Exploring the Role of 5'UTRs in the Regulation of Gene Expression in E. coli Using Synthetic and Native Libraries

Fan Chen

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de Toulouse

---

Présentée et soutenue par

Fan CHEN

Le 6 juillet 2022

Exploration du rôle des 5'UTR dans la régulation de l'expression des gènes chez E. coli par des banques synthétiques et natives.

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :
TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

Thèse dirigée par
Laurence GIRBAL et Sébastien NOUAILLE

Jury

M. Philippe BOULOC, Rapporteur
M. Guillaume CAMBRAY, Rapporteur
Mme Laurence GIRBAL, Co-directrice de thèse
Mme Muriel COCAIGN-BOUSQUET, Présidente

Preface

纵有千古，横有八荒。

前途似海，来日方长。

It is getting late, and the sun is still shining. At this moment, I draw this book to a close. Looking out of the window, I see leaves swaying in the wind and a beluga XL skimming across the pink sky. I am feeling peaceful and leisurely exactly. Today is nothing special from the other days, but confidence increased in my heart for the future and the world. This confidence comes from what I have learned and experienced over the past four years while warmth and caring permeated this journey.

学贵得师，亦贵得友

First and foremost, I want to thank my two supervisors. Laurence and Sébastie (Séb), provided me with many precious opportunities and constant support during my Ph.D. not only academically but also in life. There are many things they taught me, but nothing was more precious than their enthusiasm to push the boundaries of human knowledge. If not for their dedication, motivation, and energy, this study and many others not covered in this thesis would undoubtedly never have achieved fruition.

我有一壶酒，可以慰风尘!

I am indebted to China Scholarship Council (CSC), Institut National des Sciences Appliquées de Toulouse (INSA-Toulouse), Toulouse Biotechnology Institute (TBI) and BLADE (BacteriaL Adaptation, Diversity and Engineering; EAD4) team - to CSC for providing financial support; to my dear department for providing a high-quality environment; to my lovely college for providing all the events and opportunities that helped me to experience the French culture.

In the end, to my dear motherland!

Résumé

Face à des changements d'environnements, les bactéries adaptent leur métabolisme en reprogrammant l'expression de leur répertoire de gènes. L'objectif de mon doctorat était d'approfondir nos connaissances de la régulation de l'expression génique chez les bactéries. Nous avons concentré nos travaux sur le rôle des 5'UTR chez *E. coli*. Les 5'UTR sont les séquences transcrites mais non traduites situées aux extrémités 5' des ARNm.

Nous avons développé une approche utilisant des séquences 5'UTR synthétiques pour analyser leur rôle à trois niveaux, que sont la traduction, la transcription et la dégradation de l'ARNm. Nous avons confirmé la contribution multiniveau des 5'UTR dans le contrôle de l'expression génique au niveau de l'initiation de la traduction, de la transcription et/ou de la stabilité de l'ARNm et montré le degré de dépendance vis-à-vis de la séquence du gène rapporteur en aval. Ensuite, nous avons joué sur la régulation de la traduction et montré les conséquences sur la concentration et la stabilité de l'ARNm.

La façon dont les 5'UTR régulent l'expression génique en réponse à des changements environnementaux n'est encore que partiellement comprise. Nous avons conçu et construit une banque exhaustive de 2547 5'UTR natives de *E. coli* fusionnées à un gène rapporteur fluorescent. Le rôle des 5'UTR natifs a été exploré en caractérisant cette banque dans différentes conditions environnementales. Nous avons démontré que les 5'UTR régulent directement et efficacement l'expression des gènes en aval et contribue ainsi à l'adaptation d'*E. coli* à son adaptation.

Abstract

When facing changing environments, bacteria have to adapt their metabolism by modifying the expression pattern of their gene repertoire. The goal of my PhD was to provide additional fundamental insights into gene expression regulation in bacteria. We have focused our work on the role of 5'UTRs in gene expression regulation in *E. coli*. 5'UTRs are the transcribed but untranslated sequences located in the 5'ends of the mRNAs.

We first developed an approach using synthetic 5'UTR sequences to analyze their role at three levels, namely translation, transcription, and mRNA degradation. We confirmed the multilevel contribution of 5'UTRs in the control of gene expression at the level of translation initiation, transcription, and/or mRNA stability and showed the degree of dependence on the downstream reporter gene sequence. Then we played with the 5'UTR-mediated translation regulation and showed the consequences on mRNA concentration and stability.

How 5'UTRs regulate gene expression in response to environmental changes is still only partially understood. We designed and constructed a full-size library of 2547 native 5'UTR sequences from *E. coli* fused to a fluorescent reporter gene. The role of native 5'UTRs was explored by challenging the native 5'UTR library to grow in changing environmental conditions. We demonstrated that the 5'UTRs directly and efficiently regulate downstream gene expression and thus contribute to the *E. coli* adaption to changing environments.

TABLE OF CONTENTS

Introduction

When facing changing environments, cells have to adapt their metabolism by modifying the expression pattern of their gene repertoire. The control of gene expression encompasses a wide range of regulatory mechanisms that occur at each step of gene expression, from transcription initiation to post-translational modifications of proteins. However, it is not yet fully understood how this multi-level process of gene expression is coordinated and how it responds to changing growth conditions. The goal of my PhD was to provide additional fundamental insights into gene expression regulation in bacteria. In turn, a better understanding of gene expression regulation will support the development of molecular tools in biotechnology and synthetic biology.

We have chosen to focus our work on the role of 5'UTRs in gene expression regulation in *Escherichia coli*. 5'UTRs are the transcribed but untranslated sequences located in the 5'ends of the mRNAs, starting from the transcription start site and ending at the first nucleotide of the translational start codon. The bibliographic review provides an overview of the elements located in the 5'UTR known to be involved in gene expression regulation in *E. coli*. It clearly demonstrates that the 5'UTR of an mRNA is a region rich in regulatory elements of gene expression that contribute to the regulation of transcription, mRNA stability and/or translation.

Although many studies were conducted to better understand 5'UTR-mediated regulation of gene expression in *E. coli*, integration of the three levels of 5'UTR-mediated regulation of transcription, mRNA stability and translation is rarely performed. Therefore, in Chapter 1, we developed an approach using synthetic 5'UTR sequences to analyze their role in regulating gene expression in *E. coli* at the three levels of translation, transcription and mRNA degradation. The 5'UTR-mediated regulation of translation initiation, transcription and/or mRNA stability is discussed by analyzing the protein expression level, mRNA concentration and stability of three reporter genes under the control of a set of synthetic 5'UTRs. The results confirm the multilevel contribution of 5'UTRs in the control of gene expression at the level of translation initiation, transcription and/or mRNA stability, and show the degree of dependence on the downstream reporter gene sequence. Furthermore, in Chapter 3, we played with the 5'UTR-mediated translation regulation and followed the consequences on mRNA concentration and stability.

In addition, how 5'UTR-mediated gene expression regulations in response to environmental changes is still only partially understood. Therefore, in Chapter 2, we designed and constructed a full-size library of native 5'UTR sequences from *E. coli* fused to a fluorescent reporter gene. The role of native 5'UTRs

in regulating gene expression for adaptation of *E. coli* cells to different growth conditions was explored by challenging the native 5'UTR library to grow in changing environmental conditions (temperature and medium composition). The results demonstrate that the 5'UTR-dependent regulation directly and efficiently alters the downstream gene expression and thus contributes to *E. coli* adaption to changing environments.

In the last section of the manuscript, we conclude on the role of 5'UTRs in gene expression regulations at the three levels of translation initiation, transcription and mRNA stability and in response to changing growth conditions. These results open perspectives for further experimental work that could be initiated to understand 5'UTR-mediated regulation further. We also discuss how 5'UTR sequences identified in our study could be used as new molecular tools to better control the expression of genes of interest in *E. coli* for biotechnology applications.

BIBLIOGRAPHIC REVIEW

1 Global regulation of gene expression in *E. coli*

As one of the most important model organisms, *E. coli* plays an extremely important role in modern life science research, especially in the fields of molecular genetics and bioengineering with its clear genetic background and manipulation superiority (Blount, 2015). This research uses the non-pathogenic *E. coli* K-12 MG1655.

Gene expression is the fundamental process of synthesis of functional gene products guided by genetic information. The products can be protein or non-coding RNA (ncRNA) such as transfer RNA (tRNA) and ribosomal RNA (rRNA). Through regulation, gene expression is ordered in time and space and can adapt to a changing environment. Specifically, gene expression regulation encompasses a wide range of sophisticated mechanisms that are involved at every step of gene expression, from transcription initiation to post-translational modifications of proteins (Figure 1). The main factors influencing the expression process of a functional protein can be related to sequence features (e.g. promoter or codon bias), molecule stability (for mRNA and protein) and molecule-molecule interactions (e.g. DNA and RNA polymerase, mRNA and ribosomes).

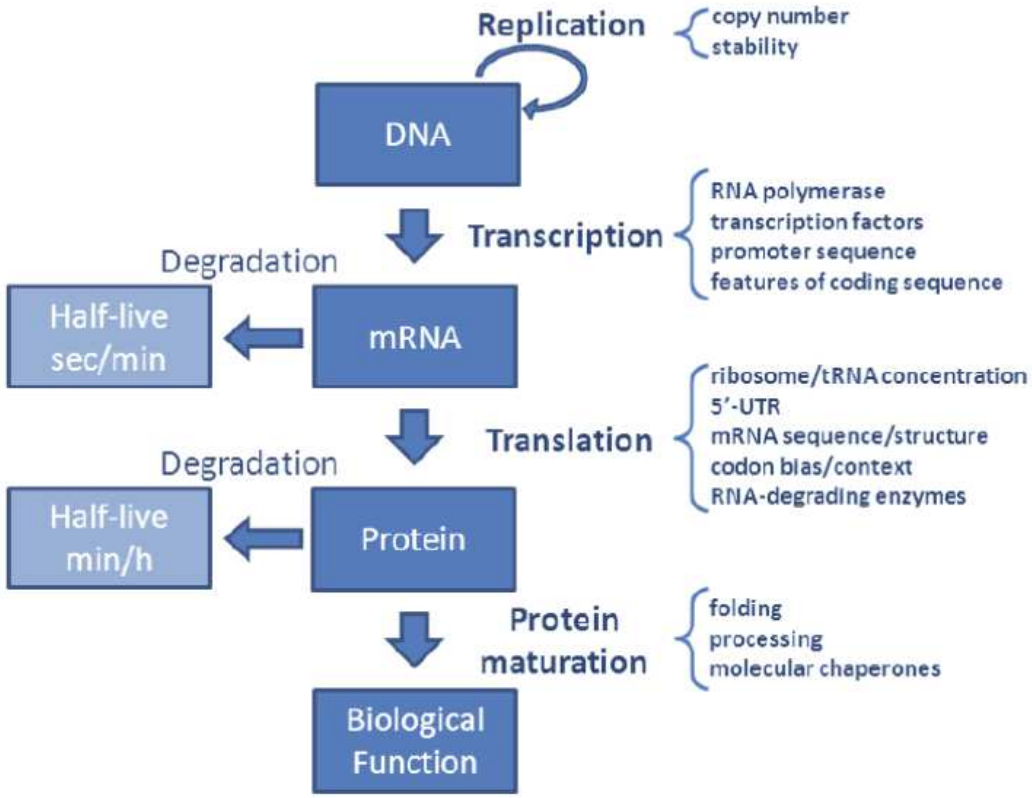


Figure 1. Factors influencing gene expression in *E. coli* (Kucharova, 2012)**.** Multiple positive and negative factors regulate each step of gene expression, from DNA to proteins with biological functions. In terms of the overall picture of gene expression, mRNA abundance is a key determinant of protein levels. mRNA abundance is the result of a balance between its transcription and degradation during

synthesis and reflects the number of template molecules available for translation. We will show in this literature review how the 5'UTR (untranslated transcribed region of an mRNA) has been shown to be as a potential 'regulatory-hub' not only for translation of the mRNA molecule but also for its stability and synthesis.

1.1 Process of transcription in *E. coli*

1.1.1 RNA polymerase

Transcription is the initial step in gene expression, where genetic information is transferred from DNA to RNA under the catalysis of RNA polymerase (RNAP). Since the discovery of RNAP in 1960, a wealth of information about its structure and function has been obtained (Hurwitz, 2005). Contrary to eukaryotes, only one type of core RNAP exists in bacteria and ensures the transcription of all RNA species. It is a large molecule (~400 KDa) estimated to be present at about 3600-6000 copies in *E. coli* for the transcription of ~ 4453 genes identified in the genome (Sun et al., 2019). The *E. coli* RNAP core enzyme (E) is constituted of five subunits (α2ββ'ω, ***Figure 2***), which can combine with one of the seven σ factors to form a holoenzyme. Although without σ factor the RNAP core enzyme still has catalytic activity, it has no selectivity to identify promoter sequences to initiate correct transcription initiation (Fredrick & Helmann, 1997).

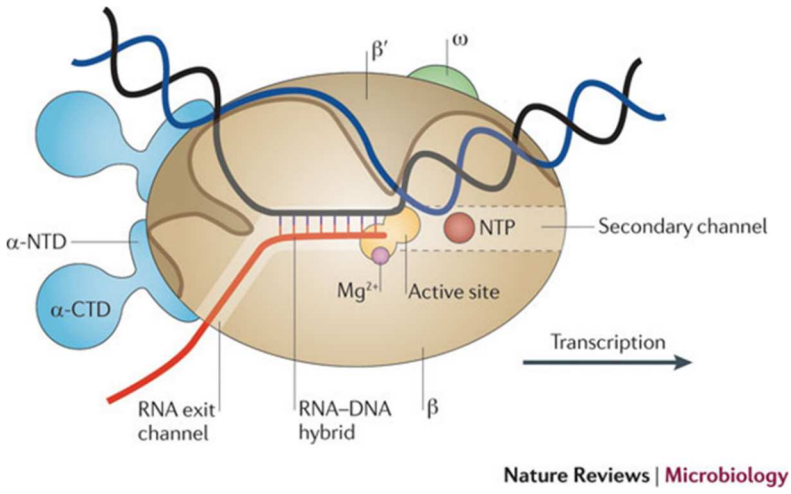


Figure 2. Schematic model of the RNAP during transcription elongation (Santangelo & Artsimovitch, 2011). The *E. coli* RNAP core enzyme (composed of an α-dimer, a β-subunit, a β'-subunit and an ω-subunit) is bound to the DNA duplex composed of the template strand (black) and the non-template stand (blue), and the nascent RNA (red).

The α subunit, encoded by the *rpoA* gene, is present in the RNAP as a homodimer and is composed of two structural domains (αNTD and αCTD in the N-terminal and C-terminal parts,

respectively), which interact for assembling the β and β' subunits and also interact with transcription factors to initiate and regulate transcription (Murakami et al., 1997).

The β and β' subunits are encoded by the *rpoB* and *rpoC* genes, respectively (Aboshkiwa et al., 1992). They are the two largest subunits of the core RNAP (E), accounting for 60 % of the total mass (Severinov et al., 1997). The β and β' subunits generate a cleft allowing DNA access to the active site of the enzyme. The β' subunit binds the DNA template and opens the double-stranded DNA, while the β subunit interacts with the incoming nucleotide (Sutherland & Murakami, 2018). Transcription initiation can be inhibited by rifampicin, which is widely used to study RNA degradation. Crystal structure and biochemical data suggest that rifampicin activity is due to its binding to the pocket of RNAP β subunit within the DNA/RNA channel but away from the active site. This spatial occlusion physically blocks the path of the elongating RNA (Campbell et al., 2001).

The ω subunit is encoded by the *rpoZ* gene, it helps the folding of the β' subunit and assists RNAP assembly (Patel et al., 2020).

In *E. coli*, seven σ factors have been identified: σ^{70} (or σ^{D}), σ^{54} (or σ^{N}), σ^{38} (or σ^{S}), σ^{32} (or σ^{H}), σ^{28} (or σ^{F}), σ^{24} (or σ^{E}), σ^{18} (or σ^{I}) (Cho et al., 2014). The σ factors contain the promoter recognition domains that trigger specificity to the RNAP core enzyme to bind and initiate transcription at the appropriate location of the DNA template. After transcription initiation, the σ factor dissociates from the transcription initiation complex (Feklístov et al., 2014). The main σ factor σ^{70}, considered a housekeeping sigma factor, is involved in the transcription of most genes in exponential growth phase (Paget & Helmann, 2003). Conversely, the other sigma factors are considered as alternative, engaged in the transcription of dedicated genes related to stress responses, adaptive responses or particular processes (Ishihama, 2000). Some of the functions transcribed by alternative sigma factors are summarized in ***Table 1***.

Table 1. Components of the RNAP and their functions in *E. coli* (Bacun-Druzina et al., 2011)

Subunit		Gene	Function
α		*rpoA*	required for assembly of the enzyme; interacts with some regulatory proteins; involved in catalysis
β		*rpoB*	involved in catalysis, chain initiation and elongation
β’		*rpoC*	binds to the DNA template
ω		*rpoZ*	required to restore denatured RNAP *in vitro* to its fully functional form
	σ^{70}	*rpoD*	transcription of most genes during the exponential phase
	σ^{54}	*rpoN*	nitrogen-regulated gene transcription
	σ^{38}	*rpoS*	gene expression during the starvation and stationary phases
σ	σ^{32}	*rpoH*	heat-shock gene transcription
	σ^{28}	*rpoF*	expression of flagellar and chemotaxis genes
	σ^{24}	*rpoE*	response to the extra cytoplasmic and extreme heat stress
	σ^{19}	*fecI*	regulation of the *fec* genes for iron dicitrate transport

1.1.2 Basic processes of transcription

The mechanism of the transcription process has undergone extensive research for a long time. Overall, the transcription process could be considered as the three sequential stages of initiation, elongation and termination.

Transcription initiation

The process of transcription begins with a multi-step (Ruff et al., 2015). The RNAP recognizes and binds to DNA elements within a promoter sequence, which triggers a series of conformational changes (Browning & Busby, 2004). Therefore, each step on the path from free RNAP and promoter to the final transcriptionally competent complex is an opportunity to regulate transcription. The currently recognized mechanism of transcription initiation can be separated in three steps (Chen et al., 2020; Davis et al., 2007; Glyde et al., 2018).

First, the σ factor and the core of multiple subunits (α2ββ'ω) assemble into the complete holoenzyme RNAP, searching and specifically recognizing the double-stranded promoter elements on the DNA to form an initially closed complex (RPc). This leads to DNA bend, which wraps RNAP and facilitates DNA melting and its insertion into the active-site cleft of RNAP (Saecker et al., 2011).

In a second step, the initial closed complex shift into an open complex (RPo). This step undergoes a series of conformational changes. Meanwhile, the DNA unwound to form a transcription bubble at the nucleotide position around -11 to +3 from the transcription initiation site. The polymerase 'clamps' fully onto the DNA lying in the DNA channel. That would permit the template DNA strand to access the active site in the cleft of RNAP (Hook-Barnard & Hinton, 2007).

Finally, after the stable open complex (RPo) is completely formed, RNAP starts *de novo* RNA synthesis (Mazumder & Kapanidis, 2019). Transcription process turns from initiation to elongation phase after the process known as promoter escape (Carpousis & Gralla, 1980). However, before achieving full promoter escape, multiple failed start-up cycles occur named abortive initiations (Hsu, 2009). Abortive initiation is a common transcriptional process that occurs widely in eukaryotes and prokaryotes, during which RNAP enters a cycle of synthesis and release of short mRNA transcripts (2-12 ntd in length, although abortive transcripts as long as 15–17 ntd have been reported) (Hsu, 2002). After the release of short RNA products, RNAP returns to the open complex state, subsequently reinitiating RNA synthesis. Abortive initiation continues to occur until a 9 to 11 ntd long RNA is synthesized (Revyakin et al., 2006). If RNAP can successfully escape from the promoter region and the abortive process, RNAP enters the elongation phase. The promoter clearance depends on an equilibrium between abortive and efficient transcription initiation but the exact determinants are not

fully known. The proposed mechanism is based on interaction between RNAP and DNA. The RNAP remains stationary with respect to the upstream side of DNA during the abortive initiation, while unwinding downstream DNA provides the energy to disrupt the interaction between RNAP and the promoter for escape (Hsu, 2002). This equilibrium between abortive and productive transcription seems to depend on the promoter and the nature of the initially transcribed sequence*, i.e.,* the 5'UTR sequence for mRNA (Mazumder & Kapanidis, 2019).

Transcription elongation

The transcription process moves to the elongation phase when RNAP has escaped from the promoter. The transcription elongation complex oscillates between the pre- and post-translocated states driven by thermal fluctuations, whereas the growing RNA and incoming NTP substrates direct the assembly towards forward translocation (Bar-Nahum et al., 2005; Douglas et al., 2020; Vassylyev et al., 2007). *E. coli* RNAP is like a powerful molecular motor, transcribing at an average rate of 20-90 ntd/s (Malinen et al., 2012). In general, a single elongation cycle can describe as the following three basic states (***Figure 3***):

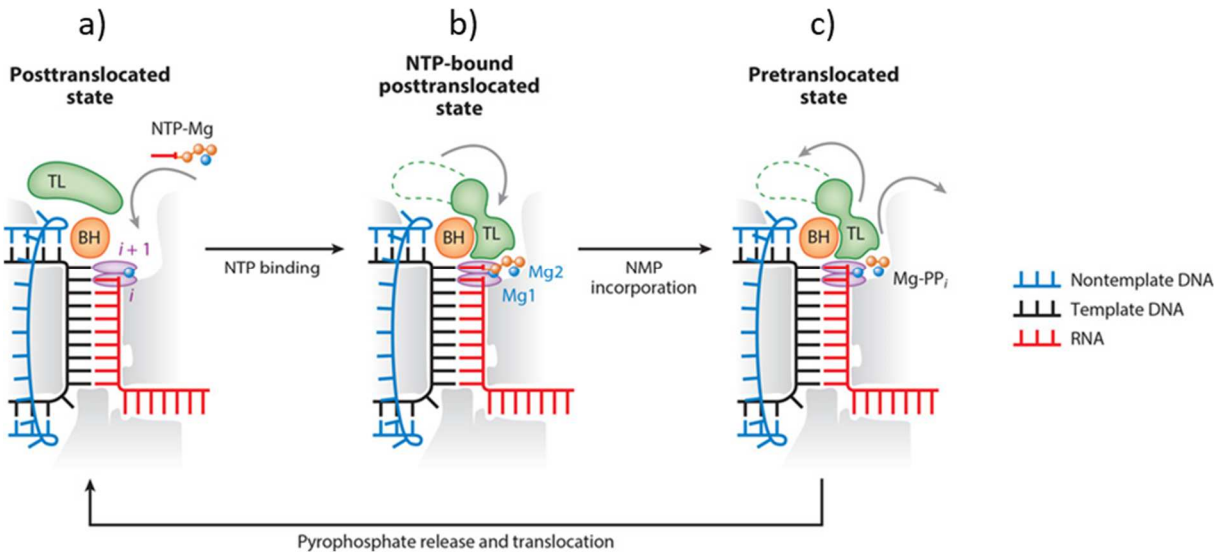


Figure 3. Schematic overview of the nucleotide addition cycle in transcription elongation (Belogurov & Artsimovitch, 2015)**.** BH, bridge helix; NMP, nucleoside monophosphate; NTP, nucleoside triphosphate; PPi, pyrophosphate. a). Post-translocated state. Following the elongation cycle, the active catalysis site was accessible due to the forward translocation of RNAP. b). NTP-bound post-translocated state. A template-complementary NTP binds to the RNA-DNA hybrid. Meanwhile, multiple conformational changes of RNAP components promote nucleoside addition (Hein & Landick, 2010; J. Zhang et al., 2010). Simultaneously, a NMP (from a templated NTP substrate) reacts with the 3'-OH of the growing RNA chain. c). Pre-translocated state. An NMP is added to the growing RNA during the catalytic reaction, and PPi is released. The active site opens, and the RNAP maintains an 8- to 9-base pair (bp) RNA-DNA hybrid within the 11- to 12-bp melted DNA bubble; forward movement translocates into the next elongation cycle.

Transcriptional pausing

However, transcriptional elongation is not a smooth and continuous process, and some transcription pauses can occur due to the presence of DNA-binding proteins, lesions in the template strand or nucleotide misincorporations (Belogurov & Artsimovitch, 2015; Roberts et al., 2008). The speed of RNAP elongation is highly modulated by accelerations, decelerations, pausing and even complete terminations due to mRNA sequence features or several general and operon-specific factors (Kammerer et al., 1986; Malinen et al., 2012; Nickels et al., 2004; Wade & Struhl, 2008). Some studies have reported that RNAP pauses at various loci can be in the order of tens of seconds or longer (Herbert et al., 2006; Kireeva & Kashlev, 2009). Functionally, pausing can provide time for the elongation complex to recruit regulatory factors, participate to efficient coupling of transcription and translation and is an obligatory step in termination (Artsimovitch & Landick, 2002; Landick, 2006; Richardson, 1991). It is widely accepted that sequence-specific interactions of DNA and RNA with RNAP initially trigger most of the pauses (Landick, 2006; Neuman et al., 2003; Saba et al., 2019; Weixlbaumer et al., 2013).

Two classes of pauses have been described (***Figure 4***). The first one is the backtrack pause, in which the transcribing RNAP moves backwards along with the template, reversing the translocation steps that assemble the RNA chain but not depolymerize the chain itself. The second one is the hairpin pause, in which the formation of a nascent RNA hairpin in the RNA exit channel, wedging open the clamp in the pre-translocated hairpin pause (Artsimovitch & Landick, 2002).

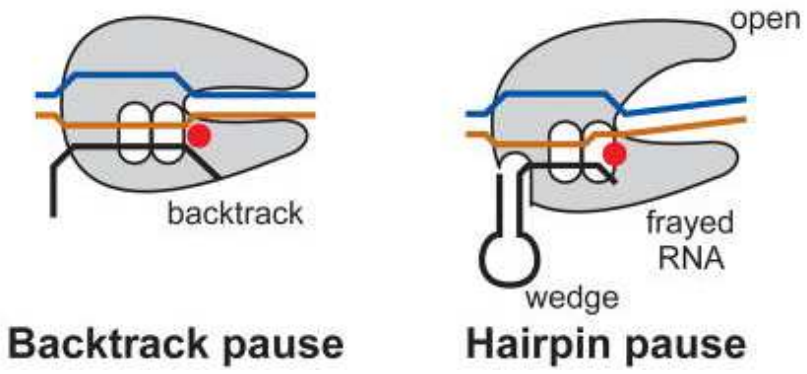


Figure 4. Schematic representation of the model for transcriptional pausing (Weixlbaumer et al., 2013). DNA (blue and brown lines), mRNA (black lines), RNAP (gray clamp), and bridge helix (red circle). The kinked bridge helix (due to RNAP backtrack or an RNA pause hairpin formed) blocks template access to the active site, resulting in transcription pausing.

Transcription termination

The last step in transcription is the termination, consisting in the dissociation of the transcription complex to release the full-length transcribed RNA and the RNAP from DNA. The two main transcription termination mechanisms described in *E. coli* are the Rho-dependent and Rho-independent terminations (***Figure 5***).

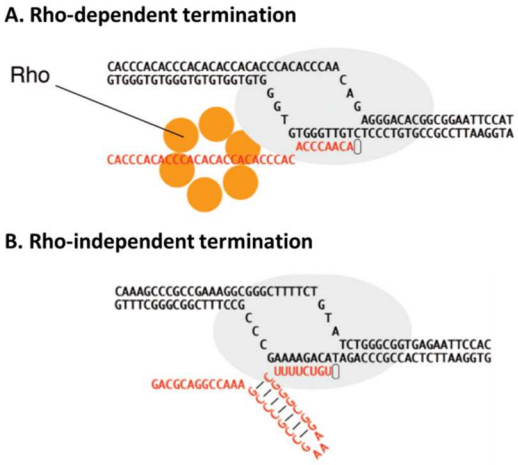


Figure 5. Two mechanisms of transcription termination (Park & Roberts, 2006). A) Rho-dependent termination. B) Rho-independent termination process. See below for a detailed description.

Rho-dependent termination is ensured by the action of the Rho factor, the main actor for transcription termination (Banerjee et al., 2006). This protein is an ATP-dependent RNA translocase, which binds to nascent RNA and pulls it out from the transcription complex (Roberts et al., 2008). Structurally, Rho is a broken homohexameric ring (a "lock washer" structure), which provides a central cavity for RNA to pass through (Skordalakes & Berger, 2003). Rho protein binds with the RNA transcript and moves along the RNA polymerase in 5'-3' direction. When Rho reaches the transcription bubble, it pulls DNA/RNA hybrid apart by dissociating the hydrogen bounds between the DNA template and the RNA transcript and releases the transcript from the transcription bubble. Rho terminates transcription in response to DNA signals transcribed into RNA called the Rho-dependent terminators. There are two fundamental features of a Rho-dependent terminator: a proximal Rho binding site called the '*rut* site' (for Rho utilization) and a distal sequence comprising the termination zone (Banerjee et al., 2006). The Rho binding site is a highly degenerated region of ~ 70-80 ntd that is rich in cytosine and lacking secondary structures (Alifano et al., 1991; Bear et al., 1988; Morgan et al., 1985). The *rut* sites are highly degenerated and therefore difficult to identify (González-González et al., 2017). When Rho scans the RNA, the presence of translating ribosomes on mRNA prevents the Rho-dependent transcription termination.

The Rho-independent termination process (also called intrinsic termination) does not require trans-acting factors and is based on hairpin formation (J. Chen et al., 2019). The intrinsic terminator has two distinctive fundamental features. One is the presence of a GC-rich dyad repeat for the formation of a stable GC-rich hairpin located about 8-9 ntd upstream of the RNA release site. Another feature is a stretch of A's in the template DNA strand, which encodes an adjacent U-rich segment right after the hairpin forming region (Belogurov & Artsimovitch, 2015; Peters et al., 2011, 2011). A three-step model described this transcription termination process: i) The A rich sequence in the template triggers a pausing of RNAP; ii) The pausing provides time for the hairpin formation which then destabilizes the transcription complex by shortening the DNA-RNA hybrid and weakening interactions; and iii) the transcript is pulled out from the transcription complex under the force of hairpin formation, the transcript is released and the transcription complex dissociates (Roberts, 2019; Roberts et al., 2008).

1.2 Process of mRNA degradation in *E. coli*

mRNA degradation involves multiple enzymes and cofactors in all organisms' life activities. mRNA degradation and associated regulation are considered as a strategy for cells to actively respond to changing environments by modulating the concentration of individual mRNAs and then reshaping their transcriptome. For individual mRNAs of *E. coli*, various studies at the omics-scale have measured a wide disparity of half-lives (Bernstein et al., 2002; H. Chen et al., 2015; Esquerré et al., 2014). Half-lives vary from less than 1 minute for the most unstable to more than 30 min for the most stable in extreme cases, but the vast majority (~80%) of mRNAs have half-lives from 1 to 5 min. Three classes of proteins are involved in mRNA degradation: the endoribonucleases, which cleave the mRNA inside the molecule; the exoribonucleases, which degrade mRNA by their extremities; and the oligoribonuclease, which degrades short mRNA fragments into mononucleotides. Many endo-and exo-ribonucleases have been characterized in *E. coli* as well as their mechanisms. Next, we will summarize the main representatives and their mode of action.

1.2.1 Proteins involved in mRNA degradation

Many proteins are involved in RNA degradation. ***Table 2*** lists endoribonucleases, exoribonucleases, the oligoribonuclease, scaffold proteins and some chaperones involved in the degradation of mRNA in *E. coli.* Degradation of mRNA is not completely blocked by inactivation of a single RNA-degrading enzyme as many enzymes have redundant activities (Arraiano et al., 2010; Houseley & Tollervey, 2009).

Table 2. Enzymes and cofactors involved in mRNA degradation in *E. coli*

Name	Gene	Mode of action	Specificity
RNase E	*rne*	Endoribonuclease	monophosphorylated 5'end dependent hydrolase Degradosome scaffold
RNase III	*rnc*	Endoribonuclease	Acting on double-stranded structures
RNase G	*rng*	Endoribonuclease	Homologue of RNase E
RNase LS	*rnlA*	Endoribonuclease	Related to *rnlAB* TA system
RNase P	*rnpA, rnpB*	Endoribonuclease	Specifically cleaves a small number of polycistronic operon mRNAs in intercistronic regions.
RNase Z/BN	*rnz*	Endoribonuclease/ 3' →5' exoribonuclease	Double functions of endo- and 3'exo-ribonucleases
RNase R	*rnr*	3' →5' exoribonuclease	Hydrolytic cleavage not subjected to stem-loop as long as there is a single-stranded 3' end
PNPase	*pnp*	3' →5' exoribonuclease	Phosphorolytic cleavage of diphosphate nucleoside terminal Inhibited by 3'end secondary structure
RNase II	*rnb*	3' →5' exoribonuclease	Hydrolase

			Inhibited by 3'end secondary structure
Oligoribonuclease	*orn*	3' →5' exoribonuclease	Specific for small oligoribonucleotides
RhlB	*rhlB*	Helicase	Untie the double-stranded structure
RhlE	*rhlE*	Helicase	Untie the double-stranded structure
SrmB	*srmB*	Helicase	Untie the double-stranded structure
CsdA	*csdA*	Helicase	Untie the double-stranded structure
RNA diphosphohydrolase	*rppH*	Phosphatase	Monophosphorylation of 5'end triphosphate
Hfq	*hfq*	RNA-binding protein	Recruits ncRNA and RNase E complementary pairing mRNA
Poly(A) polymerase	*pcnB*	Polymerase	Polyadenylation of 3'end, destabilizes mRNA
MazF/ChpBK/PemK	*mazF/chpBK/ pemK*	mRNA interferase	Specific cutting sites: ACA, ACY (Y is U, A or G) and UAH (H is C, A or U)
RelE	*relE*	mRNA interferase	Ribosome-related factor, promotes mRNA cleavage at ribosomal A site
HicA	*hicA*	mRNA interferase	Randomly cuts mRNA without any ribosome involvement
MqsR	*mqsR*	mRNA interferase	GCU specific cutting site
YoeB	*yoeB*	mRNA interferase	Combines with the 50S subunit in the 70S ribosome and cleaves efficiently at ribosomal A site

Ribonucleases are not only involved in the mRNA degradation process. They are also required for all processes of RNA metabolism including ribosomal RNA and tRNA maturation, degradation, turnover and quality control (Bechhofer & Deutscher, 2019; Deutscher, 2015). RNases that contribute to mRNA degradation are roughly divided into two types: endoribonuclease and exoribonuclease.

1.2.1. A) Endoribonucleases

The main endoribonucleases in *E. coli* are RNase E, RNase III, and RNase G. Others such as RNase P, RNase LS, and RNase Z have more specific activities related to small number of transcripts or to specific transcripts, and will not be described here (***Table 2***).

RNase E, the central ribonuclease in *E. coli*

As an essential endoribonuclease in *E. coli*, RNase E is considered the main degradative enzyme, which triggers the degradation of more than 50%-60% of mRNA during the exponential growth phase (Bechhofer & Deutscher, 2019; Stead et al., 2011). It was initially discovered that RNase E prefers to specifically cleave single-stranded mRNA in AU-rich regions (McDowall et al., 1994). The cleavage sites of RNase E can be found in many positions all along single-stranded mRNA (Belasco, 2017). Analyzes of the primary sequences and putative secondary structures around 22,033 RNase E potential cleavage sites show little secondary structure, but an enrichment in AU-rich sequences exists around cleavage sites (Chao et al., 2017). A minimal 5-nucleotide RNase E consensus cleavage site was found, with 'RN↓WUU' as a core motif (with R as G/A, was A/U, and N is any nucleotide) and a strong preference for uridine at the +2 position (Chao et al., 2017). The cleavage ability of RNase E is also hindered by physical obstacles between the 5' end and the cleavage sites, such as bound ribosomes, RNA binding proteins or small RNAs (Richards & Belasco, 2019).

RNase E: two different pathways

There are two initial pathways for RNase E-mediated mRNA degradation: the 5' end-dependent pathway and the direct-access pathway (***Figure 6***). No matters the pathway, the initiated mRNA degradation mechanism results in two consequences: i) the fragments lacking the protection of the 3' end stem-loop are generated and are rapidly degraded by 3 'exoribonucleases; ii) the fragments exposing the monophosphate 5' end are generated and are further cleaved by endoribonucleases.

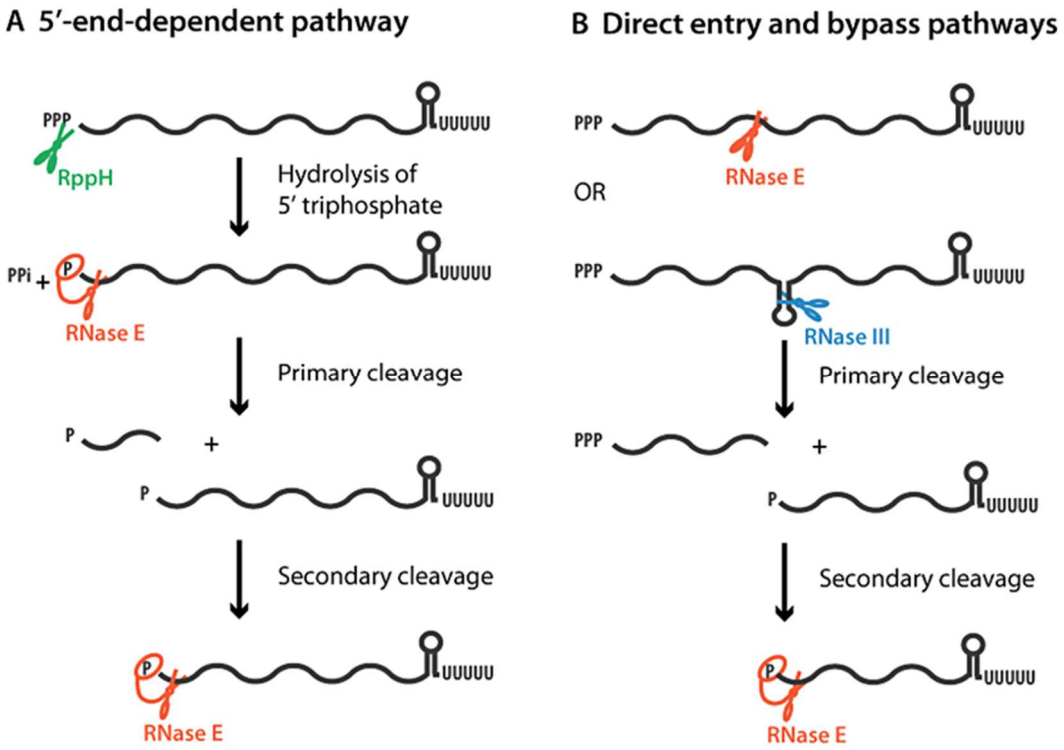


Figure 6. The two RNase E-mediated mRNA degradation pathways in *E. coli* (Bouvier & Carpousis, 2011). RppH (green scissors), RNase E (red scissors), RNase III (blue scissors), and mRNA (black lines). See below for a detailed description.

❖ 5′ end-dependent pathway (**Figure 6A**)

In this pathway, RNA pyrophosphate hydrolase (RppH) is responsible for the monophosphorylation of the triphosphate 5′ end (Richards & Belasco, 2019). mRNAs with a tri-phosphorylated 5′end are generated by transcription but tri-phosphorylated 5′ends are poor substrates for RNase E. By contrast, RNase E activity is increased for 5′ mono-phosphorylated ends (Mackie, 1998). The 5′ end-dependent access requires thus the conversion of the 5′-terminal triphosphate to monophosphate by RppH (Bandyra et al., 2018). This conversion stimulates RNase E scanning at the 5′ mono-phosphorylated ends and the search along the linear single-stranded of cleavage sites to initiate the degradation pathway. The cleavage sites are sometimes far downstream of the 5′ mono-phosphorylated end (***Figure 7B***). The crystal structure of RNase E reveals that the N-terminal catalytic domain forms a homotetramer (***Figure 7A***) which organized as a dimer of dimers (Callaghan, Marcaida, et al., 2005; Callaghan, Redko, et al., 2005). It contains two sites that interact with the substrate, one is the catalytic center and the other is the ‘pocket’ structure used to bind the monophosphorylated 5′ end. The ‘mouse-trap’ model has been proposed: after the ‘pocket’ captures the monophosphorylated 5′ end, the dimers clamp down and move on the substrate allowed the catalytic site to better identify and cut (Bandyra & Luisi, 2018; Koslover et al., 2008; Mackie, 2013).

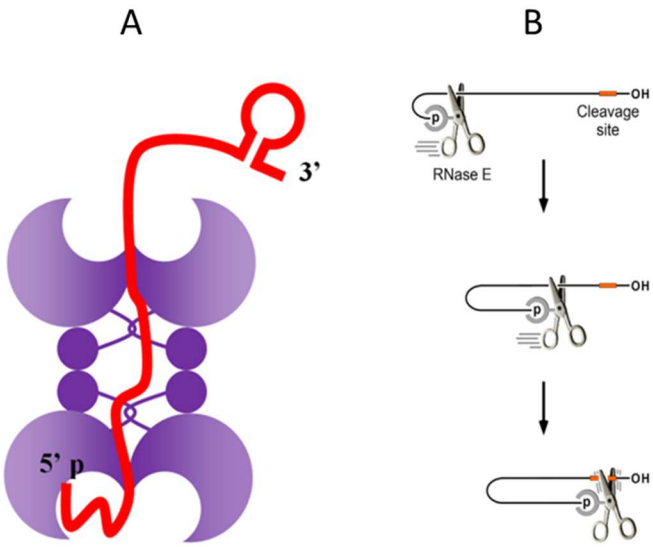


Figure 7. Illustration of RNase E initiation of mRNA degradation by 5' end-dependent pathway. (A). Interactions of RNase E (purple) with RNA (red) (Bandyra et al., 2018). (B). Scanning and cleavage of 5'-monophosphorylated RNA by RNase E (Richards & Belasco, 2019).

❖ Direct-access pathway (Figure 6B).

Although the 5' end-dependent pathway has a prominent contribution to the rapid degradation of mRNA, it has also been found that there is a 5' end-independent direct-access pathway to initiate degradation. RNase E can rapidly cleave some transcripts regardless of the phosphorylation status of the 5' end (Baker & Mackie, 2003; Hankins et al., 2007; Joyce & Dreyfus, 1998). For instance, *cspA* mRNA triphosphate 5' end is not subject to RppH monophosphorylation, but this mRNA can still be recognized and cleaved by RNase E (Kime et al., 2009). In addition, *epd-phk* mRNA cleavage by RNase E is also an example of a direct-access pathway (Bardey et al., 2005; Kime et al., 2009).

RNase E and the degradosome

From the structural point of view, RNase E can be divided into two domains (Bandyra & Luisi, 2018). The N-terminal domain (NTD) encompasses the endonucleolytic active site, and the C-terminal domain (CTD) is used as the scaffold of the degradosome. RNase E contains a membrane anchor domain that localizes the degradosome at the inner membrane surface by interactions between its alkaline hydrophobic residues and lipid membrane (Bandyra et al., 2013; Bandyra & Luisi, 2018). The localization at the membrane surface participates to the regulation of RNA degradation. Detachment of RNase E from the inner cytoplasmic membrane by deletion of the membrane anchor domain leads to an overall slowdown of mRNA degradation (Hadjeras et al., 2019).

The degradosome is a supramolecular complex (***Figure 8***) which plays a vital role in mRNA degradation and is the leading participant in the post-transcriptional regulation of gene expression in

E. coli (Tejada-Arranz et al., 2020). RNase E is the core component of the degradosome. Its CTD region provides a scaffold for the recruitment of other enzymes (Górna et al., 2012). The four core components of the degradosome are RNase E, the 3'exoribonuclease PNPase, the RhlB helicase, and the glycolytic enzyme enolase (Bernstein et al., 2002). The composition of the degradosome is variable and may be different under the stationary phase or temperature stress, but it contains at least one RNA helicase in addition to the RNase E (Hadjeras et al., 2019). The highly flexible assembly of degradosome is considered as an advantage for better adaptation to different scenarios (Bandyra & Luisi, 2018; Hadjeras et al., 2019). For example, CsdA has been reported to be associated with the degradosome during cold shock (Charollais, 2004). A study demonstrated that during the exponential phase and the stationary phase, there are many other proteins (PAP I, Hfq, SrmB, Hrp A, and RNase R) observed in the degradosome besides the four core components (Bruce et al., 2018; Carabetta et al., 2010; Purusharth et al., 2005).

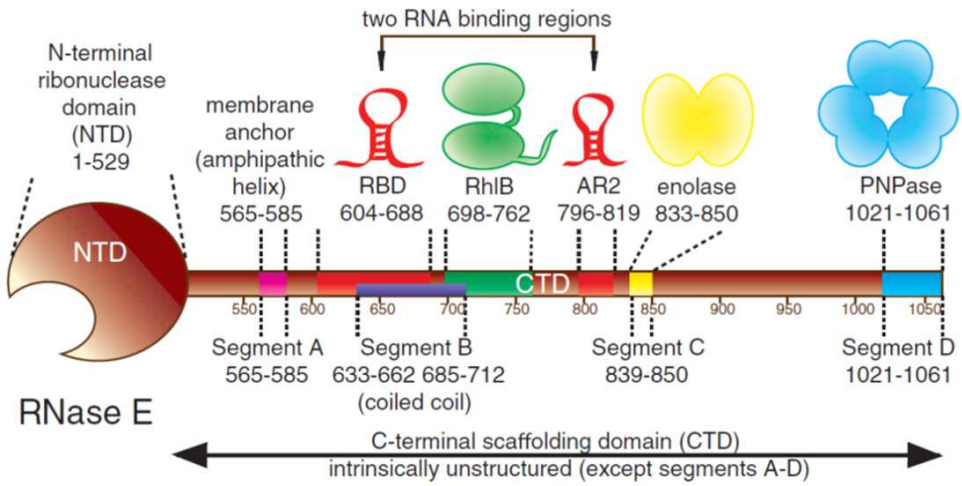


Figure 8. Schematic representation of degradosome principle components (Górna et al., 2012). The degradation components and the sites on which they bind on RNase E are marked with the corresponding colors. RNase E (brown), membrane anchoring sequence (pink), two RNA binding regions (red), RhlB (green), enolase (yellow), and PNPase (blue).

Other main endoribonucleases

RNase III encoded by the *rnc* gene cleaves double-stranded RNA. It can affect the stability of mRNA by eliminating potential protective stem-loop structures (Court et al., 2013). For example, RNase III affects the synthesis of the exoribonuclease PNPase by cutting the stem-loop structure in the *rpsO-pnp* polycistronic operon (Kushner, 2002). RNase III also contributes to the processing of the double-stranded stem located in 30S rRNA polycistronic transcripts to release 17S, 25S, and 9S rRNA precursors, as well as the embedded tRNAs (Court et al., 2013). The involvement of RNase III in the degradation of mRNA appears quite limited in *E. coli* (Court et al., 2013). RNase III can be used as a secondary participant to regulate the stability of approximately 10% of the mRNAs in *E. coli* (Bechhofer & Deutscher, 2019).

RNase G is the homologue of RNase E, with an N-terminal structure highly similar to RNase E but lacks the scaffold C-terminal region of the degradosome (Aït-Bara et al., 2015). Although RNase G has some functional homology with RNase E, it cannot fully compensate for RNase E activity (Umitsuki et al., 2001) since *E. coli* cells lacking RNase E activity are unviable (Hughes, 2016).

The other endoribonucleases, RNase P (Li & Altman, 2003), RNase LS (Otsuka & Yonesaki, 2005), RNase Z (Dutta & Deutscher, 2009) and mRNA interferases (MazF, RelE, MqsR, etc.) show limited relevant function in mRNA degradation in *E. coli.*

1.2.1. B) Exoribonucleases

E. coli cells contain several 3' exoribonucleases but no 5' exoribonuclease (Bechhofer & Deutscher, 2019; Kushner, 2002; Mathy et al., 2007). The type 3'-exoribonuclease mainly includes PNPase, RNase II, RNase R, RNase BN and the oligoribonuclease Orn.

PNPase participates to the degradation of mRNA as its inactivation leads to an increase in transcript half-life (Mohanty & Kushner, 2003). The level of PNPase is controlled by RNase III as described above. In strains lacking RNase III, PNPase levels increase tenfold (Portier et al., 1987; Takata et al., 1987). PNPase cleaves mRNA from the 3' end, releases the terminal diphosphate nucleoside and is only active on single-stranded mRNAs because it is strongly inhibited by secondary structures (Bechhofer & Deutscher, 2019; Cheng & Deutscher, 2005). PNPase can be present in the cell either in association with the degradosome or in association with the poly(A) polymerase I (PAP I) and the chaperone Hfq (Mohanty et al., 2004).

