE buffer for 5 h at 150 V. Next, the bands were visualized by UV shadowing of the gel. RNA was then extracted from the polyacrylamide matrix by crush and soak. Briefly, the gel slices were crushed with a filter tip in a microcentrifuge tube. Then, 500 µL of TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) and 120 U of RNase Inhibitor were added. The reactions were left tumbling at 37 °C overnight.

The following day the crush and soak procedure was continued. Initially the sample was centrifuged for 5 min at maximum speed and the supernatant collected. Next, 250 µL of TE buffer were added again to the solution. After 2 h at 37 °C, centrifugation allowed to isolate the new supernatant, which was combined with the first one obtained. Finally, the solution was ethanol precipitated (with addition of sodium acetate) and resuspended in 20 µL DEPC-water.

4. Crosslinking reaction

Photo-crosslinking was performed with 7.5 µM puromycin-contained linker on the purified RNA in buffer XL (100 mM KCl, 1 mM spermidine, 1 mM EDTA pH 8.0, 20 mM HEPES-KOH pH 7.5). The volume of the reaction was 50 µL. The linker was annealed to the mRNA for 3 min at 70 °C in the thermocycler, and the reaction was cooled to 25 °C in 5

min. Then, the sample was transferred into a 96 well transparent plate kept on ice and irradiated with a UV lamp at 365 nm for 20 min. Finally, the RNA was ethanol precipitated with addition of 3 M Sodium acetate, pH 5.5 and 2.5 V of ethanol 100%. After centrifugation, the pellet was washed with 70% ethanol and centrifuged again. Final resuspension was in 10 μ L of DEPC- treated water. In every selection cycle, crosslinking efficiency was tested by native 2% agarose gel on denaturated samples. A 1 μ L aliquot was tested before proceeding with the following steps of selection. Denaturation was due to the formaldehyde-containing loading dye. The sample was incubated at 70 °C for 10 min prior loading. Gel was run in 1% TAE buffer at 120 V for 30 min. An RNA ladder (low range, i.e. 100-1000 bp) was also included.

5. *Translation and fusion formation*

During the selection, the PURE system was exploited as *in vitro* transcription and translation system. Reactions (25.5 μ L) were carried out with the total RNA (crosslinked and uncrosslinked). A direct quantification of the RNA amount was not performed. However, estimations were done according agarose gel analysis on the crosslinked sample. To the PURE system reaction, 20 U RNase inhibitor and 20.4 μ Ci 35 S-Met were supplemented. In rounds 1, 2, 5, 6, 7 and 8 100 μ M MG was included. Thus in these cycles fusion formation was expected. Conversely rounds 3 and 4 were performed in absence of MG. Incubation time of the translation sample at 37 °C was 30 min, as previously established. However, to increase the stringency of the selection, rounds 7 and 8 were performed incubating just 5 min at 37 °C. Thereafter mRNA-peptide fusions were produced by addition of 531 mM KCl and 50 mM MgCl₂ and incubating for 5 min at room temperature. The fusion complexes were stored at -20 °C unless immediately processed.

6. *RNA isolation via oligo (dT) column*

After fusion formation, the crosslinked RNA-peptide complexes were purified from translation components via an Oligo (dT) affinity column. 100 μ L oligo (dT) magnetic beads

were exploited. Beads were washed three times with 1 mL with oligo (dT) binding buffer (1 M NaCl, 10 mM EDTA, 0.2% wt/vol Triton X-100, 20 mM Tris-HCl pH 8, without β -mercaptoethanol). Finally the resin was resuspended in 540 μ L binding buffer. After incubating 10 min in rotation at 4 °C, the translation reaction was added. The mixture was tumbled for 1 h at 4 °C. Thereafter, upon magnetic separation, the supernatant was removed and beads were washed six times with 1 mL oligo (dT) binding buffer each, and then six times with 1 mL oligo (dT) washing buffer. Fractions were collected. Next, four elution of 400 μ L each were carried out. Oligo (dT) purification was assessed by 35 S scintillation counting of fraction aliquot (1 μ L sample in 1 mL scintillation liquid). Next, elution fractions were combined and ethanol precipitated by addition of 3 M Sodium acetate, pH 5.5 and 2.5 V of ethanol 100%. Sample was centrifuged for 30 min, the pellet then was washed with 70% ethanol and centrifuged again. Finally the sample was resuspended in 20 μ L of DEPC- treated water.

7. Reverse transcription and selection column

Samples were reverse transcribed for 30 min at 42 °C. The reaction (40 μ L) exploited 4 U of Superscript II reverse transcriptase, 0.5 mM dNTPs, 0.1 U/ μ L RNase inhibitor, 50 mM Tris-HCl pH 8.3, 3 mM MgCl₂, 500 nM specific reverse primer. An aliquot was measured by scintillation counting. The reverse transcribed sample was transferred to an anti-FLAG peptide resin. 50 μ L of magnetic beads were initially prepared by washing for three times with 1 mL of FLAG clean buffer (100 mM glycine pH 3.5 and 0.25% wt/vol Triton X-100) and with 3 x 1 mL FLAG binding buffer (150 mM KCl, 0.01% wt/vol Triton X-100, 5 mM β -mercaptoethanol, 50 mM HEPES-KOH pH 7.4). Separation was done by magnetic support. Finally, beads were resuspended with 500 μ L of FLAG binding buffer. Reverse transcribed sample was added and left for 1 h at 4 °C while tumbling. Next, beads were washed six times with 1 mL FLAG binding buffer. Elution was performed with 2 x 100 μ L of FLAG binding buffer containing 2 equivalents of FLAG peptide. The mixture was left tumbling for 20 min 4 °C each time prior to separation. Fractions were collected and

analyzed by scintillation counting. For rounds in which MG was present in the translation reaction, both the elutions were combined and then processed. For cycles 3 and 4, where MG was not added in translation, the first two washes of the FLAG binding beads were collected, since the functional sequence were not retained on the column. However, molecules that would not form fusions efficiently would also be present in the washes. Thereafter, the sample was ethanol precipitated by addition of sodium acetate and ethanol (see above). Final resuspension was done in 10 μ L sterile water (not DEPC-treated for avoiding interference in the amplification reaction).

8. PCR Amplification reaction

In the first cycle 10 ng of the elution product were amplified in a 50 μ L reaction. 0.2 mM dNTPs, 0.5 μ M each primer, 1 U Phusion DNA polymerase in buffer High Fidelity were exploited. The amplification was carried out for 20 cycles, with an annealing temperature of 71 °C. The PCR was performed according to a the three step protocol of the Phusion polymerase manufacturer (Finzymes). The amount of starting material was changed during the selection rounds. From second to fifth cycle included, the duplex template was reduced to 100 pg. Thereafter only 0.15 pg were used. The amount variation of starting material in amplification is discussed in the next section. Finally, the amplification product was gel extracted on a 2% low melting Nusieve GTG gel. The gel was run in TBE buffer for 30 min. Bands were then column purified.

3.6 Selection results and discussion

Eight rounds of selection were performed in order to develop a MG responsive riboswitch. The progresses in the selection cycles were followed by ^{35}S -Methionine scintillation counting. From the cpm values obtained by scintillation, the number of molecules was calculated. The calculation was done via a standard curve with free ^{35}S amino acid for every experiment. Moreover, the decay factor was considered. The total counts were obtained dividing the counts in a PURE system reaction by the decay factor of ^{35}S -Met. Through Specific Activity estimation, the pmol of incorporated methionine was determined. The final number of molecules was obtained as the sum of the elution fractions considered in the experiment. While oligo (dT) column fractions were assessed by standard curve calculation, FLAG purified samples were determined by proportion with the molecules resulting after oligo (dT) column and cpm. Cpm recorded by the scintillator counter were corrected for the sample volume prior to analysis.

In order to evaluate the selection progress, the total number of molecules retrieved in every round at the end of the FLAG binding column was estimated. Initially, an increase in the molecules obtained round after round was expected. However, results were not showing any pattern through the selection (Fig. 3.20). The number of molecules was almost constant around the value 10^9 for first cycles, then a drop was observed in the last cycles. This decrease was possibly due to the reduced time allowed for translation to occur in rounds 7 and 8.

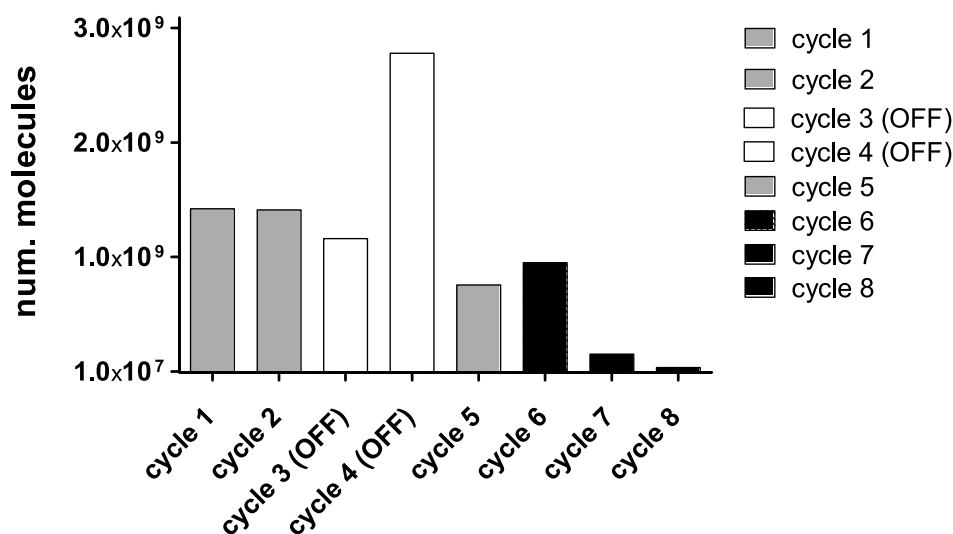


Fig 3.20. Number of molecules retrieved for each selection cycle at the end of the FLAG affinity column. Data were calculated from ³⁵S scintillation counting. OFF cycles (white) were performed without MG in the translation reaction, thus washing fractions were collected and analyzed. Black color indicates rounds where stringent conditions were applied in the selection.

Considering the selection results, the number of molecules was not taken into account as parameter for evaluation of the progression. Generally, in functional nucleic acids selection, a comparison between the fraction of the pool entering the selection step and the eluted sample is evaluated. In other words, what enters into the affinity column is related to the output of the column in terms of binding ratio. Calculations of the ratio between the radioactivity recorded on an aliquot of reversed transcribed sample, i.e. what enters into the FLAG affinity column, and the sum of elutions cpm were performed only for rounds where MG was added in the translation reactions (Fig. 3.21). These positive cycles were only considered since in the third and fourth rounds, i.e. where no ligand was present, The fractions not bound to the FLAG column were retrieved. Thus, a binding ratio could not be calculated. Again, no significance in the ratio was achieved. Some values obtained were high, indicating a ratio of 100% molecules retrieved. Moreover, the cpm radioactivity in the

reverse transcription sample was not the same as the sum of the cpm obtained reading the FLAG column fractions. Thus a different way to measure directly each step was needed.

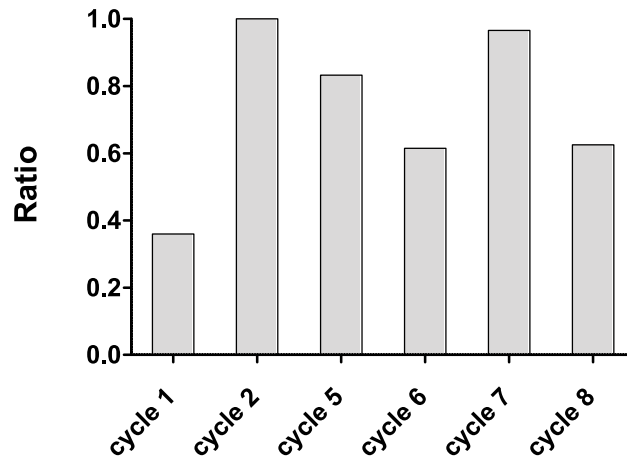


Fig. 3.21 Ratio of recorded radioactivity in the FLAG column elution over the reverse transcribed sample signal. Only positive selection cycles were evaluated. Ratio equal to 1 meant that cpm counts in the reversed transcribed sample were the same as those cpm measured on the elution sample. This result was not possible considering the radioactivity present in other fractions in FLAG column.

Since the comparison between radioactivity in the reverse transcribed and eluted samples was obtained by independent measurements and done on samples with different volumes, thus the error due to volume correction of the cpm could affect the ratio, the estimation was not taken into account. Next, the total radioactivity recorded in the FLAG column was considered as reference parameter and a new ratio was calculated. The meaning of this estimation was to consider just the FLAG column results as possible indicator of enrichment of sequences in the pool. This parameter was also free from calculation upon the number of molecules, which could have introduced bias in the estimations. Again, only the ON cycles were included in the analysis (Fig. 3.22), since crosslinked RNA that did not produce a fusion complex would be collected in the first two washing fractions of the FLAG peptide binding purification. The result of this analysis were finally consistent with stringent conditions applied during the selection progress. For example, the second cycle decrease in radioactivity recorded could be caused by the lower

number of eluted molecule that were amplified at the end of the first selection cycle.

However, the percentages of binding in the FLAG column were low and an exponential increase was not achieved.

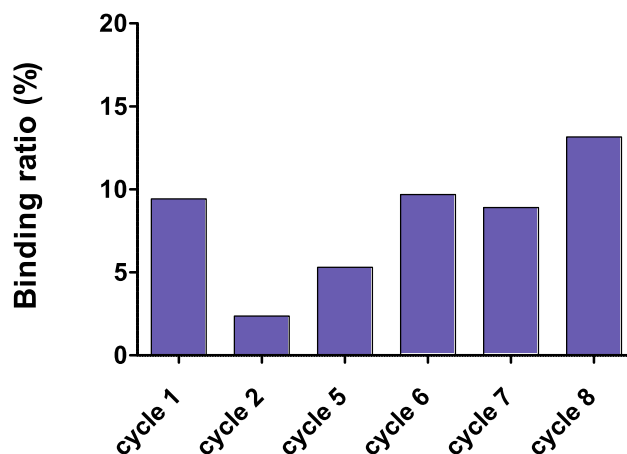


Fig. 3.22 Binding ratio in the FLAG affinity column. The ratio was calculated as the fraction of ^{35}S radioactivity measured in the elution samples over the total radioactivity recorded per cycle, i.e. the sum of cpm in FLAG column. Only ON cycles were considered. Unfortunately, the binding percentage was low and no enrichment was shown.

Together the results substantially highlighted a failure of the mRNA display selection for a new MG responsive riboswitch. Confirmation was obtained by next generation sequencing (NGS) data. NGS was carried out on the amplified product after cycle 8 with an ION Torrent Personal Genome Machine sequencer (CIBIO Core Facility, Trento). The data set consisted of 5740 sequences, which were initially aligned and evaluated by a quality score. Unfortunately, sequencing results showed that the frequencies of the nucleotides in the randomized positions were almost equivalent and close to a random distribution (Fig. 3.23). Thus the analysis did not indicate the presence of specific repeated domain that could highlight a selection convergence on a candidate, or a cluster of sequences.

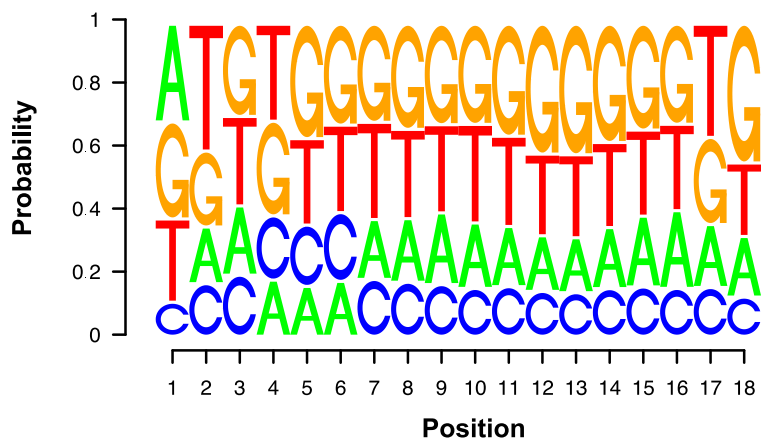


Fig. 3.23 NGS sequencing results. Nucleotide frequencies confirmed the absence of a common sequence or subsequence. For each of the 18 randomized position in the library, the frequency of a specific nucleotide was close to uniform distribution. Data analysis was performed by Dr. Michele Forlin.

At the end of the selection and considering that the selection for a new riboswitch had not been successful, some improvements could be taken into account for mRNA display evolution of riboregulators. Probably, the most important point to evaluate was that the amount of crosslinked RNA varied before entering the translation. In other words, a quantitative step was necessary before the translation reaction. The agarose gel was used as reference for assessing crosslinked RNA presence and for quantification via band intensity analysis (via ImageJ software). However, despite gel bands quantification, all the RNA was added in the translation step. Translation and consequent fusion formation were the most important steps in the mRNA display technique when selecting for a riboswitch. Functionality of the RNA sequence was determined at this point. The following steps were purifications for isolation of the RNA-peptide complex. A variation in the number of molecules entering the translation could affect the final output. The same conclusion was suggested also for other two steps: the amplification reaction and transcription. Even if the concentrations of templates exploited for each reaction were known, parameters were modified through the selection rounds. Again, variations influenced the following steps, affecting the number of molecules in each cycle.

Considering the amplification reaction, some improvements should be pursued. At the end of every round, 20 cycles of PCR on a very small part of the elution product were performed. Thus only some sequences were exponentially amplified, influencing the pool behaviour in the selection. In order to avoid the over-amplification of a limited number of sequences, all the sample resulting from the previous step should be used and the amplification should be performed for a lower number of cycles. This condition need to be screened in order to avoid also possible unspecific bands. In the selection, a low intensity band, higher molecular weight specie was always visualized on the agarose gel at the end of the amplification reaction (Fig. 3.18, panel a). To prevent possible carry-on of unwanted molecules and DNA hetero-duplex formation the amplification had to be optimized.

A limitation that the mRNA display technique presented was the high number of steps and intermediate purifications. Excluding the issue about time required for one cycle, complete recovery of the sample in these phases was not always successful, thus causing loss of potentially functional sequences. For example, PAGE purification of the transcribed RNA could in principle be substituted by a column purification. Commercial kits were available, even if a possible limitation is the length of the RNA. When designing the library sequence, a shorter length construct was preferred (145 bp), in order to avoid secondary structures. Nevertheless, a 200 bp RNA sequence could be efficiently purified via columns. Moreover, reorganization of the cycle steps would also improve the final output, e.g. reverse transcription and amplification could be performed together at the end. The choice of reverse transcribe all the molecules before entering the FLAG affinity column was done to reduce possible interaction between the RNA and the resin. However, performing reverse transcription and amplification together would avoid another ethanol precipitation purification after the oligo (dT) column. Furthermore, quality control of the mRNA-displayed peptide would need to be carried out, optimizing SDS PAGE analysis.

In conclusion, mRNA display could not be used to select successfully for riboswitches. Taking into account the improvements just described, the result could have

been different. Nevertheless, the method presented many steps that made the mRNA display procedure costly and time consuming. For these reasons, a new approach for *in vitro* riboswitch selection was exploited (see Chapter 4).

Chapter 4.

In vitro selection of RNA-based sensors via ligand triggered strand displacement

This chapter presents an original manuscript that is in preparation for submission (sections 4.1- 4.2).

Title of the manuscript: An in vitro selection for RNA sensors based on small molecule triggered strand displacement

Laura Martini, Adam J. Meyer, Jared W. Ellefson, John N. Milligan, Michele Forlin, Andrew D. Ellington & Sheref S. Mansy

This chapter describes an approach for *in vitro* selection of riboswitches different from the mRNA display method presented in Chapter 3. mRNA display is a protein-based method. Here is introduced a novel procedure for the *in vitro* selection of RNA sequences that exploits directly the conformational shift induced by a ligand binding in the structure of the RNA sensor. The technique is based on strand displacement reaction.

