



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# **Engineering post-transcriptional regulation of gene expression with RNA-binding proteins**

**Author:**

Roswitha Dolcemascolo

**Thesis advisor:**

Dr. Guillermo Rodrigo Tárrega

**UPV advisor:**

Prof. Carmelo López del Rincón

València, November 2023



To my grandfather Totò

“What I cannot create, I do not understand”

Richard P. Feynman



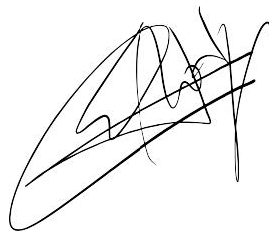
**Dr. Guillermo J. Rodrigo Tárrega**, doctor en Biotecnología por la Universitat Politècnica de València (UPV) y Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC) en el Instituto de Biología Integrativa y de Sistemas (I2SysBio), centro mixto del CSIC y de la Universitat de València Estudi General (UEG),

CERTIFICA:

Que Dña. Roswitha Dolcemascolo, graduada en Biotecnología por la Università degli studi di Palermo, ha realizado bajo su supervisión el trabajo titulado “Engineering post-transcriptional regulation of gene expression with RNA-binding proteins”, que presenta para optar al grado internacional de Doctora en Biotecnología por la Universitat Politècnica de València.

Y para que así conste a los efectos oportunos, firma el presente certificado

Dr. Guillermo Rodrigo Tárrega



En València, Noviembre 2023.



# TABLE OF CONTENTS

<b>ABBREVIATIONS</b>	<b>9</b>
<b>SUMMARY</b>	<b>11</b>
<b>RESUMEN</b>	<b>13</b>
<b>RESUM</b>	<b>15</b>
<b>INTRODUCTION</b>	<b>17</b>
Synthetic Biology	17
The first synthetic genetic devices based on transcription regulation	23
Synthetic biology chassis	26
Design principles and challenges for engineering synthetic biology circuits	30
Post-transcriptional regulation	33
RNA-binding proteins	35
<b>REFERENCES</b>	<b>40</b>
<b>OBJECTIVES</b>	<b>49</b>
<b>CHAPTER 1</b>	<b>51</b>
<b>Gene regulation by a protein translation factor at the single-cell level</b>	<b>51</b>
<b>1. INTRODUCTION</b>	<b>53</b>
<b>2. MATERIALS AND METHODS</b>	<b>55</b>
2.1. Strains, plasmids, and reagents	55
2.2. Growth curves	57
2.3. Flow cytometry	57
2.4. Deterministic mathematical modelling	58
2.5. Stochastic mathematical modelling	59
2.6. Numerical simulations	61
<b>3. RESULTS</b>	<b>62</b>
3.1. Regulation of translation with an RNA-binding protein in single cells	62
3.2. Noise analysis in transcription and translation regulation	66
3.3. Examination of global effects on regulated gene expression	69
3.4. Integrative modeling of the deterministic and stochastic dynamics	73
<b>4. DISCUSSION</b>	<b>76</b>
<b>REFERENCES</b>	<b>79</b>
<b>CHAPTER 2</b>	<b>83</b>
<b>Repurposing the mammalian RNA-binding protein Musashi-1 as an allosteric translation repressor in bacteria</b>	<b>83</b>
<b>1. INTRODUCTION</b>	<b>85</b>

<b>2. MATERIALS AND METHODS</b>	<b>88</b>
2.1. Strains, plasmids, and reagents	88
2.3. Real-time fluorescence quantification in solid medium	91
2.4. Flow cytometry	92
2.5. Purification of a Musashi protein	92
2.6. Binding kinetics assays of protein-RNA interactions	93
2.7. Gel electrophoresis	94
2.8. Microscopy	95
2.9. Mathematical modeling	95
2.10. Molecular visualization <i>in silico</i>	95
<b>3. RESULTS</b>	<b>96</b>
3.1. A Musashi protein can down-regulate translation in bacteria	96
3.2. Mechanistic insight into the engineered regulation based on a protein-RNA interaction	101
3.3. A mathematical model captured the dynamic response of the system	106
3.4. Rational redesign of the targeted transcript to enhance the dynamic range of the response	111
3.5. The regulatory activity of a Musashi protein in bacteria can be externally controlled by a fatty acid	113
<b>4. DISCUSSION</b>	<b>117</b>
<b>REFERENCES</b>	<b>121</b>
<b>GENERAL DISCUSSION</b>	<b>127</b>
<b>REFERENCES</b>	<b>133</b>
<b>CONCLUSIONS</b>	<b>135</b>
<b>ACKNOWLEDGEMENTS</b>	<b>137</b>



## ABBREVIATIONS

**aa:** amino acid

**AD:** activation domain

**AHL:** acyl-homoserine lactone

**AiiA:** AHL-lactonase

**araC:** arabinose regulator protein

**aTc:** anhydrotetracycline

**AU:** arbitrary units

**BIOFAB:** international open facility  
advancing biotechnology

**cI:** phage-encoded  $\lambda$  repressor protein

**CNS:** central nervous system

**CRISPR:** clustered regularly  
interspaced short palindromic repeats

**CSD:** cold-shock domain

**CsrA:** carbon storage regulator

**CV:** coefficient of variation

**DBDs:** DNA-binding domains

**DEAD motif:** Asp-Glu-Ala-Asp  
motif

**DNA:** deoxyribonucleic acid

**dsRBD:** double-stranded RNA-  
binding domain

**eBFP2:** enhanced blue fluorescent  
protein 2

**EDEMP cycle:** Entner-Doudoroff  
Embden-Meyerhof-Parnas cycle

**eIF4G:** eukaryotic translation  
initiation factor

**FadD:** long-chain acyl-CoA  
synthetase

**FadE:** acyl-CoA dehydrogenase

**FadL:** long-chain fatty acid transport  
protein

**GO:** Gene Ontology

**GRAS:** generally recognized as safe

**Hfq:** host factor I protein

**hnRNPs:** heterogeneous nuclear  
ribonucleoproteins

**iGEM:** international genetically  
engineered machine

**IPTG:** isopropyl- $\beta$ -D-  
thiogalactopyranoside

**K<sub>D</sub>:** dissociation constant

**k<sub>ON</sub>:** protein-RNA association rate  
constant

**k<sub>OFF</sub>:** protein-RNA dissociation rate  
constant

**LacI:** lactose repressor protein

**LB:** Luria Bertani broth

**LuxI:** *N*-acyl-1-homoserine lactone  
synthase

**LuxR** transcriptional factor for  
quorum-sensing control of  
luminescence

**KH:** homology domain

**MIT:** Massachusetts Institute of  
Technology

**m-*numb*:** mammalian *numb*

**mRNA:** messenger RNA

**miRNA:** micro RNA

**MSI-1:** Musashi-1 protein

**MSI-1\*:** synthetic Musashi-1 protein

**MS2CP:** MS2 coat protein

**M9:** minimal medium

**NES:** nuclear export signal

**NLS:** nuclear localization signal

**OB:** oligonucleotide/oligosaccharide binding family

**ORFs:** open reading frames

**PABP:** poly(A)-binding protein

**PCR:** polymerase chain reaction

**Pfam:** protein family (database)

**PLlac:** PL-based promoter repressed by LacI

**PTMs:** post-translational modifications

**RBPs:** RNA binding proteins

**RBS:** ribosome binding site

**RNA:** ribonucleic acid

**r-proteins:** ribosomal proteins

**RRM:** RNA recognition motif

**SELEX:** systematic evolution of ligands by exponential enrichment

**sfGFP:** superfolder green fluorescent protein

**SOPs:** sensory organ precursors cells

**crRNA:** cis-repressed RNA

**ssRBDs:** single-stranded RNA structure predictions

**sRNAs:** small RNAs

**tRNAs:** transfer RNAs

**taRNA:** trans-activating RNA

**TALEs:** transcription activator-like effectors

**TC:** tetracycline

**TetR:** tetracycline repressor protein

**TFs:** transcription factors

**TFBs:** transcription factor binding sites

**TX-TL:** transcription-translation cell-free expression system

**UTRs:** untranslated regions

**UV:** ultraviolet

**yemGFP:** monomeric yeast-enhanced green fluorescent protein

# SUMMARY

Synthetic biology seeks to design and construct new biological systems with desired functions. Circuits based on transcriptional control have been preponderant in the field following the pioneering work of the toggle switch and repressilator. However, to advance the creation of transformative technologies using synthetic genetic circuits, a blend of dependable control mechanisms throughout the genetic information flow is essential. This combination is necessary to attain the level of integrability and functional complexity observed in nature. In this regard, circuits based on post-transcriptional regulation have recently gained attention. In particular, the great programmability of RNA has been exploited to create regulatory circuits for biosensing of environmental signals or for controlling metabolic pathway in bioproduction. In this thesis, in contrast, we propose to exploit RNA-binding proteins to engineer synthetic circuits that operate at the level of translation in the bacterium *Escherichia coli*. This thesis intends to study how noise emerges and propagates when gene expression is regulated by a translation factor, and the expansion of the synthetic biology toolbox with new characterization of suitable RNA-binding proteins.

On the one hand, we engineered a post-transcriptional control circuit using the phage MS2 coat protein. Through meticulous single-cell level monitoring of both the regulator and the regulated gene, we quantified the dynamic behavior of the system, as well as their stochasticity. While previous efforts focused on understanding noise propagation in transcriptional regulations, the stochastic behavior of genes regulated at the translation level remain largely unknown. Our data revealed that a protein translation factor enabled strong repression at the single-cell level, buffered noise propagation from gene to gene, and led to a nonlinear sensitivity to global perturbations in translation. These findings significantly enhanced our

understanding of stochastic gene expression and provided foundational design principles for synthetic biology applications.

On the other hand, we harnessed the RNA recognition motif (RRM), the most prevalent RNA-binding domain in nature, despite its predominance in eukaryotic phyla, to engineer an orthogonal post-transcriptional control system in *Escherichia coli*. Leveraging the mammalian RNA-binding protein Musashi-1, which contains two canonical RRM, we developed a sophisticated circuit. Musashi-1 functioned as an allosteric translation repressor through its specific interactions with the N-terminal coding region of messenger RNA, exhibiting responsiveness to fatty acids. Comprehensive characterization at both population and single-cell levels highlighted a significant fold change in reporter expression. Molecular insights were gleaned through in vitro binding kinetics and in vivo functionality assessments of a series of RNA mutants. This work showcased the adaptability of RRM-based regulation to simpler organisms, introducing a novel regulatory layer for translation control in prokaryotes, ultimately expanding the horizons of genetic manipulation.

## RESUMEN

La biología sintética tiene como objetivo diseñar y construir nuevos sistemas biológicos con funciones deseadas. Los circuitos basados en el control transcripcional han tenido preponderancia en este campo tras el trabajo pionero del toggle switch y del repressilator. Sin embargo, para avanzar en la creación de tecnologías transformadoras que utilicen circuitos genéticos sintéticos, es esencial una combinación de mecanismos de control confiables en todo el flujo de la información genética. Esta combinación es necesaria para alcanzar el nivel de integrabilidad y complejidad funcional observado en la naturaleza. En tal sentido, recientemente han ganado atención los circuitos basados en regulación postranscripcional. En particular, se ha aprovechado la gran programabilidad de ARN para crear circuitos reguladores para la biodetección de señales ambientales o para controlar la vía metabólica en la bioproducción. En esta tesis, por el contrario, proponemos explotar las proteínas de unión a ARN para diseñar circuitos sintéticos que operen a nivel de traducción en la bacteria *Escherichia coli*. Esta tesis pretende estudiar como surge y se propaga el ruido cuando la expresión genética está regulada por un factor de traducción, y la ampliación de la caja de herramientas de la biología sintética con una nueva caracterización de proteínas de unión a ARN adecuadas.

Por un lado, hemos diseñado un circuito de control postranscripcional utilizando la proteína de cápside del fago MS2. Mediante una meticulosa monitorización a nivel unicelular tanto del regulador como del gen regulado, hemos cuantificado el comportamiento dinámico del sistema, así como su estocasticidad. Si bien los esfuerzos anteriores se centraron en comprender la propagación del ruido en las regulaciones transcripcionales, el comportamiento estocástico de los genes regulados a nivel de la traducción sigue siendo en gran medida desconocido. Nuestros datos han revelado que un factor de traducción de proteínas ha permitido una fuerte represión a nivel unicelular, ha amortiguado la propagación del ruido de

un gen a otro y ha conducido a una sensibilidad no lineal a las perturbaciones globales en la traducción. Estos descubrimientos han mejorado significativamente nuestra comprensión de la expresión genética estocástica y han proporcionado principios de diseño fundamentales para aplicaciones de biología sintéticas.

Por otro lado, aprovechamos el motivo de reconocimiento de ARN (RRM), el dominio proteico de unión a ARN más prevalente en la naturaleza, a pesar de su predominio en los filos eucariotas, para diseñar un sistema de control postranscripcional ortogonal en *Escherichia coli*. Aprovechando la proteína de unión a ARN de mamífero Musashi-1, que contiene dos RRM canónicos, desarrollamos un circuito sofisticado. Musashi-1 ha funcionado como represor de la traducción alostérico a través de su interacción específica con la región codificante N-terminal del ARN mensajero, mostrando capacidad de respuesta a los ácidos grasos. La caracterización integral tanto a nivel poblacional como unicelular ha destacado un cambio significativo en la expresión del reportero. Se obtuvieron conocimientos moleculares a través de la cinética de unión *in vitro* y evaluaciones de funcionalidad *in vivo* de una serie de mutantes de ARN. Este trabajo ha mostrado la adaptabilidad de la regulación basada en RRM a organismos más simples, introduciendo una nueva capa regulatoria para el control de la traducción en procariotas y, en última instancia, ampliando los horizontes de la manipulación genética.

# RESUM

La biologia sintètica té per objectiu dissenyar i construir nous sistemes biològics amb funcions desitjades. Els circuits basats en el control transcripcional han tingut preponderància en aquest camp després del treball pioner del toggle switch i del repressilator. Tot i això, per avançar en la creació de tecnologies transformades que utilitzin circuits genètics sintètics, és essencial una combinació de mecanismes de control fiables en tot el flux de la informació genètica. Aquesta combinació és necessària per assolir el nivell d'integrabilitat i complexitat funcional observat a la natura. En aquest sentit, recentement han guanyat atenció els circuits basats en regulació posttranscripcional. En particular, s'ha aprofitat la gran programabilitat d'ARN per crear circuits reguladors per a la biodetecció de senyals ambientals o per controlar la via metabòlica a la bioproducció. En aquesta tesi, per contra, proposem explotar les proteïnes d'unió a ARN per dissenyar circuits sintètics que operin a nivell de traducció al bacteri *Escherichia coli*. Aquesta tesi pretén estudiar com sorgeix i es propaga el soroll quan l'expressió genètica està regulada per un factor de traducció, il·luminant la caixa d'eines de la biologia sintètica amb una nova caracteriació de proteïnes d'unió a ARN adequades.

D'una banda, hem dissenyat un circuit de control postranscripcional utilitzant la proteïna de càpsid del fag MS2. Mitjançant una meticulosa monitorització a nivell unicel·lular tant del regulador com del gen regulat, hem quantificat el comportament dinàmic del sistema, així com la seva estocasticitat. Tot i que els esforços anteriors es van centrar a comprendre la propagació del soroll en les regulacions transcripcionals, el comportament estocàstic dels gens regulats a nivell de la traducció continua sent en gran mesura desonegut. Les nostres dades han revelat que un factor de traducció de proteïnes ha permès una forta repressió a nivell unicel·lular, ha esmorteït la propagació del soroll d'un gen a un altre i ha conduït a una sensibilitat no lineal a les pertorbacions globals a la traducció. Aquests descobriments han

millorat significativament la nostra comprensió de l'expressió genètica estocàstica i han proporcionat principis de disseny fonamentals per a aplicacions de biologia sintètiques.

D'altra banda, aprofitem el motiu de reconeixement d'ARN (RRM), el domini proteic d'unió a ARN més prevalent a la natura, malgrat el seu predomini als talls eucariotes, per dissenyar un sistema de control posttranscripcional ortogonal a *Escherichia coli*. Aprofitant la proteïna d'unió a ARN de mamífers Musashi-1, que conté dos RRM canònics, hem desenvolupat un circuit sofisticat. Musashi-1 va funcionar com un repressor de la traducció al·lostèric a través de la seva interacció específica amb la regió codificant N-terminal de l'ARN missatger, mostrant capacitat de resposta als àcids grassos. La caracterització integral tant a nivell poblacional com unicèl·lular va destacar un canvi significatiu a l'expressió de l'informador. S'obtingueren coneixements moleculars a través de la cinètica d'unió *in vitro* i avaluacions de funcionalitat *in vivo* d'una sèrie de mutants d'ARN. Aquest treball va mostrar l'adaptabilitat de la regulació basada en RRM a organismes més simples, introduint una nova capa regulatòria per al control de la traducció en procariotes i, en darrer terme, ampliant els horitzons de la manipulació genètica.



# INTRODUCTION

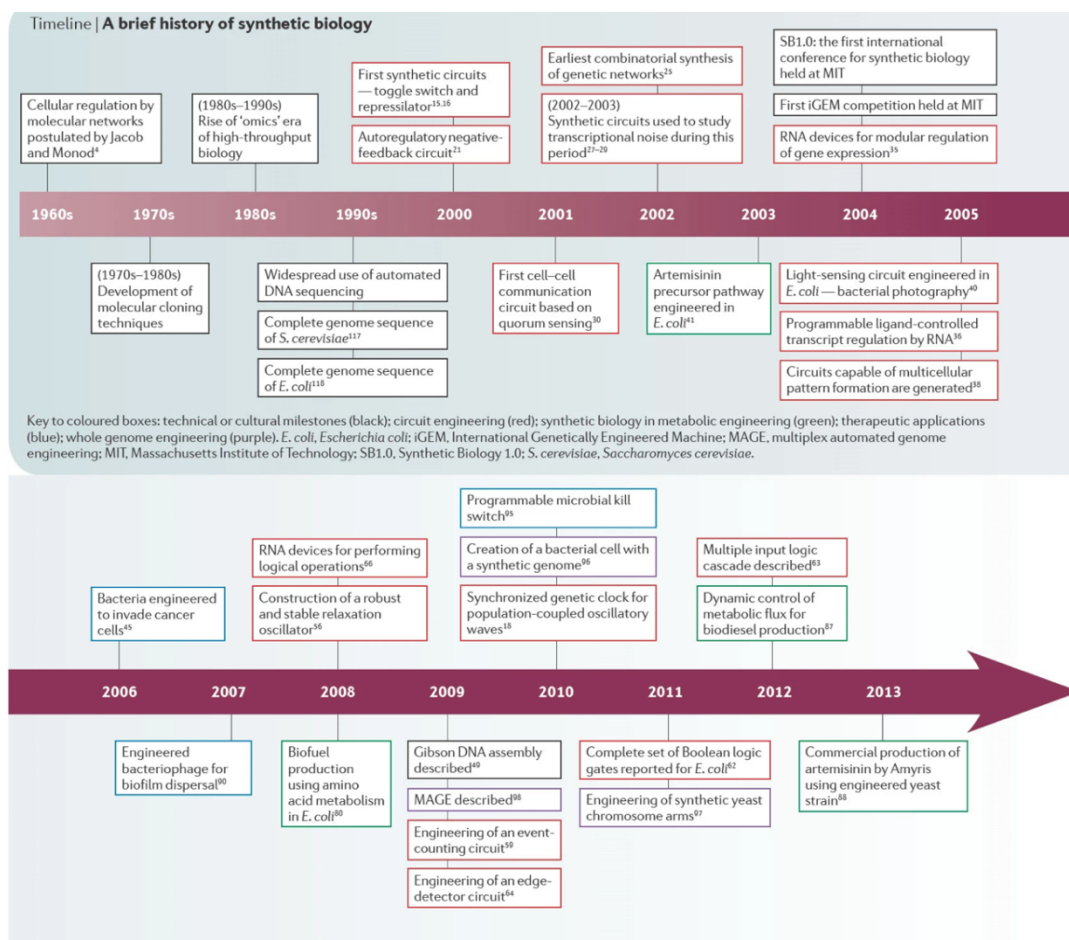
## Synthetic Biology

Synthetic biology is an emerging area of research that combines the investigative nature of biology with the constructive nature of engineering in order to study living organisms [1]. Similar descriptions have been offered by other commissions and studies. For example, a joint opinion by three scientific committees of the European Commission emphasizes the role of design and engineering approaches [2]. A report of the Secretariat of the Convention on Biological Diversity suggests that while there is no agreed international definition, the key features of synthetic biology include “the novo synthesis of genetic material and an engineering-based approach to develop components, organisms and products” [3].

Although a consensus has yet to be reached on a precise definition of synthetic biology, many disciplines (physics, chemistry, mathematics, engineering, and computer science) are applied simultaneously for a useful purpose. As a result, the use of molecular biology tools and techniques to engineer cellular behavior has emerged as a broad identity for the field, and a set of engineering approaches and laboratory practices have developed, along with a community culture. Much of the foundational work in the field has been carried out in *Escherichia coli* as a model organism. This microbial system remains central in several focal areas of the field, including complex circuit design, metabolic engineering, minimal genome construction and cell-based therapeutic strategies.

Modern synthetic biology started with the creation of simple artificial gene regulatory circuits in living cells that carry out functions in an analogous manner to electrical circuits [4, 5]. Notably, the crucial events were the construction of the first toggle switch by Collins and colleagues [6], and the construction of the first genetic oscillator by Elowitz and Leibler [7] in *E. coli* (Fig. 1).

The main goal of synthetic biology is to gain a better understanding of how to regulate pathways and gene expression in living organisms. To date, genetic tools must be developed to obtain a desired cell function with the synthetic regulatory network. Regulatory networks constructed by synthetic biology approach are used to either investigate natural biological phenomena or to furnish cells with a new function. In bacteria, this has led to a variety of circuit designs implemented at the transcriptional and post-transcriptional levels both in single cell and in a population of cells. Many genetic elements, such as promoter, transcription factor binding sites (TFBs), ribosome binding sites (RBSs), terminators, and DNA vectors, can be synthetically combined in the expression systems to customize the desired regulation.



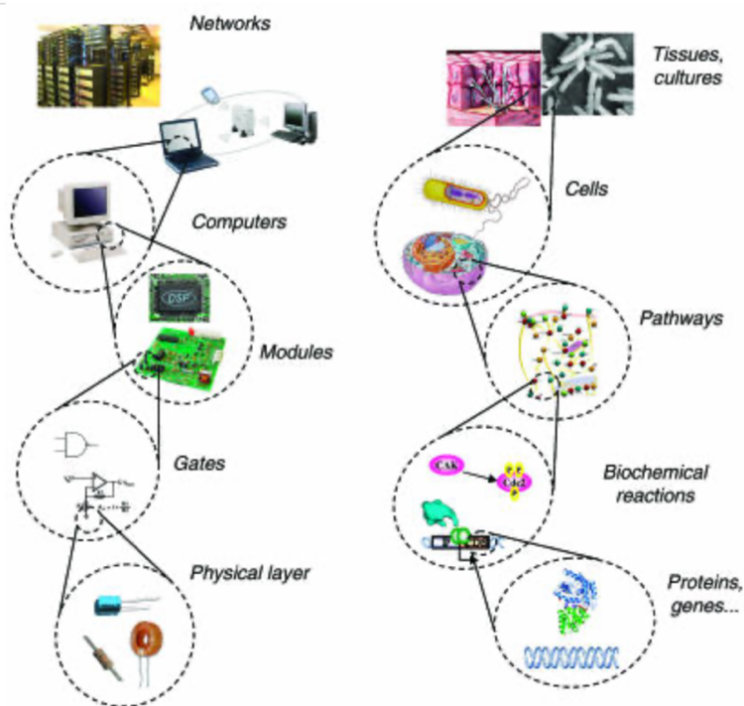
**Fig. 1** Timeline of the history of synthetic biology from [11].

The roots of synthetic biology can be traced to the seminal work of Francois Jacob and Jacques Monod in 1961 [8]. For the first time, a gene circuit consisting of several functional blocks capable of responding to environmental factors was shown in bacterial cell. Insights from their studies of the *lac* operon in *E. coli* led them to assume the existence of regulatory circuits that establish the response of a cell to its environment. According to the landmark publication of Monod and Jacob, fundamental cellular processes as differentiation and protein regulation are accomplished through signaling pathway resident at level of a gene. Therefore, the ability to assemble new regulatory systems from molecular elements was soon envisioned [9] but a more concrete vision based on programmed gene expression began to take shape only when the molecular details of transcriptional regulation in bacteria were uncovered [10].

The period from 1961 to 2000 was characterized by the accumulation of knowledge in the field of molecular biology. In the 1970s and 1980s, the development of molecular cloning and PCR techniques for the genetic manipulation offered a technical means to engineer artificial gene regulation. In the 1990s, biologists and computer scientists began to combine experimentation and computation to reverse-engineer cellular networks. This 'scaling-up' of molecular biology generated the field of systems biology. Such an approach could be used to study the functional organization of natural systems and to create artificial regulatory networks [11].

In the time immediately after the publication of the toggle switch and repressilator several studies focused on circuit engineering in order to correlate network architecture to quantitative behavior [12]. Despite being primarily focused on circuit engineering, research at this early stage started to expand gene regulatory networks. The development of the first cell-cell communication circuits portended a shift in future years toward artificial microbial consortia [13]. The first attempts to control of the behavior rewiring post-translational regulation utilizing scaffold

proteins and protein-protein interaction domains were also shown in *Saccharomyces cerevisiae* [14].

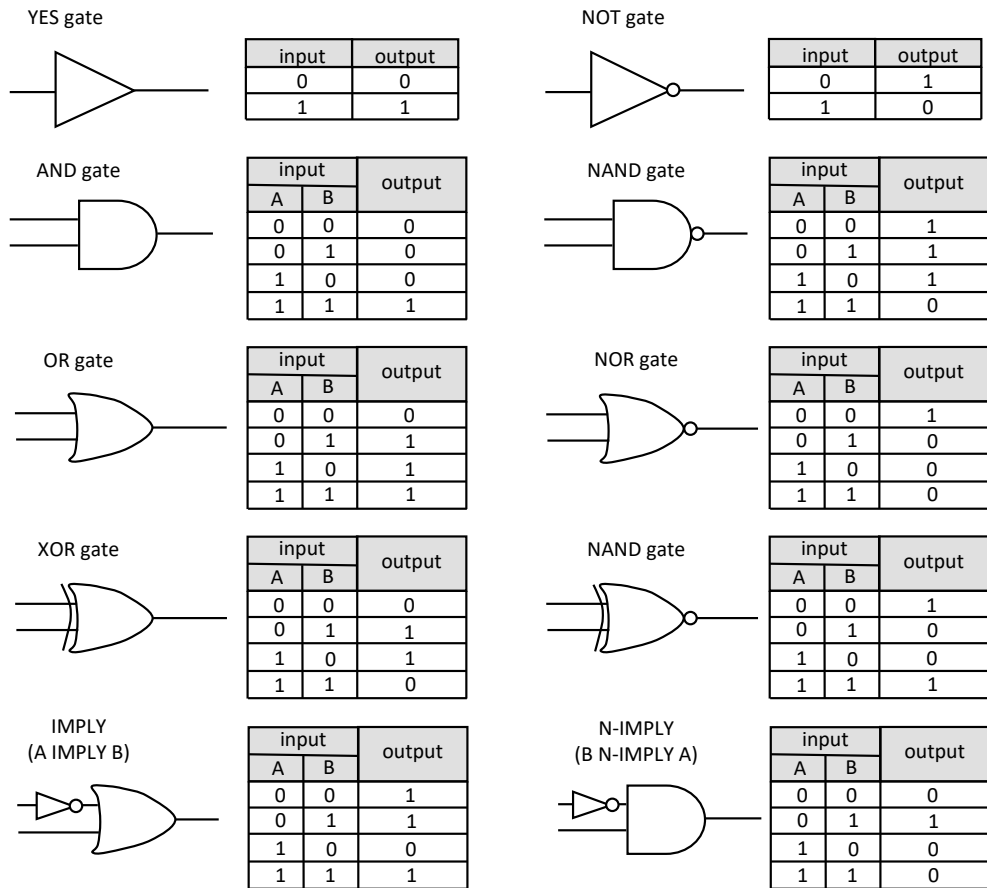


**Fig. 2** Analogy between computer science and synthetic biology from [19].

In the middle of the 2000s, the synthetic biology field significantly grew. Synthetic Biology 1.0 (SB1.0), the first global conference in the field, took place in the summer of 2004 at the Massachusetts Institute of Technology (MIT, US). It helped to establish an identifiable community and inspire work on the design, construction, and characterization of biological systems with the long-term of whole genomic engineering [15,16]. Questions concerning the compatibility of the two areas of science raised when the highly multidisciplinary community began to work together and concepts from modern engineering were incorporated into the traditional molecular biology. Could synthetic biology evolve into a sophisticated engineering-discipline, comparable to electrical or mechanical engineering? Could practices like parts standardization and concepts like abstraction hierarchies be mapped onto biological systems? To answer these questions, a number of synthetic

parts including promoters [17], regulatory proteins and RNA (i.e., riboregulators and riboswitches) [18], have been successfully engineered and characterized in a number of hosts including bacteria.

As well as generating insights into natural processes, synthetic biology aims to obtain a specific behaviour in living organisms including drug production and delivery, bioengineering, bioremediation, biosensing, and biofuel production. For many, the goal and method of synthetic biology can be conceptualized by the hierarchical structure of computer engineering (Fig. 2) [19]. Within the metaphor, the simplest components at the bottom of the computer architecture are transistors and resistors that are connected to form devices such as Boolean logic gates. The way cells process information can be compared to digital control system. Such systems are based on logic gates, which define a response (output) by combining two inputs. Each input or output can determine the state of the synthetic circuit that can be “on” or “off” giving the symbol of “1” or “0” used in the binary code. A variety of logic gates are obtained with two inputs. Boolean logic functions include AND, NOR, OR, XOR gates, and all their possible combinations. In these gates, inputs are read out by a sensor. Then, a computational core sets them a value of “0” or “1”. If the combination of these values satisfies the system requirements, the output is then executed. Each gate can be defined by symbols and a truth table (Fig. 3). Applying to a living cell, thresholds of input and output values must be precisely defined. In synthetic biology, these components can be represented by basic elements such as genes and regulatory proteins (promoters and repressors) that interact in biochemical reactions. At the beginning of the synthetic biology era, these devices were pieced together to form modules to achieve specific tasks, such as switching and oscillation [10]. Recently, modules were brought together to form larger scale systems (i.e., networks) [20]. This set of modules that are restricted to a single cell, or distributed over a number of cells, interact each other to produce a coherent behavior. Throughout the information transmission, functional spatiotemporal patterns can be achieved for a purpose.



**Fig. 3** Synthetic circuit behaviors based on Boolean logic gates.

As a result of increasing efforts in the characterization and development of biological parts used for the assembly of synthetic systems, the number of available parts expanded with the need to catalogue them. First catalogue started by the iGEM registry (Registry for Standard Biological Parts) with 20,000 parts listed. Subsequently, many other registers have been established (i.e., the International Open Facility Advancing Biotechnology -BIOFAB, the BioBrick Foundation -BBF, and the Joint BioEnergy Institute Inventory of Composable Elements -JBEI-ICEs) and commercially available to scientists and general public for free.

## The first synthetic genetic devices based on transcription regulation

The first synthetic genetic oscillator, called the repressilator, was designed by Elowitz and Leibler to implement a particular function in *E. coli* [7]. The repressilator uses a simple design, resembling a game of rock, paper, scissors. The key components are repressor proteins, which bind to specific DNA sequences to inhibit gene expression (Fig. 3a). The three transcriptional repressor systems are made by LacI from *E. coli* that inhibits the transcription of a second repressor gene, *tetR*, from the tetracycline-resistance transposon Tn10, whose protein product in turn inhibits the expression of a third gene, *cI* from  $\lambda$  phage. Finally, CI inhibits LacI expression. Thus, three repressor factors are configured to display a cycle where each one represses the expression of the next one. In addition, one of the repressors inhibits the expression of a gene encoding a green fluorescent reporter. Such configuration results in a negative feedback loop that can lead to temporal oscillation in the concentrations of each of its components and the behavior of the loop can be designed by a simple transcriptional regulation. In other words, with this configuration an increase in concentration of one repressor protein causes a decrease in the second, leading to an increase in the third, thereby decreasing the first.