Unlike PNPase, RNase II and RNase R cleave mRNA via a hydrolytic reaction and release the terminal monophosphate nucleoside (Cheng & Deutscher, 2005; Mohanty & Kushner, 2018). However, the degradation of mRNAs by RNase R is not affected by the presence of a structured stem-loop as long as there is a single-stranded 3' end (at least a 7-nt long) (Hossain et al., 2016). Therefore, it has

been suggested that RNase R can be used as both exoribonuclease and helicase-like to resolve secondary structures (Awano et al., 2010; Hossain et al., 2016). The level of RNase II is regulated by PNPase, RNase III and RNase E: PNPase regulates by degrading *rnb* mRNA (gene encoding RNase II) (Zilhão et al., 1996); RNase III indirectly regulates by controlling the level of PNPase (Nicholson, 2014); and RNase E is directly involved in the degradation of the *rnb* mRNA (Arraiano et al., 2010). In addition, there is the 3' exoribonuclease RNase BN (also called RNase Z) (Dutta & Deutscher, 2009). RNase BN is a ribonuclease that has both functions of endo- and 3' exonucleases on model RNAs. After the mRNA is cleaved by the above exoribonucleases to short oligonucleotides of 2~4 ntd in length, the oligoribonuclease activity encoded by the *orn* gene continues the degradation of these fragments into mononucleotides (Ghosh & Deutscher, 1999; Zhang et al., 1998).

1.2.1. C) Other components related to mRNA degradation

Helicases

mRNA is a single-stranded molecular structure, but secondary structures with double-stranded regions are present (Ding et al., 2014; Rouskin et al., 2014; Wan et al., 2014). These structures can modify mRNA degradation by hindering access and/or activities of RNases. Helicases are a family of ATPase-dependent proteins that remodel RNA and RNA-protein complexes by unfolding the mRNA molecule to facilitate ribonuclease access (Jarmoskaite & Russell, 2014). Four DEAD-box RNA helicase paralogs (RhlB, RhlE, SrmB, and CsdA) identified in *E. coli* are relevant to the mRNA degradation pathway (Bhaskaran & Russell, 2007; Mohanty & Kushner, 2018). They have two identical core domains and flanked by variable regions (Worrall et al., 2008). There is evidence that the unwinding of folded mRNA by RhlB promotes cleavages by RNase E and PNPase (Bandyra et al., 2018; Khemici et al., 2005). RhlE, SrmB, and CsdA can also bind to RNase E but at different binding sites than RhlB (Charollais, 2004; Proux et al., 2011; Trubetskoy et al., 2009).

Chaperone Hfq and ncRNA

Hfq is an important chaperone involved in the regulation of RNA degradation. This protein can associate with regulatory small RNAs, facilitate their interaction with the target mRNA, and then facilitate the degradation of the target mRNA by RNase E. As an example, Hfq interacts with the small RNA r*yhB* to downregulate a set of iron-storage and iron-using protein expressions (***Figure 9***). The Hfq-sRNA association stimulates the base pairing with the target mRNA (Bandyra et al., 2012; Tsai et al.,

2012; Valentin-Hansen et al., 2004). Note that the binding of Hfq to RNase E does not depend on ncRNA, but the ncRNA triggers the binding of the complex to the target mRNA (T. Morita et al., 2006).

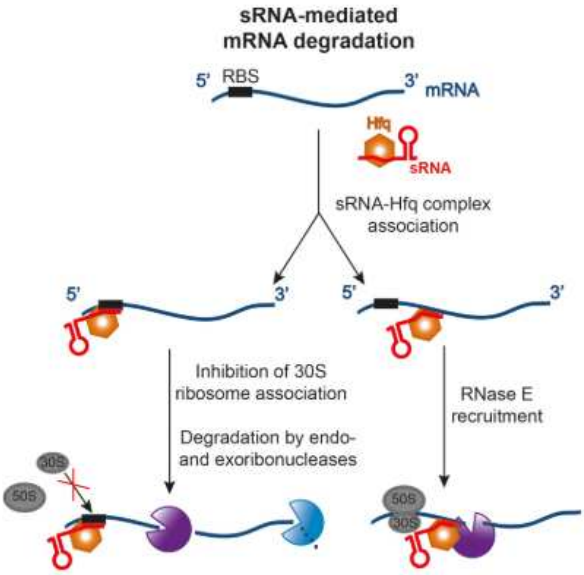


Figure 9. mRNA degradation by the RNase E-Hfq-RyhB complex (Bandyra & Luisi, 2018). First, sRNA (Ryhb) combined with Hfq to form a complex. Hfq protects sRNA from RNase E degradation prior to target pairing, and as a cofactor recruit functional proteins for target regulation (top branch). The sRNA-Hfq complex recognizes and binds a complementary sequence in the translation initiation region of mRNA, which prevents ribosome access to the mRNA (left branch). Thus, the naked mRNA is degraded by endo and exoribonucleases. The sRNA-Hfq complex can also bind with the coding region of mRNA, recruiting RNase E for degradation (right branch).

Poly(A) polymerase

For a long time, it was widely accepted that polyadenylation was merely related to eukaryotic mRNA. However, poly(A) tails have been identified on some bacterial RNAs, and experimental results have demonstrated that polyadenylation plays a role in the stability of *E. coli* mRNA (Cohen, 1995; Kushner, 2002; Xu & Cohen, 1995). Compared to polyadenylation in eukaryotes, the mechanism of polyadenylation in *E. coli* does not require the presence of a specific sequence in the mRNA (Sarkar, 1997). The activity of poly(A) polymerase (PAP I, encoded by *pcnB*) is to add A-tails at the 3'end of transcripts (Maes et al., 2016). Transcriptomic analyses of the wild-type strain and *pcnB* deletion mutant show that more than 90% of *E. coli* transcripts have undergone a certain degree of adenylation by PAP I during the exponential growth leading to an increase in mRNA stability in the *pcnB* mutant (Mohanty & Kushner, 2006).

The addition of A-tails by PAP I at the 3' end of the transcripts makes them better substrates for degradation by the 3' exoribonucleases (***Figure 10***). Transcripts with Rho-independent termination have a stem-loop structure at their 3' end, which strongly inhibits the degradation ability of 3' exoribonucleases. PAP I, by adding the poly(A) tail to the 3' end, provides a binding handle, which

ensures the degradation efficiency by 3' exoribonucleases (Maes et al., 2016; Mohanty & Kushner, 1999; O'Hara et al., 1995; Yehudai-Resheff & Schuster, 2000). However, this pathway is much weaker and slower than the above described two RNase E access pathways.

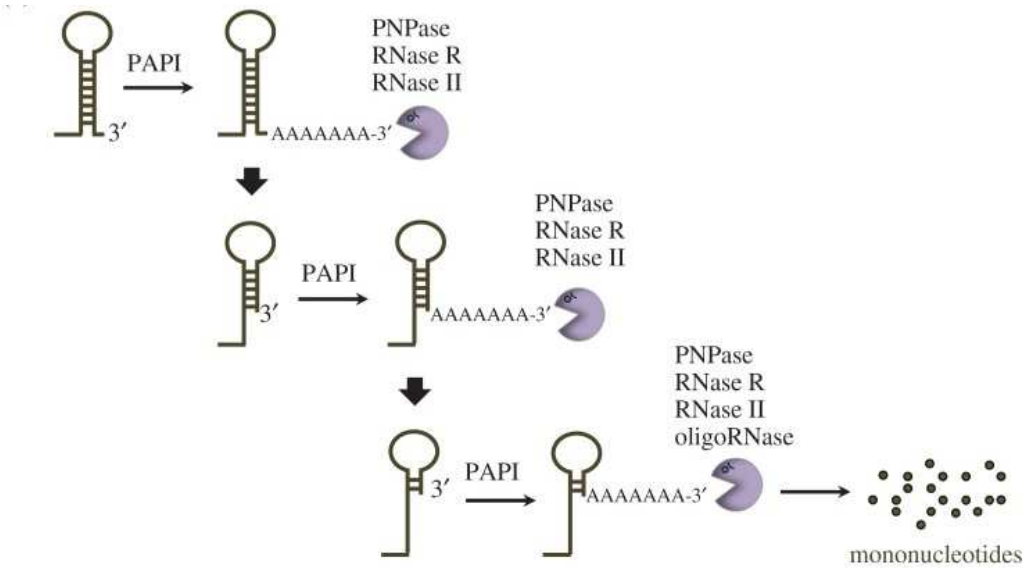


Figure 10. The PAP I-dependent degradation of structured RNAs in *E. coli* (Hajnsdorf & Kaberdin, 2018). Stable stem–loop structures at the 3'end of transcripts are resistant to the action of major exoribonucleases, but their degradation can be facilitated by polyadenylation. Repeated cycles of polyadenylation and subsequent 3' –5' degradation of the polyadenylated species by exoribonucleases and the oligoribonuclease (oligoRNase) yield mononucleotides.

mRNA interferase

mRNA interferases (part of the Toxin-Antitoxin systems) are endoribonucleases which can specifically cleave mRNAs. Several mRNA interferases have been found in *E. coli* such as MazF, RelE, MqsR, YoeB, HicA… (Mohanty & Kushner, 2018; Yamaguchi et al., 2009; Y. Zhang et al., 2005). Their specific cleavage sites are described in ***Table 1***. As an example, the mechanism of MazF action in the Toxin-Antitoxin system (MazE-MazF) is described in ***Figure 11***.

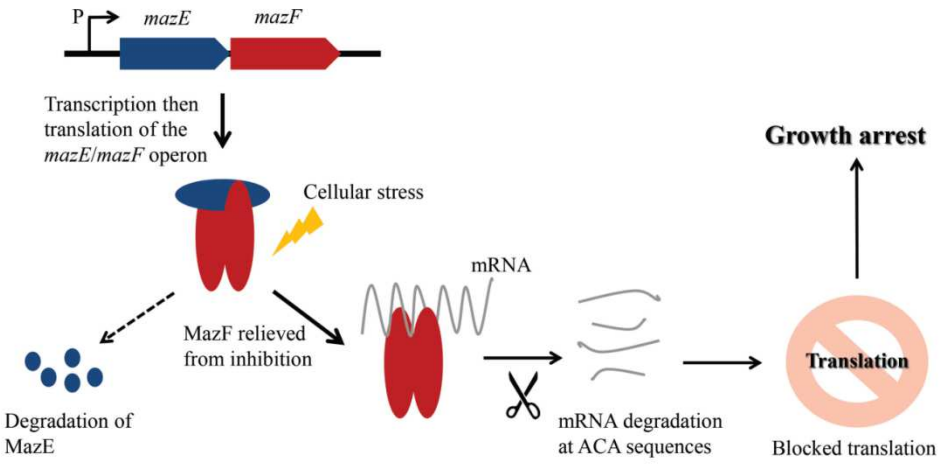


Figure 11. Mode of action of the MazF mRNA interferase ((Al-Hinai et al., 2012). In the MazE-MazF system, the toxin MazF and the antitoxin MazE are encoded by the same operon. The operon is either negatively autoregulated by MazE or a MazE-MazF complex. Generally, MazF and MazE are co-transcribed in an operon and neutralize toxicity through protein-protein interaction, but MazE is far less stable than MazF. Therefore, any effect that disrupts the toxin-antitoxin equilibrium will release the MazF in the cell. MazF then cleaves the target mRNA at ACA sequences, leading to growth arrest, followed by cell death.

1.2.2 mRNA degradation is a complex cellular process

Overall, the previous sections showed that the mRNA degradation is a complex process in *E. coli* that implies many partners. To summarize, ***Figure 12*** describes the generally accepted pathways (Arraiano et al., 2010; Hui et al., 2014; Mohanty & Kushner, 2018). Under most circumstances, mRNA degradation is initiated from internal cleavage by RNase E, but this is not absolute (Arraiano et al., 2010; Baker & Mackie, 2003). There are also internal cleavages determined by RNase P, RNase Z/ BN, and RNase III, of which RNase III is responsible for internal cleavage of the mRNA structured part (stem-loop) to initiate degradation (Mohanty & Kushner, 2018). After initiation of degradation, ribonucleases, degradosome, and mRNA interferases further cleave the fragments or undergo other auxiliary functions (such as polyadenylation, unwinding) before cutting. The full-length mRNA is cleaved into oligoribonucleotides after these series of actions. Eventually, it is completely degraded into mononucleotides by the oligoribonuclease (Ghosh & Deutscher, 1999; Zhang et al., 1998).

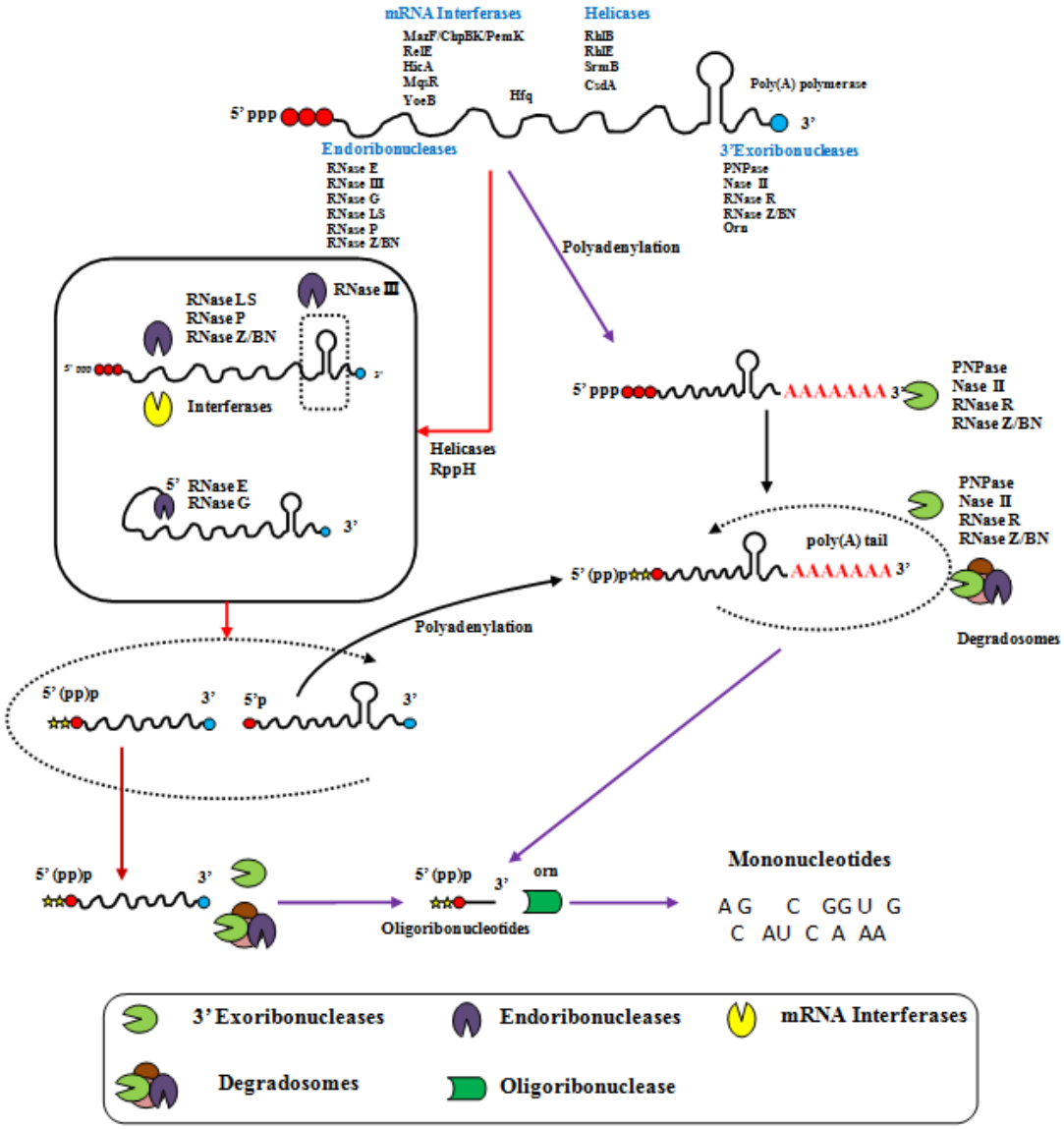


Figure 12. Schematic representation of mRNA degradation pathways in *E. coli.* The overall mRNA degradation mechanism is depicted. The enzymes involved and the sites on which they act on mRNA are annotated. Under most circumstances, mRNA degradation is initiated from internal cleavage and results in two consequences: i) the fragments lacking the protection of the 3' end stem-loop are generated and are rapidly degraded by 3' exoribonucleases; ii) the fragments exposing the monophosphate 5' end are generated and are further cleaved by endoribonucleases. The details have been described in previous sections.

1.3 Process of translation in *E. coli*

1.3.1 Ribosome

Translation is the last step of gene expression. Ribosomes translate the genetic information contained in mRNA into the amino acid sequence of proteins. Ribosome biogenesis *per se* and recycling is a complex series of processes involving mRNA, tRNAs and a number of translation factors. Structurally, *E. coli* ribosome sediment as 70S particles formed by two subunits, 30S and 50S. The 30S subunit consists of 16S rRNA (1542 ntd) and 21 proteins (Ghosh & Joseph, 2005). The 50S subunit consists of 23S rRNA (2904 ntd), 5S rRNA (120 ntd) and 36 proteins (Graf et al., 2017). The ribosome is a dynamic molecular machine with constantly changing conformation, and its proteins undergo diverse rearrangements during the translation steps (Bock et al., 2018; Schmeing & Ramakrishnan, 2009).

1.3.2 Basic processes of translation

Like transcription, translation also can be roughly divided into three stages: translation initiation, elongation, and termination.

Translation initiation

Translation initiation is the rate-limiting and highly regulated step of translation. It ensures the first codon-anticodon interaction into the peptidyl (P) site of the small ribosomal subunit (Gao et al., 2003) (***Figure 13***).

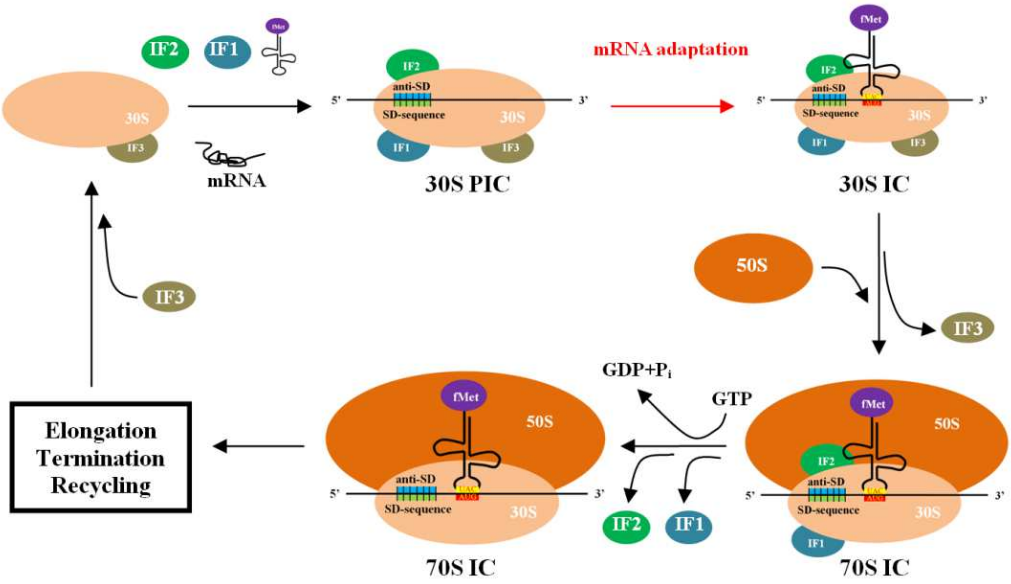


Figure 13. Illustration of the translation initiation (adapted from Laursen et al., 2005; Simonetti et al., 2009). See details as below.

The 30S ribosomal subunit, three initiation factors (IF1-3), and the aminoacylated and formylated initiator tRNA (fMet-tRNA^{fMet}) are assembled with the mRNA in a standby position to form the 30S pre-initiation complex (30S PIC). The anti-SD sequence of 16S rRNA and the SD sequence of the mRNA are base-paired to fix the 30S ribosomal subunit on the mRNA. IF1 and IF3 play a role in stabilizing the 30S ribosomal subunit. GTP-dependent IF2 brings fMet-tRNA^{fMet} to the 30S PIC in a codon-independent manner. With the rate-limiting conformation changes of 30S PIC, the start codon and fMet-tRNA^{fMet} adjust the mRNA to the appropriate site through codon-anticodon interaction to form a more stable mature 30S initiation complex (30S IC). IF1 and IF3 are released from the complex. IF2 stimulates the 50S ribosomal subunit to associate with the 30S ribosomal subunit and subsequently also released from the complex, while the GTP bound to IF2 is hydrolyzed. The final formation of the 70S initiation complex (70S IC) is assembled, and the ribosome is ready to enter the elongation phase of translation (Laursen et al., 2005; Simonetti et al., 2009; Vimberg et al., 2007).

Translation elongation

The assembly of the large and the small subunits creates three sites in the ribosomes involved in peptidyl chain synthesis. The aminoacyl site A is the receptor for the tRNA that loads specific amino acids, the peptidyl site P loads amino acids onto the growing peptide chain, and the exit site E releases deacylated tRNA (Márquez et al., 2002) (***Figure 14***).

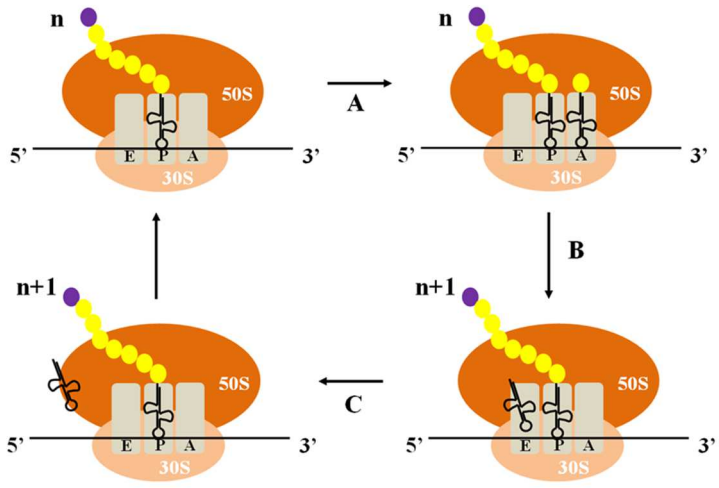


Figure 14. Illustration of three tRNA binding sites on the ribosome during translation elongation (adapted from Burkhardt et al., 1998). The chart depicted the relative positions of a tRNA and A/P/E-sites in the 70S ribosome. A) tRNA carrying the amino acid enters the A site. B) The ribosome moves one ntd forward on the mRNA, and the peptide chain extends one amino acid. The previous tRNA exits the P site and enters the E site, and the newly entered tRNA moves from the A site to the P site. C) The E site tRNA dissociates from the elongation complex, and the vacant A site awaits the arrival of the newly tRNA carrying amino acid for the next elongation cycle.

Translation elongation is facilitated by three elongation factors: EF-Tu functions with GTP and mediates the entry of the aminoacyl tRNA into a free site of the ribosome (Burnett et al., 2014); EF-G

catalyzes the translocation of the tRNA and mRNA from the ribosome at the end of each round of polypeptide elongation (Liu et al., 2014); and EF-Ts catalyzes the release of GDP from EF-Tu, allowing EF-Tu to bind to a new GTP molecule (Wieden et al., 2002). The translation elongation involves repetitive cycles of decoding, peptide bond formation, and translocation (Ramakrishnan, 2002). Decoding ensures that the correct aa-tRNA is selected at the A-site. Peptide bond formation means that the peptidyl-tRNA at the P site and the aa-tRNA at the A site react to form a peptide bond. As the translation process progresses, 70S moves on the mRNA in the 5'-3' direction. Simultaneous, the nascent polypeptide chain is switched from the A site to the P site in 70S, and then the deacetylated tRNA is transferred from the P site to the E site and is released from the ribosome. This elongation process repeats the cycle until the stop codon is encountered and translation stops.

Translation termination

Translation termination occurs when a stop codon (UAA, UAG, or UGA) on the mRNA enters the A-site (Rodnina, 2012). Two release factors RF1 and RF2 recognize the stop codon: RF1 terminates with stop codons UAA and UAG, while RF2 terminates with UAA and UGA (Scolnick et al., 1968). In addition, the release factor RF3 accelerates the dissociation of RF1 and RF2 from the ribosome after RF1 and RF2 trigger the hydrolysis of the peptidyl-tRNA at the P site (Freistroffer et al., 1997). This also allows the ribosome to recycle into subunits for the next round of protein synthesis (Graf et al., 2018).

2 5’UTR and gene expression regulation

2.1 Presentation of the 5’UTR region

The 5’ UTR (also known as the leader sequence or leader RNA) (***Figure 15***) is the transcribed but untranslated sequence located in the initial part of the mRNA, starting from the transcription start site and ending at the nucleotide preceding the start codon. With a length ranging from 0 to 700 ntd, the most frequent length is between 25 and 35 bp in *E. coli* (Kim et al., 2012). It should be noted that some mRNAs have no 5’UTR at all or have only a very short 5’UTR and they are therefore called leaderless mRNAs (Beck & Moll, 2018). We can also note that the 5’ UTR is included in the translation initiation region (TIR) which additionally contains the proximal 5’ coding sequence of the mRNA (***Figure 15***).

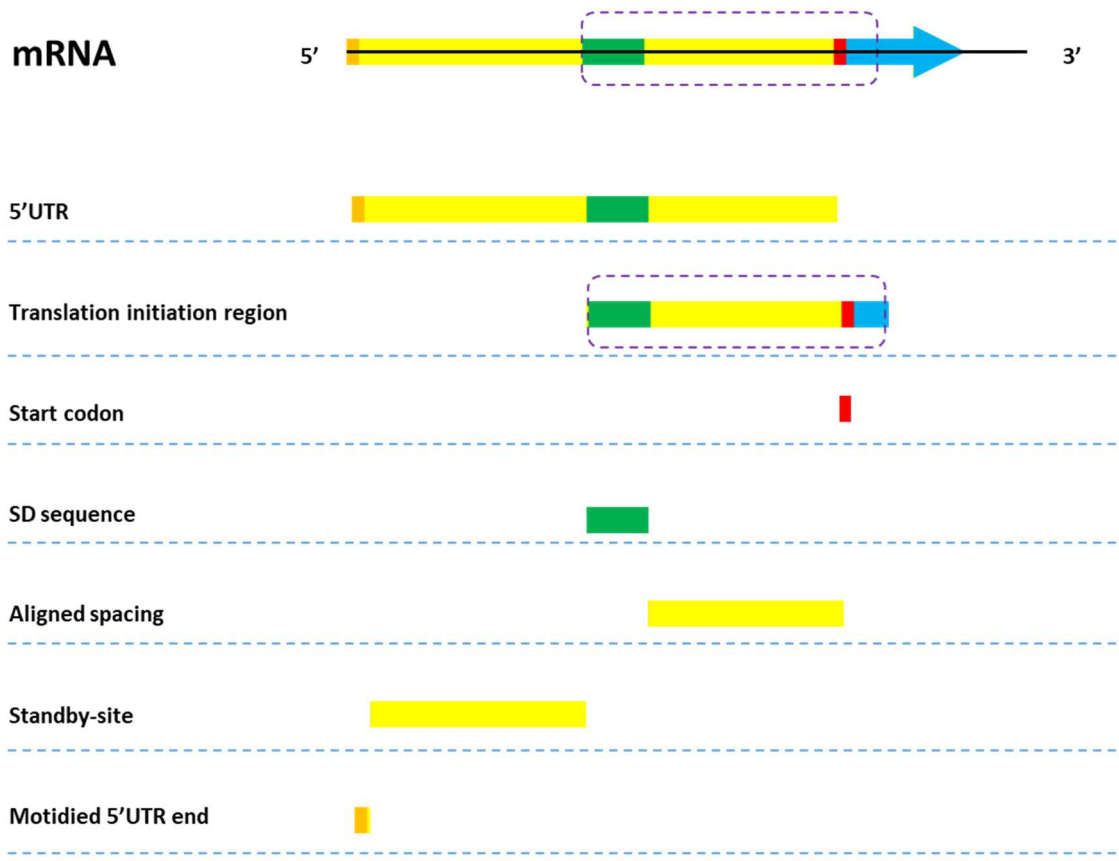


Figure 15. Key 5' UTR elements of *E. coli* mRNA. Regulatory elements in the 5'UTR are shown independently, marked with different colors. See the next section for a detailed description of each 5'UTR element.

Initially, 5'UTR was associated with translation initiation through the Shine-Dalgarno (SD) sequence (this SD sequence - with or not some adjacent nucleotides - is also called RBS for ribosome binding site). As progress has been made in deciphering the regulatory mechanism related to the 5'UTR, more and more *cis*-acting elements and *trans*-acting factors related to the 5'UTR have been identified as being involved in the regulation of translation but also of transcription and mRNA degradation. In the following sections, we will review how some 5'UTR-related elements regulate translation, transcription and/or mRNA degradation.

2.2 Regulatory elements in the 5'UTR

2.2.1 Shine-Dalgarno (SD) sequence

The 5'UTR contains part or the entire UAAGGAGGU sequence known as Shine-Dalgarno sequence, which is complementary to the region (anti-SD or aSD sequence) at the 3'end of 16S rRNA (Shine & Dalgarno, 1974). Recruitment of ribosomes through complementary pairing of the SD sequence with the 3'end of the 16S rRNA initiates translation. Most of *E. coli* SD sequences are between 4 to 8 ntd in length (the complementarity of mRNA / rRNA is not interrupted by unpaired bases) (Shultzaberger et al., 2001). The results of free energy calculations for all upstream regions of 1,159 *E. coli* genes that possibly form a duplex structure with the 16S rRNA indicate that the average length of SD: aSD is 6.3 ntd (Schurr et al., 1993).

Unlike eukaryotes, bacterial mRNAs can be polycistronic. Thus, making the correct recognition of any potential translation initiation site along the mRNA by the ribosome is particularly critical for bacterial gene expression regulation. As mentioned above, the SD sequence facilitates the ribosome binding with mRNA to initiate the translation. Particularly, stronger SD sequences improve translation efficiency when the start codon is not an AUG, or when the start site is masked by secondary structure (Olsthoorn et al., 1995; Weyens et al., 1988). Numerous studies have demonstrated that modifications of the SD and its adjacent sequences are responsible for over 1000-fold changes in protein expression (Ringquist et al., 1992). Studies have shown that experimentally lengthening the SD sequence does not improve the translation efficiency, or even decreases it. In *E. coli*, a six-nucleotide SD (AGGAGG) has a higher initiation efficiency than shorter or longer sequences at 37°C (Vimberg et al., 2007). The lower translation initiation efficiency of the relatively short SD sequence is assumed to be due to the weaker SD:aSD binding affinity. In contrast, the relatively long SD sequence likely results in redundant SD:aSD affinity, which causes ribosome stalling at the translation initiation phase. Furthermore, the optimal SD sequences for high translation initiation efficiency are temperature dependent rather than growth rate dependent. The shortest SD sequence (GGAGG) is preferred at lower temperature (20°C).

2.2.2 Aligned spacing

The aligned spacing defines the sequence between SD and the start codon and it is also a regulatory parameter of gene expression through the control of translation initiation efficiency. The optimal length of the aligned spacing in *E. coli* is 5-9 ntd (Hartz et al., 1991; E. S. Komarova et al., 2020; Osterman et al., 2013). Toeprinting experiments suggest that each specific SD sequence has an optimal aligned spacing adapted to itself for optimal mRNA-tRNA-ribosome interaction. Indeed, lengthening the aligned spacing inhibits ribosomal translocation (Wakabayashi et al., 2020). Conversely, an aligned

spacing of less than 4 ntd prevents the formation of a stable pre-translocation complex (Devaraj & Fredrick, 2010). Furthermore, a study using aligned spacing libraries reveals that many efficiently translated mRNAs have A-rich sequences with U residues in the spacer region, e.g, with AAAU, AAUA, AUAA, AUAU sequence variants (E. S. Komarova et al., 2020), despite the fact that AU-rich enhancers are generally considered to be located upstream of SD sequences. One explanation for the efficiently translated mRNAs which adenosine preference of the aligned spacing is that AU-rich sequences enhance the interaction with the ribosome. Although the aligned spacing does not participate in base pairing with the 16S rRNA, it is presumed that it can form a specific structure with the ribosome (Hussain et al., 2016).

2.2.3 Initiation transcription sequence (ITS) and promoter-like sequence

The ITS corresponds to the 20 first nucleotides of the 5'UTR. It can affect the transcription initiation process by influencing the efficiency with which RNAP escapes from the promoter and continues elongation (Heyduk & Heyduk, 2018). Transcription initiation is a multistep process that begins when RNAP recognizes and binds to DNA elements within a promoter sequence and ends when RNAP escapes from the promoter and continues through elongation. This process involves multiple cycles of failed start of synthesis and release of short (2-15 ntd) transcripts called abortive initiation until promoter release by RNAP (Hsu, 2002; Wade & Struhl, 2008). The transcription process moves from the initiation to the extension phase only if a threshold length of 9 to 11 ntd RNA is synthesized and the promoter escape is successfully completed (Mazumder & Kapanidis, 2019; Saecker et al., 2011). The composition (especially positions +1 and +2) and structuration of the first 10 nucleotides of ITS influences promoter escape the most. In addition, sequences similar to the −10 promoter element (TATAAT) in 5'UTRs can induce a σ70-dependent transcription pause after the promoter escape (Brodolin et al., 2004; Hatoum & Roberts, 2008; Nickels et al., 2004) and thus reduced transcription.

2.2.4 Standby-site and S1 ribosomal protein

The 30S ribosomal complex will first land non-specifically on the mRNA backbone prior to the SD:aSD interaction (Espah Borujeni et al., 2014). The initial landing platform upstream of the SD sequence in 5'UTR is referred to as the standby-site (de Smit & van Duin, 2003). The standby-site is usually structured, so the 30S ribosomal complex unfolding the mRNA is critical. Studies have shown that a variety of ribosomal proteins, and in particular the S1 ribosomal protein, exert RNA helicase activity when the 30S ribosome docks with the mRNA (Qu et al., 2012; Takyar et al., 2005; Yusupova et al., 2006). Cryo-EM analysis indicates that S1 might bind 11 nucleotides upstream of the SD sequence

(Sengupta et al., 2001). S1 provides RNA-melting activity to the exit site of the 30S decoding channel and confers some plasticity on the ribosome to initiate mRNA translation, which is essential for docking and unfolding standby-site of mRNA (Duval et al., 2013).

In terms of gene expression regulation, S1 is essential for translation of many mRNAs (Lauber et al., 2012; Sørensen et al., 1998; Studer & Joseph, 2006; Subramanian, 1983). The binding of S1 to the standby-site is particularly vital for translation of mRNAs lacking or with weak SD sequences (Komarova et al., 2005; Tzareva et al., 1994). In contrast, S1 is nonessential for mRNAs with optimal SD sequence and weakly structured RBS (Duval et al., 2013). Studies have shown that AU-rich sequences in the standby-site can increase translation efficiency by an order of magnitude in *E. coli* (Osterman et al., 2013). S1 also contributes in the regulation of mRNA stability. The insertion of an AU-rich standby-site can increase mRNA stability (Komarova et al., 2005), although AU-rich sequences are also potential cleavage targets for RNase E (Lin-Chao et al., 1994; McDowall et al., 1994). One explanation is that S1 wins the competition with RNase E for binding to the AU-rich site. Another possibility is that the facilitation of translation initiation by AU-rich standby-site results in greater ribosomal occupancy of the mRNA, thereby protecting the transcript from RNases attack.

2.2.5 Modified 5'UTR end

In the standard *de novo* transcription initiation, RNAP uses a nucleoside triphosphate (NTP) as a substrate, resulting in a stable triphosphate structure at the 5' end of the RNA (Barvík et al., 2017). Non-canonical substrates can compete with NTPs for transcription initiation by RNAP, such as nanoRNAs and non-canonical initiating nucleotides leading to modified 5'UTR end.

NanoRNAs

NanoRNAs (2-4 ntd long oligoribonucleotides) are mainly derived from the degradation of mRNA by the oligoribonuclease Orn (Liao et al., 2018). When a nanoRNA acts as a primer for transcription initiation, it introduces a 5'-terminal hydroxyl or 5'-terminal monophosphate into the nascent RNA. It has been shown that nanoRNA-mediated transcription initiation is especially prevalent in the stationary phase of *E. coli* growth (Vvedenskaya et al., 2012).

With respect to gene expression regulation, how nanoRNA-mediated transcription initiation affects gene expression is still unknown. One hypothesis is that this mechanism alters the phosphorylation status of the mRNA 5'end, thus regulating transcript stability (Nickels, 2012). Another hypothesis is that the introduction of nanoRNAs leads to a shift in the position of the transcription initiation site, which affects promoter function and regulates transcription (Nickels & Dove, 2011).

Non-canonical initiating nucleotides

In vitro experiments have demonstrated that nicotinamide adenine dinucleotide (NAD^+), NADH and dephosphorylated coenzyme A (dp-CoA), flavin adenine dinucleotide (FAD), UDP-glucose, and UDP-N-acetylglucosamine can be used as non-canonical initiating nucleotides and thus capping the RNA 5'end (Cahová et al., 2015, 2015; Julius & Yuzenkova, 2019; Luciano & Belasco, 2015). Besides, multiple unmethylated, monomethylated and dimethylated forms of dinucleoside polyphosphates (Np_nNs)-RNA caps were identified in RNAs isolated from *E. coli* (Benoni et al., 2020). The use of non-canonical initiating nucleotides was proposed to depend on sigma factors and promoters (Skalenko et al., 2021).

In terms of gene expression regulation, capping of the 5'end of the mRNA contributes to the regulation of its stability and transcription efficiency. For instance, it was observed that NAD-modified RNA 5'-capping strongly stabilized RNAs against 5'-processing by RppH and against endonucleolytic cleavage by RNase E (Cahová et al., 2015). *In vitro* studies have shown that *E. coli* RNAP appears to prefer dinucleoside tetraphosphates (Np_4Ns) to NTPs for initiating transcription, i.e. it uses Np4A up to nine times more efficiently than ATP (Luciano et al., 2019; Luciano & Belasco, 2020). The physiological roles of using non-canonical initiating nucleotides for transcription initiation in *E. coli* still need to be further explored.

2.2.6 Structured 5'UTR: RNA thermometer (RNAT) and Riboswitch

Due to the single-stranded structure of the mRNA, the 5'UTR potentially tends to be highly structured. The RNA can fold back on itself to form hairpins composed of spirals covered by loops, which known as secondary structure (Garst et al., 2011). The structured motifs of the SD, aligned spacing and standby-site determine translation initiation efficiency and mRNA quantity through their impact on the accessibility of ribosomes and RNases to mRNA. However, the secondary structure of RNA has an intrinsic propensity to fold dynamically (Chiaruttini & Guillier, 2020). Dynamic folding of 5'UTR can act as *cis*-acting elements in the regulation of gene expression in response to environmental stimuli. RNAT and riboswitch are two *cis*-acting elements related to secondary structures located in the 5'UTR (Chowdhury et al., 2006). The main difference between the two is that RNAT senses physical signals whereas riboswitch senses chemical signals and relies on interaction with their ligands.

RNAT

In terms of gene expression regulation, the secondary structures of RNAT rapidly respond to temperature changes by folding or exposing RBS, thereby, influencing the accessibility of ribosomes to mRNA and determining gene expression or repression. One of the well-studied RNATs in *E. coli* is the

5'UTR upstream coding sequence of the heat-shock alternative sigma factor RpoH (Morita et al., 1999). As illustrated in the following (***Figure 16***), the disruption of secondary structure by high temperature exposes the ribosome binding site and, thereby allows the synthesis of heat shock proteins.

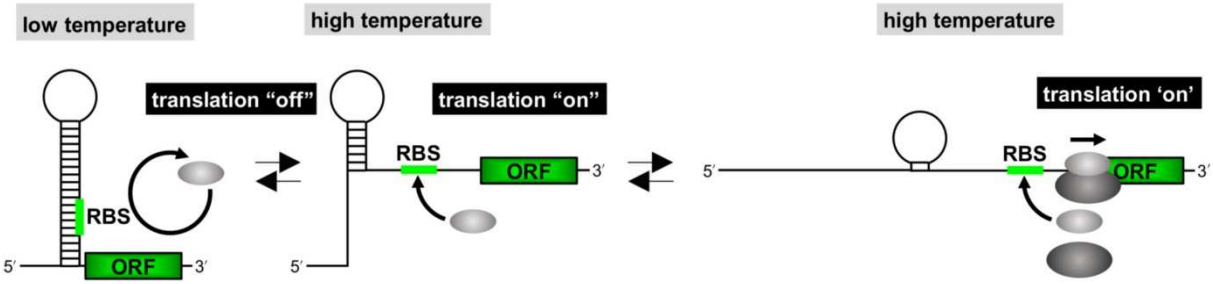


Figure 16. Mechanism of translation initiation control by heat-sensitive RNAT (Chiaruttini & Guillier, 2020). This scheme described the dynamic equilibrium between translation "off" and "on". (*Left*) At low temperature, the RBS is occluded by the folded secondary structure and the 30S ribosome (in light gray) cannot bind the mRNA, translation is "off". (*Middle and right*) While the temperature rises, the fold gradually melts open and the RBS is exposed. The accessibility of the 30S ribosome increases and allows it to bind with the 50S ribosome to form the complete translation initiation complex, then translation is "on".

Conversely, an example of RNAT sensitive to low temperatures in *E. coli* is the expression of the cold shock protein CspA with an increase in gene expression under cold shock (***Figure 17***).

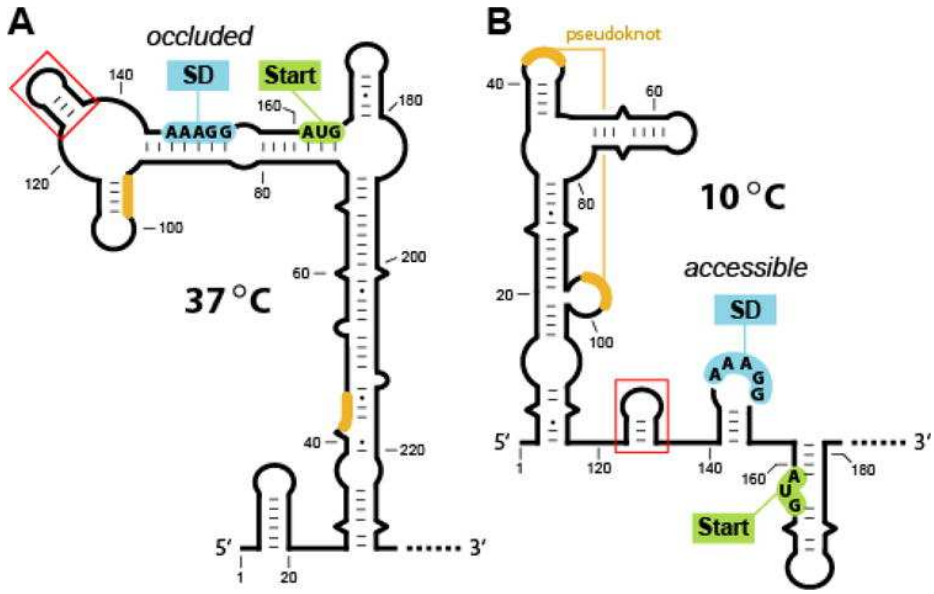


Figure 17. Mechanism of translation initiation control by cold-sensitive RNAT (Breaker, 2010; Giuliodori et al., 2010). (*Left*) At an optimal growth temperature around 37°C, nascent transcripts of *cspA* form a 160 ntd stem-loop spanning the entire 5'UTR and the front 60 ntd coding sequence. This structurally isolates the SD sequence from the start codon and obscures the RBS. (*Right*) While at 10°C, the mRNA undergoes structural rearrangement, with the SD sequence and start codon AUG sequestered into separate weakly hairpins, respectively. Therefore, the RBS is accessible for the ribosome to initiate translation.

Riboswitch

Riboswitch consists of two domains: the aptamer domain and the expression platform (Tucker & Breaker, 2005). The aptamer domain acts as the receptor that specifically binds the ligand. Ligand binding enables riboswitch to switch between two different secondary structures in response to regulatory signals. To date, a wide variety of native riboswitches have been identified in *E. coli* in response to different ligands: FMN (Pedrolli et al., 2015), guanidinium (Sherlock et al., 2017), adenosylcobalamin (Gallo et al., 2008), molybdopterin (Regulski et al., 2008), thiamine diphosphate (Serganov et al., 2006), Mn^{2+} (Dambach et al., 2015), spermidine (Yoshida et al., 1999), etc.

Although the sequence, structure and ligand of these riboswitches differ considerably, they are all involved in gene expression regulation mainly at the two levels of transcription termination and translation initiation (Sharma et al., 2022). For transcription regulation, riboswitch secondary structure may create an intrinsic transcription terminator, resulting in intrinsic transcription termination in response to environmental "signal". For translation regulation, a model below illustrates the mechanism of riboswitch-mediated translation initiation control (***Figure 18***).

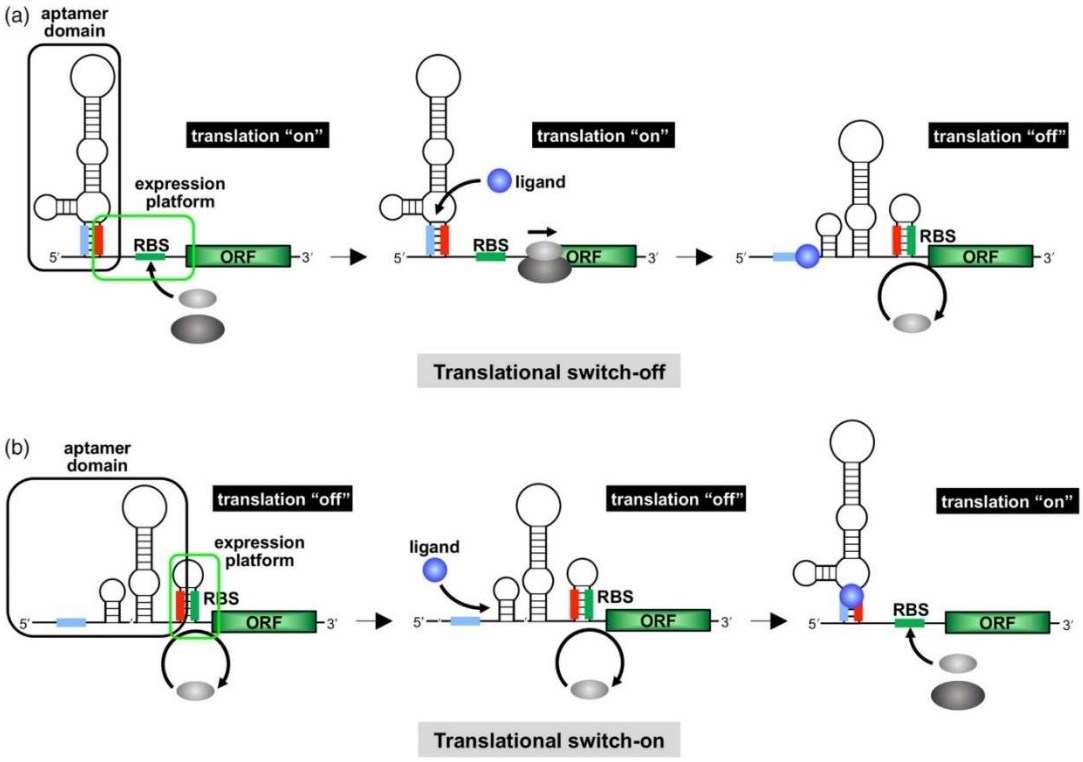


Figure 18. **Mechanism of riboswitch-mediated control of translation initiation** (Chiaruttini & Guillier, 2020)**.** (a) Translation switch-off. In the absence of ligand, the anti-RBS sequence (red) is obscured in the secondary structure of the aptamer domain, and the RBS in the expression platform is accessible for the 30S ribosome. With the ligand (blue sphere) binding to the aptamer domain, the riboswitch stabilizes into another conformation. The anti-RBS sequence, originally located in the aptamer domain, is released and pairs with the RBS in the expression platform resulting in the RBS being inaccessible to the 30S ribosome. (b) Translation switch-off. In the absence of ligand, the RBS is trapped in the stem-loop by the anti-RBS sequence (red), and the mRNA is inaccessible to the 30S ribosome. With the ligand binding to the aptamer domain, the conformational change releases the RBS resulting in the RBS being accessible to the 30S ribosome.