Initially the chapter presents the strand displacement reaction, and how the ligand induced conformation shift in a RNA sensor, e.g. thiamine pyrophosphate (TPP) responsive riboswitch, is able to trigger strand displacement. Then, an *in vitro* selection to identify a TPP dependent RNA sensor is described. The results establish that it is possible to *in vitro* select new RNA sensors by conformational change (induced by the ligand binding) triggering of strand displacement. Thereafter, this technique for the selection of novel sensors is applied to select RNA sequences responsive to three TPP analogues.

4.1 A ligand induced conformational shift triggers strand displacement for RNA sensor selections

Functional nucleic acid sequences often transition between different conformational states that correlate with different levels of activity. For example, naturally occurring attenuator sequences and riboswitches exploit conformational changes in response to metabolite availability to control gene expression. The construction of analogous, artificial systems *in vitro* has been difficult in part because typical nucleic acid selections are designed to either enrich sequences that bind specific molecules^{62,63} or display specific catalytic activity⁸¹ without constraint on conformation. However, by combining ligand-binding aptamer and catalytic ribozyme domains, non-natural nucleic acids that display conformationally dependent activity can be designed^{57,82} and selected.^{45,47,83,84}

Rather than having ligand binding to an aptamer domain control the activity of a conjugated ribozyme, the conformational change induced by ligand binding could be used

to drive coupled strand displacement reactions⁸⁵ (Fig. 4.1). Strand displacement simply exploits the ability of single-strand DNA or RNA to displace one strand of a preexisting duplex nucleic acid.⁸⁶ Often the initial duplex nucleic acid is modified with a fluorophore - quencher pair and thus functions as a reporter of the reaction.^{87,88} Displacement is kinetically driven by complementarity between sequences of the reporter and the displacing nucleic acid. The reporter typically contains a small, single-strand region (toehold) that largely mediates the initiation of strand displacement. The rate of the reaction depends upon the length and sequence composition of the toehold,⁸⁹ which is usually between 4 and 8 nucleotides long with an optimum centered around 6 nucleotides.⁸⁷ Strand displacement is widely used in DNA nanotechnology,⁹⁰ including structural DNA assemblies,⁹¹ dynamic autonomous devices,^{92,93} and isothermal nucleic acid amplification methods.^{94,95} However, nearly all reactions based on strand displacement are driven by the initial binding of an oligonucleotide to a toehold sequence. In other words, it is simply the presence or absence of the oligonucleotide that dictates whether the strand displacement reaction proceeds or not. Nevertheless, the binding of an oligonucleotide sequence does induce a conformational change that modulates the activity of a target nucleic acid.⁹⁶⁻⁹⁹

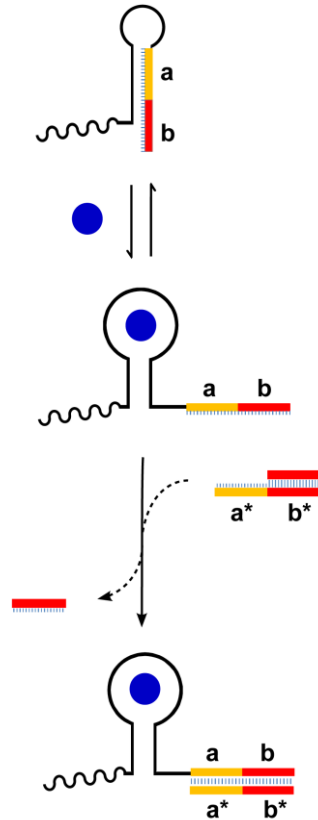


Fig. 4.1 Ligand binding to the RNA triggers a strand displacement reaction. When domain a of the RNA becomes accessible due to the binding of TPP (circle), the strand displacement reaction can proceed. The interaction between the RNA and the DNA reporter is based on sequence complementarity. The reporter is partly double stranded with a single strand toehold region (a*) that interacts with the complementary sequence of domain a of the RNA. Upon toehold hybridization, RNA domain b binds to the complementary domain b* of the reporter, displacing the shorter strand.

An *in vitro* selection strategy was developed to identify RNAs that mediate strand displacement in response to the binding of a small molecule. The RNA library was based on the thiamine pyrophosphate (TPP) riboswitch.⁶⁷ Ligand binding to the riboswitch liberated a single strand region capable of hybridizing to and displacing a duplex reporter coupled to magnetic beads. Iterative rounds of selection showed improved strand displacement activity. This fully *in vitro* methodology allowed for the selection of RNA molecules based on ligand induced conformational changes.

4.1.1 Methods

1. Chemicals, Oligonucleotides, DNA and RNA production

All chemicals were purchased from Sigma-Aldrich, unless otherwise indicated. Oligonucleotides were from Integrated DNA technology. DNA templates for transcription were prepared by PCR of overlapping oligonucleotides with AccuPrime Pfx DNA polymerase (Life Technologies). PCR products were purified by extraction from an agarose gel consisting of 1:1 SeaKem GTG and NuSieve GTG agaroses (Lonza) with Wizard SV Gel and PCR Clean-Up System (Promega). A list of all the constructs is included in Table 4.1.

Transcription buffers were prepared with diethyl pyrocarbonate (DEPC)-treated water. Transcription reactions were in a final volume of 50 μ L and contained 10 pmol DNA, T7 RNA polymerase buffer (35 mM MgCl₂, 2 mM spermidine, 200 mM HEPES, pH 7.5), 40 mM DTT, 5 μ g BSA (New England BioLabs), 5 mM each ribonucleotide (NEB), 20 U Human placenta RNase inhibitor (NEB), 50 U yeast inorganic pyrophosphatase (NEB), 150 U T7 RNA polymerase (NEB). Reactions were incubated at 37 °C for at least 4 h. Samples were then treated with RNase-free DNase I (NEB) for 1 h at 37 °C. Next, RNA was purified from a 6% denaturing (7 M Urea) PAGE. The acrylamide-bis acrylamide solution (19:1) was from Bio-Rad. The RNA was visualized by UV shadowing and isolated by crush-soak. Briefly, the excised RNA band was crushed and left overnight at 37 °C with tumbling in 500 μ L TE buffer (1 mM EDTA, 10 mM Tris- HCl, pH 7.5). The samples were centrifuged and the supernatants filtered through 0.45 μ m Ultrafree-MC centrifugal filters (Millipore). Finally, the samples were ethanol precipitated and resuspended in 30 μ L DEPC-treated water. Concentrations were determined by UV absorbance with a NanoDrop 1000 Spectrophotometer (Thermo Scientific). RNA molecules were stored at -20 °C.

Table 4.1 DNA sequences used.

Fluorophore- quencher reporter*	
Rep F	5'-/5FluorT/ GTGATGGT GCGATCCCATAGTTAATTTCTCCT-3'
Rep Q	5'-AATTA ACTATGGGATCGCACCATCAC/3IABkFQ/-3'
Selection reporter	
Rep Biot	5'-/52-Bio/ GTGATGGT GCGATCCCATAGTTAATTTCTCCT-3'
Rep Displ	5'-AATTA ACTATGGGATCGCACCATCAC-3'
oligo (dT) ₂₁	5'-TTTTTTTTTTTTTTTTTTTTTTT-3'
DNA Constructs	
sequence ON	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGCTGACTTACATTATGACATCGAAAAT AGTACTGGAACCAACTGCAGTAC <u>AGGAGAA</u> TTAACTATGGGATCGCACCATCAC-3'
sequence OFF	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGCTGACTTACATTATGACATCGTCTCC TGTA CTCAAACCATCACTAAAGTAC <u>AGGAGAA</u> TTAACTATGGGATCGCACCATCAC-3'
+ThiM#2	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGAATTGTGAGCGGATAACAATTGAAT TCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGA T <u>CTGG</u> GATAATGCCAGCGTAGGGAAGCTATTACAAGA <u>AG</u> ATC <u>AGGAGAA</u> TTAACTATGGG ATCGCACCATCAC-3'
+ThiM#2 mut1	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGAATTGTGAGCGGATAACAATTGAAT TCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGA T <u>CTGG</u> GATAATGCCAGCGTAGGGAAGCTATTACAAGAT <u>TC</u> ATC <u>AGGAGAA</u> TTAACTATGGG ATCGCACCATCAC-3'
+ThiM#2 mut2	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGAATTGTGAGCGGATAACAATTGAAT TCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGA T <u>GAGG</u> GATAATGCCAGCGTAGGGAAGCTATTACAAGAT <u>TC</u> ATC <u>AGGAGAA</u> TTAACTATGGG ATCGCACCATCAC-3'
Library 4N	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGAATTGTGAGCGGATAACAATTGAAT TCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGA T <u>NN</u> GGATAATGCCAGCGTAGGGAAGCTATTACAAGAN <u>NN</u> ATC <u>AGGAGAA</u> TTAACTATGG GATCGCACCATCAC-3'
Seq8	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGAATTGTGAGCGGATAACAATTGAAT TCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGA T <u>CTGG</u> GATAATGCCAGCGTAGGGAAGCTATTACAAGAC <u>CC</u> ATC <u>AGGAGAA</u> TTAACTATGGG ATCGCACCATCAC-3'

*RBS sequences are underlined. Nucleotides corresponding to randomized positions are in blue. Modifications of the DNA sequence are indicated with IDT (Integrated DNA Technology) nomenclature, where 5FluorT indicates the addition of a fluorescein molecule to the 5'-end, and 3IABkFQ represents the introduction of an Iowa black quencher molecule to the 3'-end of the oligonucleotide.

2. Strand displacement real-time detection

Strand displacement reactions were assembled as previously reported.⁸⁸ Briefly, a dsDNA reporter consisting of two unequal length oligonucleotides was designed to allow for real-time fluorescence measurements. The smaller oligonucleotide (Rep Q) was tagged with an Iowa black quencher molecule, whereas the longer oligonucleotide (Rep F) contained a 5'-fluorescein molecule (Table 4.1). Rep Q was hybridized to Rep F by an annealing step before final reaction assembly. Initially, a 10 μ M stock containing 2:1 Rep Q:Rep F was prepared. The reporter stock was diluted 10-fold into the annealing reaction (1 μ M reporter final concentration). The annealing was performed in TNaK buffer (140 mM NaCl, 5 mM KCl, 20 mM Tris HCl, pH 7.5) containing 1 μ M oligo (dT)₂₁. The solution was subjected to 5 min at 90 °C and slowly cooled to 25 °C at a rate of 0.1 °C/s. The annealed reporter was then ready for the strand displacement reaction assembly.

The strand displacement reaction was initiated by the addition of 100 nM RNA in a total volume of 20 μ L. Strand displacement was performed in TNaK buffer, 1 μ M oligo (dT)₂₁, 5 mM MgCl₂, 50 nM annealed reporter. TPP, when present, was added to the reaction mixture at 100 μ M, unless otherwise indicated. Reactions were performed in a 384-well black plate (NUNC) and fluorescence was recorded with a TECAN Safire plate reader at 37 °C for at least 2 h. Plates were covered with a thin sealing foil (Lightcycler, Roche). Excitation and emission wavelengths were 485 and 520 nm, respectively.

3. Selection by strand displacement

The DNA library was assembled by PCR of overlapping oligonucleotides. The sequence of the library was of the +ThiM#2 TPP responsive riboswitch⁶⁷ with four randomized positions (Table 4.1) and was purified as described above for template DNA constructs. Transcription reactions used 2 pmol of DNA template and were run and purified as described in *Chemicals, Oligonucleotides, DNA and RNA production*. 20 pmol of library RNA was used for each round of selection. Each round consisted of four steps, including

three negative selections in the absence of TPP and a final positive selection in the presence of TPP (Fig. 4.2). The reporter used for the selection was the same as described above for the real time fluorescence measurements except that the reporter did not contain a fluorophore-quencher pair and that the longer oligonucleotide contained a biotin molecule to allow separation via streptavidin magnetic beads (Life Technologies). For each round, the first negative step was performed by incubating the strand displacement reaction at 37 °C for 2 h. The reaction volume was 100 µL in TNaK buffer, 1 µM oligo (dT)₂₁, 5 mM MgCl₂, 50 nM reporter. 20 µL streptavidin magnetic beads were washed three times with 150 µL BWBT buffer (0.2 M NaCl, 1 mM EDTA, 0.1% (vol/vol) Tween 20, 10 mM Tris-HCl, pH 7.4) and once with the selection buffer (TNaK buffer supplemented with 5 mM MgCl₂). Finally, the beads were resuspended in the strand displacement reaction and incubated at room temperature for 20 min. After magnetic separation, the supernatant was retained and the magnetic beads discarded. 20 µL fresh reporter solution (1 µM) and 5 mM MgCl₂ were then added to the retained supernatant and the negative selection was repeated in the same way except that the incubation of the strand displacement reaction was carried out at 37 °C for 1 h. Isolation of the supernatant from the beads was performed as indicated above. The third negative selection followed the same procedure as the second. Next, 100 µM TPP and 10 µL of the reporter solution were added to the supernatant and incubated for 2 h at 37 °C. 10 µL washed streptavidin magnetic beads were added and incubated at room temperature for 20 min. To enrich for sequences that bound the reporter, the supernatant was discarded after magnetic separation. The beads were washed four times with 150 µL of selection buffer. The beads were then resuspended directly in the reverse transcription PCR reaction (Life Technology) that additionally contained the primers necessary for amplification. 10 cycles of PCR were performed with a Bio-Rad thermocycler. 2 pmol of the resulting DNA were then transcribed, and 20 pmol of this RNA was used for the next round of selection.

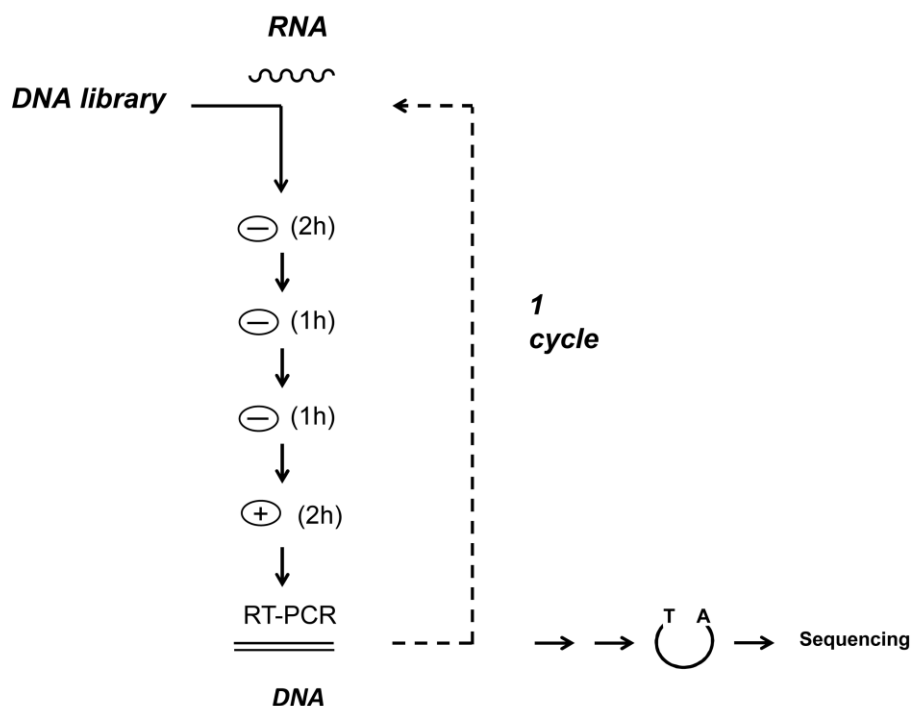


Fig. 4.2 The strand displacement based selection procedure. A DNA library was initially assembled. After transcription, the RNA pool was subjected to consecutive strand displacement reactions in the presence (+) and absence (-) of TPP with incubation times as indicated. After each incubation, the isolation of desired sequences was done via streptavidin bead separation. Finally, reverse transcription and amplification of the selected sequences were carried out. This procedure was repeated for every round of selection. An aliquot of the DNA amplification reaction was then subjected to TA cloning and sequenced.

4. Evaluation of selection cycles

Aliquots from each round of selection were assessed for strand displacement activity by real time fluorescence following the same procedure as described above. Additionally, the DNA pool after each round was cloned and sequenced. 10 ng DNA was amplified with Taq DNA polymerase (NEB) to add a deoxyadenosine to the 3'-end of the fragments. The products were gel extracted and column purified with the Wizard SV Gel and PCR Clean-Up System (Promega). Next, the DNA was ligated into pCR 2.1 according to the TA cloning kit protocol (Life Technologies). Colonies were chosen by blue-white screening with LB supplemented with 156 μ M X-gal and sent for sequencing at University of Texas ICMB Core Facilities- DNA Sequencing. Enriched sequences after three rounds of

selection were further evaluated by their ability to control protein expression. Selected sequences were inserted into the 5'-untranslated region of a gene coding for the yellow fluorescent protein mYPet. Each construct was assembled by overlapping PCR, gel extracted, and purified with the Wizard SV Gel and PCR Clean-Up kit. *In vitro* transcription-translation was performed using the PURE system (NEB). The reactions (20 μ L) contained 250 ng of double strand linear template, 16 U human placenta RNase inhibitor (NEB) and, when present, 1 mM TPP. Fluorescence was recorded over time with a TECAN Infinite 200 plate reader with excitation and emission wavelengths of 510 nm and 540 nm, respectively. Reactions were monitored for 3 h at 37 °C in a 384-well black plate (NUNC) covered by a thin adhesive foil.

4.2 Results and Discussion

4.2.1 Small molecule binding to RNA can induce strand displacement

To construct a ligand responsive strand displacement reaction, a class of RNA molecules that couples ligand binding with conformational changes that effect a biological response was used. Translational control riboswitches modulate the accessibility of a ribosome binding site (RBS) in a manner dependent on ligand binding.⁵⁰ Although the natural *E. coli* ThiM thiamine pyrophosphate (TPP) riboswitch conceals the RBS upon ligand binding, Nomura and Yokobayashi selected for a variant (+ThiM#2) in which the RBS is liberated to activate translation when bound to TPP.⁶⁷ To test whether the changes in accessibility of the RBS induced by TPP binding to the TPP riboswitch could be exploited in order to regulate strand displacement, a reporter with a toehold sequence complementary to the RBS was designed. The reporter additionally contained a 5'-fluorophore (fluorescein) on the longer strand and a 3'-quencher (Iowa black) on the shorter strand so that strand displacement would result in an increase of fluorescence. First, two test constructs with the same RBS sequence as the +ThiM#2 TPP riboswitch were prepared that display accessible

(ON) and inaccessible (OFF) RBS sites. Then the performance of the reporter was assessed with each of these sequences individually. The two constructs reacted differently with the reporter, with the ON construct resulting in 4.20 ± 0.03 -fold greater fluorescence than the OFF-construct (Figure 4.3).

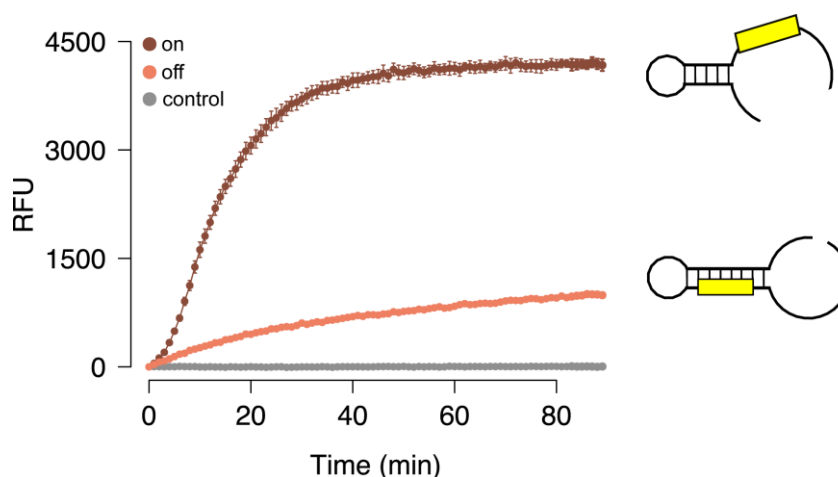


Fig. 4.3 A FRET modified reporter can distinguish between RNA sequences with an accessible (ON) and an inaccessible (OFF) RBS. 100 nM RNA ON or OFF was incubated with 50 nM of a FRET modified reporter (Rep Q:Rep F, Table 4.1) in TNaK buffer. Strand displacement occurred when the RBS sequence of either ON or OFF constructs hybridized with the complementary sequence of the toehold (yellow rectangle) reporter. The control reaction did not include ON or OFF RNA. The sequences of all the nucleic acids can be found in Table 4.1.