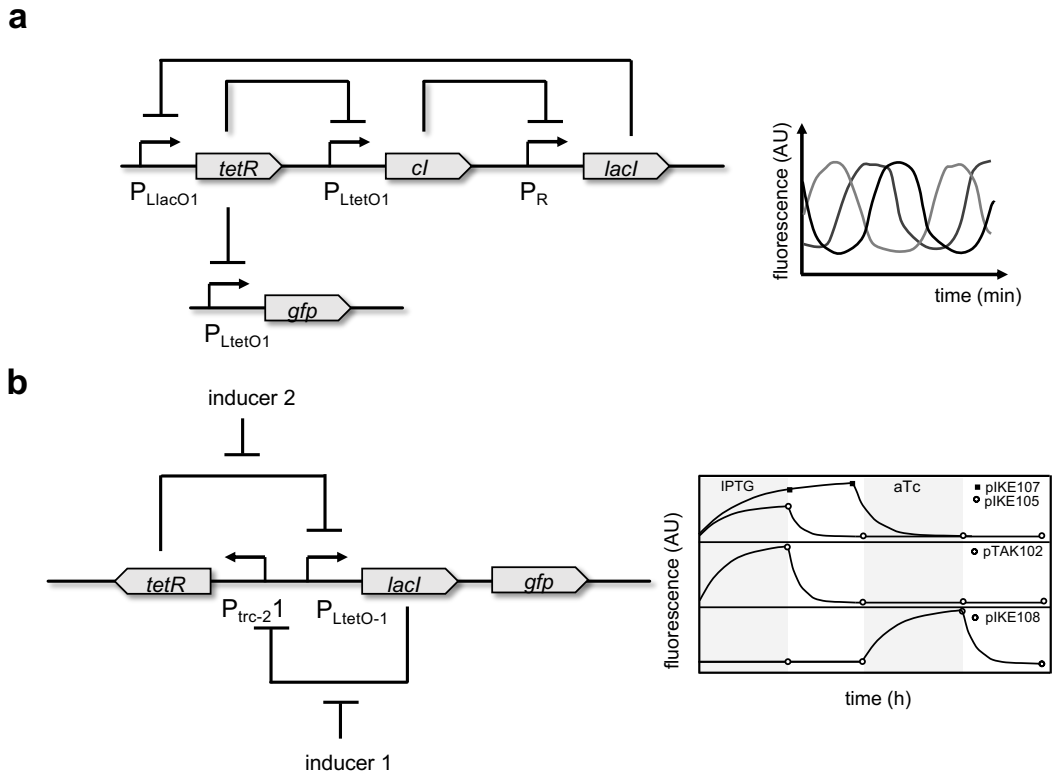
Since the design of the first repressilator, a wide variety of synthetic oscillators have been constructed. Those synthetic networks incorporated alternative designs, coupling between cells and other features. For instance, a dual-feedback oscillator was engineered with tunable oscillatory periods (13 min). Based on previous design, the *E. coli* components were used to design a fast, robust and persistent genetic oscillator. A hybrid promoter  $p_{lac/ara-1}$  controls the expression of *araC*, *lacI* and *yemGFP* to form three co-regulated transcription modules [21]. Another example is a genetic clock engineered with global intercellular coupling capable of generating synchronized oscillations in growing population of cell. Such synchronized oscillator is based on elements of the quorum sensing machineries in *Vibrio fischeri* (*luxI*) and *Bacillus thuringiensis* (*aitA*), and *yemGFP*, under control of

three identical copies of the *luxI* promoter [22]. The network architecture is similar to the motif used in previous synthetic oscillator designs. An activator its own repressor. In this case, LuxI synthase enzymatically produces an acyl-homoserine lactone (AHL) which is a small molecule that can diffuse across the membrane and binds intracellularly to LuxR (constitutively produced). The LuxR-AHL complex forms a transcriptional activator for the *luxI* promoter. By contrast, the promoter is negatively regulated by AiiA by catalyzing the degradation of AHL. Moreover, a genetic network was engineered in cell-free system prepared from an *E. coli* extract called TX-TL [23] accurately reproducing the same pattern of gene expression in live *E. coli* cells [24].

The toggle switch is another engineering approach to the study of gene expression. The synthetic circuit constructed by Collins and colleagues, contains two promoters that drive the expression of mutually inhibitory transcriptional repressors (*lacI* and *cI*) resulting in a bistable genetic circuit in which only one of the two genes is active at a given time (Fig. 3b) [6]. As a result, each promoter is inhibited by the repressor that is transcribed by the opposing promoter. The toggle does not require any specialized promoters. However, in the pivotal study of Collins and colleagues two classes of toggle switch plasmids were constructed (termed pTAK and pIKE). Both classes use the transcriptional Lac repressor (*lacI*) in conjunction with the P<sub>trc-2</sub> promoter for one promoter-repressor pair. This class of plasmid was switched between states using isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). For the second promoter-repressor pair, a temperature-sensitive  $\lambda$  repressor (*cIts*) was used in conjunction with the P<sub>Ls1con</sub> promoter. This class of plasmids was switched between states by IPTG or a thermal pulse. Alternatively, the Tet repressor (tetR) was used in conjunction with the P<sub>LtetO-1</sub> promoter. This class of plasmid was switched between states using IPTG or a pulse of anhydrotetracycline (aTc). Indeed, the toggle exhibits bistability due to the mutually inhibitory arrangement of the repressor genes. In absence of inducers, two stable states are possible: one in which promoter 1 transcribes repressor 2, and one in which promoter 2 transcribes repressor 1.



Switching is accomplished by transiently introducing an inducer of the currently active repressor. The inducer permits the opposing repressor to be maximally transcribed until it stably represses the originally active promoter.



**Fig. 4 a)** Repressilator regulatory system. The system is made of three genes connected in a feedback loop. On the left, the synthetic circuit constructed by Elowitz and Leibler; on the right, repressilator dynamics [7]. **b)** On the left, toggle switch design; on the right, bistability behaviour [6].

The successful design of the first synthetic networks showed that a computing-like behavior can be used into biological systems. In both repressilator and toggle switch, basic transcriptional regulatory elements were assembled to realize the biological equivalents of electronic memory storage and timekeeping.

## **Synthetic biology chassis**

Synthetic biology was originally built using a small set of organisms that were highly adapted to laboratory conditions. In classical engineering terms, a chassis is defined as the load-bearing framework of an object supporting the weight and structure of the object for its correct function. The term can evoke the basic frame of a car to which a number of components in response to specifications or customer's desires. This concept started to be used in the early 2000s by the synthetic biology community referring to a platform that acts as a framework and support for biological parts, and it lies between frontline bioengineering and the traditional recombinant DNA technology [25,26]. A chassis contains at least four elements: the physical container of the genetic constructs, the genomic skeleton, the cellular machinery necessary for the gene expression, and the spatial scaffold for the genetic and metabolic reactions. Therefore, a live organism qualifies as the chassis serving as a foundation to physically house genetic components and support them by providing the resource to function. Any implanted components of the chassis are orthogonal and then programmable.

The choice of a suitable chassis organism is crucial in synthetic biology, as it determines the available molecular machinery, cellular processes, and physiological characteristics that can be harnessed for engineering purposes. Notably, the choice of a chassis organism depends on the specific application and desired outcomes. Researchers and engineers are continuously exploring and expanding the range of potential chassis organisms, including non-traditional candidates like cyanobacteria, mammalian cells, and even synthetic organisms engineered from scratch.

Ideally, a synthetic biology chassis should possess several key characteristics:

i. Genetic tractability. The chassis organism should have a well-understood and easily manipulable genome. It should allow for efficient genetic engineering, such as DNA insertion, deletion, or modification, to facilitate the introduction of desired genetic circuits or pathways.

- ii. Stability and compatibility. The chassis organism should maintain genetic stability and compatibility with the engineered genetic constructs over multiple generations. This ensures that the desired traits or functions are reliably passed on and maintained over time.
- iii. Robust growth and proliferation. A chassis organism should exhibit fast growth and reproduction rates, enabling rapid production of engineered systems. This characteristic is particularly important when scaling up synthetic biology processes for industrial or commercial applications.
- iv. Modularity. A modular chassis allows for the incorporation of different genetic modules or parts without significant interference or crosstalk. This modular design facilitates the construction of complex genetic circuits and systems, allowing for greater control and predictability.
- v. Safety and biocontainment. Chassis organisms should be engineered with appropriate safety measures to prevent unintended release into the environment or potential harm to other organisms. This includes containment mechanisms to restrict the survival or spread of the chassis outside of controlled laboratory conditions.

Due to its largest available toolkit of computational design programs, genetic parts and regulatory elements, *E. coli* is undoubtedly the most used chassis in synthetic biology. *E. coli* is the laboratory workhorse, it is widely exploited as a cell factory for microbial production of pharmaceuticals [27,28], enzymes, and biofuel (including alcohol-, fatty acid-, terpenoid- based biofuel [29,30]). Indeed, *E. coli* has played a pivotal role in the development of genetics and molecular biology field of science. It is therefore not surprising that synthetic biology has been created upon it and *E. coli* is the protagonist in the field. However, scientific breakthroughs have been achieved from simple gene mutations to the insertion of rationally designed, complex synthetic circuits and creation of synthetic genomes. Indeed, a set of

functionalities are unavailable in this system (epigenetic control, post-translational protein modifications).

Although the versatile microbial chassis of *E. coli* continues to dominate the field, this bacterium may no longer be the adequate host for the sophisticated genetic design of synthetic biology, and other set of chassis should be required to cover a range of applications. To date, novel synthetic biology toolboxes have grown rapidly in the last decade including *Bacillus subtilis*, *Pseudomonas putida*, *Streptomyces sp*, and *S. cerevisiae*. These organisms possess many desirable traits, such as well-characterized genomes, ease of genetic manipulation, and established tools and techniques for engineering. Among these, *B. subtilis* is a Gram-positive spore-forming member of the phylum Firmicutes that was discovered during the second world war to treat dysentery. It is commonly used for studies in genetics, cellular metabolism, and cell factory for secondary metabolite production (polyketides, non-ribosomal peptides) for industrial application and ability to survive under harsh environments [31,32,33,34]. *B. subtilis* is an attractive chassis for the assembly of heterologous DNA parts in a genome scale and cloning system [35], for biosensors development [36,37]. The soil-dwelling *Pseudomonas putida* is a ubiquitous rhizosphere bacterium certified as “generally recognized as safe” (GRAS). Because of the high tolerance to organic solvents, redox stress, capability to degrade aromatic compounds, and ease of genetic manipulation, *P. putida* has emerged as a platform for industrial processes (biotransformation) [38,39]. Due to its central carbon metabolic network that involves the recycling of triose phosphates to hexose phosphates (called EDMP cycle) and versatile metabolism, the soil-dwelling bacterium *P. putida* is an optimal microbial cell factory for metabolic engineering applications [40]. The high tolerance to oxidative stress makes this bacterium suitable for several technical applications, such as biofuel production [41]. Physiological robustness, metabolic versatility, and high tolerance to stress are key features of this specie. Due to these properties, it is a well-established host for cloning and gene expression. Stable cloning has been demonstrated by the use of

Tn5-derived transposon system for DNA integration into the genome [42]. Moreover, genetic tools have been developed to tune gene expression using genome integration and reporter system and identifying synthetic promoters from a library with different expression levels [43]. The Gram-positive, soil-dwelling *Streptomyces* species belong to the *Actinobacteria* family. *Streptomyces* species have a diverse secondary metabolism producing bioactive compounds such as antibiotics, antifungals, and anticancer [44,45,46,47]. Because of the natural genomic information, *Streptomyces sp* have a great potential to produce novel secondary metabolites.

Nevertheless, there are some issues about the physiological and metabolic properties of bacterial chassis for constructing synthetic reliable platforms. Bacteria lack of membrane-boundaries in their cytoplasm displaying an open architecture where proteins and nucleic acids organize to catalyze the proper biochemical reactions [48]. Thus, the spatial organization of the bacterial metabolism with the link between the DNA organization, the location of the products and the interaction between the enzymes in a pathway is a challenge [47]. Indeed, over time proteins aggregate and or can be cleaved by proteases in the process of cell ageing [49]; cells can adopt the shape of biofilm where the biochemical properties of individual cell are boosted [50]. For those reason, synthetic biology is nowadays extended to eukaryotic organisms. The budding yeast *S. cerevisiae* is the optimal eukaryotic organism for metabolic engineering platform [51]. *S. cerevisiae* is the first experimental model to study eukaryotic biological system as an ideal chassis for biochemical production due to its ability to survive at low temperature, pH changes, and phage attack [52]. The *S. cerevisiae* toolkit developed involved the construction and implementation of synthetic promoters [53].

## **Design principles and challenges for engineering synthetic biology circuits**

Design principles play a crucial role in engineering circuits in the field of synthetic biology. In this context, interconnected networks of genetic elements that control the flow of information and regulate the behavior of biological systems are created. Indeed, the design of synthetic biology circuits with desired features from gene components is a non-trivial process. Compared to the standard engineering disciplines, it needs the design, construction, testing and verification of the process making it not trivial and often unpredictable. Therefore, the following approaches are described to lead the engineering of synthetic circuits.

- i. Rational design. A synthetic system is experimentally designed, and a computational model is developed for the analysis of the system behavior. A fine-tuning of the engineered system is performing by the model until the desired behavior is achieved. The repressilator [7] and the toggle switch [6] are an example of the application of this principle. Notably, many efforts using a trial-and-error method are needed to obtain the desired circuitry due to the model parameters and the resulting effect on the behavior.
- ii. Direct evolution. Random mutagenesis obtained by PCR and screening of the desired mutants is used for this method. For instance, an original genetic circuit was evolved to the functional one by mutating the coding sequence of a gene and its RBS, and the transcriptional factor LuxR to improve the quorum sensing signal.
- iii. Combinatorial synthesis. Circuit variants with parts in predefined different combinations and configurations are constructed, and the functional ones with the desired behaviour are then selected. This combinatorial method was first shown in a study of random shuffling the connectivity of three transcriptional factors (TetR, LacI, CI) and their corresponding promoters. The resulting synthetic networks displayed phenotypic behavior of NAND, NOR, and NOT logic gates.
- iv. Hybrid approaches obtained by combining the three methods mentioned above.

Engineering gene circuit is a complex and challenging area of Synthetic Biology that involves designing biological systems to perform specific functions. These circuits can be used in a variety of applications, from medicine to biotechnology. However, there are several challenges related to engineering gene circuit including orthogonality, and modularity achievement, assembly of modules, noise minimization.

i.Modularity. It involves the decomposition of complex systems into smaller, independent functional modules. Each module performs a specific task and can be combined with other modules to achieve more complex behaviors. Modularity allows for easy assembly, modification, and testing of different circuit components, enhancing flexibility and scalability. Therefore, genetic parts can be combined and modified to create new functions. However, most of the parts and modules used for building synthetic genetic systems are not standardized and quantitatively characterized lacking modularity and reusability. Indeed, the behavior of a module characterized in one context can change in new conditions. Therefore, new methods for characterization need to be developed to reduce such context dependency and increase its reusability.

ii.Orthogonality. It refers to the independence of different components of a circuit. The newly added synthetic circuits should do not interfere with the existing ones in the designed genetic system as well as the genetic background of the host. In synthetic biology, orthogonal components allow for the modular assembly of circuits without cross-talk or unwanted interactions. Orthogonal genetic parts ensure predictable and reliable circuit behavior, enabling the design of complex circuits with minimal interference. Although synthetic systems are supposed to function similarly to their electronic counterparts, one difference is that, unlike electronic digital circuits, the individual components of biological devices are not connected by wires. The interactions between biological parts depends on the chemical specificity between them. The current synthetic biology toolkit contains only a small repertoire of orthogonal regulatory elements, such as LacI, TetR, and

CI regulatory proteins and their promoters. Therefore, there is the need to expand the number of currently available orthogonal modules.

iii. Functional assembly of modules. The current design principles (rational design, evolution, combinatorial synthesis) have been shown to work in designing small scale biological systems and lots of trial-and-error efforts.

iv. Standardization. Standardization of genetic parts and components is essential for effective circuit design in synthetic biology. Standardized parts, such as promoters, terminators, and coding sequences, have well-defined properties and are characterized and documented. This standardization enables the predictable behavior and easy exchangeability of genetic elements, simplifying circuit design and facilitating collaboration among researchers.

v. Feedback Control. It is a key principle used to regulate and maintain desired circuit behavior. For instance, negative feedback loops can be incorporated into circuits to sense changes in system parameters and adjust the output accordingly. This principle helps maintain stability, robustness, and homeostasis within the biological system, allowing circuits to respond and adapt to dynamic environmental conditions.

vi. Noise. Another challenge in engineering gene circuits is the variation problem, which is related to evolution and noise in biological systems. Biological systems are inherently noisy due to stochastic processes and fluctuations. To mitigate the effects of noise, circuit designers in synthetic biology employ strategies such as amplification, signal averaging, and stochastic modeling. These techniques help improve the reliability and precision of circuit responses, ensuring consistent and predictable behavior. Notably, the host chassis may introduce mutations into the engineered biological circuits, and the entire synthetic system could collapse. Indeed, noise can cause the designed circuit to behave in an unstable and heterogeneous manner. Protein concentrations in cell can fluctuate leading to stochasticity in gene expression and cell to cell variation.



vii. **Modulation and Tunability.** Synthetic biology circuits often require the ability to modulate and control their behavior. Designers incorporate regulatory elements that allow the dynamic control of gene expression, such as inducible promoters, tunable promoters, and regulatory motifs. Modulation and tunability provide the flexibility to adjust circuit output in response to external signals, allowing for fine-tuning of system behavior.

viii. **Robustness and Stability.** Designing circuits with robustness and stability is essential to ensure reliable performance in varying conditions. Robustness refers to the ability of a circuit to maintain its function despite perturbations or parameter variations. Stability ensures that the circuit remains in a desired state without spontaneous transitions. By incorporating robustness and stability features, synthetic biology circuits can withstand fluctuations and uncertainties in the biological environment.

### **Post-transcriptional regulation**

In the cell, gene expression occurs in two essential steps: transcription and translation which, in prokaryotes, are tightly coupled in time and space [54]. Notably, the transcription event is tightly regulated by transcription factors (TFs). Natural TFs typically contain two DNA-binding domains (DBDs) that bind a target site on the gene promoter, and activation domain (AD) that activates transcription by interacting with the basal transcription machinery of the cell. Over the years, synthetic regulation of gene expression predominantly focused on the transcription level engineering promoters or TFs. For instance, the strength of promoters can be altered changing the length of the spacer region between the -35 and -10 box, or around and in the -35 box, to the -10 box [55,56,57]. Indeed, zinc-fingers proteins and transcription activator-like effectors (TALEs) are the first gene editing components that act through DNA-protein interaction [58,59]. However, this mode of function requires a laborious engineering of proteins for different genomic sequences. Nevertheless, this approach paved the way for researchers to a new era of gene

regulation developing the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems as powerful tool [60].

Although most tools have so far been developed to regulate gene expression at the transcriptional level engineering gene circuits with DNA-protein interactions, a limited number of artificial regulators are available for the post-transcriptional regulation. To date, tools that develop a programmable and versatile platform for gene regulation are needed. Recent advances in RNA biology are inspiring the use of RNA components in the design of synthetic systems [61,62]. Due to its versatile regulation of gene expression, RNA-based gene circuits are now fundamental components of the synthetic biology toolbox. In addition to RNA synthesis (transcription), the control points of gene expression that occur in prokaryotes are RNA degradation and protein synthesis (translation). Notably, for each control point classes and mechanisms of RNA-based gene regulation exist. Regarding the translational control, small RNAs (sRNAs) are defined as non-coding-RNA molecules that bind the ribosome binding site (RBS) or their target mRNA, causing competition with the ribosome for binding to that region. Indeed, sRNAs generally act as repressor in gene silencing. In bacteria, regulatory RNAs naturally comprise a size range between 20 and 400 nucleotides in length; they do not encode proteins, but they modulate transcription, translation, RNA stability in response to environmental changes. sRNAs directly target specific mRNA through base-pairing interactions at the RBS or start codon region [63,64]. mRNA is a cis-repressed RNA (crRNA) by the presence of a 5' stem loop structure that hold the RBS. Activation of translation can be obtained by the presence of a trans-activating RNA (taRNA) that target the stem loop structure exposing the RBS. Moreover, these riboregulators have been further engineered to provide orthogonal crRNA and taRNA pairs that have been combined in synthetic counter devices [65] and computationally implemented with an algorithm in a structure-guided design [66]. Alternatively, riboswitches positively or negatively control gene expression in response to ligands. Riboswitches are made of two structurally linked domains: an aptamer, and a

platform that is sensitive to the ligand [67,68,69]. RNAs that act as enzymes, called ribozymes, have also the ability to control translation by the RNA cleavage or ligation [70]. For the regulation of targeted gene expression, a key role is played by non-regulatory RNAs, however, they are limited. To date, RNA-binding proteins (RBPs) are of crucial importance in the post-transcriptional regulation of gene expression. These proteins are numerous and diverse, they contain regions with RNA-binding domain function, and auxiliary domain that mediate protein-protein interaction and subcellular targeting [71,72]. The two best studied bacterial RBPs are CsrA and Hfq. CsrA predominantly compete with the ribosome for binding to the RBS of its targets; Hfq generally mediates the interaction between sRNAs and mRNAs, but it is involved in a variety of other mechanisms to post-transcriptionally control gene expression [73]. However, additional proteins have been recently identified to control gene expression at post-transcriptional level adapting the susceptibility to RNAses, affecting RNA stability, and modulating RBS accessibility. Moreover, RNA-binding proteins can be involved in gene expression regulation acting as a scaffold and participating in the formation of ribonucleoprotein (RNP) complexes. Works published over the last years have extended our understanding of the structure and the function of the RNA-binding domains (RBDs). At the structural level, RBPs often exhibit a high degree of modularity and are composed of multiple repeats of a few small domains. RNA recognition motif (RRM) [74], and hnRNP K homology domain (KH) [75] are structural examples by which such proteins regulate RNA metabolism and, consequently, gene expression.

### **RNA-binding proteins**

RNA-binding proteins (RBPs) exist in all living organisms playing a crucial role for most cellular processes. RBPs were first characterized using biochemical techniques, such as gel electrophoresis and ultraviolet (UV)-crosslinked nuclear extracts or RNA affinity purification coupled with mass spectrometry and/or immunodetection [76,77]. First analysis using single-stranded RNA structure predictions (ssRBDs)

identified ~500 RBPs in mammals suggesting their role in controlling a complex regulatory network [78,79]. Subsequently, bioinformatic analysis supported by the Gene Ontology (GO) project, database entries (Pfam) and structural features reported by literature estimated ~1900 human RBPs [80]. The most conserved RBPs are a set of ribosomal proteins (r-proteins) that were already present at the time of the last common ancestor [81]. Prokaryotic and eukaryotic organisms use RBPs both as structural components of complexes (ribosome), and regulators of cellular processes (RNA synthesis, modification, translation, processing, and decay). As most of proteins, the structure determines the function. Notably, RBPs contain deeply conserved RNA-binding domains (RBDs) across bacteria and eukaryotes—which allow them to interact with their ligands. Based on their specific RBDs and their ability to bind to RNA, RBPs are classified in families. The most common RBDs found in bacteria are the S1 domain [82], the cold-shock domain (CSD) of the oligonucleotide/oligosaccharide binding (OB) superfamily [83], the Sm and Sm-like domains [84], the RNA recognition motif (RRM) [85], the K homology (KH) domain [86], the double-stranded RNA-binding domain (dsRBD) [87] and the PAZ and PIWI domains [88]. Regarding the classes of domains found in eukaryotes, some of RBDs are shared with prokaryotes, such as the RRM, the KH and the dsRBD. In addition, zinc-finger domain, and DEAD motif were also identified. RBPs control all aspects of gene expression at post-transcriptional level, including RNA editing, transport, and mRNA turnover.

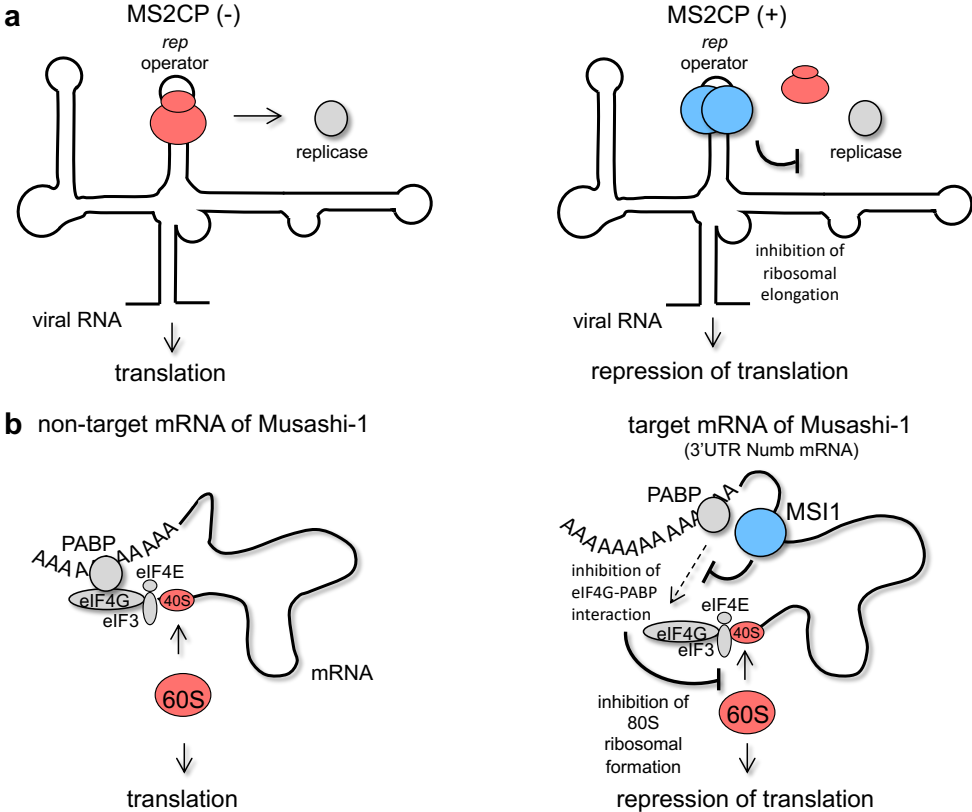
In this thesis, the role of two RBPs on gene expression regulation at post-transcriptional level was studied. The focus was on the viral MS2 coat protein, and the mammalian Musashi-1.

Bacteriophage MS2 is a small spherical virus that causes the infection of *E. coli*, and their genomic RNA serves for synthesis of viral proteins. The icosahedral virus contains 180 copies of a coat protein forming a shell around a single-stranded RNA molecule. The coat protein is made of 129 amino acid residues which, in its monomeric form, has a molecular mass of 13.7 kDa. The assembled particle has a

triangulation number of  $T=3$  and a diameter of 275 Å. In addition to phage-particles packaging and host cell infection, MS2 plays a role in gene regulation. MS2 binds to a specific stem-loop structure in the single-stranded viral RNA acting as translational repressor of the phage-encoded replicase gene in infected cells [89]. The viral RNA genome is made of 3569 nucleotides, it encodes a maturation or A protein, the coat protein promoter, a replicase subunit and a lysis protein. During the infection, coat protein binds to the translation initiation region of the replicase cistron preventing the initiation of translation by ribosome. Studies of X-ray crystallography showed that the MS2 coat protein and RNA form two different complexes depending on the protein concentration and the RNA-protein ratio [90]. In the form of homodimer, MS2 binds to an RNA stem-loop structure of 19 nucleotides that contain the initiation codon of the phage replicase gene [91,92]. Therefore, the active repressor is a dimer, and one RNA molecule is bound by a repressor dimer at saturation. Over the years, the MS2 coat protein has been engineered for many applications such as the study of protein-RNA interaction *in vivo*, and trafficking of delivered-mRNA (Fig. 5a).

Musashi-1 (MSI-1) is a eukaryotic protein that is expressed in both vertebrates and invertebrates. MSI-1 belongs to a family of neural RNA binding protein (RBPs) which was originally discovered in 1994 by Nakamura and colleagues [93] in *Drosophila* as a progenitor cell fate regulator. In *Drosophila*, MSI-1 plays a key role in regulating asymmetrical division sensory neural precursors cells. When *msi* gene is mutated and its function is lost, the sensory organ precursors cells (SOPs) appear different in shape leading in a double-bristle phenotype. For this reason, the name of the gene reflects similarity of this phenotype to the martial portrayals of the two-swords fighting style originated by the Japanese warrior Miyamoto Musashi born in the 15<sup>th</sup> century. The *msi* gene is evolutionarily conserved and MSI-1 protein is a conserved marker for neural progenitor cells [94,95]. MSI-1 was isolated as a mammalian homologue of *Drosophila* Musashi; the family protein consists of two orthologs (MSI-1 and MSI-2) in humans and mice. Both homologs post-transcriptionally regulate the expression of target genes that are involved in cell

fate determination and cancer development. Notably, MSI-1 expression pattern in the central nervous system and primary protein structure is conserved among vertebrates and invertebrates, and the human and mouse proteins share sequence identity [93,96,97,98].



**Fig. 5** Schematics of the downregulatory function of **a)** the coat protein MS2 on the viral replicase, and **b)** the MSI-1 on the target mRNA in mammalian cells.

In humans, MSI-1mRNA and protein are highly expressed in the progenitor cells of the central nervous system (CNS) (astrocytes) playing a role in the stemness maintenance [99,100]. By binding the 3'-untranslated region (UTR) of the mammalian *numb* (m-numb) messenger RNA (mRNA), MSI-1 is naturally involved in the regulation of the Notch signaling pathway and the development of the central nervous system [101]. The MSI-1 protein consists of 362 amino acid (aa) residues, and it contains two conserved RNA recognition motifs (RRMs) in its N-terminal region, each containing a nuclear localization signal (NLS), and a putative nuclear

export signal (NES) in its C-terminal region [102,103]. The *in vitro* selection method (SELEX) demonstrated that MSI-1 is able to block translation of the target mRNA by binding to (G/A) $U_n$ AGU sequence motif ( $n=1-3$ ) at 3' untranslated regions (UTRs). In a natural context, MSI-1 inhibits the initiation of translation of target mRNAs by competing with eIF4G factor for binding to poly(A) binding protein (PABP) preventing the formation of 80s ribosome complex (Fig. 5b).

## REFERENCES

1. Committee on Science, Technology, and Law, Policy and Global Affairs, and Board on Life Sciences, Division on Earth and Life Sciences, National Academy of Engineering and National Research Council. (2013). Positioning synthetic biology to meet the challenges of the 21st century: Summary report of a six academies symposium series. Washington DC: National Academies Press.
2. Breitling R, Takano E, Gardner TS (2015). Judging synthetic biology risks. *Science*, 347,107.
3. Convention on Biological Diversity. (2015). Synthetic Biology. Technical series no.82. *Montreal: Secretariat of the Convention on Biological Diversity*.
4. McAdams HH, Shapiro L (1995) Circuit simulation of genetic networks. *Science*, 269, 650-656.
5. McAdams HH, Arkin A (2000) Towards a circuit engineering discipline. *Curr Biol*, 10, R318-R320.
6. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403, 339-342.
7. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, 403, 335-338.
8. Monod J, Jacob F (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol*, 26, 389-401.
9. Jacob F, Monod J (1961) On the regulation of gene activity. *Cold Spring Harb Symp Quant Biol*, 26, 193-211.
10. Ptashne M, Johnson AD, Pabo CO (1982) A genetic switch in a bacterial virus. *Sci Am*, 247, 128-130.
11. Cameron D, Bashor C, Collins J (2014) A brief history of synthetic biology. *Nat Rev Microbiol* 12, 381-390.
12. Hasty J, McMillen D, Collins JJ (2002) Engineered gene circuits. *Nature* 420, 224-230.
13. Weiss, R, Knight TF (2001) DNA Computing. *Springer*, 1-16.
14. Park SH, Zarrinpar A, Lim WA (2003) Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science*, 299, 1061-1064.
15. Ball P (2004) Synthetic biology: starting from scratch. *Nature*, 431, 624-626.
16. Ferber D (2004) Synthetic biology. Microbes made to order. *Science*, 303, 158-161.
17. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I<sub>1</sub>-I<sub>2</sub> regulatory elements. *Nucleic Acids Res*, 25, 1203-1210.