2.2.7 5'UTR binding factors: small non-coding RNAs (sRNAs) and RNA-binding proteins

5'UTR can be the target of *trans*-acting factors such as sRNAs and RNA-binding proteins involved in the regulation of gene expression.

sRNAs and sRNA chaperones

Regulatory sRNAs that act as *trans*-acting factor are either transcribed independently or processed from the non-target mRNAs (Villa et al., 2018). Hundreds of sRNA-encoding genes have been annotated in *E. coli* (Barquist & Vogel, 2015; Hör et al., 2020). To date, most of these characterized sRNAs are independently transcribed, but some are also processed from the 5'UTR and 3'UTR of the mRNAs (Adams & Storz, 2020). It has been suggested that RNA fragments produced after RNase cleavages or premature transcription termination can also function as regulatory sRNAs (Adams et al., 2021).

The sRNAs can pair with the "seed region" in the 5'UTR with complete or incomplete complementarity, and perform a variety of gene expression regulations at the levels of translation, mRNA degradation and transcription. Many, but not all, sRNAs require the assistance of chaperone proteins to function properly. The three major broad-spectrum sRNA chaperones in *E. coli* are Hfq, CsrA and ProQ. Analysis of hundreds of sRNA in *E. coli* showed that about 25% of them rely on Hfq to carry on their functions (Vogel & Luisi, 2011; Zhang et al., 2013). Hfq protects sRNA from RNase degradation prior to target pairing and facilitates the annealing of sRNA-mRNA duplexes (Holmqvist & Vogel, 2013). As regulators of gene expression, sRNAs can bind to the RBS to obstruct ribosome access, or structurally modify the 5'UTR to facilitate ribosome assembly and translation (***Figure 19C and D***) (Lease et al., 1998; Morita et al., 2006). Second, sRNAs can recruit RNase E to promote 5'end-dependent degradation or recruit chaperone proteins to obstacle RNase E scanning and cleavage from the 5' end and thereby extend the life of the mRNA (***Figure 19B***) (Prévost et al., 2011; Richards & Belasco, 2019). Third, sRNA in association (with or without Hfq) can mediate anti-termination transcription by suppressing Rho-dependent premature transcription termination at the 5'UTR (Rabhi et al., 2011; Sedlyarova et al., 2016).

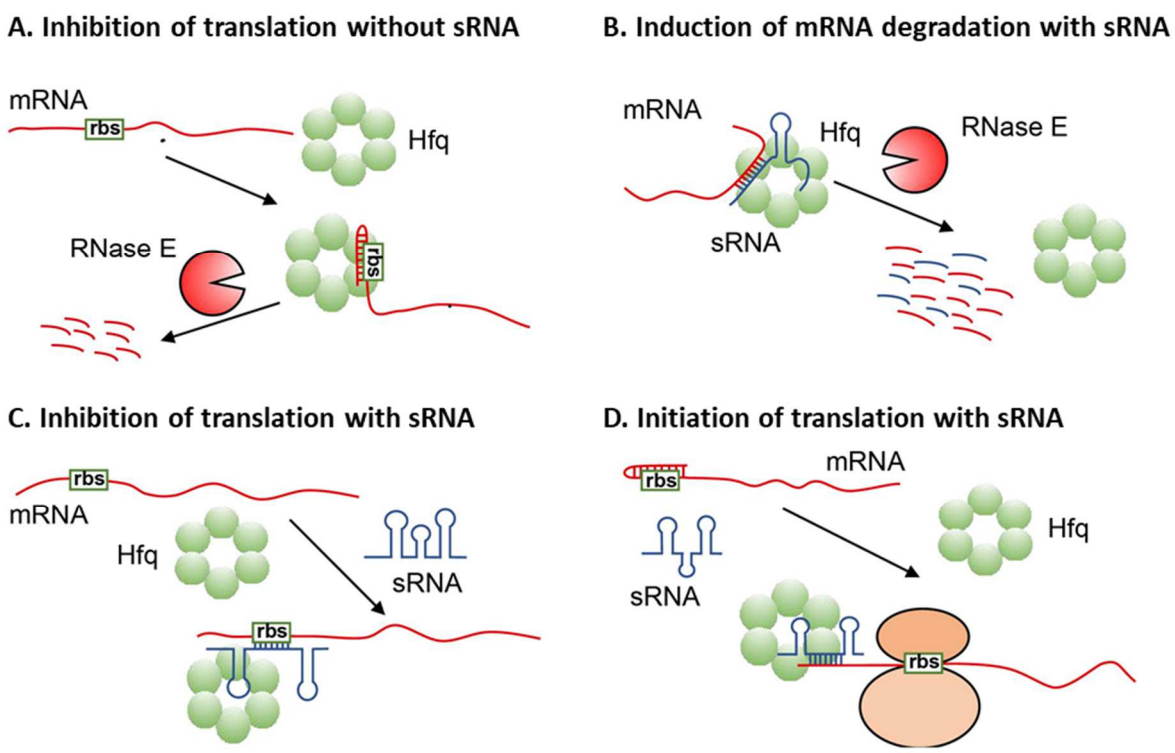


Figure 19. The 5'UTR-related regulatory mechanism of Hfq (Fu et al., 2021).
A. Hfq binds directly to the mRNA 5'UTR and rearranges its second structure to form a stem-loop block ribosome access, thus causing mRNA rapid degradation by RNase E in the 5'end-dependent pathway (J. Chen & Gottesman, 2017; Ellis et al., 2015; Vytvytska et al., 2000).
B. Hfq facilitates the mRNA-sRNA hybrid and recruit RNase E to target mRNAs of sRNAs, resulting in them simultaneously rapid degradation (Ikeda et al., 2011).
C. Hfq facilitates the mRNA binding to the SD sequence of mRNA, thus obstacle the accessibility of ribosome to mRNA and inhibit the downstream translation process (Desnoyers & Massé, 2012).
D. Hfq facilitates annealing of sRNA-mRNA upstream of the SD sequence, leading to stem-loop unfold within the 5'UTR, thereby activating the RBS accessibility for the ribosome to initiate translation (Kawamoto et al., 2006).

Action of 5'UTR binding proteins independently of sRNAs

Binding of Hfq directly to the 5' UTR of mRNA can inhibit translation initiation through structural rearrangements of the ribosome binding site and lead to mRNA cleavage by RNase E (***Figure 19A***) (Salvail et al., 2013). Similarly, binding of CsrA to the 5' UTR mRNA can inhibit translation initiation by competing with the ribosome, activate translation initiation in the case of a riboswitch, and stabilize the mRNA by preventing cleavage by RNase E (***Figure 20***) (Vakulskas et al., 2015).

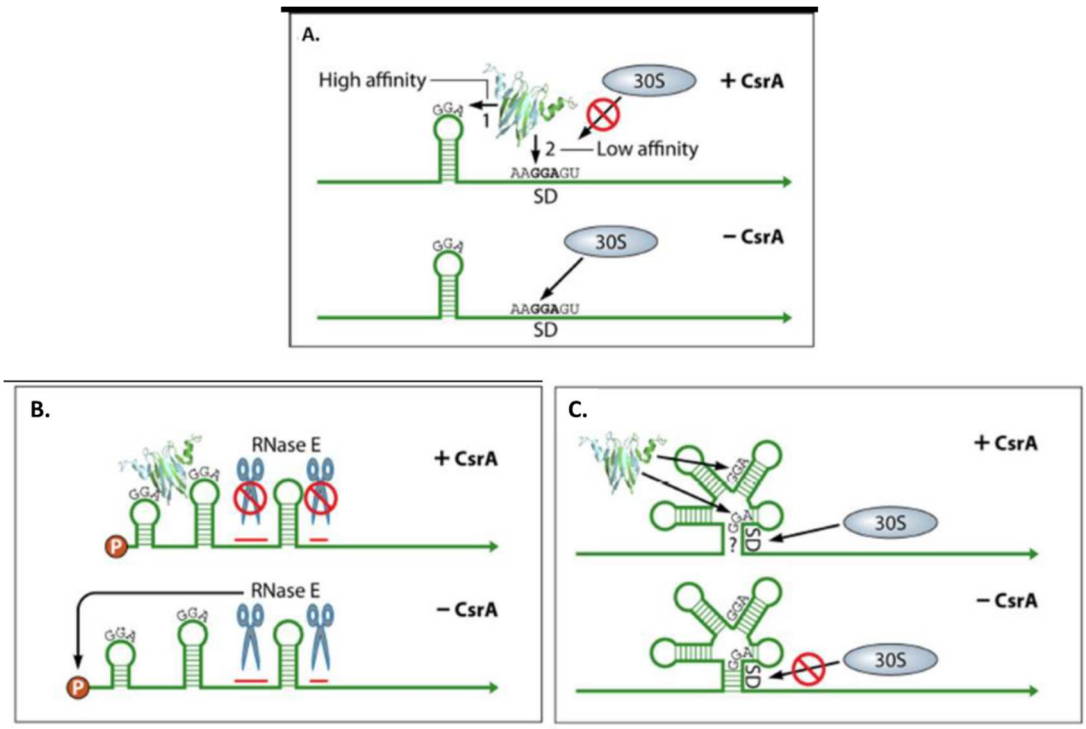


Figure 20. Models for repression and activation by CsrA (Vakulskas et al., 2015).
A. Inhibition of translation by competition with the ribosome (case of *glgC* regulation). In the presence of CsrA (+CsrA), the protein binds to two recognition sites of varying low affinity. Binding results in sequestration of the SD sequence and prevents ribosome attachment. In the absence of CsrA (-CsrA), translation proceeds normally.
B. Mechanism of stabilization of *flhDC*. In the presence of CsrA (+CsrA), the protein binds to two recognition sites preventing recognition of the phosphorylated 5'-end by RNase E. In the absence of CsrA (-CsrA), RNase E degrades the mRNA.
C. Activation of *moaA* translation by riboswitch binding. In the presence of CsrA (+CsrA), the protein binds to the riboswitch resulting in the SD sequence being accessible to ribosomes. In the absence of CsrA (-CsrA), the SD sequence remains sequestered in the riboswitch.

2.3 Summary of the role of 5'UTR elements in gene expression regulation

In this section, we have tried to group the regulatory elements of the 5'UTR according to the targeted process, translation, transcription or mRNA degradation.

2.3.1 5'UTR elements involved in the regulation of translation

In ***Table 3***, we have summarized the elements of the 5'UTR acting on translation regulation, as described in Section 2.2. We can observe that most of the elements discussed in the previous section (except the ITS and the use of non-canonical substrates for transcription) participate in the regulation of translation by controlling the initiation rate.

Table 3: 5'UTR elements involved in translation regulation.

5'UTR element	**Translation**
SD	Translation initiation: affects complementarity with the anti-SD sequence on the 16S rRNA
Aligned spacing	Translation initiation: affects ribosome translocation on the mRNA before finding the correct start codon
Standby-site and S1 ribosomal protein	Translation initiation: facilitates the recognition of anti-SD by the SD sequence
RNAT, Riboswitch, sRNAs and sRNA chaperones	Translation initiation: affect ribosome accessibility by exposing or occluding the RBS
5'UTR binding proteins independent of sRNAs	Translation initiation: competition with the ribosome

Currently, several tools have been developed for predicting translation initiation rates based on the 5'UTR features like RBS Calculator (Salis et al., 2009), RBS Designer (Na & Lee, 2010) and UTR Designer (Seo et al., 2013). These algorithms are established on different statistical thermodynamic models that estimate the translation initiation rate.

2.3.2 5'UTR elements involved in the regulation of transcription

The 5' UTR elements acting on transcription regulation, as described in Section 2.2, are listed in ***Table 4***. Five elements participate in transcription regulation at the initiation and termination levels.

Table 4: 5'UTR elements involved in transcription regulation.

5'UTR element	**Transcription**
ITS and promoter-like sequence	Transcription initiation: affect how RNAP escapes from the promoter or RNAP pausing
nanoRNAs	Transcription initiation: shift of transcription start codon
Non-canonical initiating nucleotides	Transcription initiation: competition with NTP
Riboswitch	Transcription termination: create an intrinsic transcription terminator
sRNAs and sRNA chaperones	Transcription termination: anti-termination by competition with the Rho protein

2.3.3 5'UTR elements involved in the regulation of mRNA degradation

The 5' UTR elements involve in mRNA degradation regulation are listed in ***Table 5***. They contribute to both protection and destabilization of the transcript. It can be noted that some 5'UTR elements could be indirectly involved in mRNA degradation regulation. For instance, secondary structures such as RNAT and riboswitch can promote or block ribosome binding and thereby indirectly affect transcript stability by altering ribosome protection. This indirect effect on mRNA degradation related to ribosome protection could be present for all 5' UTR elements acting on translation regulation.

Table 5: 5'UTR elements involved in mRNA degradation regulation.

5'UTR element	**mRNA degradation**
Standby-site and S1 ribosomal protein	Protection by competition with RNase E or by an increase in ribosome density
NanoRNAs, Non-canonical initiating nucleotides	Protection from 5'end-dependent RNase E attack
sRNAs and sRNA chaperones	Destabilization by RNase E recruitment or protection by sRNA chaperone binding
5'UTR binding proteins independent of sRNAs	Protection by masking RNase E cleavage site

In conclusion, this literature review has clearly demonstrated that the 5'UTR of an mRNA is a region rich in regulatory elements of gene expression. These 5'UTR elements are related to primary sequence, 5'end capping, secondary structure, and accessibility for intermolecular interactions with ribosomes, sRNAs and proteins. 5'UTR regulatory elements influence gene expression in all three processes of translation, transcription or mRNA degradation. We can notice that one element can be a regulator of more than one process (for example, sRNAs regulate the three processes).

3 Objectives of the PhD work

The bibliography review clearly demonstrates that 5'UTRs are important elements in mRNA lifestyle to regulate gene expression. It also shows that most of the current research on 5'UTR-mediated regulations merely focuses on one dimension alone (e.g., modifying a 5'UTR to explore its regulation of translation initiation efficiency). However, it is known that protein expression level can be determined by translation efficiency, but also by mRNA concentration, which itself results from the balance between transcription and mRNA degradation. Many researches on 5'UTR-mediated translation regulation often overlook changes in other cellular parameters such as mRNA concentration, transcription and mRNA stability. Therefore, it is necessary to integrate the regulations mediated by 5'UTR in the three dimensions of transcription, mRNA stability, and translation.

My PhD work is fully in line with this perspective. We will better characterize the link between 5'UTR-mediated regulations of translation initiation, and the mRNA concentration and stability. First, we will develop a synthetic approach based on a set of 5'UTRs homogenous in length but not in sequence, that cover the whole range of theoretical translation initiation rate. Fused to three different reporter genes, the aim will be to measure different parameters such as protein expression, mRNA concentration and mRNA stability. These data will be integrated to identify how, and at which level, the 5'UTRs regulate gene expression: do 5'UTRs act on translation, transcription and/or stability of the mRNA? Do these regulations depend on the downstream gene or not? Second, I will modulate the 5'UTR-mediated translation regulation and follow the effects on mRNA concentration and stability. This work will be part of a more global project of the team dealing with the interplay between translation, transcription and degradation of mRNAs.

The literature review also clearly demonstrates that most of the elements located in the 5'UTR respond to environmental changes. However, the full role of 5'UTRs in adaptation is not clearly understood. To address this issue, we will develop an original approach using an exhaustive library of native *E. coli* 5'UTRs cultured in different growth environments. We will design and construct a library that will contain all the regulatory elements (already functionally defined or not) of the native 5'UTRs.

The library will be challenged under different growth conditions to identify the 5'UTRs that regulate gene expression in response to the environment and 5'UTR elements that may be related to adaptation.

Overall, my PhD work will aim at investigating if 5'UTRs are "regulatory-hubs" of gene expression in *E. coli* by acting as "multiple controllers" of gene expression and responding to constantly changing environment. This work should lead to a better understanding of the role of 5'UTRs in gene expression regulation in *E. coli* and provide new perspectives for the development of molecular tools in the fields of biotechnology and synthetic biology.

References

Aboshkiwa, M., al-Ani, B., Coleman, G., & Rowland, G. (1992). Cloning and physical mapping of the Staphylococcus aureus rplL, rpoB and rpoC genes, encoding ribosomal protein L7/L12 and RNA polymerase subunits beta and beta'. *Journal of General Microbiology*, *138*(9), 1875–1880. https://doi.org/10.1099/00221287-138-9-1875

Adams, P. P., Baniulyte, G., Esnault, C., Chegireddy, K., Singh, N., Monge, M., Dale, R. K., Storz, G., & Wade, J. T. (2021). Regulatory roles of Escherichia coli 5' UTR and ORF-internal RNAs detected by 3' end mapping. *ELife*, *10*, e62438. https://doi.org/10.7554/eLife.62438

Adams, P. P., & Storz, G. (2020). Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1863*(7), 194524. https://doi.org/10.1016/j.bbagrm.2020.194524

Aït-Bara, S., Carpousis, A. J., & Quentin, Y. (2015). RNase E in the γ-Proteobacteria: Conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains. *Molecular Genetics and Genomics*, *290*(3), 847–862. https://doi.org/10.1007/s00438-014-0959-5

Al-Hinai, M. A., Fast, A. G., & Papoutsakis, E. T. (2012). Novel System for Efficient Isolation of Clostridium Double-Crossover Allelic Exchange Mutants Enabling Markerless Chromosomal Gene Deletions and DNA Integration. *Applied and Environmental Microbiology*, *78*(22), 8112–8121. https://doi.org/10.1128/AEM.02214-12

Alifano, P., Rivellini, F., Limauro, D., Bruni, C. B., & Carlomagno, M. S. (1991). A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell*, *64*(3), 553–563. https://doi.org/10.1016/0092-8674(91)90239-u

Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., Moreira, R. N., Pobre, V., Reis, F. P., Saramago, M., Silva, I. J., & Viegas, S. C. (2010). The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiology Reviews*, *34*(5), 883–923. https://doi.org/10.1111/j.1574-6976.2010.00242.x

Artsimovitch, I., & Landick, R. (2002). The Transcriptional Regulator RfaH Stimulates RNA Chain Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand. *Cell*, *109*(2), 193–203. https://doi.org/10.1016/S0092-8674(02)00724-9

Awano, N., Rajagopal, V., Arbing, M., Patel, S., Hunt, J., Inouye, M., & Phadtare, S. (2010). *Escherichia coli* RNase R Has Dual Activities, Helicase and RNase. *Journal of Bacteriology*, *192*(5), 1344–1352. https://doi.org/10.1128/JB.01368-09

Bacun-Druzina, V., Butorac, A., Mrvčić, J., Dragicevic, T., & Stehlik-Tomas, V. (2011). Bacterial Stationary-Phase Evolution. *Food Technology and Biotechnology*, *49*.

Baker, K. E., & Mackie, G. A. (2003). Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*. *Molecular Microbiology*, *47*(1), 75–88. https://doi.org/10.1046/j.1365-2958.2003.03292.x

Bandyra, K. J., Bouvier, M., Carpousis, A. J., & Luisi, B. F. (2013). The social fabric of the RNA degradosome. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1829*(6–7), 514–522. https://doi.org/10.1016/j.bbagrm.2013.02.011

Bandyra, K. J., & Luisi, B. F. (2018). RNase E and the High-Fidelity Orchestration of RNA Metabolism. *Microbiology Spectrum*, *6*(2), 6.2.23. https://doi.org/10.1128/microbiolspec.RWR-0008-2017

Bandyra, K. J., Said, N., Pfeiffer, V., Górna, M. W., Vogel, J., & Luisi, B. F. (2012). The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Molecular Cell*, *47*(6), 943–953. https://doi.org/10.1016/j.molcel.2012.07.015

Bandyra, K. J., Wandzik, J. M., & Luisi, B. F. (2018). Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E. *Molecular Cell*, *72*(2), 275-285.e4. https://doi.org/10.1016/j.molcel.2018.08.039

Banerjee, S., Chalissery, J., Bandey, I., & Sen, R. (2006). Rho-dependent Transcription Termination: More Questions than Answers. *Journal of Microbiology (Seoul, Korea)*, *44*(1), 11–22. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1838574/

Bardey, V., Vallet, C., Robas, N., Charpentier, B., Thouvenot, B., Mougin, A., Hajnsdorf, E., Régnier, P., Springer, M., & Branlant, C. (2005). Characterization of the molecular mechanisms involved in the differential production of erythrose-4-phosphate dehydrogenase, 3-phosphoglycerate kinase and class II fructose-1,6-bisphosphate aldolase in Escherichia coli. *Molecular Microbiology*, *57*(5), 1265–1287. https://doi.org/10.1111/j.1365-2958.2005.04762.x

Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A., & Nudler, E. (2005). A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell*, *120*(2), 183–193. https://doi.org/10.1016/j.cell.2004.11.045

Barquist, L., & Vogel, J. (2015). Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annual Review of Genetics*, *49*(1), 367–394. https://doi.org/10.1146/annurev-genet-112414-054804

Barvík, I., Rejman, D., Panova, N., Šanderová, H., & Krásný, L. (2017). Non-canonical transcription initiation: The expanding universe of transcription initiating substrates. *FEMS Microbiology Reviews*, *41*(2), 131–138. https://doi.org/10.1093/femsre/fuw041

Bear, D. G., Hicks, P. S., Escudero, K. W., Andrews, C. L., McSwiggen, J. A., & von Hippel, P. H. (1988). Interactions of Escherichia coli transcription termination factor rho with RNA: II. Electron microscopy and nuclease protection experiments. *Journal of Molecular Biology*, *199*(4), 623–635. https://doi.org/10.1016/0022-2836(88)90306-3

Bechhofer, D. H., & Deutscher, M. P. (2019). Bacterial ribonucleases and their roles in RNA metabolism. *Critical Reviews in Biochemistry and Molecular Biology*, *54*(3), 242–300. https://doi.org/10.1080/10409238.2019.1651816

Beck, H. J., & Moll, I. (2018). Leaderless mRNAs in the Spotlight: Ancient but Not Outdated! *Microbiology Spectrum*, *6*(4), 6.4.02. https://doi.org/10.1128/microbiolspec.RWR-0016-2017

Belasco, J. G. (2017). Ribonuclease E: Chopping Knife and Sculpting Tool. *Molecular Cell*, *65*(1), 3–4. https://doi.org/10.1016/j.molcel.2016.12.015

Belogurov, G. A., & Artsimovitch, I. (2015). Regulation of Transcript Elongation. *Annual Review of Microbiology*, *69*(1), 49–69. https://doi.org/10.1146/annurev-micro-091014-104047

Benoni, R., Culka, M., Hudeček, O., Gahurova, L., & Cahová, H. (2020). Dinucleoside Polyphosphates as RNA Building Blocks with Pairing Ability in Transcription Initiation. *ACS Chemical Biology*, *15*(7), 1765–1772. https://doi.org/10.1021/acschembio.0c00178

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S., & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, *99*(15), 9697–9702. https://doi.org/10.1073/pnas.112318199

Bhaskaran, H., & Russell, R. (2007). Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature*, *449*(7165), 1014–1018. https://doi.org/10.1038/nature06235

Blount, Z. D. (2015). The unexhausted potential of *E. coli*. *ELife*, *4*, e05826. https://doi.org/10.7554/eLife.05826

Bock, L. V., Kolář, M. H., & Grubmüller, H. (2018). Molecular simulations of the ribosome and associated translation factors. *Current Opinion in Structural Biology*, *49*, 27–35. https://doi.org/10.1016/j.sbi.2017.11.003

Bouvier, M., & Carpousis, A. J. (2011). A tale of two mRNA degradation pathways mediated by RNase E: RNase E pathways. *Molecular Microbiology*, *82*(6), 1305–1310. https://doi.org/10.1111/j.1365-2958.2011.07894.x

Breaker, R. R. (2010). RNA Switches Out in the Cold. *Molecular Cell*, *37*(1), 1–2. https://doi.org/10.1016/j.molcel.2009.12.032

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., & Heumann, H. (2004). The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. *Nature Structural & Molecular Biology*, *11*(6), 551–557. https://doi.org/10.1038/nsmb768

Browning, D. F., & Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, *2*(1), 57–65. https://doi.org/10.1038/nrmicro787

Bruce, H. A., Du, D., Matak-Vinkovic, D., Bandyra, K. J., Broadhurst, R. W., Martin, E., Sobott, F., Shkumatov, A. V., & Luisi, B. F. (2018). Analysis of the natively unstructured RNA/protein-recognition core in the Escherichia coli RNA degradosome and its interactions with regulatory RNA/Hfq complexes. *Nucleic Acids Research*, *46*(1), 387–402. https://doi.org/10.1093/nar/gkx1083

Burkhardt, N., Jünemann, R., Spahn, C. M. T., & Nierhaus, K. H. (1998). Ribosomal tRNA Binding Sites: Three-Site Models of Translation. *Critical Reviews in Biochemistry and Molecular Biology*, *33*(2), 95–149. https://doi.org/10.1080/10409239891204189

Burnett, B. J., Altman, R. B., Ferguson, A., Wasserman, M. R., Zhou, Z., & Blanchard, S. C. (2014). Direct Evidence of an Elongation Factor-Tu/Ts·GTP·Aminoacyl-tRNA Quaternary Complex. *The Journal of Biological Chemistry*, *289*(34), 23917–23927. https://doi.org/10.1074/jbc.M114.583385

Cahová, H., Winz, M.-L., Höfer, K., Nübel, G., & Jäschke, A. (2015). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature*, *519*(7543), 374–377. https://doi.org/10.1038/nature14020

Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., & Luisi, B. F. (2005). Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature*, *437*(7062), 1187–1191. https://doi.org/10.1038/nature04084

Callaghan, A. J., Redko, Y., Murphy, L. M., Grossmann, J. G., Yates, D., Garman, E., Ilag, L. L., Robinson, C. V., Symmons, Martyn. F., McDowall, K. J., & Luisi, B. F. (2005). "Zn-Link": A Metal-Sharing Interface that Organizes the Quaternary Structure and Catalytic Site of the Endoribonuclease, RNase E. *Biochemistry*, *44*(12), 4667–4675. https://doi.org/10.1021/bi0478244

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell*, *104*(6), 901–912. https://doi.org/10.1016/S0092-8674(01)00286-0

Carabetta, V. J., Silhavy, T. J., & Cristea, I. M. (2010). The Response Regulator SprE (RssB) Is Required for Maintaining Poly(A) Polymerase I-Degradosome Association during Stationary Phase. *Journal of Bacteriology*, *192*(14), 3713–3721. https://doi.org/10.1128/JB.00300-10

Carpousis, A. J., & Gralla, J. D. (1980). Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry*, *19*(14), 3245–3253. https://doi.org/10.1021/bi00555a023

Chao, Y., Li, L., Girodat, D., Förstner, K. U., Said, N., Corcoran, C., Śmiga, M., Papenfort, K., Reinhardt, R., Wieden, H.-J., Luisi, B. F., & Vogel, J. (2017). In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Molecular Cell*, *65*(1), 39–51. https://doi.org/10.1016/j.molcel.2016.11.002

Charollais, J. (2004). CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Research*, *32*(9), 2751–2759. https://doi.org/10.1093/nar/gkh603

Chen, H., Shiroguchi, K., Ge, H., & Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *E scherichia coli* . *Molecular Systems Biology*, *11*(1), 781. https://doi.org/10.15252/msb.20145794

Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A. Y., Olinares, P. D. B., Saecker, R. M., Winkelman, J. T., Maloney, M. F., Chait, B. T., Ross, W., Gourse, R. L., Campbell, E. A., & Darst, S. A. (2020). Stepwise Promoter Melting by Bacterial RNA Polymerase. *Molecular Cell*, *78*(2), 275-288.e6. https://doi.org/10.1016/j.molcel.2020.02.017

Chen, J., & Gottesman, S. (2017). Hfq links translation repression to stress-induced mutagenesis in *E. coli*. *Genes & Development*, *31*(13), 1382–1395. https://doi.org/10.1101/gad.302547.117

Chen, J., Morita, T., & Gottesman, S. (2019). Regulation of Transcription Termination of Small RNAs and by Small RNAs: Molecular Mechanisms and Biological Functions. *Frontiers in Cellular and Infection Microbiology*, *9*, 201. https://doi.org/10.3389/fcimb.2019.00201

Cheng, Z.-F., & Deutscher, M. P. (2005). An Important Role for RNase R in mRNA Decay. *Molecular Cell*, *17*(2), 313–318. https://doi.org/10.1016/j.molcel.2004.11.048

Chiaruttini, C., & Guillier, M. (2020). On the role of mRNA secondary structure in bacterial translation. *WIREs RNA*, *11*(3). https://doi.org/10.1002/wrna.1579

Cho, B.-K., Kim, D., Knight, E. M., Zengler, K., & Palsson, B. O. (2014). Genome-scale reconstruction of the sigma factor network in Escherichia coli: Topology and functional states. *BMC Biology*, *12*(1), 4. https://doi.org/10.1186/1741-7007-12-4

Chowdhury, S., Maris, C., Allain, F. H.-T., & Narberhaus, F. (2006). Molecular basis for temperature sensing by an RNA thermometer. *The EMBO Journal*, *25*(11), 2487–2497. https://doi.org/10.1038/sj.emboj.7601128

Cohen, S. N. (1995). Surprises at the 3' end of prokaryotic RNA. *Cell*, *80*(6), 829–832. https://doi.org/10.1016/0092-8674(95)90284-8

Court, D. L., Gan, J., Liang, Y.-H., Shaw, G. X., Tropea, J. E., Costantino, N., Waugh, D. S., & Ji, X. (2013). RNase III: Genetics and Function; Structure and Mechanism. *Annual Review of Genetics*, *47*(1), 405–431. https://doi.org/10.1146/annurev-genet-110711-155618

Dambach, M., Sandoval, M., Updegrove, T. B., Anantharaman, V., Aravind, L., Waters, L. S., & Storz, G. (2015). The Ubiquitous yybP-ykoY Riboswitch Is a Manganese-Responsive Regulatory Element. *Molecular Cell*, *57*(6), 1099–1109. https://doi.org/10.1016/j.molcel.2015.01.035

Davis, C. A., Bingman, C. A., Landick, R., Record, M. T., & Saecker, R. M. (2007). Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences*, *104*(19), 7833–7838. https://doi.org/10.1073/pnas.0609888104

de Smit, M. H., & van Duin, J. (2003). Translational Standby Sites: How Ribosomes May Deal with the Rapid Folding Kinetics of mRNA. *Journal of Molecular Biology*, *331*(4), 737–743. https://doi.org/10.1016/S0022-2836(03)00809-X

Desnoyers, G., & Massé, E. (2012). Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes & Development*, *26*(7), 726–739. https://doi.org/10.1101/gad.182493.111

Deutscher, M. P. (2015). How bacterial cells keep ribonucleases under control. *FEMS Microbiology Reviews*, *39*(3), 350–361. https://doi.org/10.1093/femsre/fuv012

Devaraj, A., & Fredrick, K. (2010). Short spacing between the Shine-Dalgarno sequence and P codon destabilizes codon-anticodon pairing in the P site to promote +1 programmed frameshifting. *Molecular Microbiology*, *78*(6), 1500–1509. https://doi.org/10.1111/j.1365-2958.2010.07421.x

Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, *505*(7485), 696–700. https://doi.org/10.1038/nature12756

Douglas, J., Kingston, R., & Drummond, A. J. (2020). Bayesian inference and comparison of stochastic transcription elongation models. *PLOS Computational Biology*, *16*(2), e1006717. https://doi.org/10.1371/journal.pcbi.1006717

Dutta, T., & Deutscher, M. P. (2009). Catalytic Properties of RNase BN/RNase Z from Escherichia coli*. *Journal of Biological Chemistry*, *284*(23), 15425–15431. https://doi.org/10.1074/jbc.M109.005462

Duval, M., Korepanov, A., Fuchsbauer, O., Fechter, P., Haller, A., Fabbretti, A., Choulier, L., Micura, R., Klaholz, B. P., Romby, P., Springer, M., & Marzi, S. (2013). Escherichia coli Ribosomal Protein S1 Unfolds Structured mRNAs Onto the Ribosome for Active Translation Initiation. *PLOS Biology*, *11*(12), e1001731. https://doi.org/10.1371/journal.pbio.1001731

Ellis, M. J., Trussler, R. S., & Haniford, D. B. (2015). A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition. *Nucleic Acids Research*, *43*(13), 6511–6527. https://doi.org/10.1093/nar/gkv584

Espah Borujeni, A., Channarasappa, A. S., & Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, *42*(4), 2646–2659. https://doi.org/10.1093/nar/gkt1139

Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A. J., Girbal, L., & Cocaign-Bousquet, M. (2014). Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Research*, *42*(4), 2460–2472. https://doi.org/10.1093/nar/gkt1150

Feklístov, A., Sharon, B. D., Darst, S. A., & Gross, C. A. (2014). Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annual Review of Microbiology*, *68*(1), 357–376. https://doi.org/10.1146/annurev-micro-092412-155737

Fredrick, K., & Helmann, J. D. (1997). RNA polymerase sigma factor determines start-site selection but is not required for upstream promoter element activation on heteroduplex (bubble) templates. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 4982–4987. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC24617/

Freistroffer, D. V., Pavlov, M. Y., MacDougall, J., Buckingham, R. H., & Ehrenberg, M. (1997). Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *The EMBO Journal*, *16*(13), 4126–4133. https://doi.org/10.1093/emboj/16.13.4126

Fu, Y., Yu, Z., Zhu, L., Li, Z., Yin, W., Shang, X., Chou, S.-H., Tan, Q., & He, J. (2021). The Multiple Regulatory Relationship Between RNA-Chaperone Hfq and the Second Messenger c-di-GMP. *Frontiers in Microbiology*, *12*, 689619. https://doi.org/10.3389/fmicb.2021.689619

Gallo, S., Oberhuber, M., Sigel, R. K. O., & Kräutler, B. (2008). The Corrin Moiety of Coenzyme B12 is the Determinant for Switching the btuB Riboswitch of *E. coli*. *ChemBioChem*, *9*(9), 1408–1414. https://doi.org/10.1002/cbic.200800099

Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., Van Roey, P., Agrawal, R. K., Harvey, S. C., Sali, A., Chapman, M. S., & Frank, J. (2003). Study of the Structural Dynamics of the *E. coli* 70S Ribosome Using Real-Space Refinement. *Cell*, *113*(6), 789–801. https://doi.org/10.1016/S0092-8674(03)00427-6

Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: Structures and Mechanisms. *Cold Spring Harbor Perspectives in Biology*, *3*(6), a003533. https://doi.org/10.1101/cshperspect.a003533

Ghosh, S., & Deutscher, M. P. (1999). Oligoribonuclease is an essential component of the mRNA decay pathway. *Proceedings of the National Academy of Sciences*, *96*(8), 4372–4377. https://doi.org/10.1073/pnas.96.8.4372

GHOSH, S., & JOSEPH, S. (2005). Nonbridging phosphate oxygens in 16S rRNA important for 30S subunit assembly and association with the 50S ribosomal subunit. *RNA*, *11*(5), 657–667. https://doi.org/10.1261/rna.7224305

Giuliodori, A. M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., Gualerzi, C. O., & Pon, C. L. (2010). The cspA mRNA Is a Thermosensor that Modulates Translation of the Cold-Shock Protein CspA. *Molecular Cell*, *37*(1), 21–33. https://doi.org/10.1016/j.molcel.2009.11.033

Glyde, R., Ye, F., Jovanovic, M., Kotta-Loizou, I., Buck, M., & Zhang, X. (2018). Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. *Molecular Cell*, *70*(6), 1111-1120.e3. https://doi.org/10.1016/j.molcel.2018.05.021

González-González, A., Hug, S. M., Rodríguez-Verdugo, A., Patel, J. S., & Gaut, B. S. (2017). Adaptive Mutations in RNA Polymerase and the Transcriptional Terminator Rho Have Similar Effects on Escherichia coli Gene Expression. *Molecular Biology and Evolution*, *34*(11), 2839–2855. https://doi.org/10.1093/molbev/msx216

Górna, M. W., Carpousis, A. J., & Luisi, B. F. (2012). From conformational chaos to robust regulation: The structure and function of the multi-enzyme RNA degradosome. *Quarterly Reviews of Biophysics*, *45*(2), 105–145. https://doi.org/10.1017/S003358351100014X

Graf, M., Arenz, S., Huter, P., Dönhöfer, A., Nováček, J., & Wilson, D. N. (2017). Cryo-EM structure of the spinach chloroplast ribosome reveals the location of plastid-specific ribosomal proteins and extensions. *Nucleic Acids Research*, *45*(5), 2887–2896. https://doi.org/10.1093/nar/gkw1272

Graf, M., Huter, P., Maracci, C., Peterek, M., Rodnina, M. V., & Wilson, D. N. (2018). Visualization of translation termination intermediates trapped by the Apidaecin 137 peptide during RF3-mediated recycling of RF1. *Nature Communications*, *9*(1), 3053. https://doi.org/10.1038/s41467-018-05465-1

Hadjeras, L., Poljak, L., Bouvier, M., Morin-Ogier, Q., Canal, I., Cocaign-Bousquet, M., Girbal, L., & Carpousis, A. J. (2019). Detachment of the RNA degradosome from the inner membrane of *Escherichia coli* results in a global slowdown of mRNA degradation, proteolysis of RNase E and increased turnover of ribosome-free transcripts. *Molecular Microbiology*, *111*(6), 1715–1731. https://doi.org/10.1111/mmi.14248

Hajnsdorf, E., & Kaberdin, V. R. (2018). RNA polyadenylation and its consequences in prokaryotes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1762), 20180166. https://doi.org/10.1098/rstb.2018.0166

Hankins, J. S., Zappavigna, C., Prud'homme-Généreux, A., & Mackie, G. A. (2007). Role of RNA Structure and Susceptibility to RNase E in Regulation of a Cold Shock mRNA, cspA mRNA. *Journal of Bacteriology*, *189*(12), 4353–4358. https://doi.org/10.1128/JB.00193-07

Hartz, D., McPheeters, D. S., & Gold, L. (1991). Influence of mRNA determinants on translation initiation in Escherichia coli. *Journal of Molecular Biology*, *218*(1), 83–97. https://doi.org/10.1016/0022-2836(91)90875-7

Hatoum, A., & Roberts, J. (2008). Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. *Molecular Microbiology*, *68*(1), 17–28. https://doi.org/10.1111/j.1365-2958.2008.06138.x

Hein, P. P., & Landick, R. (2010). The bridge helix coordinates movements of modules in RNA polymerase. *BMC Biology*, *8*(1), 141. https://doi.org/10.1186/1741-7007-8-141

Herbert, K. M., La Porta, A., Wong, B. J., Mooney, R. A., Neuman, K. C., Landick, R., & Block, S. M. (2006). Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell*, *125*(6), 1083–1094. https://doi.org/10.1016/j.cell.2006.04.032

Heyduk, E., & Heyduk, T. (2018). DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Research*, *46*(9), 4469–4486. https://doi.org/10.1093/nar/gky172

Holmqvist, E., & Vogel, J. (2013). A small RNA serving both the Hfq and CsrA regulons. *Genes & Development*, *27*(10), 1073–1078. https://doi.org/10.1101/gad.220178.113

Hook-Barnard, I. G., & Hinton, D. M. (2007). Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters. *Gene Regulation and Systems Biology*, *1*, 117762500700100. https://doi.org/10.1177/117762500700100020

Hör, J., Di Giorgio, S., Gerovac, M., Venturini, E., Förstner, K. U., & Vogel, J. (2020). Grad-seq shines light on unrecognized RNA and protein complexes in the model bacterium Escherichia coli. *Nucleic Acids Research*, *48*(16), 9301–9319. https://doi.org/10.1093/nar/gkaa676

Hossain, S. T., Malhotra, A., & Deutscher, M. P. (2016). How RNase R Degrades Structured RNA. *Journal of Biological Chemistry*, *291*(15), 7877–7887. https://doi.org/10.1074/jbc.M116.717991

Houseley, J., & Tollervey, D. (2009). The Many Pathways of RNA Degradation. *Cell*, *136*(4), 763–776. https://doi.org/10.1016/j.cell.2009.01.019

Hsu, L. M. (2002). Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1577*(2), 191–207. https://doi.org/10.1016/S0167-4781(02)00452-9

Hsu, L. M. (2009). Monitoring Abortive Initiation. *Methods (San Diego, Calif.)*, *47*(1), 25–36. https://doi.org/10.1016/j.ymeth.2008.10.010

Hughes, D. (2016). Using the power of genetic suppressors to probe the essential functions of RNase E. *Current Genetics*, *62*(1), 53–57. https://doi.org/10.1007/s00294-015-0510-1

Hui, M. P., Foley, P. L., & Belasco, J. G. (2014). Messenger RNA Degradation in Bacterial Cells. *Annual Review of Genetics*, *48*(1), 537–559. https://doi.org/10.1146/annurev-genet-120213-092340

Hurwitz, J. (2005). The Discovery of RNA Polymerase. *Journal of Biological Chemistry*, *280*(52), 42477–42485. https://doi.org/10.1074/jbc.X500006200

Hussain, T., Llácer, J. L., Wimberly, B. T., Kieft, J. S., & Ramakrishnan, V. (2016). Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell*, *167*(1), 133-144.e13. https://doi.org/10.1016/j.cell.2016.08.074

Ikeda, Y., Yagi, M., Morita, T., & Aiba, H. (2011). Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli: Hfq-binding site within RNase E. *Molecular Microbiology*, *79*(2), 419–432. https://doi.org/10.1111/j.1365-2958.2010.07454.x

Ishihama, A. (2000). Functional Modulation of *Escherichia Coli* RNA Polymerase. *Annual Review of Microbiology*, *54*(1), 499–518. https://doi.org/10.1146/annurev.micro.54.1.499

Jarmoskaite, I., & Russell, R. (2014). RNA Helicase Proteins as Chaperones and Remodelers. *Annual Review of Biochemistry*, *83*(1), 697–725. https://doi.org/10.1146/annurev-biochem-060713-035546

Joyce, S. A., & Dreyfus, M. (1998). In the absence of translation, RNase E can bypass 5′ mRNA stabilizers in Escherichia coli. *Journal of Molecular Biology*, *282*(2), 241–254. https://doi.org/10.1006/jmbi.1998.2027

Julius, C., & Yuzenkova, Y. (2019). Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate. *WIREs RNA*, *10*(2), e1512. https://doi.org/10.1002/wrna.1512

Kammerer, W., Deuschle, U., Gentz, R., & Bujard, H. (1986). Functional dissection of Escherichia coli promoters: Information in the transcribed region is involved in late steps of the overall process. *The EMBO Journal*, *5*(11), 2995–3000. https://doi.org/10.1002/j.1460-2075.1986.tb04597.x

Kawamoto, H., Koide, Y., Morita, T., & Aiba, H. (2006). Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Molecular Microbiology*, *61*(4), 1013–1022. https://doi.org/10.1111/j.1365-2958.2006.05288.x

Khemici, V., Poljak, L., Toesca, I., & Carpousis, A. J. (2005). Evidence *in vivo* that the DEAD-box RNA helicase RhlB facilitates the degradation of ribosome-free mRNA by RNase E. *Proceedings of the National Academy of Sciences*, *102*(19), 6913–6918. https://doi.org/10.1073/pnas.0501129102

Kim, D., Hong, J. S.-J., Qiu, Y., Nagarajan, H., Seo, J.-H., Cho, B.-K., Tsai, S.-F., & Palsson, B. Ø. (2012). Comparative Analysis of Regulatory Elements between Escherichia coli and Klebsiella pneumoniae by Genome-Wide Transcription Start Site Profiling. *PLoS Genetics*, *8*(8), e1002867. https://doi.org/10.1371/journal.pgen.1002867

Kime, L., Jourdan, S. S., Stead, J. A., Hidalgo-Sastre, A., & McDowall, K. J. (2009). Rapid cleavage of RNA by RNase E in the absence of 5′ monophosphate stimulation: Rapid 5′ monophosphate-independent cleavage by *E. coli* RNase E. *Molecular Microbiology*, *76*(3), 590–604. https://doi.org/10.1111/j.1365-2958.2009.06935.x

Kireeva, M. L., & Kashlev, M. (2009). Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(22), 8900–8905. https://doi.org/10.1073/pnas.0900407106

Komarova, A. V., Tchufistova, L. S., Dreyfus, M., & Boni, I. V. (2005). AU-Rich Sequences within 5J Untranslated Leaders Enhance Translation and Stabilize mRNA in Escherichia coli. *J. BACTERIOL.*, *187*, 6.

Komarova, E. S., Chervontseva, Z. S., Osterman, I. A., Evfratov, S. A., Rubtsova, M. P., Zatsepin, T. S., Semashko, T. A., Kostryukova, E. S., Bogdanov, A. A., Gelfand, M. S., Dontsova, O. A., & Sergiev, P. V. (2020). Influence of the spacer region between the Shine–Dalgarno box and the start codon for fine-tuning of the translation efficiency in *Escherichia coli*. *Microbial Biotechnology*, *13*(4), 1254–1261. https://doi.org/10.1111/1751-7915.13561

Koslover, D. J., Callaghan, A. J., Marcaida, M. J., Garman, E. F., Martick, M., Scott, W. G., & Luisi, B. F. (2008). The Crystal Structure of the Escherichia coli RNase E Apoprotein and a Mechanism for RNA Degradation. *Structure*, *16*(8), 1238–1244. https://doi.org/10.1016/j.str.2008.04.017

Kucharova, V. (2012). Expression of recombinant proteins in Escherichia coli: The influence of the nucleotide sequences at the 5´ ends of target genes. In *236*. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/245828

Kushner, S. R. (2002). MRNA Decay in *Escherichia coli* Comes of Age. *Journal of Bacteriology*, *184*(17), 4658–4665. https://doi.org/10.1128/JB.184.17.4658-4665.2002

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions*, *34*(6), 1062–1066. https://doi.org/10.1042/BST0341062

Lauber, M. A., Rappsilber, J., & Reilly, J. P. (2012). Dynamics of Ribosomal Protein S1 on a Bacterial Ribosome with Cross-Linking and Mass Spectrometry *. *Molecular & Cellular Proteomics*, *11*(12), 1965–1976. https://doi.org/10.1074/mcp.M112.019562

Laursen, B. S., Sørensen, H. P., Mortensen, K. K., & Sperling-Petersen, H. U. (2005). Initiation of Protein Synthesis in Bacteria. *Microbiology and Molecular Biology Reviews*, *69*(1), 101–123. https://doi.org/10.1128/MMBR.69.1.101-123.2005

Lease, R. A., Cusick, M. E., & Belfort, M. (1998). Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proceedings of the National Academy of Sciences*, *95*(21), 12456–12461. https://doi.org/10.1073/pnas.95.21.12456

Li, Y., & Altman, S. (2003). A specific endoribonuclease, RNase P, affects gene expression of polycistronic operon mRNAs. *Proceedings of the National Academy of Sciences*, *100*(23), 13213–13218. https://doi.org/10.1073/pnas.2235589100

Liao, H., Liu, M., & Guo, X. (2018). The special existences: NanoRNA and nanoRNase. *Microbiological Research*, *207*, 134–139. https://doi.org/10.1016/j.micres.2017.11.014

Lin-Chao, S., Wong, T. T., McDowall, K. J., & Cohen, S. N. (1994). Effects of nucleotide sequence on the specificity of rne-dependent and RNase E-mediated cleavages of RNA I encoded by the pBR322 plasmid. *The Journal of Biological Chemistry*, *269*(14), 10797–10803.