Since the reporter was capable of distinguishing between two RNA constructs engineered to fold differently, the following step was to determine if strand displacement could discriminate between the same construct that adopted different conformational states in response to ligand binding. The +ThiM#2 riboswitch selected by Nomura and Yokobayashi⁶⁷ was incubated with the same reporter tested above and increasing concentrations of TPP and monitored by fluorescence over time (Fig. 4.4, panel b). Fluorescence increased with increasing TPP concentration, with 100 μ M TPP resulting in 4.6 ± 0.2 -fold greater fluorescence than in the absence of TPP after 4 h. To confirm that the differences in strand displacement activity were due to TPP induced conformational changes of the riboswitch, two mutant sequences of +ThiM#2 were evaluated. The helix of

the riboswitch that blocks the RBS in the absence of TPP was destabilized in +ThiM#2 mut1, and the residues required for TPP binding were removed in +ThiM#2 mut2 (Table 4.1).⁶⁷ Both mutant constructs were previously characterized *in vivo* through assays that monitored the control of gene expression in response to TPP.⁶⁷ Consistent with these earlier *in vivo* results, *in vitro* strand-displacement activity increased in the presence and absence of TPP for +ThiM#2 mut1, and strand displacement activity was no longer distinguishable between the presence or absence of TPP for +ThiM#2 mut2 (Fig. 4.4, panel d).

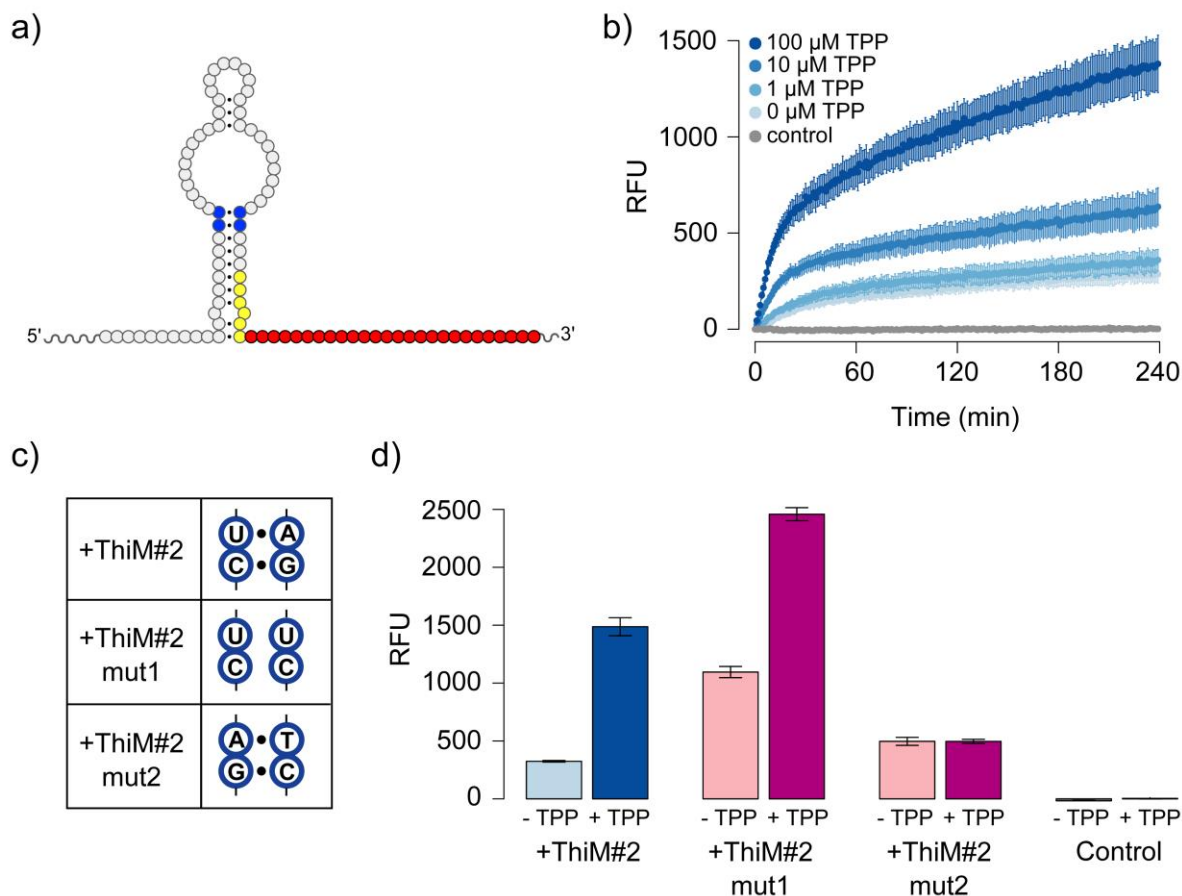


Fig. 4.4 TPP binding to the riboswitch triggers a strand displacement reaction. (a) Schematic view of the +ThiM#2 riboswitch highlighting positions involved in mediating strand displacement. The RBS (yellow) is complementary to the toehold sequence of the reporter. +ThiM#2 mut1 and +ThiM#2 mut2 mutation positions are shown in blue. The sequence that hybridizes with the reporter is shown in red. (b) The +ThiM#2 riboswitch was characterized by strand displacement with the FRET based DNA reporter. 100 nM +ThiM#2 RNA was incubated with 50 nM FRET reporter, 1 μM oligo (dT)₂₁, 5 mM MgCl₂ in TNaK buffer. Reactions were carried out in the presence of increasing concentration of TPP for 4 h at 37 °C. In the control reaction RNA and TPP were not included. (c) Nucleotides that were mutated in the +ThiM#2, +ThiM#2 mut1, and +ThiM#2 mut2 constructs are shown. (d) +ThiM#2 mutant activity in performing strand displacement was compared to the +ThiM#2 riboswitch. The reaction conditions were 100 nM RNA, 50 nM FRET reporter, 1 μM oligo (dT)₂₁, 5 mM MgCl₂ in TNaK buffer either in the presence or absence of 100 μM TPP for 4 h at 37 °C. Fluorescence was recorded with a TECAN Safire plate reader. All the reported results were the average of triplicates.

Most strand displacement reactions reported thus far are solely built with DNA and are not responsive to small molecules. Exceptions include the incorporation of DNA aptamers to control strand displacement in response to ATP and arginine amide ligands.⁸⁵

More recently, several RNA-based strand displacement systems were developed that are triggered by hybridization of specific oligonucleotide sequences.⁹⁶⁻⁹⁸ The data obtained show that both are possible together, that is, small molecules can trigger strand displacement by binding directly to RNA. There are several potential uses of strand displacement reactions. For example, the coupling of small molecule triggered strand-displacement with catalytic hairpin assembly (CHA)⁹⁴ could potentially serve as a sensitive platform for analyte detection. An alternative use of strand displacement would be to use the reporter as bait to pull out ligand responsive sequences from a complex pool of RNA.

4.2.2 Strand displacement is an effective method of selecting for RNA-based sensors

To determine whether strand displacement could be used to identify sequences from a pool of RNA molecules that change conformation in response to ligand binding, a library was constructed based on the +ThiM#2 riboswitch. The library contained four randomized positions chosen to disrupt riboswitch activity. The randomized positions were the same as the mutant positions found in +ThiM#2 mut1 and mut2 constructs.⁶⁷ Active sequences were enriched through binding to the same reporter described above except that the reporter was tagged with a biotin molecule and did not contain a FRET pair. In this arrangement, active sequences bound the reporter in the presence of TPP and were isolated with streptavidin magnetic beads. Conversely, in the absence of TPP, active sequences did not hybridize with the reporter. Iterative selection cycles designed to deplete sequences that bound the reporter in the absence of TPP and enriched sequences that bound in the presence of TPP were used to isolate sequences with TPP responsive strand displacement activity (Fig. 4.2).

Three rounds of selection were performed and the overall activity of the pool was monitored by strand displacement. An aliquot of the RNA output from each round of selection was incubated with the FRET modified reporter and TPP and measured by fluorescence spectroscopy. After 4 h at 37 °C the fluorescence changed by -5.4%, +64%, and +120% for rounds one, two, and three, respectively, with respect to the starting pool (Figure 4.5 and Fig. 4.6). The initial library fluorescence was similar in the presence and absence of TPP, consistent with the starting pool not containing detectable levels of ligand responsive activity. Further, fluorescence in the absence of TPP was lower after each round of selection than the starting RNA pool (Figure 4.5, Fig. 4.6).

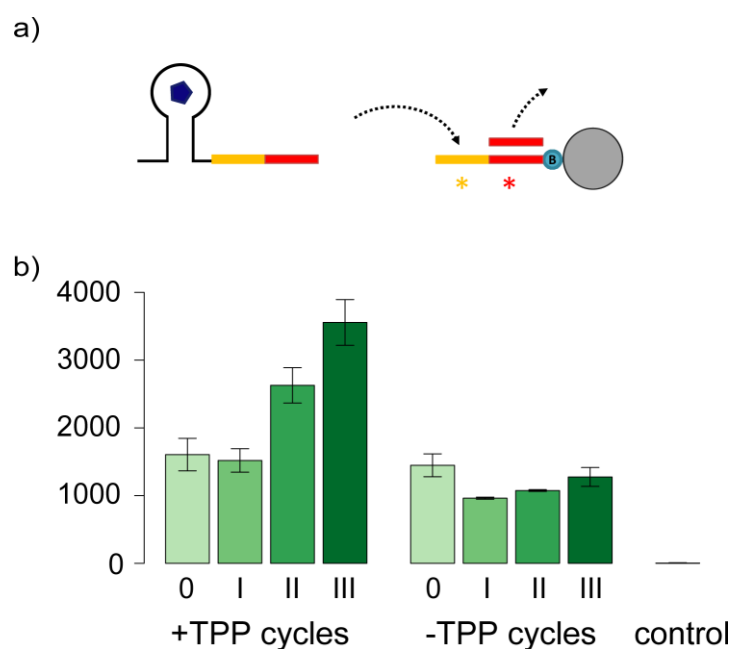


Fig. 4.5 Successive rounds of selection improved TPP responsive strand displacement activity. (a) During the selection, the reporter contained a biotin molecule so that active sequence could be isolated from the pool with streptavidin beads. (b) After each cycle of selection, the activity of the RNA pool was assessed by incubating an aliquot with the reporter modified with a FRET pair. Reactions were run either in the presence or absence of 100 μ M TPP for 4 h at 37 °C. RNA and TPP were not added to the control reaction. The average of three measurements is shown.

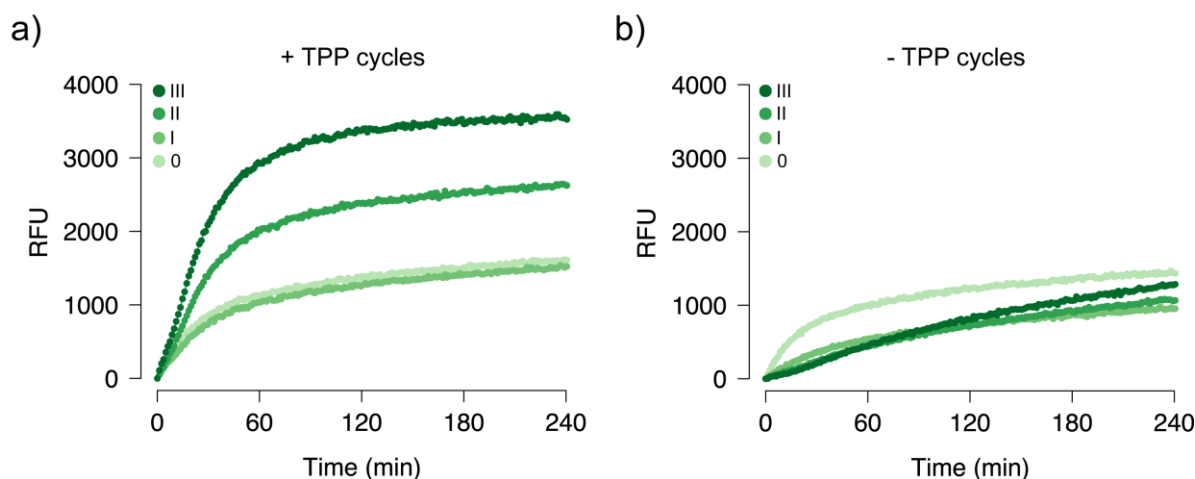


Fig. 4.6 The RNA pool after each round of selection was tested for strand displacement activity with the FRET reporter. Reactions (20 μ L) were performed with 100 nM RNA, 50 nM FRET reporter, 1 μ M oligo (dT)₂₁, 5 mM MgCl₂ in TNaK buffer. +TPP measurements included 100 μ M TPP in the reactions. Fluorescence was monitored for 4 h at 37 $^{\circ}$ C with a TECAN Safire plate reader. Round 0 was the initial RNA pool.

To gain more insight into the progression of the selection, aliquots after each round of selection were sequenced. The starting pool had the highest diversity of sequences (all 10 sequences were different). Conversely, after three rounds of selection, one sequence (Seq8) represented 40% of 20 sequences (Fig. 4.7, Table 4.2). Seq8 contained the same sequence as the +ThiM#2 riboswitch at the first two randomized positions (CU), but a CC in place of the AG found in the +ThiM#2 riboswitch for the last two randomized positions. The full +ThiM#2 sequence did not appear in any of the sequenced samples. Similarly, the non-functional +ThiM#2 mut2 sequence was not observed, consistent with a selection that enriched for TPP binding RNA sequences. However, the +ThiM#2 mut1 sequence (Seq6) was identified in samples taken after each round of selection with a frequency of 0.1 (Table 4.2). In addition to Seq6 and Seq8, Seq1 and Seq7 were enriched with a frequency of 0.15 after three rounds of selection.

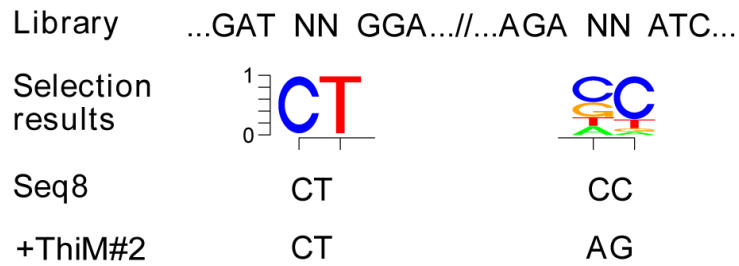


Figure 4.7. Three rounds of selection resulted in an enrichment of sequences different from the +ThiM#2 riboswitch. Twenty clones were sequenced to determine the enrichment at the four randomized positions of the library after the third round of selection. Seq8 was the most highly represented sequence. The +ThiM#2 riboswitch sequence is shown for comparison. Additional sequence information for the other rounds of selection is provided in Fig. 4.8 and Table 4.2.

Table 4.2. Frequencies of DNA molecules after each round of selection.

		Frequency			
		round 0	round 1	round 2	round 3
Seq1	...ACCTGAT CT GGATAA....GA AC ATCAGGAG...	0	0.1	0	0.15
Seq2	...ACCTGAT CT GGATAA....GAG A ATCAGGAG...	0	0.1	0.2	0.05
Seq3	...ACCTGAT CT GGATAA....GAT G ATCAGGAG...	0	0	0.1	0.05
Seq6	...ACCTGAT CT GGATAA....GAT C ATCAGGAG...	0	0.1	0.1	0.1
Seq7	...ACCTGAT CT GGATAA....GAG T ATCAGGAG...	0	0	0	0.15
Seq8	...ACCTGAT CT GGATAA....GAC CC ATCAGGAG...	0	0.1	0	0.4
Seq14	...ACCTGAT CC GGATAA....GAC CC ATCAGGAG...	0	0	0	0.05
Seq15	...ACCTGAT CT GGATAA....GAG C ATCAGGAG...	0	0	0	0.05
+ThiM#2	...ACCTGAT CT GGATAA....GA AG ATCAGGAG...	0	0	0	0
+ThiM#2 mut2	...ACCTGAT G AGGATAA....GAT C ATCAGGAG...	0	0	0	0

Frequencies of DNA molecules after each round of selection. At the end of each round of selection, 10 clones were sequenced. After round three an additional 10 clones were sequenced for a total of 20. Nucleotides corresponding to the randomized positions in the pool are in blue. Round 0 indicates the sequencing results from an aliquot of the initial DNA library. Seq6 is identical to +ThiM#2 mut 1 sequence. +ThiM#2 and +ThiM#2 mut2 frequencies are included for comparison.

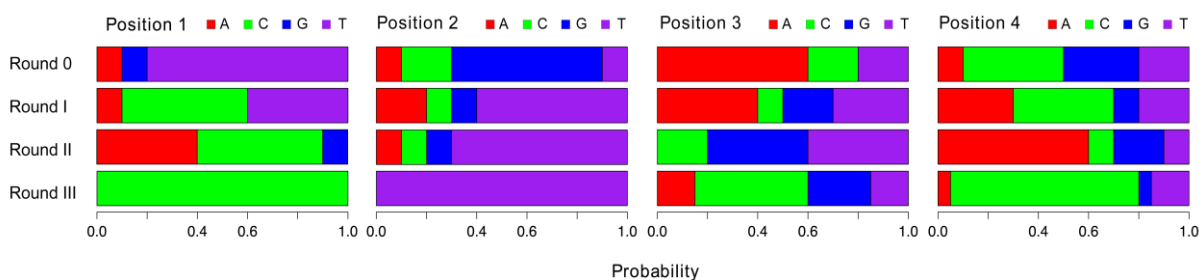


Fig. 4.8 Calculated nucleotide frequencies for every randomized position in the pool after each round of selection. The position number indicates each randomized nucleotide ordered from 5'- to 3'-ends of the library. Round 0 represents the sequencing results obtained from the initial library. A clear evolution towards a specific nucleotide was observed for positions 1 and 2, whereas positions 3 and 4 show less convergence.

The most abundant sequence, Seq8, showed the strongest strand displacement activity of all the sequences tested. The addition of 100 μ M TPP to a solution containing Seq8 and the FRET modified reporter resulted in a 7.0 ± 0.3 -fold increase of fluorescence over the same reaction in the absence of TPP (Fig. 4.9). Seq8 was more active in strand displacement than the original +ThiM#2 riboswitch from which the library was based, which showed a 4.6 ± 0.2 -fold increase in fluorescence upon the addition of 100 μ M TPP. The strand displacement activity of Seq1, Seq6 (+ThiM#2 mut1), and Seq7 were similar with changes in fluorescence of 2.5 ± 0.1 -fold, 2.4 ± 0.2 -fold, and 1.8 ± 0.1 -fold, respectively, upon the addition of TPP (Fig. 4.10). The fact that the selection gave rise to a sequence with better strand displacement activity than +ThiM#2 was consonant with the applied selective pressures. Seq8 was enriched with an *in vitro* method based on strand displacement activity. Conversely, the +ThiM#2 riboswitch was identified through an *in vivo* selection based on the control of gene expression and not strand displacement.