18. Dueber JE, Yeh BJ, Chak K, Lim WA (2003) Reprogramming control of an allosteric signaling switch through modular recombination. *Science*, 301, 1904-1908.
19. Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol*, 2, 2006-0028.
20. O'Malley MA, Dupré J (2005) Fundamental issues in systems biology. *BioEssays*, 12, 1270-1276.
21. Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. *Nature*, 456, 516-519.
22. Danino T, Mondragón-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. *Nature*, 463, 326-330.
23. Shin J, Noireaux V (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J Biol Eng*, 4, 1-9.
24. Niederholtmeyer H, Sun ZZ, Hori Y, Yeung E, Verpoorte A, Murray RM, Maerkl SJ (2015) Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife*, e09771.
25. Danchin A, Sekowska A (2013) Constraints in the design of the synthetic bacterial chassis. *Meth Microbiol*, 40, 39-67.
26. Keasling JD (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol*, 3, 64-76.
27. Goeddel DV, Kleid DG, Bolivar F, Riggs AD (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci USA*, 76, 106-110.
28. Chang MC, Eachus RA, Trieu W, Ro DK, Keasling JD (2007) Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nat Chem Biol*, 3, 274-277.
29. Nielsen DR, Leonard E, Yoon SH, Tseng HC, Yuan C, Prather KL (2009) Engineering alternative butanol production platforms in heterologous bacteria. *Metab Eng*, 11, 262-273.
30. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 451, 86-89.
31. Song Y, Fu G, Chen J, Li Q, Xie N, Zheng P, Sun J, Zhang D (2016) Promoter screening from *Bacillus subtilis* in various conditions hunting for synthetic biology and industrial applications. *PLoS One*, 11, e0158447.
32. Ma Y, McClure DD, Somerville MV, Proschogo NW, Dehghani F, Kavanagh JM, Coleman NV (2019) Metabolic engineering of the MEP pathway in *Bacillus subtilis* for increased biosynthesis of menaquinone-7. *ACS Synthetic Biology*, 8, 1620-1630.

33. Ulrich N, Nagler K, Laue M, Cockell CS, Setlow P, Moeller R (2018) Experimental studies addressing the longevity of *Bacillus subtilis* spores – The first data from a 500-year experiment. *PLoS One*, 13, e0208425.
34. Spizizen J (1958) Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *Proceedings of the National Academy of Science of the United States of America*, 44, 1072-1078.
35. Watanabe S, Shiwa Y, Itaya M, Yoshikawa M (2012) Complete sequence of the first chimera genome constructed by cloning the whole genome of *Synechocystis* strain PCC6803 into the *Bacillus subtilis* 168 genome. *J of Bacteriology*, 194, 7007.
36. Daszczuk A, Dessalegne Y, Drenth I, Hendriks E, Jo E, van Lente, T, Oldebusten A, Parrish J, Poljakova W, Purwanto AA, van Raaphorst R, Boonstra M, van Heel A, Herber M, van der Meulen S, Siebring J, Sorg RA, Heinemann M, Kuipers OP, Veening JW (2014) *Bacillus subtilis* biosensor engineered to assess meat spoilage. *ACS synthetic biology*, 3, 999-1002.
37. Zhang W, Zhu X, Wang Y, Li T (2022) *Bacillus subtilis* chassis in biomanufacturing 4.0. *J. Chem. Technol. Biotechnol*, 97, 2665-2674.
38. Inoue A, Horikoshi K (1989) A *Pseudomonas* thrives in high concentrations of toluene. *Nature*, 227, 264-265.
39. Jiménez JI, Miñambres B, García JL, Díaz E (2002) Genomic analysis of the aromatic catabolic pathway from *Pseudomonas putida* KT2440. 4, 824-841; Beckham GT, Johnson CW, Karp EM, Salvachúa, Vardon DR (2016) Opportunities and challenges in biological lignin valorization. *Curr Opin Biotechnol*, 42, 40-53.
40. Nickel PI, Chavarría M, Fuhrer T, Sauer U, de Lorenzo V (2015) *Pseudomonas putida* KT2440 strain metabolizes glucose through a cycle formed by enzymes of the Entner-Doudoroff, Embden-Meyerhof-Parnas, and pentose phosphate pathways. *J. Biol Chem*, 290, 25920-25932.
41. Valencia LE, Incha MR, Schmidt M, Pearson AN, Thompson MG, Roberts JB, Mehling M, Yin K, Sun N, Oka A, Shih PM, Blank LM, Gladden J, Keasling, JD (2022) Engineering *Pseudomonas putida* KT2440 for chain length tailored free fatty acid and oleochemical production. *Commun Biol*, 5, 1363.
42. de Lorenzo V (1994) Designing microbial systems for gene expression in the field. *Trends Biotechnol*, 12, 365-371.
43. Zobel S, Benedetti I, Eisenbach L, de Lorenzo V, Wierckx N, Blank LM (2015) Tn7-based device for calibrated heterologous gene expression in *Pseudomonas putida*. *ACS synthetic biology*, 4, 1341-1351.

44. Komatsu M, Uchiyama T, Ōmura S, Cane DE, Ikeda H (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc Natl Acad Sci USA*, 107, 2646-2651.
45. Newman DJ, Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat Prod*, 70, 461-477.
46. Nguyen HT, Pokhrel AR, Pham VTT, Dhakal D, Lim HN, Jung HJ, Kim TS, Yamaguchi, Sohng JK (2020) *Streptomyces* sp. VN1, a producer of diverse metabolites including non-natural furan-type anticancer compound. *Sci Rep*, 10, 1756.
47. Dahl JU, Gray MJ, Jakob U (2015) Protein quality control under oxidative stress conditions. *J. Mol Biol*, 427, 1549-1563.
48. Surovtsev IV, Jacobs-Wagner C (2018) Subcellular organization: a critical feature of bacterial cell replication. *Cell*, 172, 1271-1293.
49. Conrado RJ, Varner JD, DeLisa MP (2008) Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr Opin Biotechnol*, 19, 492-499.
50. Benedetti I, de Lorenzo V, Nikel PI (2016) Genetic programming of catalytic *Pseudomonas putida* biofilms for boosting biodegradation of haloalkanes. *Metabolic engineering*, 33, 109-118.
51. Lian J, Mishra S, Zhao H (2018) Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: new tools and their applications. *Met Eng*, 50, 85-108.
52. Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature reviews microbiology*, 12, 355-367.
53. Feng X, Marchisio MA (2021) *Saccharomyces cerevisiae* promoter engineering before and during the synthetic biology era. *Biology*, 10, 504.
54. McGary K, Nudler E (2013) RNA polymerase and the ribosome: the close relationship. *Curr Opin Microbiol*, 16, 112-117.
55. Rossi JJ, Soberon X, Marumoto Y, McMahan J, Itakura K (1983) Biological expression of an *Escherichia coli* consensus sequence promoter and some mutant derivatives. *Proc Natl Acad Sci USA*, 80, 3203-3207.
56. Horwitz M, Loeb LA (1988) DNA sequences of random origin as probes of *Escherichia coli* promoter architecture. *J Biol Chem*, 263, 14724-14731.
57. Jaurin B, Grundström T, Normark S (1982) Sequence elements determining ampC promoter strength in *E. coli*. *EMBO J*, 1, 875-881.
58. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc Natl Sci USA*, 93, 1156-1160.

59. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326, 1509.
60. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.
61. Thompson KM, Syrett HA, Knudsen SM, Ellington AD (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol.*, 2, 21.
62. 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.
63. Takayanagi K, Maeda S, Mizuno T (1991) Expression of micF involved in porin synthesis in *Escherichia coli*: two distinct cis-acting elements respectively regulate micF expression positively and negatively. *FEMS Microbiol Lett*, 83, 39-44.
64. Urban JH, Vogel J (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. *PLoS Biol*, 6, e64.
65. Friedland AE, Luu TK, Wang X, Shi D, Church G, Collins JJ (2009) Synthetic gene networks that count. *Science*, 324, 1199-1202.
66. Rodrigo G, Landrain TE, Jaramillo A (2012) De novo automated design of small RNA circuits for engineering synthetic riboregulators in living cells. *Proc Natl Acad Sci USA*, 109, 15271-15276.
67. Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol*, 59, 487-517.
68. Wang JX, Lee ER, Morales DR, Lim J, Breaker RR (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell*, 29, 691-702.
69. Wrinkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419, 952-956.
70. Tang J, Breaker RR (1997) Rational design of allosteric ribozymes. *Chem Biol*, 4, 453-459.
71. Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, 265: 615-621.
72. Kiledjian M, Burd CG, Portman DS, Gorlach M, Dreyfuss G (1994) Structure and function of hnRNP proteins. *RNA-protein Interactions*, 127-149.
73. Adamson DN, Lim HN (2011) Essential requirements for robust signaling in Hfq dependent small RNA network. *PLoS Comput Biol*, 7, e1002138.

74. Dreyfuss G, Swanson MS, Pinol-Roma S (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci*, 13, 86-91.
75. Valverde R, Edwards L, Regan L (2008) Structure and function of KH domains. *FEBS J*, 275, 2712-2726.
76. Dreyfuss G, Choi YD, Adam SA (1984) Characterization of heterogeneous nuclear RNA-protein complexes *in vivo* with monoclonal antibodies. *Mol Cell Biol*, 4, 1104-1114.
77. Pinol-Roma S, Choi YD, Matunis MJ, Dreyfuss G (1988) Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev*, 2, 215-227.
78. McKee AE, Minet E, Stern C, Riahi S, Stiles CD, Silver PA (2005). A genome-wide *in situ* hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC dev biol*, 5, 1-9.
79. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res*, 39, D301-D308.
80. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000). Gene ontology: tool for the unification of biology. *Nat genet*, 25, 25-29.
81. Fox GE (2010) Origin and evolution of the ribosome. *Cold Spring Harb perspect biol*, 2, a003483.
82. Hajnsdorf E, Boni IV (2012) Multiple activities of RNA-binding proteins S1 and Hfq. *Biochimie*, 94, 1544-1553.
83. Chaikam V, Karlson DT (2010) Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. *BMB Rep*, 43, 1-8.
84. Updegrove TB, Zhang A, Storz G (2016) Hfq: the flexible RNA matchmaker. *Curr Opin Microbiol*, 30, 133-138.
85. Koonin EV, Makarova KS (2013) CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol*, 10, 679-686.
86. Nicastrò G, Taylor IA, Ramos A (2015) KH-RNA interactions: back in the groove. *Curr Opin Struct Biol*, 30, 63-70.
87. Masliah G, Barraud P, Allain FH (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci*, 70, 1875-1895.

88. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J (2014) The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol*, 21, 743–753.
89. Peabody DS (1993) The RNA binding site of bacteriophage MS2 coat protein. *The EMBO J*, 12, 595-600.
90. Valegård K, Liljas L, Fridborg K, Unge T (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, 345, 36-41.
91. Golmohammadi R, Valegård K, Fridborg K, Liljas L (1993) The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J Mol Biol*, 234, 620-639.
92. Bernardi A and Spahr PF (1972) Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. *Proc Nat Acad Sci USA*, 10, 3033-3037.
93. Nakamura M, Okano H, Blendy JA, Montell C (1994) Musashi, a neural RNA-binding protein required for drosophila adult external sensory organ development. *Neuron*, 13, 67-81.
94. Kanemura Y, Yamasaki M, Mori K, Fujikawa H, Hayashi H, Nakano A, Matsumoto T, Tamura K, Arita N, Sakakibara S, Ohnishi T, Fushiki S, Nakamura Y, Imai T, Okano H (2001) Musashi1, an evolutionarily conserved neural RNA-binding protein, is a versatile marker of human glioma cells in determining their cellular origin, malignancy, and proliferative activity. *Differentiation*, 69, 141-152.
95. Bley N, Hmedat A, Müller S, Rolnik R, Rausch A, Lederer M, Hüttelmaier S (2021) Musashi-1-A stemness RBP for cancer therapy? *Biology (Basel)*, 10, 407.
96. Okano H, Imai T, Okabe M (2002) Musashi: a translational regulator of cell fate. *JCell Sci*, 115, 1355-1359.
97. Yoda A, Sawa H, Okano H (2000) MSI-1, a neural RNA-binding protein, is involved in mle mating behavior in *Caenorhabditis elegans*. *Genes Cells*, 5, 885-895.
98. Sakakibara S, Imai T, Hamaguchi K, Okabe M, Aruga J, Nakajima K, Yasutomi D, Nagata T, Kurihara Y, Uesugi S, Miyata T, Ogawa M, Mikoshiba K, Okano H (1996) Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev Biol*, 176, 230-242.
99. Toda M, Iizuka Y, Yu W, Imai T, Ikeda E, Yoshida K, Kawase T, Kawakami Y, Okano H, Uyemura K (2001) Expression of the neural RNA-binding protein Musashi1 in human gliomas. *Glia*, 34, 1-7.

100. Kaneko Y, Sakakibara S, Imai T, Suzuki A, Nakamura Y, Sawamoto K, Ogawa Y, Toyama T, Okano H (2000) Musashi1: evolutionarily conserved markers for CNS progenitor cells including neural stem cells. *Dev Neurosci*, 22, 138–152.
101. Imai T, Tokunaga A, Yoshida T, Hashimoto M, Mikoshiba K, Weinmaster G, Nakafuku M, Okano H (2001) The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol*, 21, 3888–900.
102. Iwaoka R, Nagata T, Tsuda K, Imai T, Okano H, Kobayashi N, Katahira M (2017) Backbone and side chain assignment of the second RNA-binding domain of Musashi-1 in its free form and in complex with 5-mer RNA. *Biomol NMR Assign*, 11, 265-268.
103. Nagata T, Kanno R, Kurihara Y, Uesugi S, Imai T, Sakakibara S, Okano H, Katahira M (1999) Structure, backbone dynamics and interaction with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi. *J Mol Biol*, 287, 315-330.





# OBJECTIVES

Post-transcriptional control systems based on RNA-protein interactions could be of special interest in synthetic biology. However, more work is required to understand in quantitative terms this type of regulation and to know what elements are available for such an engineering effort. Therefore, this thesis has the following objectives:

- 1) To study the dynamic behavior, both deterministic and stochastic (single-cell data), of a post-transcriptional gene expression program based on a suitable RNA binding protein.
- 2) To study how an RNA-binding protein from mammals works in a bacterium, which has much simpler gene expression machinery and intracellular organization, and serves to engineer an orthogonal post-transcriptional control system.



# CHAPTER 1

## **Gene regulation by a protein translation factor at the single-cell level**

This work has been published in a peer-reviewed journal.

See the full citation:

Dolcemascolo R, Goiriz L, Montagud-Martínez R, and Rodrigo G. (2022) Gene regulation by a protein translation factor at the single-cell level. *PLoS Computational Biology*, 18, e1010087.

<https://doi.org/10.1371/journal.pcbi.1010087>

My contribution to this work:

I made the genetic constructs and performed the characterization experiments under the supervision of G. Rodrigo. I also contributed to the writing of the manuscript.

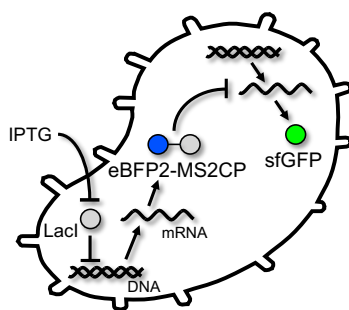


# 1. INTRODUCTION

The ability to map a given genotype to its corresponding phenotype is perhaps the biggest pursuit in molecular biology [1], especially in the post-genomic era, as it can provide fundamental insight and predictive power on natural evolution, with clear applications in biomedicine and ecology. However, it is well established that the very same genotype can lead to phenotypic heterogeneity in a non-changing environment [2]. This is the consequence of the inherent stochasticity of the different biochemical reactions that are needed for gene expression [3]. While stochastic events are often seen as undesirable, as they are when the optimal gene expression levels are lost [4], we now realize that a noisy gene expression can also be advantageous for the cell population to face environmental changes or induce time-dependent behaviors [5]. In this regard, substantial progress has been made over the last years to quantitatively understand and model this non-genetic variability (noise). However, there are still numerous questions regarding the mechanisms that produce and regulate noise in gene expression.

Motivated by the prevalence of transcriptional regulations in the cell [6], previous work focused on studying the emergence and propagation of noise in genes regulated transcriptionally [3,7]. For example, we now appreciate that some promoters can generate bursts of expression as a consequence of a stochastic switching in their activity [8], that the sign of the regulation determines the best way to extract information from the environment [9], and that the stochastic fluctuations can inform about the underlying regulation when time is considered [10]. In addition, recent studies also focused on post-transcriptional regulations implemented by small non-coding RNAs (in particular, by microRNAs in eukaryotic cells) [11,12]. These studies concluded that microRNAs, by controlling the messenger RNA (mRNA) abundance, can suppress part of the noise generated at the level of transcription, hence resulting in ideal genetic elements to engineer robust circuits. Nevertheless,

studies on cell-to-cell variability when protein expression is regulated at the level of translation are scarce, especially when the regulation is exerted by a translation factor. We just know that structured 5' untranslated regions (UTRs) can generate noise in protein expression [13], which can even be tuned by *trans*-acting small RNAs [14]. The importance of studying how stochastic gene expression is generated and regulated at different levels in the genetic information flow lies in the fact that living cells implement highly intricate circuitries for multiple signal integration that allow displaying a variety of phenotypes. Certainly, this signal integration becomes easier and more scalable if different layers are exploited (*e.g.*, transcriptional, translational, and post-translational), and this is precisely what has evolved in nature. Only by understanding the particularities of each regulatory mode, we can rationalize the impact of gene expression on the cell behavior. As highlighted before, more studies on stochastic gene expression regulated at a layer other than transcription are mandatory, especially because there are important phenotypes in nature that arise as a consequence of changing expression translationally.



**Fig. 1** Schematics of the synthetic genetic circuit implemented in a bacterial cell. IPTG is the external molecule that controls the expression of the protein translation factor (eBFP2-MS2CP). sfGFP is the final output of the system.

In this work, we exploited the bacteriophage MS2 coat protein (MS2CP) as a translation factor [15] to engineer a basic synthetic regulatory circuit from which to study stochastic gene expression when it is regulated translationally (Fig. 1).

In the natural context, in addition to be a structural protein to form the virion, MS2CP blocks the translation of the viral replicase upon binding to an RNA hairpin in the

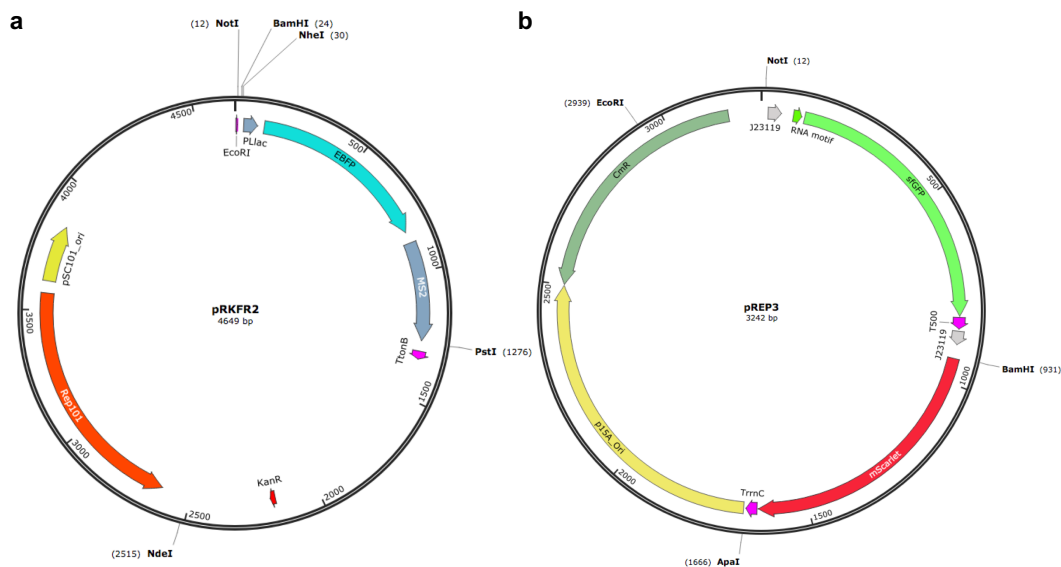
corresponding 5' UTR that contains the ribosome binding site (RBS) and the start codon [16]. Over the years, MS2CP has been used for many applications due to its strong binding affinity to RNA, such as the subcellular tracking of mRNAs with time and space [17], the study of protein-RNA interactions *in vivo* [18], the development of CRISPR scaffold RNAs for programmable transcription regulation (CRISPR stands for clustered regularly interspaced short palindromic repeats) [19], and the construction of nanoscale architectures that can serve, for example, to improve enzymatic reactions [20]. With our engineered circuit, we examined gene expression in single cells by using a double reporter system to monitor both the regulator (MS2CP) and the regulated gene, and we also developed a mathematical model to provide a predictive quantitative foundation of the system.

## 2. MATERIALS AND METHODS

### 2.1. Strains, plasmids, and reagents

*E. coli* Dh5 $\alpha$  was used for cloning purposes by following standard procedures. To express our genetic circuit, *E. coli* MG1655-Z1 (*lacI*<sup>+</sup> and *tetR*<sup>+</sup>; kindly gifted by M.B. Elowitz) was used. This strain was co-transformed by electroporation with two plasmids, called pRKFR2 (kanR, pSC101-E93G ori) and pREP3 (camR, p15A ori). pRKFR2 contains the gene coding for MS2CP translationally fused to eBFP2 in its N terminus (eBFP2-MS2CP) under the transcriptional control of the inducible promoter PLlac. pREP3 contains the gene coding for sfGFP under the transcriptional control of the constitutive promoter J23119 and the translational control of the MS2CP-recognizing RNA motif (Fig. 2). The genetic cassettes were synthesized by IDT. LB medium was used for both overnight and characterization cultures. Kanamycin and chloramphenicol were used at the concentration of 50  $\mu\text{g}/\text{mL}$  and 34  $\mu\text{g}/\text{mL}$ , respectively. IPTG and TC were used as inducers of the system.

The concentration gradient of IPTG that we tested was 0, 5, 10, 20, 50, 100, 200, 500, and 1000  $\mu\text{M}$ , and the concentration gradient of TC was 0, 10, 100, 200, 300, 400, 500, 700, and 1000 ng/mL. Compounds provided by Sigma.



**Fig. 2** Maps of the plasmids used to implement the synthetic gene circuit in which MS2 represses the translation of sfGFP. **a)** Map of pRKFR2 to express the MS2CP protein fused with eBFP protein from a PLlac promoter, induced with lactose or IPTG. **b)** Map of pREP3 to express the reporter sfGFP protein from a constitutive promoter (J23119), harboring a suitable RNA motif in the leader region for translation regulation.

Note that LacI is overexpressed in MG1655-Z1 to efficiently regulate the PLlac promoter in the pRKFR2 plasmid. This overexpression adds to the wild-type expression of LacI. In addition, TetR is expressed in *E. coli* MG1655-Z1 (although it does not play any regulatory role), so it will bind to TC when this inducer is used. The titration effect will be more relevant at low concentrations of TC, although overall this will only mean an effective TC concentration slightly lower. The maximal TC concentration used here reduced significantly the growth rate of the cells, suggesting a marginal effect of TetR in this case.



## 2.2. Growth curves

Cultures (2 mL) inoculated from single colonies (three replicates) were grown overnight in LB medium at 37 °C and 200 rpm. Cultures were then diluted 1:100 in fresh LB medium (2 mL) and were grown for 3 h at the same conditions to reach exponential phase ( $OD_{600}$  around 0.5). Cultures were then diluted 1:50 in fresh LB medium (200  $\mu$ L) to load a microplate (96 wells, black, clear bottom; Corning) with appropriate concentrations of IPTG and TC. The microplate was then incubated for 10 h at 37 °C and 1,000 rpm in a PST-60HL plate shaker (Biosan). Absorbance (600 nm) was measured every hour in a Varioskan Lux fluorometer (Thermo). The growth rate was calculated as the slope between absorbance (in log scale) and time during the exponential phase. Data analysis performed with MATLAB (MathWorks) and Python.

## 2.3. Flow cytometry

Cultures (2 mL) inoculated from single colonies (four replicates) were grown overnight in LB medium at 37 °C and 200 rpm. Cultures were then diluted 1:100 in fresh LB medium (2 mL) and were grown for 3 h at the same conditions to reach exponential phase ( $OD_{600}$  around 0.5). Cultures were then diluted 1:50 in fresh LB medium (200  $\mu$ L) to load a microplate (96 wells, black, clear bottom; Corning) with appropriate concentrations of IPTG and TC. The microplate was then incubated at 37 °C and 1,000 rpm in a PST-60HL plate shaker (Biosan) until cultures reached a sufficient  $OD_{600}$  (a different incubation time for each TC concentration). Cultures (6  $\mu$ L) were then diluted in PBS (1 mL). Fluorescence was measured in an LSRFortessa flow cytometer (BD); a 405 nm laser and a 450 nm filter for blue fluorescence, and a 488 nm laser and a 530 nm filter for green fluorescence. Events were gated by using the forward and side scatter signals and compensated ( $\sim 10^4$  events after this process). The mean value of the autofluorescence of the cells was subtracted in each channel to obtain a final estimate of expression. Data analysis performed with

MATLAB and Python. The mean and the variance were calculated for each distribution after removing outliers, which served to compute the noise in gene expression.

## 2.4. Deterministic mathematical modelling

*Mathematical model and data analysis were performed in collaboration with the research team at our lab. Even though my contribution in this part is limited because I am not an expert in mathematical modelling, I believe it was necessary to add the following sections for the integrity of the thesis and relevance of the topic.*

We assumed that the cellular amount of regulatory protein (eBFP2-MS2CP) is proportional to the signal of blue fluorescence, and that the amount of regulated protein (sfGFP) is proportional to the signal of green fluorescence. Thus, with IPTG and TC being the two external molecules of control, the Hill-Langmuir equations that dictate average protein expression (population measure) are

$$\begin{aligned} \langle \text{eBFP2} \rangle &= \alpha_x \frac{\rho_x + \left( \frac{\text{IPTG}}{\theta_i} \right)^{n_i}}{1 + \left( \frac{\text{IPTG}}{\theta_i} \right)^{n_i}} \\ \langle \text{sfGFP} \rangle &= \alpha_y \frac{1 + \rho_y \left( \frac{\langle \text{eBFP2} \rangle}{\theta_x} \right)^{n_x}}{1 + \left( \frac{\langle \text{eBFP2} \rangle}{\theta_x} \right)^{n_x}}, \quad (1) \end{aligned}$$

where  $\alpha_x$  is the maximal protein level from the PLlac promoter (in presence of IPTG),  $\rho_x$  the transcriptional repression fold by LacI,  $\alpha_y$  the maximal protein level from the constitutive J23119 promoter, and  $\rho_y$  the translational repression fold by eBFP2-MS2CP. In addition,  $\theta_i$  is the effective dissociation constant between LacI and IPTG,  $n_i$  the effective degree of cooperativity of LacI,  $\theta_x$  the effective dissociation constant between eBFP2-MS2CP and the cognate RNA motif embedded within the sfGFP mRNA, and  $n_x$  the effective degree of cooperativity of eBFP2-MS2CP. By using our data, the adjusted parameter values are  $\alpha_x = 1,890$  AU,  $\rho_x =$

0.225,  $\theta_i = 116 \mu\text{M}$ ,  $n_i = 2.38$ ,  $\alpha_y = 29,000 \text{ AU}$ ,  $\rho_y = 0.016$ ,  $\theta_x = 610 \text{ AU}$ , and  $n_x = 5$  (upon varying IPTG, with no TC).

Moreover, if  $\mu$  denotes the actual cell growth rate, which is modulated by TC and dictates the dilution rate of the proteins, it turns out that the following Michaelis-Menten equation

$$\mu = \frac{\mu_0}{1 + \frac{\text{TC}}{\theta_c}}, \quad (2)$$

where  $\mu_0$  is the maximal cell growth rate (in absence of TC) and  $\theta_c$  the half maximal inhibitory concentration of TC. From the data, we obtained  $\mu_0 = 1.2 \text{ h}^{-1}$  and  $\theta_c = 526 \text{ ng/mL}$  (upon varying TC, with no IPTG).

## 2.5. Stochastic mathematical modelling

The expressions for the variances in gene expression (or fluorescence signal) can be derived by following some basic calculations [3,7]. We followed a Langevin formalism, which consists in introducing in the right-hand side of the differential equation a series of stochastic processes, and the mean-field approximation for the analytical treatment, which consists in assuming that the different fluctuation amplitudes depend on the deterministic solution. In brief, the variance for a given gene and induction condition can be decomposed into three different variances according to the nature of the molecular noise source. The variance coming from extrinsic noise can be assumed to scale with the square of the expression level (then leading to a constant term in gene expression noise), the variance from intrinsic noise with the expression level, and the variance from the regulatory protein noise with the square of the derivative of the transfer function (in terms of protein synthesis rate). In particular, we can write

$$\begin{aligned}
CV_{\text{eBFP2}}^2 &= \underbrace{\eta_x^2}_{\text{extrinsic}} + \underbrace{\frac{\beta_x}{\langle \text{eBFP2} \rangle}}_{\text{intrinsic}} + \underbrace{\frac{1}{2} \left( \frac{\alpha_x (1 - \rho_x) n_i \left( \frac{\text{IPTG}}{\theta_i} \right)^{n_i - 1}}{\theta_i \left( 1 + \left( \frac{\text{IPTG}}{\theta_i} \right)^{n_i} \right)^2} \right)^2}_{\text{regulation}} \frac{\eta_{\text{lac}}^2}{\langle \text{eBFP2} \rangle^2} \\
CV_{\text{sfGFP}}^2 &= \underbrace{\eta_y^2}_{\text{extrinsic}} + \underbrace{\frac{\beta_y}{\langle \text{sfGFP} \rangle}}_{\text{intrinsic}} + \underbrace{\frac{1}{2} \left( \frac{\alpha_y (1 - \rho_y) n_x \left( \frac{\langle \text{eBFP2} \rangle}{\theta_x} \right)^{n_x - 1}}{\theta_x \left( 1 + \left( \frac{\langle \text{eBFP2} \rangle}{\theta_x} \right)^{n_x} \right)^2} \right)^2}_{\text{regulation}} \frac{\gamma_y \langle \text{eBFP2} \rangle^2}{\langle \text{sfGFP} \rangle^2} CV_{\text{eBFP2}}^2, \quad (3)
\end{aligned}$$

where  $\eta_x^2$  and  $\eta_y^2$  are two empirical constants that quantify the levels of noise of extrinsic nature on eBFP2-MS2CP and sfGFP, respectively. Also,  $\eta_{\text{lac}}^2$  is a constant that measures the noise in LacI expression,  $\beta_x$  and  $\beta_y$  the Fano factors of noise of intrinsic nature for eBFP2-MS2CP and sfGFP, respectively, and  $\gamma_y$  a constant that accounts for the difference between fluorescence and number of molecules. Note that  $\langle \text{eBFP2} \rangle = \langle \text{MS2CP} \rangle$  and  $CV_{\text{eBFP2}}^2 = CV_{\text{MS2CP}}^2$ . Note also that when a strong repression occurs at the level of translation the transcriptional noise can be neglected and then  $\beta_y$  can be considered constant (*i.e.*, the Fano factor, in number of molecules per cell, can be approached by 1). By using our data, the adjusted parameter values are  $\eta_x^2 = 0.246$ ,  $\beta_x = 45.6$  AU,  $\eta_{\text{lac}}^2 = 6,470 \mu\text{M}^2$ ,  $\eta_y^2 = 0.127$ ,  $\beta_y = 61.9$  AU, and  $\gamma_y = 0.0233$  (upon varying IPTG, with no TC).

Finally, we assumed that the stochastic gene expression follows a Gamma distribution [28]. Then, the probability for a given expression level reads

$$\begin{aligned}
P(\text{eBFP2}) &= \frac{\text{eBFP2}^{a_x - 1} e^{-\text{eBFP2}/b_x}}{\Gamma(a_x) b_x^{a_x}} \\
P(\text{sfGFP}) &= \frac{\text{sfGFP}^{a_y - 1} e^{-\text{sfGFP}/b_y}}{\Gamma(a_y) b_y^{a_y}}, \quad (4)
\end{aligned}$$

where  $a_x$  and  $a_y$  are the Gamma shape parameters for eBFP2-MS2CP and sfGFP, respectively, and  $b_x$  and  $b_y$  the Gamma scale parameters. Importantly, by knowing

that for a Gamma distribution  $ab$  is the mean and  $ab^2$  the variance, these two parameters can be defined as

$$\begin{aligned} a_x &= \frac{1}{CV_{\text{eBFP2}}^2} \\ a_y &= \frac{1}{CV_{\text{sfGFP}}^2} \\ b_x &= \langle \text{eBFP2} \rangle CV_{\text{eBFP2}}^2 \\ b_y &= \langle \text{sfGFP} \rangle CV_{\text{sfGFP}}^2. \end{aligned} \quad (5)$$

This means that the Gamma shape parameter is directly the inverse of the noise, and that the Gamma scale parameter depends on the translation rate in the case of eBFP2-MS2CP (transcription regulation) and is nearly independent of it in the case of sfGFP (translation regulation).