Liu, G., Song, G., Zhang, D., Zhang, D., Li, Z., Lyu, Z., Dong, J., Achenbach, J., Gong, W., Zhao, X. S., Nierhaus, K. H., & Qin, Y. (2014). EF-G catalyzes tRNA translocation by disrupting interactions between decoding center and codon-anticodon duplex. *Nature Structural & Molecular Biology*, *21*(9), 817–824. https://doi.org/10.1038/nsmb.2869

Luciano, D. J., & Belasco, J. G. (2015). NAD in RNA: Unconventional headgear. *Trends in Biochemical Sciences*, *40*(5), 245–247. https://doi.org/10.1016/j.tibs.2015.03.004

Luciano, D. J., & Belasco, J. G. (2020). Np4A alarmones function in bacteria as precursors to RNA caps. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(7), 3560–3567. https://doi.org/10.1073/pnas.1914229117

Luciano, D. J., Levenson-Palmer, R., & Belasco, J. G. (2019). Stresses that raise Np4A levels induce protective nucleoside tetraphosphate capping of bacterial RNA. *Molecular Cell*, *75*(5), 957-966.e8. https://doi.org/10.1016/j.molcel.2019.05.031

Mackie, G. A. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. *Nature*, *395*(6703), 720–723. https://doi.org/10.1038/27246

Mackie, G. A. (2013). RNase E: At the interface of bacterial RNA processing and decay. *Nature Reviews Microbiology*, *11*(1), 45–57. https://doi.org/10.1038/nrmicro2930

Maes, A., Gracia, C., Innocenti, N., Zhang, K., Aurell, E., & Hajnsdorf, E. (2016). Landscape of RNA polyadenylation in *E. coli*. *Nucleic Acids Research*, gkw894. https://doi.org/10.1093/nar/gkw894

Malinen, A. M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M. S., & Belogurov, G. A. (2012). Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Research*, *40*(15), 7442–7451. https://doi.org/10.1093/nar/gks383

Márquez, V., Wilson, D. N., & Nierhaus, K. H. (2002). Functions and interplay of the tRNA-binding sites of the ribosome. *Biochemical Society Transactions*, *30*(2), 133–140. https://doi.org/10.1042/bst0300133

Mathy, N., Bénard, L., Pellegrini, O., Daou, R., Wen, T., & Condon, C. (2007). 5'-to-3' Exoribonuclease Activity in Bacteria: Role of RNase J1 in rRNA Maturation and 5' Stability of mRNA. *Cell*, *129*(4), 681–692. https://doi.org/10.1016/j.cell.2007.02.051

Mazumder, A., & Kapanidis, A. N. (2019). Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. *Journal of Molecular Biology*, *431*(20), 3947–3959. https://doi.org/10.1016/j.jmb.2019.04.046

McDowall, K. J., Lin-Chao, S., & Cohen, S. N. (1994). A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *Journal of Biological Chemistry*, *269*(14), 10790–10796. https://doi.org/10.1016/S0021-9258(17)34129-7

Mohanty, B. K., & Kushner, S. R. (1999). Analysis of the function of Escherichia coli poly(A) polymerase I in RNA metabolism. *Molecular Microbiology*, *34*(5), 1094–1108. https://doi.org/10.1046/j.1365-2958.1999.01673.x

Mohanty, B. K., & Kushner, S. R. (2003). Genomic analysis in Escherichia coli demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay: The role of exonucleases in *E. coli* mRNA decay. *Molecular Microbiology*, *50*(2), 645–658. https://doi.org/10.1046/j.1365-2958.2003.03724.x

Mohanty, B. K., & Kushner, S. R. (2006). The majority of Escherichia coli mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic Acids Research*, *34*(19), 5695–5704. https://doi.org/10.1093/nar/gkl684

Mohanty, B. K., & Kushner, S. R. (2018). *Enzymes Involved in Posttranscriptional RNA Metabolism in Gram-Negative Bacteria*. 16.

Mohanty, B. K., Maples, V. F., & Kushner, S. R. (2004). The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli: Hfq control of polyadenylation. *Molecular Microbiology*, *54*(4), 905–920. https://doi.org/10.1111/j.1365-2958.2004.04337.x

Morgan, W., Bear, D. G., Litchman, B. L., & von Hippel, P. H. (1985). RNA sequence and secondary structure requiresments for rho-dependent transcription termination. *Nucleic Acids Research*, *13*(10), 3739–3754. https://doi.org/10.1093/nar/13.10.3739

Morita, M. T., Tanaka, Y., Kodama, T. S., Kyogoku, Y., Yanagi, H., & Yura, T. (1999). Translational induction of heat shock transcription factor ç32: Evidence for a built-in RNA thermosensor. *Genes & Development*, *13*(6), 655–665. http://genesdev.cshlp.org/content/13/6/655

Morita, T., Mochizuki, Y., & Aiba, H. (2006). Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proceedings of the National Academy of Sciences*, *103*(13), 4858–4863. https://doi.org/10.1073/pnas.0509638103

Murakami, K., Kimura, M., Owens, J. T., Meares, C. F., & Ishihama, A. (1997). The two α subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. *Proceedings of the National Academy of Sciences*, *94*(5), 1709–1714. https://doi.org/10.1073/pnas.94.5.1709

Na, D., & Lee, D. (2010). RBSDesigner: Software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics (Oxford, England)*, *26*(20), 2633–2634. https://doi.org/10.1093/bioinformatics/btq458

Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J., & Block, S. M. (2003). Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. *Cell*, *115*(4), 437–447. https://doi.org/10.1016/S0092-8674(03)00845-6

Nicholson, A. W. (2014). Ribonuclease III mechanisms of double-stranded RNA cleavage: Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdisciplinary Reviews: RNA*, *5*(1), 31–48. https://doi.org/10.1002/wrna.1195

Nickels, B. E. (2012). A new way to start: NanoRNA-mediated priming of transcription initiation. *Transcription*, *3*(6), 300–304. https://doi.org/10.4161/trns.21903

Nickels, B. E., & Dove, S. L. (2011). NanoRNAs: A Class of Small RNAs That Can Prime Transcription Initiation in Bacteria. *Journal of Molecular Biology*, *412*(5), 772–781. https://doi.org/10.1016/j.jmb.2011.06.015

Nickels, B. E., Mukhopadhyay, J., Garrity, S. J., Ebright, R. H., & Hochschild, A. (2004). The σ70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. *Nature Structural & Molecular Biology*, *11*(6), 544–550. https://doi.org/10.1038/nsmb757

O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E., & Kushner, S. R. (1995). Polyadenylylation Helps Regulate mRNA Decay in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(6), 1807–1811. http://www.jstor.org/stable/2366880

Olsthoorn, R. C. L., Zoog, S., & Duin, J. (1995). Coevolution of RNA helix stability and Shine-Dalgarno complementarity in a translational start region. *Molecular Microbiology*, *15*(2), 333–339. https://doi.org/10.1111/j.1365-2958.1995.tb02247.x

Osterman, I. A., Evfratov, S. A., Sergiev, P. V., & Dontsova, O. A. (2013). Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Research*, *41*(1), 474–486. https://doi.org/10.1093/nar/gks989

Otsuka, Y., & Yonesaki, T. (2005). A Novel Endoribonuclease, RNase LS, in Escherichia coli. *Genetics*, *169*(1), 13–20. https://doi.org/10.1534/genetics.104.033290

Paget, M. S., & Helmann, J. D. (2003). The σ70 family of sigma factors. *Genome Biology*, *4*(1), 203. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC151288/

Park, J.-S., & Roberts, J. W. (2006). Role of DNA bubble rewinding in enzymatic transcription termination. *Proceedings of the National Academy of Sciences*, *103*(13), 4870–4875. https://doi.org/10.1073/pnas.0600145103

Patel, U. R., Gautam, S., & Chatterji, D. (2020). Validation of Omega Subunit of RNA Polymerase as a Functional Entity. *Biomolecules*, *10*(11), 1588. https://doi.org/10.3390/biom10111588

Pedrolli, D., Langer, S., Hobl, B., Schwarz, J., Hashimoto, M., & Mack, M. (2015). The ribB FMN riboswitch from Escherichia coli operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *The FEBS Journal*, *282*(16), 3230–3242. https://doi.org/10.1111/febs.13226

Peters, J. M., Vangeloff, A. D., & Landick, R. (2011). Bacterial Transcription Terminators: The RNA 3′-End Chronicles. *Journal of Molecular Biology*, *412*(5), 793–813. https://doi.org/10.1016/j.jmb.2011.03.036

Portier, C., Dondon, L., Grunberg-Manago, M., & Régnier, P. (1987). The first step in the functional inactivation of the Escherichia coli polynucleotide phosphorylase messenger is a ribonuclease III processing at the 5′ end. *The EMBO Journal*, *6*(7), 2165–2170. https://doi.org/10.1002/j.1460-2075.1987.tb02484.x

Prévost, K., Desnoyers, G., Jacques, J.-F., Lavoie, F., & Massé, E. (2011). Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes & Development*, *25*(4), 385–396. https://doi.org/10.1101/gad.2001711

Proux, F., Dreyfus, M., & Iost, I. (2011). Identification of the sites of action of SrmB, a DEAD-box RNA helicase involved in Escherichia coli ribosome assembly: RRNA mutations that bypass SrmB requirement. *Molecular Microbiology*, *82*(2), 300–311. https://doi.org/10.1111/j.1365-2958.2011.07779.x

Purusharth, R. I., Klein, F., Sulthana, S., Jäger, S., Jagannadham, M. V., Evguenieva-Hackenberg, E., Ray, M. K., & Klug, G. (2005). Exoribonuclease R Interacts with Endoribonuclease E and an RNA Helicase in the Psychrotrophic Bacterium Pseudomonas syringae Lz4W*. *Journal of Biological Chemistry*, *280*(15), 14572–14578. https://doi.org/10.1074/jbc.M413507200

Qu, X., Lancaster, L., Noller, H. F., Bustamante, C., & Tinoco, I. (2012). Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(36), 14458–14463. https://doi.org/10.1073/pnas.1208950109

Rabhi, M., Espéli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V., & Boudvillain, M. (2011). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *The EMBO Journal*, *30*(14), 2805–2816. https://doi.org/10.1038/emboj.2011.192

Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell*, *108*(4), 557–572. https://doi.org/10.1016/S0092-8674(02)00619-0

Regulski, E. E., Moy, R. H., Weinberg, Z., Barrick, J. E., Yao, Z., Ruzzo, W. L., & Breaker, R. R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Molecular Microbiology*, *68*(4), 918–932. https://doi.org/10.1111/j.1365-2958.2008.06208.x

Revyakin, A., Liu, C., Ebright, R. H., & Strick, T. R. (2006). ABORTIVE INITIATION AND PRODUCTIVE INITIATION BY RNA POLYMERASE INVOLVE DNA SCRUNCHING. *Science (New York, N.Y.)*, *314*(5802), 1139–1143. https://doi.org/10.1126/science.1131398

Richards, J., & Belasco, J. G. (2019). Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Molecular Cell*, *74*(2), 284-295.e5. https://doi.org/10.1016/j.molcel.2019.01.044

Richardson, J. P. (1991). Preventing the synthesis of unused transcripts by rho factor. *Cell*, *64*(6), 1047–1049. https://doi.org/10.1016/0092-8674(91)90257-Y

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., & Gold, L. (1992). Translation initiation in Escherichia coli: Sequences within the ribosome-binding site. *Molecular Microbiology*, *6*(9), 1219–1229. https://doi.org/10.1111/j.1365-2958.1992.tb01561.x

Roberts, J. W. (2019). Mechanisms of Bacterial Transcription Termination. *Journal of Molecular Biology*, *431*(20), 4030–4039. https://doi.org/10.1016/j.jmb.2019.04.003

Roberts, J. W., Shankar, S., & Filter, J. J. (2008). RNA Polymerase Elongation Factors. *Annual Review of Microbiology*, *62*(1), 211–233. https://doi.org/10.1146/annurev.micro.61.080706.093422

Rodnina, M. V. (2012). Quality control of mRNA decoding on the bacterial ribosome. In A. Marintchev (Ed.), *Advances in Protein Chemistry and Structural Biology* (Vol. 86, pp. 95–128). Academic Press. https://doi.org/10.1016/B978-0-12-386497-0.00003-7

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., & Weissman, J. S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, *505*(7485), 701–705. https://doi.org/10.1038/nature12894

Ruff, E., Record, M., & Artsimovitch, I. (2015). Initial Events in Bacterial Transcription Initiation. *Biomolecules*, *5*(2), 1035–1062. https://doi.org/10.3390/biom5021035

Saba, J., Chua, X. Y., Mishanina, T. V., Nayak, D., Windgassen, T. A., Mooney, R. A., & Landick, R. (2019). The elemental mechanism of transcriptional pausing. *ELife*, *8*, e40981. https://doi.org/10.7554/eLife.40981

Saecker, R. M., Record, M. T., & deHaseth, P. L. (2011). Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *Journal of Molecular Biology*, *412*(5), 754–771. https://doi.org/10.1016/j.jmb.2011.01.018

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, *27*(10), 946–950. https://doi.org/10.1038/nbt.1568

Salvail, H., Caron, M.-P., Bélanger, J., & Massé, E. (2013). Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. *The EMBO Journal*, *32*(20), 2764–2778. https://doi.org/10.1038/emboj.2013.205

Santangelo, T. J., & Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. *Nature Reviews Microbiology*, *9*(5), 319–329. https://doi.org/10.1038/nrmicro2560

Sarkar, N. (1997). POLYADENYLATION OF mRNA IN PROKARYOTES. *Annual Review of Biochemistry*, *66*(1), 173–197. https://doi.org/10.1146/annurev.biochem.66.1.173

Schmeing, T. M., & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature*, *461*(7268), 1234–1242. https://doi.org/10.1038/nature08403

Schurr, T., Nadir, E., & Margalit, H. (1993). Identification and characterization of *E. coli* ribosomal binding sites by free energy computation. *Nucleic Acids Research*, *21*(17), 4019–4023. https://doi.org/10.1093/nar/21.17.4019

Scolnick, E., Tompkins, R., Caskey, T., & Nirenberg, M. (1968). Release factors differing in specificity for terminator codons. *Proceedings of the National Academy of Sciences of the United States of America*, *61*(2), 768–774. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC225226/

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., & Nudler, E. (2016). SRNA-Mediated Control of Transcription Termination in *E. coli*. *Cell*, *167*(1), 111-121.e13. https://doi.org/10.1016/j.cell.2016.09.004

Sengupta, J., Agrawal, R. K., & Frank, J. (2001). Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. *Proceedings of the National Academy of Sciences*, *98*(21), 11991–11996. https://doi.org/10.1073/pnas.211266898

Seo, S. W., Yang, J.-S., Kim, I., Yang, J., Min, B. E., Kim, S., & Jung, G. Y. (2013). Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metabolic Engineering*, *15*, 67–74. https://doi.org/10.1016/j.ymben.2012.10.006

Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R., & Patel, D. J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*, *441*(7097), 1167–1171. https://doi.org/10.1038/nature04740

Severinov, K., Mooney, R., Darst, S. A., & Landick, R. (1997). Tethering of the Large Subunits of Escherichia coli RNA Polymerase*. *Journal of Biological Chemistry*, *272*(39), 24137–24140. https://doi.org/10.1074/jbc.272.39.24137

Sharma, A., Alajangi, H. K., Pisignano, G., Sood, V., Singh, G., & Barnwal, R. P. (2022). RNA thermometers and other regulatory elements: Diversity and importance in bacterial pathogenesis. *Wiley Interdisciplinary Reviews. RNA*, e1711. https://doi.org/10.1002/wrna.1711

Sherlock, M. E., Malkowski, S. N., & Breaker, R. R. (2017). Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry*, *56*(2), 352–358. https://doi.org/10.1021/acs.biochem.6b01270

Shine, J., & Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(4), 1342–1346. https://doi.org/10.1073/pnas.71.4.1342

Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., & Schneider, T. D. (2001). Anatomy of Escherichia coli ribosome binding sites. *Journal of Molecular Biology*, *313*(1), 215–228. https://doi.org/10.1006/jmbi.2001.5040

Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B. P., & Yusupov, M. (2009). A structural view of translation initiation in bacteria. *Cellular and Molecular Life Sciences*, *66*(3), 423–436. https://doi.org/10.1007/s00018-008-8416-4

Skalenko, K. S., Li, L., Zhang, Y., Vvedenskaya, I. O., Winkelman, J. T., Cope, A. L., Taylor, D. M., Shah, P., Ebright, R. H., Kinney, J. B., Zhang, Y., & Nickels, B. E. (2021). Promoter-sequence determinants and structural basis of primer-dependent transcription initiation in Escherichia coli. *Proceedings of the National*

Academy of Sciences of the United States of America, *118*(27), e2106388118. https://doi.org/10.1073/pnas.2106388118

Skordalakes, E., & Berger, J. M. (2003). Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell*, *114*(1), 135–146. https://doi.org/10.1016/s0092-8674(03)00512-9

Sørensen, M. A., Fricke, J., & Pedersen, S. (1998). Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in Escherichia coli in vivo11Edited by D. Draper. *Journal of Molecular Biology*, *280*(4), 561–569. https://doi.org/10.1006/jmbi.1998.1909

Stead, M. B., Marshburn, S., Mohanty, B. K., Mitra, J., Castillo, L. P., Ray, D., van Bakel, H., Hughes, T. R., & Kushner, S. R. (2011). Analysis of Escherichia coli RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Research*, *39*(8), 3188–3203. https://doi.org/10.1093/nar/gkq1242

Studer, S. M., & Joseph, S. (2006). Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Molecular Cell*, *22*(1), 105–115. https://doi.org/10.1016/j.molcel.2006.02.014

Subramanian, A.-R. (1983). Structure qnd Functions of Ribosomal Protein S1. In W. E. Cohn (Ed.), *Progress in Nucleic Acid Research and Molecular Biology* (Vol. 28, pp. 101–142). Academic Press. https://doi.org/10.1016/S0079-6603(08)60085-9

Sun, Z., Cagliero, C., Izard, J., Chen, Y., Zhou, Y. N., Heinz, W. F., Schneider, T. D., & Jin, D. J. (2019). Density of σ70 promoter-like sites in the intergenic regions dictates the redistribution of RNA polymerase during osmotic stress in Escherichia coli. *Nucleic Acids Research*, *47*(8), 3970–3985. https://doi.org/10.1093/nar/gkz159

Sutherland, C., & Murakami, K. S. (2018). An Introduction to the Structure and Function of the Catalytic Core Enzyme of *Escherichia coli* RNA Polymerase. *EcoSal Plus*, *8*(1), ecosalplus.ESP-0004-2018. https://doi.org/10.1128/ecosalplus.ESP-0004-2018

Takata, R., Mukai, T., & Hori, K. (1987). RNA processing by RNase III is involved in the synthesis of Escherichia coli polynucleotide phosphorylase. *Molecular and General Genetics MGG*, *209*(1), 28–32. https://doi.org/10.1007/BF00329832

Takyar, S., Hickerson, R. P., & Noller, H. F. (2005). MRNA helicase activity of the ribosome. *Cell*, *120*(1), 49–58. https://doi.org/10.1016/j.cell.2004.11.042

Tejada-Arranz, A., de Crécy-Lagard, V., & de Reuse, H. (2020). Bacterial RNA Degradosomes: Molecular Machines under Tight Control. *Trends in Biochemical Sciences*, *45*(1), 42–57. https://doi.org/10.1016/j.tibs.2019.10.002

Trubetskoy, D., Proux, F., Allemand, F., Dreyfus, M., & Iost, I. (2009). SrmB, a DEAD-box helicase involved in Escherichia coli ribosome assembly, is specifically targeted to 23S rRNA in vivo. *Nucleic Acids Research*, *37*(19), 6540–6549. https://doi.org/10.1093/nar/gkp685

Tsai, Y.-C., Du, D., Domínguez-Malfavón, L., Dimastrogiovanni, D., Cross, J., Callaghan, A. J., García-Mena, J., & Luisi, B. F. (2012). Recognition of the 70S ribosome and polysome by the RNA degradosome in Escherichia coli. *Nucleic Acids Research*, *40*(20), 10417–10431. https://doi.org/10.1093/nar/gks739

Tucker, B. J., & Breaker, R. R. (2005). Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology*, *15*(3), 342–348. https://doi.org/10.1016/j.sbi.2005.05.003

Tzareva, N. V., Makhno, V. I., & Boni, I. V. (1994). Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions. *FEBS Letters*, *337*(2), 189–194. https://doi.org/10.1016/0014-5793(94)80271-8

Umitsuki, G., Wachi, M., Takada, A., Hikichi, T., & Nagai, K. (2001). Involvement of RNase G in *in vivo* mRNA metabolism in *Escherichia coli*: Role of RNase G in mRNA metabolism. *Genes to Cells*, *6*(5), 403–410. https://doi.org/10.1046/j.1365-2443.2001.00430.x

Vakulskas, C. A., Potts, A. H., Babitzke, P., Ahmer, B. M. M., & Romeo, T. (2015). Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiology and Molecular Biology Reviews: MMBR*, *79*(2), 193–224. https://doi.org/10.1128/MMBR.00052-14

Valentin-Hansen, P., Eriksen, M., & Udesen, C. (2004). The bacterial Sm-like protein Hfq: A key player in RNA transactions: Hfq, an Sm-related protein. *Molecular Microbiology*, *51*(6), 1525–1533. https://doi.org/10.1111/j.1365-2958.2003.03935.x

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., & Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, *448*(7150), 163–168. https://doi.org/10.1038/nature05931

Villa, J. K., Su, Y., Contreras, L. M., & Hammond, M. C. (2018). SYNTHETIC BIOLOGY OF SMALL RNAS AND RIBOSWITCHES. *Microbiology Spectrum*, *6*(3), 10.1128/microbiolspec.RWR-0007–2017. https://doi.org/10.1128/microbiolspec.RWR-0007-2017

Vimberg, V., Tats, A., Remm, M., & Tenson, T. (2007). Translation initiation region sequence preferences in Escherichia coli. *BMC Molecular Biology*, *8*(1), 100. https://doi.org/10.1186/1471-2199-8-100

Vogel, J., & Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nature Reviews. Microbiology*, *9*(8), 578–589. https://doi.org/10.1038/nrmicro2615

Vvedenskaya, I. O., Sharp, J. S., Goldman, S. R., Kanabar, P. N., Livny, J., Dove, S. L., & Nickels, B. E. (2012). Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. *Genes & Development*, *26*(13), 1498–1507. https://doi.org/10.1101/gad.192732.112

Vytvytska, O., Moll, I., Kaberdin, V. R., von Gabain, A., & Bläsi, U. (2000). Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes & Development*, *14*(9), 1109–1118. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC316587/

Wade, J. T., & Struhl, K. (2008). The transition from transcriptional initiation to elongation. *Current Opinion in Genetics & Development*, *18*(2), 130–136. https://doi.org/10.1016/j.gde.2007.12.008

Wakabayashi, H., Warnasooriya, C., & Ermolenko, D. N. (2020). *Extending the spacing between the Shine-Dalgarno sequence and P-site codon reduces the rate of mRNA translocation* [Preprint]. Biochemistry. https://doi.org/10.1101/2020.04.16.045807

Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R. C., Snyder, M. P., Segal, E., & Chang, H. Y. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, *505*(7485), 706–709. https://doi.org/10.1038/nature12946

Weixlbaumer, A., Leon, K., Landick, R., & Darst, S. A. (2013). Structural Basis of Transcriptional Pausing in Bacteria. *Cell*, *152*(3), 431–441. https://doi.org/10.1016/j.cell.2012.12.020

Weyens, G., Charlier, D., Roovers, M., Piérard, A., & Glansdorff, N. (1988). On the role of the Shine-Dalgarno sequence in determining the efficiency of translation initiation at a weak start codon in the car operon of Escherichia coli K12. *Journal of Molecular Biology*, *204*(4), 1045–1048. https://doi.org/10.1016/0022-2836(88)90061-7

Wieden, H.-J., Gromadski, K., Rodnin, D., & Rodnina, M. V. (2002). Mechanism of elongation factor (EF)-Ts-catalyzed nucleotide exchange in EF-Tu. Contribution of contacts at the guanine base. *The Journal of Biological Chemistry*, *277*(8), 6032–6036. https://doi.org/10.1074/jbc.M110888200

Worrall, J. A. R., Górna, M., Crump, N. T., Phillips, L. G., Tuck, A. C., Price, A. J., Bavro, V. N., & Luisi, B. F. (2008). Reconstitution and Analysis of the Multienzyme Escherichia coli RNA Degradosome. *Journal of Molecular Biology*, *382*(4), 870–883. https://doi.org/10.1016/j.jmb.2008.07.059

Xu, F., & Cohen, S. N. (1995). RNA degradation in Escherichia coli regulated by 3' adenylation and 5' phosphorylation. *Nature*, *374*(6518), 180–183. https://doi.org/10.1038/374180a0

Yamaguchi, Y., Park, J.-H., & Inouye, M. (2009). MqsR, a Crucial Regulator for Quorum Sensing and Biofilm Formation, Is a GCU-specific mRNA Interferase in Escherichia coli. *Journal of Biological Chemistry*, *284*(42), 28746–28753. https://doi.org/10.1074/jbc.M109.032904

Yehudai-Resheff, S., & Schuster, G. (2000). Characterization of the *E. coli* poly(A) polymerase: Nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Research*, *28*(5), 1139–1144. https://doi.org/10.1093/nar/28.5.1139

Yoshida, M., Meksuriyen, D., Kashiwagi, K., Kawai, G., & Igarashi, K. (1999). Polyamine stimulation of the synthesis of oligopeptide-binding protein (OppA). Involvement of a structural change of the Shine-Dalgarno sequence and the initiation codon aug in oppa mRNA. *The Journal of Biological Chemistry*, *274*(32), 22723–22728. https://doi.org/10.1074/jbc.274.32.22723

Yusupova, G., Jenner, L., Rees, B., Moras, D., & Yusupov, M. (2006). Structural basis for messenger RNA movement on the ribosome. *Nature*, *444*(7117), 391–394. https://doi.org/10.1038/nature05281

Zhang, A., Schu, D. J., Tjaden, B. C., Storz, G., & Gottesman, S. (2013). Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets. *Journal of Molecular Biology*, *425*(19), 3678–3697. https://doi.org/10.1016/j.jmb.2013.01.006

Zhang, J., Palangat, M., & Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. *Nature Structural & Molecular Biology*, *17*(1), 99–104. https://doi.org/10.1038/nsmb.1732

Zhang, X., Zhu, L., & Deutscher, M. P. (1998). Oligoribonuclease Is Encoded by a Highly Conserved Gene in the 3'-5' Exonuclease Superfamily. *Journal of Bacteriology*, *180*(10), 2779–2781. https://doi.org/10.1128/JB.180.10.2779-2781.1998

Zhang, Y., Zhang, J., Hara, H., Kato, I., & Inouye, M. (2005). Insights into the mRNA Cleavage Mechanism by MazF, an mRNA Interferase. *Journal of Biological Chemistry*, *280*(5), 3143–3150. https://doi.org/10.1074/jbc.M411811200

Zilhão, R., Cairrão, F., Régnier, P., & Arraiano, C. M. (1996). PNPase modulates RNase II expression in Escherichia coli: Implications for mRNA decay and cell metabolism. *Molecular Microbiology*, *20*(5), 1033–1042. https://doi.org/10.1111/j.1365-2958.1996.tb02544.x

CHAPTER 1

Study of 5'UTR-mediated regulation of translation initiation, transcription and/or mRNA stability

Summary

To better understand the role of 5'UTRs in gene expression regulation in *E. coli*, we generated a set of synthetic 5'UTRs and analyzed their contribution in regulating translation, transcription, and mRNA stability of different reporter genes. Forty-one synthesized 5'UTRs, homogenous in size but with different theoretical translation initiation rates (corresponding to a wide range of RBS index calculated by the RBS Calculator algorithm), were placed under the transcriptional control of the same inducible promoter to control the expression of the reporter gene *lacZ*. At the protein level, we observed large changes in β-galactosidase activity between strains. Globally, high protein levels were associated with 5'UTRs of high RBS index. These results confirmed that translation initiation is a regulatory factor of protein level. However, the theoretical translation initiation rates did not correlate exactly with the measured protein level. This suggests that the RBS index is only an estimator of the translation initiation rate as it does not take into account additional constraints linked for instance to mRNA synthesis (transcription) and transcript half-life time (stability). At the mRNA level, since the same concentration of transcription inducer was applied to all constructions, it was expected to obtain the same mRNA concentrations between strains. Nevertheless, we measured a large variation in mRNA concentrations. These differences suggest that 5'UTRs modify either transcription and/or mRNA stability. We validated that these conclusions were still true for two other reporter genes when replacing the *lacZ* gene with the *txAbF* gene (encoding an α-L-arabinofuranosidase) and the *msfGFP* gene (encoding a fluorescent protein) on a selection of eight representative synthetic 5'UTRs. To quantify the contribution of regulation of transcript stability in mRNA concentration variation, we measured the mRNA half-life time for each combination of the three reporter genes and the 8 5'UTRs. Analysis of changes in mRNA half-life showed that 5'UTRs strongly affected mRNA stability and that this effect was dependent on the downstream gene coding sequence. Using the regulatory coefficient theory, which correlates variation in mRNA concentration to variation in transcription and/or transcript stability, we showed that mRNA concentration of the reported gene could be controlled mainly by transcription or stability regulation, or by a shared control between transcription and degradation depending on the 5'UTR–reporter combination. In conclusion, this study confirms the multilevel contribution of 5'UTRs in the control of gene expression. We have identified the regulations mediated by the 5'UTRs at the level of translation initiation, transcription, and/or mRNA stability and showed their degree of dependence on the reporter gene sequence.

Introduction

When facing changing environments, *E. coli* cells have to adapt their metabolism by modifying the expression pattern of their gene repertoire. The control of gene expression encompasses a wide range of regulatory mechanisms that occur at each step of gene expression, from transcription initiation to post-translational modifications of proteins. mRNA is a central molecule of gene expression control. The 5'UTR, untranslated transcribed region of the mRNA emerged as potential "regulatory-hubs" for translation, mRNA stability, and synthesis (transcription).

As well known, 5'UTR is crucial for efficient translation. It is generally considered that the important features are the SD sequence (Shine & Dalgarno, 1974) and A/U richness, which helps to minimize the formation of strong secondary structures of ribosome binding site (Berg et al., 2012; Kozak, 2005; Simonetti et al., 2009). A well-positioned SD/Anti-SD pairing and reduced secondary structure in sequences flanking the start codon and SD are the hallmarks of highly expressed genes *in E. coli* (Abolbaghaei et al., 2017; Prabhakaran et al., 2015). In addition, the regions upstream of the SD sequence and the distance between the SD sequence with the initiation codon are also critical, they can be considered as the "reinforce elements" (Abolbaghaei et al., 2017; Viegas et al., 2018; Wakabayashi et al., 2020). Many studies in bacteria have also shown that 5'UTR associated diverse "signals" to tune protein expression levels, such as the presence of ligands, regulatory proteins, or small RNAs. They bind to relevant sites on the 5'UTR, e.g. riboswitch (Kang et al., 2014; Kirchner et al., 2017), and block or facilitate ribosome recruitment to mRNA (Malina et al., 2005; Storz et al., 2004). In bacteria, the two prominent examples of such global RBPs are CsrA and Hfq (Chiaruttini & Guillier, 2020; Romeo et al., 1993) involved in post-transcriptional regulation of up to 25% and 20% of all mRNAs in *E. coli,* respectively (Holmqvist & Vogel, 2018).

Moreover, numerous studies have also shown that 5'UTRs contribute to transcript stability. Some researchers suggested that the effect of 5'UTR on translation indirectly alters the stability of mRNA (Iost & Dreyfus, 1995; Komarova et al., 2005). High ribosomal occupancy protects the mRNA from rapid degradation. In particular, mutations in the ribosome binding site (RBS) which reduce translation initiation efficiency, and hence the efficiency of the overall translation, accelerate mRNA degradation (Iost & Dreyfus, 1995; McCormick et al., 1994; Yarchuk et al., 1991, 1992). The first cleavage of mRNA by endoribonucleases is generally considered as the rate-limiting step of mRNA turnover, and therefore elements such as stem-loop structures in the 5'UTR are considered to be important for mRNA stability by protecting from RNase entry at the 5' end (Berg et al., 2012; Condon, 2007; Picard et al., 2009).

Furthermore, it has been shown that 5'UTRs can also influence transcription. The initially transcribed sequence (ITS) corresponding to the 20 first nucleotides of the 5'UTR was reported to

affect transcription initiation (Heyduk & Heyduk, 2018). Transcription initiation is a multistep process that begins when RNAP recognizes and binds to DNA elements within a promoter sequence and ends when RNAP escapes from the promoter and continues through elongation. Using a library of ITS variants, ITS was shown to increase or slow RNAP promoter escape and thus to participate to the level of transcription (Heyduk & Heyduk, 2018). The composition (especially positions +1 and +2) and structuration of the first 10 nucleotides of ITS influence the most the promoter escape. Also, sequences similar to the −10 promoter element (TATAAT) within the 5'UTRs have been demonstrated to induce a σ70-dependent transcription pause after the promoter escape (Brodolin et al., 2004; Hatoum & Roberts, 2008; Nickels et al., 2004), and thus reduced transcription.

To better understand the role of 5'UTRs in gene expression regulation in *E. coli*, we investigated their contribution in regulating translation, transcription, and mRNA stability, respectively. We generated a set of synthetic 5'UTRs covering a large range of theoretical translation initiation rates fused to three different reporter genes. We confirmed for all reporters the control of gene expression by 5'UTR-dependent regulation of translation initiation. In addition, we showed that 5'UTRs were also involved in the regulation of the level of transcription and/or the stability of the transcript in a manner dependent on the sequence of the downstream reporter gene. Overall, these results suggest that the 5'UTRs are capable of acting as the "regulatory-hubs" of gene expression.

Materials and Method

1. Ribosome Binding Site calculators

The software RBS Calculator (http://voigtlab.ucsf.edu/software) was used for synthetic 5'UTRs design. RBS Calculator uses a statistical thermodynamic model considering Gibbs free energies for key molecular interactions in translation initiation to give an estimation of translation initiation rate. RBS Calculator can be used in two ways: first, to predict the translation initiation rate of each start codon on an mRNA sequence (reverse engineering); second, to design synthetic 5'UTR sequences (containing a RBS) to rationally control the translation initiation rate (forward engineering). We used this second function for the design of synthetic 5'UTRs. RBS Calculator designs 5'UTRs with theoretical RBS index from 0.1 to 100 000 arbitrary units. A second software, UTR Designer (https://sbi.postech.ac.kr/utr_designer) has been developed for similar purposes. UTR Designer uses a very similar model as RBS Calculator, considering the changes in Gibbs free energy before and after 30S ribosome binding to mRNA transcript. In addition, UTR Designer subdivides the state transition of the 30S binds to mRNA into two scenarios, i.e., the 30S ribosome can either locate directly to the transcription initiation region (TIR) on the mRNA or slides into the TIR through a multi-step process. UTR Designer calculates the RBS index from 1 to 1 000 000 arbitrary units.

2. Design of synthetic 5'UTRs

We designed a large set of synthetic 5'UTRs to be compared with a reference 5'UTR. The reference 5'UTR is originated from the pBAD-*lacZ* control plasmid from the Invitrogen expression system pBAD-his/myc used for the production of heterologous proteins in *E. coli*. This reference 5'UTR is 33 nt long and has an efficient SD sequence (AGGAGG) and an RBS index of 33 000. To cover the large range of RBS indexes (0.1 to 100 000 arbitrary units) designed by RBS Calculator, we selected 10 targeted RBS indexes (100, 250, 500, 1000, 1500, 2500, 5000, 33 000, 66 000 and 100 000 arb. unit). The design of 5'UTR was performed considering the 150 first nucleotides of the *lacZ* coding sequence. For each target RBS index, 200 unique 5'UTR sequences of 33 nt long were generated by RBS Calculator. In each RBS index class, 4 sequences were manually selected, two with high GC% and two with low GC%, without strong folding within the 5'UTR sequence or at the beginning of the *lacZ* coding sequence. Including the reference, this led to a set of 41 synthetic 5'UTR sequences covering a wide range of RBS indexes (***Table S1***).

3. Plasmid construction and reporter replacement

Synthetic 5'UTRs were introduced in place of the reference 5'UTR in the pBAD-*lacZ* plasmid by PCR. For each 5'UTR, a pair of primers was designed, which contains half of the sequence of the 5'UTR to be cloned and extended with a sequence able to hybridize with the recipient plasmid. The full plasmid was amplified by PCR using *Phusion* polymerase (New England Biolabs). Amplicons were gel-purified, 5'ends were phosphorylated with T4 polynucleotide kinase (30 min, 37°C, New England Biolabs), and self-ligated with T4 DNA ligase (overnight, 16°C). The ligation mix was used to transform *E. coli* DH5α (New England Biolabs). Primers used for 5'UTR cloning are listed in ***Table S2***. Plasmid DNA was isolated using the QIAprep Miniprep kit (Qiagen) and constructs were verified by sequencing (Eurofins).

For gene reporter replacement, the CDS from *msfGFP* (encoding the monomeric super-folder green fluorescent protein) and *txAbF* (encoding an α-L-arabinofuranosidase) were PCR amplified. The whole plasmid backbones containing the 8 selected 5'UTRs fused to *lacZ* were PCR amplified from the end of the *lacZ* coding sequence to the 5'UTR. Amplicons were gel purified, the *msfGFP* or *txAbF* CDS inserts were phosphorylated, ligated with the backbones, and transferred in *E. coli* DH5α. Correct fusions between the 5'UTRs and the reporters were verified by sequencing. The primers used are listed in ***Table S2***.

4. Bacterial strains, growth and induction conditions

All cloning steps were performed using NEB 5-α (New England Biolabs) *E. coli* strain and cultures were grown in LB or LB-agar. For characterization, all constructs were transferred in our reference strain (MET 346), a derivative strain of DLT 2202 (MG 1655 *ΔaraFGH*, *Ωpcp18::araE533*) in which the chromosomal copy of *lacZ* was deleted (Ah-Seng et al., 2013; Nouaille et al., 2017). In this strain, the pBAD promoter is proportionally induced by the concentration of arabinose, without heterogeneity between cells for the induction level.

All the strains were routinely grown in M9 minimal medium (Esquerré et al., 2014) supplemented with ampicillin 100 µg/ml at 37°C, 150 rpm except for other statements. Cultures were inoculated from overnight cultures at an initial OD_{600nm} of 0.1.

For evaluation of induction levels, arabinose was added in exponentially growing cultures (OD_{600}=0.6) with serial dilutions for final concentration from 0.00001% to 0.1% arabinose. For physiological characterizations, strains carrying *lacZ* and *txAbF* were induced with 0.001% arabinose, and strains

carrying *msfGFP* induced with 0.01% arabinose. Samplings were performed 30 min after arabinose induction.

5. Protein level determination

Measurements of β-galactosidase and α-L-arabinofuranosidase activity. Strains carrying the lacZ gene were grown as described above. 3 mg of cells (dry weight) were collected and immediately kept on ice. Cells were harvested and washed twice with cold 0.2% KCl, resuspended in 1 ml of breaking buffer (15 mM Tris tris 400 mM/ tricarballylate 1, 4.5% Glycerol, 0.9 mM MgCl2, 0.2 mM DTT; pH=7.2), transferred in screw capped tube containing 0.1 g glass beads. Cells were disrupted with FAST PREP by 6 cycles (6.5 m/s, 30 s) with 1 min on ice between each cycle. After centrifugation (13200 rpm, 15 min), supernatants containing soluble proteins were used for quantification. All measurements were carried out on 3 biological and technical replicates.

Total protein was assayed using the Bradford method with bovine serum albumin as the protein standard (Bradford, 1976).

The β-Galactosidase activity was determined by the colorimetric method using o-nitrophenyl-β-D-galactopyranoside (ONPG) as a substrate (Held, 2007).

The α-L-arabinofuranosidase activity assay was modified from the colorimetric method of the discontinuous assay (Bissaro et al., 2014). The determination uses 4-Nitrophenyl-α-L-arabinofuranoside (pNP-Araf, colorless), which is hydrolyzed by α-L-arabinofuranosidase to release p-nitrophenyl (pNP, yellow). 100 µL samples diluted in sodium phosphate (100 mM) were placed in the 96-well microplate. The assay was initiated by the addition of 100 µl of 2X assay buffer. Assay buffer (1X) consists of 100 mM sodium phosphate, pH=7; 1 mM $MgCl_2$; 50 mM β-mercaptoethanol and 10 mM pNP-Araf. All absorbance determinations were performed at 401 nm using a microplate reader (SpectraMax Plus 384). After the assay was initiated by the addition of the 2X assay buffer, kinetic readings were initiated immediately with absorbance determinations made every 20 seconds for a total of 30 minutes at 45°C. The unit of α-L-arabinofuranosidase activity is defined by the catalysis of 1 µmol of pNP per minute.

Fluorescence measurements. Specific fluorescence was defined as the relative fluorescence unit divided by the corresponding OD_{600nm} (GFP/OD_{600}) and determined by the end-point quantification method. Strains carrying the *msfGFP* gene were grown and inducted as described above. 100 µL

cultures were serially diluted 1:2 with cold M9 medium four times on a microplate (final volume 300 µL) on ice, and both OD_{600nm} and fluorescence were quantified. Fluorescence and cell density (OD600) were measured on Bio TeK Synergy H1 microplate reader (BioTek, Winooski, USA). The reference strain MET 346 (5'UTR_33k_31) was used as the negative control and its fluorescence intensity was subtracted as background. The excitation and emission wavelengths of GFP are 475 and 510 nm. Measurements were carried out on 3 biological and technical replicates.

6. mRNA concentration and half-life determination

Sampling. Cultures were grown and transcription induced as described above for 30 min. 3 mg of cells (dry weight) were harvested and flash frozen in liquid nitrogen. These samples are reference points (T0) for the half-life determination procedure. To arrest transcription initiation, rifampicin (500 mg/l) was added and cells (3 mg dry weight) were harvested at 0.3, 0.6, 1, 1.5, 2, 3, 4, 5, 6, 7.5, 10, 12.5 and 15 min after and flash frozen in liquid nitrogen. Samples were stored at −80°C until mRNA extraction. For stability measurements, samples corresponding to T0 plus 5 points after rifampicin addition were extracted.

mRNA extraction and quality control. All mRNA extractions were performed using the RNeasy Mini extraction kit (Qiagen) following the recommendations of the manufacturer. Briefly, after thawing on ice, the culture sample was centrifuged (10 min, 5000 rmp, 4°C). The cell pellet was resuspended in 500 µl of buffer RLT supplemented with 0.01% β-mercaptoethanol (Sigma) and the mixture was transferred into a tube containing 0.1 glass beads. Cells were disrupted at 4°C by 3 cycles of 30 s with a FastPrep-24 instrument (MP Biomedicals) and centrifuged for 10 min at 13200 g at 4°C. Total RNA concentration was quantified a Nanodrop spectrophotometer (Thermo Fisher Scientific). Potential DNA contamination co-purified with RNA, was eliminated using the TURBO DNA-freeTM Kit (Ambion) on 50 µg total RNA in a final volume of 100 µL according to the manufacturer's recommendations. RNAs were quantified using a Nanodrop spectrophotometer and their integrity was certified with Bioanalyzer 2100 (Agilent) with the RNA 6000 Nano LabChip kit (Agilent). RNAs were stored at −80°C until required.

cDNA synthesis. 5 µg of total RNA were mixed with 1 µl of random primers (500 ng/µl; Life Technologies), and sterile water to a final volume of 24 µl. The mixture was heated at 70°C for 5 min and then immediately cooled to 4°C. Then 300 units SuperScript II reverse transcriptase were added with 0.5 M DTT, 15 mM dNTP Mix and 5x first strand buffer (Life Technologies) in a total volume of 50 µl. After 10 min incubation at 25°C, reverse transcription was performed at 42°C for 1h followed by an

inactivation step at 70°C for 15 min. The RNA-cDNA hybrids were then degraded by addition of 1 µl RNase H (2 units, 20 min, 37°C; Life Technologies). cDNAs were further purified via Illustra Microspin G-25 columns (GE Healthcare) according to the manufacturer's instructions.

Primer design and validation. Primers for qPCR were designed using Vector NTI advance v11 (Life Technologies) with as constraints a melting temperature of 59-61°C, a length of 20-22 bp and 50-67% GC content and leading to amplicon sizes ranging from 75 to 148 bp. The reaction efficiency of each pair of primers was tested as a single amplicon on serial dilutions of *lacZ*, *txAbF*, *msfGFP* containing plasmid as a matrix, depending on the primer pairs analyzed. Primer pairs were validated if the PCR was quantified between 90 and 110% efficiency over the dilution range tested. For each reporter gene (*lacZ, txAbF,* and *msfGFP*), three primer pairs were designed and distributed equally along the CDSs (beginning, middle, and end of the sequence). Housekeeping *ihfB* gene (integration host factor β-subunit) was used as the internal normalization control.

Real-time quantitative PCR. The cDNAs were 1:10 serially diluted and the qRT-PCR was performed using SYBR Green Supermix (Bio-Rad) on a LightCycler 480 II thermal cycler system (Roche, Mannheim, Germany). qRT-PCR was performed using the following temperature program: 5 min at 95°C for pre-incubation; then 45 PCR cycles of 10 s at 95°C for denaturation, 10 s at 60°C for annealing, and 10 s at 72°C for elongation. The melting curve analysis consisted of 1 min at 65°C followed by heating to 95°C with a ramp rate of 1°C/9s. For mRNA stability measurements, the qRT-PCR was done using Fluidigm High-throughput method (Biomark) at the Gentiane platform (Clermont Ferrand, INRAE). To determine the mRNA concentration and half-lives, 3 technical replicates and 3 biological replicates were performed for each primer pair.

7. Data analysis and statistical treatment

mRNA concentration and half-life. For mRNA concentration quantifications by qRT-PCR, the Pfaffl analysis method was applied (Pfaffl, 2001). After normalization by fold change of the *ihfB* normalization gene, results were expressed as differences (n-fold) between the tested strain and the reference strain (MET 346, 5'UTR_33k_31). Results are expressed as means of n-folds with a standard deviation of the mean.

mRNA half-life was determined by qRT-PCR in different conditions (*lacZ*, *txAbF*, or *msfGFP* fused with 8 selected 5'UTRs respectively) with one biological replicate and 3 technical replicates of three primer pairs. The cycle threshold (Ct) values were plotted as a function of time after rifampicin addition. Hereafter, the linear regression coefficient (k) of Ct versus time were calculated for each

mRNA species. Since the Ct values were very sensitive to small changes in concentration, we estimated that it was not possible to accurately estimate any delay in transcript degradation after rifampicin addition. Therefore, the mRNA half-life ($t_{1/2}$) was calculated from the linear regression coefficient k corresponding to the degradation rate constant as a function of time with the relation $t_{1/2} = 1/k$. Only slopes with R^2>0.9 were considered.

Degradational regulation coefficient. The degradational regulation coefficient (*ρD*) is defined as the contribution of stability regulation in the control of an mRNA concentration (Esquerré et al., 2014). The calculation of *ρD* between two strains with different mRNA concentrations shows if the variation of mRNA concentration is due to a modification of stability. Assuming that a steady state was established, the *ρD* between two strains (the strain of interest and the reference strain) can be calculated as the negative value of the slope of the double-logarithmic plot of the degradation rate constant *k* against the initial mRNA concentration (before rifampicin treatment) in the two compared strains. The function is as follows: $\rho D = -\frac{dlnk}{dln[mRNA]}$ where *dlnk* and *dln[mRNA]* are respectively the variations of the logarithmic degradation rate constant *k* and the mRNA concentration between the strain of interest and the reference strain. Three classes of regulatory control of the mRNA concentration have been defined depending on the *ρD* value: *ρD*<0.4, class I, low *ρD* value indicates that stability regulation does not significantly contribute to variation in mRNA concentration. In this case, variation of mRNA concentration is mainly due to regulation of transcription so class I corresponds to a mainly transcriptional control of the mRNA concentration; 0.4<*ρD*<0.6, class II, stability regulation contributes to variation in mRNA concentration but at the similar level than transcription regulation. Class II corresponds to a shared control of the mRNA concentration between transcription and degradation; *ρD*>0.6, class III, high *ρD* value indicates that stability regulation is the primary contribution to variation in mRNA concentration. Therefore, class III corresponds to a mainly degradational control of the mRNA concentration.

Results

1. Analysis of 5'UTR effect on β-galactosidase protein level

We first determined the effect of the synthetic 5'UTRs fused to the *lacZ* gene on the production of β-galactosidase. To do this, we measured the specific β-galactosidase activity of the 41 strains developed in this work. Since all 41 constructs are identical except for the 5'UTR, one can assume that transcription and concentration of *lacZ* mRNA will be similar between strains and thus any changes in specific β-galactosidase activity between strains should be due to changes in translation. The quantification of the β-galactosidase activity for the 41 strains (***Figure 1***) showed many differences in protein production between strains. For instance, the specific β-galactosidase activity varies by a factor of 100 between the highest value (12.1±1.4 µmol/min/mg of protein) and one of the lowest value (0.18±0.03 µmol/min/mg of protein). In addition, some strains did not produce β-galactosidase at all, or at a level under the detection limits. The large differences in β-galactosidase expression between strains were expected by the different RBS indexes we imposed with the design. Globally, we can observe a tendency of increased protein expression with an increased RBS index. The top 10 5'UTR that mediated the β-galactosidase highly expressed (5'UTR_100k_41, 5'UTR_66k_37, 5'UTR_33k_32, 5'UTR_33k_31, 5'UTR_66k_36, 5'UTR_5k_25, 5'UTR_100k_40, 5'UTR_100k_39, 5'UTR_66k_34) generally had higher theoretical translation initiation rates (RBS index ≥ 5 000 arb. unit). However, we found that measured protein levels did not exactly correspond to RBS indexes. Some 5'UTRs with a high RBS index resulted in low levels of protein, and inversely. Furthermore, 5'UTRs with the same theoretical translation initiation rate resulted in different protein levels. For example, constructs with RBS indexes around 33 000 arb. unit 5'UTR_33k_32 (9.29±1.3 µmol/min/mg per protein) and 5'UTR_33k_26 (0.32±0.02 µmol/min/mg per protein) showed a difference in protein levels of about 25-fold. Therefore, these results indicate that the RBS indexes of the 5'UTRs that estimate the theoretical translation initiation rates, correlated only partially with the protein levels.