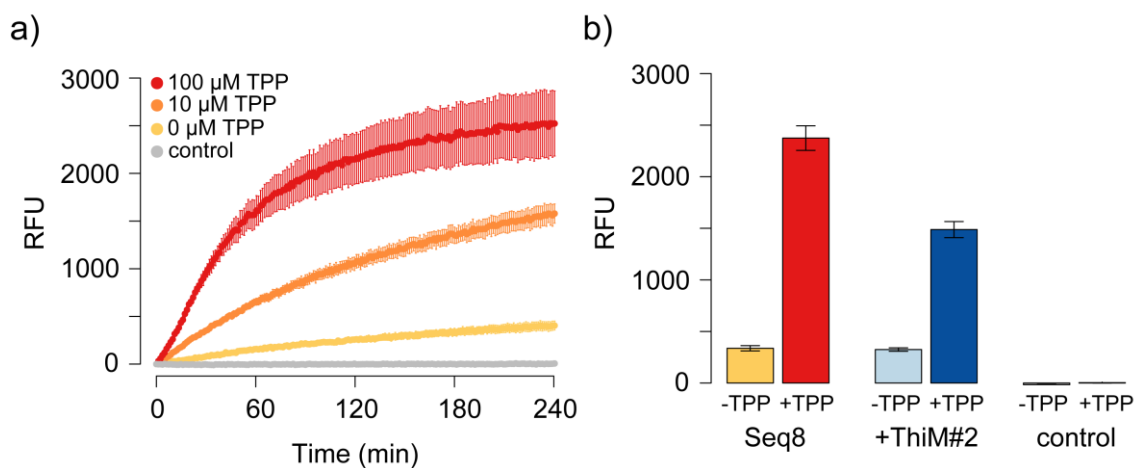


Fig. 4.9. Seq8 is optimized for strand displacement. (a) Seq8 showed TPP dependent strand displacement activity. The reactions contained 100 nM RNA Seq8, 50 nM FRET reporter, 1 μ M oligo (dT)₂₁, 5 mM MgCl₂, TNaK buffer, and varying concentrations of TPP (0, 10 μ M and 100 μ M). In the control reaction the RNA was omitted. Fluorescence was recorded with TECAN Safire plate reader for 4 h at 37 °C. (b) Seq8 is more effective at strand displacement than +ThiM#2. 100 nM Seq 8 RNA and +ThiM#2 RNA were added respectively to reactions (20 μ L) containing 50 nM FRET reporter, 1 μ M oligo (dT)₂₁, 5 mM MgCl₂ in TNaK buffer. Data were collected in the presence and absence of 100 μ M TPP for 4 h at 37 °C with a TECAN Safire plate reader. Control reactions lacked RNA. The average of three measurements is displayed.

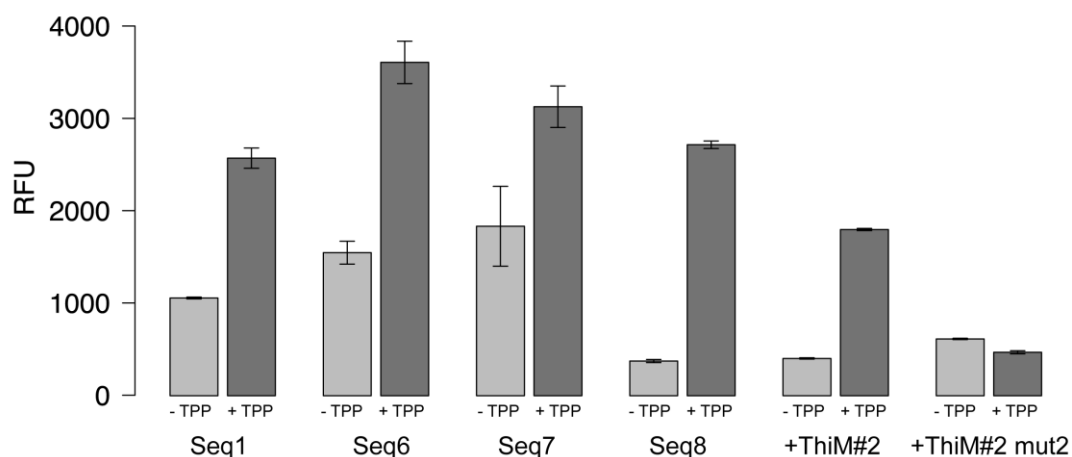


Fig. 4.10 Strand displacement activity of sequences after three round of selections. 20 μ L reactions containing 100 nM RNA with 50 nM DNA FRET reporter, 1 μ M oligo (dT)₂₁, 5 mM MgCl₂, in TNaK buffer. Fluorescence was recorded with a TECAN Safire plate reader for 4 h at 37 °C. 100 μ M TPP, when present, was added to the reaction. Seq6 was the same sequence as +ThiM#2 mut1. +ThiM#2 and +ThiM#2 mut2 sequences were also tested for comparison.

Since +ThiM#2 has both strand displacement and riboswitch activity, the sequences identified through a selection based on strand displacement were probed whether they could also function as riboswitches. The same sequences tested for strand displacement activity were placed in the 5'-untranslated region of a gene encoding yellow fluorescent protein (YFP) and gene expression was evaluated. All of the sequences enriched after three rounds of selection showed between 2- and 3-fold increased YFP synthesis in the presence of TPP (Fig. 4.11). For comparison, YFP synthesis increased 7.3 ± 0.3 -fold in the presence of TPP when under the control of the +ThiM#2 riboswitch. Conversely, the previously characterized inactive riboswitch +ThiM#2 mut2 showed no difference in protein expression in the presence or absence of TPP. Therefore, the strand displacement based selection strategy did give rise to riboswitch activity, albeit greatly diminished in comparison to previously selected, artificial riboswitches.^{55,56,100} However, naturally occurring riboswitches do not typically regulate large differences in protein synthesis and instead mediate more subtle effects in response to ligand availability. In fact, of the few natural riboswitches tested *in vitro* for their ability to control protein output,^{37,101} the effect of ligand binding was similar to that observed for the sequences isolated from the selection based on strand displacement.

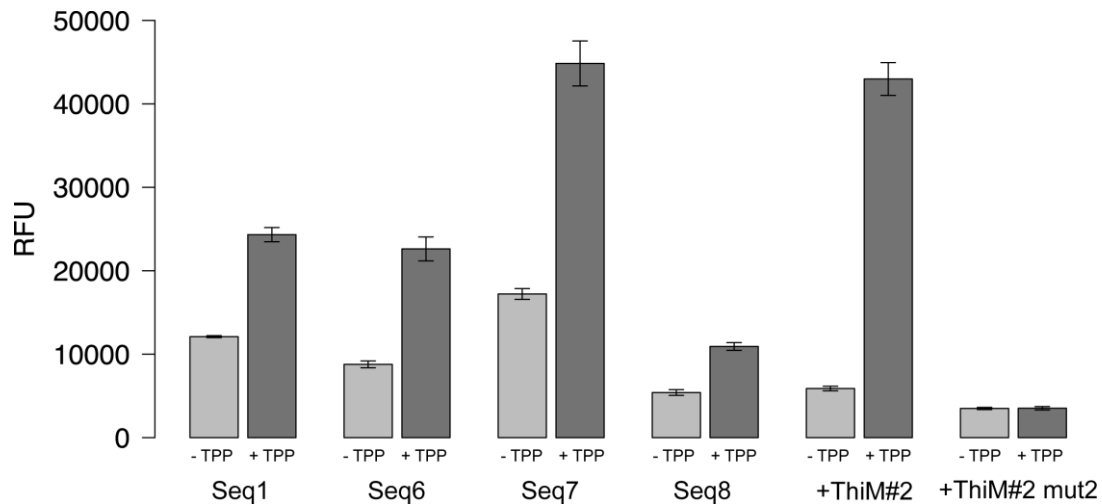


Fig. 4.11 Sequences identified after the third round of selection are assessed for *in vitro* translation control activity. Each sequence was inserted into the 5' untranslated region of a gene coding for yellow fluorescent protein (YPet). *In vitro* transcription-translation exploited the PURE system (NEB). The total reaction volume was 20 μ L and included 250 ng of double strand linear DNA template, 16 U human placenta RNase inhibitor and, when present, 1 mM TPP. Fluorescence was recorded over time at 37 $^{\circ}$ C in a TECAN Infinite 200 plate reader for 3 h with a 384-well plate covered by a thin adhesive foil. Seq6 is identical to +ThiM#2 mut1 sequence.

These results show that it is possible to use a strand displacement-based selection strategy to enrich sequences that undergo specific conformational changes in response to ligand binding. Based on the described methodology, it should be possible to select for strand displacement reactions controlled by small molecules other than TPP. Ligand responsive strand displacement opens new opportunities for the implementation of RNA circuits and nucleic acid-based nanotechnology, broadening the field beyond nucleic acid triggered systems. Further, since sequences that are active in mediating strand displacement are also capable of controlling gene expression, this methodology or a derivative may be useful in identifying artificial riboswitches. More examples are needed to determine if strand displacement-based selections can reliably function as a fully *in vitro* riboswitch selection strategy. If so, a wider variety of synthetic riboswitches would become available to construct engineered living^{49,61} and artificial cellular systems.^{33,100}

4.3 *In vitro* selection of TPP analogs- responsive RNA sensors

The *in vitro* selection technique based on strand displacement was successfully applied to functional sequence discovery in a small pool of 256 variants. However, the selection was carried out for a thiamine responsive RNA sensor, and the initial library sequence was based on the TPP +ThiM#2 responsive riboswitch. Although the initial pool diversity in functionality was assessed via the FRET coupled reporter, selection for a TPP analogue would be the first real application of the proposed method.

For this reason, a new library was designed starting from the TPP riboswitch previously exploited. Since new binding properties were sought, a rational design approach was applied for the identification of randomized nucleotide positions. The TPP riboswitch structure was considered in detail.¹⁰² Initially, the structure of the natural *thiM* box of *E. coli* was considered. This natural riboswitch differed from the +ThiM#2 TPP riboswitch previously exploited for strand displacement characterization⁶⁷ for the actuator part, i.e. the RBS and proximal elements. However, the binding pocket of the TPP riboswitch presented the same sequence. The TPP molecule inserts in the RNA structure by formation of non-covalent bonds and hydrogen interactions with its 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) moiety, which intercalates with pocket residues. On the contrary, the central thiazol is not recognized by the RNA. The pyrophosphate molecule is stabilized by interactions with a second RNA pocket via two Mg²⁺ ions. Thus, the interaction is not direct on the nucleic acids sequence. The overall TPP orientation is perpendicular to the RNA helices of the binding pocket. Starting from these crystallography investigations,¹⁰² the bases that were mostly involved in the binding of TPP were identified. A library containing 14 randomized nucleotides was then designed. Both the HMP and magnesium ions binding pockets were included in the randomized part.

The first ligand exploited was the TPP analog oxythiamine. This choice was due to the high similarity of oxythiamine presented towards the HMP moiety and thiazol group of

TPP (Fig. 4.12). Then, due to the negative results obtained in the oxythiamine selection and to a new structure analysis, another two ligands were tested, including an ATP analog and an ADP analog. The property that characterized these two molecules, rather than oxythiamine, was the presence of phosphate groups, which made the analogs more similar to the TPP molecule in spatial length. The structure investigation is presented in detail in the results section.

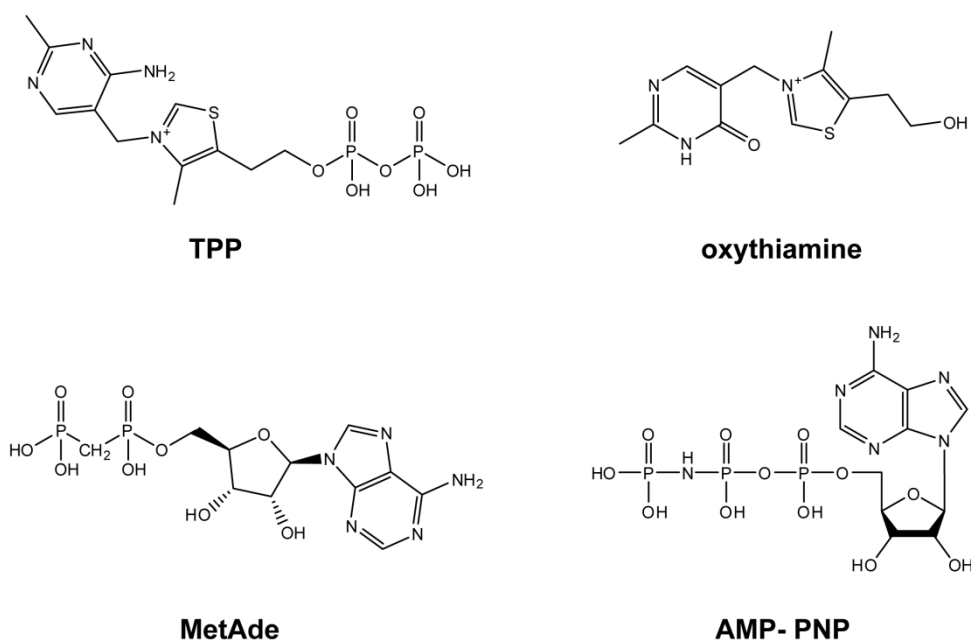


Fig. 4.12 Chemical structures of the ligands exploited for selections. TPP, oxythiamine, the ADP analog Adenosine 5'-(α,β -methylene)diphosphate (MetAde) and the ATP analog Adenosine 5'-(β,γ -imido)triphosphate (AMP- PNP) are presented.

4.3.1 Material and method

The selection was carried out accordingly to the proposed method for *in vitro* selection of RNA sensors based on strand displacement triggering. Materials and methods were described previously (see section 4.1.1). Oxythiamine, Adenosine 5'-(α,β -methylene)diphosphate (ADP analog, MetAde thereafter) and Adenosine 5'-(β,γ -imido)triphosphate lithium salt hydrate (ATP analog, AMP- PNP thereafter) were purchased from Sigma Aldrich.

Each TPP analog was exploited as ligand in positive selection steps. The library complexity was 10^8 different molecules, considering the fourteen nucleotides that were randomized (4^{14}). The library design was based on the +ThiM#2 TPP riboswitch sequence (Table 4.3). Again, amplification reaction of overlapping oligonucleotides was performed for library assembly. Transcription of the pool was carried out in 100 μ L of T7 transcription buffer with 10 pmol of PCR product for 1 h at 37 °C. RNA was PAGE purified and 20 pmol were exploited for the strand displacement selection. The selection procedure was repeated as previously reported.

Table 4.3 14 N library DNA sequence for evolution of new RNA sensors

Constructs*	
+ThiM#2 riboswitch	5'- ATAAATTAATACGACTCACTATAGGGAGAGGAGGGAATTGTGAGCGGATAACAATTGAAT TCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGA TCTGGATAATGCCAGCGTAGGGAAGCTATTACAAGAAGATC <u>CAGGAGAAATTA</u> ACTATGGG ATCGCACCATCAC-3'
Library 14 N	5'- ATAAATTAATACGACTCACTATAGGGAGAAGAGGGAATTGTGAGCGGATAACAATTGAATT CAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGC NNNNNN NATACCCGTATCAC NNNA TCTGGATAAT NCCAGNNTN NGGAAGCTATTACAAGAAGATC <u>CAGGAGAAATTA</u> ACTATGGG ATCGCACCATCAC-3'

*RBS sequence is underlined.

1. Oxythiamine selection

Eleven cycles of selection were carried out. For rounds I and II, three negative selection of 2 h, 1 h and 1 h, respectively, were performed. Then a positive selection cycle, i.e. in presence of 100 μ M oxythiamine, was carried out for 2 h. From rounds II to round VI, the incubation time of the strand displacement reaction in the presence of oxythiamine was reduced to only 1 h. This change increased the stringency of selection conditions. In round VII, the strand displacement in presence of the ligand was carried out for 30 min. However, since very small recovery was obtained, round VIII was again incubated for 1 h during the positive selection step. Finally, from round IX to XI, only the positive selection was

performed with a 1 h incubation, since molecules showing high activity in presence of the ligand were sought. At the end of each round, 10 amplification cycles were performed, excluding round VII where 15 cycles of PCR were exploited. From the second round on, the transcription reaction was scaled down to 40 μ L and almost 3.3 pmol of amplified DNA were added in the reaction.

2. *MetAde and AMP- PNP selection*

These selections were performed starting from the 14N library as described for the oxythiamine selection. The concentrations exploited were 10 μ M and 100 μ M for MetAde and AMP- PNP respectively. Six rounds of selection were carried out for the MetAde selection, while three rounds were performed with AMP- PNP ligand. 20 pmol of RNA were maintained fixed for addition in the strand displacement reaction. The selection steps included three negative and a positive one. Incubation times were different between each cycle (Table 4.4) and was modulated according to the result of strand displacement activity assessed in the previous round.

Table 4.4 Incubation times in presence and absence of ligand during selection procedures

Selection incubation	MetAde selection rounds						AMP- PNP selection rounds		
	I	II	III	IV	V	VI	I	II	III
no ligand	1 h	30 min	30 min	1 h	1 h	1 h	1 h	1 h	1 h
no ligand	30 min	15 min	15 min	30 min	30 min	30 min	30 min	30 min	30 min
no ligand	30 min	15 min	15 min	30 min	30 min	30 min	30 min	30 min	30 min
plus ligand	30 min	15 min	7.5 min	7.5 min	7.5 min	7.5 min	30 min	15 min	15 min

Selection steps were performed in presence and absence of the ligand for each round of selection. Incubation of the strand displacement reaction at 37 °C was carried out for the indicated time. Time modulation reflected in variation in stringency conditions of the selection.

3. *Evaluation of the rounds of selection*

In the three selections, all the cycles were monitored with a FRET reporter for strand displacement activity in the presence and absence of the ligand. Measurements were performed with a TECAN plate reader and with a qPCR machine (Rotor Gene, Quiagen). Differences between the readings of the two instruments were due firstly to the filter-based detection of the qPCR machine, while the plate reader was based on a monochromator module, secondly to the spinning of the samples while the measurement was taken at the Rotor Gene. However, data were comparable because the values were a ratio of activity in the presence and absence of the specific ligand.

Moreover, aliquots of samples obtained at the end of each round were sent for sequencing. Generally, not all the cycles performed were sequenced. Identification of sequencing samples was carried out by preliminary evaluation of strand displacement activity on the FRET reporter.

4.3.2 Results and discussion

1. In vitro selection of an oxythiamine responsive RNA sensor

The *in vitro* selection for developing an oxythiamine responsive RNA sensor was the first selection carried out. Aliquots of the pool were measured with a FRET reporter while the selection was ongoing. These measurements allowed to change conditions, e.g. stringency, dependent upon the observed results. In particular, since strand displacement was triggered by the conformational shift induced by the ligand binding, the FRET reporter-based experiments were exploited to verify if the binding of oxythiamine to RNA sequences had occurred. After the first three cycles, the sequences were still not able to switch in the presence and absence of the ligand (Fig. 4.13). However, the fluorescence signal after the third cycle was similar to the original library fluorescence, indicating that no progression in

the selection was achieved. Cycle IV displayed initially a moderate difference between activity in the presence and absence of oxythiamine. Surprisingly, the switching was not maintained also after round V. This result was opposite to what observed with the first 4 N selection for the TPP responsive RNA sensor development. In that selection, the activity increased proportionally upon rounds iteration in presence of the ligand. Here, the fluorescence signal was lost in cycle V. Nevertheless, considering that the complexity of the pool exploited was higher the selection was continued. Fortunately, round VI presented again a small amount of switching activity.