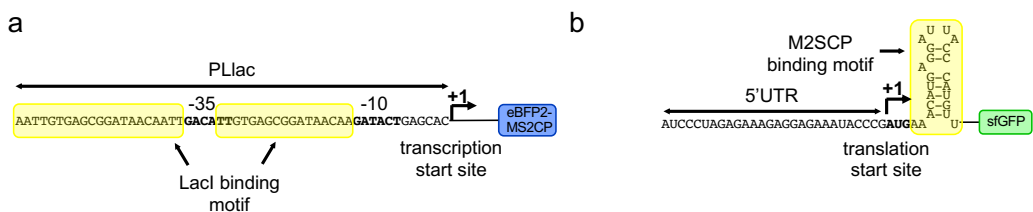
## 2.6. Numerical simulations

The system of stochastic differential equations was solved numerically to obtain stochastic trajectories of mRNA and protein concentrations. For that, we followed an integration scheme previously described [43]. The colored stochastic processes (for extrinsic and regulation noise) were obtained from independent white stochastic processes. The system was solved in one-time interval with the routine ode45s from MATLAB, considering constant the stochastic fluctuations in that interval. The values of the fluctuations were updated in each interval with the previous mRNA and protein concentrations. Negative concentration values were avoided.

### 3. RESULTS

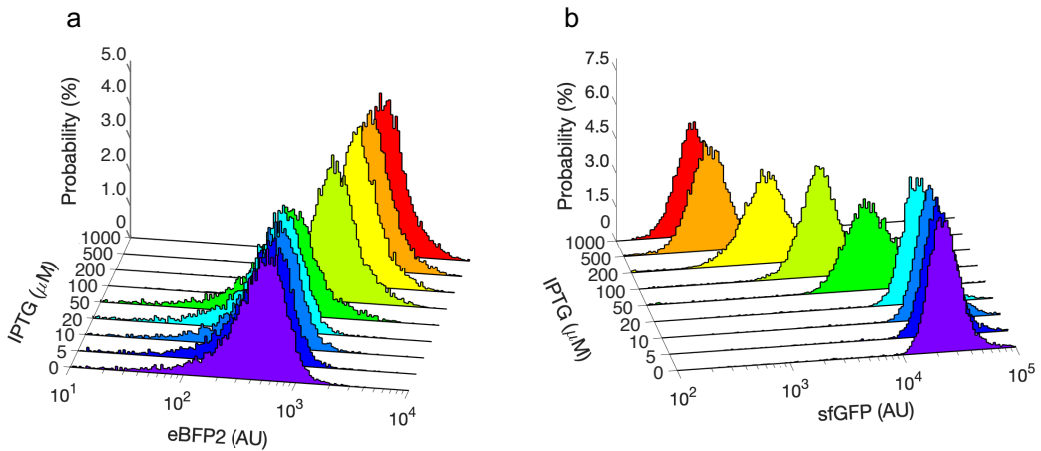
#### 3.1. Regulation of translation with an RNA-binding protein in single cells

We engineered a synthetic genetic system in *E. coli* in which the RNA-binding protein MS2CP acts as a protein translation factor (Fig. 1,3). MS2CP was expressed from a synthetic PL-based promoter repressed by LacI (named as PLlac) [21] in a medium copy number plasmid (about 80 copies/cell). This allowed controlling the expression of the regulator (at the transcriptional level) with isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). In addition, we fused the enhanced blue fluorescent protein 2 (eBFP2) [22] to the N terminus of MS2CP (leading to eBFP2-MS2CP) in order to monitor its expression (Fig. 3a). As a regulated element, here we used the superfolder green fluorescent protein (sfGFP) [23], which was expressed from a constitutive promoter in a low copy number plasmid (about 15 copies/cell). The wild-type RNA motif recognized by MS2CP (with a dissociation constant of about 3 nM) [24] was placed in frame just after the start codon of sfGFP. In this way, MS2CP can block the progression of the ribosome on the regulated gene in the initial phase [15]. This mode of action differs from the natural one, in which MS2CP prevents translation initiation rather than elongation [16]. The resulting circuit behaves like an inverter considering IPTG as input and sfGFP as output, MS2CP being an internal regulator that operates at the level of translation.



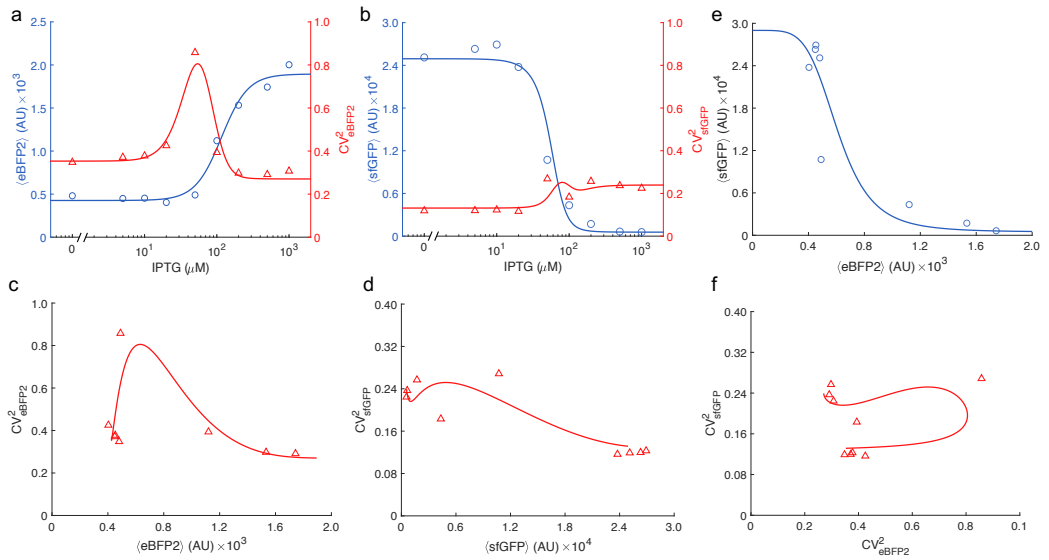
**Fig. 3** Schematics of the synthetic genetic circuit. **a)** sequence details of the *cis*-regulatory region (DNA level) for transcriptional regulation (PLlac promoter). **b)** sequence details of the *cis*-regulatory region (RNA level) for post-transcriptional regulation (MS2CP RNA motif).

We performed single-cell measurements of blue and green fluorescence by flow cytometry for a concentration gradient of IPTG (9 conditions) in order to quantitatively study the stochastic regulatory dynamics of this engineered system (Fig. 4a,b).



**Fig. 4** a) Histograms of single-cell fluorescence for eBFP2 (fused to the regulatory protein) for different induction conditions with IPTG. b) Histograms of single-cell fluorescence for sfGFP (the regulated protein) for different induction conditions with IPTG.

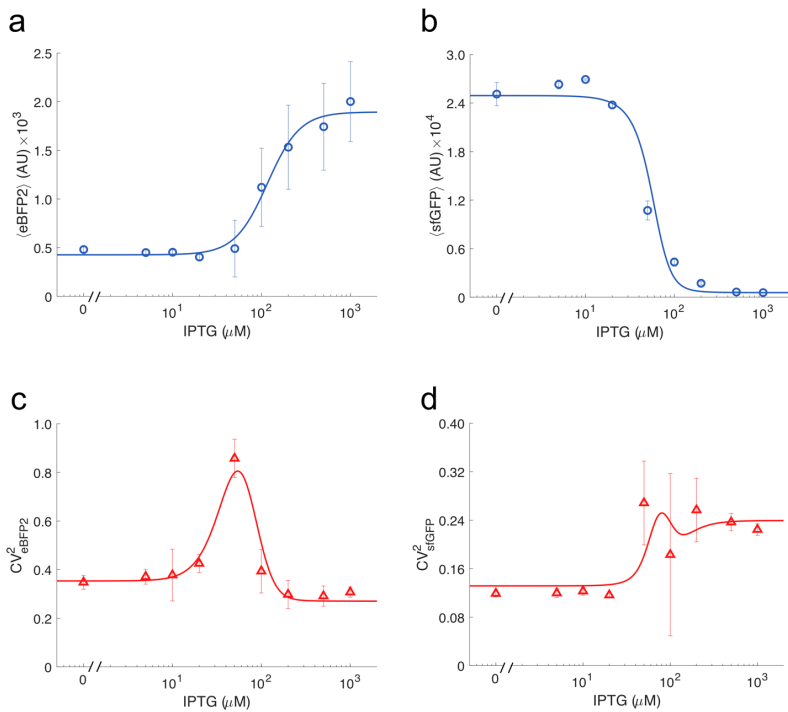
We found a substantial down-regulation of sfGFP (about 50-fold in expression) as a consequence of the action of MS2CP on the cognate mRNA (Fig. 4b). From these data, we calculated the mean and the noise of expression for both eBFP2-MS2CP and sfGFP (the noise as the square of the coefficient of variation) [3], which were represented as a function of IPTG (Fig. 5a,b). The mean gives the average position of the population, and the noise is a measure of the cell-to-cell variability. These measurements were repeated for different populations, finding consistency in the results (Fig. 6a-d).



**Fig. 5** **a)** Mean and noise of expression for eBFP2 as a function of IPTG. **b)** Mean and noise for sfGFP as a function of IPTG. **c)** Noise for eBFP2 as a function of mean expression. **d)** Noise for sfGFP as a function of mean expression. **e)** Transfer function of the post-transcriptional regulation in terms of mean expression. **f)** Transfer function of the post-transcriptional regulation in terms of noise. In the plots, points correspond to calculations from the experimental data, while solid lines to predictions with the mathematical model.

We then constructed a mathematical model relying on a series of algebraic equations from basics on the biochemistry of gene expression and molecular noise propagation [7]. To derive these mathematical expressions for the mean and the noise, we constructed a system of stochastic differential equations for mRNA and protein expression following the Langevin formalism. The rates of concentration changes were subject to stochastic fluctuations of intrinsic and extrinsic nature. This system was analytically solved in steady state with the mean-field approximation for the fluctuations. With a suitable parameterization, our model was able to recapitulate with reasonable agreement the values of mean expression and noise for both eBFP2-MS2CP and sfGFP, highlighting the functional form of the different dose-response curves. In particular, the mean expressions follow Hill-Langmuir equations and the noises non-monotonous curves presenting a maximum at an intermediate IPTG concentration.





**Fig. 6 a)** Mean of eBFP2 expression as a function of IPTG. **b)** Mean of sfGFP expression as a function of IPTG. **c)** Noise of eBFP2 expression as a function of IPTG. **d)** Noise of sfGFP expression as a function of IPTG. Points correspond to the values of the population shown in the main figures. Error bars correspond to standard errors calculated from four different populations. Solid lines correspond to predictions with the mathematical model.

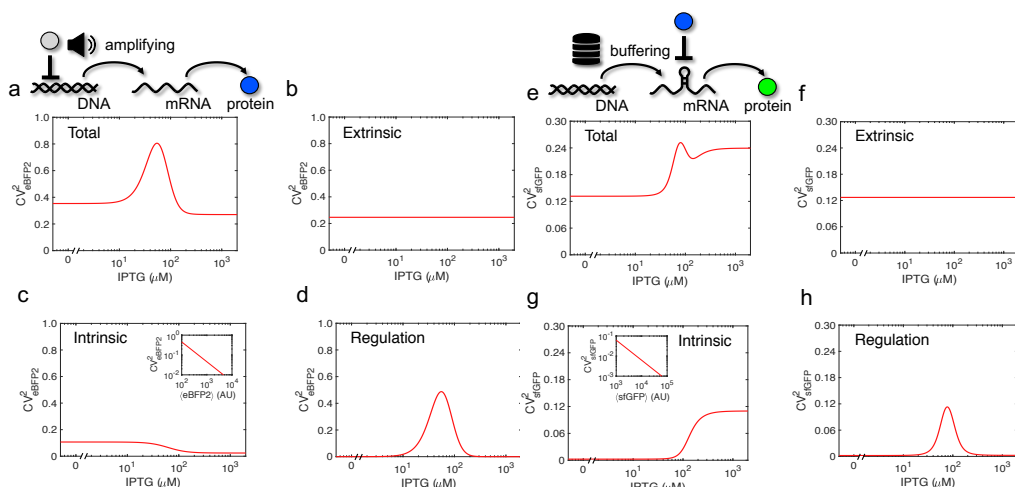
Indeed, the peak-like noise curve is a consequence of a sigmoidal dynamics at the population level. We also observed that the noise levels in sfGFP are lower than in eBFP2 for all IPTG concentrations. In addition, we represented the noise *versus* the mean to show the stochastic expression scaling laws of the system (Fig. 5c,d). The model was also explicative about the nonlinear transfer functions in terms of mean expression regulation (Fig. 5e) and noise propagation (*i.e.*, how the noise of eBFP2-MS2CP impacts on the noise of sfGFP; Fig. 5f). Together, these results indicated that the protein translation factor is a suitable element to control expression and that the cell-to-cell variability emerged at this level can be predicted with certain accuracy.

### 3.2. Noise analysis in transcription and translation regulation

To further analyze the stochastic behavior, we looked inside the noise. That is, we inspected how a particular noise level is achieved. For that, we first decomposed the total noise of both eBFP2-MS2CP and sfGFP into three fundamental components: extrinsic noise, intrinsic noise, and regulation noise. Extrinsic noise comes from replication and variability in the cellular machinery, intrinsic noise is a consequence of a low number of molecules, and regulation noise accounts for the noise that is propagated from the regulator to the regulated gene [3,7,25]. In previous work, the regulation noise has been considered as a part of the extrinsic noise.

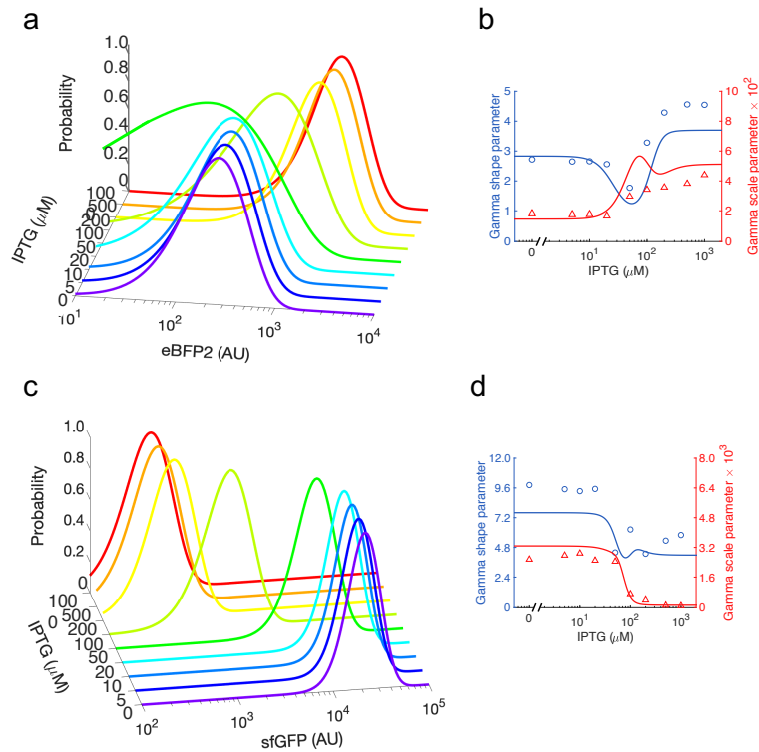
Here, we separate this component to study more in detail the stochastic gene expression when it is regulated. Assuming independence between the different stochastic sources, we were able to end with compact mathematical expressions for the noise in which the different components were identified, although at the cost of introducing some inaccuracies since the extrinsic noise may correlate responses. Along the IPTG gradient and according to our mathematical model, the extrinsic noise of the system is constant, the intrinsic noise decreases in the case of eBFP2-MS2CP and increases in the case of sfGFP (as this noise scales inversely with the expression level), and the regulation noise follows a peak-like curve (Fig. 7a-h).

Even though for both eBFP2-MS2CP and sfGFP the functional form of the regulation noise is similar, peaking at 50-75  $\mu\text{M}$  IPTG, the maximal noise level is much lower (about four times) in the case of sfGFP. This suggested that with a translational control the noise of the regulator is buffered, *i.e.*, the fluctuations of MS2CP expression are manifested on sfGFP expression in lower extent than the fluctuations of LacI expression on MS2CP expression. This is because in a scenario of translational control the regulated gene is constantly transcribed at high levels, where fluctuations in the number of mRNAs per cell are small in comparison with the mean quantity that is produced.



**Fig. 7** **a)** Model-based calculation of the total noise in eBFP2 expression as a function of IPTG. On top, schematics of the amplification effect by the transcriptional regulation. **b-d)** Decomposition of the total eBFP2 noise into extrinsic, intrinsic, and regulation noise components. **e)** Model-based calculation of the total noise in sfGFP expression as a function of IPTG. On top, schematics of the buffering effect by the post-transcriptional regulation. **f-h)** Decomposition of the total sfGFP noise into extrinsic, intrinsic, and regulation noise components. Insets in c,g) show the scaling of the intrinsic noise with the mean expression.

Thus, the transcriptional noise is not significant. In addition, the regulation enters at the level of translation, which prevents the typical amplification process of the noise of the regulator that occurs with a transcriptional control [26]. Indeed, in such a scenario, the transcription rate can be quite low when the promoter is repressed, thereby leading to substantial fluctuations in mRNA amount in comparison with the mean production. Furthermore, in the post-transcriptional case, the fluctuations in mRNA abundance can partly be absorbed by the effect of the translation factor, controlling the number of mRNAs available for translation. This has already been discussed in the case of regulatory RNAs [27], but it also applies to the case of a protein translation factor. Consequently, we can argue that the noise in the regulated gene is reduced when the regulation occurs at the level of translation. only the particular noise value. In addition, we aimed at predicting the shape of the whole distribution of protein expression and not To this end, we considered a Gamma distribution, which has been shown to describe quite well the stochasticity of genetic systems [28] and which emerges from *ab initio* calculations [29].



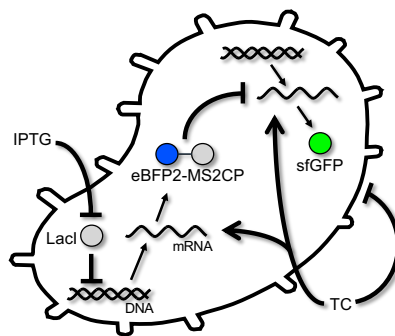
**Fig. 8** **a)** Predicted eBFP2 fluorescence distribution for different induction conditions with IPTG (Gamma distributions). **b)** Gamma shape and scale parameters for eBFP2. **c)** Predicted sfGFP fluorescence distribution for different induction conditions with IPTG (Gamma distributions). **d)** Gamma shape and scale parameters for sfGFP. In plots b,d points correspond to calculations from the experimental values of mean and noise, while solid lines to predictions with the mathematical model.

The distribution of protein expression is instrumental to appreciate the degree of heterogeneity in the production with time and from cell to cell (assuming ergodicity). Here, by defining the Gamma shape parameter as the inverse of the noise (equal to the mean square divided by the variance) and the Gamma scale parameter as the product between the mean and the noise (*i.e.*, an effective Fano factor), we were able to predict the distributions for both eBFP2-MS2CP and sfGFP (Fig. 8a,d). This was done with the values of mean expression and noise given by the mathematical model. As a result of a transcriptional control, the Gamma scale parameter for eBFP2-MS2CP depends on the translation rate; but in the case of sfGFP, as this element is controlled at the level of translation, the Gamma scale parameter is nearly independent of that rate. Importantly, these theoretical

distributions were very close to those fitted directly against the experimental data (data not shown), although some discrepancies were observed between the data and the model at intermediate IPTG concentrations. Overall, this highlighted the generality of the Gamma distribution to describe genetic systems regulated at both transcriptional and translational levels.

### 3.3. Examination of global effects on regulated gene expression

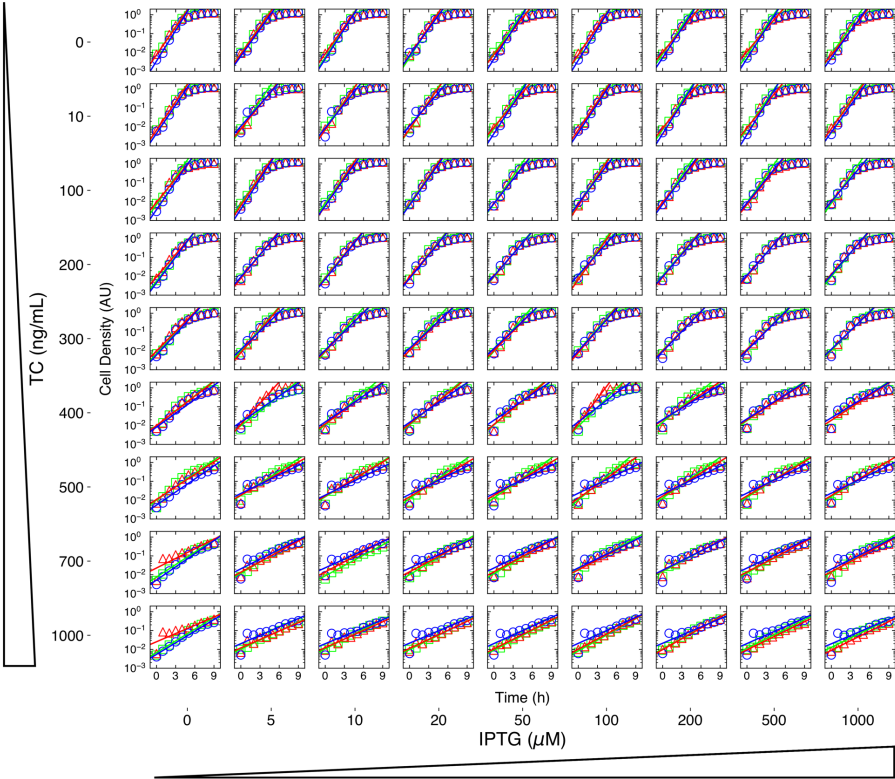
Subsequently, we decided to study how global perturbations can impact the single-cell response of the system. To this end, we considered the effect of a global signal affecting translation. Here, we used sublethal concentrations of tetracycline (TC), a bacteriostatic antibiotic known to inhibit the formation of active ribosomes (Fig. 9) [30].



**Fig. 9** Extended schematics of the gene regulatory system in which TC further modulates it through its negative impact on translation rate (global effects).

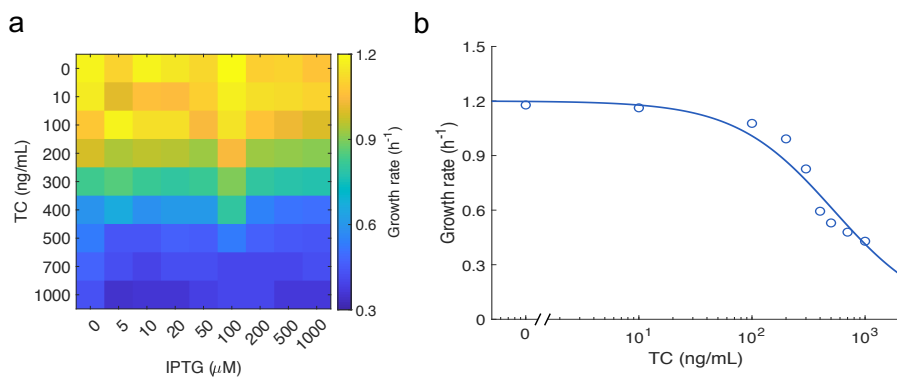
Paradoxically, this inhibition leads to an increase in translation rate as a result of an over-production of ribosomes (a global response mechanism in bacteria against this type of antibiotics) [31,32]. That is, the cell is able to sense that a substantial amount of ribosomes is being inhibited upon binding to TC and produces more. In particular, TC binds to the 30S subunit and interferes with the transfer RNAs (tRNAs). In turn, the cell growth rate is compromised due to the action of TC. Importantly, this parameter has been shown to modulate the mean and noise of gene expression [33,34], so we decided to exploit it as a predictor variable. Over a two-

dimensional concentration gradient of IPTG and TC (81 conditions), we first generated growth curves (Fig. 10).



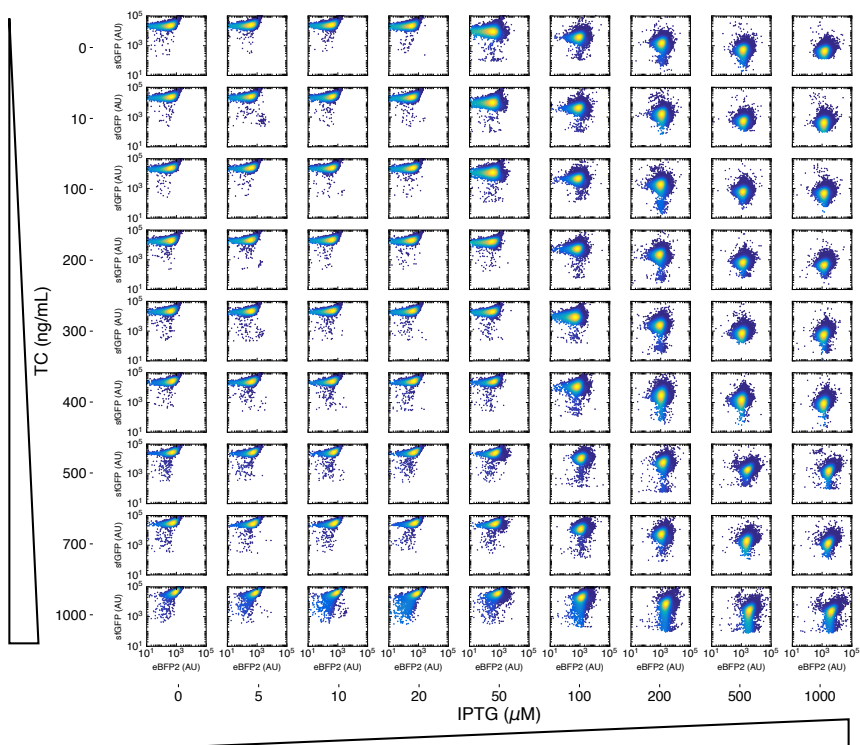
**Fig. 10** Predicted growth curves. Three different populations (blue, red, and green) were monitored with time. Points correspond to absorbance values, while solid lines come from fitted exponential trends.

Basically, only TC showed a significant impact on growth rate, with a maximal reduction of almost 3-fold, which was well explained by a Michaelis-Menten function (Fig. 11a,b).



**Fig. 11** a) Heatmap of the mean growth rate as a function of IPTG and TC. b) Dose-response curve between growth rate and TC. Points correspond to experimental data, while solid line comes from the mathematical model.

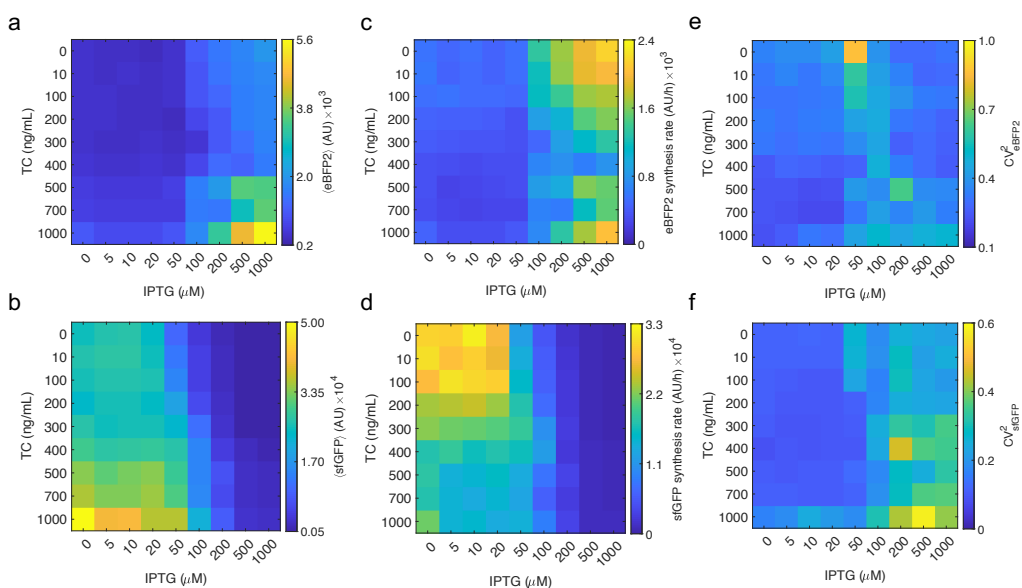
In parallel, we performed single-cell measurements of blue and green fluorescence for each condition (Fig. 12).



**Fig. 12** Projected two-dimensional histograms of single-cell fluorescence for eBFP2 and sfGFP for different induction conditions with IPTG and TC.

We observed that the mean expression levels of both eBFP2-MS2CP and sfGFP remained almost constant at low TC concentrations, but they increased

significantly from 500 ng/mL TC, irrespective of the induction with IPTG (Fig. 13a,b). Because protein expression comes from the ratio between the protein synthesis rate (accounting for both transcription and translation) and the growth rate (in the case of stable proteins, as it is the case here), this indicated that the protein synthesis rate of both eBFP2-MS2CP and sfGFP scales with the growth rate (Fig. 12c,d). It was interesting to note here the logical NOR behavior of the sfGFP synthesis rate and the difference between protein expression and synthesis rate. In addition, we calculated the noise levels for each condition (Fig. 13e,f). We observed that the regulation noise decreases with TC for both eBFP2-MS2CP and sfGFP, as well as that TC leads to a substantial increase in the sfGFP noise when this gene is fully repressed by MS2CP.



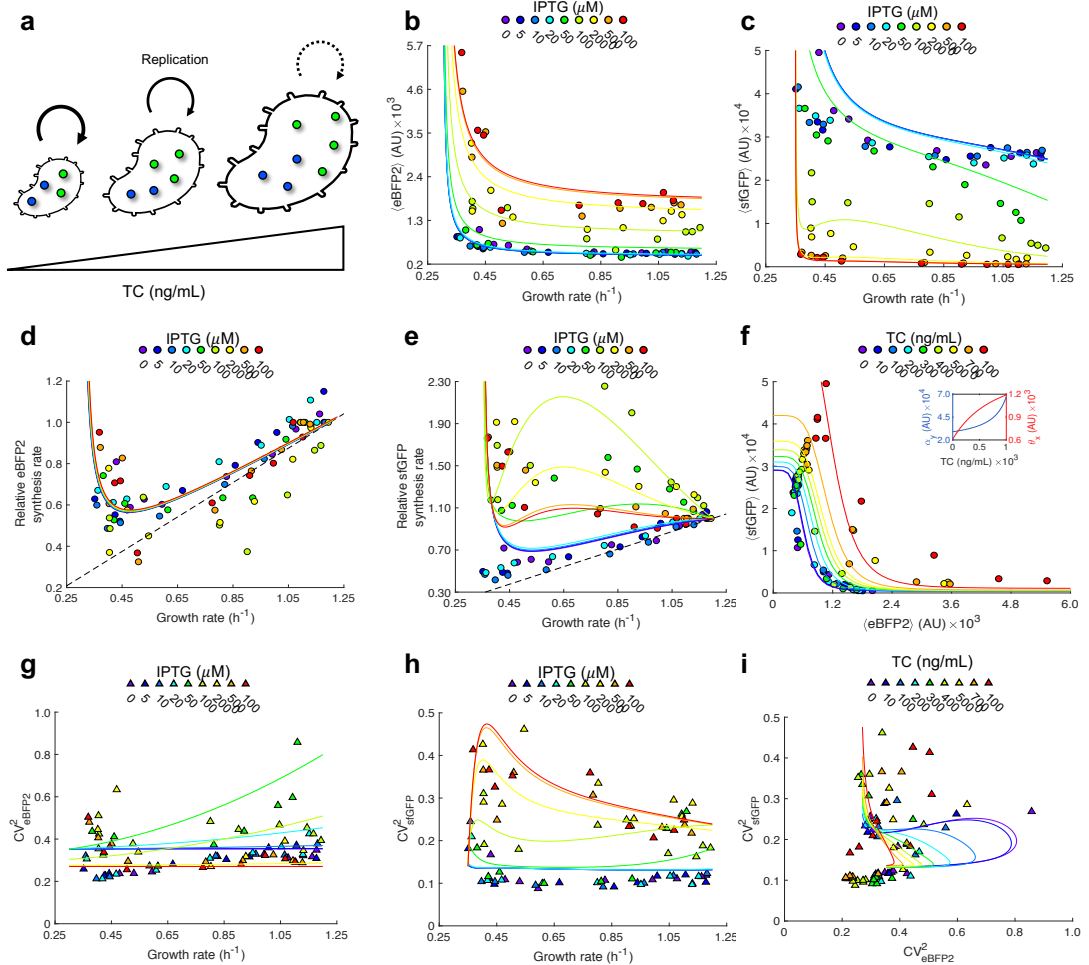
**Fig. 13** **a)** Heatmap of the mean eBFP2 fluorescence as a function of IPTG and TC. **b)** Heatmap of the mean sfGFP fluorescence as a function of IPTG and TC. **c)** Heatmap of the mean eBFP2 synthesis rate as a function of IPTG and TC. **d)** Heatmap of the mean sfGFP synthesis rate as a function of IPTG and TC. **e)** Heatmap of the sfGFP noise as a function of IPTG and TC.