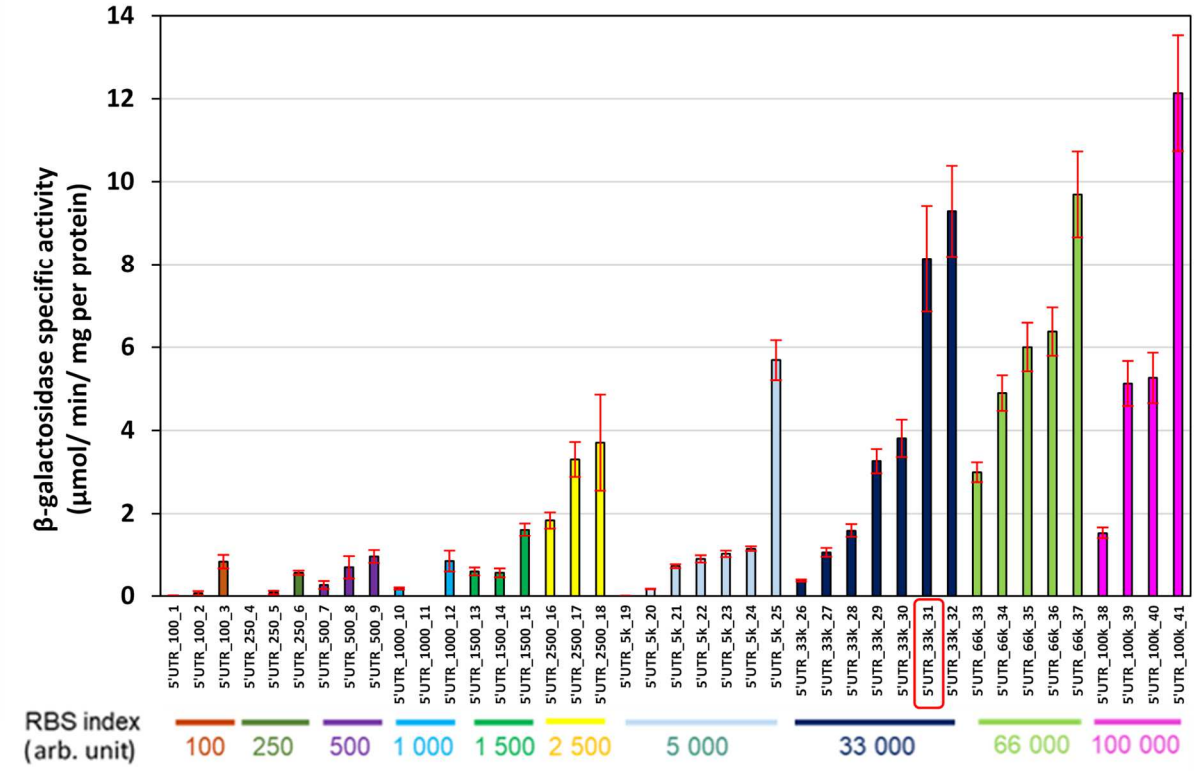


Figure 1. Activity of β-galactosidase of 41 synthetic 5'UTRs ranged by RBS index from 100-100 000 arb. unit. Each cluster of columns with the same color refers to a specific RBS index. The reference (5'UTR_33k_31) were highlighted with red box. Error bars represent standard deviation (n=3).

2. Prediction of the theoretical translation initiation rate

The fact that protein levels did not fully correlate with the RBS index estimated by RBS Calculator may be due to incorrect or inaccurate determination of the theoretical translation initiation rate by this software. To investigate this possibility, another translation initiation rate predicting software, UTR Designer was used. We plotted the experimentally measured β-galactosidase activity as a function of the RBS indexes of the 41 synthetic 5'UTRs given by the two software (***Fig 2***). Similar to RBS Calculator, the 5'UTRs predicted by UTR Designer to have high RBS indexes correlated with high protein expressions although the RBS indexes provided by the two software for one specific 5'UTR can be very different. For instance, for 5'UTR_100k_41 which corresponds to the highest experimentally measured β-galactosidase activity, RBS Calculator predicts a top-ranked RBS index, while UTR Designer predicts an RBS index ranked only 32. Globally, ***Figure 2*** shows that the correlation between predictions of RBS index and measurements of protein level is better with RBS Calculator than with UTR designer. We cannot exclude that if we had used a larger number of 5'UTRs with more differences in composition, such as length, GC%, potential secondary structures, the RBS index predictions given by either software could have been more correlated to the protein level.

However, these software estimate the theoretical translation initiation rate based on statistical thermodynamic models on mRNA molecule itself and in interactions with the ribosome but they do not consider other cellular parameters such as transcription, stability, and mRNA concentration. Numerous researches have reported that the 5'UTRs can influence mRNA stability, with an impact on mRNA concentration into the cell (Cetnar & Salis, 2021; Hui et al., 2014; Jia et al., 2020; Tuller & Zur, 2015). Because the 5'UTRs of the synthetic library were designed only based on the theoretical translation initiation rate, we cannot rule out the possibility that the modifications of the 5'UTR sequences also resulted in changes in mRNA levels in addition to changes in translation initiation.

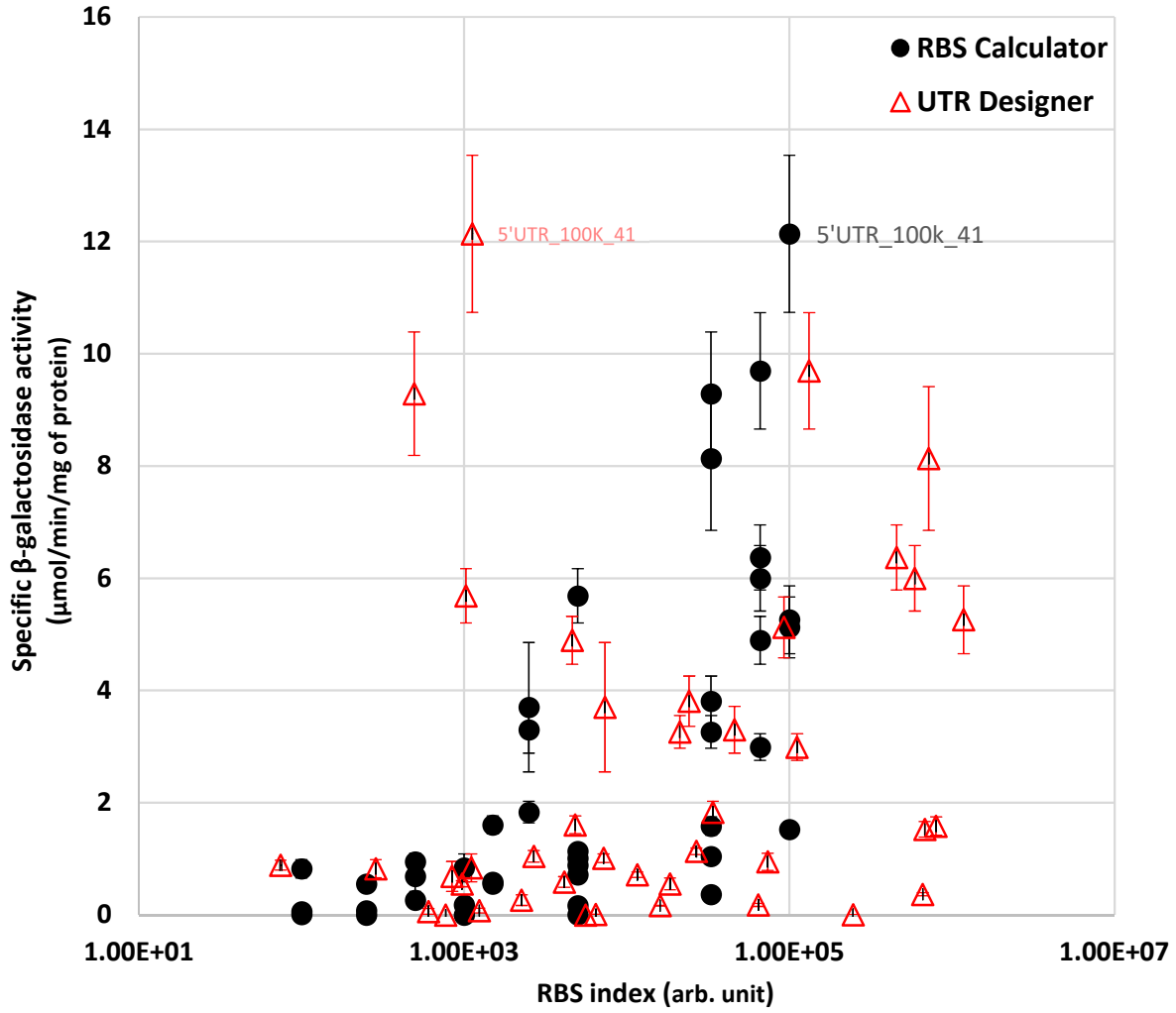


Figure 2. Correlation of the protein level of the 41 constructs with the RBS index. The experimental protein level (specific β-galactosidase activity) is plotted as a function of the theoretical translation initiation rate (RBS index) calculated using RBS Calculator (black circles) and UTR Designer (red triangles). Error bars represent standard deviation of β-galactosidase activity (n=3).

3. Effect of 5'UTRs on *lacZ* mRNA level and mRNA-protein correlation

To investigate whether the different 5'UTRs in our synthetic library could also lead to changes in mRNA levels, we measured the *lacZ* mRNA concentration of the 41 strains with different synthetic 5'UTRs (***Fig 3***). We observed a large variation in mRNA concentrations. For instance, the highest mRNA concentration (5'UTR_66k_34) was over 40-fold higher than one of the lowest (5'UTR_1500_13). Moreover, some mRNA concentrations (e.g. 5'UTR_250_4, 5'UTR_1000_10) were extremely low. It was not expected to observe such large variations in *lacZ* mRNA levels because all constructs used the same promoter induced with the same concentration of arabinose. These results allow us to conclude that 5'UTRs themselves can have a significant effect on the final mRNA concentration.

More precisely, nine 5'UTRs (5'UTR_5k_25, 5'UTR_33k_28, 5'UTR_33k_30, 5'UTR_33k_32, 5'UTR_66k_34, 5'UTR_66k_36, 5'UTR_66k_37, 5'UTR_100k_40, 5'UTR_100k_41) led to an increase in *lacZ* mRNA concentrations and 31 displayed a reduced concentration compared to the reference (5'UTR_33k_31). We can note that eight of the nine 5'UTRs with increased mRNA concentration have an RBS index similar to or above the RBS index of the reference 5'UTR (33k arb. unit) indicating a global trend of higher mRNA concentrations for higher RBS indexes. However, we did not find a clear correlation between mRNA concentration and RBS index. For instance, the 5'UTRs with RBS indexes around 33 000 arb. unit displayed a large range of *lacZ* mRNA concentrations (***Figure 3***). Within this group, up to 8-fold difference in mRNA concentration was observed (5'UTR_33k_32 vs 5'UTR_33k_26). When we took into account the corresponding β-galactosidase activities (***Figure 1***), we observed that within this group, 5'UTR_33K_32 had the highest mRNA concentration (2.27±0.46 fold) and also the highest protein level (9.29±1.3 µmol/min/mg of protein). But this relationship between mRNA level and protein level was not valid for all constructs: the mRNA concentration of 5'UTR_33k_29 (0.31±0.1 fold) was 25% that of 5'UTR_33k_28 (1.18±0.1 fold), but the protein level of 5'UTR_33k_29 (3.26±0.3 µmol/min/mg of protein) increased 2-fold compared to 5'UTR_33k_28 (1.58±0.2 µmol/min/mg of protein).

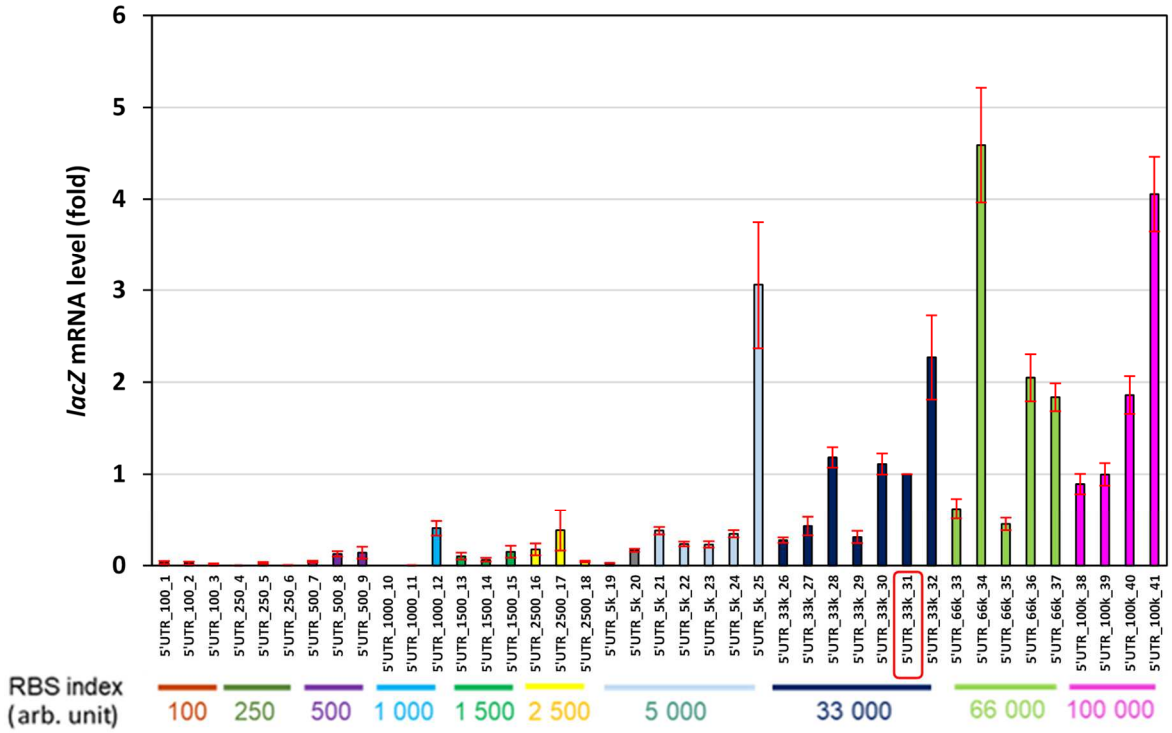


Figure 3. The *lacZ* mRNA concentration of 41 synthetic 5'UTRs ranged by RBS index from 100-100 000 arb. unit. The *lacZ* mRNA concentration measured in each strain is represented as Fold change compared to the reference (5'UTR_33k_31, highlighted with red box). Strains are clustered by RBS index (10 ranges from 100 to 100 000 arb. unit) with color code below the values. On the horizontal coordinate, the 41 constructs with different 5'UTRs are ranged according to the RBS index. Error bars represent standard deviation (n=3).

To provide a more comprehensive picture, we plotted the specific β-galactosidase activity versus the *lacZ* mRNA concentration for all the 41 constructs ***(Fig 4)***. We observed a moderate linear correlation between protein levels and *lacZ* mRNA concentrations (coefficient of determination R^2=0.47). A correlation was expected because the more concentrated an mRNA is, the more protein can be synthesized. However, the low level of correlation indicates that some variations in β-galactosidase activity cannot be attributed to changes in mRNA concentration. We can find constructs that have similar mRNA levels but different levels of protein. In this case, variation in protein levels can be associated with variations in the translation initiation mediated by the different 5'UTRs. On the other hand, some strains displayed variations in mRNA concentrations but a constant protein level. This situation could be explained by antagonist effects of variation in mRNA concentration and translation initiation efficiency.

In conclusion, analyses of variations in mRNA concentration and protein level show that 5'UTRs impact gene expression not only at the translational level but also at the level of mRNA concentration. The variations in mRNA concentration could be associated with modifications of either transcription or stability of the mRNA. This point has been investigated an presented in the next sections.

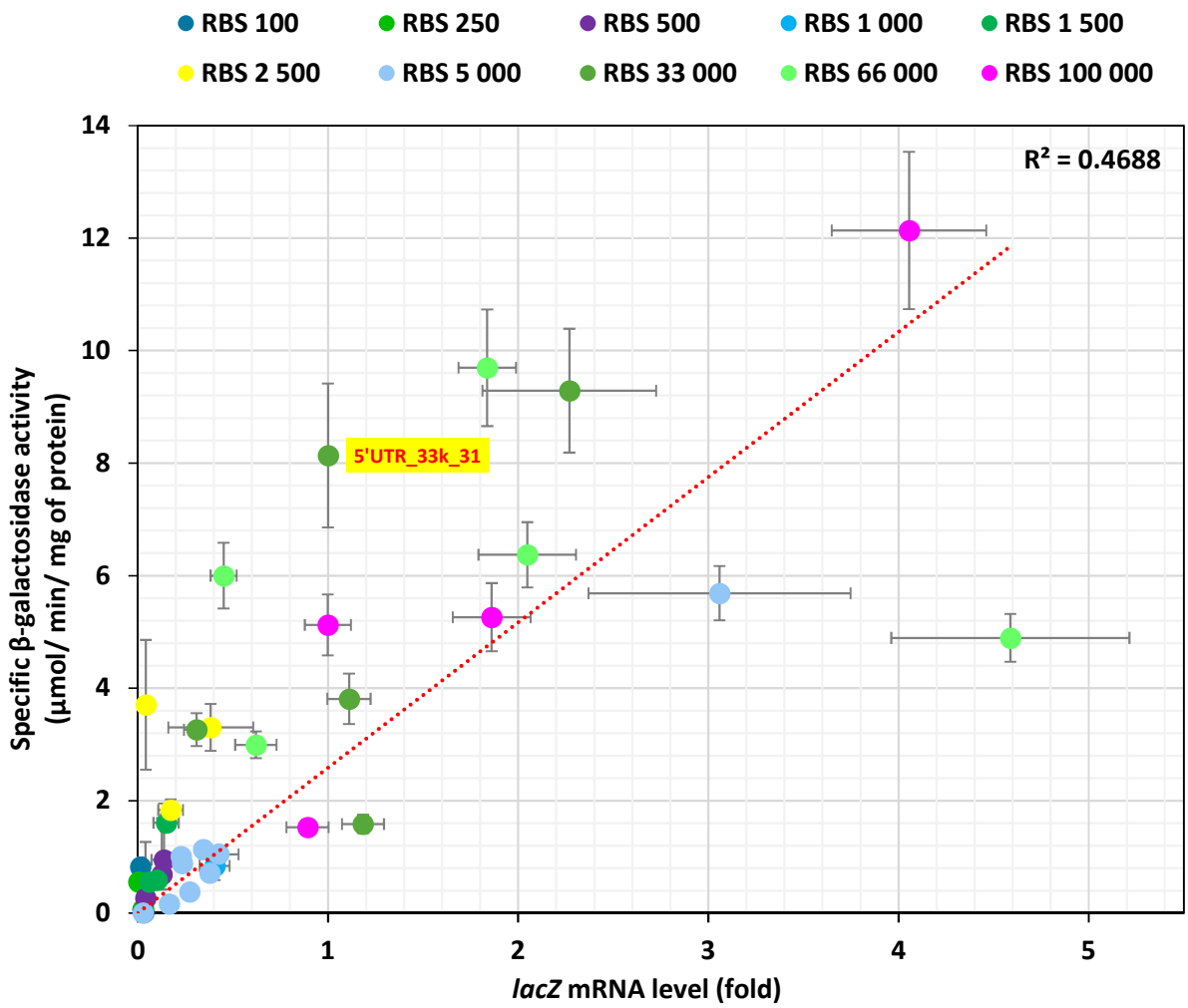


Figure 4. Correlation of specific β-galactosidase activity with *lacZ* mRNA concentration for all the 41 constructs. The mRNA concentration is expressed as Fold change compared to the reference strain (5'UTR_33k_31, highlighted in yellow). Circles of the same color represent constructs with the same RBS index given by RBS Calculator (the colors are the same as in figure 3). Red dashed line represents the regression line with the corresponding R^2 determination coefficient. Error bars represent standard deviations (n=3).

4. Effect of 5'UTRs on stability and transcription of *lacZ* mRNA

Since all 41 constructs are under control of the same promoter induced with the same level of inducer (0.001% arabinose), we expected that all constructs have the same transcription initiation. Thus, changes in *lacZ* mRNA concentrations between constructs are likely due to regulations related to the different 5'UTR sequences. Knowing that mRNA concentration in cells results from a balance between transcription and mRNA degradation, we focus our analyses on mRNA degradation to explore whether changes in *lacZ* mRNA stability could be responsible for changes in *lacZ* mRNA concentration observed when using different 5'UTRs. We selected 8 representative synthetic constructs among the 41 initially studied. They were selected to cover the ranges of variation in β-galactosidase activity and mRNA concentration and to be positioned near the linear correlation between protein level and mRNA concentration (R^2=0.91) (***Figure 5A***). We selected these 5'UTRs with a strong correlation between β-galactosidase expression and *lacZ* mRNA concentration because in this case, their role in regulating β-galactosidase expression is primarily in the regulation of mRNA concentration and much less in the regulation of translation initiation. The eight selected constructs are 5'UTR_500_07, 5'UTR_1500_14, 5'UTR_500_09, 5'UTR_1500_15, 5'UTR_2500_16, 5'UTR_33k_29, 5'UTR_66k_35, and 5'UTR_33K_31 renamed for better comprehension as by 5'UTR1 to 5'UTR8 respectively.

We measured *lacZ* mRNA stability by quantification of its concentration decay over time after blocking transcription initiation by rifampicin addition. We observed differences in *lacZ* mRNA half-life between the 8 constructs, with half-lives varying from 2.3±0.1 min to 0.4±0.1 min (***Figure 5C***). This 5-fold variation confirmed that 5'UTRs have an effect on the stability of the transcript to which they belong. Except for 5'UTR7 and 5'UTR8, the six other constructs displayed all similar and short half-lives of approximately 0.5 min. We can note that the two constructs with the highest half-lives had also the highest mRNA concentrations (***Figure 5B***). Between strains, some variations in mRNA concentration could be related to variations in its stability. For instance, the *lacZ* mRNA half-life of 5'UTR8 is 3.8 times higher than that of 5'UTR6, while its concentration is 3.2 times higher. This suggests that mRNA stability modification could explain variation in mRNA concentration between these two strains. 5'UTR involvement in mRNA stability regulation was already described in literature: 5'UTR can protect or destabilize mRNAs by recruiting or preventing RNase E binding (Prévost et al., 2011; Richards & Belasco, 2019) or indirectly when ribosome protection is altered (Iost & Dreyfus, 1995). Nevertheless, the effect of 5'UTR on mRNA stability was not the only parameter responsible of variation in mRNA concentration. For instance, 5'UTR1 and 5'UTR6 showed similar *lacZ* mRNA half-lives, but their mRNA concentrations had a 7.6-fold difference. For the 6 constructs with similar *lacZ* mRNA half-lives, variation in mRNA stability cannot explain the observed variations in mRNA concentration. In these cases, the more probable explanation is that the 5'UTR could have also an effect on transcription. 5'UTR involvement

in regulation of transcription initiation and termination is already described in literature, for example via the initiation transcription sequence, ITS (Heyduk & Heyduk, 2018) or the binding of sRNAs (Rabhi et al., 2011; Sedlyarova et al., 2016).

Taken together, our results show that changing the 5'UTR can have an effect on mRNA concentration, stability, or both. In addition, for some 5'UTRs, the variation in mRNA concentration could not be due to a variation in stability, meaning that the 5'UTR had to have a direct effect on the level of transcription. In the next section, we analyzed whether the effects of the 5'UTR on both stability and mRNA concentration are dependent or not on the downstream reporter gene.

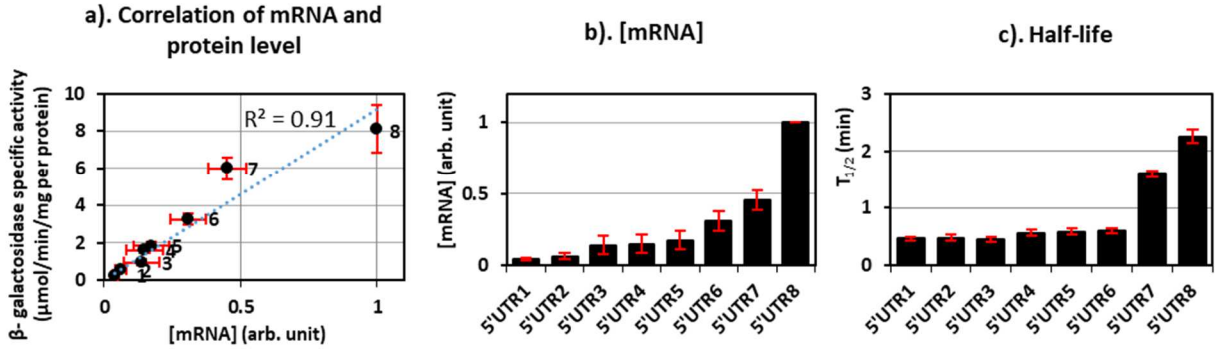


Figure 5. Concentration and stability of the *lacZ* mRNA of the 8 selected constructs and level of the corresponding β-galactosidase protein. a) Correlation of β-galactosidase activity and *lacZ* mRNA concentration. Numbers 1-8 represent the constructs with 5'UTR 1 to 8, respectively. *lacZ* mRNA concentration is expressed as fold change compared to the reference strain 5'UTR8. The blue dashed line represents the linear regression curve and the coefficient of determination is indicated (R^2=0.91); b) *lacZ* mRNA concentration as fold change compared to the reference strain 5'UTR8; c) Half-life of *lacZ* mRNA in min for the 8 constructs. For all, error bars represent standard deviation (n= 3).

5. Reporter gene replacement

We next analyzed whether the modifications of mRNA stability and concentration observed with *lacZ* as reporter gene are somewhat dependent on the reporter, meaning that the reporter sequence participates in the regulations, or independent, suggesting that modifications are fully accountable to the 5'UTRs. To address these questions, we replaced the *lacZ* gene with two other unrelated reporters: the *txAbF* gene encoding an α-L-arabinofuranosidase and *msfGFP* gene encoding a fluorescent protein for the 8 different 5'UTRs previously analyzed. For all constructs, we measured the protein level, mRNA concentration, and mRNA stability (***Figure 6***).

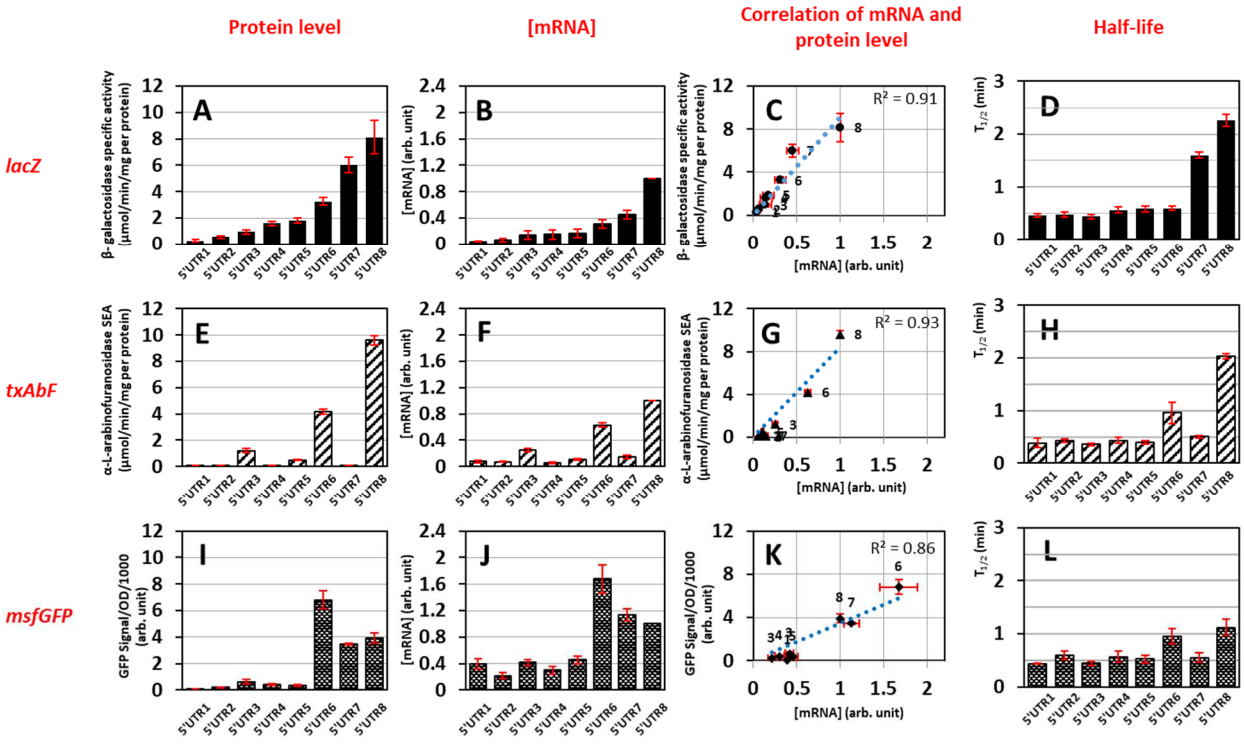


Figure 6. Protein levels and mRNA concentration of the constructs with 8 different 5'UTRs and 3 different reporters *(lacZ, txAbF* and *msfGFP).* For Fig 6A, B, C and D related to *lacZ*, the values are those shown previously in Fig 5A, B and C. A) Specific β-galactosidase activity. B) *lacZ* mRNA concentration as a fold change relative to 5'UTR8-*lacZ*. C) Correlation between *lacZ* mRNA concentration and β-galactosidase activity. D) *lacZ* mRNA half-life. E) Specific α-L-arabinofuranosidase activity. F) *txAbF* mRNA concentration as a fold change relative to 5'UTR8-*txAbF*. G) Correlation between *txAbF* mRNA concentration and α-L-arabinofuranosidase activity. H) *txAbF* mRNA half-life. I) msfGFP fluorescence/OD. J) *msfGFP* mRNA concentration as a fold change relative to 5'UTR8-*msfGFP*. K) Correlation between *msfGFP* mRNA concentration and GFP fluorescence/OD. L) *msfGFP* mRNA half-life. For C, G, K, the blue dashed line represents the linear regression and the coefficient of determination (R^2) is indicated on the graph. For all, error bars represent standard deviation: n=3 for mRNA concentration and half-life determinations, n=9 (biological triplicates and technical triplicates) for β-galactosidase, TxAbF and GFP activity.

Reporter replacement affects the 5'UTR-mediated regulation of protein level

For all constructs, we determined the level of the corresponding protein. Similar to what we observed with the β-galactosidase (***Figure 6A***), the protein levels of TxAbF (***Figure 6E***) and msfGFP (***Figure 6I***) were different between strains with different 5'UTRs. Furthermore, the effect of a given 5'UTR on the protein level differed with the reporter. When comparing the three reporters, the 5'UTRs leading to the highest and lowest protein levels were not the same. The highest β-galactosidase and TxAbF protein levels were obtained with 5'UTR8 while for msfGFP it was with 5'UTR6. The lowest TxAbF level was measured for 5'UTR2 while it was for 5'UTR1 for the β-galactosidase and msfGFP. Some 5'UTR leading to middle-level expression of β-galactosidase led to nearly undetectable levels of TxAbF (5'UTR1-2-4-7).

Taken together, these results strongly suggest that the 5'UTR effect on protein level is somehow dependent on the combination of the 5'UTR sequence and the downstream gene coding sequence.

Reporter replacement affects the 5'UTR-mediated regulation of mRNA level

We compared the mRNA concentration for all constructs of 5'UTRs fused to *lacZ*, *txAbF* and *msfGFP* genes (***Figure 6B, 6F, 6J***). Similar to what was observed with *lacZ* (***Figure 6B)***, the concentrations of *txAbF* (***Figure 6F***) and *msfGFP* (***Figure 6J***) mRNAs were different between strains with different 5'UTRs (although all constructs were under the same transcriptional control of the inducible promoter). When comparing the three reporters, the 5'UTRs corresponding to the highest and lowest mRNA concentrations were not the same. The highest mRNA concentration was found with 5'UTR8 for *lacZ* and *txAbF* but with 5'UTR6 for *msfGFP*. The lowest mRNA concentration was obtained with 5'UTR1 for *lacZ* but with 5'UTR4 for *txAbF* and 5'UTR2 for *msfGFP*. Globally, we can notice that for each 5'UTR, the level of mRNA is dependent on the reporter gene. This suggests that the control of the mRNA concentration depends on the combination of the 5'UTR sequence and the downstream gene coding sequence

Reporter replacement maintains the correlation between protein and mRNA levels

When we compared the correlation between protein level and mRNA concentration for the three reporters (***Figure 6C, 6G, 6K***), we always found a good correlation between the two parameters. For the 3 reporters and 5'UTR, the mRNA concentration remained the major parameter to control protein expression. In other words, 5'UTRs selected for regulating mRNA concentration rather translation in the case of *lacZ*, have the same behavior in the case of the two other reporters *txAbF* and *msfGFP*.

Reporter replacement affects the 5'UTR-mediated regulation of mRNA stability

For all constructs, we analyzed the mRNA stability. Similar to what was observed with *lacZ* (***Figure 6D***), the mRNA half-life of *txAbF* (***Figure 6H***) and *msfGFP* (***Figure 6L***) were different between strains with different 5'UTRs. The 5'UTR associated with the highest half-life of each reporter mRNA was the same 5'UTR8, but for the second longest, it was with 5'UTR7 for *lacZ* instead of 5'UTR6 for *txAbF* and *msfGFP*. We can note also that the amplitude between the most and the least stable mRNA constructs was different depending on the reporter, with 5-fold for *lacZ*, 6-fold for *txAbF*, and 2.5-fold for *msfGFP*.

Similar to what was observed for *lacZ*, the variation of mRNA stability can explain to some extent the variations of mRNA concentrations for *txAbF* and *msfGFP*. This point will be analyzed in more detail in next section.

6. Effects of 5'UTRs on transcription and/or stability to regulate mRNA level depend on the reporter gene

To study the degree of involvement of mRNA stability regulation in mRNA concentration changes, we compared the variations in mRNA half-life and concentration. First, we graphically observe how differences in 5'UTR can affect the transcription process and/or the stability of mRNA, and in turn mRNA concentration (***Figure 7***). Three cases can be described: (i) a variation in concentration is related to a variation in stability (***Figure 7G; red***). For example, the half-life of 5'UTR8-*lacZ* is higher than 5'UTR4-*lacZ* (***Figure 7A***), and it is associated with a higher mRNA concentration of 5'UTR8-*lacZ* than of 5'UTR4-*lacZ* (***Figure 7B***); (ii), the mRNA concentration variation is not related to modification of stability but is under the control of transcription (***Figure 7G; blue***). For example, the half-lives of 5'UTR3-*txAbF* and 5'UTR1-*txAbF* are similar (***Figure 7 C***) while the mRNA concentration is increased with 5'UTR3-*txAbF.* This means that transcription of 5'UTR3-*txAbF* was higher than that of 5'UTR1-*txAbF* (***Figure 7D***); and (iii) both variations in mRNA stability and transcription are responsible for variation in mRNA concentration (***Figure 7G; green***). For example, the half-life of 5'UTR8-*msfGFP* is higher than that of 5'UTR7-*msfGFP* (***Figure 7E***), but their mRNA concentrations are similar (***Figure 7F***). In this situation, transcription of 5'UTR8-*msfGFP* must have been lower than that of 5'UTR7-msfGFP to counteract the effect of an increase in mRNA stability and to result in similar mRNA concentrations of 5'UTR8-*msfGFP* and 5'UTR7-*msfGFP*. In this case, variation in mRNA concentrations results from both variation in mRNA stability and transcription.

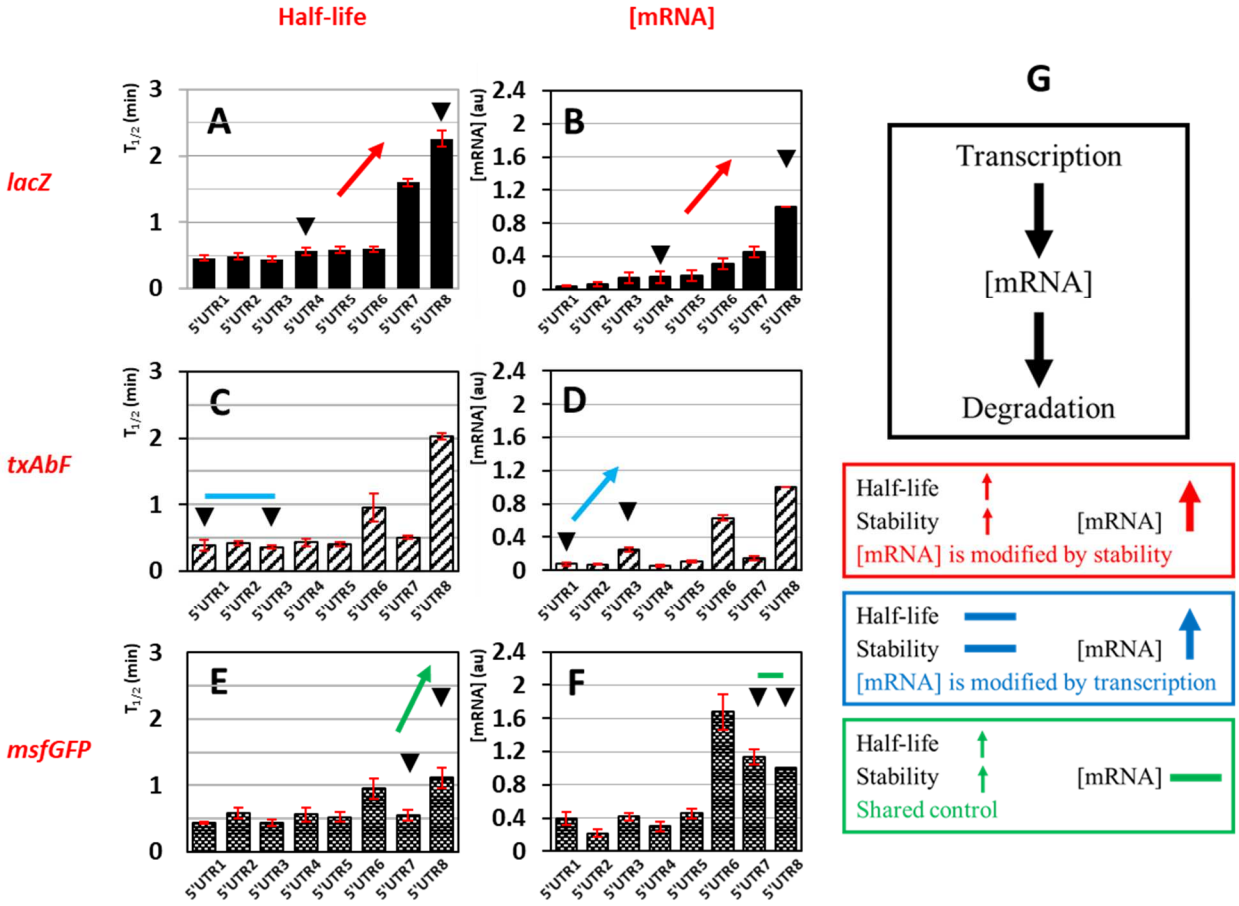


Figure 7. Graphical interpretation of the effects of 5'UTRs on mRNA stability and concentration. The data of mRNA half-lives (A, C and E) and concentrations (B, D and F) of 24 combinations of 8 5'UTRs (5'UTR1-8) and 3 reporters (lacZ, txAbF and msfGFP) are those shown in Figure 6**.** The standard deviation is marked with red error bars (n=3). **G)** Scheme of the dynamic relationship between the stability and concentration of an mRNA. For all the graphs, the red arrow exemplifies a case where mRNA concentration is regulated by stability, a blue arrow a case where mRNA concentration is regulated by transcription, and a green arrow a case where mRNA concentration is regulated by a shared control of stability and transcription.

Second, to quantify more precisely how variations in mRNA concentration depend on variation in transcription and/or stability, we calculated the degradational regulation cofficient (*ρD*) defined as the contribution of stability regulation in the control of an mRNA concentration. For a given reporter gene, *ρD* is calculated using the equation given in the Material and Methods from the measured variations in mRNA concentrations and half-lives between two strains and we arbitrarily chose 5'UTR6 as the reference for the comparisons (***Table 1***).

Table 3. Degradational regulation coefficient ρD of constructs with 8 different 5'UTRs and 3 different reporters.

Reporter		5'UTR1	5'UTR2	5'UTR3	5'UTR4	5'UTR5	5'UTR6	5'UTR7	5'UTR8
lacZ	*ρD*	0.13	0.13	0.37	0.08	0.03	-	2.6	1.13
txAbF	*ρD*	0.44	0.37	1.08	0.33	0.48	-	0.44	1.61
msfGFP	*ρD*	0.54	0.23	0.56	0.31	0.46	-	1.39	0.31

For $\rho D<0.4$ (blue), variation in mRNA concentration is mainly due to regulation of transcription relative to the reference (5'UTR6);

For $0.4<\rho D<0.6$ (green), variation in mRNA concentration is due to a shared control by transcription and degradation regulations relative to the reference (5'UTR6);

For $\rho D>0.6$ (red), variation in mRNA concentration is mainly due to regulation of degradation relative to the reference (5'UTR6).

From the *ρD* values, we classified each combination of 5'UTR-reporter in one of the three regulatory types: mainly controlled by transcription or stability, or by a shared control. We found that for a given reporter gene, different 5'UTRs can result in various types of regulation. For example, for *txAbF with* 5'UTR2, 5'UTR3 and 5'UTR5, variations in mRNA concentration were mainly regulated by transcription, mainly regulated by degradation and under a shared control, respectively. In addition, a given 5'UTR can result in similar or different types of regulation depending on the reporter gene. 5'UTR2 and 5'UTR4 were always associated with variations of mRNA concentration under a transcriptional control, regardless of the downstream reporter. This suggests that 5'UTR2 and 5'UTR4 did not have an important effect on the stability of the mRNA. In contrast, 5'UTR3 was associated with the three types of control, transcriptional, degradational, and shared control depending on the reporter gene. In this case, the effect of the 5'UTR3 seems to depend on the downstream reporter gene. We can note that we did not identify a 5'UTR that was associated with a degradational control for all reporters.

Conclusion and perspectives

In this research, we confirmed that mRNA concentration and translation initiation are regulatory factors of protein level. We provided evidence that 5'UTRs play a key role in regulating gene expression in *E. coli* at the level of translation but also at the levels of transcription and mRNA degradation. Multilevel regulations mediated by 5'UTRs were also reported in the literature: for example, variation in transcript production and translational level in a study on the heterologous protein expression controlled by different *Pseudomonas putida Pm* 5'UTR variants (Berg et al., 2012) and variation in translation yield, mRNA concentration and half-life in a study of *lacZ* expression using 5'UTRs with AU-rich sequences inserted upstream of the SD sequence (Komarova et al., 2005).

We observed that the β-galactosidase expression level was globally correlated with the theoretical translation initiation rates (RBS index) of our synthetic 5'UTRs but within a group of a similar RBS index, discrepancy in protein level was measured. Therefore, the theoretical translation initiation rate (given by the softwares RBS Calculator and 5'UTR Designer) cannot be used alone to accurately predict the real protein expression level. Our results demonstrated that additional 5'UTR-related effects influence the protein level. We measured 5'UTR-related effects on mRNA concentration as shown by the large variation in reporter mRNA concentrations for different 5'UTRs at the same concentration of transcription inducer. More precisely, we showed that 5'UTRs could regulate mRNA concentration by changing its stability. The observed overall tendency of mRNA concentration increasing with increased RBS index supported the assumption of an indirect effect of 5'UTRs on mRNA stability. 5'UTRs related to efficient translation initiation rates (RBS Index) may recruit more ribosomes on mRNA, and the high ribosome occupancy may in turn protect the transcript from rapid degradation by RNases, thereby increasing the mRNA concentration (Iost & Dreyfus, 1995). However, we cannot exclude a more direct effect of the 5'UTR sequence on transcript stability that could modulate the accessibility of cleavage sites to RNases. We also showed that 5'UTRs can regulate mRNA concentration by changing its transcription. A possible mechanism could be that 5'UTR regulates transcription by acting on the process of promoter escape (Heyduk & Heyduk, 2018). Additional analysis of the nucleotide composition and structure of the ITS present in our synthetic 5'UTRs needs to be performed to validate this assumption. Changing the reporter sequence led to the conclusion that 5'UTR-mediated effects on transcription, mRNA degradation and translation depend on the combination of the 5'UTR and the gene coding sequence. The context dependency of 5'UTR element was already reported: a prokaryotic ribosome binding site that initiates translation for one gene coding sequence might not function at all with another coding sequence (Mutalik et al., 2013; Salis et al., 2009; Zwick et al., 2013). The context dependency of regulatory elements is a big challenge for synthetic

biology that wishes to make the various standardized regulatory parts truly plug-and-play whatever the context (downstream and upstream sequences, bacterial host).

A limitation of this study is that transcription regulation was estimated indirectly (by measuring changes in mRNA concentration and half-life), but not quantified directly experimentally. We therefore plan to measure the transcription rate *in vitro* for combinations of 5'UTR-reporter. This will give us information on the effect of each 5'UTR on transcription only, without the cumulative effect of degradation like *in vivo*. A second possibility will be to use a (stable) tRNA positioned at the 3'end of the mRNA molecule as a marker of the level of transcription *in vivo* (Iost & Dreyfus, 1995).

From this study, we cannot yet edit precise and general rules of mRNA concentration and protein synthesis regulation by 5'UTRs. It would be useful to increase the number of constructs and select 5'UTRs representative of specific regulations: only translational regulation (i.e. 5'UTRs leading to similar mRNA concentration but different protein expression), regulation of mRNA concentration mainly due to stability regulation (i.e. 5'UTRs related to degradational control) or mainly due to transcription regulation (i.e. 5'UTRs related to transcriptional control regulation). Analysis of their sequence and structural characteristics might help us to better understand the underlying molecular mechanisms of each process. However, transcription, mRNA degradation, and translation are spatiotemporally interactive, and any changes to either may trigger a domino effect (Dahan et al., 2011). The possibility that an initial regulation of the 5'UTR on transcript concentration (via transcription or mRNA stability, independently) or on translation may have a subsequent potential impact on other processes complicating a systemic understanding of the overall effect of the 5'UTR on the regulation of gene expression.