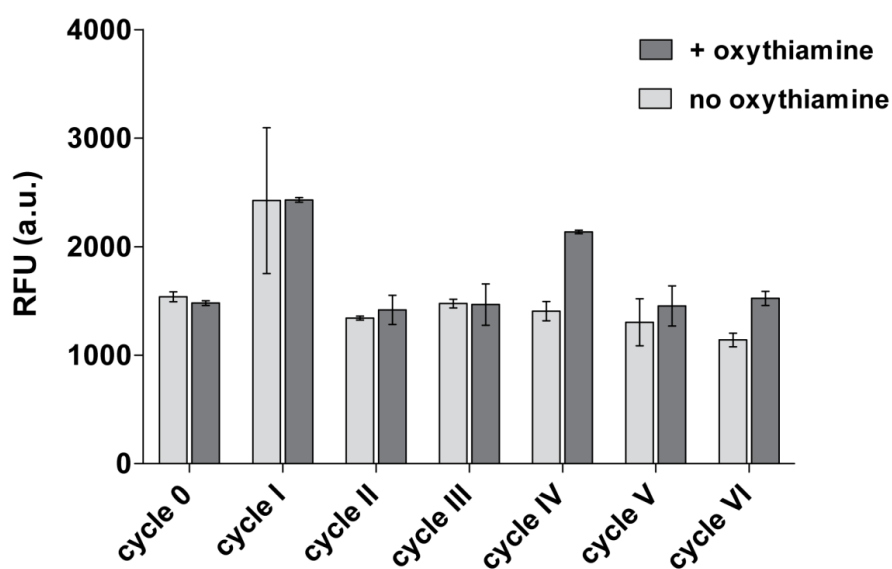


Fig.4.13 Oxythiamine selection progression was assessed with the FRET pair reporter for strand displacement activity. The RNA pool behaviour in presence and absence of the ligand after six rounds of selection is shown. Data are averages of duplicates, measurements were performed at TECAN plate reader. Results evidenced that no different activity was recorded in presence or absence of 100 μ M oxythiamine after 4 h reaction at 37 $^{\circ}$ C. Small switching activity was observed after cycle IV. However, the signal was almost lost at the end of the fifth cycle.

Both rounds IV and VI were sent for sequencing. 40 clones were sequenced after round IV, while only 10 clones after the sixth round. Results demonstrated that any identical sequence was obtained after the fourth cycle. Conversely, after the fourth round, two similar sequences (Seq4 and Seq7) were identified. The similarity was due to 9 nucleotides out of 14 randomized positions in the original pool. Considering that the number of sequencing

samples was low, further investigation and activity assessment in strand displacement of these sequences was carried out. FRET based strand displacement reactions was performed with sequences 4 and 7 in the presence and absence of 100 μ M oxythiamine (Fig. 4.14). The results showed a difference in the behavior of the two sequences. In particular, Seq4 fluorescent signal was comparable with the initial library activity. Conversely, Seq7 had a very low activity. This distinction was due only to five base pair diversity in the two sequences. However, both did not demonstrate to be able to modulate strand displacement upon ligand presence.

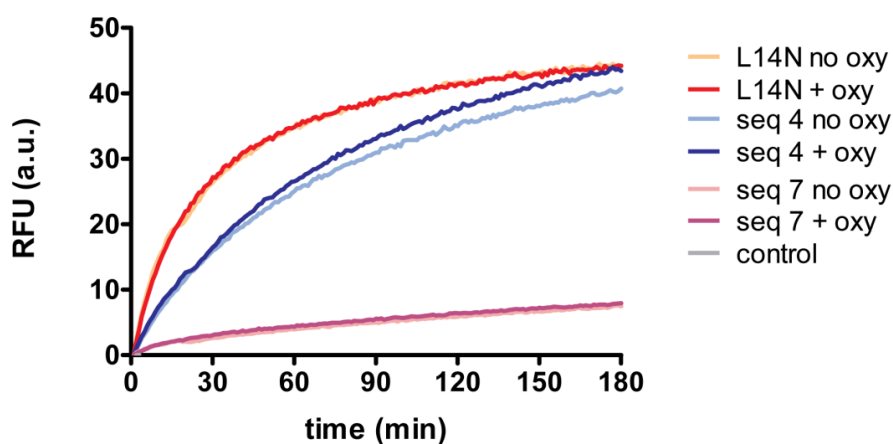


Fig.4.14 Strand displacement characterization of sequences 4 and 7. These clones were identified by sequencing for similarity in the sequence. Strand displacement was performed with the FRET reporter monitoring for 3 h at 37°C in presence or absence of 100 μ M oxythiamine. Data were collected with the Rotor gene instrument. Seq4 and Seq7 presented a different behaviour, due to only 5 nucleotides variation between the two sequences.

Since any of the tested sequences emerged to be active from strand displacement analysis and sequencing results showed how the pool was not converging towards a specific sequence pattern, further cycles of selection were carried out. Initially, more stringent conditions were applied in round VII. The incubation time in the positive step, where molecules were selected for strand displacement activity in presence of oxythiamine, was decreased to 30 min. However, the amplification step at the end of the cycle failed to show DNA bands. Thus, the cycle was repeated with 1 h incubation instead than 30 min.

After cycle VIII, the overall behaviour was again tested via strand displacement. Still, the activity of the pool was almost absent and strand displacement was not occurring in presence of oxythiamine (Fig. 4.15). On the other hand, the signal in absence of the ligand was an optimal result for the development of an RNA sensor with low background. In order to drive the RNA pool towards enrichment of molecules with low background but which activates in presence of oxythiamine, the strand displacement selections in absence of ligand were removed. The last three cycles (IX, X, XI) were carried out with a single selection reaction in which oxythiamine was added. Incubation was performed for 1 h at 37 °C. While in cycle IX the aliquot of RNA pool was unvaried in ability of performing strand displacement, in rounds X and XI enrichment was observed for molecules with improved activity in strand displacement (Fig. 4.15, panel b). However, the increase was shown for both presence and absence of oxythiamine. Thus a pool response for selection of more accessible sequences, i.e. with RBS in open configuration, was achieved, but no switching in conformation due to ligand binding occurred.

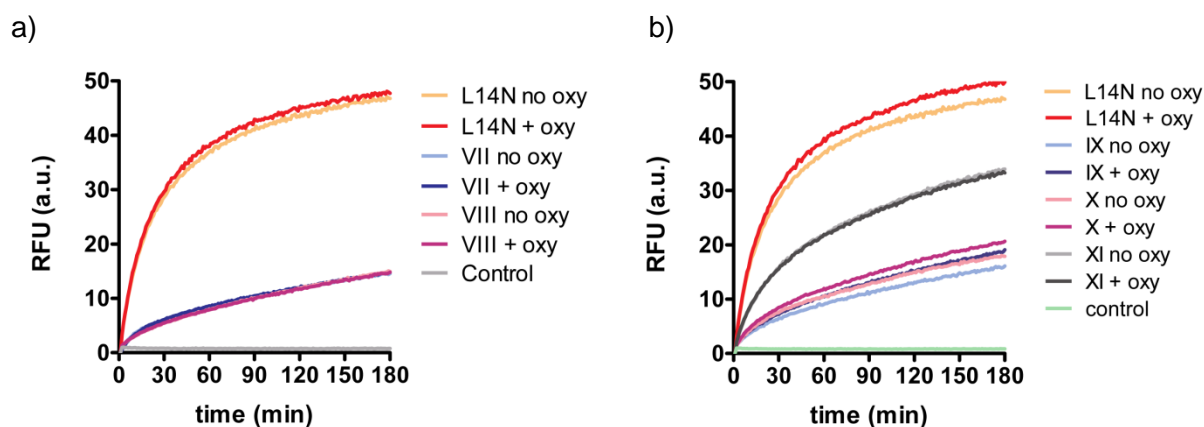


Fig. 4.15 Strand displacement evaluation of oxythiamine selection cycles in presence and absence of ligand. Strand displacement was performed with a FRET reporter, monitoring at Rotor Gene for 3 h at 37 °C. When present, 100 μ M oxythiamine was exploited. Data are average of measurements. Panel a) Aliquots of the RNA pool were tested after cycles VII and VIII. Activity was compared to the library (L14N), i.e. cycle 0 of the selection. Panel b) Cycles IX, X, XI are shown. These rounds were performed without negative selection steps. The activity increased, although no difference between presence and absence of the ligand was observed.

At the end of the last round, 30 samples were sent for sequencing. Data were analyzed by clustering based on the similarity of the fourteen residues randomized. The similarity parameter was considered instead of the identity of sequences in order to relate clones which presented small variations in nucleotides in each position. For this reason, a higher score was assigned for identical nucleotides (score of 1), while for changes of, for example, a pyrimidine base with another pyrimidine a lower score (0.5) was set. The resulting dendrogram indicated the height of difference between the sequences (Fig. 4.16). The more two sequences were similar, the lower would be in the graph. Moreover, high similarity would also be indicated by how close the sequences would be compared to other ones. In other words, this algorithm allowed to cluster the sequences in families and to evaluate simultaneously the proximity of each sequence towards another belonging to the same family. The identification of families depends on the height grade that is permitted, i.e. how much two sequences are considered to be related.

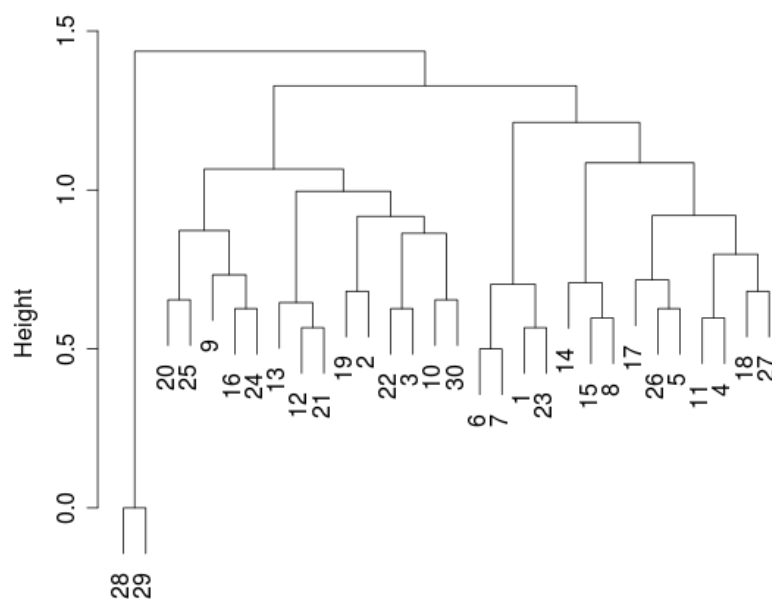


Fig. 4.16 Dendrogram of similarity of sequences resulting from sequencing analysis after the XI cycle of the oxythiamine selection. The height indicated how different two sequences were in the 14 nucleotide positions initially randomized. At lower height, the similarity increased. For example, sequences 28 and 29 were identical. A score of 1 was set for identical nucleotides, purine or pyrimidine substitution was assigned of a score value of 0.5. Proximity of sample indicated that the sequences were similar one to the other. The analysis on sequencing results was done by Dr. Michele Forlin.

Out of 30 sequences sent for sequencing, two were identical (Seq28 and 29). Then, other seven clones were tested for strand displacement capability. Sequences were chosen in order to cover many of the families identified by setting the height around 0.5. However, the strand displacement reactions with the FRET reporter showed different signals for activity, but none of the analyzed sequence was able to modulate strand displacement accordingly to presence and absence of the ligand. Neither Seq28 displayed activity (Fig. 4.17).

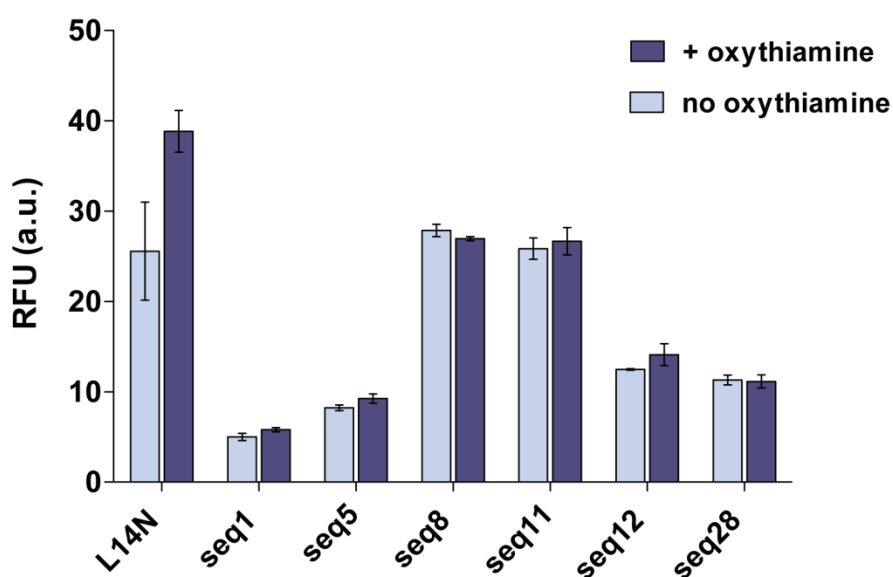


Fig. 4.17 Strand displacement activity of some sequences identified in after round XI. The FRET reporter was exploited, reactions were incubated at 37 °C for 3 h in presence and absence of 100 μM oxythiamine. Data (duplicates) were recorded at the Rotor Gene. Results were compared to the initial library activity (L14N). No difference in presence and absence of the ligand was observable.

From the characterization of sequences obtained after round XI and from the results of the RNA pool in strand displacement reaction, the selection for oxythiamine was suspended. A possible reason for its failure was given by a second analysis of the TPP riboswitch structure. From crystallography studies of the TPP RNA sensor with TPP analogs,¹⁰³ it was possible to infer that the oxythiamine structure was too small for effectively bind in the riboswitch pocket. Although the riboswitch structure was partially

randomized, probably the TPP binding structure was not enough disrupted. Thus, oxythiamine would enter into the first pocket for the HMP moiety, but the molecule would be too short for binding on the other side of the RNA pocket. This hypothesis had been demonstrated by electron density calculation of the TPP riboswitch with analogs as thiamine monophosphate and pyrithiamine, which present a single or no phosphate group respectively.¹⁰³ When pyrithiamine was exploited as ligand in solving the crystallography complex of the *E. coli thiM* box, the lack of phosphate group determined a large instability on the second binding pocket of the TPP sensor.

These structural considerations gave a possible explanation for why oxythiamine selection did not produce a positive sequence. Based on these hypotheses, new selections were planned exploiting molecules which were longer than oxythiamine and which presented phosphate groups, in order to be more suitable for the binding to the RNA.

2. In vitro selections of a MetAde- and a AMP- PNP- responsive RNA sensors

MetAde and AMP-PNP selections were performed on the 14 N library previously exploited. Six rounds of selection were carried out for MetAde RNA sensor selection and three rounds for the AMP-PNP sensor development. For the MetAde selection, stringent conditions in the strand displacement reaction in presence of the ligand were applied from the second round on. Incubation time was reduced to 7.5 min, in order to increase chances of evolution of fast switching RNA sensor. Based on the oxythiamine results of evolution of molecules with very low activity in the absence of the ligand, the increased stringency would allow to enrich from the beginning of the selection only for sequences with high binding properties. This strategy would avoid carry-over of molecules that interacts in strand displacement independently of the conformational switch.

After six cycles, MetAde selection results did not show any conformational change triggered strand displacement activity in presence of the ligand (Fig. 4.18). Sequencing of

16 clones after round VI showed two identical sequences; however, again no difference in the presence and absence of MetAde was observed while performing strand displacement.

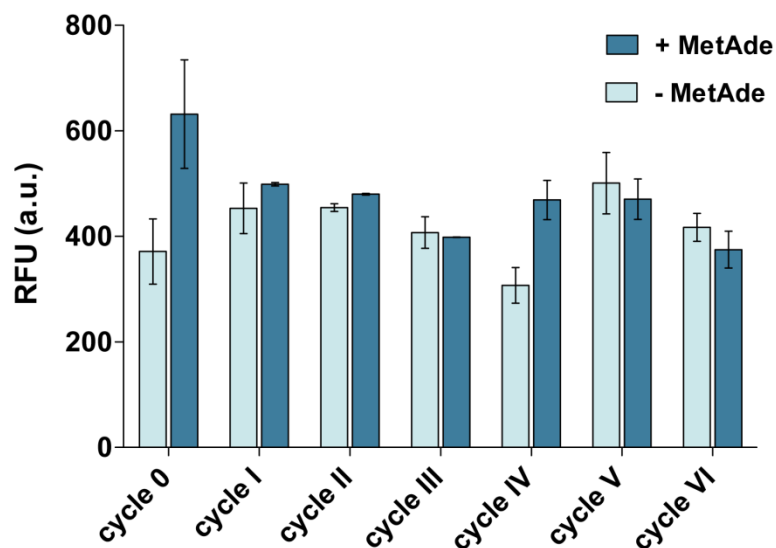


Fig. 4.18 MetAde selection progression was monitored by strand displacement reaction with a FRET reporter. Average of two measurement is displayed. The data were collected with a TECAN plate reader, incubating the reaction at 37 °C for 3 h. No switching ratio in presence and absence of 10 μ M MetAde was observed. All the cycles presented less activity than the starting pool at cycle 0 (L14N).

The AMP-PNP selection gave the same result of MetAde selection. Cycles were performed almost simultaneously with the last cycles of MetAde selection, thus after 3 rounds the selection for the ATP analog-responsive RNA sensor was interrupted. The overall pool behaviour, characterized by strand displacement activity, showed a signal lower than the library fluorescence and almost no difference in the presence and absence of AMP-PNP (Fig. 4.19). Sequencing of 16 samples at the end of the third round did not show any identical sequence, thus no clones were further evaluated in strand displacement.

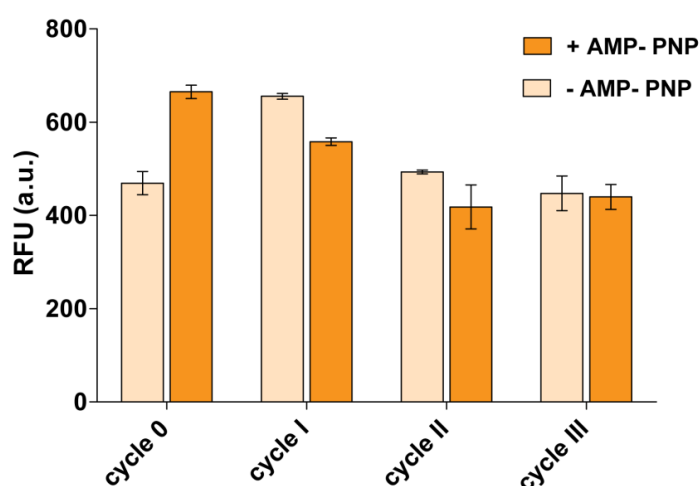


Fig. 4.19 Strand displacement characterization of AMP- PNP selection. Data (duplicates) were recorded at TECAN plate reader incubating 3 h at 37 °C. The selection was interrupted after three cycles since no switching activity in presence and absence of 100 μ M AMP- PNP was shown.

These results showed how both the selections for long molecules carrying phosphate groups failed to develop a specific RNA sensor. In the oxythiamine selection, a possible cause for not achieving ligand induced triggering of the strand displacement reaction was the length of the molecule. However, for MetAde and AMP-PNP the same hypothesis could not be considered. Both molecules were long, with moiety similar to the HMP ring that TPP presented and the phosphate groups could interact with the other helix of the RNA binding pocket. Nevertheless, the three selections were performed starting from the same 14N library. Thus another possible reason for no observable ligand induced strand displacement was the initial pool lacking of activity. 10 clones of the library were sent sequencing and results confirmed that the pool was constituted by random sequences. However, no binding activity was achieved for any molecule selection. In the TPP 4N selection after three rounds the binding activity, measured as a strand displacement fluorescence output, was increased by +125%. Sequencing at the end of the selection showed how the nucleotides involved in TPP binding were almost 100% present in the clones analyzed. Exploiting a 14 N library, these results were never achieved. For this

reason, it would be possible to start again the selection from different libraries. In particular, the design could reduce the number of randomized position in order to focus on the evolution of just part of the binding pockets. Considering the ligand structure, only a small portion of the RNA sequence would then be randomized. This approach would still be based on an *in vitro* selection for development of the RNA sensor, but a rational method would be exploited first in order to increase molecule binding.