### 3.4. Integrative modeling of the deterministic and stochastic dynamics

Importantly, we noted that the cell volume increases as a consequence of TC (Fig. 14a, which agrees with previous observations [35]). This entails the necessity of a higher number of MS2CP molecules per cell to repress the target gene. Moreover, it is known that the number of total ribosomes decreases linearly with the growth rate when it is modulated by an antibiotic (*i.e.*, the slower the replication, the larger the number of ribosomes) [31]. This leads to a reciprocal function with the growth rate to describe the translation rate of a given gene [32]. That is, the translation rate scales inversely with the growth rate. It is also known that the transcription rate (mRNA production) scales linearly with the growth rate [33]. In terms of number of proteins per cell, this effect is cancelled out by the effective dilution due to cell division. By introducing into our mathematical model these dependencies, we were able to predict with relatively good agreement the impact of TC on mean expression for both eBFP2-MS2CP and sfGFP (Fig. 14b,c).

To further comprehend the interplay between gene regulation and cell growth, we used the model to predict the protein synthesis rate. Remarkably, we found that for eBFP2-MS2CP the same curve is explicative for all induction conditions with IPTG, provided the values are relativized to the case of no TC (Fig. 13d). This is because eBFP2-MS2CP is regulated transcriptionally, and in turn the transcription factor of the system (LacI) is expressed constitutively and modulated by IPTG in terms of activity at the post-translational level. Hence, there is a decoupling between the regulation and the effect of growth rate on expression. A minimum in eBFP2-MS2CP synthesis rate was found at a growth rate of about  $0.45 \text{ h}^{-1}$ , which comes from the fact that the protein synthesis rate is modeled by a rational function with the growth rate. In essence, at low TC concentrations, the transcription rate is reduced, but the translation rate remains almost constant. However, at high TC concentrations, the translation rate very much increases and dominates over the transcription rate.



**Fig. 14** Detailed analysis of stochastic gene expression modulated by growth rate. **a)** As TC increases, cells grow slower and bigger. **b)** Mean eBFP2 expression as a function of growth rate for each IPTG condition. **c)** Mean sfGFP expression as a function of growth rate for each IPTG condition. **d)** Relative mean eBFP2 synthesis rate as a function of growth rate for each IPTG condition. **e)** Relative mean sfGFP synthesis rate as a function of growth rate for each IPTG condition. In **d,e)**, the values are relative to the case TC = 0, and the dashed line corresponds to a linear dependence (which comes from a null model in which the translation rate is not affected by the growth rate). **f)** Transfer function of the post-transcriptional regulation in terms of mean expression for each TC condition. The inset shows the effect of TC on the sfGFP mRNA synthesis rate (included in  $\alpha_y$ , as the growth rate decreases) and the effective dissociation constant between eBFP2-MS2CP and sfGFP mRNA ( $\theta_x$ , as the volume increases) according to the mathematical model. **g)** eBFP2 noise as a function of growth rate for each IPTG condition. **h)** sfGFP noise as a function of growth rate for each IPTG condition. **i)** Transfer function of the post-transcriptional regulation in terms of noise for each TC condition. In plots **b-i)**, points correspond to calculations from the experimental data, while solid lines to predictions with the mathematical model.

By contrast, we found that the relative sfGFP synthesis rate strongly depends on IPTG and that this dependence is well captured by the model (Fig. 14e). In this case, the number of MS2CP molecules per cell changes with IPTG, so there is a coupling between the translational regulation and the effect of growth rate on expression. While at low IPTG concentrations the relative sfGFP synthesis rate follows the aforementioned trend for eBFP2-MS2CP, at high IPTG concentrations there is a maximum at a growth rate of about  $0.65 \text{ h}^{-1}$  (it was particularly pronounced at the intermediate value of  $100 \mu\text{M}$ ). Fig. 14f illustrates how the transfer function in terms of mean expression varies with the TC concentration (*i.e.*, by increasing the maximal expression level and shifting the inflexion point towards the right).

Finally, we applied the model to project the noise in protein expression. In this case, we needed to introduce a phenomenological dependence with the growth rate on three noise-related parameters to explain the data. In particular, we set that the noise in LacI expression scales with the square of the growth rate (*i.e.*, LacI expression varies from cell to cell in greater extent when cells grow faster) and that the extrinsic noise of both eBFP2-MS2CP and sfGFP scales with the inverse of the growth rate by following the translation rate (*i.e.*, the extrinsic noise is higher at lower growth rates, which seems in tune with recent experiments characterizing genome-wide noisy expression levels [36]). While for eBFP2-MS2CP the noise decreases or remains constant with the growth rate (Fig. 14g), for sfGFP the noise presents a more complex trend (almost constant at low IPTG concentrations and with a maximum at high IPTG concentrations; Fig. 14h). In turn, Fig. 14i illustrates how the transfer function in terms of noise varies with the TC concentration, showing how the belly shape is reduced with TC, which indicates that noise propagation through the translation factor is less significant (*i.e.*, it is masked) when the growth rate is low. That is, the regulation noise term, which quantifies how much fluctuation sfGFP perceives from MS2CP, becomes smaller than the other noise terms (intrinsic and extrinsic) with TC. Arguably, at high growth rates, when global perturbations are small, a translation factor is superior to a transcription factor because it is able to

regulate gene expression without transmitting much noise. However, this is not the case at low growth rates, when global perturbations become substantial, due to a poor signal-to-noise ratio. Together, these results highlight the complex impact of growth rate in the system and serve to appreciate how global and local regulatory mechanisms interplay in the cell.

## 4. DISCUSSION

Our development follows previous work on exploiting RNA-binding proteins as translation factors to engineer gene circuits [15]. In this work, we focused on quantitatively studying the stochastic behavior of this type of circuits (*i.e.*, noise generation and propagation in gene expression when it is regulated at the level of translation). Our results show that the protein-RNA interaction in this case leads to a significant down-regulation in expression of about 50-fold by blocking the progression of the ribosome on the target mRNA, which is comparable to transcriptional fold-changes. In addition, a general mathematical framework was shown suitable to describe the stochastic behavior in regulations exerted by both transcription and translation factors. Noise propagation from gene to gene is buffered when the regulator acts at the level of translation, as the amplification process by transcription is avoided, and a Gamma distribution properly parameterized can provide deep analytical explanations about the resulting cell-to-cell variability [28]. By modulating the cellular growth rate, we also reported an interplay between global and local regulatory mechanisms in the cell that affect both the mean expression and noise levels. It is important to notice that the growth rate that we measured here corresponds to an average of the population (as we calculate it having monitored absorbance with time for a culture). Nevertheless, each cell grows differently,

especially when the culture is under the effect of TC. In this regard, a mathematical model incorporating such heterogeneous growth might explain better the observed noise patterns [37]. Another limitation of our work is the assumption that the noise sources are independent, which allowed us to derive a compact mathematical expression for the noise of sfGFP. This is not strictly true since eBFP2-MS2CP and sfGFP share extrinsic noise sources [7]. Yet, we expect reliability in the conclusions derived from this study.

Here, we exploited the viral protein MS2CP to implement the regulatory system, but in principle other RNA-binding proteins might be used. For example, the bacteriophage PP7 coat protein or the *Mycobacterium* enzyme PyrR [16] are suitable elements from which to engineer orthogonal systems. In fact, given the plethora of RNA-binding proteins in nature, especially in eukaryotes [38], and noting that the regulatory mechanism only requires a tight protein-RNA interaction to interfere with the ribosome, multiple implementations might be achieved.

Our mathematical model is general enough to describe these eventual implementations. We only expect to change the kinetic parameters for each particular protein, preserving the functional form. In principle, each RNA motif will lead to a different translation rate. In this regard, the predictability of the system might be strengthened by using the RBS calculator [39]. We also anticipate that the use of tandem repeats of the RNA motif might enhance the regulatory fold-change of the system. Furthermore, since RNA is a very versatile molecule, RNA-binding proteins can regulate gene expression through a variety of mechanisms acting post-transcriptionally, including the regulation of translation initiation, translation elongation, transcription termination, and RNA stability [40]. In prokaryotes, the regulation of translation initiation by controlling RBS accessibility is a widespread mechanism, but in eukaryotes the blockage of translation elongation has been observed in the case of Argonaute proteins [41]. Arguably, a blockage in the initial phase by MS2CP is key in our synthetic system. Further work should analyze how those other mechanisms generate and propagate noise.

In sum, our work provides new quantitative insights about the stochastic behavior in genes regulated translationally by RNA-binding proteins. Protein translation factors can integrate some advantages distinctively attributed to proteins (as transcription factors), such as the ability to transduce small signals and achieve large dynamic ranges, or to small RNAs, such as the ability to produce rapid responses and buffer transcriptional noise [27]. Furthermore, our work paves the way for engineering gene regulatory circuits with greater integrability and then sophistication. Certainly, the combination of different layers within the genetic information flow (*i.e.*, transcription and translation) leads to an easier integration of signals to achieve a given function [42]. Therefore, we envision that RNA-binding proteins will be of great utility in synthetic biology in the close future to face biotechnological and biomedical challenges.

## REFERENCES

1. Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* 12: 204-213.
2. Balázsi G, van Oudenaarden A, Collins JJ (2011) Cellular decision-making and biological noise: from microbes to mammals. *Cell* 144: 910-925.
3. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99: 12795-12800.
4. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol* 2: e137.
5. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467: 167-173.
6. Beckett D (2001) Regulated assembly of transcription factors and control of transcription initiation. *J Mol Biol* 314: 335-352.
7. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307: 1965-1969.
8. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4: e309.
9. Libby E, Perkins TJ, Swain PS (2007) Noisy information processing through transcriptional regulation. *Proc Natl Acad Sci USA* 104: 7151-7156.
10. Dunlop MJ, Cox RS, Levine JH, Murray RM, Elowitz MB (2008) Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat Genet* 40: 1493-1498.
11. Siciliano V, Garzilli I, Fracassi C, Criscuolo S, Ventre S, di Bernardo D (2013) MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nat Commun* 4: 2364.
12. Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Blüthgen N, Marks DS, van Oudenaarden A (2015) MicroRNA control of protein expression noise. *Science* 348: 128-132.
13. Dacheux E, Malys N, Meng X, Ramachandran V, Mendes P, McCarthy JE (2017) Translation initiation events on structured eukaryotic mRNAs generate gene expression noise. *Nucleic Acids Res* 45: 6981-6992.
14. Rodrigo G (2018) Post-transcriptional bursting in genes regulated by small RNA molecules. *Phys Rev E* 97: 032401.
15. Katz N, Cohen R, Solomon O, Kaufmann B, Atar O, Yakhini Z, Goldberg S, Amit R (2019) Synthetic 5' UTRs can either up- or downregulate expression upon RNA-binding protein binding. *Cell Syst* 9: 93-106.
16. Babitzke P, Baker CS, Romeo T (2009) Regulation of translation initiation by

- RNA binding proteins. *Annu Rev Microbiol* 63: 27-44.
17. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. *Mol Cell* 2: 437-445.
  18. Graindorge A, Pinheiro I, Nawrocka A, Mallory AC, Tsvetkov P, Gil N, Carolis C, Buchholz F, Ulitsky I, Heard E, Taipale M, Shkumatava A (2019) In-cell identification and measurement of RNA-protein interactions. *Nat Commun* 10: 5317.
  19. Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160: 339-350.
  20. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333: 470-474.
  21. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25: 1203-1210.
  22. Ai HW, Shaner NC, Cheng Z, Tsien RY, Campbell RE (2007) Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* 46: 5904-5910.
  23. Pédelaq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24: 79-88.
  24. Lim F, Peabody DS (1994) Mutations that increase the affinity of a translational repressor for RNA. *Nucleic Acids Res* 22: 3748-3752.
  25. Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature* 439: 861-864.
  26. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6: 451-464.
  27. Mehta P, Goyal S, Wingreen N (2008) A quantitative comparison of sRNA-based and protein-based gene regulation. *Mol Syst Biol* 4: 221.
  28. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329: 533-538.
  29. Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett* 97: 168302.



30. Wilson DN (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol* 12: 35-48.
31. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330: 1099-1102.
32. Borkowski O, Goelzer A, Schaffer M, Calabre M, Mäder U, Aymerich S, Jules M, Fromion V (2016) Translation elicits a growth rate-dependent, genome-wide, differential protein production in *Bacillus subtilis*. *Mol Syst Biol* 12: 870.
33. Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139: 1366-1375.
34. Keren L, van Dijk D, Weingarten-Gabbay S, Davidi D, Jona G, Weinberger A, Milo R, Segal E (2015) Noise in gene expression is coupled to growth rate. *Genome Res* 25: 1893-1902.
35. Harris LK, Theriot JA (2016) Relative rates of surface and volume synthesis set bacterial cell size. *Cell* 165: 1479-1492.
36. Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, van Nimwegen E (2021) Genome-wide gene expression noise in *Escherichia coli* is condition-dependent and determined by propagation of noise through the regulatory network. *PLoS Biol* 19: e3001491.
37. Ray JCJ, Wickersheim ML, Jalihal AP, Adeshina YO, Cooper TF, Balázsi G (2016) Cellular growth arrest and persistence from enzyme saturation. *PLoS Comput Biol* 12: e1004825.
38. Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume AM, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J, Hentze MW (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* 6: 10127.
39. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27: 946-50.
40. van Assche E, van Puyvelde S, Vanderleyden J, Steenackers PH (2015) RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol* 6: 141.
41. Iwakawa HO, Tomari Y (2013) Molecular insights into microRNA-mediated translational repression in plants. *Mol Cell* 52: 591-601.
42. Rosado A, Cordero T, Rodrigo G (2018) Binary addition in a living cell based on riboregulation. *PLoS Genet* 14: e1007548.
43. Rodrigo G, Carrera J, Jaramillo A (2011) Computational design of synthetic regulatory networks from a genetic library to characterize the designability of dynamical behaviors. *Nucleic Acids Res* 39: e13



## CHAPTER 2

### **Repurposing the mammalian RNA-binding protein Musashi-1 as an allosteric translation repressor in bacteria**

This work has been published in a peer-reviewed journal.

See the full citation:

Dolcemascolo R, Heras-Hernández M, Goiriz L, Montagud-Martínez R, Requena-Menéndez A, Ruiz R, Pérez-Ràfols A, Higuera-Rodríguez RA, Pérez-Ropero G, Vranken WF, Martelli T, Kaiser W, Buijs J, Rodrigo G (2023) Repurposing the mammalian RNA-binding protein Musashi-1 as an allosteric translation repressor in bacteria. *eLife*, in press.

<https://doi.org/10.7554/eLife.91777.1>

My contribution to this work:

I made most of the genetic constructions used in this work. I performed the functional characterization experiments under the supervision of G. Rodrigo. I performed the dynamic analysis during my secondment in Sweden. I also contributed to the data analysis, the figures preparation, and the manuscript writing.



# 1. INTRODUCTION

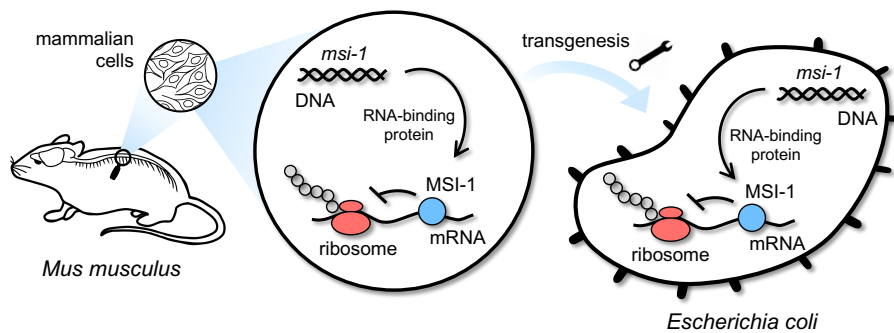
Gene regulation at the post-transcriptional level is pervasive in living organisms of ranging complexity [1-4]. Indeed, the ability to regulate the genetic information flow at different points appears instrumental to maximize the integration of intrinsic and extrinsic signals, which enables an efficient information processing by the organisms. However, the solutions implemented in prokaryotes and eukaryotes greatly differ. In prokaryotes, small RNAs (sRNAs) regulate messenger RNA (mRNA) stability and translation initiation [1], supported by a series of RNA-binding proteins that act globally [2]. Regulatory proteins of specific scope in these simple organisms mainly operate in the transcriptional layer [5], what is aligned with the models presented in the early times of molecular biology [6]. By contrast, eukaryotes deploy a sizeable number of RNA-binding proteins with a variety of functions [4] that participate in the regulation of mRNA turnover, transport, splicing, and translation in a gene-specific manner and also at a global scale. In animals, in particular, most RNA-binding proteins contain RNA recognition motifs (RRMs) [7]. RRM are small globular domains of about 90 amino acids that fold into four antiparallel  $\beta$ -strands and two  $\alpha$ -helices, which can bind to single-strand RNAs with sufficient affinity and specificity to control biological processes [8].

Yet, while important to attain functional diversity in the post-transcriptional layer in animals, RRM are not prevalent in all organisms. In fact, the scarcity of RRM-containing proteins in prokaryotes and the often-unknown functional role of those identified by bioinformatic methods [9] question if RRM can readily work in organisms with much simpler gene expression machinery and intracellular organization. If so, this would raise the potential to use RRM-RNA interactions as an orthogonal layer to engineer gene regulation in prokaryotes. To address these intriguing questions, we adopted a synthetic biology approach where a specific RRM-containing protein was incorporated in a bacterium in order to engineer a post-

transcriptional control module. Synthetic biology has highlighted how living cells can be (re)programmed through the assembly of independent genetic elements into functional networks for a variety of applications in biotechnology and biomedicine [10]. Yet, synthetic biology can also be used to disentangle natural systems and to probe hypotheses about biological function [11].

In previous work, some proteins with the ability to recognize RNA have been exploited as translation factors in bacteria for a gene-specific regulation [12-14]. The first instance was the tetracycline repressor protein (TetR), which naturally functions as a transcription factor, by means of the selection of synthetic RNA aptamers [12]. The bacteriophage MS2 coat protein (MS2CP) [13] and eukaryotic Pumilio homology domains [14] were also used in synthetic circuits. Alternatively, a wide palette of post-transcriptional control systems based on sRNAs have been developed in recent years to program gene expression in bacteria [15]. Of note, these systems are amenable to be combined with regulatory proteins to attain complex dynamic behaviors [16]. A heterologous RRM-containing protein with definite regulatory activity, in addition to provide empirical evidence on the adaptability of such RNA-binding domains to different genetic backgrounds, would enlarge the synthetic biology toolkit [17], boosting applications in which high orthogonality, expression fine-tuning, and signal integrability are required features. In addition, RRMs can themselves be allosterically regulated, opening up new avenues for post-transcriptional regulation by small molecules.

In this work, the mammalian RNA-binding protein Musashi-1 (MSI-1) [18] was used as a translation repressor in the bacterium *Escherichia coli* (Fig. 1). MSI-1 belongs to an evolutionarily conserved family of RRM-containing proteins, of which a member was first identified in *Drosophila melanogaster* [19]. MSI-1 contains two RRMs in the N-terminal region (RRM1 and RRM2) and recognizes the RNA consensus sequence  $RU_nAGU$  on the nanomolar affinity scale [20]. Importantly, MSI-1 can be allosterically inhibited by fatty acids (in particular, 18-22-carbon w-9 monounsaturated fatty acids) [21].



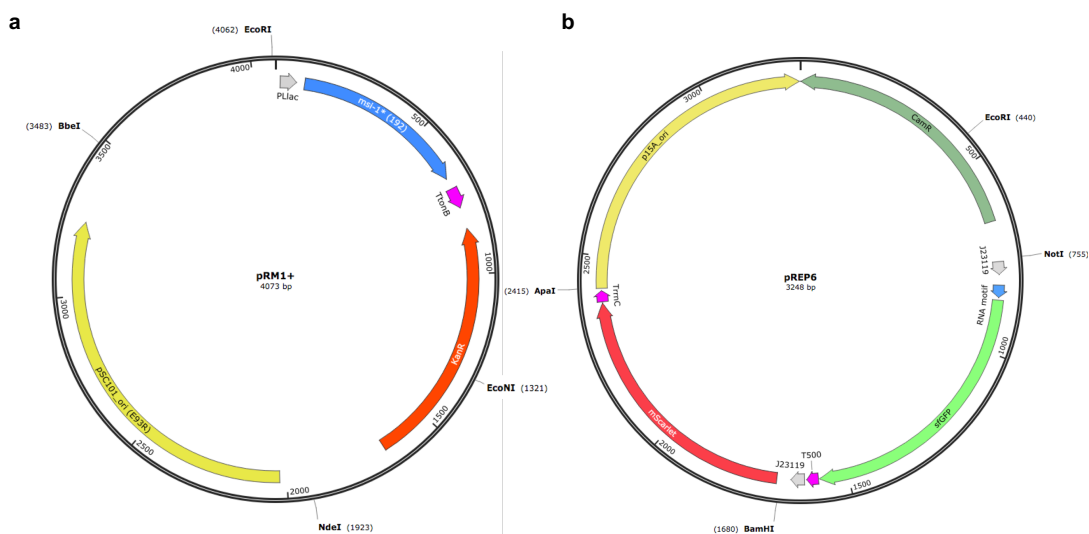
**Fig. 1** Overview of the biotechnological development. In mammals, MSI-1 binds to the 3' UTR of its target mRNA to repress translation. Here, the *M. musculus* gene coding for MSI-1 was moved to *E. coli* (transgenesis) to implement a synthetic regulation system at the level of translation.

In mammals, MSI-1 is mainly expressed in stem cells of neural and epithelial lineage and plays crucial roles in differentiation, tumorigenesis, and cell cycle regulation [18]. Notably, MSI-1 regulates Notch signaling by repressing the translation of a key protein in the pathway [20]. Hence, rather than moving genetic elements from simple to complex organisms, as it is normally done (*e.g.*, the TetR-aptamer module was implemented in simple eukaryotes [22]), we reversed the path by moving an important mammalian gene (from *Mus musculus*) to *E. coli*. Some eukaryotic factors have already been implemented in bacteria to regulate gene expression at different levels [14,23], but the case of RRM-containing proteins has remained elusive. In the following, we present quantitative experimental and theoretical results on the response dynamics of a synthetic gene circuit in which MSI-1 works as an allosteric translation repressor. There, MSI-1 is transcriptionally controlled by the lactose repressor protein (LacI), and translation regulation by MSI-1 is accomplished by means of a specific interaction with an mRNA (encoding a reporter protein) that harbors a suitable binding motif in its leader coding region.

## 2. MATERIALS AND METHODS

### 2.1. Strains, plasmids, and reagents

*E. coli* Dh5a was used for cloning purposes following standard procedures. To express our genetic circuit for functional characterization, *E. coli* MG1655-Z1 cells (*lacI*<sup>+</sup>, *tetR*<sup>+</sup>) were used. This strain was co-transformed with two plasmids, called pRM1+ (KanR, pSC101-E93R ori; leading to ~230 copies/cell) [60] and pREP6 (CamR, p15A ori; leading to ~15 copies/cell) (Fig. 2).



**Fig. 2** Maps of the plasmids used to implement the synthetic gene circuit in which MSI-1\* represses the translation of sfGFP. **a)** Map of pRM1+ to express the MSI-1\* protein from a PLlac promoter, induced with lactose or IPTG. **b)** Map of pREP6 to express the reporter sfGFP protein from a constitutive promoter (J23119), harboring a suitable RNA motif in the leader region for translation regulation.

On the one hand, pRM1+ was obtained by cloning a truncated coding region of the *M. musculus* MSI-1 protein (the first 192 amino acids, containing the two RRM; UniProt #Q61474; termed MSI-1\*). This gene was under the transcriptional control of the inducible promoter PLlac. On the other hand, pREP6 was obtained by cloning the coding region of sfGFP with an RNA sequence motif recognized by MSI-1. The coding region of mScarlet was also present in the plasmid. These two genes



were under the control of the constitutive promoter J23119 in two different transcriptional units.

**Table 1.** List of plasmids used in this work.

name	insert feature	backbone feature	reference
pRM1+	PLlac: <i>msi-1*</i>	KanR, pSC101(E93R) ori	this work
pREP6	J23119: <i>sfGFP</i> (with RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP6-mut1	J23119: <i>sfGFP</i> (with mutated RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP6-mut2	J23119: <i>sfGFP</i> (with mutated RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP6-mut3	J23119: <i>sfGFP</i> (with mutated RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP6-mut4	J23119: <i>sfGFP</i> (with mutated RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP6-mut5	J23119: <i>sfGFP</i> (with mutated RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP7	J23119: <i>sfGFP</i> (with RNA motif for MSI-1* binding and consensus sequences within RBS)	CamR, p15A ori	this work
pREP4	J23119: <i>sfGFP</i> (with minimal RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP4b	J23119: <i>sfGFP</i> (with less structured RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pREP4b3x	J23119: <i>sfGFP</i> (with 3x less structured RNA motif for MSI-1* binding)	CamR, p15A ori	this work
pRKFR2	PLlac: <i>eBFP2</i>	KanR, pSC101(E93K) ori	[34]
pGio	T7p: <i>msi-1*</i>	KanR, pUC ori	this work

To perform the dynamic assays with LigandTracer (Ridgeview), *E. coli* BL21(DE3) cells (*lacI<sup>+</sup>*, *T7pol<sup>+</sup>*) were used. This strain was also co-transformed with pRM1+ and pREP6. In addition to the original RNA sequence motif, five point-mutated sequences were designed and cloned in pREP6 (Fig. 6a). Additional RNA sequence motifs were cloned in front of *sfGFP* for control experiments (the resulting plasmids were named pREP4, pREP4b, pREP4b3x, and pREP7). In particular, pREP4b3x incorporates three RNA motifs in tandem after the start codon, and

pREP7 has two R<sub>U</sub>NAGU repeats flanking the RBS and a full RNA motif after the start codon. Suitable genetic cassettes to obtain the final constructions were synthesized by IDT. Table 1 lists all plasmids used in this work.

To purify a recombinant Musashi protein, *E. coli* BL21-Gold(DE3) cells (*lacI*<sup>+</sup>, *T7pol*<sup>+</sup>) were used. A truncated coding region of the human MSI-1 protein (the first 200 amino acids; UniProt #O43347; also termed MSI-1\* abusing of notation) was cloned under the control of a T7pol promoter into the plasmid pET29b (KanR, pUC ori). Luria-Bertani (LB) medium was used for the overnight cultures and M9 minimal medium (1X M9 minimal salts, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.05% thiamine, 0.05% casamino acids, 1% glycerol or 0.4% glucose) for the characterization cultures. M9-glucose medium was only used for real-time fluorescence quantification in liquid medium with IPTG. LB-agar was used for real-time fluorescence quantification in solid medium. Kanamycin and chloramphenicol were used at a concentration of 50 µg/mL and 34 µg/mL, respectively. Lactose and IPTG were used as the inducers of the system (controlling the expression of MSI-1\* in *E. coli*) at a concentration of 5, 10, 20, 50, 100, 200, 500, or 1000 µM. Oleic acid was used as the allosteric inhibitor of MSI-1\* at a concentration of 20 mM in the *in vivo* assays (both in liquid and solid medium). In the *in vitro* assays, oleic acid was used at a concentration of 0.01, 0.1, 0.2, 0.5, 0.7, 1, 1.5, or 2 mM. It was neutralized with NaOH and used in a medium containing 0.5% tergitol NP-40.

Compounds provided by Merck.

## 2.2. Bulk fluorometry

Cultures (2 mL) inoculated from single colonies (in triplicate) were grown overnight in LB medium with shaking (220 rpm) at 37 °C. Cultures were then diluted 1:100 in fresh M9 medium (200 µL) with the appropriate inducer (lactose or IPTG). The microplate (96 wells, black, clear bottom; Corning) was incubated with shaking (1300 rpm) at 37 °C up to 8-10 h (to reach an OD<sub>600</sub> around 0.5-0.7). At different

times, the microplate was assayed in a Varioskan Lux fluorometer (Thermo) to measure absorbance (600 nm), green fluorescence (excitation: 485 nm, emission: 535 nm), and red fluorescence (excitation: 570 nm, emission: 610 nm). To characterize the time-course response of the system, cultures were grown to exponential phase and then diluted before adding the inducer (to minimize the response lag). Mean background values of absorbance and fluorescence, corresponding to M9 medium, were subtracted to correct the signals. Normalized fluorescence was calculated as the slope of the linear regression between fluorescence and absorbance (assuming fluorophore maturation faster than cell doubling time and no proteolytic degradation) [61]. The mean value of normalized fluorescence corresponding to non-transformed cells was then subtracted to obtain a final estimate of expression. In addition, cell growth rate was calculated as the slope of the linear regression between the logarithm of background-subtracted absorbance and time in the exponential phase.

### **2.3. Real-time fluorescence quantification in solid medium**

Cultures (2 mL) inoculated from single colonies (in triplicate) were grown overnight in LB medium with shaking (220 rpm) at 37 °C. The overnight culture was plated (15 µL) in areas A and D of a MultiDish 2x2 plate (Ridgeview) coated with LB-agar. IPTG was added in areas A and B of the dish at the final concentration of 1 mM. Area C was kept free of cells/inducers as a reference. The dish was then placed in the rotating support of the LigandTracer instrument (Ridgeview) and incubated at 37 °C for 24 h. The fluorescence from sfGFP and mScarlet was quantified with time in the seeded areas of the dish using the BlueGreen (excitation: 488 nm, emission: 535 nm) and OrangeRed (excitation: 568 nm, emission: 620 nm) detectors. The readouts of the opposite parts of the dish were subtracted to correct the signals.

## **2.4. Flow cytometry**

Cultures (2 mL) inoculated from single colonies (three replicates) were grown overnight in LB medium with shaking (220 rpm) at 37 °C. Cultures were then diluted 1:100 in fresh LB medium (200 µL) to load a microplate (96 wells, black, clear bottom; Corning) with the appropriate concentrations of lactose (0, 100, 1000 µM) and oleic acid (0, 20 mM). The microplate was then incubated with shaking (1300 rpm) at 37 °C until cultures reached a sufficient OD<sub>600</sub>. Cultures (6 µL) were then diluted in PBS (1 mL). Fluorescence was measured in an LSRFortessa flow cytometer (BD) using a 488 nm laser and a 530 nm filter for green fluorescence. Events were gated by using the forward and side scatter signals and compensated (~10<sup>4</sup> events after this process). The mean value of the autofluorescence of the cells was subtracted to obtain a final estimate of expression. Data analysis performed with MATLAB (MathWorks).