REFERENCES

Abolbaghaei, A., Silke, J. R., & Xia, X. (2017). How Changes in Anti-SD Sequences Would Affect SD Sequences in Escherichia coli and Bacillus subtilis. *G3: Genes|Genomes|Genetics*, *7*(5), 1607–1615. https://doi.org/10.1534/g3.117.039305

Ah-Seng, Y., Rech, J., Lane, D., & Bouet, J.-Y. (2013). Defining the role of ATP hydrolysis in mitotic segregation of bacterial plasmids. *PLoS Genetics*, *9*(12), e1003956. https://doi.org/10.1371/journal.pgen.1003956

Berg, L., Kucharova, V., Bakke, I., Valla, S., & Brautaset, T. (2012). Exploring the 5'-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible Pm promoter in Escherichia coli. *Journal of Biotechnology*, *158*(4), 224–230. https://doi.org/10.1016/j.jbiotec.2011.07.012

Bissaro, B., Saurel, O., Arab-Jaziri, F., Saulnier, L., Milon, A., Tenkanen, M., Monsan, P., O'Donohue, M. J., & Fauré, R. (2014). Mutation of a pH-modulating residue in a GH51 α-l-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1840*(1), 626–636. https://doi.org/10.1016/j.bbagen.2013.10.013

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254. https://doi.org/10.1006/abio.1976.9999

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., & Heumann, H. (2004). The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. *Nature Structural & Molecular Biology*, *11*(6), 551–557. https://doi.org/10.1038/nsmb768

Cetnar, D. P., & Salis, H. M. (2021). Systematic Quantification of Sequence and Structural Determinants Controlling mRNA stability in Bacterial Operons. *ACS Synthetic Biology*, *10*(2), 318–332. https://doi.org/10.1021/acssynbio.0c00471

Chiaruttini, C., & Guillier, M. (2020). On the role of mRNA secondary structure in bacterial translation. *WIREs RNA*, *11*(3). https://doi.org/10.1002/wrna.1579

Condon, C. (2007). Maturation and degradation of RNA in bacteria. *Current Opinion in Microbiology*, *10*(3), 271–278. https://doi.org/10.1016/j.mib.2007.05.008

Dahan, O., Gingold, H., & Pilpel, Y. (2011). Regulatory mechanisms and networks couple the different phases of gene expression. *Trends in Genetics*, *27*(8), 316–322. https://doi.org/10.1016/j.tig.2011.05.008

Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A. J., Girbal, L., & Cocaign-Bousquet, M. (2014). Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Research*, *42*(4), 2460–2472. https://doi.org/10.1093/nar/gkt1150

Hatoum, A., & Roberts, J. (2008). Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. *Molecular Microbiology*, *68*(1), 17–28. https://doi.org/10.1111/j.1365-2958.2008.06138.x

Held, P. (2007, February 26). *Kinetic Analysis of ß-Galactosidase Activity using PowerWave™ HT Microplate Spectrophotometer and Gen5™ Data Analysis Software*. https://www.biotek.com/resources/application-notes/kinetic-analysis-of-galactosidase-activity-using-powerwave-ht-microplate-spectrophotometer-and-gen5-data-analysis-software/

Heyduk, E., & Heyduk, T. (2018). DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Research*, *46*(9), 4469–4486. https://doi.org/10.1093/nar/gky172

Holmqvist, E., & Vogel, J. (2018). RNA-binding proteins in bacteria. *Nature Reviews Microbiology*, *16*(10), 601–615. https://doi.org/10.1038/s41579-018-0049-5

Hui, M. P., Foley, P. L., & Belasco, J. G. (2014). Messenger RNA Degradation in Bacterial Cells. *Annual Review of Genetics*, *48*(1), 537–559. https://doi.org/10.1146/annurev-genet-120213-092340

Iost, I., & Dreyfus, M. (1995). The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO Journal*, *14*(13), 3252–3261. https://doi.org/10.1002/j.1460-2075.1995.tb07328.x

Jia, L., Mao, Y., Ji, Q., Dersh, D., Yewdell, J. W., & Qian, S.-B. (2020). Decoding mRNA translatability and stability from the 5' UTR. *Nature Structural & Molecular Biology*, *27*(9), 814–821. https://doi.org/10.1038/s41594-020-0465-x

Kang, Z., Zhang, C., Zhang, J., Jin, P., Zhang, J., Du, G., & Chen, J. (2014). Small RNA regulators in bacteria: Powerful tools for metabolic engineering and synthetic biology. *Applied Microbiology and Biotechnology*, *98*(8), 3413–3424. https://doi.org/10.1007/s00253-014-5569-y

Kirchner, M., Schorpp, K., Hadian, K., & Schneider, S. (2017). An in vivo high-throughput screening for riboswitch ligands using a reverse reporter gene system. *Scientific Reports*, *7*(1), 7732. https://doi.org/10.1038/s41598-017-07870-w

Komarova, A. V., Tchufistova, L. S., Dreyfus, M., & Boni, I. V. (2005). AU-Rich Sequences within 5J Untranslated Leaders Enhance Translation and Stabilize mRNA in Escherichia coli. *J. BACTERIOL.*, *187*, 6.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, *361*, 13–37. https://doi.org/10.1016/j.gene.2005.06.037

Malina, A., Khan, S., Carlson, C. B., Svitkin, Y., Harvey, I., Sonenberg, N., Beal, P. A., & Pelletier, J. (2005). Inhibitory properties of nucleic acid-binding ligands on protein synthesis. *FEBS Letters*, *579*(1), 79–89. https://doi.org/10.1016/j.febslet.2004.06.103

McCormick, J. R., Zengel, J. M., & Lindahl, L. (1994). Correlation of Translation Efficiency with the Decay of lacZ mRNA in Escherichia coli. *Journal of Molecular Biology*, *239*(5), 608–622. https://doi.org/10.1006/jmbi.1994.1403

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., & Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, *10*(4), 354–360. https://doi.org/10.1038/nmeth.2404

Nickels, B. E., Mukhopadhyay, J., Garrity, S. J., Ebright, R. H., & Hochschild, A. (2004). The σ70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. *Nature Structural & Molecular Biology*, *11*(6), 544–550. https://doi.org/10.1038/nsmb757

Nouaille, S., Mondeil, S., Finoux, A.-L., Moulis, C., Girbal, L., & Cocaign-Bousquet, M. (2017). The stability of an mRNA is influenced by its concentration: A potential physical mechanism to regulate gene expression. *Nucleic Acids Research*, *45*(20), 11711–11724. https://doi.org/10.1093/nar/gkx781

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), 45e–445. https://doi.org/10.1093/nar/29.9.e45

Picard, F., Dressaire, C., Girbal, L., & Cocaign-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. *Comptes Rendus Biologies*, *332*(11), 958–973. https://doi.org/10.1016/j.crvi.2009.09.005

Prabhakaran, R., Chithambaram, S., & Xia, X. (2015). Escherichia coli and Staphylococcus phages: Effect of translation initiation efficiency on differential codon adaptation mediated by virulent and temperate lifestyles. *The Journal of General Virology*, *96*(Pt 5), 1169–1179. https://doi.org/10.1099/vir.0.000050

Prévost, K., Desnoyers, G., Jacques, J.-F., Lavoie, F., & Massé, E. (2011). Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes & Development*, *25*(4), 385–396. https://doi.org/10.1101/gad.2001711

Rabhi, M., Espéli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V., & Boudvillain, M. (2011). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *The EMBO Journal*, *30*(14), 2805–2816. https://doi.org/10.1038/emboj.2011.192

Richards, J., & Belasco, J. G. (2019). Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Molecular Cell*, *74*(2), 284-295.e5. https://doi.org/10.1016/j.molcel.2019.01.044

Romeo, T., Gong, M., Liu, M. Y., & Brun-Zinkernagel, A. M. (1993). Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *Journal of Bacteriology*, *175*(15), 4744–4755. https://doi.org/10.1128/jb.175.15.4744-4755.1993

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, *27*(10), 946–950. https://doi.org/10.1038/nbt.1568

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., & Nudler, E. (2016). SRNA-Mediated Control of Transcription Termination in *E. coli*. *Cell*, *167*(1), 111-121.e13. https://doi.org/10.1016/j.cell.2016.09.004

Shine, J., & Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(4), 1342–1346. https://doi.org/10.1073/pnas.71.4.1342

Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B. P., & Yusupov, M. (2009). A structural view of translation initiation in bacteria. *Cellular and Molecular Life Sciences*, *66*(3), 423–436. https://doi.org/10.1007/s00018-008-8416-4

Storz, G., Opdyke, J. A., & Zhang, A. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. *Current Opinion in Microbiology*, *7*(2), 140–144. https://doi.org/10.1016/j.mib.2004.02.015

Tuller, T., & Zur, H. (2015). Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Research*, *43*(1), 13–28. https://doi.org/10.1093/nar/gku1313

Viegas, S. C., Apura, P., Martínez-García, E., de Lorenzo, V., & Arraiano, C. M. (2018). Modulating Heterologous Gene Expression with Portable mRNA-Stabilizing 5′-UTR Sequences. *ACS Synthetic Biology*, *7*(9), 2177–2188. https://doi.org/10.1021/acssynbio.8b00191

Wakabayashi, H., Warnasooriya, C., & Ermolenko, D. N. (2020). *Extending the spacing between the Shine-Dalgarno sequence and P-site codon reduces the rate of mRNA translocation* [Preprint]. Biochemistry. https://doi.org/10.1101/2020.04.16.045807

Yarchuk, O., Iost, I., & Dreyfus, M. (1991). The relation between translation and mRNA degradation in the lacZ gene. *Biochimie*, *73*(12), 1533–1541. https://doi.org/10.1016/0300-9084(91)90188-7

Yarchuk, O., Jacques, N., Guillerez, J., & Dreyfus, M. (1992). Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *Journal of Molecular Biology*, *226*(3), 581–596. https://doi.org/10.1016/0022-2836(92)90617-s

Zwick, F., Lale, R., & Valla, S. (2013). Combinatorial engineering for heterologous gene expression. *Bioengineered*, *4*(6), 431–434. https://doi.org/10.4161/bioe.24703

Supplementary data

Table S4. Sequence and RBS index of the 41 synthetic 5'UTR

Oligo Name	RBS index (au)	DNA Sequence (-33 nt)	Note
5'UTR_100_01	100	TCTTTCCTTGTCCTCATCCATAGATAAAGGATA	
5'UTR_100_02		ACGTCCGCCCGCAGCGTAGCTTGGCGGATAACC	
5'UTR_100_03		TGTTCCGGTCTCCTTCGCTCCTTCTACCTTTCT	
5'UTR_250_04	250	CAACTACAAAACGACTTTCTTTGTTCCGAAGTA	
5'UTR_250_05		CGACTCGACCCGCCTGTTGTCATCGGGACAGCG	
5'UTR_250_06		TAGACCAGTCTCCTCAAGACCTTGCTGTACACC	
5'UTR_500_07	500	CAAAACTCCGCTCTCATAGGAGCGAACGGATAA	Same to $5'UTR_1$
5'UTR_500_08		CTTAACCCATTGGTTAAATAATTCTGTAAATCT	
5'UTR_500_09		CGAGCTGTCCGGGGGCCAGTAAGGAGTAACCAG	Same to $5'UTR_3$
5'UTR_1000_10	1000	TTACGAAGTTAGCCCTGACAATTAGAGCCCGAC	
5'UTR_1000_11		GGCCACTCTGGAGCTGAAAAAGGACGGGGCAAC	
5'UTR_1000_12		TGAAACCTCAAAACGTTAACAAATATATTCATC	
5'UTR_1500_13	1500	CCAAAGAGCAGACGACCACAAAGACTATCACTC	
5'UTR_1500_14		CCTTCTAGACCTCCCCGTCAGGAGCCACGGTTC	Same to $5'UTR_2$
5'UTR_1500_15		CAATATCGCCCATACAAGAAGTAAAAATAGTCT	Same to $5'UTR_4$
5'UTR_2500_16	2500	GACCAATTGGCCCGCCCAGAAGGAGGCTGGCAT	Same to $5'UTR_5$
5'UTR_2500_17		GAGAATATCAAACTTAATATTGAGAATCCCTGT	
5'UTR_2500_18		CATTGCTGCTAGTCTAGAGCCCGAGGAGTATAT	
5'UTR_5k_19	5000	CCTCGTCCTCCATATAAAAGGGGAGGAGTCAAG	
5'UTR_5k_20		TAACCCTAATAGAAAATTATTAAGCTATATATA	
5'UTR_5k_21		ATAGCGGACACTCCGCAAGACCTACGAGGAACC	
5'UTR_5k_22		CAACAACATAAAATAATCTAAAGCTAATTTTAA	
5'UTR_5k_23		AGTATCGATAAGAGCTAAGGAAACCCCTTCAAA	

5'UTR_5k_24		GCACTACCCCTACAAATTTCGCTAGAGAGTAAC	
5'UTR_5k_25		ACCAAGACGATAAATTAAGGGAGGACAAGAATC	
5'UTR_33k_26	33000	CAATCGCCCCCAACAAGAAGATAAGGAGGCGCC	
5'UTR_33k_27		TGACGGAGGTCGCAAGATCAGGGAGGGACGAAG	
5'UTR_33k_28		CAACGAAGGTATTAAAAATAAGATAGGAATAAA	
5'UTR_33k_29		TGAAAAAAACTATATAGATTTATAAGGGGTAAC	Same to $5'UTR_6$
5'UTR_33k_30		TGGGGAACTTTCATTTTTAAGGTATAGGTCAGT	
5'UTR_33k_31		ACCCGTTTTTTGGGCTAACAGGAGGAATTAACC	Same to $5'UTR_8$
5'UTR_33k_32		AGAACGCGTCAACAAACATTAAAGAGGTCAATC	
5'UTR_66k_33	66000	CATTCAAATTAAAACTTATATAAGAAGGTAGAA	
5'UTR_66k_34		GTAAGCTATCCCATCAAGTAATAAGGAGTTCCA	
5'UTR_66k_35		CACCCGACTACCCCCAAGAATTAAGGAGGCGAG	Same to $5'UTR_7$
5'UTR_66k_36		GGACTCGCACGAGTCAGAACAGGAGGAGGATAA	
5'UTR_66k_37		ATAAGATAAATCAAAAAAGTAAGGATAAAAACC	
5'UTR_100k_38	100000	CCTTCGCGAAGAGCGACAAAATAAGGAGGGCTT	
5'UTR_100k_39		TTTCAAGATTAAGAGAAATAAGGAAAGGTAAAA	
5'UTR_100k_40		CGCGTAATCCGCGAATAAAAAGGAGGAGGAAAC	
5'UTR_100k_41		TAACCCTATTCATTTAAGGAGGTAACCTAAACC	

Table S2. Primers used in this study

Primer number	Sequence 5'-3'	Function Description
1	ATGGATCCACTAGTAACGGCCGCC	Forward primer for cloning 5'UTR into P_{BAD}-*LacZ*
2	TATCCTTTATCTATGGATGAGGACAAGGAAAGA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer to cloning 5'UTR_100_01 into P_{BAD}-*LacZ*
3	GGTTATCCGCCAAGCTACGCTGCGGGCGGACGT ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_100_02 into P_{BAD}-*LacZ*
4	AGAAAGGTAGAAGGAGCGAAGGAGACCGGAACA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_100_03 into P_{BAD}-*LacZ*

5	TACTTCGGAACAAAGAAAGTCGTTTTGTAGTTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_250_04 into P_{BAD}-*LacZ*
6	CGCTGTCCCGATGACAACAGGCGGGTCGAGTCG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_250_05 into P_{BAD}-*LacZ*
7	GGTGTACAGCAAGGTCTTGAGGAGACTGGTCT AATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_250_06 into P_{BAD}-*LacZ*
8	TTATCCGTTCGCTCCTATGAGAGCGGAGTTTTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_500_07 into P_{BAD}-*LacZ*
9	AGATTTACAGAATTATTTAACCAATGGGTTAAG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_500_08 into P_{BAD}-*LacZ*
10	CTGGTTACTCCTTACTGGCCCCCGGACAGCTCG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_500_09 into P_{BAD}-*LacZ*
11	GTCGGGCTCTAATTGTCAGGGCTAACTTCGTAA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_1000_10 into P_{BAD}-*LacZ*
12	GTTGCCCCGTCCTTTTTCAGCTCCAGAGTGGCC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_1000_11 into P_{BAD}-*LacZ*
13	GATGAATATATTTGTTAACGTTTTGAGGTTTCA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_1000_12 into P_{BAD}-*LacZ*
14	GAGTGATAGTCTTTGTGGTCGTCTGCTCTTTGG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_1500_13 into P_{BAD}-*LacZ*
15	GAACCGTGGCTCCTGACGGGGAGGTCTAGAAGG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_1500_14 into P_{BAD}-*LacZ*
16	AGACTATTTTTACTTCTTGTATGGGCGATATTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_1500_15 into P_{BAD}-*LacZ*
17	ATGCCAGCCTCCTTCTGGGCGGGCCAATTGGTC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_2500_16 into P_{BAD}-*LacZ*
18	ACAGGGATTCTCAATATTAAGTTTGATATTCTC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_2500_17 into P_{BAD}-*LacZ*
19	ATATACTCCTCGGGCTCTAGACTAGCAGCAATG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_2500_18 into P_{BAD}-*LacZ*
20	CTTGACTCCTCCCCTTTTATATGGAGGACGAGG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_19 into P_{BAD}-*LacZ*

21	TATATATAGCTTAATAATTTTCTATTAGGGTTA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_20 into P_{BAD}-*LacZ*
22	GGTTCCTCGTAGGTCTTGCGGAGTGTCCGCTAT ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_21 into P_{BAD}-*LacZ*
23	TTAAAATTAGCTTTAGATTATTTTATGTTGTTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_22 into P_{BAD}-*LacZ*
24	AGTATCGATAAGAGCTAAGGAAACCCCTTCAAA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_23 into P_{BAD}-*LacZ*
25	GCACTACCCCTACAAATTTCGCTAGAGAGTAAC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_24 into P_{BAD}-*LacZ*
26	ACCAAGACGATAAATTAAGGGAGGACAAGAATC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_5k_25 into P_{BAD}-*LacZ*
27	GGCGCCTCCTTATCTTCTTGTTGGGGGCGATTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_26 into P_{BAD}-*LacZ*
28	CTTCGTCCCTCCCTGATCTTGCGACCTCCGTCA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_27 into P_{BAD}-*LacZ*
29	TTTATTCCTATCTTATTTTTAATACCTTCGTTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_28 into P_{BAD}-*LacZ*
30	GTTACCCCTTATAAATCTATATAGTTTTTTTCA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_29 into P_{BAD}-*LacZ*
31	TGGGGAACTTTCATTTTTAAGGTATAGGTCAGT ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_30 into P_{BAD}-*LacZ*
32	ACCCGTTTTTTGGGCTAACAGGAGGAATTAACC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_31 into P_{BAD}-*LacZ*
33	AGAACGCGTCAACAAACATTAAAGAGGTCAATC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_33k_32 into P_{BAD}-*LacZ*
34	TTCTACCTTCTTATATAAGTTTTAATTTGAATG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_66k_33 into P_{BAD}-*LacZ*
35	GTAAGCTATCCCATCAAGTAATAAGGAGTTCCA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_66k_34 into P_{BAD}-*LacZ*

36	CTCGCCTCCTTAATTCTTGGGGGTAGTCGGGTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_66k_35 into P_{BAD}-*LacZ*
37	TTATCCTCCTCCTGTTCTGACTCGTGCGAGTCC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_66k_36 into P_{BAD}-*LacZ*
38	GGTTTTTATCCTTACTTTTTTGATTTATCTTAT ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_66k_37 into P_{BAD}-*LacZ*
39	AAGCCCTCCTTATTTTGTCGCTCTTCGCGAAGG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_100k_38 into P_{BAD}-*LacZ*
40	TTTTACCTTTCCTTATTTCTCTTAATCTTGAAA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_100k_39 into P_{BAD}-*LacZ*
41	GTTTCCTCCTCCTTTTTATTCGCGGATTACGCG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_100k_40 into P_{BAD}-*LacZ*
42	TAACCCTATTCATTTAAGGAGGTAACCTAAACC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for cloning 5'UTR_100k_41 into P_{BAD}-*LacZ*
43	ATGAACGTGGCAAGCCGGGTAGTCGT	Forward primer for amplifying the CDS of TxAbF
44	TCAGTGGTGGTGGTGGTGGTGCTCG	Reverse primer for amplifying the CDS of TxAbF
45	ATGAGCAAAGGAGAAGAACTTTTCACTGGAG	Forward primer for amplifying the CDS of msfGFP
46	TTATTTGTAGAGCTCATCCATGCCATGTGTAATC	Reverse primer for amplifying the CDS of msfGFP
47	GTTTAAACGGTCTCCAGCTTGGC	Forward primer for amplifying the plasmid backbone for CDS replacement
48	TTATCCGTTCGCTCCTATGAGAGCGGAGTTTTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_1 for CDS replacement
49	GAACCGTGGCTCCTGACGGGGAGGTCTAGAAGG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_2 for CDS replacement
50	CTGGTTACTCCTTACTGGCCCCCGGACAGCTCG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_3 for CDS replacement

51	AGACTATTTTTACTTCTTGTATGGGCGATATTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_4 for CDS replacement
52	ATGCCAGCCTCCTTCTGGGCGGGCCAATTGGTC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_5 for CDS replacement
53	GTTACCCCTTATAAATCTATATAGTTTTTTTCA ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_6 for CDS replacement
54	CTCGCCTCCTTAATTCTTGGGGGTAGTCGGGTG ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_7 for CDS replacement
55	ACCCGTTTTTTGGGCTAACAGGAGGAATTAACC ATGGAGAAACAGTAGAGAGTTGCG	Reverse primer for amplifying the plasmid backbone from 5'UTR_8 for CDS replacement
56	GTCGTGACTGGGAAAACCCTGG	Forward primer for qRT-PCR analysis of *lacZ* beginning part
57	AACTGTTGGGAAGGGCGATCG	Reverse primer for qRT-PCR analysis of *lacZ* beginning part
58	AACAACTTTAACGCCGTGCGCT	Forward primer for qRT-PCR analysis of *lacZ* middle part
59	CACCATGCCGTGGGTTTCAATA	Reverse primer for qRT-PCR analysis of *lacZ* middle part
60	CAGCTGGCGCAGGTAGCAGAG	Forward primer for qRT-PCR analysis of *lacZ* end part
61	GGCAGATCCCAGCGGTCAAA	Reverse primer for qRT-PCR analysis of *lacZ* end part
62	CGTAAACGCCGACAGGGTGAA	Forward primer for qRT-PCR analysis of *txAbF* beginning part
63	CGGAATCGGCGAATCTTCTCC	Reverse primer for qRT-PCR analysis of *txAbF* beginning part
64	ACTACGGCGACAACAAGCTGCA	Forward primer for qRT-PCR analysis of *txAbF* middle part
65	AGCGACAGCCCGTGCATGA	Reverse primer for qRT-PCR analysis of *txAbF* middle part
66	GGCGGACGGCAAGATCCAC	Forward primer for qRT-PCR analysis of *txAbF* end part
67	GTCAGCGTCGTGCCGGTTG	Reverse primer for qRT-PCR analysis of *txAbF* end part

68	GGCACAAATTTTCTGTCCGTGGA	Forward primer for qRT-PCR analysis of *msfGFP* beginning part
69	TGACAAGTGTTGGCCACGGAAC	Reverse primer for qRT-PCR analysis of *msfGFP* beginning part
70	TTTCAAAGATGACGGGACCTACAAGA	Forward primer for qRT-PCR analysis of *msfGFP* middle part
71	TCGAGTTTGTGTCCAAGAATGTTTCC	Reverse primer for qRT-PCR analysis of *msfGFP* middle part
72	CGATGGCCCTGTCCTTTTACCA	Forward primer for qRT-PCR analysis of *msfGFP* end part
73	GCCATGTGTAATCCCAGCAGCA	Reverse primer for qRT-PCR analysis of *msfGFP* end part
74	GCCAAGACGGTTGAAGATGC	Forward primer for qRT-PCR analysis of internal normalization control *ihfB* gene
75	CAAAGAGAAACTGCCGAAACC	Reverse primer for qRT-PCR analysis of internal normalization control *ihfB* gene

CHAPTER 2

Role of native 5'UTRs in the regulation of gene expression for adaptation of *E. coli* to different growth conditions

Summary

To identify if 5'UTRs are used in bacterial cells to control gene expression in response to environmental changes, we developed an original approach in *E. coli* based on the construction of an exhaustive native 5'UTR library challenged in different growth conditions. A large number of 5'UTRs (2547) were cloned upstream a reporter gene (encoding msfGFP) and under the control of an inducible promoter. After transcription induction of the *msfGFP* gene in exponential phase in different environmental conditions, cells were sorted by flow cytometry into six windows, from very low msfGFP level (lowest fluorescence) to extremely high level (highest fluorescence), reflecting the gene expression level. Then, the identities of the 5'UTRs present in each expression fraction were determined by next generation sequencing and mapped onto the *E. coli* genome. We developed a procedure to select 5'UTRs with a well-positioned peak of the read count within the six windows. Comparison of the peak positions of each selected 5'UTR in different growth conditions let us to identify 5'UTRs that contribute or not to *msfGFP* expression regulation by changing the level of msfGFP protein depending on the environment. We identified relationships between 5'UTR-related regulations of msfGFP level and the function of 5'UTR-associated genes when *E. coli* cells respond to environmental changes, showing the involvement of 5'UTR in canonical gene regulation. In some cases, riboswitch and binding of the CsrA protein were likely involved in the 5'UTR-related regulatory mechanisms. This study identified 5'UTRs that could be used in the field of biotechnology to control gene expression in *E. coli.*

Introduction

Bacteria have to constantly regulate their gene expression to adapt to changing environments like temperature, environment composition, availability of nutrients, and many other parameters. Adaptation is achieved by modifying gene expression to fine-tune cellular components to the metabolic requirements of the new environment. For instance, in response to environmental stresses such as heat and osmotic stress, the expression of hundreds of *E. coli* genes is modified (Bartholomäus et al., 2016). Changes in gene expression are often associated with changes at the transcription level, but also at the post-transcriptional level or both (Heyduk & Heyduk, 2018; Nouaille et al., 2017; Picard et al., 2009; Tuller & Zur, 2015).

The 5'UTR sequences are considered as potential multi-level regulators of gene expression (Kaberdin & Bläsi, 2006; Tietze & Lale, 2021). At the translational level, the 5'UTR is important for gene expression because it contains the Ribosome Binding Site (RBS), which triggers translation initiation by interaction with the 16S part of the ribosome (Kaminishi et al., 2007). 5'UTRs are also involved in the control of mRNA stability through the 5'-end dependent entry of RNase E, a key endoribonuclease of the mRNA degradation pathway in *E. coli* (Carpousis et al., 2009). In the 5'-end dependent mRNA

degradation pathway, the phosphorylation of the 5'end of the 5'UTR and the obstacles encountered by RNase E in scanning the cleavage site determine the regulation of mRNA stability (Luciano et al., 2019; Richards & Belasco, 2019). The 5'UTR is also involved in the regulation of transcription, including abortive initiation and premature termination. For the 5'UTR-mediated abortive initiation of transcription, the initial transcription sequence (ITS) corresponding to the first 20 nucleotides of the 5'UTR has been shown to increase or delay the escape of RNA polymerase from the promoter, thereby participating in regulation at the transcriptional level (Heyduk & Heyduk, 2018). The 5'UTR is also the common target of the Rho-dependent terminators that mediated premature termination of transcription, and many genes rely on sRNA-mediated anti-termination to control premature transcription termination in *E. coli* (Sedlyarova et al., 2016).

The regulation of gene expression by 5'UTRs appears to be under the control of environmental factors. Growth temperatures can affect the affinity of SD: aSD resulting in changes in translation initiation. For instance, the optimal SD sequence with the highest translation initiation rate in *E. coli* differs at 37°C and 20°C (Vimberg et al., 2007). Furthermore, secondary structures in the 5'UTR are dynamic and can fold or unfold upon stimulation by environmental factors, thereby exposing or blocking RBS, RNase cleavage sites, binding sites of sRNA and ribosome binding proteins (RBPs) to affect translation, mRNA stability, and transcription (Chiaruttini & Guillier, 2020; Del Campo et al., 2015; Hollands et al., 2012; Villa et al., 2018). For example, we can cite the temperature-sensitive RNATs (RNA thermometers) (Sharma et al., 2022) and the metabolite-sensitive riboswitches (Nudler & Mironov, 2004). In addition, the 5'capping mechanism, which utilizes non-canonical substrates to initiate transcription, may also play a role in 5'UTR-mediated adaptation in bacteria (Luciano et al., 2019; Luciano and Belasco, 2020). It was hypothesized that bacteria use methylated caps as protection against RNA degradation under starvation conditions (Hudeček et al., 2020).

Nevertheless, the full role of 5'UTRs in adaptation is not clearly understood. Are these sequences involved in differential gene expression depending on the growth environment? To address this issue, we developed an exhaustive library of native *E. coli* 5'UTRs cloned upstream of a fluorescent reporter gene. The library was sorted on fluorescent level and each 5'UTR in each fraction was identified. Then the library was grown in different media and at different temperatures. We identified the 5'UTRs that did or did not modulate fluorescent levels between different growth environments. We searched for relationships between 5'UTR-mediated regulation of reporter gene expression and function of the 5'UTR-associated gene.

Materials and Methods

1. Design and synthesis of 5'UTR library

The genome of *E. coli* K-12 MG1655 (GenBank assembly accession: GCA_000005845.2) was used as a reference for the design of the native 5'UTR library. Two studies have previously experimentally determined genome-scale boundaries of 5'UTRs by identifying transcriptional start sites (TSS) (Kim et al., 2012; Mendoza-Vargas et al., 2009). From these combined data sets, we selected the 5'UTRs of monocistronic genes and genes positioned first in operons. In the case of more than one TSS identified for a gene, the shortest 5'UTRs were selected. We restricted selection of 5'UTRs with a maximum length of 300 ntd due to technical constraints for synthesis. This led to a set of 2547 unique 5'UTRs. The distribution of the 2547 native 5'UTRs per length is represented in ***Figure 1***.

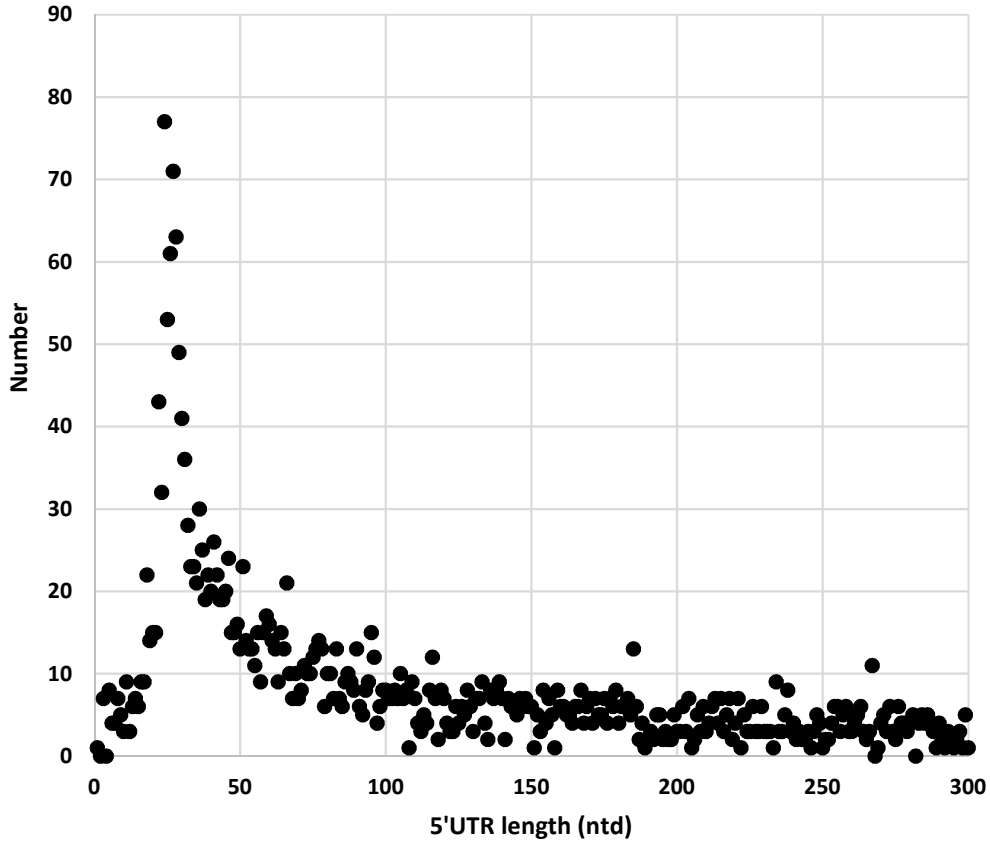


Figure 1. Length distribution of the 2547 5'UTRs. Each dot represents the number of 5'UTR in the synthesised library for each length.

For molecular assembly into the expression vector, each 5'UTR sequence was extended by 20 nucleotides on either side, corresponding respectively to the 3' end of the pBAD promoter sequence and the 5' end of *msfGFP* coding sequence: (ACTCTCTACTGTTTCTCCAT$N_{(n)}$ATGAGCAAAGGAG

AAGAACT_where N corresponds to the designed native 5'UTR sequence). The 2547 sequences were synthesized as ssDNA by IDT (Integrated DNA Technologies) and provided as an oligonucleotide pool.

2. Bacterial strains, vectors and reagents

All analyses were performed in MET 345 strain, the *E. coli* DCT2022 strain (MG1655 *ΔaraFGH, Ωpcp18::araE533)*, with a deleted chromosomal *lacZ* copy (Nouaille et al., 2017). This strain allows linear induction of the pBAD expression system without cellular heterogeneity within the culture. Unless otherwise indicated, sub-clonings were performed in *E. coli* DH5α (New England Biolabs) before transfer into MET 345. Reagents for molecular biology were purchased from New England Biolabs and used as recommended. Sanger sequencings were performed by Eurofins (Mix2Seq). The pMET 296 was used for the 5'UTR library establishment (see below). This plasmid is a derivative of P_{BAD}-lacZ-cmyc/his (Invitrogen). It is composed of the *msfGFP* gene encoding the monomeric superfolder green fluorescent protein under the control of the P_{BAD} inducible promoter. All cultures for molecular biology were routinely grown in LB medium supplemented with 100 µg/mL of ampicillin (Sigma) and incubated at 37°C and 150 rpm agitation.

3. Construction of the native 5'UTR library

The 2547 5'UTRs were provided as ssDNA at low concentrations. For cloning, dsDNA sufficient amount was obtained by PCR using primers targeting the common 20 ntd present upstream and downstream of the 5'UTR sequences (primers 1103 and 1104, ***Table S1***). The PCR mix consisted of 0.01 pmol template (ssDNA), 50 pmol of each primer, 1 U Q5 DNA polymerase, 400 µM dNTP mix, 10 µL of the 10x buffer provided by the manufacturer, and nuclease-free water to a final volume of 50 µL. The reaction was performed under the following conditions: 30 s at 98°C followed by 16 cycles of 10 s at 98°C, 30 s at 50°C and 10 s at 72°C, followed by a final extension at 72°C for 2 min. The resulting PCR products were cleaned up by 10 U exonuclease I to remove excess primers at 37°C for 15 min. Exonuclease I was inactivated at 80°C for 15 min.

The recipient vector was generated from pMET 296 by PCR amplification of the whole plasmid without the 5'UTR portion, using two primers (primers 1085 and 397, ***Table S1***) complementary to those used for amplifying the native 5'UTRs set. The PCR product was gel-purified and quantified (NanoDrop) before use.

Integration of the 5'UTR set in the plasmid backbone was performed by molecular assembly. The insert to vector molar ratio was 5:1 and assembly was performed in a mix with 10 µL of HiFi DNA Assembly

Master Mix (New England Biolabs), and sterile water to a final volume of 20 µl. The reaction mix was incubated at 50°C for 60 min and stored at -20°C until use.

The assembly reaction product (2 µL) was transferred by electroporation into NEB DH5-α electrocompetent *E. coli* cells. Upon electroporation, cells were recovered in SOC medium at 37°C for 1 h with shaking at 200 rpm, plated on LB agar dishes supplemented with ampicillin for plasmid selection, and grown overnight at 37°C. In parallel, serial dilutions of the electroporated cells were plated to estimate transformation efficiency. Four independent transformations were performed to obtain a sufficient number of clones for representative coverage of the library. We estimated from the plating of serial dilutions that the four library subsets contained 4.5×10^4 clones representing an 18-fold coverage of the library size. All transformants were scraped, resuspended in LB supplemented with ampicillin and grown overnight. A mix of all plasmids was extracted using Midiprep Kits (Thermo Fisher). The plasmids library was transferred into MET 345 by electroporation, a total of 6.8×10^5 clones was obtained, scraped, resuspended in LB for overnight culture, and the library was stored as 1 ml aliquots with glycerol (40% final) at −80°C.

4. Cell growth for fluorescence measurements

For small-scale characterization, an aliquot of the library was used to inoculate 5 ml of LB with ampicillin for overnight culture. After serial dilutions and plating, 96 clones were randomly selected for growth in a microplate containing 200 µL LB supplemented with ampicillin 100 µg/mL and 0.01% arabinose. Plates were incubated in Bio Tek Synergy H1 microplate reader (Bio Tek, Winooski, USA) at 37°C with high orbital shaking for 12 hours. Every 10 minutes, the OD_{600} and the fluorescence (475 nm excitation filter and a 510 nm emission filter) were recorded.

For cell sorting experiments, cultures were performed either on chemically defined medium M9 supplemented with glucose (Esquerré et al., 2014) at 20, 37 or 42°C, or on LB at 37°C, under shaking. Four different conditions were performed depending on the medium and temperature, hereafter referred to as M9_37°C, M9_42°C, M9_20°C, and LB_37°C. Overnight cultures of the whole library were used to inoculate 150 ml of medium supplemented with ampicillin at an initial OD_{600} of 0.1. The OD of induction with 0.01% of arabinose and sampling time were adapted to the growth condition. For M9_37°C and M9_42°C, cells were induced when the OD_{600} reached 0.5-0.6, sampling after 2.5 h (OD=2-2.5); For M9_20°C, cells were induced when OD_{600} reached 0.2, sampling after 16 h (OD=0.7); For LB_37°C, cells were induced when OD_{600} reached 0.6, sampling after 1 h (OD=1). After induction, 10 mL of cultures were collected, cells were washed and resuspended in cold phosphate buffer saline

(PBS), diluted to OD_{600} of 0.1 in cold PBS corresponding to ~ $1x10^8$ CFU/mL and kept on ice for no more than a few hours until cell sorting procedure.

5. Cell sorting by FACS

Cells cultivated under different environmental conditions were subjected to cell sorting based on their fluorescent levels using the Cell sorter MoFlo Astrios EQ (Beckman Coulter) at the TWB FACS platform (Toulouse). Cells spanning the whole fluorescent distribution were sorted into 6 distinct windows: 10% of the population with the lowest fluorescence, 10% of the population with the highest fluorescence, and the rest was divided into 4 equal groups containing each 20% of the remaining fluorescence distribution. Before sorting, the background auto fluorescence of *E. coli* was subtracted. Depending on the condition and initial cell concentration, from 10 000 to 100 000 cells per window were collected and transferred to LB with ampicillin but without arabinose for overnight propagation (37°C, 150 rpm). The plasmid population contained in each fraction was recovered by miniprep preparation (Qiagen miniprep).

6. Sequencing and mapping

Sequencing and processing of NGS data were performed by the GeT-Biopuces platform (INSA-TBI Toulouse) using the Ion Torrent technology (Thermo Fisher Scientific).

Quality control of DNA for sequencing. Extracted DNA was evaluated by gel electrophoreses, and the nucleic acid concentrations were quantified by Qubit™ (Thermo Fisher Scientific). The DNA fragments corresponding to 5'UTRs were PCR amplified, and purified on E-Gel (Thermo Fisher Scientific) to obtain homogeneous fragments of 350-450 bp in size to meet sequencing requirements.

Sequencing. For each sequencing library preparation, 5 ng of DNA were used with NEXTflex Cell Free DNA-seq kit (PerkinElmer) and NEXTflex DNA Barcodes (PerkinElmer) according to the manufacturer's recommendations. The prepared sequencing libraries were quantified with the Qubit™ (dsDNA HS Assay Kit) and the Bioanalyzer 2100 (Agilent-High Sensitivity kit). After pooling, sequencing was performed using the Ion Torrent technology.

Mapping. All raw reads were trimmed for adapter sequences and barcode sequences using cutadapt. Trimmed reads were aligned and mapped onto the reference genome (*Escherichia_coli_str*_k_12_ substr_mg1655. ASM584v2) with the "map4" algorithm of TMAP module of the Ion Torrent Suite

Software v5.12.1 (Thermo Fisher Scientific). The number of reads by gene was counted with HTSeq-count v0.11.2 (PMID: 25260700).

7. Gene Ontology enrichment analyses

Gene Ontology (GO) term enrichment was used to link 5'UTR-mediated gene expression regulation to 5'UTR-associated gene function using the "biological processes" category. The associations between genes and GO-terms were obtained from EcoCyc https://ecocyc.org/. The database applicable to this study was reduced to the 2547 5'UTR-associated genes. Significantly enriched GO terms categories were identified by applying an adjusted p-value cutoff of 0.01.

Results

1. Small scale quality control of the library

Development and use of the 5'UTR library required to obtain a library as diverse as possible to contain the targeted 2547 native 5'UTRs. Since in a second step, we will have to be able to visualize as much fluorescence range of the clones as possible, we first decided to evaluate the potentialities of the library at a very small scale, by testing a few randomly selected clones.

96 clones were randomly selected for 'Real-time' monitoring of fluorescence intensity over growth on a microplate. The variation of OD_{600} with time reflects the growth of each clone in microwell (***Figure 2***). All but two clones ended the exponential phase after about 6-7 hours. One explanation could have been that the differences in growth were due to various levels in msfGFP produced between clones, mobilizing different levels of cellular resources. However, this explanation seems unlikely because we did not observe a correlation between growth rate and msfGFP production.

Fluorescence intensity was monitored simultaneously. Dividing the fluorescence intensity by OD at each reading point provides the specific fluorescence level of the individual cell over growth. As expected, the specific msfGFP fluorescence intensity for all clones reached a peak at the end of the exponential phase (***Figure 3***). However, the absolute specific msfGFP fluorescence intensities were different between clones by more than 16 times (***Figure 4***), confirming differences in msfGFP production capabilities between the few tested representatives of our library. From these experiments, we selected the sampling time for cell sorting in late exponential phase to obtain the most comprehensive range of msfGFP fluorescence intensity.

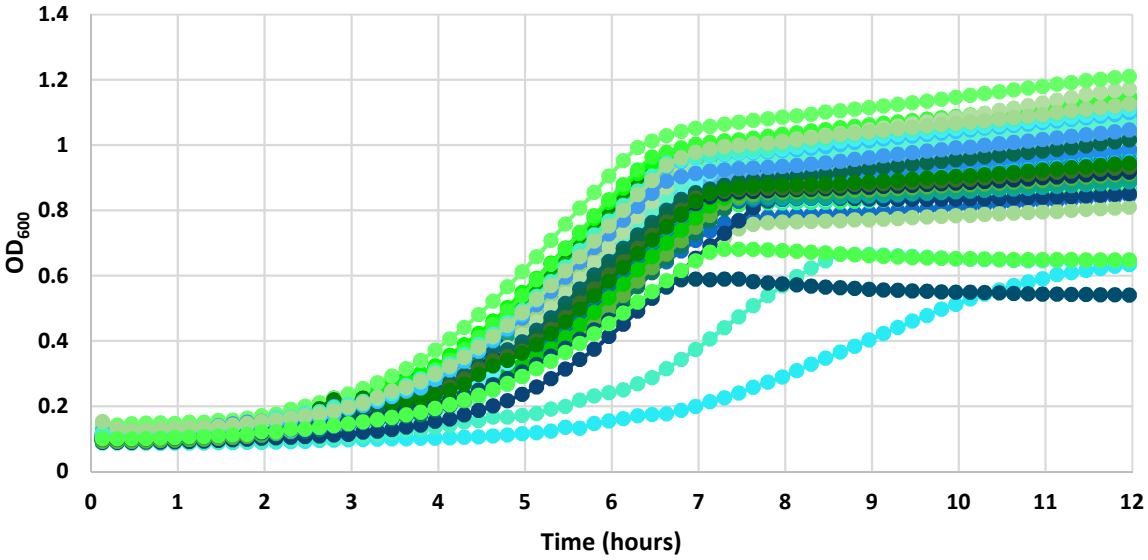


Figure 2. Growth of the 96 randomly selected clones. The OD values at 600 nm of 96 clones were monitored on microplates every 10 minutes for 12 hours. Dotted lines with different colors represent different clones.

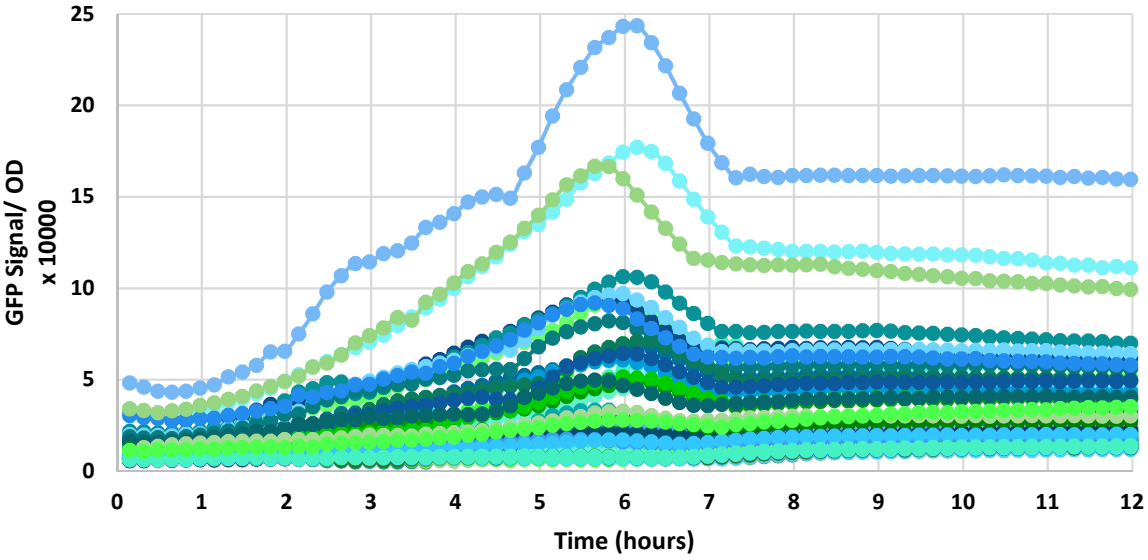


Figure 3. Fluorescence intensity per cell of the 96 randomly selected clones. GFP fluorescence intensity divided by OD is plotted as a function of time during real-time kinetic analyses of *E. coli* cells harboring plasmids with different 5'UTRs. Fluorescence intensity was monitored at 475nm excitation and 510 nm emission wavelength every 10 minutes for 12 hours. Dotted lines with different colors represent different clones.

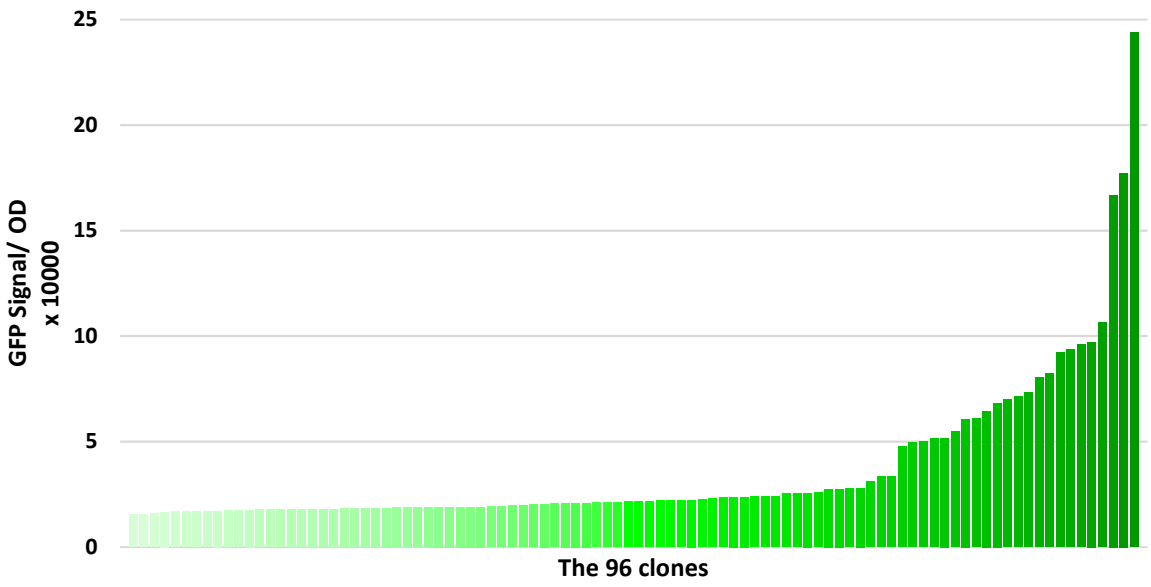


Figure 4. Fluorescent intensity per cell at time 6.15 h of the 96 randomly selected clones. The 96 clones harboring plasmids with different 5'UTRs were ranked in ascending order of msfGFP fluorescence intensity per cell measured at time 6.15 h.

To quickly confirm on a small-scale the representativeness of the 5'UTR library the absence of major bias, we randomly selected 7 of 96 clones and sequenced their 5'UTR (***Table 1***). We identified unique 5'UTRs with a length between 18 and 117 ntd and in agreement with their theoretical length. (***Table 1***).

Table 1. Comparison of the sequencing results of 7 clones with the theoretical native 5'UTRs

5'UTR- msfGFP	Sequenced size	Theoretical size	5'UTR-associated gene
clone 1	18	18	*basR*
clone 2	19	19	*frmR*
clone 3	26	26	*rraA*
clone 4	36	36	*yebW*
clone 5	46	46	*ydhP*
clone 6	48	48	*folB*
clone 7	117	117	*ybjG*

Together, these results suggest that our library is suitable for further characterization as it covers a large range of msfGFP fluorescence.

2. Library expression profile in various environmental conditions

The study aims to identify 5'UTRs responsible for gene expression modifications in response to environmental modification. We tested the effect of changes in medium composition, by comparing growth in rich medium (LB) and chemically defined M9 medium. It was previously shown that a large set of genes had their expression modified between the two media (Thomason et al., 2015). In addition, we tested growth at different temperatures as this parameter can modulate gene expression at different levels. Some gene expressions are specifically up or down regulated in response to temperatures stresses (Bartholomäus et al., 2016), and their 5'UTR may participate in such regulations. Temperature can also modify gene expression through different foldings of 5'UTRs, affecting downstream gene expression (Kortmann & Narberhaus, 2012; Narberhaus et al., 2006). Therefore, the entire native 5'UTR library was grown in 4 growth conditions, M9_20°C, M9_37°C, M9_42°C, and LB_37°C.