Chapter 5.

Conclusions

This dissertation presented the building of cellular mimics that sense and respond to external stimuli, focusing on the application and the *in vitro* selection of the sensor element. The theophylline and the adenine- responsive riboswitches were initially exploited for the assembly. The synthetic riboswitch responsive to the small molecule theophylline was demonstrated to function as a translation controller in a cell-free environment. Transcription and translation machinery made of purified components was employed, i.e. PURE system. Encapsulation in emulsion and vesicle compartments of the genetic DNA element encoding the riboswitch and the PURE system for protein production allowed to build a cellular mimic that was able to sense theophylline from the external environment. Through riboswitch processing, theophylline presence in the environment activated the production of a yellow fluorescence protein in the compartment.

Furthermore, a cellular mimic able to translate a chemical message for bacteria was built. Again, the sensor element exploited was the theophylline responsive riboswitch, whose activation controlled the expression of a pore forming protein. In the presence of theophylline, the cellular mimic was able to release IPTG, which was initially included in the compartment. The release causes activation in bacteria, and a response could be detected. Thus a cellular mimic acted as translator of a signal (theophylline), that bacteria could not directly recognize, into a chemical message (IPTG) that *E. coli* could sense.

Although the riboswitch was the mediator of efficacy of the cellular mimic, the riboregulator represented also the limiting feature for possible applications. In other words, the exploited riboswitch was able only to respond to theophylline. In order to expand the system sensing possibility and to engineer cellular mimics able to sense and respond to any molecule of choosing, selection of new RNA sensors was considered. mRNA display *in vitro* selection method was initially tested for development of a riboswitch responsive to malachite green. Unfortunately, after eight cycles of selection the results did not show that the initial library converged to any sequence (or family). Thus a new approach was developed. The selection strategy was defined as a small ligand induced conformational

change in the RNA structure of the sensor to trigger a toehold mediated strand displacement on a specifically engineered reporter. Through the screening of a small library of sequences, the strand displacement selection allowed to identify new RNA sensor variants responsive to TPP. These sequences could perform strand displacement when TPP was present in the solution, consistent with the selection pressures applied. Moreover, frequent sequences identified by sequencing analysis at the end of the selection also demonstrated that the original TPP riboswitch was able to control translation. Although protein expression modulation was lower compared to artificial riboswitches, the activity was comparable to the regulation typically observed with naturally occurring riboregulators. Thus the strand displacement-based *in vitro* selection method allowed to identify sequences with strand displacement ability and riboswitch activity. These RNA molecules could be applied to protein expression regulation as well as to nucleic acid *in vitro* circuit design.

However, investigation of *in vitro* selections described in this dissertation highlighted how the development of new riboswitches was a difficult task. Library design was crucial for the final output. For example, strand displacement selections for TPP analogs failed probably for the lack of functionality in the initial library pool. For this reason a combination of computational approaches and successive *in vitro* screenings would increase the chances of a successful selection. Nevertheless, the application of newly selected sensors, as the riboswitches originated from strand displacement selection, would greatly expand the cellular mimic spectra of sensing and responding towards different targets.

5.1 Future perspective

The future step in this work would be the engineering of already selected RNA sensors to detect analog molecules via strand displacement *in vitro* selection. Up to now, many RNA sensors have been well established and characterized in the aptamer region. Structural calculations and data are available to elucidate the ligand induced conformational

change of the aptamer. Thus, specific sequence elements and nucleotides involved in ligand binding can be targeted for *in vitro* strand displacement-based selections. Furthermore, strand displacement-based selection can be applied not only to RNA sensor, but also to the high number of already selected DNA aptamers, for converting these aptamer to be responsive to different targets.

Finally, the development of new RNA sensors that possess strand displacement capability would allow for the construction of regulatory circuits dependent on the presence of a small molecule. Cascade events in a cellular mimic could simply be controlled by the addition in the external environment of the ligand and thus activate regulatory circuits in the compartment. Controllers subjected to strand displacement could be engineered for any sensor developed in a high, sequence-specific way.

Bibliography

- (1) Cameron, D. E.; Bashor, C. J.; Collins, J. J. *Nat. Rev. Microbiol.* **2014**, *12*, 381.
- (2) Zhang, F.; Carothers, J. M.; Keasling, J. D. *Nat. Biotechnol.* **2012**, *30*, 354.
- (3) Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; Polichuk, D. R.; Teoh, K. H.; Reed, D. W.; Treynor, T.; Lenihan, J.; Fleck, M.; Bajad, S.; Dang, G.; Dengrove, D.; Diola, D.; Dorin, G.; Ellens, K. W.; Fickes, S.; Galazzo, J.; Gaucher, S. P.; Geistlinger, T.; Henry, R.; Hepp, M.; Horning, T.; Iqbal, T.; Jiang, H.; Kizer, L.; Lieu, B.; Melis, D.; Moss, N.; Regentin, R.; Secrest, S.; Tsuruta, H.; Vazquez, R.; Westblade, L. F.; Xu, L.; Yu, M.; Zhang, Y.; Zhao, L.; Lievens, J.; Covello, P. S.; Keasling, J. D.; Reiling, K. K.; Renninger, N. S.; Newman, J. D. *Nature* **2013**, *496*, 528.
- (4) Forster, A. C.; Church, G. M. *Genome Res.* **2007**, *17*, 1.
- (5) Forlin, M.; Lentini, R.; Mansy, S. S. *Curr. Opin. Chem. Biol.* **2012**, *16*, 586.
- (6) Boyle, P. M.; Silver, P. A. *J. R. Soc. Interface* **2009**, *6 Suppl 4*, S535.
- (7) Jewett, M. C.; Forster, A. C. *Curr. Opin. Biotechnol.* **2010**, *21*, 697.
- (8) Caschera, F.; Noireaux, V. *Curr. Opin. Chem. Biol.* **2014**, *22C*, 85.
- (9) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409*, 387.
- (10) Deamer, D. W. *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 239.
- (11) Hanczyc, M. M.; Fujikawa, S. M.; Szostak, J. W. *Science* **2003**, *302*, 618.
- (12) Zhu, T. F.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 5705.
- (13) Noireaux, V.; Libchaber, A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17669.
- (14) Monnard, P. A.; Deamer, D. W. *Orig. Life Evol. Biosph.* **2001**, *31*, 147.
- (15) Nishimura, K.; Matsuura, T.; Sunami, T.; Fujii, S.; Nishimura, K.; Suzuki, H.; Yomo, T. *RSC Adv.* **2014**, *4*, 35224.
- (16) Tawfik, D. S.; Griffiths, A. D. *Nat. Biotechnol.* **1998**, *16*, 652.
- (17) Griffiths, A. D.; Tawfik, D. S. *Trends Biotechnol.* **2006**, *24*, 395.
- (18) Levy, M.; Ellington, A. D. *ACS Chem. Biol.* **2008**, *15*, 979.
- (19) Long, M. S.; Jones, C. D.; Helfrich, M. R.; Mangeney-Slavin, L. K.; Keating, C. D. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5920.
- (20) Dewey, D. C.; Strulson, C. A.; Cacace, D. N.; Bevilacqua, P. C.; Keating, C. D. *Nat. Commun.* **2014**, *5*, 4670.
- (21) Adams, D. W.; Errington, J. *Nat. Rev. Microbiol.* **2009**, *7*, 642.
- (22) Osawa, M.; Anderson, D. E.; Erickson, H. P. *Science* **2008**, *320*, 792.
- (23) Cabre, E. J.; Sanchez-Gorostiaga, A.; Carrara, P.; Roperio, N.; Casanova, M.; Palacios, P.; Stano, P.; Jimenez, M.; Rivas, G.; Vicente, M. *J. Biol. Chem.* **2013**, *288*, 26625.
- (24) Baumgart, T.; Hess, S. T.; Webb, W. W. *Nature* **2003**, *425*, 821.
- (25) Andes-Koback, M.; Keating, C. D. *J. Am. Chem. Soc.* **2011**, *133*, 9545.
- (26) Zhu, T. F.; Adamala, K.; Zhang, N.; Szostak, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 9828.
- (27) Kuruma, Y.; Stano, P.; Ueda, T.; Luisi, P. L. *Biochim. Biophys. Acta* **2009**, *1788*, 567.
- (28) Oberholzer, T.; Albrizio, M.; Luisi, P. L. *ACS Chem. Bio.* **1995**, *2*, 677.
- (29) Mencia, M.; Gella, P.; Camacho, A.; de Vega, M.; Salas, M. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18655.
- (30) Li, Y.; Kim, H. J.; Zheng, C.; Chow, W. H.; Lim, J.; Keenan, B.; Pan, X.; Lemieux, B.; Kong, H. *Nucleic Acids Res.* **2008**, *36*, e79.
- (31) Forster, A. C.; Church, G. M. *Mol. Syst. Biol.* **2006**, *2*, 45.
- (32) Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. *Science* **2011**, *332*, 209.
- (33) Ichihashi, N.; Usui, K.; Kazuta, Y.; Sunami, T.; Matsuura, T.; Yomo, T. *Nat. Commun.* **2013**, *4*, 2494.

- (34) Kurihara, K.; Tamura, M.; Shohda, K.; Toyota, T.; Suzuki, K.; Sugawara, T. *Nat. Chem.* **2011**, *3*, 775.
- (35) Del Bianco, C.; Mansy, S. S. *Acc. Chem. Res.* **2012**, *45*, 2125.
- (36) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751.
- (37) Martini, L.; Mansy, S. S. *Chem. Commun.* **2011**, *47*, 10734.
- (38) Torre, P.; Keating, C. D.; Mansy, S. S. *Langmuir* **2014**, *30*, 5695.
- (39) Skerker, J. M.; Perchuk, B. S.; Siryaporn, A.; Lubin, E. A.; Ashenberg, O.; Goulian, M.; Laub, M. T. *Cell* **2008**, *133*, 1043.
- (40) Sallee, N. A.; Yeh, B. J.; Lim, W. A. *J. Am. Chem. Soc.* **2007**, *129*, 4606.
- (41) Shin, J.; Noireaux, V. *ACS Synth. Biol.* **2012**, *1*, 29.
- (42) Meyer, A. J.; Ellefson, J. W.; Ellington, A. D. *ACS Synth. Biol.* **2014**.
- (43) Liang, J. C.; Bloom, R. J.; Smolke, C. D. *Mol. Cell* **2011**, *43*, 915.
- (44) Bayer, T. S.; Smolke, C. D. *Nat. Biotechnol.* **2005**, *23*, 337.
- (45) Soukup, G. A.; Breaker, R. R. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3584.
- (46) Ogawa, A.; Maeda, M. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3156.
- (47) Wieland, M.; Benz, A.; Klauser, B.; Hartig, J. S. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 2715.
- (48) Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247.
- (49) Culler, S. J.; Hoff, K. G.; Smolke, C. D. *Science* **2010**, *330*, 1251.
- (50) Winkler, W. C.; Breaker, R. R. *ChemBioChem* **2003**, *4*, 1024.
- (51) Lemay, J.-F.; Desnoyers, G.; Blouin, S.; Heppell, B.; Bastet, L.; St-Pierre, P.; Massé, E.; Lafontaine, D. A. *PLoS Genet.* **2011**, *7*.
- (52) Topp, S.; Gallivan, J. P. *ACS Chem. Biol.* **2010**, *5*, 139.
- (53) Serganov, A.; Nudler, E. *Cell* **2013**, *152*, 17.
- (54) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952.
- (55) Lynch, S. A.; Gallivan, J. P. *Nucleic Acids Res.* **2009**, *37*, 184.
- (56) Mishler, D. M.; Gallivan, J. P. *Nucleic Acids Res.* **2014**, *42*, 6753.
- (57) Ogawa, A.; Maeda, M. *ChemBioChem* **2008**, *9*, 206.
- (58) Wieland, M.; Hartig, J. S. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 2604.
- (59) Mandal, M.; Breaker, R. R. *Nat. Struct. Mol. Biol.* **2004**, *11*, 29.
- (60) Muranaka, N.; Yokobayashi, Y. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 4653.
- (61) Topp, S.; Gallivan, J. P. *J. Am. Chem. Soc.* **2007**, *129*, 6807.
- (62) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
- (63) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
- (64) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611.
- (65) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550.
- (66) Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. *ACS Chem. Biol.* **2007**, *14*, 173.
- (67) Nomura, Y.; Yokobayashi, Y. *J. Am. Chem. Soc.* **2007**, *129*, 13814.
- (68) Ogawa, A. *RNA* **2011**, *17*, 478.
- (69) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12297.
- (70) Keefe, A. D.; Szostak, J. W. *Nature* **2001**, *410*, 715.
- (71) Hanes, J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4937.
- (72) Liu, R.; Barrick, J. E.; Szostak, J. W.; Roberts, R. W. *Methods Enzymol.* **2000**, *318*, 268.
- (73) Kurz, M.; Gu, K.; Lohse, P. A. *Nucleic Acids Res.* **2000**, *28*, E83.
- (74) Shultzaberger, R. K.; Bucheimer, R. E.; Rudd, K. E.; Schneider, T. D. *J. Mol. Biol.* **2001**, *313*, 215.
- (75) Lentini, R.; Forlin, M.; Martini, L.; Del Bianco, C.; Spencer, A. C.; Torino, D.; Mansy, S. S. *ACS Synth. Biol.* **2013**, *2*, 482.
- (76) Lynch, S. A.; Topp, S.; Gallivan, J. P. *Methods Mol. Biol.* **2009**, *540*, 321.
- (77) Grate, D.; Wilson, C. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6131.
- (78) Babendure, J. R.; Adams, S. R.; Tsien, R. Y. *J. Am. Chem. Soc.* **2003**, *125*, 14716.
- (79) Seelig, B. *Nat. Protoc.* **2011**, *6*, 540.

- (80) Muranaka, N.; Hohsaka, T.; Sisido, M. *Nucleic Acids Res.* **2006**, *34*, e7.
- (81) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411.
- (82) Win, M. N.; Smolke, C. D. *Science* **2008**, *322*, 456.
- (83) Robertson, M. P.; Ellington, A. D. *Nucleic Acids Res.* **2000**, *28*, 1751.
- (84) Soukup, G. A.; Emilsson, G. A.; Breaker, R. R. *J. Mol. Biol.* **2000**, *298*, 623.
- (85) Wieland, M.; Benz, A.; Haar, J.; Halder, K.; Hartig, J. S. *Chem. Commun.* **2010**, *46*, 1866.
- (86) Yurke, B.; Turberfield, A. J.; Mills, A. P., Jr.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605.
- (87) Li, Q.; Luan, G.; Guo, Q.; Liang, J. *Nucleic Acids Res.* **2002**, *30*, E5.
- (88) Li, B.; Ellington, A. D.; Chen, X. *Nucleic Acids Res.* **2011**, *39*, e110.
- (89) Zhang, D. Y.; Winfree, E. *J. Am. Chem. Soc.* **2009**, *131*, 17303.
- (90) Zhang, D. Y.; Seelig, G. *Nat. Chem.* **2011**, *3*, 103.
- (91) Seeman, N. C. *Nano Lett.* **2010**, *10*, 1971.
- (92) Yan, H.; Zhang, X.; Shen, Z.; Seeman, N. C. *Nature* **2002**, *415*, 62.
- (93) Yin, P.; Choi, H. M.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318.
- (94) Jiang, Y. S.; Li, B.; Milligan, J. N.; Bhadra, S.; Ellington, A. D. *J. Am. Chem. Soc.* **2013**, *135*, 7430.
- (95) Compton, J. *Nature* **1991**, *350*, 91.
- (96) Bhadra, S.; Ellington, A. D. *RNA* **2014**, *20*, 1183.
- (97) Bhadra, S.; Ellington, A. D. *Nucleic Acids Res.* **2014**, *42*, e58.
- (98) Green, Alexander A.; Silver, Pamela A.; Collins, James J.; Yin, P. *Cell* **2014**, *159*, 925.
- (99) Akter, F.; Yokobayashi, Y. *ACS Synth. Biol.* **2014**.
- (100) Lentini, R.; Santero, S. P.; Chizzolini, F.; Cecchi, D.; Fontana, J.; Marchioretto, M.; Del Bianco, C.; Terrell, J. L.; Spencer, A. C.; Martini, L.; Forlin, M.; Assfalg, M.; Dalla Serra, M.; Bentley, W. E.; Mansy, S. S. *Nat. Commun.* **2014**, *5*, 4012.
- (101) Lemay, J. F.; Desnoyers, G.; Blouin, S.; Heppell, B.; Bastet, L.; St-Pierre, P.; Masse, E.; Lafontaine, D. A. *PLoS Genet.* **2011**, *7*, e1001278.
- (102) Serganov, A.; Polonskaia, A.; Phan, A. T.; Breaker, R. R.; Patel, D. J. *Nature* **2006**, *441*, 1167.
- (103) Edwards, T. E.; Ferre-D'Amare, A. R. *Structure* **2006**, *14*, 1459.

Appendix

In the appendix is inserted the paper

Fluorescent Proteins and in vitro Genetic Organization for Cell-Free Synthetic Biology

Roberta Lentini, Michele Forlin, **Laura Martini**, Cristina Del Bianco, Amy C. Spencer, Domenica Torino, and Sheref S. Mansy; *ACS Synthetic Biology* **2013**, 9, 482-9.

My contribution to the work regarded the sequences design and construction, editing of the manuscript prior to publication.

** Reprinted with the permission of the copyright holder American Chemical Society*

Fluorescent Proteins and *in Vitro* Genetic Organization for Cell-Free Synthetic Biology

Roberta Lentini, Michele Forlin, Laura Martini, Cristina Del Bianco, Amy C. Spencer, Domenica Torino, and Sheref S. Mansy*

CIBIO, University of Trento, via delle Regole 101, 38123 Mattarello (TN), Italy

Supporting Information

ABSTRACT: To facilitate the construction of cell-free genetic devices, we evaluated the ability of 17 different fluorescent proteins to give easily detectable fluorescence signals in real-time from *in vitro* transcription-translation reactions with a minimal system consisting of T7 RNA polymerase and *E. coli* translation machinery, i.e., the PUREsystem. The data were used to construct a ratiometric fluorescence assay to quantify the effect of genetic organization on *in vitro* expression levels. Synthetic operons with varied spacing and sequence composition between two genes that coded for fluorescent proteins were then assembled. The resulting data indicated which restriction sites and where the restriction sites should be placed in order to build genetic devices in a manner that does not interfere with protein expression. Other simple design rules were identified, such as the spacing and sequence composition influences of regions upstream and downstream of ribosome binding sites and the ability of non-AUG start codons to function *in vitro*.

KEYWORDS: cell-free, fluorescent protein, transcription-translation, ribosome binding site, synthetic biology



The majority of synthetic biology research makes use of a living chassis that provides for the necessary but poorly characterized biological components required for life. Conversely, a smaller community of synthetic biologists has begun to build cell-like systems with a nonliving, cell-free chassis.^{1–7} Although the cell-free branch of synthetic biology has progressed more slowly, success could provide for new technologies with several beneficial features. For example, the resulting cellular mimics would consist of fully defined components. Therefore, it should be possible to build a complete mathematical model describing the cellular mimic that could aid in designing new features. Additionally, potentially technologically problematic features of life, such as evolution, could be intentionally removed by building systems that do not replicate.