## **2.5. Purification of a Musashi protein**

Cells were grown in LB medium with shaking at 37 °C until OD<sub>600</sub> reached 0.6-0.8. Subsequently, the expression of MSI-1\* was induced with 0.5 mM IPTG. Cells were incubated at 37 °C for 4 h and harvested by centrifugation at 7500 rpm for 15 min at 4 °C. The cell pellet was resuspended in a lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol, with protease inhibitor cocktail), ruptured by sonication, and separated by centrifugation at 30,000 rpm for 35 min at 4 °C. The soluble fraction was collected and treated with a 5% polyethylenimine solution in order to remove DNA/RNA attached to the protein. Resuspension of the protein was done in 20 mM Tris-HCl, pH 9.0, with protease inhibitor cocktail. Soluble protein was filtered with a 0.22 µm membrane and purified by ion exchange chromatography using an Anion exchange Q FF 16/10 column previously equilibrated in alkaline buffer. The protein was collected on the flow-through. The protein was filtered and further purified to homogeneity by size exclusion chromatography using a Hi load 26/60 Superdex 75

pg column previously equilibrated in alkaline buffer with NaCl. The purified fractions were collected and buffer exchange chromatography was performed using a HiPrep 26/10 Desalting column previously equilibrated with the final buffer (20 mM MES, pH 6.0, 100 mM NaCl, 0.5 mM EDTA, with protease inhibitor cocktail). Purification performed at Giotto.

## **2.6. Binding kinetics assays of protein-RNA interactions**

Binding experiments of a purified MSI-1\* protein against different RNA ligands were performed using the switchSENSE proximity sensing technology [33,62] and a suitable adapter chip on the heliX biosensor platform (Dynamic Biosensors). The adapter chip consists of a microfluidic channel with two gold electrodes functionalized with fluorophore-decorated DNA nanolevers that serve as linkers between the gold surface and the ligand of interest. A constant negative voltage is applied to the electrodes to keep the DNA nanolevers in an upright position. Binding between the injected analyte (MSI-1\*) and the ligand attached to the sensor surface (RNA) leads to the alteration of the chemical surrounding of the dye, which results in a fluorescence change. Fluorescence change of the dye in real time describes the binding kinetics of the molecule of interest. Kinetic experiments consisted of a protein association phase (5 min) and a dissociation phase (15 min) in which the chip was rinsed with a buffer (50 mM Tris-HCl, 0.5 mM EDTA, 140 mM NaCl, 0.05% Tween 20, 1 mM TCEP, pH 7.2). A flow rate of 100  $\mu$ L/min was applied and a sampling rate of 1 Hz was used. Six different RNA ligands (original and 5 mutants) were attached to the 5' end of a generic 48 nt DNA ligand strand, which is part of the DNA linker system on the heliX adapter chip surface. All oligonucleotides were synthesized by Ella Biotech. The ligand strand was hybridized with an adapter strand carrying the fluorophore. Different fluorophores were tested towards their sensitivity for protein-RNA interactions. The green fluorophore Gb showed the most significant signal change. The other half of the adapter strand is complementary to a DNA anchor strand, which is pre-attached to the chip surface. The immobilization of the

RNA used a standard functionalization procedure on the heliX device. Kinetic rate constants and affinities were obtained by fitting the experimental data with theoretical binding models implemented in the heliOS software (Dynamic Biosensors). Exponential decay models were used. As a negative control to check for unspecific protein-RNA binding, the single-strand RNA sequence CGGCGCCGC was used (without any binding motif). All data were referenced with a blank run and with the negative control.

## 2.7. Gel electrophoresis

Mobility shift assays with a purified MSI-1\* protein and its cognate RNA motif were performed. The RNA motif was generated by *in vitro* transcription with the TranscriptAid T7 high yield transcription kit (Thermo) from a DNA template. It was then purified using the RNA clean and concentrator column (Zymo) and quantified in a NanoDrop. Bovine serum albumin (BSA) was used as a control protein (at 30  $\mu$ M). Reactions with different combinations of elements were prepared (MSI-1\* at 45  $\mu$ M, RNA at 11  $\mu$ M, and oleic acid at 1 mM). Reactions with concentration gradients of MSI-1\* (from 0 to 45  $\mu$ M) and oleic acid (from 0 to 2 mM) were also performed. Reactions were incubated for 30 min at 37 °C. Reaction volumes were then loaded in 3% agarose gels prepared with 0.5X TBE and stained using RealSafe (Durviz). Gels ran for 45 min at room temperature applying 110 V. The GeneRuler ultra-low range DNA ladder (10-300 bp, Thermo) was used. This staining served to reveal the RNA and oleic acid (free or in complex with the MSI-1\* protein) [63,64]. In addition, gels were soaked for 10 min in the Coomassie blue stain (Fisher) at room temperature with shaking to reveal the proteins. Gels were then soaked in a destaining solution overnight to remove the excess of blue stain. Pictures were taken with the Imager2 gel documentation system (VWR).

## 2.8. Microscopy

LB-agar plates seeded with *E. coli* MG1655-Z1 cells co-transformed with pRM1+ and pREP6 or pREP7 were grown overnight at 37 °C. Lactose (1 mM) and oleic acid (20 mM) were used as supplements. The plates were irradiated with blue light and images were acquired with a 2.8 Mpixel camera with a filter for green fluorescence in a light microscope (Leica MSV269). The commercial software provided by Leica was used to adjust the visualization of the differential fluorescence among plates. The fluorescence intensity of the colonies was quantified with Fiji [65].

## 2.9. Mathematical modeling

On the one hand, Hill equations were used to empirically model sfGFP expression with lactose/IPTG, eBFP2 expression with lactose, and sfGFP expression with eBFP2 expression. On the other hand, a system of ordinary differential equations was developed to model the dynamic response of the synthetic gene circuit from a bottom-up approach. The system accounted for the intracellular mRNA and protein concentrations, considering a scenario of equilibrium to model both LacI-DNA and MSI-1\*-RNA binding. Parameter values were obtained by nonlinear fitting against our experimental data.

## 2.10. Molecular visualization *in silico*

The RMM1 of MSI-1 protein structure determined by nuclear magnetic resonance was downloaded from the UniProt database ([www.uniprot.org](http://www.uniprot.org)) [66]. A 3D structure of the RNA motif subsequence involving the two  $RU_nAGU$  repeats was predicted with the RNAComposer software [67]. The oleic acid molecule was downloaded from the ChemSpider database ([www.chemspider.com](http://www.chemspider.com)). All the molecules were loaded, visualized, colored, trimmed (where necessary), and manually docked using the open source PyMol software (Schrödinger; [pymol.org](http://pymol.org)).

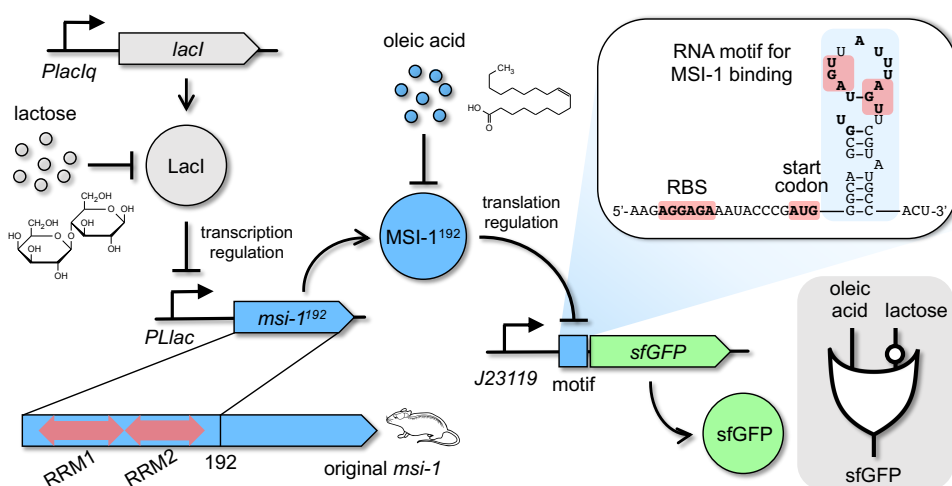
## 3. RESULTS

### 3.1. A Musashi protein can down-regulate translation in bacteria

From the amino acid sequence of *M. musculus* MSI-1, we generated a nucleotide sequence with codons optimized for *E. coli* expression. Knowing that the C-terminus of MSI-1 is of low structural complexity [24], we cloned a truncated version of the gene encompassing the first 192 amino acids, which include the two RRM, to implement our synthetic circuit (Fig. 3). The resulting protein (termed MSI-1\*) was expressed from a synthetic PL-based promoter repressed by LacI (termed PLlac) [25] lying in a high copy number plasmid. This allowed controlling the expression of the heterologous RNA-binding protein at the transcriptional level with lactose or isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) in a genetic background over-expressing LacI. As a regulated element, we used the superfolder green fluorescent protein (sfGFP) [26], which was expressed from a constitutive promoter (J23119) lying in a low copy number plasmid. An RNA motif obtained by affinity elution-based RNA selection (SELEX) containing two copies of the consensus recognition sequence (*viz.*, GUUAGU and AUUUAGU) [20] was placed in frame after the start codon of *sfGFP*. This motif folds into a stem-loop structure that allows stabilizing the exposure of the recognition sequence to the solvent. In this way, MSI-1\* can block the progression of the ribosome on the regulated gene in the initial phase. This mode of action differs from the natural one in mammals, in which MSI-1 binds to the 3' untranslated region of its target mRNA (*Numb*) to repress translation by disrupting the activation function of the poly(A)-binding protein [27].

Here, considering lactose (or IPTG) and oleic acid as the two inputs and sfGFP as the output, MSI-1\* being an internal allosteric regulator operating at the post-transcriptional level, an IMPLY gate would model the logic behavior of the resulting circuit (*i.e.*, sfGFP would only turn off with lactose and without oleic acid in the medium).





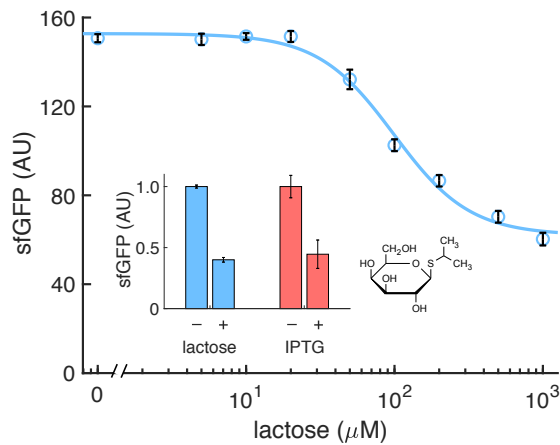
**Fig. 3** Schematic of the synthetic gene circuit engineered in *E. coli*. A truncated version of MSI-1 (termed MSI-1\*) was expressed from the PLac promoter to be induced with lactose (or IPTG) in a genetic background over-expressing LacI. sfGFP was used as a reporter expressed from a constitutive promoter (J23119) and under the control of a suitable RNA motif recognized by MSI-1\* in the leader coding region of the transcript (*viz.*, located after the start codon). The activity of MSI-1\* could in turn be allosterically inhibited by oleic acid. In electronic terms, this circuit implements an IMPLY logic gate. The inset shows the predicted secondary structure of the leader region of the reporter mRNA. Within the motif (blue shaded), the consensus recognition sequences (RU<sub>n</sub>AGU) are bolded and the minimal cores (UAG) are marked in red. System implemented with pRM1+ and pREP6.

We first characterized by bulk fluorometry the dose-response curve of the system using a lactose concentration gradient up to 1 mM. Our data show that MSI-1\* down-regulated sfGFP expression by 2.5-fold (Fig. 4). The repression of sfGFP as a function of lactose was modelled by the following Hill equation

$$(\text{sfGFP}) = \frac{A_1}{1 + \left(\frac{(\text{Lactose})}{K_1}\right)^{n_1}} + B_1,$$

where  $K_1$  is the regulatory coefficient,  $n_1$  the Hill coefficient,  $A_1 + B_1$  the maximal expression level, and  $B_1$  the basal expression level at full repression. In the case of sfGFP, its concentration is given by the normalized green fluorescence signal in arbitrary units (AU). The adjusted parameter values are  $A_1 = 90.7$  AU,  $B_1 = 62.1$  AU,  $K_1 = 99.1$   $\mu\text{M}$ , and  $n_1 = 1.70$ . Fitting a Hill equation, we obtained a regulatory

coefficient of 99  $\mu\text{M}$  (lactose concentration at which the repression is half of the maximal) and a Hill coefficient of 1.7.

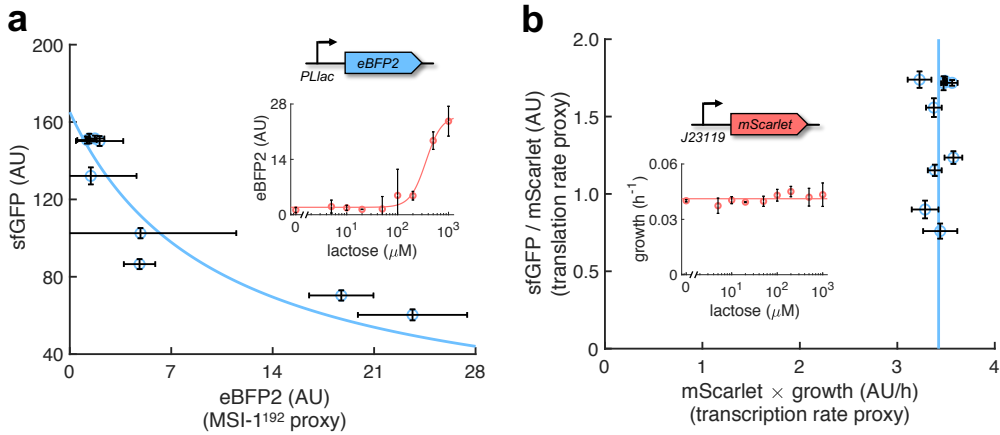


**Fig. 4** Dose-response curve of the system using lactose as inducer (up to 1 mM). The inset shows the dynamic range of the response using lactose or IPTG (1 mM).

We also observed that IPTG (a synthetic compound) triggered a very similar response (Fig. 4, inset). Also, the following Hill equation models the repression of sfGFP by IPTG:

$$(\text{sfGFP}) = \frac{A_2}{1 + \left(\frac{(\text{IPTG})}{K_2}\right)^{n_2}} + B_2,$$

To further inspect the activity of the RNA-binding protein, we filtered out the transcriptional regulatory effect. For that, we expressed the enhanced blue fluorescent protein 2 (eBFP2) [28] from the PLlac promoter to obtain the corresponding dose-response curve with lactose. In this way, eBFP2 expression was a proxy of MSI-1\* expression, which allowed representing the transfer function of the engineered regulation (Fig. 5a).

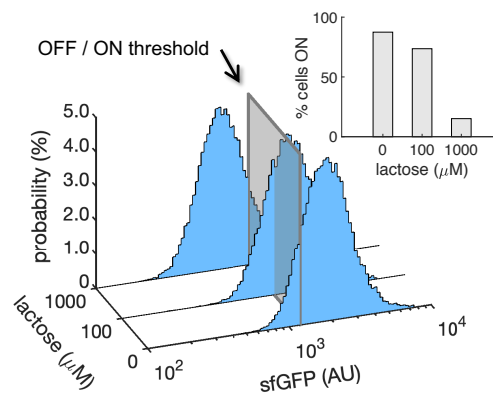


**Fig. 5 a)** Transfer function of the system (between sfGFP and MSI-1\*). The inset shows the dose-response curve of eBFP2 expressed from the PLlac promoter (proxy of MSI-1\* expression) with lactose. **b)** Scatter plot of the dynamic response of the system in the Crick space (translation rate *vs.* transcription rate). The dose-response curve of mScarlet expressed from the J23119 promoter with lactose was used to perform the decomposition (vertical line fitted to 48 AU/h). The inset shows the growth rate of the cells for each induction condition (horizontal line fitted to 0.55 h<sup>-1</sup>). In all cases, points correspond to experimental data, while solid lines come from adjusted mathematical models. Error bars correspond to standard deviations ( $n = 3$ ).

A Hill equation with no cooperative binding (*i.e.*, Hill coefficient of 1) explained the data with sufficient agreement, suggesting that only one protein interacted with a given mRNA (*i.e.*, each RRM of MSI-1\* binds to a consensus sequence repeat, in agreement with a previous structural model [24]). We also measured the cell growth rate for all induction conditions, finding that the values were almost constant (Fig. 5b, inset). This indicates that the expression of the mammalian protein did not produce a significant burden to the bacterial cell. In simple terms, protein expression comes from the product of the transcription and translation rates of the gene.

Hence, we examined such a decomposition in the case of sfGFP expression regulated by MSI-1\*. Of note, the low copy number plasmid harbors an additional transcriptional unit to express the monomeric red fluorescent protein mScarlet [29] from a constitutive promoter (J23119). We then monitored its expression profile with lactose. Assuming that *sfGFP* and *mScarlet* were equally transcribed, as they were expressed from the same promoter, and that the translation rate of *mScarlet* was

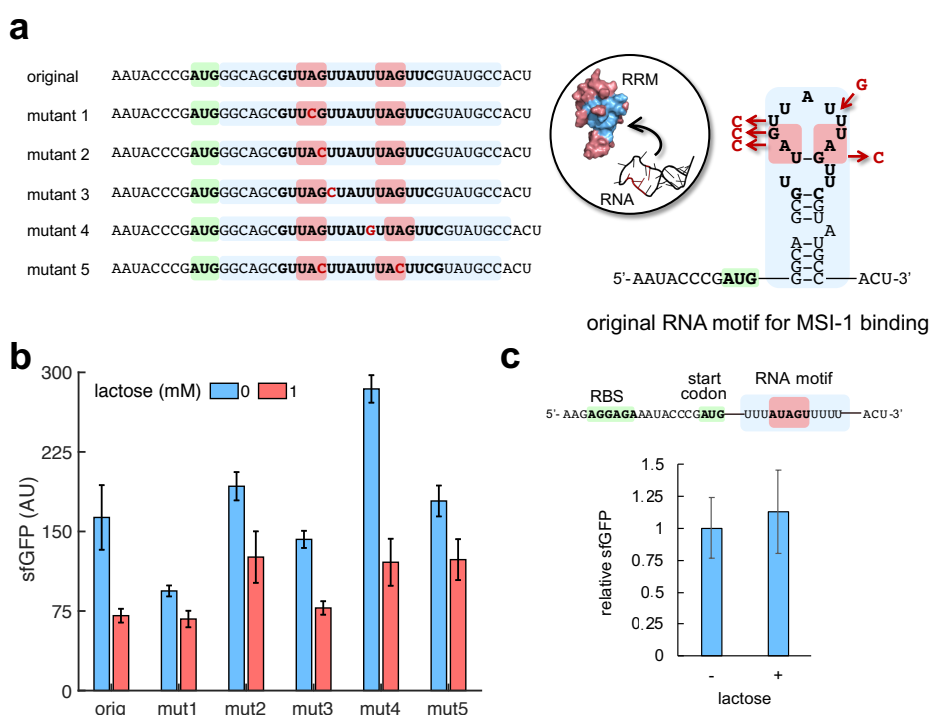
constant, the product of mScarlet expression and cell growth rate was considered a proxy of the transcription rate of *sfGFP*. Moreover, the ratio of sfGFP and mScarlet expressions was a proxy of the translation rate of *sfGFP* [30]. This served us to represent the dynamics of the system in a plane defined as translation rate vs. transcription rate (termed Crick space [31]), highlighting that the change in sfGFP expression with lactose comes indeed from translation regulation (Fig. 5b). Finally, to evaluate the heterogeneity of the response within a bacterial population, we performed single-cell measurements of sfGFP expression by flow cytometry. Unimodal distributions able to shift in response to lactose were observed (Fig. 6). Setting a threshold to categorize expression, we found that the percentage of cells in the ON state dropped from 87% to 15% upon addition of 1 mM lactose (Fig. 6, inset). In sum, our results show that MSI-1\* can regulate translation in a specific manner in *E. coli*, and hence that eukaryotic regulators can be borrowed to be functional elements in prokaryotes.



**Fig. 6** Probability-based histograms of sfGFP expression from single-cell data for different lactose concentrations. The inset shows the percentage of cells in the ON state (sfGFP expressed), according to a specified threshold, for each lactose concentration. AU, arbitrary units.

### 3.2. Mechanistic insight into the engineered regulation based on a protein-RNA interaction

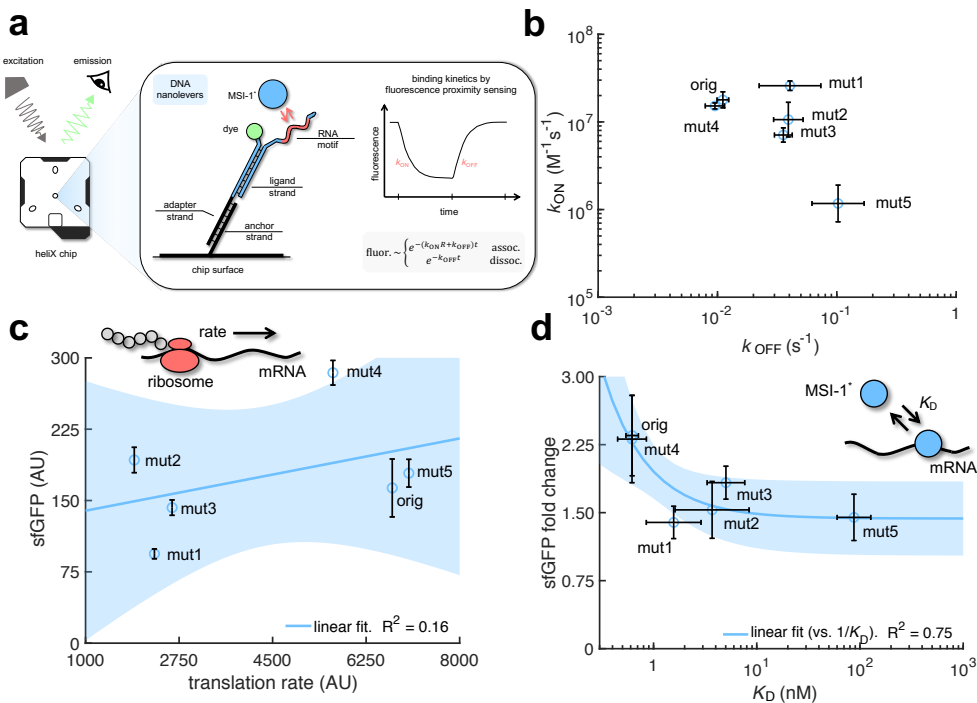
We then introduced a series of point mutations into the SELEX RNA motif to assess their effect over the regulatory activity of the RRM-containing protein (Fig. 7a,9a). These mutations change the consensus recognition sequence of at least one repeat. A characterization of all systems revealed that the mutations affected both the maximal level and fold change of sfGFP expression (Fig. 7b).



**Fig. 7 a)** Sequences and predicted secondary structures of the different RNA motif variants for MSI-1 binding analyzed in this work. Point-mutations indicated in red. Three-dimensional representations of the RRM1 and RNA motif are also shown. Within the RRM1, the region that recognizes the RNA is shown in blue. **b)** Dynamic range of the response of the different genetic systems using lactose (1 mM). **c)** Characterization of the system response with lactose using pREP4 as a reporter plasmid (induction with 1 mM lactose). Error bars correspond to standard deviations ( $n = 3$ ). On the top, sequence of the leader RNA of sfGFP in this case.

Of note, a single point mutation in one repeat leading to  $RU_nCGU$  (mutant 1) was quite detrimental for the MSI-1\*-based regulation (only 1.4-fold reduction in sfGFP expression). This agrees with the prior observation that,

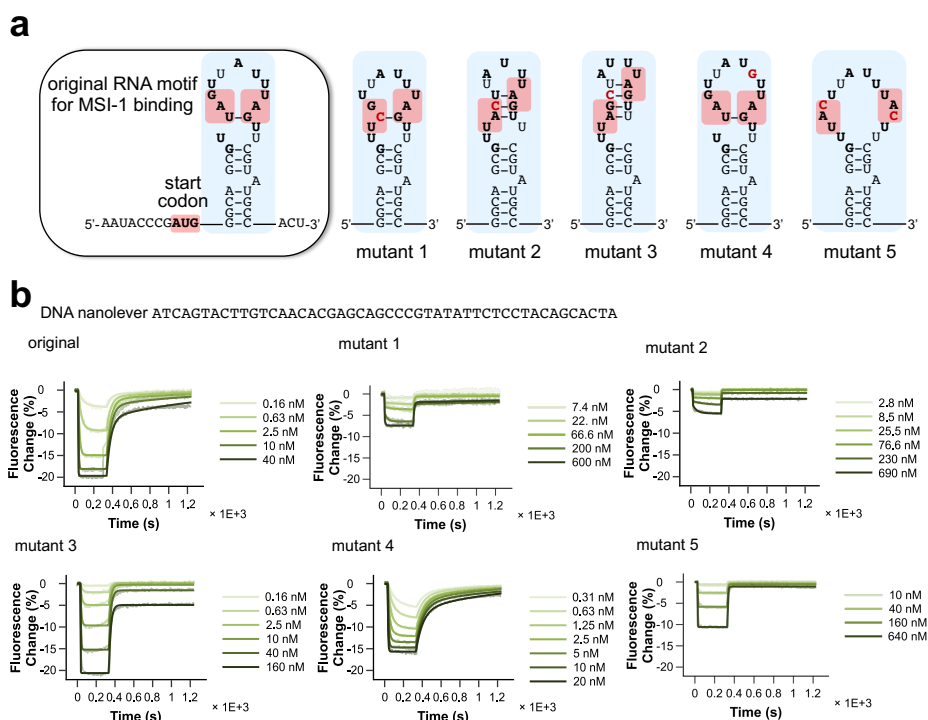
within the consensus sequence, UAG is a minimal core that determines the specific recognition by MSI-1 [32].



**Fig. 8** Schematics of the heliX biosensor platform. **a** A double-strand DNA nanolever was immobilized on a gold electrode of the chip. The nanolever carried a fluorophore in one end and the RNA motif for MSI-1 binding in the other. Binding between MSI-1\* (injected analyte) and RNA led to a fluorescence change, whose monitoring in real time served to extract the kinetic constants that characterize the interaction. **b** Scatter plot of the experimentally-determined kinetic constants of association and dissociation between the protein and the RNA for all systems (original and 5 mutants). Means and deviations calculated in log scale (geometric). **c** Correlation between the maximal sfGFP expression level (in absence of lactose) and the translation rate predicted by RBS calculator. Linear regression performed. **d** Correlation between the fold change in sfGFP expression and the dissociation constant ( $K_D$ ). Deviations calculated by propagation. Linear regression performed (vs.  $1/K_D$ ). Blue shaded areas indicate 95% confidence intervals. In all cases, error bars correspond to standard deviations ( $n = 3$ ). AU, arbitrary units.

A double point mutation changing the minimal cores of the two repeats (UAC rather than UAG; mutant 5) also resulted in a detrimental action, but not to a greater extent. We also engineered a new reporter system with a minimal RNA motif consisting of a single copy of the shortest possible consensus sequence (AUAGU), but its characterization showed no apparent regulation by MSI-1\* (Fig. 7c). Taken together, two copies of the consensus sequence seem necessary for a successful regulation of protein expression.

To relate the cellular effects with protein-RNA interactions, we obtained a purified MSI-1\* preparation in order to perform *in vitro* binding kinetics assays. For that, a gene coding for a truncated version of the human MSI-1 was expressed from a T7 polymerase promoter in *E. coli*. With respect to the *M. musculus* version, this protein only differs in one residue of RRM2, which is the subsidiary domain for RNA recognition (note also that the human and mouse proteins recognize the same consensus sequence [32]; data not shown).



**Fig. 9** Characterization of different mutant RNA motifs in terms of binding kinetics against the MSI-1\* protein. **a**) Predicted secondary structures of original and mutant RNA motifs; mutations are marked in red. **b**) Binding and unbinding kinetic curves for the different RNA sequences (representative samples). The kinetic constants were extracted from mono-exponential model fits. For the original, mutant 1, mutant 2, and mutant 4 RNA motifs, a bi-exponential model was also explored to describe the binding kinetics, reflecting two types of interactions, although with little improvement.

To avoid the necessity of labelling the molecules of interest and allow working with very low amounts of protein and RNA, we used the switchSENSE technology, which allows measuring molecular dynamics on a chip (Fig. 8a) [33]. Fig. 9b summarizes the resulting protein-RNA association and dissociation rates ( $k_{\text{ON}}$  and  $k_{\text{OFF}}$ , respectively; see also Table 2).

**Table 2.** Inferred kinetic constants ( $k_{\text{ON}}$ ,  $k_{\text{OFF}}$ ) and the resulting dissociation constant ( $K_{\text{D}}$ ) for each sequence and replicate.

RNA motif ligand	replicate	$k_{\text{on}}$ ( $\text{M}^{-1}\text{s}^{-1}$ )	$k_{\text{off}}$ ( $\text{s}^{-1}$ )	$K_{\text{D}}$ (nM)
orig	1	$(23.600 \pm 0.400) \times 10^6$	$(10.10 \pm 0.10) \times 10^{-3}$	$0.427 \pm 0.007$
orig	2	$(17.000 \pm 0.300) \times 10^6$	$(13.00 \pm 0.10) \times 10^{-3}$	$0.760 \pm 0.010$
orig	3	$(14.200 \pm 0.400) \times 10^6$	$(10.40 \pm 0.10) \times 10^{-3}$	$0.728 \pm 0.010$
mut1	1	$(22.600 \pm 2.400) \times 10^6$	$(33.10 \pm 1.40) \times 10^{-3}$	$1.470 \pm 0.170$
mut1	2	$(25.100 \pm 3.500) \times 10^6$	$(22.10 \pm 0.90) \times 10^{-3}$	$0.880 \pm 0.100$
mut1	3	$(30.600 \pm 3.900) \times 10^6$	$(90.90 \pm 1.40) \times 10^{-3}$	$2.970 \pm 0.380$
mut2	1	$(16.700 \pm 2.000) \times 10^6$	$(40.80 \pm 2.20) \times 10^{-3}$	$2.440 \pm 0.3200$
mut2	2	$(12.600 \pm 3.300) \times 10^6$	$(27.20 \pm 2.90) \times 10^{-3}$	$2.160 \pm 0.600$
mut2	3	$(5.700 \pm 1.710) \times 10^6$	$(54.40 \pm 3.20) \times 10^{-3}$	$9.550 \pm 2.930$
mut3	1	$(5.700 \pm 0.210) \times 10^6$	$(45.40 \pm 0.50) \times 10^{-3}$	$7.970 \pm 0.300$
mut3	2	$(9.000 \pm 0.200) \times 10^6$	$(31.50 \pm 0.20) \times 10^{-3}$	$3.510 \pm 0.100$
mut3	3	$(6.950 \pm 0.530) \times 10^6$	$(31.60 \pm 1.00) \times 10^{-3}$	$4.540 \pm 0.370$
mut4	1	$(14.600 \pm 0.000) \times 10^6$	$(8.80 \pm 0.04) \times 10^{-3}$	$0.600 \pm 0.008$
mut4	2	$(17.000 \pm 0.400) \times 10^6$	$(12.30 \pm 0.10) \times 10^{-3}$	$0.723 \pm 0.018$
mut4	3	$(14.200 \pm 0.300) \times 10^6$	$(7.96 \pm 0.07) \times 10^{-3}$	$0.550 \pm 0.014$
mut5	1	$(0.596 \pm 0.000) \times 10^6$	$(60.70 \pm 2.50) \times 10^{-3}$	$102.000 \pm 11.000$
mut5	2	$(1.540 \pm 0.190) \times 10^6$	$(87.60 \pm 3.60) \times 10^{-3}$	$57.000 \pm 7.300$
mut5	3	$(1.760 \pm 0.180) \times 10^6$	$(202.00 \pm 5.00) \times 10^{-3}$	$115.000 \pm 12.00$

In the case of the original RNA motif, we found an association rate of  $1.1 \text{ nM}^{-1}\text{min}^{-1}$ , which means that a single regulator molecule would take 1-3 min to find its target in the cell, and a residence time of the protein on the RNA of 1.5 min (given by  $1/k_{\text{OFF}}$ ). Of note, the reported value of  $k_{\text{ON}}$  is relatively close to the upper limit imposed by the diffusion rate ( $\sim 1 \text{ nM}^{-1}\text{s}^{-1}$ ). This fast rate suggests that MSI-1\* is able to find its target mRNA in *E. coli*, competing with ribosomes and ribonucleases, and then achieve translation regulation. We also found that a single



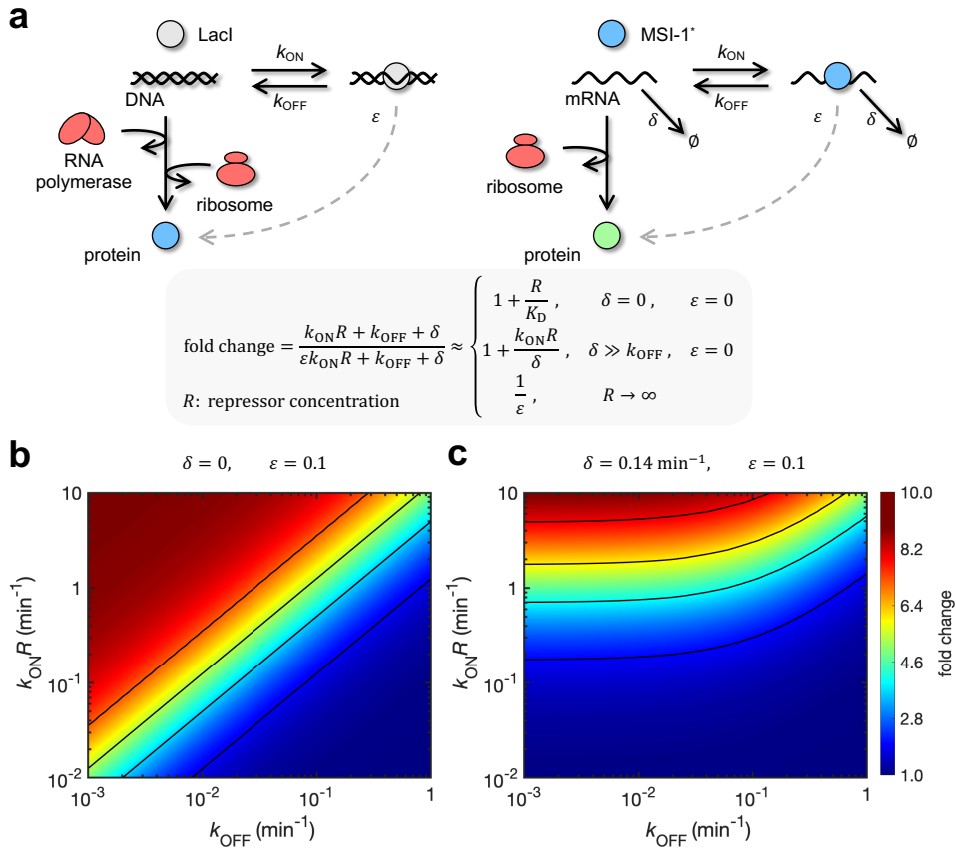
mutation in one of the two UAG minimal cores (mutants 1 and 2) led to similar association but faster dissociation (almost 4 times faster dissociation), whereas a double mutation affecting the two cores (mutant 5) disturbed both phases (almost 15 times slower association and 10 times faster dissociation). The dissociation constant ( $K_D = k_{\text{OFF}}/k_{\text{ON}}$ ) was 0.62 nM for the original system, while 87 nM for mutant 5. The switchSENSE technology allowed revealing that affinity on the subnanomolar scale, refining a previous estimate of 4 nM obtained by gel shift assays [20].