We first determined the level of transcription induction required to obtain a maximal range of msfGFP fluorescence from the library. We compared two induction levels (0.01% arabinose and 0.1% arabinose) to no induction. The msfGFP fluorescence clouds were measured by flow cytometry for the three induction levels. Similar results were obtained for the 4 growth conditions. An example of fluorescence distribution in M9_37°C is shown in ***Figure 5***. Without induction, the whole population was clustered at low fluorescence intensity. This signal reflected the autofluorescence background of cells or a possible weak leak of the P_{BAD} promoter. With increasing concentration of arabinose to induce transcription, we noticed an expansion of the fluorescence range in proportion to the level of transcription induction. This expansion over a 3-log scale reflects the wide range of msfGFP expression in our library. Although maximum expansion was achieved with 0.1% arabinose induction, we chose to apply a middle level induction (0.01%) because we wondered if a too high induction level might hide some of the regulatory effects of 5'UTRs.

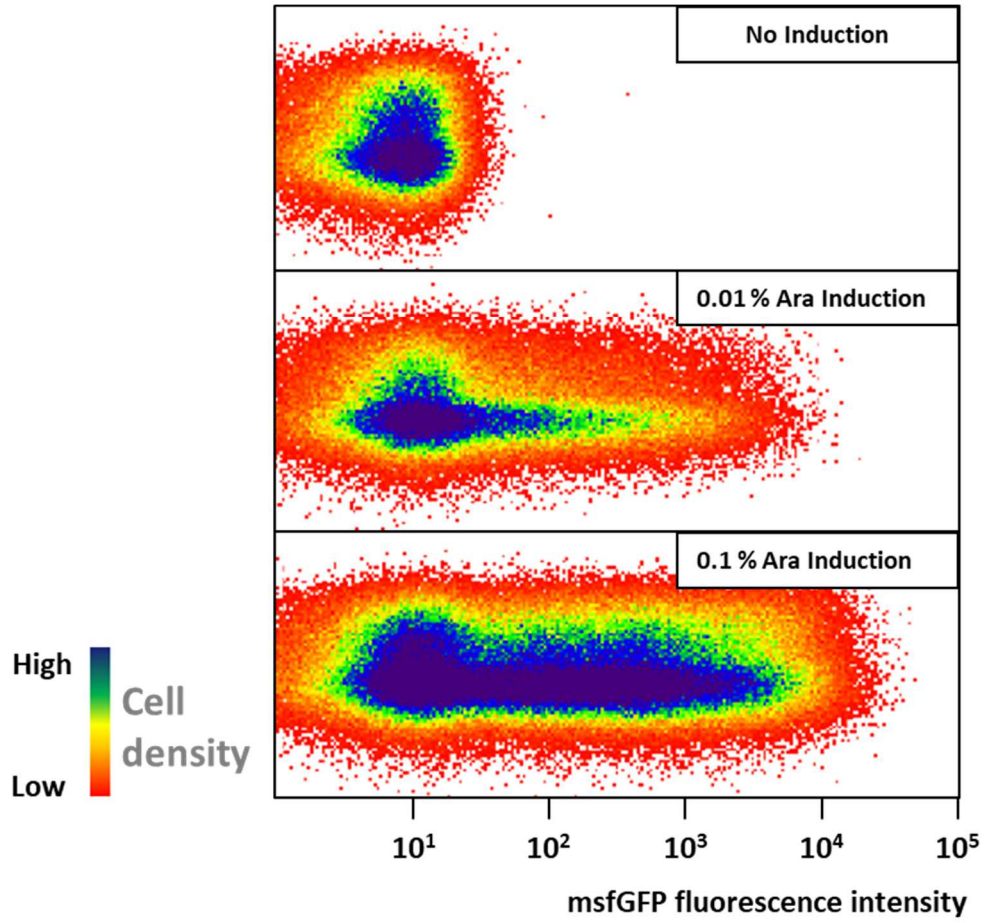


Figure 5. Cloud of msfGFP fluorescence in M9_37°C induced by different arabinose concentrations. Transcription in *E. coli* cells harboring plasmids with different 5'UTRs was induced (0.01% arabinose and 0.1% arabinose) or not at M9_37°C. The horizontal coordinate presents the msfGFP fluorescence intensity after subtracting background autofluorescence. The dots from red to dark blue represent low to high cell density.

The whole 5'UTR library was cultivated in the 4 growth conditions, induced with 0.01% arabinose and the msfGFP fluorescence intensity was monitored (***Figure 6***). The clouds of fluorescence intensity of cells under each growth condition was widely distributed between low and high signals. The distribution of fluorescence was similar between conditions, except for growth at 20°C with a reduction of about 1-log of the fluorescence range. These results show that our library of native 5'UTRs results in a vast range of msfGFP expression that differs under different growth condition. This likely reflects 5'UTR-dependent regulation of msfGFP expression affected by changing environments.

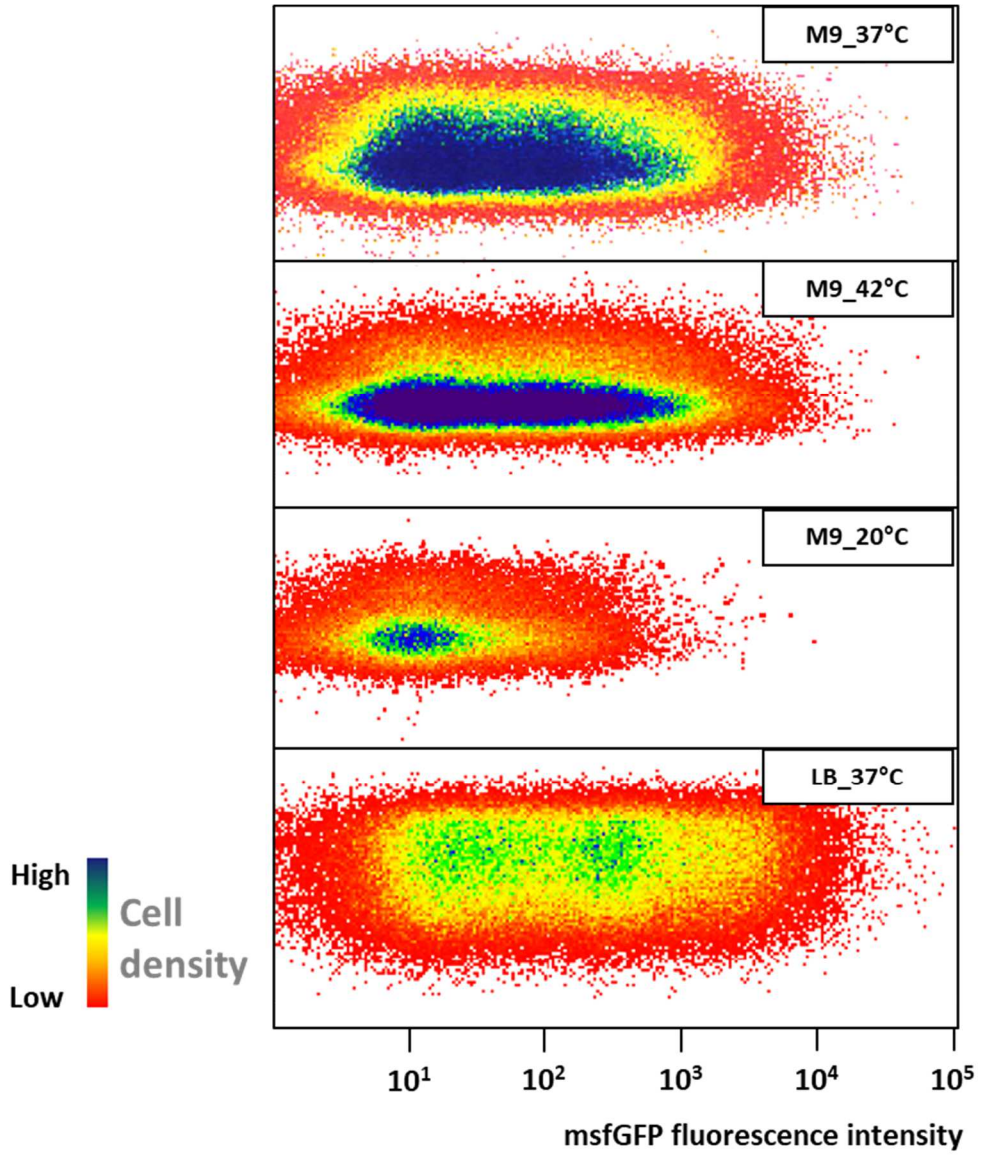


Figure 6. Clouds of msfGFP fluorescence of cells grown under different conditions. *E. coli* containing the 5'UTR librarywere cultivated under four growth conditions (M9_37°C, M9_42°C, M9_20°C, LB_37°C). The dots from red to dark blue represent low to high cell density.

3. Cell sorting by msfGFP expression windows

The project aims to identify which 5'UTRs are present in each expression window in each growth condition. If some 5'UTRs are always present in the same window regardless of the growth environment, this will suggest that these 5'UTRs are not involved in gene expression regulation; and if some 5'UTRs are present in different windows depending on growth environment, this will suggest that they contribute to gene expression regulation.

Because the range of fluorescence were different under different growth conditions, we could not separate the sub-populations into windows based on intensity criteria. We based the window selection on the number of cells relative to the total number of cells. For each condition, 6 windows were defined. The W1 fraction contained 10 % of the cells with the lowest fluorescence intensity. The W6 contained 10% of the cells with the highest fluorescence intensity. In between, the remaining cells were divided into 4 equal fractions containing each 20 % of the library (***Figure 7***). In each growth condition, the 6 sub-libraries were propagated and their plasmid DNA was purified for sequencing.

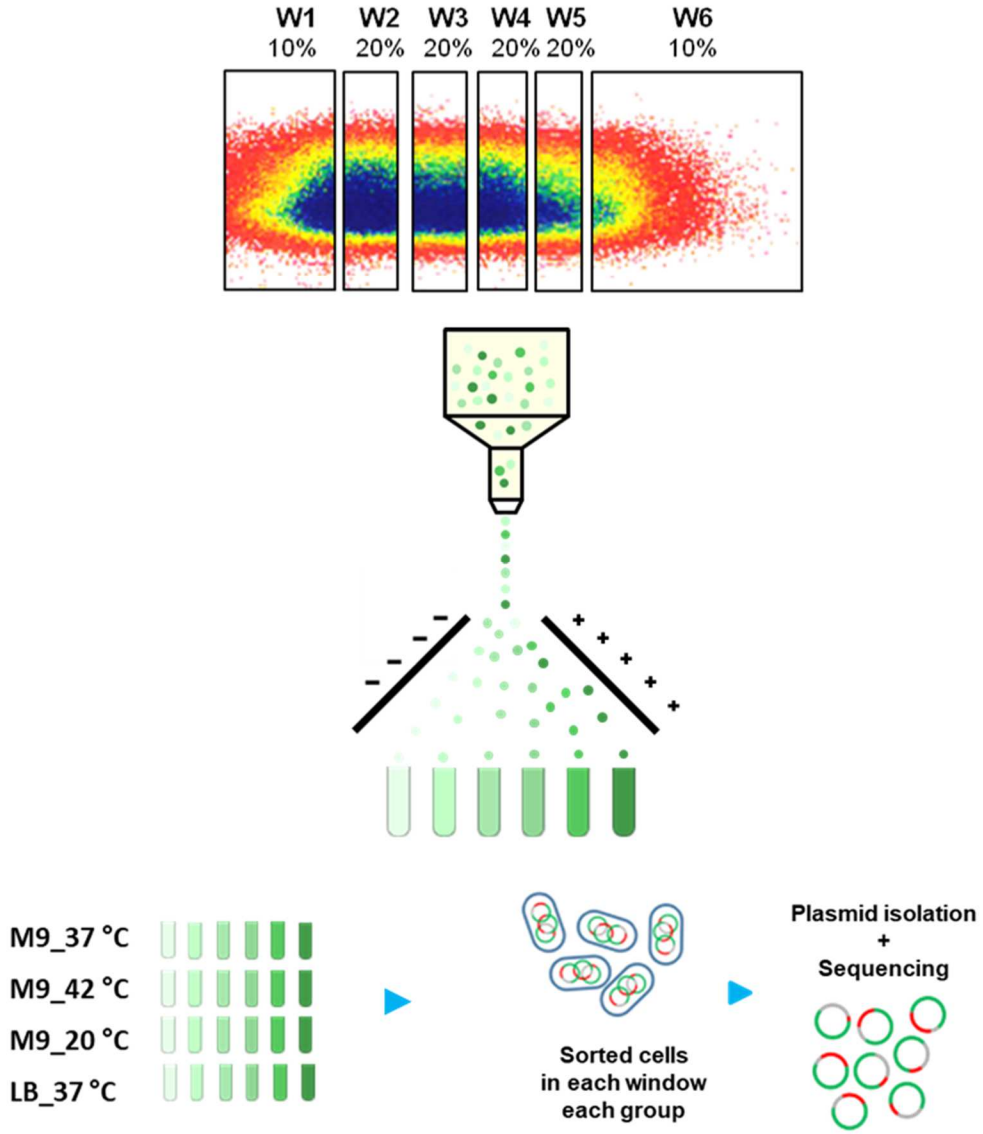


Figure 7. Schematic representation of the workflow used for library sorting. In each growth condition, cells were sorted and the 6 sub-libraries were propagated and their plasmid DNA was purified for sequencing.

4. Adjustment of the sequencing method for a library heterogeneous in size

The requirement for NGS sequencing is to start from DNA fragments homogenous in size of about 350-450 bp but in our study, the 5'UTRs are heterogeneous with a length spanning from 1 to 300 ntd. Therefore, we developed an amplification method to re-homogenize the size of the DNA fragments without losing information.. As described in ***Figure 8***, **a)** we designed ten pairs of primers to amplify the 5'UTRs from the plasmids. The primers used are listed in ***Table S1***. Each primer pair hybridized on the vector backbone upstream and downstream of the 5'UTRs. The pair located at the closest position of 5'UTRs amplified the complete 5'UTR sequences with the addition of 50 ntd from the vector, resulting in amplicons from 51 to 351 ntd. Each primer pair increased the amplicon size by 50nt compared to the previous one. **b)** Each sample was amplified individually with the 10 pairs of primers. **c)** Once mixed, amplicons were from 40 nt to 790 nt long. **d)** Each 5'UTR independently of its initial length was now present in an amplicon from 350 to 450 ntd long. This amplification method was applied to the 24 samples generated this study (6 windows x 4 growth conditions). An example of amplification products for window 1 of M9_37°C is given in ***Figure 9***. The length range of the ten sets of dsDNA amplification products from 40-340 bp to 480-790bp was in line with expectations of a 300 nt range (***Figure 9A***). After pooling and gel purification, the length of the re-homogenized dsDNA was distributed from 350 bp to 450 bp. (***Figure 9B***) consistent with the requirements for sequencing.

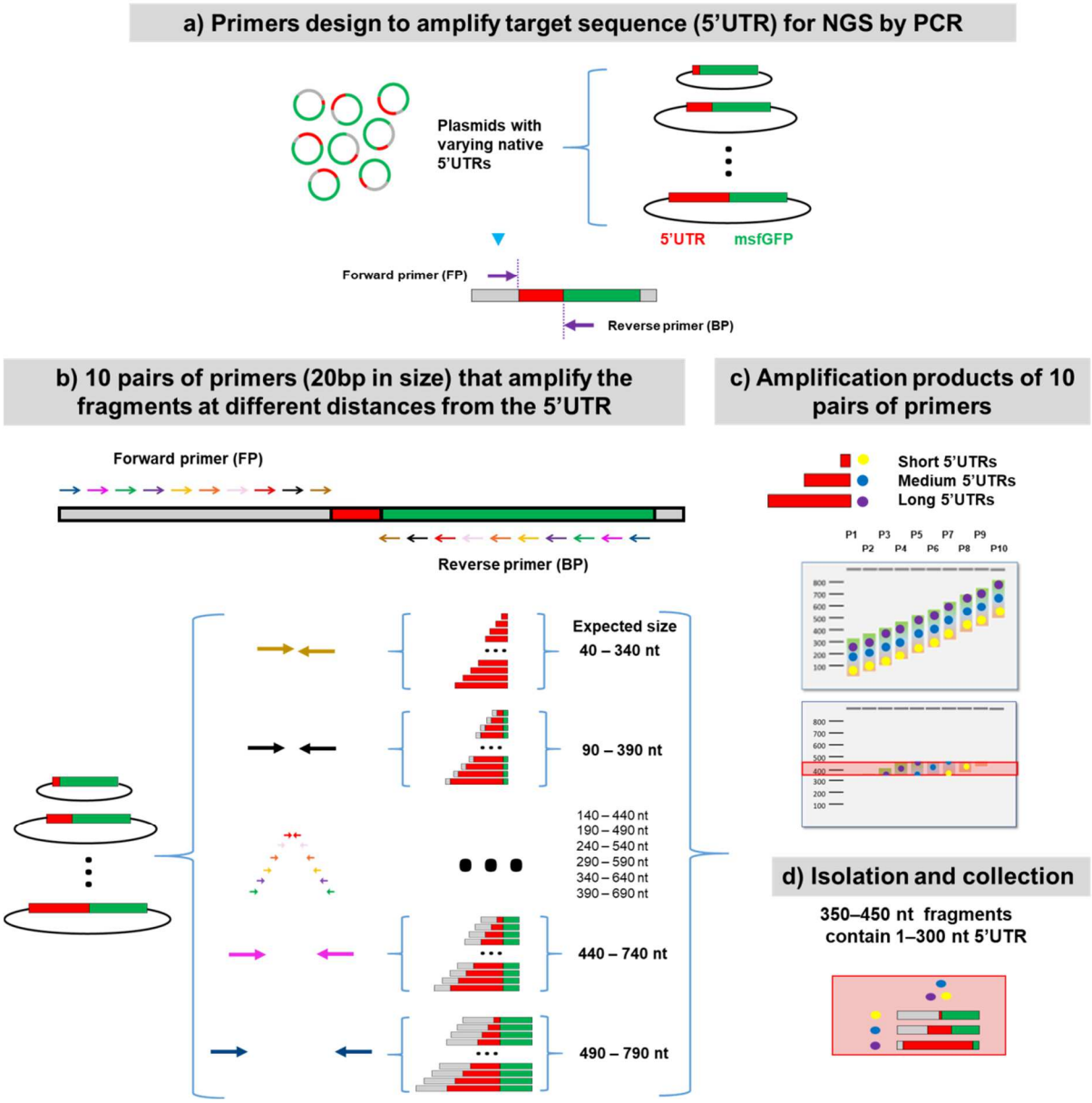


Figure 8. Strategy to re-homogenize the size of DNA fragments for sequencing. The red band represents the 5'UTRs in the library, the green band represents the msfGFP gene, and the grey one represents other sequences of the plasmid. Arrows in different colors display the 10 pairs of primers that amplify the fragments at different distances from the 5'UTR. Each dsDNA obtained after amplification contains an intact 5'UTR. Yellow dots represent short 5'UTRs, blue dots represent medium size 5'UTRs, and purple dots represent long 5'UTRs.

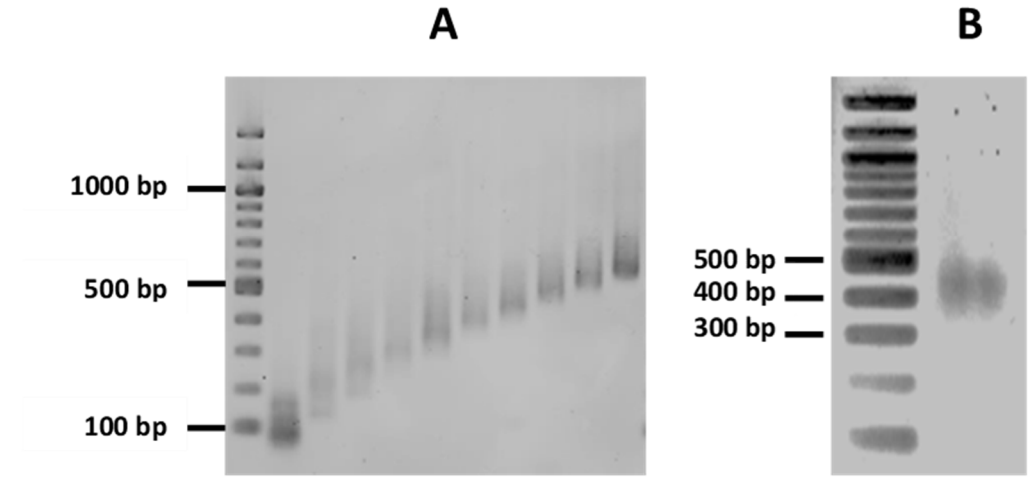


Figure 9. Size distribution of dsDNA fragments amplified from window 1 sample of M9_37°C. A) Ten pairs of primer that amplified the heterogeneous library; B) re-homogenized purified dsDNA.

5. Sequencing quality control and representativeness of the 5'UTRs in the library

Sequencing, read trimming, alignment, and counting of the 24 samples were performed by the TBI GeT platform as described in the Materials and Methods section. ***Table 2*** shows the total read number per window and per condition and the sum per growth condition. The minimum sum of reads was obtained for M9_20°C (around 175 000) and corresponded in theory to ~70 reads per 5'UTR sequence. The minimum number of reads in a window was above 22 000 and was found for window 3 of M9_20°C. Assuming that all the 2547 5'UTRs were present in this sample, this would represent in theory more than 8 reads per 5'UTR. In reality it should be higher because the 5'UTRs were spread across the 6 windows. We can therefore conclude that sequencing was deep enough to allow relevant analyses.

Table 2. Total read number in each window of each growth condition

	W1	W2	W3	W4	W5	W6	SUM
M9_37°C	24626	24626	50388	25039	29164	43147	196329
M9_42°C	46716	59405	30819	37660	39808	54059	268467
M9_20°C	33436	31548	22493	25924	30992	30642	175035
LB_37°C	33428	27474	32830	28812	26211	27760	176515

We then analyzed whether there might be a technical bias that could skew the comparisons, such as the representativeness of the library. For all 5'UTRs, we plotted the number of reads versus the length for each of the four conditions (***Figure 10***). We observed that the maximum number of reads decreased

with increasing 5'UTR length, indicating that the shorter the 5'UTR, the easier it was to obtain reads. We have not identified the cause of this distribution, perhaps the process of library construction or the re-homogenization of the 5'UTRs. The most important facts was that all four conditions showed similar distribution, indicating no significant differences between conditions and in this study we did not need to compare the number of reads between different 5'UTRs.

In addition, we mapped the number of reads per sequenced 5'UTRs as a function of their length (***Figure 11***). When comparing the length distribution of the sequenced 5'UTRs with the one of the synthesized 5'UTRs (***Figure 1*** and in insert), we found the same shape and no major differences between the four conditions. We did not obtain any sequenced 5'UTRs shorter than nine nucleotides in any of the four groups, although 36 such 5'UTRs were synthesized. This was not surprising, and expected as internal control, as a minimal length of UTR is required to sustain translation initiation by the presence of SD sequence.

We can conclude that our sequencing data seem representative of the library we aimed to clone, that the representativeness seems similar at the global scale between the four conditions and that no major technical bias has been introduced during the construction procedures.

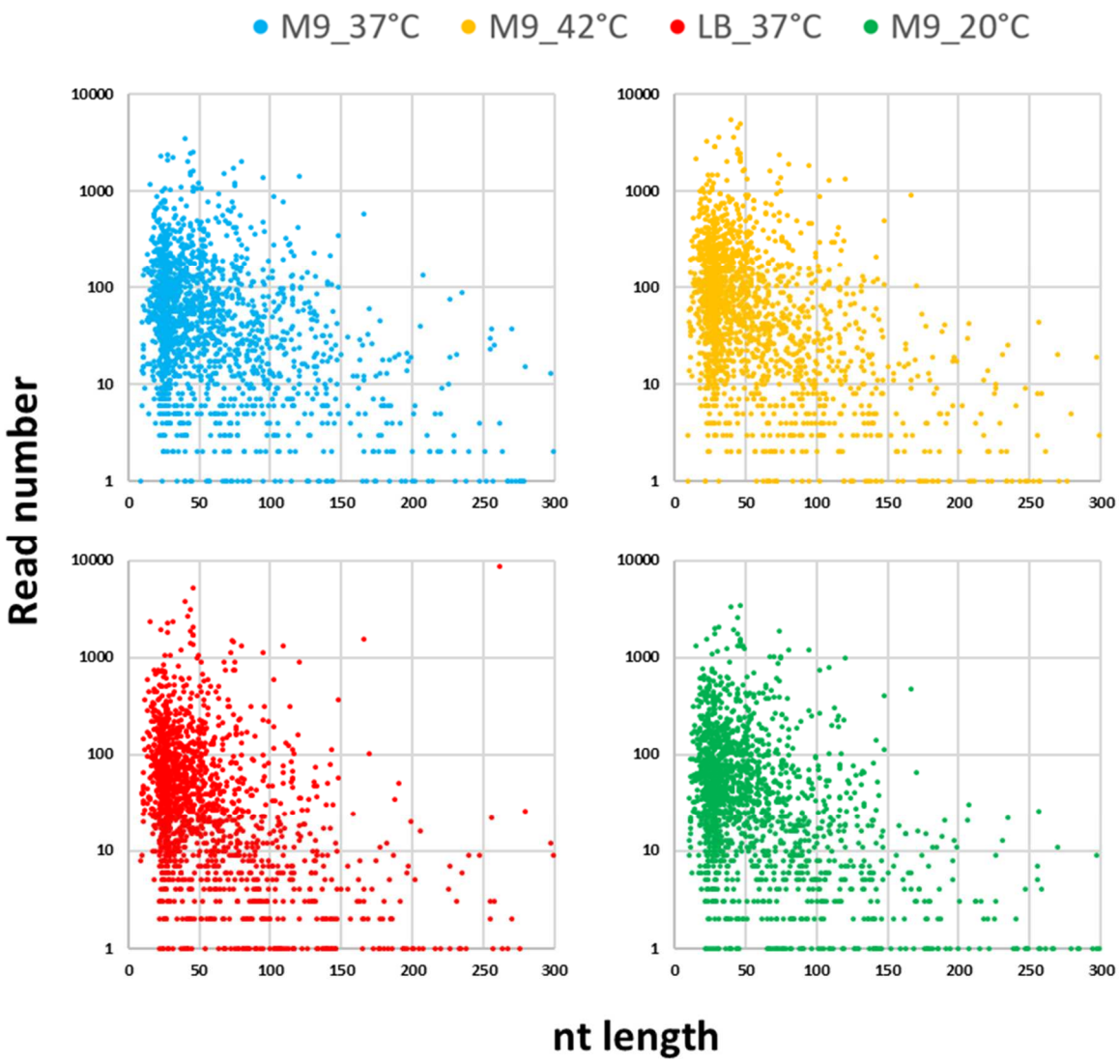


Figure 10. Distribution of read numbers of each 5'UTR as a function of length in four growth conditions. Each point represents a sequenced 5'UTR, the abscissa is its nucleotide length, and the ordinate is the number of reads of this 5'UTR obtained by sequencing. Four subgraphs with different colors stand for different growth conditions.

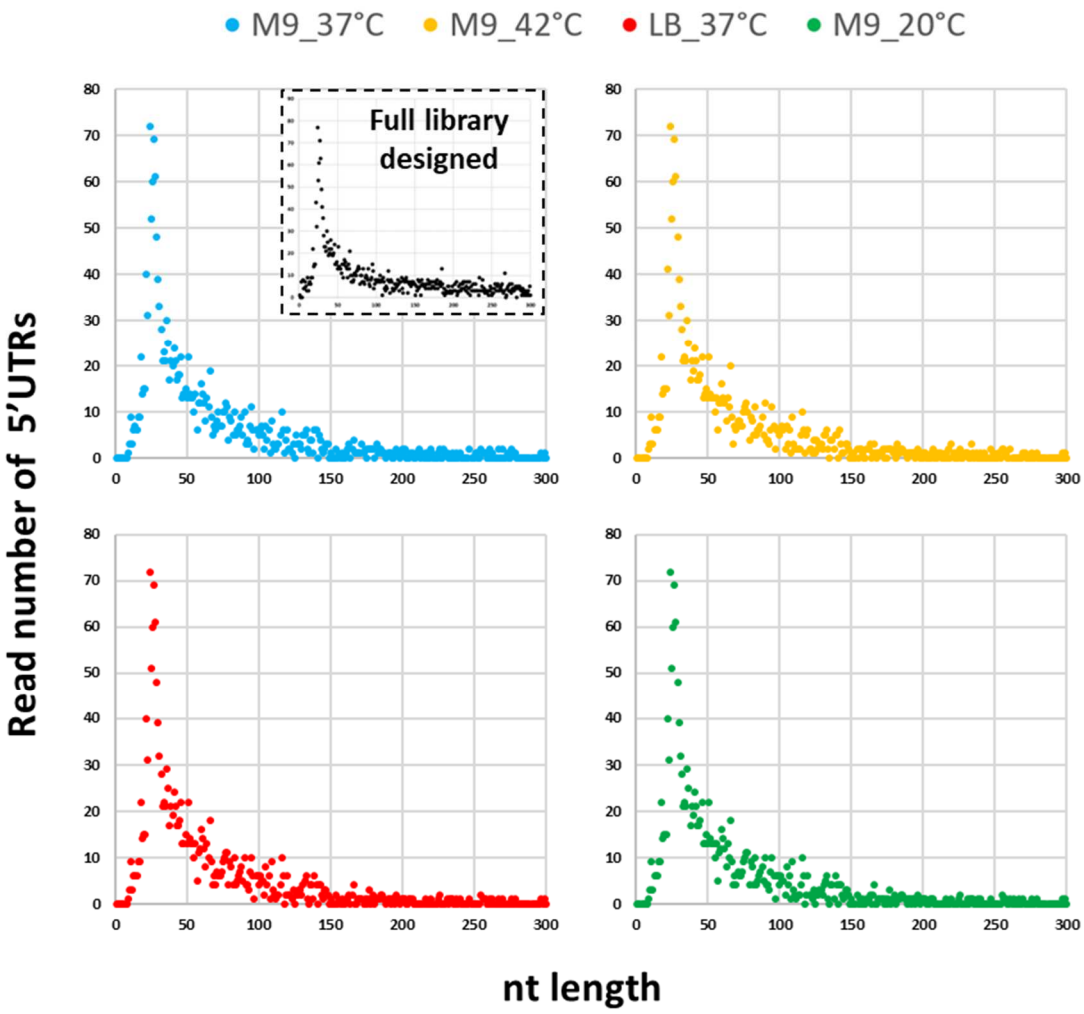


Figure 11. Number of the reads of 5'UTRs according to their length in four growth conditions. Each dot represents the read number of a sequenced 5'UTR. Four subgraphs with different colors stand for different growth conditions. The length distribution of the synthesized 2547 5'UTRs (the one of **Figure 1**) is reminded in the top left insert.

6. Location of 5'UTRs in the six windows under each growth condition

First, we wanted to identify in which window (W1-6) each 5'UTR was located under a specific growth condition using the read number information. When looking at the read number distribution over the windows, we noticed that 5'UTRs sometimes appeared in one single window but were also often present in adjacent and sometimes in non-adjacent multiple windows (***Figure 12***). For instance, reads of *fadL* 5'UTR were only present in window six, so the *fadL* 5'UTR positioning in window 6 was clear. In contrast, positioning of *maeA* and *msbA* 5'UTR was less obvious with almost equal read numbers in adjacent windows 4 and 5 for *maeA* and in non-adjacent windows 1, 3 and 4 for *msbA.* Therefore, to

consider all possible read number distributions, we developed a 4-step strategy to position the peak of 5'UTR read number within the six windows (***Figure 13***).

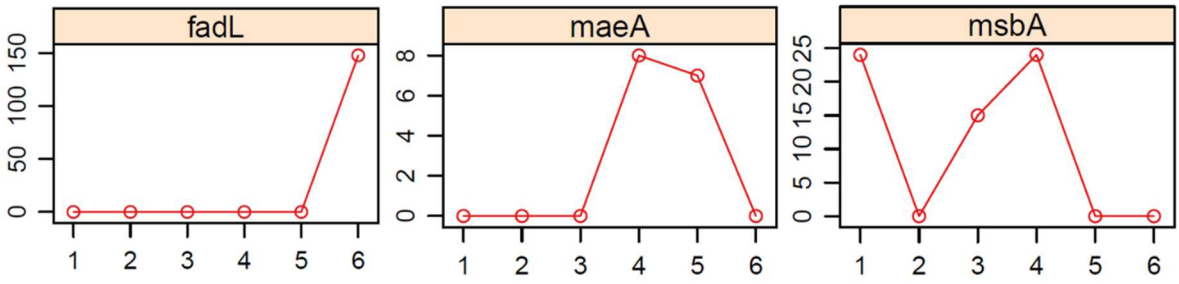


Figure 12. Examples of distribution of 5'UTR read number in the six windows (M9_37°C). The abscissa represents the six windows, and the ordinate displays the number of reads of the 5'UTR in each window.

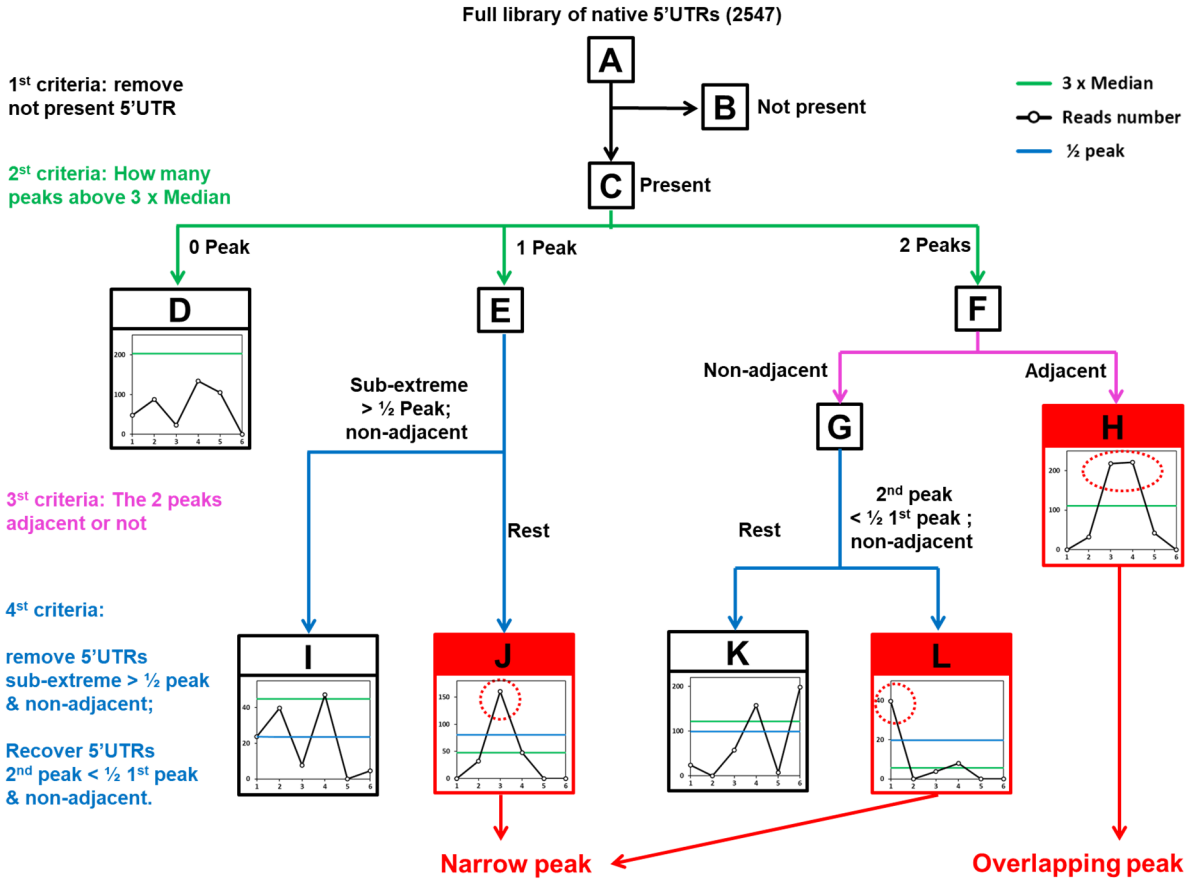


Figure 13. Strategy to position 5'UTRs in the 6 windows. Examples of 5'UTR read number distribution are shown in each step.

- The 1st step was to identify which 5'UTRs were present or not in a given growth condition. The full list of 2547 native 5'UTRs (***Figure 13, A***) was separated into two lists: 5'UTRs not present when no reads were mapped to the 5'UTR sequence (***Figure 13, B***); and 5'UTRs present when reads could be mapped to the 5'UTR sequence (***Figure 13, C***).

- The 2nd step was to classify the present 5'UTRs by their number of reads above the 3 times the median. Three classes were defined, 5'UTRs with 0 read number above 3 times the median, so 5'UTRs were declared with no peak of read number (***Figure 13, D***); 5'UTRs with 1 read number above 3 times the median, so 5'UTRs were declared with 1 peak of read number (***Figure 13, E***); 5'UTRs with 2 read numbers above 3 times the median, so 5'UTRs were declared with 2 peaks of read number (***Figure 13, F***). Due to the chosen threshold, no more than 2 read numbers were observed above 3 times the median in all 5'UTRs.

- The 3rd step was to identify the relative position of the 2 peaks in the case of group F. The 5'UTRs were further separated into 2 lists: when the 2 peaks were adjacent (***Figure 13, H***) and when the 2 peaks were not adjacent (***Figure 13, G***).

- The 4th step was implemented to refine the methods by considering particular read number distributions. We observed that in 5'UTRs with 1 peak (***Figure 13, E***), a second read number (sub-extreme read number) non-adjacent could be relatively high, decreasing the accuracy of the 5'UTR positioning. So we introduced a new criterion to remove 5'UTRs with a sub-extreme read number above half the highest read number (***Figure 13, I***) to only keep 5'UTRs with no high sub-extreme read number (***Figure 13, J***). In the same way, we also found in the group of 5'UTRs with 2 non-adjacent peaks (***Figure 13, G***) that a sub-extreme read number could be relatively high compared to the highest read number (***Figure 13, K***). So we also implemented a criterion to remove 5'UTRs with a sub-extreme read number above half the value of the extreme read number (***Figure 13, K***) to only keep 5'UTRs with no high sub-extreme read number (***Figure 13, L***).

After the 4-step strategy of peak positioning, all the 2547 native 5'UTRs were finally categorized into 7 lists named ***B, D, H, I, J, L*** and ***K***. For the following analyses, we kept the 5'UTRs with well-positioned peaks of read number corresponding to lists ***J*** and ***L*** of a narrow peak in one window and list ***H*** of a large peak overlapping two adjacent windows.

For instance, when we applied the 4-step strategy of peak positioning to the reference condition M9 medium at 37°C, we obtained the following distribution of 5'UTRs in the different lists: ***List M9_37°C_B*** included 810 5'UTRs not present in this growth condition. ***List M9_37°C_D*** included 226 5'UTRs with 0 peaks. ***List M9_37°C_H*** included 892 5'UTRs with a large peak overlapping two windows. ***List M9_37°C_I*** included 10 5'UTRs with a single peak but a high non-adjacent sub-extreme peak. ***List M9_37°C_J*** and ***List M9_37°C_L*** included 549 and 39 5'UTRs with a narrow peak and no high non-adjacent sub-extreme peak, respectively. ***List M9_37°C_K*** included 5 5'UTRs with 2 non-adjacent peaks but a high non-adjacent sub-extreme peak. The list **J+L+H** of selected 5'UTRs with well-positioned peaks of read counts included 1480 5'UTRs in M9 medium at 37°C.

7. Number of selected 5'UTRs

The 4-step strategy of peak positioning was applied to the four growth conditions: M9_37°C, M9_42°C, M9_20°C and LB_37°C. ***Table 3*** summarizes the number of 5'UTRs of some key lists. The number of 5'UTRs not present *(List B*) in each growth condition was similar, around 33% of the total 5'UTRs.

Table 3. Number of 5'UTRs in each list in the four growth conditions

	Full list	**Not present**	**Narrow peak**		**Overlapping peak**	**Selected**
	List A	List B	List J	List L	List H	List J+L+H
M9_37℃	2547	810	549	39	892	1480
M9_42℃	2547	820	549	44	895	1488
M9_20℃	2547	851	448	32	413	894
LB_37℃	2547	891	558	152	733	1443

The Venn diagram shows that 724 5'UTRs were never present in all four growth conditions (***Figure 14***). We can speculate that these 724 5'UTRs were not detected due to their absence at the end of the library construction. Indeed, we consider unlikely the hypothesis that the 5'UTRs were not detected because of a too weak fluorescence signal since we selected all the events from an extremely low fluorescence signal. Instead, the 5'UTRS were probably not present in the library due to troubles in the syntheses of ssDNA or dsDNA, the amplification step or the cloning itself. However, we cannot exclude that strong secondary structures in these 5'UTRs have impaired the library construction. We did not analyze these assumptions further.

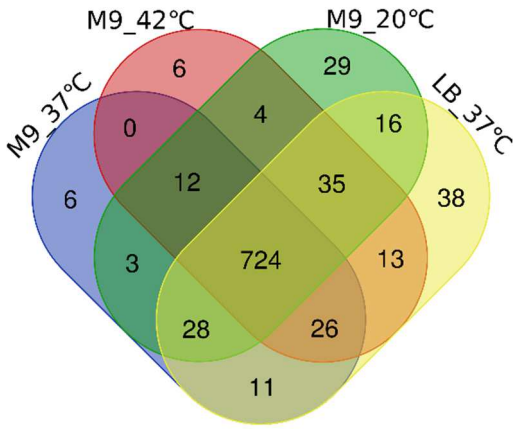


Figure 18. Venn diagram of 5'UTRs not present in each growth condition. Distribution of the 5'UTRS not detected across the four growth conditions M9_37°C (blue), M9_42°C (red), M9_20°C (green) and LB_37°C (yellow).

As mentioned above, only the 5'UTRs in list **J+L+H** with well-positioned peaks were selected for further analyses. The number of selected 5'UTRs (***Table 3***, list **J+L+H**) was similar for three conditions M9_37°C, M9_42°C and LB_37°C around 1470 (58% of the total 5'UTRs) but almost two-times lower for M9_20°C. Nevertheless, this number of selected 5'UTRs for M9_20°C was considered sufficient for the subsequent comparisons.

Then we wondered whether a minimum read number should be introduced for the selected 5'UTRs to be suitable for the subsequent comparative analysis. We plotted the percentage of 5'UTRs belonging to different ranges of read number in each window of each growth condition (***Figure 15***). The number of reads was divided into five sections 0-10 (light blue), 10-50 (orange), 50-100 (gray), 100-500 (light yellow), ≥500 (dark blue). The percentage of 5'UTRs corresponding to each read number section was plotted for each window in the four conditions. For instance, in the first window in M9_37°C, 7%, 23%, 14%, 40% and 16% of the 5'UTRs had between 0-10 reads, 10-50 reads, 50-100 reads, 100-500 reads and >500 reads, respectively. This figure shows that overall 90% of the selected 5'UTRs had greater than ten reads in each window of the four growth conditions. Therefore, we concluded that there was no need to set a minimum threshold of read number for the selected 5'UTRs.

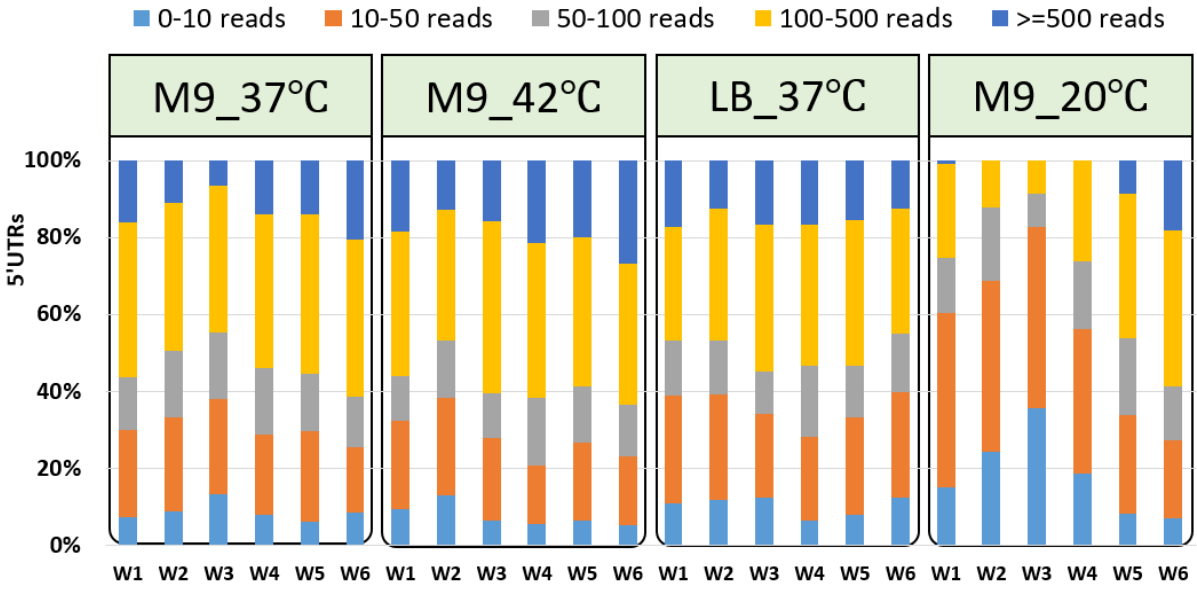


Figure 15. Percentage of 5'UTRs belonging to different ranges of read number in each window of each growth condition. The abscissa represents the six windows under four growth conditions. The ordinate displays the percentage of 5'UTRs belonging to five ranges of read number (represented in five different colors) in each window. Light blue, orange, grey, yellow, and dark blue represent the proportion of 5'UTRs with 0-10, 10-50, 50-100, 100-500, and ≥ 500 reads, respectively.

8. Differential analysis of msfGFP expression between growth conditions

Comparison of the peak positions of each selected 5'UTR between different growth conditions can lead to two conclusions. When the peaks shifted in different windows between conditions, this indicates that the 5'UTR contributed to msfGFP expression regulation. When the peaks remained in the same window regardless of the growth condition, this indicates that the 5'UTR was not involved in msfGFP expression regulation.

Pairwise comparison

We chose to study 5'UTR-mediated msfGFP expression regulation following environmental changes: growth temperature (20, 37 or 42 °C) and growth medium composition (M9 or LB). As shown in ***Figure 16***, we pairwise compared the peak position of 5'UTRs between two growth conditions using M9_37°C as the reference condition. When comparing the peak position of 5'UTRs in M9_42°C versus M9_37°C, we studied the temperature up-shift effect on 5'UTR-dependent msfGFP expression, whereas the temperature down-shift effect was investigated in the M9_20°C versus M9_37°C comparison. When comparing the peak position of 5'UTRs in M9_42°C versus M9_20°C, a sharp temperature up-shift effect was explored. On the other hand, comparing the peak position of 5'UTRs between the rich medium LB_37°C and the minimal medium M9_37°C provided the growth medium dependency of 5'UTR-related msfGFP expression.

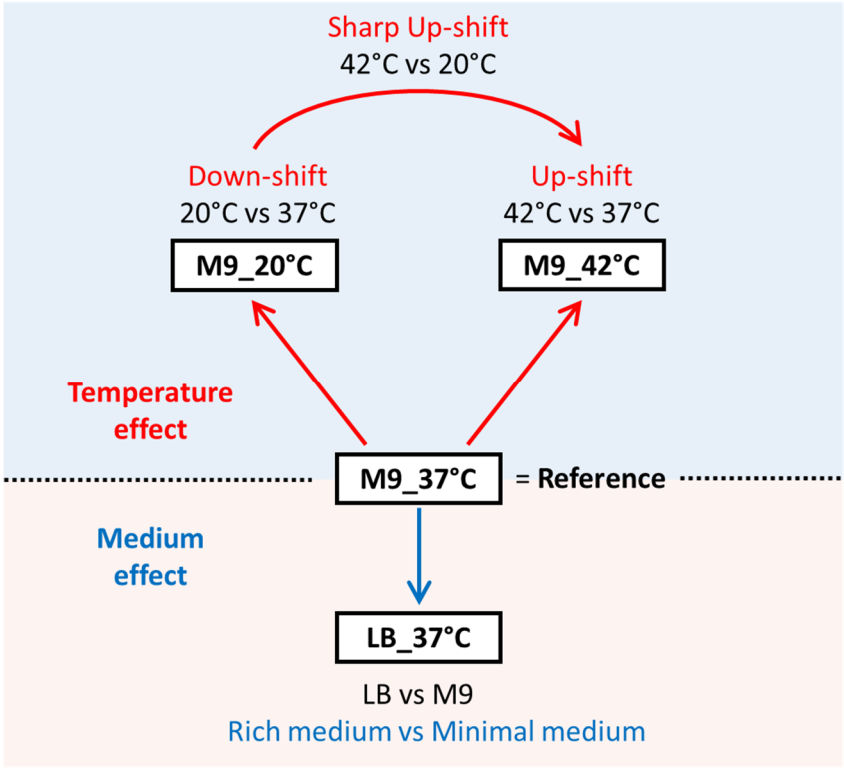


Figure 16. Pairwise comparisons of 5'UTR peak positions in different growth conditions.