A significant step forward in allowing for the construction of such well-defined, bottom-up systems came from Ueda and colleagues, who showed that coupled transcription and translation reactions can be mediated by fully defined components *in vitro*.⁸ Their system, hereafter referred to as the PUREsystem, consisted of T7 RNA polymerase and *Escherichia coli* translation machinery. Subsequent work demonstrated the compatibility of the PUREsystem with liposomes^{9,10} and with the expression of gene networks.¹¹ Nevertheless, there has been little attempt to better define the influences of genetic organization on protein output with purified transcription-translation machinery. Recently, a S30 *E. coli* cell extract translation system and the PUREsystem were used to determine the influences of different ribosome binding sites and transcriptional repressors on the synthesis of eGFP.^{12,13}

Although much is known about natural, *in vivo* genetics, much still remains unresolved. For example, the refactoring of

the T7 genome was successful in the sense that viable bacteriophage were produced; however, the refactored bacteriophage was significantly less infective.¹⁴ Similar challenges are routinely encountered when genetic elements are inserted into organisms to engineer new circuitry. Typically, many permutations are required before desired function is achieved.¹⁵ The situation is perhaps even more challenging for systems that exploit a cell-free chassis since biological parts are evolved to function optimally under the chemical conditions found *in vivo*. *In vitro* conditions are undoubtedly different. Further, unidentified molecular components necessary for activity *in vivo* may be missing from *in vitro* constructions. The design and implementation of predictable, genetically encoded cell-free systems is difficult because of the lack of cell-free chassis data coupled with an incomplete understanding of natural, *in vivo* genetics.

Here we sought to identify some practical rules for the construction of genetically encoded, cell-free systems. First, 17 different fluorescent proteins were screened for their ability to generate easily detectable fluorescence signals after *in vitro* transcription and translation with the PUREsystem. Fluorescent proteins then were expressed from a bicistronic construct to identify fluorescent protein pairs that could be used to quantify the influences of genetic organization on protein production. A series of synthetic operons that differed in the spacing and sequence between the two encoded genes, the spacing and sequence between the ribosome binding site and the start codon, and the influence of the first nucleotide position of the start codon on *in vitro* expression levels was assessed with the developed ratiometric fluorescence assay. We

Received: January 17, 2013

Published: March 8, 2013

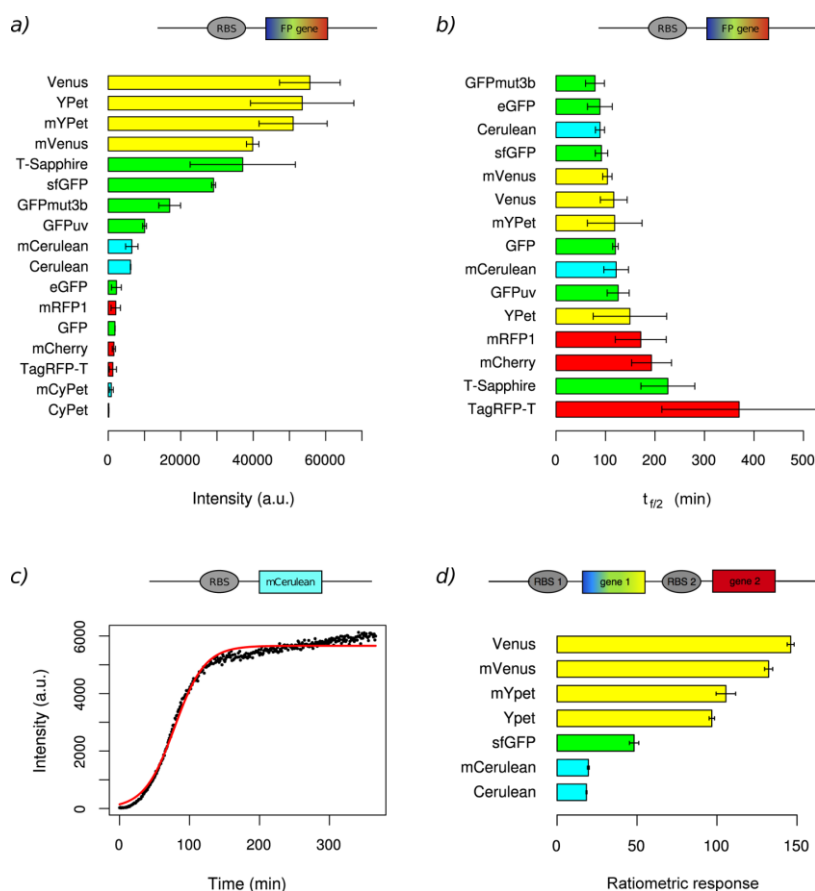


Figure 1. Fluorescence profiles of *in vitro* expressed genetic constructs at 37 °C with the PUREsystem. (a) Fluorescence intensities after 6 h of *in vitro* expression for 17 different fluorescent proteins. (b) The $t_{1/2}$ of each fluorescent protein was calculated by fitting the kinetic data to a logistic model as described in the Methods section. The $t_{1/2}$ represents the time at maximum growth. (c) The fitting of mCerulean kinetic data is shown as a representative example. The logistic model estimation is shown in red, while the black points represent measured values. A control reaction without plasmid showed no fluorescence. (d) The ratiometric response of bicistronic constructs after 6 h of *in vitro* expression. The ratiometric response was calculated by dividing the fluorescence arising from the protein encoded by gene 1 by the fluorescence resulting from the gene product of gene 2. Here gene 2 always encoded mCherry. A cartoon above each panel gives a schematic representation of the used constructs. The data shown in panels a and b are from constructs RL001A-RL013A and CD100A-CD103A. Panel c used RL005A, and panel d used RL015A-RL021A. More information on each construct is provided in Supplementary Table S1.

found that a high guanosine content inhibited translation, that sequences 5' to the ribosome binding site were more amenable to the incorporation of restriction sites for cloning, and that ribosome binding sites were most efficient when separated from the start codon by 4–9 nucleotide positions. GUG, UUG, and CUG were functional as start codons in minimal, reconstituted translation systems, although their associated expression levels were significantly reduced.

RESULTS AND DISCUSSION

***In Vitro* Expression of Fluorescent Proteins.** A total of 17 different fluorescent proteins were tested individually for their ability to give easily detectable fluorescence signals from *in vitro* transcription-translation reactions with the PUREsystem at 37 °C. Of these 17 proteins, four (mCerulean, mCyPet, mVenus, and mYPet) contained a A206K substitution to inhibit dimerization. As seen in Figure 1a, all of the tested constructs

produced easily detectable signals above background arising from the fluorescent protein except for CyPet and mCyPet. These two cyan fluorescent proteins gave slightly increased fluorescence when expressed at 30 °C (Supplementary Figure S1). Consistent with the reported brightness of each fluorescent protein,¹⁶ the yellow fluorescent proteins were associated with the most intense fluorescence, followed by the green, cyan, and red fluorescent proteins (Figure 1a). Monomeric versions of Cerulean and YPet gave fluorescence intensities within 5% of their dimeric parent construct. *In vitro* transcribed and translated Venus was 40% more intense than mVenus; however, the error associated with the single fluorescent protein measurements was too large to make meaningful conclusions. This issue was resolved by using a ratiometric method described below. After 6 h of *in vitro* transcription-translation, the mVenus concentration reached 8 μ M.

Most of the constructs gave sigmoidal shaped kinetic profiles and were complete within 6 h. The exceptions were T-Sapphire and TagRFP-T (Supplementary Figure S2), both of which did not reach their maximal fluorescence within 6 h. The fitting of the kinetic data to a logistic model was used to determine the time point at which the rate of fluorescence increase was maximal, which corresponded to the time required to reach half maximal fluorescence ($t_{1/2}$) (Figure 1b,c). Note that the $t_{1/2}$ includes all of the steps involved in converting the information encoded in DNA to a fluorescence signal and does not solely describe the last oxidation step of chromophore formation.¹⁷ The shortest $t_{1/2}$ value was 79 min for GFPmut3b, and the longest $t_{1/2}$ was over 300 min for TagRFP-T (Supplementary Table S4). The average $t_{1/2}$ values for the expression of cyan, green, yellow, and red fluorescent proteins were 105, 122, 122, and 245 min, respectively. The $t_{1/2}$ was 40% larger for mCerulean than Cerulean, whereas mVenus and mYPet had $t_{1/2}$ values 12% and 26% smaller than Venus and YPet, respectively. On the basis of fluorescence intensity and kinetic data, Cerulean, mCerulean, super folder GFP (sfGFP), Venus, mVenus, YPet, mYPet, mRFP1, and mCherry were selected for further analysis.

To reduce experimental error, we pursued the construction of a ratiometric fluorescence system based on synthetic operons that encoded two fluorescent proteins. In this way the influences of pipetting, lamp performance, and DNA template quality and concentration, among other difficult to control variables, would be removed. To build such a ratiometric system, a red fluorescent protein was desirable because the excitation and emission spectra of red fluorescent proteins are better separated from the fluorescence spectra of other fluorescent proteins. mRFP1 and mCherry were, therefore, tested in bicistronic constructs that additionally encoded sfGFP to evaluate their utility in characterizing expression levels. More specifically, small synthetic operons containing a standard T7 transcriptional promoter, a ribosome binding site (RBS), a gene encoding sfGFP followed by a sequence that encoded the red fluorescent protein and a T7 transcriptional terminator were assembled. All of the fluorescent proteins in these constructs gave reproducible and easily detectable fluorescence signals. After 6 h of expression with purified transcription-translation machinery, the ratio of sfGFP fluorescence to mRFP1 and to mCherry fluorescence was 115.1 ± 6.9 and 49.9 ± 2.4 , respectively (Supplementary Figure S3). We chose to use mCherry for the remaining experiments, because mCherry showed more intense fluorescence from the bicistronic construct and because mCherry was shown to be more photostable than mRFP1.¹⁶

We next assembled six additional synthetic operons that encoded different fluorescent proteins followed by a sequence coding for mCherry. After *in vitro* transcription and translation, the fluorescence profiles were similar to those obtained with the single fluorescent protein constructs in that the yellow fluorescent proteins were the most intense, followed by green, and cyan fluorescent proteins (Figure 1d). However, the error of each ratiometric measurement was significantly reduced (relative standard error <8%) in comparison to the data obtained from the monocistronic, single fluorescent protein constructs (relative standard error <60%, excluding TagRFP-T). The A206K substitution that inhibits protein dimerization had a small effect on fluorescence intensity. More specifically, the ratiometric response, i.e., the fluorescence intensity of the fluorescent protein tested divided by the

fluorescence intensity of mCherry, for mVenus, mYPet, and mCerulean were within 10% of the values measured for Venus, YPet, and Cerulean, respectively. The ratiometric response over time showed that stable readings could be taken after 3 h for all constructs tested (Supplementary Figure S4).

It was not clear from the outset which fluorescent proteins would perform well *in vitro* with minimal transcription-translation machinery. Although the physical characteristics of individually purified proteins, such as brightness and photostability, are useful in deciding if a protein could be suitable for a specific application, these parameters are not enough to understand if *in vitro* expression will give a robust, reproducible signal. For example, if *in vitro* produced protein is insoluble, folds slowly, or requires a long period of time for chromophore formation, then that protein would be less useful as an *in vitro* genetic reporter. Even within cells, differences in fluorescent protein behavior have been noted, particularly for multidomain proteins.¹⁸ Despite these difficulties, we found that most of the fluorescent proteins tested function satisfactorily in *in vitro* transcription-translation reactions with the PUREsystem at 37 °C. One exception is CyPet, which fails to give a significant fluorescent output. The fact that CyPet expression at 30 °C gives a better fluorescence signal is consistent with previous reports on the poor folding properties of CyPet.¹⁶ If a fluorescent protein with cyan spectral properties were desired, cerulean would be a better choice. The green fluorescent proteins are generally bright and rapidly give rise to fluorescence signals, e.g., the $t_{1/2}$ of sfGFP is 92 min. sfGFP is particularly amenable to *in vitro* transcription-translation; however, GFPmut3b performs similarly well. GFPmut3b is one of the more common fluorescent proteins used in synthetic biology. Two of the tested green fluorescent proteins fluoresce upon excitation with near-UV light. Of these two, T-Sapphire has a $t_{1/2}$ approximately 100 min longer than that of GFPuv. Therefore, GFPuv would be better for real-time detection assays than T-Sapphire. The yellow fluorescent proteins Venus and YPet are the brightest fluorescent proteins that we tested and have $t_{1/2}$ values below 150 min. Venus and YPet are excellent choices to monitor *in vitro* reactions particularly when low protein output is expected, e.g., when expressing inside of vesicles.¹⁹ YPet is more photostable,¹⁶ which could be important depending upon the nature of the planned experiments. The red fluorescent proteins mCherry and mRFP1 perform similarly well in *in vitro* transcription-translation reactions, but mCherry is more photostable. Although TagRFP-T is a highly photostable red fluorescent protein alternative, the long $t_{1/2}$ of TagRFP-T limits its usefulness.

All of the seven tested double fluorescent protein constructs performed well, and so the choice of fluorescent protein pairs depends on the specifics of the experimental setup. We found that the mVenus-mCherry pair gives easy to detect fluorescence signals and reproducible data without interference between the emission of mVenus and the emission of mCherry. Therefore, the subsequent experiments that probed the effects of genetic organization on protein production were performed with synthetic operons encoding mVenus and mCherry. However, for the remaining experiments the order of the genes was reversed so that mCherry was encoded first followed by mVenus in the bicistronic message. In this way, the lower intensity fluorescent protein, i.e., mCherry, could be used to provide the reference fluorescence signal and the influences of the region between the two genes on the expression of the

brighter fluorescent protein, i.e., mVenus, could be more easily assessed. Nevertheless, care should be taken in interpreting the resulting data. The assay can be used to characterize how changes in DNA sequence influence the ratio of the two encoded proteins. However, the assay does not differentiate between the decrease of expression of gene 1 or the increase of expression of gene 2. In other words, multiple mechanisms can give indistinguishable results.

Influence of Sequences Upstream of the Ribosome Binding Site. The first question we sought to answer was whether the number of nucleotides separating the stop codon of gene 1 from the ribosome binding site of gene 2 influenced gene expression. Therefore, constructs containing 0, 5, 20, 31, and 50 bp spacer sequences between the UAA stop codon of gene 1 and the AAGGAG RBS of gene 2 were tested (Figure 2). Although differences in expression levels were observed, the

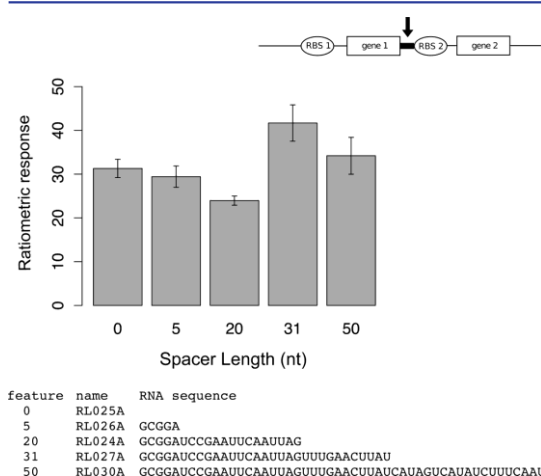


Figure 2. Influence of spacer length between an upstream gene and a downstream ribosome binding site on expression levels. The ratiometric response represents the fluorescence arising from mVenus (encoded by gene 2) divided by the fluorescence of mCherry (encoded by gene 1). Spacer lengths of 0, 5, 20, 31, and 50 nucleotides were tested. The corresponding RNA sequence for the region of interest of each construct is shown below the graph. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h.

differences did not correlate with the length of the spacer. For example, the 5 bp and the 31 bp spacer containing constructs both resulted in higher relative expression of gene 2 when compared with the 20 bp spacer. This suggested that the variance in fluorescence ratios resulted from something other than spacer length, such as sequence composition. For the remainder of the experiments, the 31 bp spacer construct (RL027A) was used as the reference.

Since the length of the spacer between gene 1 and RBS 2 did not appear to be correlated with the expression of gene 2, we wondered if the sequence composition rather than the length was responsible for the observed differences in expression. We decided to investigate the influences of sequence composition by incorporating different restriction sites immediately upstream to RBS 2. In this way we hoped to additionally identify restriction sites useful for the assembly of genetically encoded devices. Therefore, in each of the tested constructs, the 31 bp spacer length was maintained, and sequences containing a

NdeI, BamHI, NheI, EcoRI, NotI, or a scar site were incorporated. The scar site represented the sequence that results from standard BioBrick assembly in which complementary XbaI and SpeI digested products are ligated.²⁰ Additionally, the U before the AAGGAG RBS was mutated to a G, since a U residue is capable of base-pairing with 16S rRNA. A significant effect of sequence composition on the amount of protein produced was observed (Figure 3). The

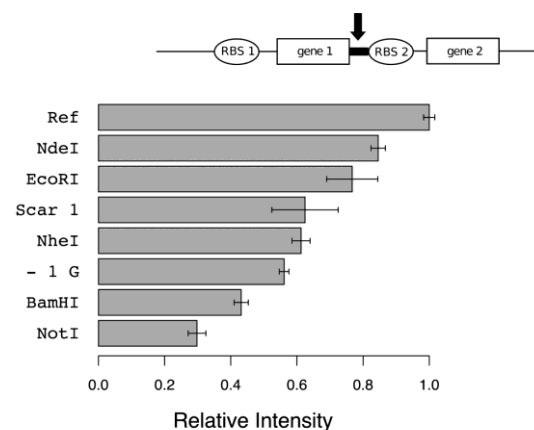


Figure 3. Influence of sequence composition upstream of the ribosome binding site on *in vitro* expression levels. The corresponding RNA sequence for the region of interest of each construct is shown below the graph. Underlined positions refer to the introduced feature. Ref refers to the reference construct RL027A, Scar 1 indicates the standard BioBrick scar sequence, and -1 G refers to the introduction of a G immediately prior to RBS 2. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. Data are plotted relative to RL027A.

introduction of a NotI site was the most inhibitory, bringing relative expression down by 70% in comparison to the reference RL027A construct. Of the restriction sites tested, NdeI and EcoRI restriction site sequences were the most conducive to high expression (84% and 77% relative expression, respectively). Removing the additional base-pair of the RBS, i.e., the U to G mutation, decreased protein production by 44%, consistent with the observed decrease in expression from the 20 bp spacer construct described above that contained the same nucleotide at this position.