To contextualize these values, we compared to the binding kinetics of MS2CP, a phage RNA-binding protein that has evolved in a prokaryotic context and that we recently exploited to study how expression noise emerges and propagates through translation regulation [34]. Previous work disclosed an association rate to the cognate RNA motif of  $0.032 \text{ nM}^{-1}\text{min}^{-1}$  and a residence time of 12 min, leading to a dissociation constant of 2.6 nM [35]. Thus, MSI-1\* would target RNA faster than MS2CP, but once this happened the phage protein would remain bound longer. Next, we tried to predict the impact of the mutations on sfGFP expression. On the one hand, we used an empirical free-energy model (RBS calculator) to obtain an estimate of the mRNA translation rate from the sequence [36]. However, only a poor correlation ( $R^2=0.16$ ) with the maximal expression level was observed (Fig. 8c), suggesting that additional variables should be considered. For example, it was surprising the higher expression level in the case of mutant 4, despite a minimal change in the structure of the RNA motif (Fig. 9a). On the other hand, when the fold change was correlated with the inverse of the dissociation constant ( $1/K_D$ , *i.e.*, the equilibrium constant) better results were obtained ( $R^2 = 0.75$ ; Fig. 8d). Mutant 1 is illustrative in this case because, even though a fast association rate was preserved ( $1.6 \text{ nM}^{-1}\text{min}^{-1}$ ), it displayed a marginal regulatory activity as a result of a shorter residence time (0.41 min). This indicates that the underlying protein-RNA interaction in the bacterial circuit was close to thermodynamic equilibrium.

### 3.3. A mathematical model captured the dynamic response of the system

*Mathematical model and data analysis were performed in collaboration with the research team at our lab. Even though my contribution in this part is limited because I am not an expert in mathematical modelling, I believe it was necessary to add the following sections for the integrity of the thesis and relevance of the topic.*

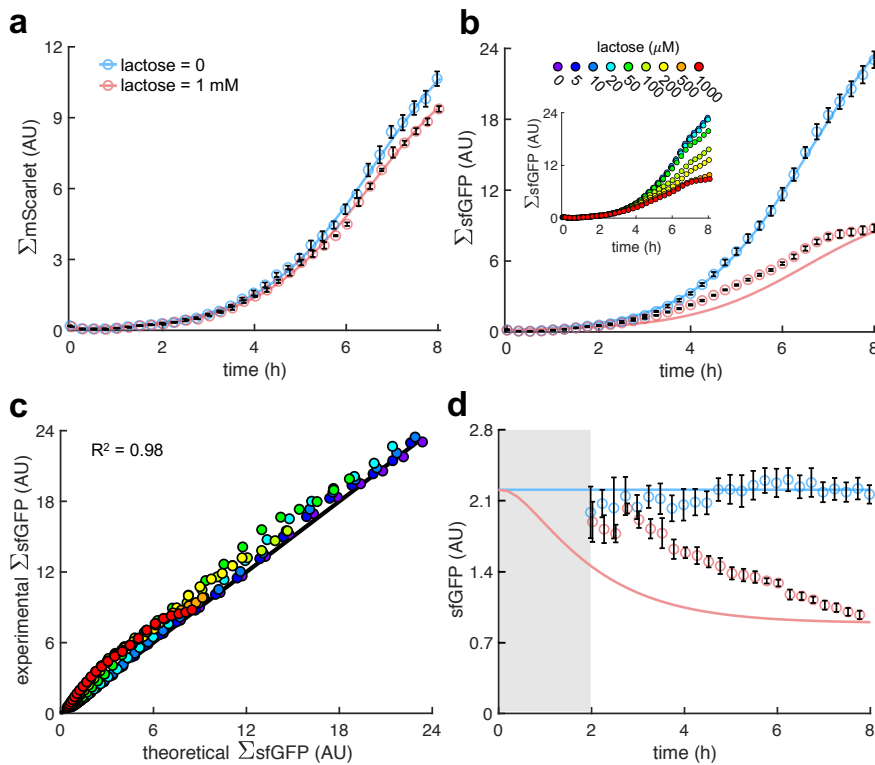
Translation regulation is more challenging than transcription regulation because mRNA is unstable compared to DNA, especially in bacteria. In *E. coli*, in particular, the average mRNA half-life is about 5 min [37]. However, it is possible to derive a common mathematical framework from which to analyze the dynamics of both regulatory modes (Fig. 10a). The fold change in protein expression is a suitable mesoscopic parameter that is directly related to the kinetic parameters that characterize the interaction in the cell [38]. Using mass action kinetics, we obtained a general mathematical description of the fold change as a function of the regulator concentration ( $R$ ), the association and dissociation rates, the leakage fraction of RNA/peptide-chain elongation, and the nucleic acid degradation rate (data not shown). To visualize the impact of the different parameters, we represented the fold change equation as a heatmap. When there is no nucleic acid degradation (DNA), a linear dependence between the first-order association rate ( $k_{\text{ON}}R$ ) and  $k_{\text{OFF}}$  is established to maintain a given fold change value (Fig. 10b), which would correspond to the case of transcription regulation. Accordingly, our model converges to the classical description of  $\text{fold} = 1 + R/K_D$ . However, if the nucleic acid degrades quickly (mRNA), the dependence between the first-order kinetic rates becomes nonlinear (Fig. 10c). Indeed, in the case of translation regulation, it is important to note that when  $k_{\text{ON}}R$  is lower than the mRNA degradation rate (*i.e.*, the mRNA is degraded faster than the protein binds), the functionality is greatly compromised. To overcome this barrier, the regulator needs to be highly expressed, as MSI-1\* is in our system (we estimate  $R > 1 \mu\text{M}$  with 1 mM lactose). Furthermore, when the residence time is much longer than the mRNA half-life (*i.e.*, the mRNA is degraded before the protein unbinds),  $K_D$  is not a suitable parameter to characterize the regulation, which is solely association-dependent, resulting in non-equilibrium thermodynamics [39].



**Fig. 10** A mathematical model captures the dynamic response of the system. **a)** Schematics of gene regulation at different levels with proteins that bind to nucleic acids (DNA or RNA). On the left, schematic of transcription regulation (*e.g.*, Lacl regulating MSI-1\* expression). On the right, schematic of translation regulation (*e.g.*, MSI-1\* regulating sfGFP expression). A general mathematical expression (grey shaded) was derived to calculate the fold change in protein expression as a function of the repressor concentration ( $R$ ), the association and dissociation rates ( $k_{ON}$  and  $k_{OFF}$ ), the elongation leakage fraction ( $\varepsilon$ ), and the nucleic acid degradation rate ( $\delta$ ). **b)** Heatmap of the fold change as a function of  $k_{ON}R$  and  $k_{OFF}$  (*i.e.*, the first-order kinetic rates that characterize the protein-DNA/RNA interaction) when  $\delta = 0$  and  $\varepsilon = 0.1$ . This would correspond to transcription regulation. **c)** Heatmap of the fold change when  $\delta = 0.14 \text{ min}^{-1}$  and  $\varepsilon = 0.1$ . This would correspond to translation regulation.

According to the aforementioned kinetic rates, this would be the case for MS2CP, but not for MSI-1\* (*i.e.*, both  $k_{ON}$  and  $k_{OFF}$  are instrumental to describe the regulation exerted by MSI-1\*). Furthermore, given the 2.5-fold down-regulation in our system, we estimated an elongation leakage fraction of 40% (using the fold change equation in the limit  $R \rightarrow \infty$ ). This leakage would come from the ability of

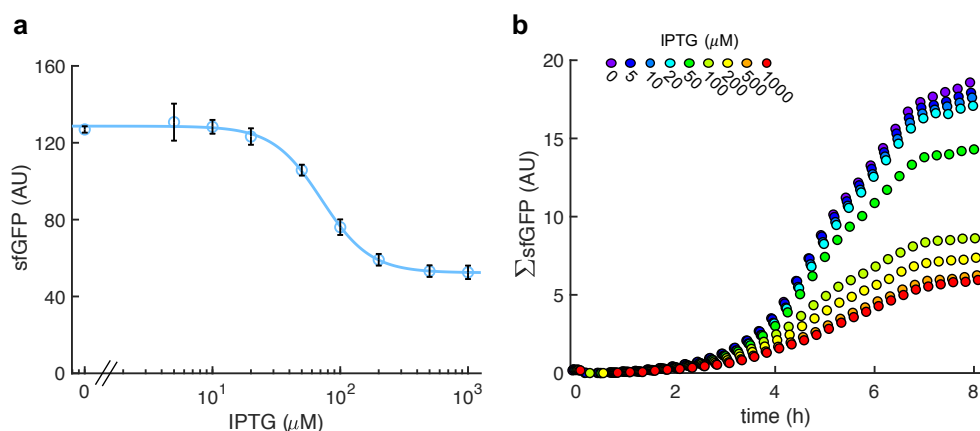
ribosomes to elongate even if MSI-1\* is bound and their ability to bind sooner to the *sfGFP* mRNA due to a conserved transcription-translation coupling mechanism [40].



**Fig. 11** A mathematical model captures the dynamic response of the system. **a)** Total red fluorescence of the cell population ( $\Sigma$ mScarlet) over time without and with 1 mM lactose. In this case, the cell growth rate was fitted to  $0.80 \text{ h}^{-1}$ . **b)** Total green fluorescence of the cell population ( $\Sigma$ sfGFP) over time without and with 1 mM lactose. The inset shows the dynamic response for different lactose concentrations. **c)** Correlation between the experimental values of SsfGFP at different times and for different lactose concentrations and the predicted values from a mathematical model that accounts for population growth and gene regulation. Data for  $t > 2 \text{ h}$ . Linear regression performed. **d)** Ratio of total green and red fluorescence as a proxy of cellular sfGFP expression over time. Ratio not represented at early times due to the high error obtained given the low number of cells present in the culture (grey shaded area). Deviations calculated by propagation. In all cases, points correspond to experimental data, while solid lines come from an adjusted mathematical model. Error bars correspond to standard deviations ( $n = 3$ ). AU, arbitrary units.

In addition, we studied the transient response of the gene circuit with lactose, as both MSI-1\* and sfGFP expressions changed with time. For that, we quantified the total red fluorescence of the cell population (Fig. 11a), which is an estimate of the total number of cells, and the total green fluorescence (Fig. 11b), which comes from the composition of population growth and gene regulation. We developed a

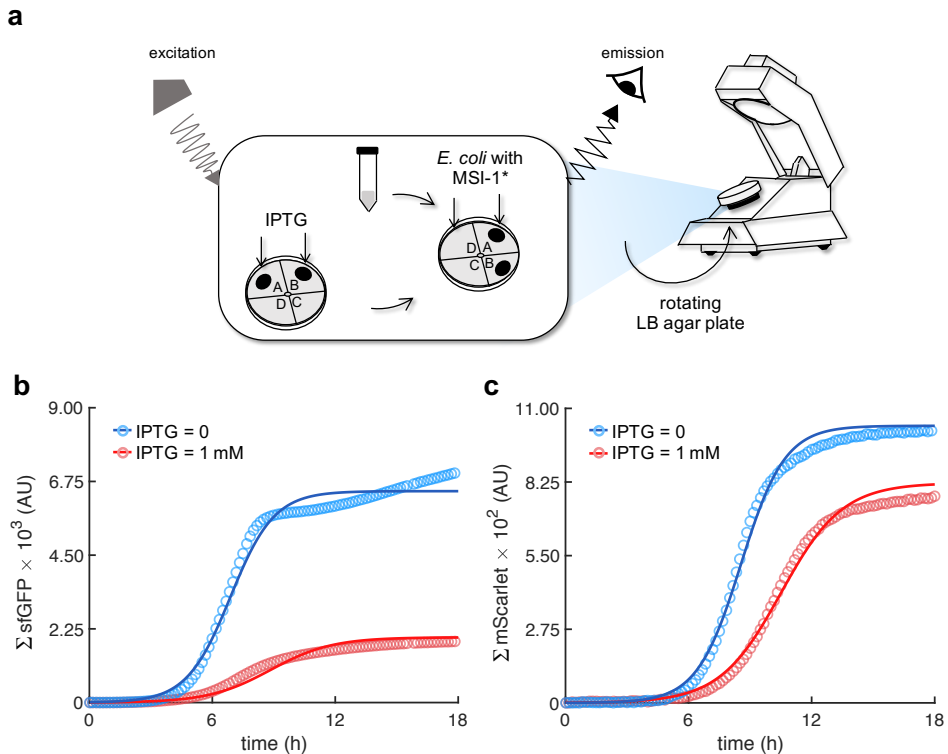
bottom-up mathematical model based on differential equations to predict sfGFP expression in the cell (data not shown), as well as a phenomenological model for the bacterial growth (data not shown). The parameter values were adjusted with the curves without and with 1 mM lactose. Then, we used the mathematical model to predict the transient responses for different intermediate lactose concentrations, finding excellent agreement with the experimental data ( $R^2=0.98$ ; Fig. 11c). We also characterized the time-course response of the circuit with IPTG, encountering similar results (Fig. 12a,b).



**Fig. 12** Characterization of the system response with IPTG (implemented with pRM1+ and pREP6). **a)** Dose-response curve. Error bars correspond to standard deviations ( $n=3$ ). **b)** Time-course response for different inducer concentrations (average of 4 clones). The fluorescence of the whole population is represented ( $\Sigma\text{sfGFP}$ ).

Moreover, to explore the maintenance of the regulatory behavior when the cell physiology changes, we characterized cells growing in solid medium with a repurposed LigandTracer technology (Fig. 13a), which initially was developed to monitor molecular interactions in real time [41]. In this case, a significant difference in the total red fluorescence was observed without and with 1 mM IPTG, suggesting that MSI-1\* expression was costly for the cell in these conditions. Besides, the total green fluorescence of the growing population was recapitulated using the model with a 2.6-fold down-regulation of cellular sfGFP expression, which is in tune with the results in liquid medium (Fig. 13b,c).

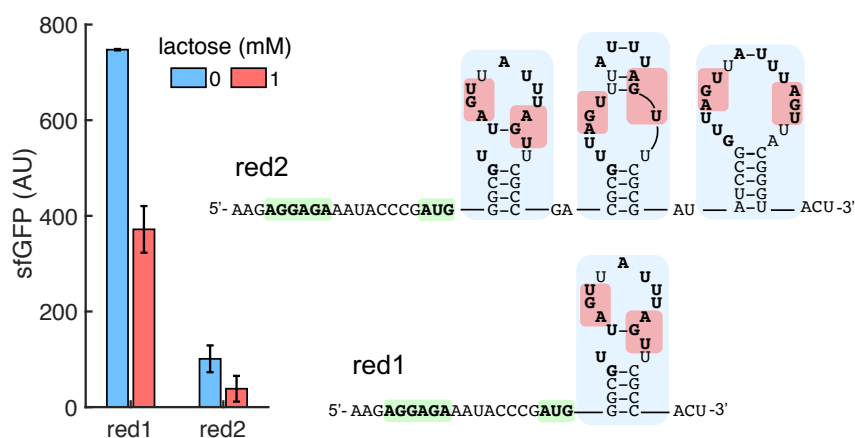
Subsequently, we analyzed the intracellular response. The time-dependent ratio of total green and red fluorescence was used as a proxy of sfGFP expression. A delay in the response is expected because MSI-1\* needs to be produced upon addition of lactose [42]. Nevertheless, our model predicted a faster response than experimentally observed (Fig. 11d). Overall, this quantitative inspection of translation regulation backs connections between molecular attributes and cellular behavior.



**Fig. 13** **a**) Schematics of the LigandTracer technology repurposed for characterizing bacterial cells expressing fluorescent proteins (implemented with pRM1+ and pREP6). **b**) Real-time green fluorescence of the whole population ( $\Sigma \text{sfGFP}$ ) upon induction with IPTG. **c**) Real-time red fluorescence ( $\Sigma \text{mScarlet}$ ). Points correspond to the experimental data, while solid lines come from an adjusted mathematical model.

### 3.4. Rational redesign of the targeted transcript to enhance the dynamic range of the response

The presence of stem-loop structures in the leader coding region contributes to lower the expression level. The more stable and closer to the start codon, the greater the impact on expression [43]. We hypothesized that, by destabilizing the RNA motif for MSI-1 binding, we would obtain an alternative regulatory system with higher expression levels. Accordingly, a new reporter system was engineered removing three base pairs from the stem, maintaining the two consensus recognition sequences. An experimental analysis revealed a 4.9-fold increase of the maximal sfGFP expression level and a 2.0-fold down-regulation with 1 mM lactose (Fig. 14, redesign 1). We then investigated the possibility of increasing the dynamic range of the response by placing three consecutive RNA motifs. However, we did not observe a greater down-regulation with 1 mM lactose (Fig. 14, redesign 2), suggesting that the additional motifs far away from the start codon had no effect; what was noticed is an effect on the maximal expression level.

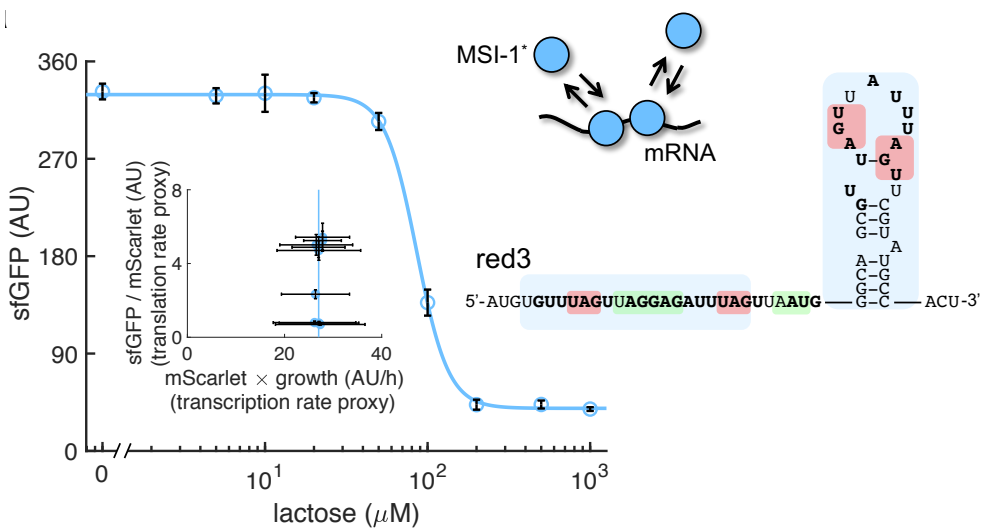


**Fig. 14** Dynamic range of the response of two redesigned genetic systems using lactose (1 mM). The predicted secondary structures of the leader regions of the reporter mRNAs are shown on the right. Redesign 1 (red1) was implemented with pREP4b and redesign 2 (red2) with pREP4b3x, which contains three MSI-1 binding sites. These stem-loop structures are less stable than the original one.

We also designed fusion proteins, with eBFP2 or MS2CP connected with a glycine-serine linker in the C-terminal region of MSI-1\*, thereby envisioning a greater ability

to interfere with the ribosome due to a bigger size of the resulting protein. However, no suitable clones were obtained despite multiple attempts, stressing the difficulty of recombinant protein expression (data not shown).

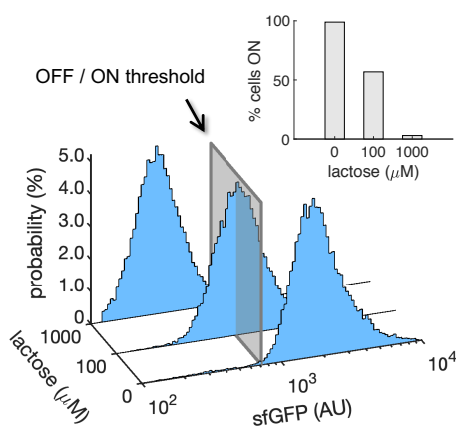
As a further strategy to enhance the dynamic range of the response, we redesigned the 5' untranslated region (UTR) of *sfGFP* to accommodate two additional  $\text{RU}_n\text{AGU}$  repeats (*viz.*, GUUUAGU and AUUUAGU) flanking the ribosome binding site (RBS), maintaining the original RNA motif after the start codon. Indeed, this is a widespread post-transcriptional regulatory strategy in prokaryotes, as it happens *e.g.* with the MS2 phage replicase [44]. We characterized by bulk fluorometry the dose-response curve of this new system, revealing an 8.6-fold down-regulation of *sfGFP* expression by MSI-1\* (Fig. 15, redesign 3).



**Fig. 15** Dose-response curve of another redesigned genetic system (redesign 3, red3) using lactose as inducer (up to 1 mM). MSI-1\* down-regulated *sfGFP* expression by 8.6-fold. The inset shows the scatter plot of the dynamic response in the Crick space (translation rate vs. transcription rate; vertical line fitted to 27 AU/h). The predicted secondary structure of the leader region of the reporter mRNA containing two MSI-1 binding sites (blue shaded) is shown on the right. In the 5' UTR, the binding site is formed by two  $\text{RU}_n\text{AGU}$  repeats that flank the RBS. In the leader coding region, the binding site is the original one. The minimal cores (UAG) are marked in red. Redesign 3 was implemented with pREP7. Points correspond to experimental data, while the solid line comes from an adjusted mathematical model. In all cases, error bars correspond to standard deviations ( $n = 3$ ).



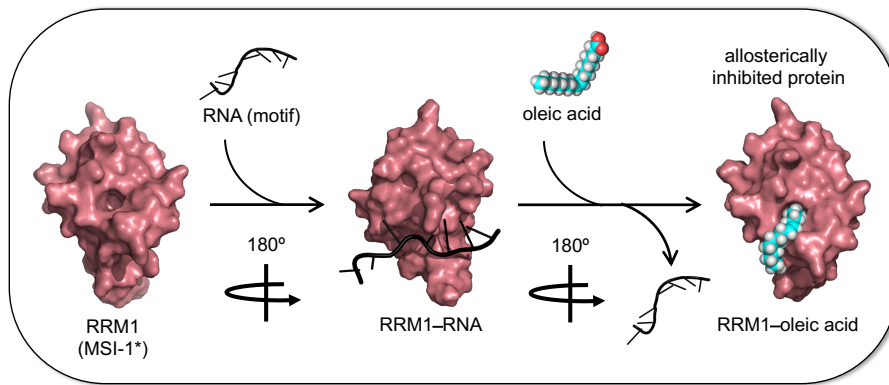
This was a substantial increase in performance with respect to the 2.5-fold down-regulation of the system shown in Fig. 3. At the single-cell level, we found a 91% of ON cells in the uninduced state that decreased to 5.3% with 1 mM lactose (Fig. 16). Taken together, our data present MSI-1\* as a powerful heterologous translation regulator in bacteria.



**Fig. 16** Probability-based histograms of sfGFP expression from single-cell data for different lactose concentrations (redesign 3). The inset shows the percentage of cells in the ON state (sfGFP expressed), according to a specified threshold, for each lactose concentration. AU, arbitrary units.

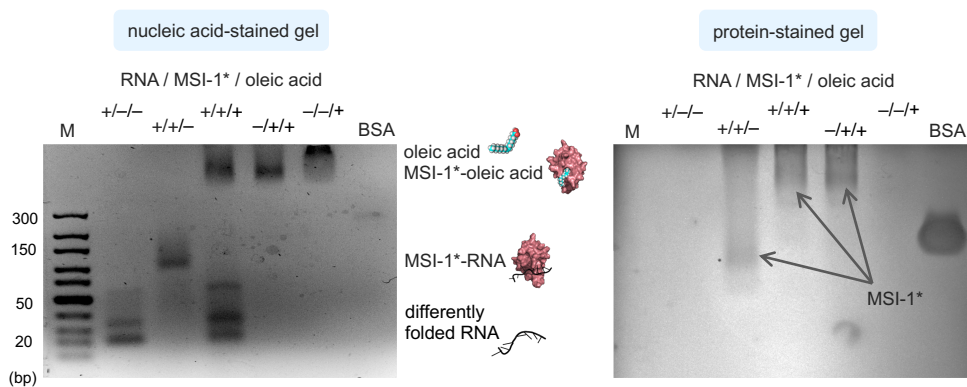
### 3.5. The regulatory activity of a Musashi protein in bacteria can be externally controlled by a fatty acid

The ability of proteins to respond to small molecules is instrumental for environmental and metabolic sensing. Previous work revealed that MSI-1 can be allosterically inhibited by w-9 monounsaturated fatty acids and, in particular, by oleic acid [21], an 18-carbon fatty acid naturally found in various animal and plant oils (*e.g.*, olive oil). Oleic acid binds to the RRM1 domain of MSI-1 and induces a conformational change that prevents RNA recognition (Fig. 17).



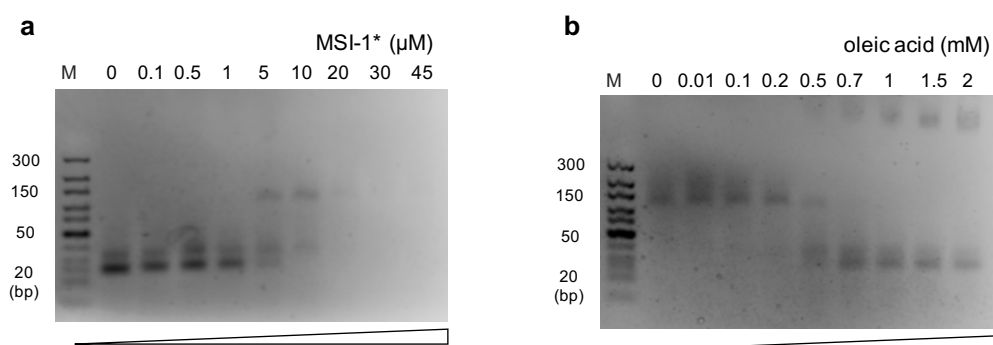
**Fig. 17** Oleic acid inhibits the regulatory activity of Musashi-1 in bacteria. Three-dimensional structural schematic of the allosteric regulation. RRM1 of MSI-1 is shown alone, in complex with the RNA motif, and in complex with oleic acid.

To gain insight about the interactions between the elements of our system, we performed gel electrophoretic assays using the purified MSI-1\* protein, the RNA motif as a label-free sRNA molecule, and oleic acid. The different mobility of the nucleic acids upon binding to proteins and the coincident staining capacity of nucleic acid and fatty acids were exploited.



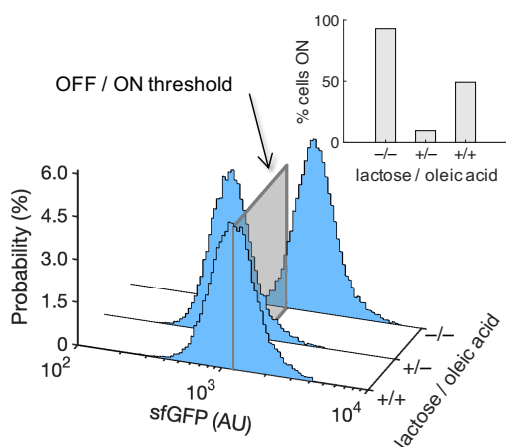
**Fig. 18** Oleic acid inhibits the regulatory activity of Musashi-1 in bacteria. **a)** Three-dimensional structural schematic of the allosteric regulation. RRM1 of MSI-1 is shown alone, in complex with the RNA motif, and in complex with oleic acid. **b)** Gel electrophoretic assay to test the allosteric inhibition of MSI-1\* with oleic acid. A purified MSI-1\* protein (45  $\mu\text{M}$ ), the RNA motif as a label-free sRNA molecule (11  $\mu\text{M}$ ), and oleic acid (1 mM) were mixed in a combinatorial way *in vitro*. On the left, nucleic acid-stained gel. On the right, protein-stained gel (Coomassie). The different formed species are indicated. M denotes molecular marker (GeneRuler ultra-low range DNA ladder, 10-300 bp, Thermo). BSA was used as a control.

We confirmed the MSI-1\*-RNA interaction using a protein concentration gradient in this *in vitro* set up (Fig. 18a), and we found that the interaction was completely disrupted in presence of 1 mM oleic acid (Fig. 18). Furthermore, using an oleic acid concentration gradient, we obtained a half-maximal effective inhibitory concentration of about 0.5 mM (Fig. 19b).



**Fig. 19** Gel electrophoretic assays to test the MSI-1\*-RNA and the MSI-1\*-oleic acid interactions (nucleic acid-stained gels). **a**) Interaction of MSI-1\* with the RNA motif. RNA added at 11 μM. **b**) Interaction of MSI-1\* with oleic acid. RNA added at 11 μM and MSI-1\* at 45 μM. M, molecular marker (GeneRuler ultra-low range DNA ladder, 10-300 bp, Thermo).

Subsequently, we assessed the effect of oleic acid over the regulatory activity of MSI-1\* expressed in *E. coli*. This bacterium has evolved a machinery to uptake fatty acids from the environment. FadL and FadD are two membrane proteins that act as transporters, and FadE is the first enzyme that processes the fatty acid via the  $\beta$ -oxidation cycle [45]. Because of the high turbidity of the cell culture observed in presence of oleic acid, we characterized the system by single-cell measurements of sfGFP expression by flow cytometry. The percentage of cells in the ON state increased from 10% (with 1 mM lactose) to 49% upon addition of 20 mM oleic acid (Fig. 20). However, the initial 93% of ON cells observed in absence of lactose was not recovered. Arguably, oleic acid was partially degraded once it entered the cell.

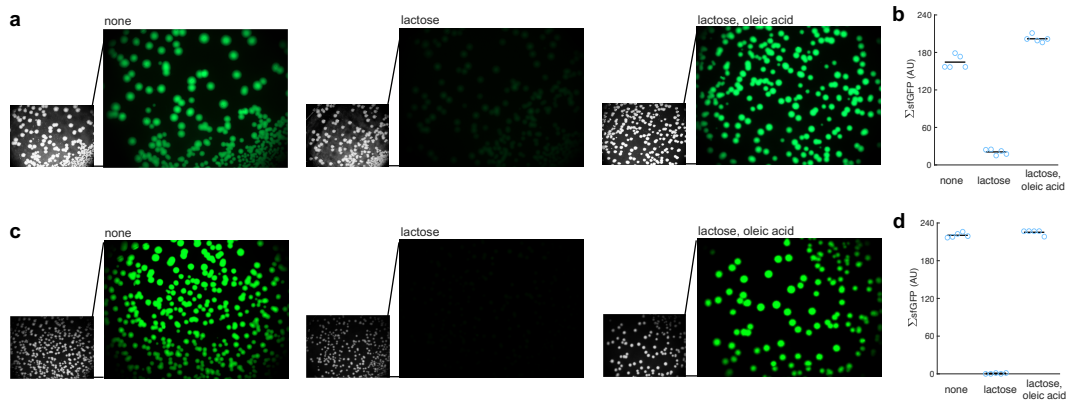


**Fig. 20** Probability-based histograms of sfGFP expression from single-cell data for different induction conditions (1 mM lactose or 1 mM lactose + 20 mM oleic acid) for the original system (implemented with pRM1+ and pREP6). The inset shows the percentage of cells in the ON state (sfGFP expressed), according to a specified threshold, for each condition.

Moreover, a controlled heterogeneity of the response with oleic acid was observed. In particular, we found a dispersion of the unimodal distribution (quantified as the Fano factor, *i.e.*, the ratio between variance and mean) 2.9-fold higher in the case of induction with 0.1 mM lactose than with 1 mM lactose and 20 mM oleic acid, despite having similar mean expression levels. As the former response sensitivity was dominated by transcription regulation and the latter by translation regulation, that result agrees with previous work suggesting that molecular noise is buffered when the control mechanism is post-transcriptional [34].

In addition, we investigated this allosteric regulation by imaging the fluorescence of bacterial colonies grown in solid medium with different inducers. In stationary phase, FadE and the rest of oxidative enzymes could be saturated with the fatty acids generated from the membrane degradation [46], oleic acid then having more time to interact with MSI-1\*. Notably, we found a substantial inhibition of the repressive action of MSI-1\* with 20 mM oleic acid (Fig. 21a,c). The system implemented with the redesign-3 reporter displayed an even better dynamic behavior in response to lactose and oleic acid (Fig. 22a,b). Conclusively, these results illustrate how the plasticity of RRM-containing proteins (*e.g.*, MSI-1) can be exploited to

engineer, even in simple organisms, gene regulatory circuits that operate in an integrated way at the transcriptional, translational, and post-translational levels.



**Fig. 21 Oleic acid inhibits the regulatory activity of Musashi-1 in bacteria.** **a)** Images of *E. coli* colonies harboring pRM1+ and pREP6. Bacteria were seeded in LB-agar plates with suitable inducers (1 mM lactose or 1 mM lactose + 20 mM oleic acid). Fluorescence and bright field images are shown. On the bottom, schematics of the working mode of the synthetic gene circuit according to the different induction conditions. **b)** Quantification of the green fluorescence of the colonies from panel a (denoted by  $\Sigma sfGFP$  as it is from populations;  $n = 5$ ). **c)** Images of *E. coli* colonies harboring pRM1+ and pREP7. **d)** Quantification of the green fluorescence of the colonies from panel c. AU, arbitrary units.

## 4. DISCUSSION

The successful incorporation of the mammalian MSI-1 protein as a translation factor in *E. coli* highlights, in first place, the versatility of RRM-containing proteins to function as specific post-transcriptional regulators in any living cell, from prokaryotes to eukaryotes. Our data show that the protein-RNA association phase is very fast, which is suitable for regulation even in cellular contexts in which RNA molecules are short-lived, such as in *E. coli* [37]. Nonetheless, it is important to stress that the kinetic parameters *in vivo* might differ from those measured *in vitro* due to off-target bindings and crowding effects [47].

Moreover, our data show that a down-regulation of translation rate up to 8.6-fold can be achieved, with an appropriate design of the target mRNA leader region, and that the engineered cell can sense oleic acid from the environment. Here, the C-terminal low-complexity domain of the native MSI-1 was discarded to create MSI-1\* [24], in order to increase solubility, even though this domain might contribute to RNA binding [48]. Interestingly, proteins associated to clustered regularly interspaced short palindromic repeats (CRISPR), which belong to the prokaryotic immune system, contain distorted RRM versions [49]. Some CRISPR proteins might have evolved, for example, from an ancestral RRM-based (palm) polymerase after duplications, fusions, and diversification. Noting that the palm domain indeed presents an RRM-like fold [50], we hypothesize that a boost of functionally diverse RRM-containing proteins took place once the polymerases were confined into the nucleus, as the pressure for efficient replication was relieved in the cytoplasm, which would provide a rationale on the unbalance noticed between eukaryotes and prokaryotes [7,51]. In second place, our results pave the way for engineering more complex circuits in bacteria with plastic and orthogonal RNA-binding proteins, such as MSI-1, capable of signal multiplexing.

Nature is a formidable reservoir of functional genetic material sculpted by evolution that can be exploited to (re)program specific living cells [10]. However, to overcome biological barriers, transgenes usually come from related organisms or cognate parasites, at the cost of limiting the potential engineering. Therefore, efforts to borrow functional elements from highly diverse organisms are suggestive (*e.g.*, regulatory proteins from mammals to bacteria), with the ultimate goal of developing industrial or biomedical applications. Notably, advances in synthetic biology have pushed the bioproduction of a wide variety of compounds in bacteria as a result of a better ability to fine tune enzyme expression [52]. Translation regulation is instrumental to this end because in multiple cases different enzymes are expressed from the same transcriptional unit (*i.e.*, operon). Previous work exploited regulatory RNAs for such a tuning [53], but the use of RNA-binding proteins as translation

factors is also appealing. We envision the application of MSI-1\* as a genetic tool for metabolic engineering. Furthermore, MSI-1\* is able to respond to fatty acids, which are ideal precursors of potential biofuels due to their long hydrocarbon chains. In particular, biofuel in the form of fatty acid ethyl ester, whose bioproduction in *E. coli* can be optimized by reengineering the regulation of the  $\beta$ -oxidation cycle with the allosteric transcription factor FadR [54]. Arguably, MSI-1\* might be used in place of or in combination with FadR for subsequent developments. However, engineering regulatory circuits for efficient bioproduction is not evident in general as the enzymatic expression levels may require fine tuning, so systems-level mathematical models need to be considered for design along with a wide genetic toolkit for implementation [52].

We anticipate that other animal RRM-containing proteins might be repurposed in *E. coli* as translation factors. Moreover, protein design might be used to reengineer MSI-1\* in order to respond to new ligands, maintaining high specificity and affinity for a particular RNA sequence, as previously done with the transcription factor LacI [55]. In addition, the Musashi protein family is of clinical importance, as in humans it is involved in different neurodegenerative disorders (*e.g.*, Alzheimer's disease) and some types of cancer [18,56,57]. Therefore, the development of simple genetic systems from which to test protein mutants, potential target mRNAs, decoying RNA aptamers, and inhibitory small molecules in a systematic manner is very relevant. Furthermore, isolating human regulatory elements would help to filter out indirect effects that likely occur in the natural context. This might lead to new therapeutic opportunities. Nevertheless, one limitation of using *E. coli* as a chassis is that some post-translational modifications (PTMs) may be lost, thereby compromising the functionality of the expressed proteins [58]. Fortunately, there are metabolic engineering efforts devoted to implement eukaryotic PTM pathways in *E. coli*, such as the glycosylation pathway [59].

In conclusion, the functionalization of RRM-containing proteins in bacteria offers exciting prospects, especially as more information becomes available on how

individual RRM domains bind to precise RNA sequences, interact with further protein domains, and respond to small molecules through allosteric effects. This work illustrates how synthetic biology, through the rational assembly of heterologous genes and designer *cis*-regulatory elements into circuits, is useful to generate knowledge about the application range of a fundamental type of proteins in nature.



## REFERENCES

1. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136: 615-628.
2. Holmqvist E, Vogel J (2018) RNA-binding proteins in bacteria. *Nat Rev Microbiol*, 16: 601-615.
3. Jonas S, Izaurralde E (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet*, 16: 421-433.
4. Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*, 582: 1977-1986.
5. Babu MM, Teichmann SA, Aravind L (2006) Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol*, 358: 614-633.
6. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3: 318-356.
7. Maris C, Dominguez C, Allain FHT (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J*, 272: 2118-2131.
8. Messias AC, Sattler M (2004) Structural basis of single-stranded RNA recognition. *Acc Chem Res*, 37: 279-287.
9. Maruyama K, Sato N, Ohta N (1999) Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. *Nucleic Acids Res*, 27: 2029-2036.
10. Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. *Nat Rev Genet*, 11: 367-379.
11. Bashor CJ, Collins JJ (2018) Understanding biological regulation through synthetic biology. *Annu Rev Biophys*, 47: 399-423.
12. Belmont BJ, Niles JC (2010) Engineering a direct and inducible protein-RNA interaction to regulate RNA biology. *ACS Chem Biol*, 5: 851-861.
13. Katz N, Cohen R, Solomon O, Kaufmann B, Atar O, Yakhini Z, Goldberg S, Amit R (2019) Synthetic 5' UTRs can either up- or downregulate expression upon RNA-binding protein binding. *Cell Syst*, 9: 93-106.
14. Cao J, Arha M, Sudrik C, Mukherjee A, Wu X, Kane RS (2015) A universal strategy for regulating mRNA translation in prokaryotic and eukaryotic cells. *Nucleic Acids Res*, 43: 4353-4362.
15. Qi LS, Arkin AP (2014) A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat Rev Microbiol*, 12: 341-354.
16. Rosado A, Cordero T, Rodrigo G (2018) Binary addition in a living cell based on riboregulation. *PLoS Genet*, 14: e1007548.

17. Shotwell CR, Cleary JD, Berglund JA (2020) The potential of engineered eukaryotic RNA-binding proteins as molecular tools and therapeutics. *Wiley Interdiscip Rev RNA*, 11: e1573.
18. Fox RG, Park FD, Koechlein CS, Kritzik M, Reya T (2015) Musashi signaling in stem cells and cancer. *Annu Rev Cell Dev Biol*, 31: 249-267.
19. Nakamura M, Okano H, Blendy JA, Montell C (1994) Musashi, a neural RNA-binding protein required for Drosophila adult external sensory organ development. *Neuron*, 13: 67-81.
20. Imai T, Tokunaga A, Yoshida T, Hashimoto M, Mikoshiba K, Weinmaster G, Nakafuku M, Okano H (2001) The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol*, 21: 3888-3900.
21. Clingman CC, Deveau LM, Hay SA, Genga RM, Shandilya SMD, Massi F, Ryder SP (2014) Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife*, 3: e02848.
22. Ganesan SM, Falla A, Goldfless SJ, Nasamu AS, Niles JC (2016) Synthetic RNA-protein modules integrated with native translation mechanisms to control gene expression in malaria parasites. *Nat Commun*, 7: 10727.
23. MacDonald CI, Seamons TR, Emmons JC, Javdan SB, Deans TL (2021) Enhanced regulation of prokaryotic gene expression by a eukaryotic transcriptional activator. *Nat Commun*, 12: 4109.
24. Iwaoka R, Nagata T, T suda K, Imai T, Okano H, Kobayashi N, Katahira M (2017) Structural insight into the recognition of r(UAG) by Musashi-1 RBD2, and construction of a model of Musashi-1 RBD1-2 bound to the minimum target RNA. *Molecules*, 22: 1207.
25. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*, 25: 1203-1210.
26. Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol*, 24: 79-88.
27. Kawahara H, Imai T, Imataka H, Tsujimoto M, Matsumoto K, Okano H (2008) Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. *J Cell Biol*, 181: 639-653.
28. Ai HW, Shaner NC, Cheng Z, Tsien RY, Campbell RE (2007) Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry*, 46: 5904-5910.

29. Bindels DS, Haarbosch L, Van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, Gadella TWJ Jr (2017) mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods*, 14: 53-56.
30. Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139: 1366-1375.
31. Hausser J, Mayo A, Keren L, Alon U (2019) Central dogma rates and the trade-off between precision and economy in gene expression. *Nat Commun*, 10: 68.
32. Zearfoss NR, Deveau LM, Clingman CC, Schmidt E, Johnson ES, Massi F, Ryder SP (2014) A conserved three-nucleotide core motif defines Musashi RNA binding specificity. *J Biol Chem*, 289: 35530-35541.
33. Cléry A, Sohler TJ, Welte T, Langer A, Allain FH (2017) switchSENSE: A new technology to study protein-RNA interactions. *Methods*, 118: 137-145.
34. Dolcemascolo R, Goiriz L, Montagud-Martinez R, Rodrigo G (2022) Gene regulation by a protein translation factor at the single-cell level. *PLoS Comput Biol*, 18: e1010087.
35. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol*, 32: 562-568.
36. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, 27: 946-50.
37. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA*, 99: 9697-9702.
38. Garcia HG, Phillips R (2011) Quantitative dissection of the simple repression input-output function. *Proc Natl Acad Sci USA*, 108: 12173-12178.
39. Goiriz L, Rodrigo G (2021) Nonequilibrium thermodynamics of the RNA-RNA interaction underlying a genetic transposition program. *Phys Rev E*, 103: 042410.
40. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P (2017) Architecture of a transcribing-translating expressome. *Science*, 356: 194-197.
41. Björke H, Andersson K (2006) Measuring the affinity of a radioligand with its receptor using a rotating cell dish with in situ reference area. *Appl Radiat Isot*, 64: 32-37.

42. Rosenfeld N, Alon U (2003) Response delays and the structure of transcription networks. *J Mol Biol*, 329: 645-654.
43. Paulus M, Haslbeck M, Watzele M (2004) RNA stem-loop enhanced expression of previously non-expressible genes. *Nucleic Acids Res*, 32: e78.
44. Babitzke P, Baker CS, Romeo T (2009) Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol*, 63: 27-44.
45. Fujita Y, Matsuoka H, Hirooka K (2007) Regulation of fatty acid metabolism in bacteria. *Mol Microbiol*, 66: 829-839.
46. Navarro-Llorens JM, Tormo A, Martínez-García E (2010) Stationary phase in gram- negative bacteria. *FEMS Microbiol Rev*, 34: 476-495.
47. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J (2012) The lac repressor displays facilitated diffusion in living cells. *Science*, 336: 1595-1598.
48. Järvelin AI, Noerenberg M, Davis I, Castello A (2016) The new (dis)order in RNA regulation. *Cell Commun Signal*, 14: 9.
49. Koonin EV, Makarova KS (2013) CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol*, 10: 679-686.
50. Anantharaman V, Iyer LM, Aravind L (2010) Presence of a classical RRM-fold palm domain in Thg1-type 3'-5' nucleic acid polymerases and the origin of the GGDEF and CRISPR polymerase domains. *Biol Direct*, 5: 43.
51. Koonin EV, Krupovic M, Ishino S, Ishino Y (2020) The replication machinery of LUCA: common origin of DNA replication and transcription. *BMC Biol*, 18: 61.
52. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY (2019) Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol*, 37: 817-837.
53. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat Biotechnol*, 31: 170-174.
54. Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol*, 30: 354-359.
55. Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, Rogers JK, Isaacs FJ, Kosuri S, Baker D, Fields S, Church GM, Raman S (2016) Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods*, 13: 177-183.

56. Montalbano M, McAllen S, Puangmalai N, Sengupta U, Bhatt N, Johnson OD, Kharas MG, Kayed R (2020) RNA-binding proteins Musashi and tau soluble aggregates initiate nuclear dysfunction. *Nat Commun*, 11: 4305.
57. Kang MH, Jeong KJ, Kim WY, Lee HJ, Gong G, Suh N, Gyorffy B, Kim S, Jeong SY, Mills GB, Park YY (2017) Musashi RNA-binding protein 2 regulates estrogen receptor 1 function in breast cancer. *Oncogene*, 36: 1745-1752.
58. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem*, 307: 249-264.
59. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, Heiss C, Azadi P, Aebi M, DeLisa MP (2012) An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat Chem Biol*, 8: 434-436.
60. Peterson J, Philips GJ (2008) New pSC101-derivate cloning vectors with elevated copy numbers. *Plasmid*, 59: 193-201.
61. Leveau JH, Lindow SE (2001) Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol*, 183: 6752-6762.
62. Langer A, Hampel P, Kaiser W, Knezevic J, Welte T, Villa V, Maruyama M, Svejda M, Jahner S, Fischer F, Strasser R, Rant U (2013) Protein analysis by time-resolved measurements with an electro-switchable DNA chip. *Nat Commun*, 4: 2099.
63. Perea W, Greenbaum NL (2020) Label-free horizontal EMSA for analysis of protein- RNA interactions. *Anal Biochem*, 599: 113736.
64. Fessenden-Raden JM (1972) Effect of fatty acids on the movement and staining of membrane proteins in polyacrylamide gel electrophoresis. *Biochem Biophys Res Commun*, 46: 1347-1353.
65. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9: 676-682.
66. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res*, 33 D154- D159.
67. Popena M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res*, 40: e112.



# GENERAL DISCUSSION

Synthetic biology aims to design, construct, and optimize biological systems for useful purposes. It uses engineering principles to build unique biological circuits and organisms with predictable and programmable behavior. In this context, biological parts (genes, proteins, and regulatory elements) are treated as composable elements that can be interconnected to form complex systems. These systems could perform a wide range of tasks, from sensing and responding to environmental changes to producing valuable compounds for various applications [1,2].

In the cell, proteins can bind to DNA to regulate transcription as well as to RNA to regulate translation. However, bacterial cells have mainly evolved to exploit transcription factors as specific gene regulators, while translation factors have remained as global modulators of expression. Consequently, transcription regulation has attracted much attention over the last years to unveil design principles of genetic organization and to engineer synthetic circuits for cell reprogramming [3,4]. Yet, while substantial work combining theory and experiments has been carried out to study how noise propagates through transcriptional regulations, the stochastic behavior of genes regulated at the level of translation is poorly understood [5].

In this thesis, we have engineered a synthetic genetic system in which a target gene is down-regulated by an RNA-binding protein acting as a translation factor (MS2CP), which in turn is regulated transcriptionally. In particular, we have exploited the phage protein to regulate the expression of a green fluorescent protein at the level of translation. Thus, we have designed a two-layer genetic system that involves transcriptional and post-transcriptional regulations. Through our synthetic circuit, we have studied the generation and propagation of noise (stochastic behavior) of the gene regulated at the level of translation. In particular, we quantified the stochasticity of the system by monitoring both the expression of the regulator and

the regulated gene at the single-cell level [6].

Importantly, our results show that with a protein translation factor i) a tight repression can be achieved in single cells, ii) noise propagation from gene to gene is buffered, iii) and the regulated gene is sensitive in a nonlinear way to global perturbations in translation. Key findings revealed a significant down-regulation in gene expression, approximately 50-fold, achieved by inhibiting ribosomal progression on the target mRNA through protein-RNA interactions. This down-regulation was comparable to transcriptional fold-changes. In addition, we developed a robust mathematical framework capable of describing stochastic behaviors influenced by both transcription and translation factors. Our results showed that a bottom-up mathematical model can be exploited to predict the transfer functions of the system. We have also shown that a Gamma distribution parameterized with mesoscopic parameters, such as the mean expression and coefficient of variation, provided a deep analytical explanation about the system, displaying enough versatility to capture the cell-to-cell variability in genes regulated both transcriptionally and translationally.

Our utilization of the viral protein MS2CP in the regulatory system suggested the potential for diverse RNA-binding proteins. Orthogonal systems could be engineered using proteins like the bacteriophage PP7 coat protein or the Mycobacterium enzyme PyrR [7]. Considering the abundance of RNA-binding proteins in nature, especially in eukaryotes, multiple implementations were conceivable. The adaptability of our mathematical model allowed us to accommodate different proteins while preserving the fundamental functional form. Predictability could be enhanced by leveraging tools like the RBS calculator, and incorporating tandem repeats of RNA motifs might augment the regulatory fold-change. RNA-binding proteins offered versatility in regulating gene expression post-transcriptionally, and analyzing other mechanisms, such as translation elongation



blockage by Argonaute proteins in eukaryotes, was a crucial avenue for future research.

In essence, our study provided in-depth quantitative insights into the stochastic behavior of genes translationally regulated by RNA-binding proteins. These translation factors offered unique advantages, bridging attributes of both proteins (like transcription factors) and small RNAs. Our work not only advanced our understanding of these regulatory mechanisms but also paved the way for more sophisticated gene regulatory circuit engineering. The integration of different layers within the genetic information flow, specifically transcription and translation, facilitates easier signal integration for achieving specific functions. We anticipated that RNA-binding proteins would play a vital role in synthetic biology, addressing various biotechnological and biomedical challenges in the near future.

In this thesis, we have additionally engineered an orthogonal post-transcriptional synthetic genetic system in *E. coli* with a mammalian RNA-binding protein [8]. In nature, the RNA recognition motif (RRM) is the most common conserved RNA-binding protein domain identified. Such conserved structural domain consists of a four-stranded antiparallel  $\beta$ -sheet packet against two  $\alpha$ -helices with specific amino acid residues that enable the contact with RNA through hydrogen bonding [9]. However, RRM-containing proteins are only prevalent in eukaryotic phyla, in which they play central regulatory roles along RNA metabolism and stability [10]. In particular, RRM-containing proteins can fine-tune protein synthesis by interacting with RNA. Although RRM-containing proteins have a pivotal role in gene expression regulation, they are scarce in prokaryotes. Thus, it would be useful to use RRM-RNA interactions as an orthogonal layer to engineer gene regulation adding a novel post-transcriptional regulatory layer in such organisms.

In particular, we have engineered an orthogonal post-transcriptional control system of gene expression in the bacterium *E. coli* with the mammalian RNA-binding protein Musashi-1, which is a stem cell marker with neurodevelopmental role that contains two canonical RRM. In the circuit, Musashi-1 is regulated transcriptionally and works as an allosteric translation repressor thanks to a specific interaction with the leader coding region of a messenger RNA and its structural plasticity to respond to fatty acids. To study the response of the regulatory system, we have presented quantitative assays and theoretical results. We fully characterized the genetic system both at the population and single-cell levels showing a significant fold change in reporter expression, and the underlying molecular mechanism by assessing the *in vitro* binding kinetics and *in vivo* functionality of a series of RNA mutants. Moreover, the dynamic response of the system has been well recapitulated by a bottom-up mathematical model.

The successful incorporation of the mammalian MSI-1 protein as a translation factor in *E. coli* underscored the versatility of RRM-containing proteins as specific post-transcriptional regulators across diverse living cells. Notably, we modified the native MSI-1 by discarding its C-terminal low-complexity domain to enhance solubility, although this domain potentially contributed to RNA binding. The rapid protein-RNA association phase, even in contexts with short-lived RNA molecules like in *E. coli*, demonstrated the potential applicability of these proteins. However, it is crucial to acknowledge the potential disparities between *in vivo* and *in vitro* kinetic parameters due to, for example, off-target bindings and crowding effects [11]. Overall, we have shown how RRM-based regulation can be adapted to simple organisms, thereby paving the way for engineering more complex circuits in prokaryotes by combining transcription and translation control with proteins. Indeed, we aimed to expand the synthetic toolbox and enlarge the repertoire of translational repressors suitable for biotechnological applications. Overall, our results elucidate the use of Musashi-1 protein in synthetic biology.

Our engineered cells, capable of sensing environmental oleic acid, showcased the utility of our synthetic biology development. Responding to fatty acids offers opportunities in optimizing biofuel production, particularly fatty acid ethyl esters, in *E. coli* through regulation of the  $\beta$ -oxidation cycle [12]. Indeed, the bioproduction of industrially-relevant compounds in microbes stands as a linchpin for the preservation of the biosphere. Unlike traditional chemical synthesis methods, microbial bioproduction offers an eco-friendly alternative, dramatically reducing the environmental footprint of various industries. By harnessing the potential of RNA-binding proteins, we might design specific pathways for the production of valuable compounds, mitigating the pressure on natural resources, conserving biodiversity and maintaining ecological equilibrium.

We also anticipate the repurposing of other animal RRM-containing proteins in *E. coli* and the use of protein design to engineer these proteins for new ligand responses. We have demonstrated that once our reporter gene is transcribed into a mRNA, RNA-binding proteins are able to bind to a specific sequence placed in the leader region of the reporter mRNA to regulate translation. Alternatively, in the case of eukaryotes, RNA-binding proteins can also participate in chromatin remodeling, in addition to RNA stability, editing, and decay [13]. In addition, understanding the native protein-RNA interactome, which refers to the network of RNA molecules that interact with specific RNA-binding proteins, would facilitate the identification of novel proteins with potential to be borrowed for synthetic gene expression regulation and reprogrammed cellular function [14].

In summary, this thesis has illuminated the potential of RNA-binding proteins (in particular, RRM-containing proteins) in bacteria, offering promising prospects for future research. As our understanding deepens regarding how individual protein domains interact with specific RNA sequences and small molecules and regarding the quantitative aspects of gene regulation at the level of translation, our ability to

implement more complex circuits will increase. This study is intended to lay the foundation for exploring the wide-ranging capabilities of RNA-binding proteins, showcasing their adaptability and versatility.

## REFERENCES

1. Stephanopoulos G (2012) Synthetic biology and metabolic engineering. *ACS synthetic biology*, 1, 514-525.
2. Zhang J, Jensen MK, Keasling JD (2015) Development of biosensors and their application in metabolic engineering. *Current opinion in chemical biology*, 28, 1-8.
3. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, 123, 1025-1036.
4. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307: 1965-1969.
5. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science*, 297, 1183-1186.
6. Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol*, 24, 545-554.
7. Babitzke P, Baker CS, Romeo T (2009) Regulation of translation initiation by RNA binding proteins. *Annual review of microbiology*, 63, 27-44.
8. Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, 265, 615-621.
9. Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, 8, 479-490.
10. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, Na H, Irimia M, Matzat LH, Dale KR, Smith SA, Yarosh CA, Kelly SM, Nabet B, Mecnas D, Li W, Laishram RS, Qiao M, Lipshitz HD, Piano F, Corbett AH, Carstens RP, Frey BJ, Anderson RA, Lynch KW, Penalva LOF, Lei EP, Fraser AG, Blencowe BJ, Morris QD, Hughes TR (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, 499, 172-177.
11. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J (2012) The lac repressor displays facilitated diffusion in living cells. *Science*, 336, 1595-1598.
12. Zhang L, Veres-Schalnat TA, Somogyi A, Pemberton JE, Maier RM (2012) Fatty acid cosubstrates provide  $\beta$ -oxidation precursors for rhamnolipid biosynthesis in *Pseudomonas aeruginosa*, as evidenced by isotope tracing and gene expression assays. *Applied and environmental microbiology*, 78, 8611-8622.
13. Perez CAG, Adachi S, Nong QD, Adhitama N, Matsuura T, Natsume T, Wada T, Kato Y, Watanabe H (2021) Sense-overlapping lncRNA as a decoy of

translational repressor protein for dimorphic gene expression. *PLoS genetics*, 17, e1009683

14. Perez-Perri JJ, Rogell B, Schwarzl T, Stein F, Zhou Y, Rettel M, Brosig A, Hentze MW (2018) Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nature communications*, 9, 4408.

# CONCLUSIONS

This PhD dissertation was intended to expand our knowledge on how translation control of gene expression with proteins in prokaryotic cells can be engineered to create a new variety of synthetic circuits. In particular, this thesis has reached the following main conclusions:

- 1) Gene expression programs can be engineered with RNA-binding proteins (*e.g.*, MS2CP or MSI-1\*) in *E. coli*, achieving sufficient dynamic range and tunability. This was accomplished through an interaction with the 5' UTR or N-terminal coding region of the target mRNA.
- 2) Through the use of a protein translation factor a tight repression can be achieved in single cells, noise propagation from gene to gene is buffered, and the regulated gene is sensitive in a nonlinear way to global perturbations in translation.
- 3) Translation regulation with proteins can be bottom-up mathematically modeled to predict dynamic responses. Moreover, Gamma distribution parameterized with mesoscopic parameters, such as the mean expression and coefficient of variation, provides a deep analytical explanation about the system, displaying enough versatility to capture the cell-to-cell variability in genes regulated both transcriptionally and translationally.
- 4) RNA-binding proteins from mammals carrying RNA recognition motifs (RRMs) can work in bacteria (*e.g.*, MSI-1\*).
- 5) RNA-binding proteins can be allosterically regulated (*e.g.*, MSI-1\* in bacteria can be externally controlled by a fatty acid). This plasticity can be exploited to engineer, even in simple organisms, gene regulatory circuits that operate in an integrated way at the transcriptional, translational, and post-translational levels.

6) Through the use of an engineered post-transcriptional mechanism, it was possible to achieve the specific regulation within an operon and the implementation of combinatorial regulation.



# ACKNOWLEDGEMENTS

My stay at the GR2.2 lab has been an incredible journey of learning.

I am deeply grateful to my supervisor Dr Guillermo Rodrigo who dedicated his time to this project, and whose profound insight and knowledge into the Synthetic Biology field steered me through this research. The environment he created is an inspiration for my future career.

Gracias para la increíble oportunidad y el tiempo pasado aquí, al principio ha sido todo “preliminar” pero al final “se ha quedado canela”.

I would like to extend my special thanks to the Spanish National Research Council (CSIC) and the Institute for Integrative Systems Biology (I2SysBio) for hosting me on this amazing adventure and providing me with an opportunity of pursuing the PhD degree at the Polytechnic University (UPV) of Valencia.

Thanks to the MSCA-ITN H2020 #813239 fellowship, funded by the European Commission for letting me be part of this incredible network.

I would like to express my gratitude to the RNAct consortium which became a family. Thanks to the supervisors Prof Dr Wim Vranken, Prof Dr Michel Sattler, Dr Wolfgang Kaiser, Dr Isaure Chauvot de Beauchêne, Dr Tommaso Martelli, Dr Jos Bujis; and the ESRs Jose Gavaldá-Garcia, Joel Roca Martínez, Hrishikesh Dhondge, Anna Kravchenko, Stefano Mocci, Niki Messini, Luca Sperotto, Anna Pérez i Ràfols, Anahí Higuera, Guillermo Pérez Roperó, for the scientific trainings and workshops around Europe even in pandemic time.

Thanks to my buddy Hrishi, and Anna P., we perfectly matched with the trio LaLaLa.

Special thanks to the project manager Dr Aitor Sánchez for the care through this research project, merci.

Many thanks to Prof. Helena Danielson to let me join her lab group at BMC. To Jos for welcoming me at Ridgeview during my secondment in Uppsala. His plentiful

experience has encouraged me in all the time of my Swedish experience. I would also thank the team of the company (Karl, Sina, Anna B, John) for sharing thoughts and ideas during fika, tack så mycket!

Thanks to the Oslogatan flatmates Guille, Alva, Daniel, Leni, Ryan, Daniel B., João, and “los españoles de Uppsala” Alba, María, Tiscar, Javi, Laura, Lorenzo, Daniel for making the cold Swedish winter warmer.

Thanks to my lab mates, colleagues, and research team María, Lucas, Rosa, Javi, Alejandro, Sara, Masoud, Pablo, Chen, Gabriele, Tom, Joan, Ricardo, for a cherished time spent together in the “drama-full lab”, and outside the lab. Special thanks to Roser and Raúl for the technical support and fruitful scientific discussion.

I am also thankful to Lia, Victor, Javier, and Carolina for being my family in Valencia.

Thanks to my Sicilian friends spread all over the world Viviana, Luca, Lucrezia, Valentina, Francesco A., Giacomo, Alessandra, Paolo, Laura, Pasquale, Annamaria, Alberto, Marta, Veronica, Claudia O., Noemi, Silvia, Francesco M., Simone, Armida, Claudia C., Rita, Naomi, for always being there even with miles apart.

Last but certainly not least, thanks to my parents for their boundless love, unconditional support, and belief in me.