Influence of Sequences Downstream of the Ribosome Binding Site. Having probed the influences of the region 5' to RBS 2, we next investigated the impact of the region 3' to RBS 2. First, we altered the spacing between RBS 2 and the start codon of gene 2 one nucleotide at a time from -2 to 15 bp. Here the spacing nomenclature followed the aligned spacing described by Chen et al.²¹ in which the RBS was aligned with

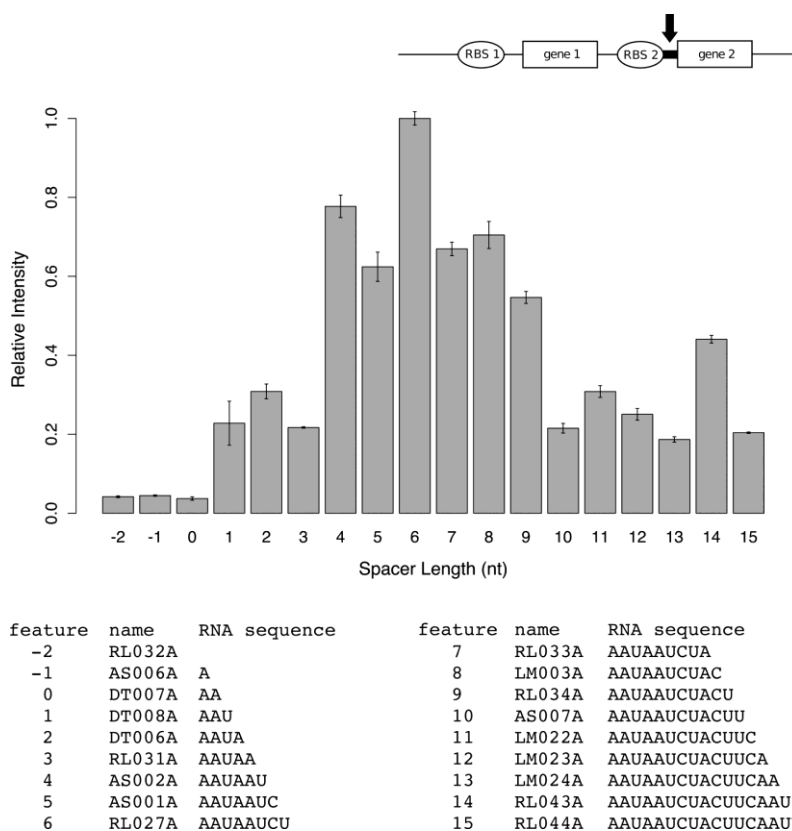


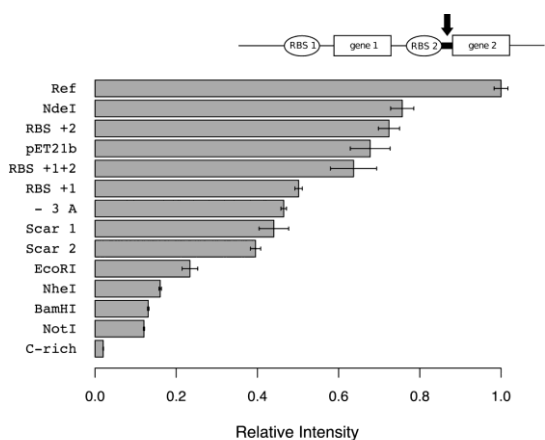
Figure 4. Ribosome binding site spacing. The influence of the aligned spacing between the ribosome binding site and the start codon is shown. The corresponding RNA sequence for the region of interest of each construct is reported below the graph. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. Data are plotted in reference to RL027A.

the anti-RBS sequence of the 16S rRNA and the position across from the last position of the anti-RBS was taken as 0 (Supplementary Figure S5). The results were consistent with previous *in vivo* studies,²² which showed a Gaussian distribution of activity with optimal aligned spacing between 4 and 9 bp (Figure 4). Spacer lengths shorter or longer than this range generally resulted in dramatically decreased protein production. For example, the 3 bp spacer produced 72% less protein than the 4 bp spacer. Similarly, the 10 bp spacer reduced protein synthesis by 60% when compared to the 9 bp spacer construct. For the specific constructs tested in this study, the 6 bp spacer produced the most protein. Since protein expression was detected with the shortest spacer tested on both sides of RBS 2, we also made a minimal construct with a 0 bp spacer between the UAA stop codon of gene 1 and RBS 2 and -2 aligned spacing between RBS 2 and the start codon of gene 2. The synthesis of mVenus from this minimally spaced construct was low but still detectable (3% relative to RL027A).

Next, we evaluated the effect of sequence composition of the region between RBS 2 and the AUG start codon of gene 2 on expression levels. This region of the reference sequence RL027A was designed to be high in A-U content and low in G content because a sequence that is known to facilitate gene expression, i.e., the T7 phage gene 10 leader sequence,²³ has similar characteristics. Sequences that contained the same

restriction sites tested above for the region upstream of RBS 2 were placed immediately upstream of the start codon of gene 2. An additional BioBrick scar site also was screened that was shorter and thus thought to interfere less with translation. The presence of an A three nucleotides upstream of the start codon was evaluated since an A at this position is frequently found in prokaryotic and eukaryotic sequences.^{24,25} A C-rich sequence was evaluated since a previous *in vitro* study²⁶ found increased expression associated with high C-content. Finally, mutations that introduced additional base-pairing with the 16S rRNA were added. The data showed a strong influence of sequence composition on protein yields with the NotI restriction site being the most inhibitory, decreasing expression by 87% (Figure 5). The NdeI restriction site was the most conducive to protein synthesis (76% relative expression). Both scar sequences resulting from BioBrick assembly performed similarly, decreasing translation by over 50%. The C-rich sequence greatly decreased protein expression by 98% relative to RL027A. Neither an A residue three nucleotides preceding the start codon nor the expansion of the RBS-anti-RBS base-pairing region increased protein production in the tested constructs.

Finally, we investigated whether other codons could substitute for the AUG start codon. In *E. coli*, GUG and UUG function as start codons at a frequency of 14% and 3%,



feature	name	RNA sequence	feature	name	RNA sequence
Ref	RL027A	AAUAUCU	Scar 1	LM010A	UACUAGAG
NdeI	LM005A	AAUACAU	Scar 2	LM011A	AAUACUAG
RBS +2	LM014A	AUUAUCU	EcoRI	LM008A	AAGAAUUC
pET21b	LM016A	AUAUACAU	NheI	LM007A	AAGCUAGC
RBS +1+2	LM015A	GUUAUCU	BamHI	LM006A	AAGGAUCC
RBS +1	LM013A	GAUAUCU	NotI	LM009A	GCGGCCCC
-3 A	LM012A	AAUAAUCU	C-rich	RL050A	CCCCUCC

Figure 5. Influence of sequence composition between the ribosome binding site and the start codon on expression levels. The corresponding RNA sequence for the region of interest of each construct is shown below the bar graph. Ref indicates the reference construct RL027A. Scar 1 is the standard BioBrick scar sequence. Scar 2 is the shorter, alternate scar sequence. -3 A indicates the introduction of an A three positions upstream of the start codon. pET21b is the same spacer sequence found in the expression vector pET21b (Novagen). RBS +1, RBS +2, and RBS +1+2 indicate RBS expansions. Each introduced feature is underlined in the corresponding sequence. Note that only half of the NdeI restriction site is shown since the remaining half overlaps with the start codon. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. Data are plotted in reference to RL027A.

respectively.²⁵ If non-AUG codons can function as start codons in minimally reconstituted systems, then these alternate start codons could be used to control protein levels. Also, knowledge regarding the functionality of non-AUG start codons could help to identify internal RBS-start codon pairs that could potentially interfere with the intended activity of genetic devices. We therefore substituted a GUG, UUG, and CUG in place of the AUG start codon and measured the production of mVenus. All of the alternate start codons produced protein, albeit at a significantly reduced level between 12% and 27% relative to the AUG start codon containing reference construct (Figure 6).

Considerations for the Assembly of *in Vitro* Genetic Systems. To determine if simple rules could be formulated that would facilitate the construction of genetically encoded, cell-free devices, the collected data were statistically analyzed. First, we sought to determine which regions were more amenable to the incorporation of restriction sites. A paired *t* test showed that sequences upstream of RBS 2 had less influence on the protein fluorescence ratios than the sequence between RBS 2 and the start codon (p -value = 0.0145). Next, sequences immediately 5' and 3' to RBS 2 (8 bp each) were considered. The resulting data from 22 synthetic operons were fit to multiple regression models that searched for first and second order interactions between base composition that

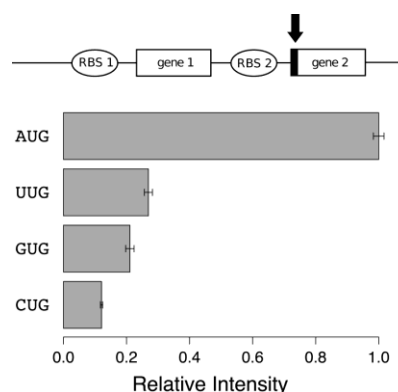


Figure 6. Alternate start codons. The ability of UUG, GUG, and CUG to function as start codons *in vitro* was evaluated. Relative intensities are averages of three replicates and plotted in reference to the AUG start codon containing construct. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. The AUG, UUG, GUG, and CUG start codon constructs were RL027A, LM019A, LM018A, and LM020A, respectively.

correlated with the measured fluorescence intensity ratios. The resulting model was statistically significant (F -test p -value = 8.79×10^{-7}) and described almost 75% of the data variability (adjusted $r^2 = 0.7453$). The estimated parameters (Supplementary Table S1) revealed a strong effect of the G content in sequence composition of the region 5' to RBS 2 (p -value < 0.001). More specifically, a high G content negatively correlated with the fluorescence ratio, whereas combined A-U-rich sequences in the region 3' to RBS 2 positively correlated with the fluorescence intensity ratio (p -value < 0.001).

Taken together, the data indicate that the nucleotide sequence between genes 1 and 2 influence protein production, but not uniformly. The spacing upstream of the RBS is not as strong of a determinant of expression levels as the spacing downstream of the RBS. The one construct that deviates from this trend (RL024A) contains a mutation that decreases the number of potential base-pairs between the mRNA and the 16S rRNA. Most of the constructs tested here contain six to seven potential base-pairing interactions between the Shine-Dalgarno (RBS) and the anti-Shine-Dalgarno site of the ribosome. The introduction of additional base-pairing does not facilitate expression, consistent with previous studies that show that on average *E. coli* mRNA RBS sequences interact with the ribosome via six base-pairs and that the strengthening of the interaction often decreases rather than increases protein synthesis.²⁷ The optimal aligned spacing between the RBS and the start codon and the functionality of alternate start codons is the same for *in vitro* protein production with the PUREsystem and for natural *E. coli* expression.

Taken together, a few simple rules for the construction of *in vitro* genetic systems can be formulated from the acquired data. Restriction sites should either be placed before the RBS, since this region is more amenable to sequence modification, or a NdeI site that overlaps with the start codon should be exploited. If high protein levels are desired, then the aligned RBS spacing should be between four and nine nucleotides and the spacer sequence should be high in A and T content and low in G content. The use of alternate start codons can be used to

significantly reduce protein synthesis, when needed, and the spacing between the end of one gene and the RBS of the next gene is not crucial. Nevertheless, the complexity of transcription and translation ensures that there are many more factors that influence gene expression than was probed here. mRNA can interact with regions of the ribosome other than the 3'-terminus of the 16S rRNA^{26–29} and the folding of mRNA significantly affects protein synthesis.^{19,30–33} Further studies with purified, *in vitro* systems likely will aid in better understanding these processes and in facilitating the synthesis of more complex cellular mimics.

METHODS

Genetic Constructs. Genes encoding the fluorescent proteins were synthesized by Genscript or Mr. Gene, except for super folder GFP (BBa_I746916), GFPmut3b (BBa_E0040), and mRFP1 (BBa_E1010), which were from the registry of standard biological parts (<http://partsregistry.org>), and eGFP, which was from Roche. Mutagenesis was either performed by Genscript or through the use of phusion site-directed mutagenesis (Finnzymes). All genes were subcloned into pET21b by restriction digestion and ligation with NdeI and BamHI, except for super folder GFP and GFPmut3b, which used NheI and BamHI sites. All constructs were confirmed by sequencing at Genechron or Eurofins MWG Operon. The DNA sequences of all the constructs used are provided in the Supporting Information (Table S2).

Transcription-Translation Reactions. Plasmids were amplified in *E. coli* DH5 α or NovaBlue and purified with Wizard Plus SV Minipreps DNA Purification System (Promega) or QIAprep Spin Miniprep Kit (Qiagen). Subsequently, the DNA was phenol-chloroform extracted, ethanol precipitated, and resuspended in deionized and diethylpyrocarbonate (DEPC) treated water. A 250 ng portion (2 nM final concentration) of DNA was used for each transcription-translation reaction with the PURExpress *in vitro* protein synthesis kit (New England BioLabs) supplemented with 20 units of human placenta RNase inhibitor (New England BioLabs). The final volume of each reaction was 25.5 μ L. Reactions were monitored by fluorescence spectroscopy with a Photon Technology International (PTI) QuantaMaster 40 UV–vis spectrofluorometer equipped with two detectors (T-format). Excitation and emission wavelengths were specific for each fluorescent protein (Supplementary Table S3). The reaction components, except for the DNA template, were assembled on ice and then incubated at 37 °C in the spectrofluorometer. Subsequently, the reaction was initiated by the addition of DNA template. Mineral oil was layered on top of each sample to inhibit evaporation during the course of the experiment. Control experiments with GFPmut3b showed that mineral oil did not influence the appearance of fluorescence. Each reaction was repeated at least three times. An Agilent 8453 UV–vis spectrophotometer was used to quantify mVenus protein concentration by using an extinction coefficient at 515 nm of 92,200 M⁻¹ s⁻¹.¹⁶

Data analysis. All statistical analyses used R statistical computing software.³⁴ The single protein construct fluorescent data were fit to

$$I(t) = \frac{K}{1 + e^{-B(t-t_{1/2})}} \quad (1)$$

where K , B , and $t_{1/2}$ were the upper asymptote, growth rate, and time of maximum growth, respectively (Supplementary Table S4). The parameters were estimated by using a nonlinear least-squares analysis with the Gauss–Newton algorithm. The mean values and standard errors were then calculated from data from three replications. The influence of spacer nucleotide composition on the fluorescence intensity was determined with multiple regression models. The models were estimated and reduced by using stepwise regression with a penalty term that was selected by minimum predictive mean squared error based on repeated cross-validation (10% leave-out). The best predictive models were obtained by using a stringent criterion (twice the Bayesian Information Criterion, BIC). We then estimated the model with such a penalty term on the whole set of operon spacer data. Paired t tests were used to test whether the restriction sites 5' or 3' to RBS 2 affected differently fluorescence intensity ratios.

ASSOCIATED CONTENT

Supporting Information

Supplementary tables and figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Tel: +39 0461 28 3438. Fax: +39 0461-283091. E-mail: mansy@science.unitn.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

RL015A, RL016A, RL018A, RL020A, and RL027A are available through Addgene. Versions of RL027A modified by the 2012 Trento iGEM team to be BioBrick compatible were deposited as BBa_K731700 and BBa_K731710 in the Registry of Standard Biological Parts (<http://partsregistry.org/>). We thank the Armenise-Harvard Foundation, the autonomous province of Trento (Ecomm), the Marie-Curie Trentino COFUND (ACS), and CIBIO for funding.

REFERENCES

- (1) Forlin, M., Lentini, R., and Mansy, S. S. (2012) Cellular imitations. *Curr. Opin. Chem. Biol.* 16, 586–592.
- (2) Forster, A. C., and Church, G. M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.* 2, 45.
- (3) Harris, D. C., and Jewett, M. C. (2012) Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. *Curr. Opin. Biotechnol.* 23, 672–678.
- (4) Ichihashi, N., Matsuura, T., Kita, H., Sunami, T., Suzuki, H., and Yomo, T. (2010) Constructing partial models of cells. *Cold Spring Harbor Perspect. Biol.* 2, 295–303.
- (5) Martos, A., Jimenez, M., Rivas, G., and Schwillie, P. (2012) Towards a bottom-up reconstitution of bacterial cell division. *Trends Cell Biol.* 22, 634–643.
- (6) Noireaux, V., Maeda, Y. T., and Libchaber, A. (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3473–3480.
- (7) Stano, P., and Luisi, P. L. (2010) Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chem. Commun. (Cambridge)* 46, 3639–3653.
- (8) Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755.

- (9) Stano, P., Kuruma, Y., Souza, T. P., and Luisi, P. L. (2010) Biosynthesis of proteins inside liposomes. *Methods Mol. Biol.* 606, 127–145.
- (10) Sunami, T., Matsuura, T., Suzuki, H., and Yomo, T. (2010) Synthesis of functional proteins within liposomes. *Methods Mol. Biol.* 607, 243–256.
- (11) Kita, H., Matsuura, T., Sunami, T., Hosoda, K., Ichihashi, N., Tsukada, K., Urabe, I., and Yomo, T. (2008) Replication of genetic information with self-encoded replicase in liposomes. *ChemBioChem* 9, 2403–2410.
- (12) Karig, D. K., Iyer, S., Simpson, M. L., and Doktycz, M. J. (2012) Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res.* 40, 3763–3774.
- (13) Shin, J., and Noireaux, V. (2012) An E. coli cell-free expression toolbox: application to synthetic gene circuits and synthetic cell. *ACS Synth. Biol.* 1, 29–41.
- (14) Chan, L. Y., Kosuri, S., and Endy, D. (2005) (2005) Refactoring bacteriophage T7. *Mol. Syst. Biol.* 1, 0018.
- (15) Kwok, R. (2010) Five hard truths for synthetic biology. *Nature* 463, 288–290.
- (16) Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–909.
- (17) Iizuka, R., Yamagishi-Shirasaki, M., and Funatsu, T. (2011) Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem.* 414, 173–178.
- (18) Chang, H. C., Kaiser, C. M., Hartl, F. U., and Barral, J. M. (2005) De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. *J. Mol. Biol.* 353, 397–409.
- (19) Martini, L., and Mansy, S. S. (2011) Cell-like systems with riboswitch controlled gene expression. *Chem. Commun. (Cambridge)* 47, 10734–10736.
- (20) Shetty, R. P., Endy, D., and Knight, T. F., Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2, 5.
- (21) Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res.* 22, 4953–4957.
- (22) Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Mol. Microbiol.* 6, 1219–1229.
- (23) Olins, P. O., Devine, C. S., Rangwala, S. H., and Kavka, K. S. (1998) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in Escherichia coli. *Gene* 73, 227–235.
- (24) Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292.
- (25) Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. (2001) Anatomy of Escherichia coli ribosome binding sites. *J. Mol. Biol.* 313, 215–228.
- (26) Barendt, P. A., Shah, N. A., Barendt, G. A., and Sarkar, C. A. (2012) Broad-specificity mRNA-rRNA complementarity in efficient protein translation. *PLoS Genet.* 8, e1002598.
- (27) Vimberg, V., Tats, A., Remm, M., and Tenson, T. (2007) Translation initiation region sequence preferences in Escherichia coli. *BMC Mol. Biol.* 8, 100.
- (28) Etchegaray, J. P., Xia, B., Jiang, W., and Inouye, M. (1998) Downstream box: a hidden translational enhancer. *Mol. Microbiol.* 27, 873–874.
- (29) Boni, I. V., Isaeva, D. M., Musychenko, M. L., and Tzareva, N. V. (1991) Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res.* 19, 155–162.
- (30) Caschera, F., Bedau, M. A., Buchanan, A., Cawse, J., de Luca, D., Gazzola, G., Hanczyc, M. M., and Packard, N. H. (2011) Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnol. Bioeng.* 108, 2218–2228.
- (31) de Smit, M. H., and van Duijn, J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.* 87, 7668–7672.
- (32) Kobori, S., Ichihashi, N., Kazuta, Y., Matsuura, T., and Yomo, T. (2012) Kinetic analysis of aptazyme-regulated gene expression in a cell-free translation system: modeling of ligand-dependent and -independent expression. *RNA* 18, 1458–1465.
- (33) Winkler, W. C., and Breaker, R. R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* 59, 487–517.
- (34) R Development Core Team (2011) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.

Acknowledgments

First of all I would like to thank my supervisor Dr. Sheref S. Mansy for giving me the opportunity to work on this project. I learnt a lot from Sheref and the door to his office was always open for help and suggestions.

Then I would like to thank Dr. Andrew D. Ellington for giving me the opportunity to spend time in his laboratory and develop an interesting project. A big acknowledgment is for the Ellington lab group, in particular Adam, John and Jared.

I would also like to thank all members of the Mansy Lab: Amy, Cristina, Paola, Domenica, Fabio and Michele which were my first companions of the PhD. Thanks for all the help and support, for having taught me so much during these years. Thanks to Silvia, for the support and the funny moments we had.

I also thank the “new” members of the lab: Simone, Claudia, Noel, Giuliano, Jason, Ilaria, and Luca. Thanks for always being present, in and out of the lab. Thanks also to Cristiana and Annachiara. A big acknowledgment is for Dario for the patience of having me at the bench close to his, and for Roberta. Thank you Robi for sharing with me the PhD adventure and for being present in the lab, in the radioactive room and in my life in general.

Thanks to my friends Ilaria and Giulia that supported me. Without you I could not have made it through. Thanks to my Austin friends Giulia and Julia, for the fun inside and outside the lab when I was there. Thanks to my housemates Lucia, Chiara, Anna and to my friends Cristina and Francesca for the support at any time of the day.

A big acknowledgment is for my family. My parents and my sister were always there for me and for encouraging me.

Finally, I would like to thank Riccardo. You have been my inspiration behind every idea and the strength to face this PhD. This result is dedicated to you.