To be able to identify the shift of 5'UTR peak position in a pairwise comparison, we decided to compare narrow and overlapping peak positions separately. Examples of these comparisons are shown in ***Figure 17***. We determined three patterns of 5'UTR peak positioning:

i) the 5'UTR peak remains in the same window upon a change in growth conditions, this is the constant pattern. It is based on two types of comparisons, narrow peak with narrow peak and overlapping peak with overlapping peak. Examples of constant 5'UTRs (***Figure 17***) are the 5'UTR of *fdhD* with a narrow peak positioned in window 1 in both the reference and test conditions and the 5'UTR of *tomB* with an overlapping peak positioned in windows 1 and 2 in both conditions.

ii) the 5'UTR peak up slides to higher windows upon a change in growth conditions, this is the up-regulated pattern. It includes three types of the peak position comparison, narrow peak with narrow peak, overlapping peak with overlapping peak and narrow peak with overlapping peak. Examples of up-regulated 5'UTRs (***Figure 17***) are the 5'UTR of *fadH* with a narrow peak positioned in window 1 in the reference condition, but in window 4 in the test condition the 5'UTR of *lptD* with an overlapping peak positioned in windows 2 and 3 in the reference condition but in windows 3 and 4 in the test condition, and the 5'UTR of *lhr* with a narrow peak positioned in window 3 in the reference condition but with an overlapping peak positioned in windows 3 and 4 in the test condition.

iii) the 5'UTRs peak down slides to lower windows upon a change in growth conditions, this is the down-regulated pattern. It is based on the same position comparisons as those used for the up-regulated pattern. Examples of down-regulated 5'UTRs (***Figure 17***) are the 5'UTR of *argG* with a narrow peak positioned in window 3 in the reference condition but in window 2 in the test condition, the 5'UTR of *rnt* with an overlapping peak positioned in windows 5 and 6 in the reference condition but windows 4 and 5 in the test condition, and the 5'UTR of *sbcB* with a narrow peak positioned in window 3 in the reference condition and an overlapping peak positioned in windows 1 and 2 in the test condition.

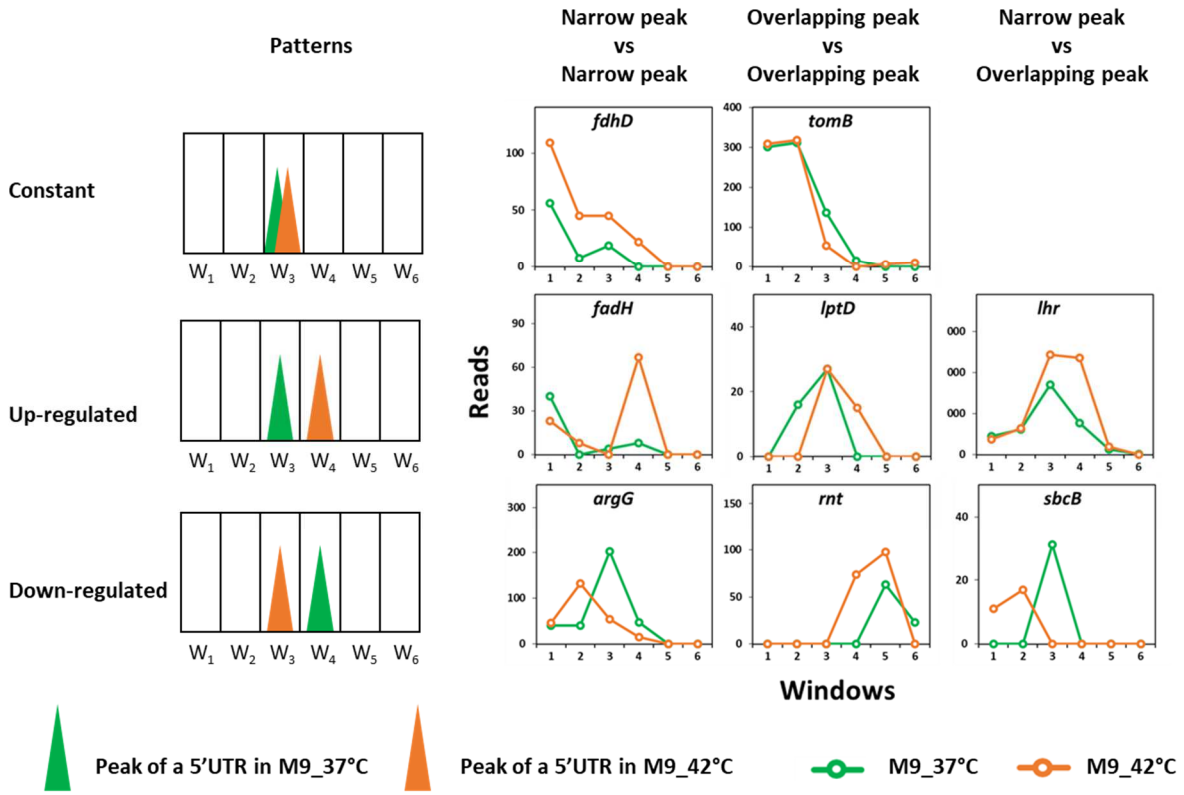


Figure 17. Examples of comparisons of narrow and overlapping peak positions in the test condition (M9_42°C) versus the reference condition (M9_37°C).

Effect of a temperature up-shift: M9_42°C versus M9_37°C

To investigate the effect of a temperature up-shift, we compared the peak position of 1311 common 5'UTRs in M9_42°C versus M9_37°C. As shown in ***Table 4***, 57 % of the 5'UTRs (755/1311) belonged to the constant pattern, demonstrating that these 5'UTRs did not affect the msfGFP expression when the growth temperature up-shifts from 37 to 42°C. 248 5'UTRs belonged to the up-regulated pattern, meaning that these 5'UTRs up-regulated the msfGFP expression after a temperature up-shift. Among them, *emrE* 5'UTR, whose peak up shifted over two windows, strongly up-regulated the msfGFP expression. 308 5'UTRs belonged to the down-regulated pattern, indicating that those 5'UTRs down-regulated the msfGFP expression when the growth temperature up-shift. Among them, 4 5'UTRs, whose peak down shifted over 2 windows were related to a strong down-regulated msfGFP expression.

Table 4 Number of 5'UTRs classified in the different patterns upon a temperature up-shift from 37 to 42°C

Constant	Up-regulated	Down-regulated
755	248 strongly up-regulated: *emrE*	308 strongly down-regulated : *ynfM, yqhG, tyrS, ydW*

To identify a putative correlation between the 5'UTR-dependent regulation and the function of the 5'UTR-associated gene, we performed a functional enrichment analysis using Gene Ontology (GO) database on the associated genes of the 5'UTRs associated to each pattern (***Table 5***). Globally, only a few functional categories were enriched in each of the 5'UTR patterns.

In the constant 5'UTRs, the functional categories *Transposition and DNA recombination* were enriched with 27 and 35 associated genes, respectively. It should be noted that the two sets of genes contained the same 26 *ins* genes, which are insertion sequences involved in transposition.

In the up-regulated 5'UTRs, two functional categories were enriched: *Response to heat* and *Base-excision repair*. The *mug* and *ung* genes of this last category are involved in DNA repair, and *mug* is reported to act in stress conditions (Mokkapati et al., 2001). The finding of the enrichment of the category *Response to heat* is interesting. It is associated to 7 genes, *tig* (trigger factor), *moaA* (GTP 3',8'-cyclase), *rhlE* (ATP-dependent RNA helicase), *yccX* (acylphosphatase), *lsrR* (DNA-binding transcriptional repressor), *clpB* (chaperone) and *loiP* (metalloprotease). Expression of *loiP* and *clpB* are known to be regulated at the transcriptional level after a temperature up-shift. Expression of *loiP* encoding an outer membrane metalloprotease is up-regulated in response to heat shock stress (Huang et al., 2009). *clpB* encodes ClpB, a member of the heat-shock induced chaperone network (Alam et al.,

2021) that suppresses or reverses the aggregation of denatured proteins after heat stress (Mogk et al., 2003; Ziętkiewicz et al., 2004). Transcription of *clpB* is induced by the heat shock σ32 factor upon temperature upshift (Kitagawa et al., 1991). Expressions of *tig, moaA, rhlE, yccX* and *lsrR* genes also contribute to the heat shock response (Krisko et al., 2014). The authors proposed that their expression is regulated at the translational level using a codon usage more optimal for high temperatures. At the post-transcriptional level, *moaA* expression is also known to be regulated by a riboswitch in its 5'UTR that binds both the molybdenum cofactor as well as the CsrA protein (Patterson-Fortin et al., 2013; Regulski et al., 2008). The CsrA protein recognizes mRNA secondary structure often located in the 5'UTR (Dubey et al., 2005). Binding of CsrA to the *moaA* 5'UTR appears to be necessary, but not sufficient for activation of translation (Patterson-Fortin et al., 2013). The expression of *lsrR* is also regulated by the binding of the CsrA protein (Mitra et al., 2016). Our results indicate that the 5'UTRs of *moaA* and *lsrR* are directly involved in the up-regulated msfGFP expression in response to heat stress. Therefore, we propose a post-transcriptional mechanism of up-regulated gene expression based on 5'UTR secondary structure influenced by heat.

In the 5'UTRs with a down-regulated pattern after temperature up-shift, we obtained two enriched functional categories with only three associated genes related to degradation of protein (*cnoX*, *rssB* and *exbB*) and DNA (*sbcB, endA* and *hofM*). Only the expression of *cnoX* is already known to response to heat stress (Kthiri et al., 2008).

Table 5. Functional category enrichment of 5'UTR-associated genes in M9_42°C vs M9_37°C.

Term	Annotated[1]	Significant[2]	*p*-value[3]
Constant			
Transposition	39	27	1.90E-07
DNA recombination	58	35	9.90E-07
Up-regulated			
Response to heat	25	7	0.0086
Base-excision repair, AP site formation	2	2	0.0098
Down-regulated			
Regulation of protein stability	3	3	0.0016
DNA catabolic process	4	3	0.0058

[1] Number of genes annotated in the functional category.
[2] Number of genes of the functional category present in the selected 5'UTR pattern.
[3] Only significantly enriched functional categories (p-value < 0.01) are shown.

Effect of a temperature down-shift: M9_20°C versus M9_37°C

To study the effect of a temperature down-shift on 5'UTR-related regulation of gene expression, we compared the peak position of 794 common 5'UTRs in M9_20°C versus M9_37°C (***Table 6***). Upon a temperature down-shift from 37 to 20°C, 40% of the 5'UTRs (319/794) belonged to the constant pattern of msfGFP expression, 189 5'UTRs belonged to the up-regulated pattern of msfGFP expression and 286 5'UTRs belonged to the down-regulated pattern of msfGFP expression. Among them, the peak positions of 9 5'UTRs shifted over two windows identifying strongly up- and down-regulations of msfGFP expression when the temperature decreases from 37 to 20°C.

Table 6. Number of 5'UTRs classified in the different patterns upon a temperature down-shift from 37 to 20°C.

Constant	Up-regulated	Down-regulated
319	189 strongly up-regulated: *mdtH yceF fhlA clpS pepT slyD talB abpA sgbE*	286 strongly down-regulated: *ydiE pssA yhaJ rimO ibsB fliA ispG rplU ygbK*.

The results of the functional category enrichment of 5'UTR-associated genes in M9_20°C vs M9_37°C are shown in ***Table 7***. No functional category was significantly enriched in the genes associated with the 319 constant 5'UTRs. In the up-regulated 5'UTRs, two functional categories were enriched. The category *Positive regulation of DNA-templated transcription* is the process of activating or increasing the frequency, rate or extent of transcription initiation and was associated with 4 genes coding for DNA-binding transcriptional regulators (*nhaR, mntR, sdiA* and *fhlA*). The category *Homoserine biosynthetic process* includes only two genes, *thrA* (fused aspartate kinase/homoserine dehydrogenase 1) and *lysC* (aspartate kinase III). The 5'UTR of *lysC* contains a riboswitch that in absence of lysine binding favors translation initiation and inhibits mRNA decay via RNase E (Caron et al., 2012). We can speculate that the riboswitch conformation may be dependent on temperature and be more favorable to gene expression at 20°C than at 37°C. In the down-regulated 5'UTRs, the category *Cellular response to hydrogen peroxide* was enriched associated with four genes (*ychH, eamA, ygiW* and *yhcN*). The expressions of *ychH* (stress-induced protein) and *eamA* (cysteine/O-acetylserine exporter) are known to be regulated at the transcriptional level.

Table 7. Functional category enrichment of 5'UTR-associated genes in M9_20°C vs M9_37°C.

Term	Annotated[1]	Significant[2]	*p*-value[3]
Constant			
-	-	-	-
Up-regulated			
Positive regulation of DNA-templated transcription, Initiation	10	4	0.0028
Homoserine biosynthetic process	2	2	0.0044
Down-regulated			
Cellular response to hydrogen peroxide	6	4	0.0015

[1] Number of genes annotated in the functional category.
[2] Number of genes of the functional category present in the selected 5'UTR pattern.
[3] Only significantly enriched functional categories (p-value < 0.01) are shown.

Effect of a sharp temperature up-shift: M9_42°C versus M9_20°C

To consider now the effect of a sharp temperature up-shift we compared the peak positions of 798 common 5'UTRs in M9_42°C versus M9_20°C (***Table 8***). Between 20 and 42 °C, 40 % of 5'UTRs (312/798) belonged to the constant pattern of msfGFP expression, 277 5'UTRs belonged to the up-regulated pattern and 209 5'UTRs belonged to the down-regulated pattern. Among them, 18 and 12 5'UTRs were associated with strong up- and down-regulations of msfGFP expression, respectively.

Table 8. Number of 5'UTRs classified in the different patterns after a sharp temperature up-shift from 20 to 42 °C.

Constant	Up-regulated	Down-regulated
	277	209
312	strongly up-regulated: *ycbZ ydiE ygbK thrL yajD abgR ybaV emrE fadH ydcF fliA glpA rimO ypeA ispG gss rplU azuC*	strongly down-regulated: *dps mdtH fhlA cmtB gspB valS clpS csiD tyrS yqcE lysC cysQ*.

The results of functional category enrichment of 5UTR-associated genes in M9_42°C vs M9_20°C are shown in ***Table 9***. When we compared the response to a sharp increase from 20 to 42°C with moderate shifts (37 to 42°C and 20 to 37 °C), we observed that the category *Response to heat* was not enriched anymore while *Homoserine biosynthetic* pathway and *Response to oxidative stress* were still present. Among the genes associated with the response to oxidative stress, the 5'UTR of the *iraD* gene encoding

an inhibitor of σ^S proteolysis is known to bind the CsrA protein to negatively regulated translation of the downstream gene (Park et al., 2017). A new stress response, *Cellular hyperosmotic response*, was enriched in the up-regulated 5'UTR pattern. This category was associated to the genes *proQ* (RNA chaperone), *osmF* (ABC transporter) and *treF* (trehalase). We have no clear interpretation for the enrichment of the functional category *Lactate metabolic process* in the constant 5'UTR pattern between 20 and 42 °C.

Table 9. Functional category enrichment of 5'UTR-associated gene in M9_42°C vs M9_20°C.

Term	Annotated[1]	Significant[2]	*p*-value[3]
Constant			
Lactate metabolic process	8	4	0.0086
Up-regulated			
Cellular hyperosmotic response	4	3	0.0039
Cellular response to oxidative stress	19	7	0.0093
Down-regulated			
Homoserine biosynthetic process	2	2	0.006

[1] Number of genes annotated in the functional category.
[2] Number of genes of the functional category present in the selected 5'UTR pattern.
[3] Only significantly enriched functional categories (p-value < 0.01) are shown.

Effect of the growth medium: LB versus M9

To analyze the effect of the growth medium composition on 5'UTR-mediated regulation of msfGFP expression, we compared the peak positions of 1235 common 5'UTRs in the rich LB medium versus the minimal M9 medium at 37°C. As shown in ***Table 10,*** when comparing LB versus M9, 40% of 5'UTRs (507/1235) belonged to the constant 5'UTR pattern of msfGFP expression, 433 5'UTRs belonged to the up-regulated 5'UTR pattern, and 295 5'UTRs belonged to the down-regulated 5'UTR pattern. Among them, 14 and 7 5'UTR exhibited peaks up shifting over two windows.

Table 10. Number of 5'UTRs classified in the different patterns in LB versus M9 at 37°C.

Constant	Up-regulated	Down-regulated
507	433 strongly up-regulated: *yceF rimL ydiV cspE ydcF sdiA yjaB fldA pepN minC malI pdeN fbp cmtB*.	295 strongly down-regulated: *ybdO yqhG entC yobA lrp rluE nfeR*

The results of functional category enrichment of 5'UTR-associated genes in LB vs M9 are shown in ***Table 11***. In the constant 5'UTR pattern, we found again an enrichment of the category *Transposition* associated to 22 *ins* genes as previously seen upon growth temperature up-shift. Two categories related to amino acid transport and associated to the same three genes (*ygaZ, livJ* and *yjeH*) were also enriched in the constant 5'UTR pattern. In the up-regulated 5'UTR pattern, the category *Regulation of transcription* was enriched corresponding to 42 DNA-binding transcriptional regulators among the 54 associated genes. We can speculate that in LB, 5'UTR-mediated regulation of these transcriptional regulators helps the cell to adapt to a large list of available carbon sources in the rich medium compared to the sole carbon source glucose in M9. In the down-regulated 5'UTR pattern, we found an enrichment of the category Regulation of cell growth associated to only three genes: *ybjN* encoding protein YbjN involved in the production of flagella and fimbriae, *prlF* which codes for an antitoxin component and *rpmA* encoding a 50S ribosomal subunit protein involved in the assembly of the 70S ribosome. Nothing is reported in the literature regarding 5'UTR-related expression regulation of these three genes.

Table 11. Functional category enrichment of 5'UTR-associated gene in LB vs M9 at 37°C.

Term	Annotated[1]	Significant[2]	*p*-value[3]
Constant			
Transposition	39	22	2.50E-07
Valine transport	3	3	0.0074
Branched-chain amino acid transport	3	3	0.0074
Up-regulated			
Regulation of transcription, DNA-templated	219	54	0.0065
Down-regulated			
Regulation of cell growth	4	3	0.0043

[1]Annotated: number of genes annotated in the functional category.
[2]Significant: number of genes of the functional category present in the selected 5'UTR pattern.
[3]Only significantly enriched functional categories (p-value < 0.01) are shown.

Four-group comparison

After pairwise comparisons that identified 5'UTRs contributing to gene expression regulation upon temperature and medium changes, we wanted to identify 5'UTRs that did not participate in gene expression regulation regardless of the growth conditions. Thus, we overlapped the 5'UTRs of the constant patterns (with no involvement in the regulation of msfGFP expression) identified in each growth condition (755 5'UTRs upon temperature up-shift, 319 5'UTRs upon temperature down-shift,

312 5'UTRs upon sharp temperature up-shift and 507 5'UTRs after a change in growth medium). As shown in the Venn diagram (***Figure 18***), 98 5'UTRs with a constant pattern in the four conditions tested, never participated in gene expression regulation. We checked the peak position of these 98 5'UTRs and found that 85 exhibited an overlapping peak located in windows 5 and 6 corresponding to high msfGFP expression. These results show that these 85 5'UTRs allow a constant high level msfGFP expression regardless of the tested growth conditions.

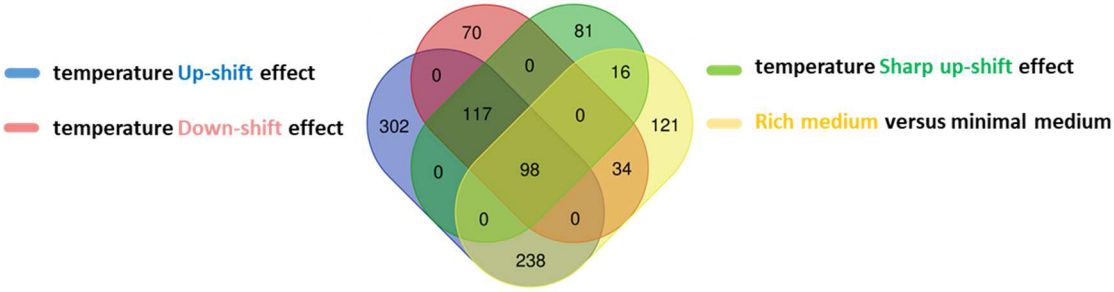


Figure 18. Venn diagram of 5'UTRs that did not participate in gene expression regulation. 5'UTRS that did not affect msfGFP expression following the effects of temperature up-shift (blue, M9_42°C vs M9_37°C), temperature down-shift (red, M9_20°C vs M9_37°C), temperature sharp up-shift (green, M9_42°C vs M9_20°C) and change in growth medium (yellow, LB_37°C vs M9_37°C).

Functional category enrichment in these 98 5'UTR-associated genes only identified a category related to cell motility (***Table 12)*** with three associated genes (*ycdX*, *flgB*, and *yfiR*).

Table 12. Functional category enrichment of associated genes to 98 5'UTRs that did not participate in gene expression regulation regardless of the growth conditions.

Term	Annotated[1]	Significant[2]	*p*-value[3]
Bacterial-type flagellum-dependent swarming motility	5	3	0.00046

[1] Number of genes annotated in the functional category.
[2] Number of genes of the functional category present in the selected 5'UTR pattern.
[3] Only significantly enriched functional categories (p-value < 0.01) are shown.

The Venn diagram also helped us to identify 5'UTRs which specifically did not response to temperature or growth medium composition: 117 5'UTRs belonged to the constant 5'UTR pattern only upon temperature effect and 121 5'UTRs belonged to the constant pattern only under growth medium effect (***Figure 18***).

9. Conclusions and perspectives

In this research, we have developed and constructed a full size library of native 5'UTR sequences from *E. coli* fused to a fluorescent reporter gene. This library aims to better understand the role of 5'UTRs in gene regulation in this bacterium, in a physiological context. The library was challenged to grow under different environmental conditions, in terms of temperature and medium composition, and we analyzed the expression level of the msfGFP fluorescent protein. Cells were sorted based on their individual fluorescence level in six windows, from very low to high fluorescent protein expression level. The identity of 5'UTRs in each window, in each growth condition, was determined by NGS.

We have shown that a large range of fluorescence was covered by our library, confirming the role of 5'UTRs in gene expression regulation. In addition, we identified that some 5'UTRs can either up-regulate or down-regulate msfGFP expression levels when challenged to specific growth conditions. We correlated 5'UTR-related abilities of expression regulation under different growth conditions with the function of the 5'UTR-associated genes. We have identified the enriched category *Response to heat* in the up-regulated msfGFP expression upon a temperature upshift, showing a direct 5'UTR-mediated response to the imposed stress. We found enriched categories related to other stress like oxidative and osmotic stress, it is a possible indication of stress cross-responses. Global cellular processes such as transcription, DNA repair and degradation, protein degradation and cell growth were also enriched in response to temperature and/or medium effects. Together, these results show that at the physiological level, these genes have part of their regulation carried by the 5'UTR sequence, independently of the original coding sequence and promoter.

Obviously, 5'UTRs are involved in gene expression regulation by the presence within their sequence of the SD sequence controlling translation initiation. The next step is to further analyze the 5'UTR sequences according to the msfGFP expression windows in which they were detected. This work will consist of an in-depth analysis of the SD sequences present in each window, to analyze on a global scale which SDs are more prone to translation. We expect to cluster highly efficient SD sequences in high msfGFP expression windows and inversely, unfavorable SD sequences in low expression windows. In addition, we will analyze the link between SD sequences and their efficiency of translation initiation as a function of the growth environment.

It was often difficult to link the 5'UTR-mediated regulation to a known post-transcriptional mechanism located in the 5'UTR. Nevertheless, the presence in some 5'UTRs of riboswitch (*moaA* and *lysC*) and CsrA binding sites (*moaA, lsdR* and *iraD*) was known and could explain in these cases the observed

regulations. Understanding in more details the mechanisms of the 5'UTR-dependent regulation will require an in-depth investigation on the 5'UTR sequence characteristics such as folding, length, presence of sequence motifs for binding of regulatory proteins and sRNAs, etc. Search for statistical correlations between these sequence parameters and the msfGFP expression regulation will help to define 5'UTR-related determinants of gene expression relation upon temperature shift or changes in medium composition.

In this research, we studied gene expression regulation by following the msfGFP protein level, but we did not have time to integrate the other levels of gene expression regulation, mainly at the mRNA level. It needs to be clarified whether the ability of the different 5'UTR-dependent expression regulations could be also related to the control of transcription and/or mRNA stability. To do so, we will perform transcriptomic experiments to analyze the concentration and lifetime of msfGFP mRNAs from the whole library under different growth conditions. If changes in msfGFP protein level are not associated with changes in mRNA levels, we will be able to conclude that the msfGFP expression was regulated at the translational level only. In contrast, if changes in msfGFP protein level correlate with changes in mRNA levels, we will conclude that the msfGFP expression was regulated at the mRNA level. In this case, 5'UTR-dependent regulation of transcription and/or mRNA stability should be involved. Combination of protein level determinations as performed in this work with future quantifications of mRNA concentrations and stabilities will increase our mechanistic knowledge of how native 5'UTRs contribute to the regulation of gene expression in *E. coli*.

Furthermore, this work has provided molecular tools in the fields of biotechnology and synthetic biology. On one hand, we identified 5'UTR sequences that can be used to control and optimize the expression of heterologous genes of interest when changing the growth temperature or medium composition. We can provide 5'UTRs that (strongly) up- and down-regulate target gene expression upon a specific shift in temperature or medium composition. Furthermore, 5'UTRs can be selected to be sensitive to one stimulus, temperature for example and not to another one, medium change in our case. On the other hand, it can be useful in synthetic biology to have 5'UTRs that lead to constant gene expression regardless of the growth conditions. This study provided 85 5'UTRs that lead to high gene expression regardless of the growth temperature and medium composition. It would be very interesting to challenge our native 5'UTR library in other growth conditions (other stress and carbon sources for instance) to identify 5'UTRs which lead to constitutive gene expression whatever the growth condition. These 5'UTRs would constitute new molecular tools to increase the robustness of a protein production process in different environments. Finally, under a given growth condition, we ranked all the native 5'UTRs according to the expression level of the gene they control. This provides

a large repertoire of regulatory sequences that can be used for custom modulation at the post-transcriptional level of the expression of any gene, or pathway, of interest.

This study provided new insights in the post-transcriptional regulation of gene expression in *E. coli.* 5'UTR-dependent regulations directly and efficiently alter the expression levels of the downstream gene and thus contribute to bacterial adaption to changing environments.

REFERENCES

Alam, A., Bröms, J. E., Kumar, R., & Sjöstedt, A. (2021). The Role of ClpB in Bacterial Stress Responses and Virulence. *Frontiers in Molecular Biosciences*, *8*, 668910. https://doi.org/10.3389/fmolb.2021.668910

Bartholomäus, A., Fedyunin, I., Feist, P., Sin, C., Zhang, G., Valleriani, A., & Ignatova, Z. (2016). Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2063), 20150069. https://doi.org/10.1098/rsta.2015.0069

Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., & Lafontaine, D. A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences*, *109*(50), E3444–E3453. https://doi.org/10.1073/pnas.1214024109

Carpousis, A. J., Luisi, B. F., & McDowall, K. J. (2009). Endonucleolytic initiation of mRNA decay in Escherichia coli. *Progress in Molecular Biology and Translational Science*, *85*, 91–135. https://doi.org/10.1016/S0079-6603(08)00803-9

Chiaruttini, C., & Guillier, M. (2020). On the role of mRNA secondary structure in bacterial translation. *WIREs RNA*, *11*(3). https://doi.org/10.1002/wrna.1579

Del Campo, C., Bartholomäus, A., Fedyunin, I., & Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLOS Genetics*, *11*(10), e1005613. https://doi.org/10.1371/journal.pgen.1005613

DUBEY, A. K., BAKER, C. S., ROMEO, T., & BABITZKE, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA–RNA interaction. *RNA*, *11*(10), 1579–1587. https://doi.org/10.1261/rna.2990205

Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A. J., Girbal, L., & Cocaign-Bousquet, M. (2014). Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Research*, *42*(4), 2460–2472. https://doi.org/10.1093/nar/gkt1150

Heyduk, E., & Heyduk, T. (2018). DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Research*, *46*(9), 4469–4486. https://doi.org/10.1093/nar/gky172

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., & Groisman, E. A. (2012). Riboswitch control of Rho-dependent transcription termination. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14), 5376–5381. https://doi.org/10.1073/pnas.1112211109

Huang, L.-H., Wang, C.-Z., & Kang, L. (2009). Cloning and expression of five heat shock protein genes in relation to cold hardening and development in the leafminer, Liriomyza sativa. *Journal of Insect Physiology*, *55*(3), 279–285. https://doi.org/10.1016/j.jinsphys.2008.12.004

Hudeček, O., Benoni, R., Reyes-Gutierrez, P. E., Culka, M., Šanderová, H., Hubálek, M., Rulíšek, L., Cvačka, J., Krásný, L., & Cahová, H. (2020). Dinucleoside polyphosphates act as 5′-RNA caps in bacteria. *Nature Communications*, *11*(1), 1052. https://doi.org/10.1038/s41467-020-14896-8

Kaberdin, V. R., & Bläsi, U. (2006). Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiology Reviews*, *30*(6), 967–979. https://doi.org/10.1111/j.1574-6976.2006.00043.x

Kaminishi, T., Wilson, D. N., Takemoto, C., Harms, J. M., Kawazoe, M., Schluenzen, F., Hanawa-Suetsugu, K., Shirouzu, M., Fucini, P., & Yokoyama, S. (2007). A Snapshot of the 30S Ribosomal Subunit Capturing mRNA via the Shine-Dalgarno Interaction. *Structure*, *15*(3), 289–297. https://doi.org/10.1016/j.str.2006.12.008

Kim, D., Hong, J. S.-J., Qiu, Y., Nagarajan, H., Seo, J.-H., Cho, B.-K., Tsai, S.-F., & Palsson, B. Ø. (2012). Comparative Analysis of Regulatory Elements between Escherichia coli and Klebsiella pneumoniae by Genome-Wide Transcription Start Site Profiling. *PLoS Genetics*, *8*(8), e1002867. https://doi.org/10.1371/journal.pgen.1002867

Kitagawa, M., Wada, C., Yoshioka, S., & Yura, T. (1991). Expression of ClpB, an analog of the ATP-dependent protease regulatory subunit in Escherichia coli, is controlled by a heat shock sigma factor (sigma 32). *Journal of Bacteriology*, *173*(14), 4247–4253. https://doi.org/10.1128/jb.173.14.4247-4253.1991

Kortmann, J., & Narberhaus, F. (2012). Bacterial RNA thermometers: Molecular zippers and switches. *Nature Reviews. Microbiology*, *10*(4), 255–265. https://doi.org/10.1038/nrmicro2730

Krisko, A., Copic, T., Gabaldón, T., Lehner, B., & Supek, F. (2014). Inferring gene function from evolutionary change in signatures of translation efficiency. *Genome Biology*, *15*(3), R44. https://doi.org/10.1186/gb-2014-15-3-r44

Kthiri, F., Le, H.-T., Tagourti, J., Kern, R., Malki, A., Caldas, T., Abdallah, J., Landoulsi, A., & Richarme, G. (2008). The thioredoxin homolog YbbN functions as a chaperone rather than as an oxidoreductase. *Biochemical*

and Biophysical Research Communications, *374*(4), 668–672. https://doi.org/10.1016/j.bbrc.2008.07.080

Luciano, D. J., Levenson-Palmer, R., & Belasco, J. G. (2019). Stresses that raise Np4A levels induce protective nucleoside tetraphosphate capping of bacterial RNA. *Molecular Cell*, *75*(5), 957-966.e8. https://doi.org/10.1016/j.molcel.2019.05.031

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juárez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J., & Morett, E. (2009). Genome-Wide Identification of Transcription Start Sites, Promoters and Transcription Factor Binding Sites in *E. coli*. *PLoS ONE*, *4*(10), e7526. https://doi.org/10.1371/journal.pone.0007526

Mitra, A., Herren, C. D., Patel, I. R., Coleman, A., & Mukhopadhyay, S. (2016). Integration of AI-2 Based Cell-Cell Signaling with Metabolic Cues in Escherichia coli. *PLOS ONE*, *11*(6), e0157532. https://doi.org/10.1371/journal.pone.0157532

Mogk, A., Deuerling, E., Vorderwülbecke, S., Vierling, E., & Bukau, B. (2003). Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Molecular Microbiology*, *50*(2), 585–595. https://doi.org/10.1046/j.1365-2958.2003.03710.x

Mokkapati, S. K., Fernández de Henestrosa, A. R., & Bhagwat, A. S. (2001). Escherichia coli DNA glycosylase Mug: A growth-regulated enzyme required for mutation avoidance in stationary-phase cells. *Molecular Microbiology*, *41*(5), 1101–1111. https://doi.org/10.1046/j.1365-2958.2001.02559.x

Narberhaus, F., Waldminghaus, T., & Chowdhury, S. (2006). RNA thermometers. *FEMS Microbiology Reviews*, *30*(1), 3–16. https://doi.org/10.1111/j.1574-6976.2005.004.x

Nouaille, S., Mondeil, S., Finoux, A.-L., Moulis, C., Girbal, L., & Cocaign-Bousquet, M. (2017). The stability of an mRNA is influenced by its concentration: A potential physical mechanism to regulate gene expression. *Nucleic Acids Research*, *45*(20), 11711–11724. https://doi.org/10.1093/nar/gkx781

Nudler, E., & Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences*, *29*(1), 11–17. https://doi.org/10.1016/j.tibs.2003.11.004

Park, H., McGibbon, L. C., Potts, A. H., Yakhnin, H., Romeo, T., & Babitzke, P. (2017). Translational Repression of the RpoS Antiadapter IraD by CsrA Is Mediated via Translational Coupling to a Short Upstream Open Reading Frame. *MBio*, *8*(4), e01355-17. https://doi.org/10.1128/mBio.01355-17

Patterson-Fortin, L. M., Vakulskas, C. A., Yakhnin, H., Babitzke, P., & Romeo, T. (2013). Dual posttranscriptional regulation via a cofactor-responsive mRNA leader. *Journal of Molecular Biology*, *425*(19), 3662–3677. https://doi.org/10.1016/j.jmb.2012.12.010

Picard, F., Dressaire, C., Girbal, L., & Cocaign-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. *Comptes Rendus Biologies*, *332*(11), 958–973. https://doi.org/10.1016/j.crvi.2009.09.005

Regulski, E. E., Moy, R. H., Weinberg, Z., Barrick, J. E., Yao, Z., Ruzzo, W. L., & Breaker, R. R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Molecular Microbiology*, *68*(4), 918–932. https://doi.org/10.1111/j.1365-2958.2008.06208.x

Richards, J., & Belasco, J. G. (2019). Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Molecular Cell*, *74*(2), 284-295.e5. https://doi.org/10.1016/j.molcel.2019.01.044

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., & Nudler, E. (2016). SRNA-Mediated Control of Transcription Termination in *E. coli*. *Cell*, *167*(1), 111-121.e13. https://doi.org/10.1016/j.cell.2016.09.004

Sharma, A., Alajangi, H. K., Pisignano, G., Sood, V., Singh, G., & Barnwal, R. P. (2022). RNA thermometers and other regulatory elements: Diversity and importance in bacterial pathogenesis. *Wiley Interdisciplinary Reviews. RNA*, e1711. https://doi.org/10.1002/wrna.1711

Tietze, L., & Lale, R. (2021). Importance of the 5′ regulatory region to bacterial synthetic biology applications. *Microbial Biotechnology*, *14*(6), 2291–2315. https://doi.org/10.1111/1751-7915.13868

Tuller, T., & Zur, H. (2015). Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Research*, *43*(1), 13–28. https://doi.org/10.1093/nar/gku1313

Villa, J. K., Su, Y., Contreras, L. M., & Hammond, M. C. (2018). SYNTHETIC BIOLOGY OF SMALL RNAS AND RIBOSWITCHES. *Microbiology Spectrum*, *6*(3), 10.1128/microbiolspec.RWR-0007–2017. https://doi.org/10.1128/microbiolspec.RWR-0007-2017

Vimberg, V., Tats, A., Remm, M., & Tenson, T. (2007). Translation initiation region sequence preferences in Escherichia coli. *BMC Molecular Biology*, *8*(1), 100. https://doi.org/10.1186/1471-2199-8-100

Ziętkiewicz, S., Krzewska, J., & Liberek, K. (2004). Successive and Synergistic Action of the Hsp70 and Hsp100 Chaperones in Protein Disaggregation *. *Journal of Biological Chemistry*, *279*(43), 44376–44383. https://doi.org/10.1074/jbc.M402405200

Supplementary data

Table S1. Primers used in this study

Primer number	Sequence 5'-3'	Function Description
1103	ACTCTCTACTGTTTCTCCAT	Forward primer for dsDNA generation and amplification of the whole 2547 native 5'UTR library
1104	AGTTCTTCTCCTTTGCTCAT	Reverse primer for dsDNA generation and amplification of the whole 2547 native 5'UTR library
1085	ATGAGCAAAGGAGAAGAACTTTTCACTGGAG	Forward primer for amplifying the plasmid backbone from pMET 296 except the 5'UTR moiety
397	ATGGAGAAACAGTAGAGAGTTGCGATAAA	Reverse primer for amplifying the plasmid backbone from pMET 296 except the 5'UTR moiety
1151	ATCCTACCTGACGCTTTTTA	Forward primer to amplify the 5'UTR library for sequencing
1152	AAGAATTGGGACAACTCCAG	Reverse primer to amplify the 5'UTR library for sequencing
1153	GCATTTTTATCCATAAGATTAGC	Forward primer to amplify the 5'UTR library for sequencing
1154	CATTAACATCACCATCTAATTC	Reverse primer to amplify the 5'UTR library for sequencing
1155	CGGCGTCACACTTTGCTATG	Forward primer to amplify the 5'UTR library for sequencing
1156	TCTCCACGGACAGAAAATTT	Reverse primer to amplify the 5'UTR library for sequencing
1157	GAAAAGTCCACATTGATTATTTG	Forward primer to amplify the 5'UTR library for sequencing
1158	TCCGTTTGTAGCATCACCTT	Reverse primer to amplify the 5'UTR library for sequencing
1159	AACAAAAGTGTCTATAATCACG	Forward primer to amplify the 5'UTR library for sequencing
1160	TGCAAATAAATTTAAGGGTGAG	Reverse primer to amplify the 5'UTR library for sequencing
1161	GACCAAAGCCATGACAAAAAC	Forward primer to amplify the 5'UTR library for sequencing

1162	CACGGAACAGGTAGTTTTCC	Reverse primer to amplify the 5'UTR library for sequencing
1163	TTAAAAGCATTCTGTAACAAAGC	Forward primer to amplify the 5'UTR library for sequencing
1164	GGTCAGAGTAGTGACAAGTG	Reverse primer to amplify the 5'UTR library for sequencing
1165	CTAACCAAACCGGTAACCCC	Forward primer to amplify the 5'UTR library for sequencing
1166	AACGGGAAAAGCATTGAACAC	Reverse primer to amplify the 5'UTR library for sequencing
1167	CTGCGTCTTTTACTGGCTCT	Forward primer to amplify the 5'UTR library for sequencing
1168	TCATGCCGTTTCATGTGATCC	Reverse primer to amplify the 5'UTR library for sequencing

CHAPTER 3

Contribution to the manuscript entitled "mRNA is destabilized throughout the molecule when translation is altered while its concentration is locally affected" (Annex 1)

Introduction

In the manuscript "*mRNA is destabilized throughout the molecule when translation is altered while its concentration is locally affected*" (Annex 1), we show that modification of translation initiation and elongation destabilizes the entire mRNA molecule and decreases (locally) its concentration. These experiments demonstrate the protective role of ribosomes on mRNAs. Furthermore, subsequent experiments highlight how the three processes of translation, mRNA degradation and transcription are deeply interconnected for quality control implemented by cells to avoid unproductive gene expression.

The following sections describe my contribution to this work regarding the modification of the efficiency of translation initiation and the consequences on mRNA concentration and stability. More particularly, I describe the selection of the SD sequences used in this work to modulate translation initiation of the *lacZ* reporter mRNA and preliminary measurements of *lacZ* mRNA concentration and stability.

1. Selection of SD sequences with predicted low translation initiation efficiency

Our goal was to provide a model mRNA with a reduced efficiency of translation initiation for subsequent exploration of ribosome protection of mRNAs. We chose to modify the SD (Shine-Dalgarno) sequences on a plasmid carrying the *lacZ* gene coding for the β-galactosidase. The SD sequence is located in the 5'UTR of an mRNA and complementary base-pairs with a region at the 3'end of 16S rRNA. The SD sequence facilitates ribosome binding to mRNA to initiate translation. Weak binding affinity between SD and 16S rRNA leads to low translation initiation efficiency. Therefore, we designed four SD variants to compare with a reference wild-type version SD-wt. SD-wt is the near-optimal SD sequence AGGAGG that sustains efficient translation initiation. The four SD variants (also 6 ntd in length) were designed to have a low complementary with the 16S rRNA sequence and displayed the following sequences CCCCGG, GGAGGT, TTATAA, and GCTCCA. All SD sequences were located in place of the AGGAGG wt SD sequence within a 33 ntd 5'UTR inserted upstream of the *lacZ* gene and under the control of the pBAD inducible promoter in the pBAD-*lacZ* plasmid originated from pBAD-his/myc (Invitrogen). Detailed construction methods can be found in the manuscript (Annexe 1). The RBS indexes corresponding to the theoretical translation initiation efficiency were predicted for the wild type and four SD variants using UTR Designer. For a detailed description of the software, see Chapter 1 Material and Methods. Prediction of theoretical translation initiation rates for the five SD sequences are shown in the table below (***Table 1***).

Table *1*. Prediction of theoretical translation initiation rates for the five SD sequences by UTR Designer.

SD	5'UTR sequence[1]	RBS index[2]
SD-wt	CCCGTTTTTTGGGCTAACAGGAGGAATTAACCATG	719 882
SD-1	CCCGTTTTTTGGGCTAACCCCCGGAATTAACCATG	11 584
SD-2	CCCGTTTTTTGGGCTAACGGAGGTAATTAACCATG	707 904
SD-3	CCCGTTTTTTGGGCTAACTTATAAAATTAACCATG	513
SD-4	CCCGTTTTTTGGGCTAACGCTCCAAATTAACCATG	213

[1] Sequences of the 33 ntd 5'UTRs. SD sequences (6 ntd) are underlined; UTR Designer predicted region that potentially binds to the 16s rRNA are marked in red; start codon is marked in blue.

[2] Predicted theoretical translation initiation rates given by the UTR Designer software.

The results show that the theoretical translation initiation efficiency of SD-2 (RBS index= 719 882) and SD-wt (RBS index= 707 904) were similar and high, SD-1 was lower than that of SD-wt (more than 60-fold), SD-3 and SD-4 were extremely lower than that of SD-wt (more than 1000-fold). The next step was to experimentally confirm that SD variants predicted to result in low translation initiation efficiency do indeed affect gene expression.

2. Effect of SD on β-galactosidase protein level and *lacZ* mRNA concentration

We first determined the production of β-galactosidase by measuring its specific activity with the SD-wt and the four SD variants (***Figure 1A***). The β-galactosidase assay is described in the Materials and Methods of the manuscript (Annex 1). The β-galactosidase activity with SD-2 was reduced to half of SD-wt and those with SD-1 and SD-3 were extremely lower than with SD-wt (more than 20-fold). With the SD-4, the β-galactosidase did not seem to be produced or at a level under the detection limit.

We also determined the effect of the four SD variants on the fold change of *lacZ* mRNA concentration compared to the reference SD-wt (***Figure 1B).*** Quantification of *lacZ* mRNA concentration by qRT-PCR is described in the Materials and Methods of the manuscript (Annex 1). The results showed a large variation in mRNA concentrations for all four SD variants compared to SD-wt. It was not expected to

observe such large variations in *lacZ* mRNA levels because the level of transcription was theoretically the same for all constructs that used the same promoter induced with the same concentration of arabinose. Since mRNA concentration is a balance between transcription and mRNA degradation, the other explanation is a modification in *lacZ* mRNA stability with the four SD variants.

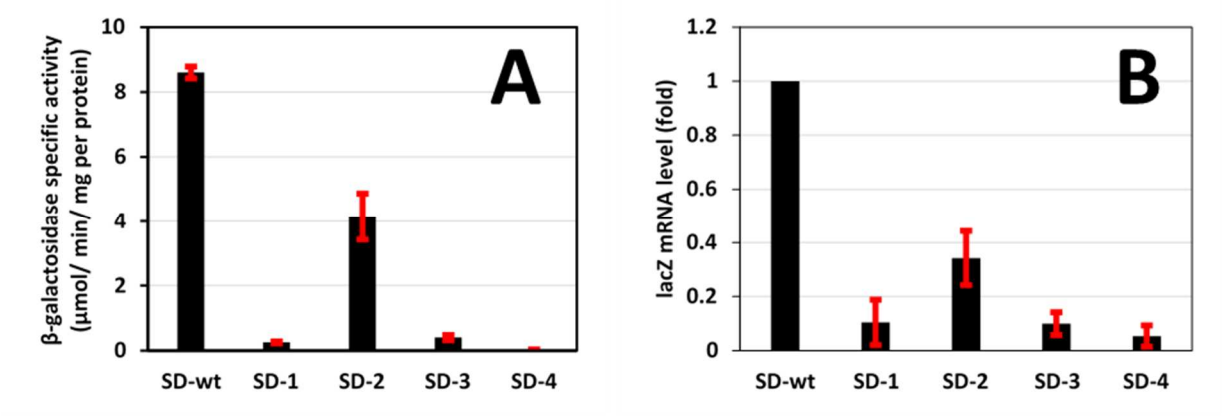


Figure *1*. Specific β-galactosidase activity and lacZ mRNA concentration with the five SD sequences. A) Specific β-galactosidase activity with the five SDs (n=5). B) Concentration in lacZ mRNA with the four SD variants as a fold change relative to the reference to SD-wt. Error bars represent standard deviation of replicates (n=24).

Altogether, these experimental results showed that the four SD variants with predicted low translation efficiency indeed led to a lower β-galactosidase expression but also to a lower *lacZ* mRNA concentration. To further investigate the effect of SD variants on the variation in *lacZ* mRNA concentration we wanted to measure *lacZ* mRNA stability. Since these experiments are time and cost consuming, we decided to select only one SD variant. Based on the prediction of theoretical translation initiation rates, SD-3 and SD-4 would be the most suitable sequences for reducing translation initiation efficiency compared to SD-wt. However, the effect of SD-4 seemed too drastic with no detectable β-galactosidase activity, so we selected SD-3. In the manuscript (Annex 1) SD-wt and SD-3 are renamed *wt-lacZ* and *SD-lacZ,* respectively.

3. Effect of SD on *lacZ* mRNA stability

To test whether the decrease in *lacZ* mRNA concentration with SD-3 was related to a variation in transcript stability, I performed preliminary measurements of the *lacZ* mRNA half-lives with SD-3 and SD-wt (***Figure 2***). Quantification of *lacZ* mRNA half-life by qRT-PCR is described in the Materials and Methods of the manuscript (Annex 1). I used three primer pairs which were designed and distributed equally along the CDSs (beginning, middle, and end of the sequence) to estimate the *lacZ* mRNA half-lives. We found that *lacZ* mRNA stability was more than 3-fold reduced with SD-3 than with SD-wt. This means that reducing translation initiation efficiency with SD-3 led to a decrease in *lacZ* mRNA stability, which in turn was responsible for the lower *lacZ* mRNA concentration. The reduction in the efficiency of translation initiation likely decreased the number of ribosomes bound on the mRNA and, therefore, the protection of mRNAs by ribosomes from RNase attack.

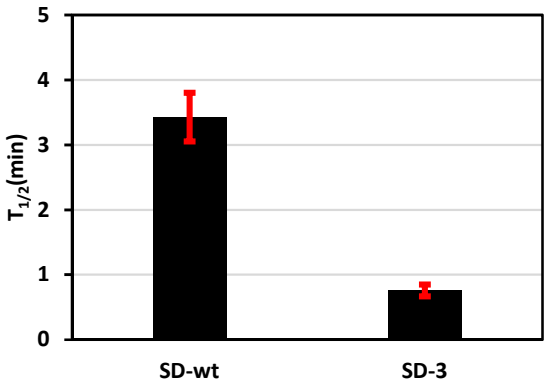


Figure 2. Half-life of lacZ mRNA with SD-wt and SD-3. Error bars represent standard deviation of replicates (n=6).

Conclusion

In order to study the coupling of mRNA degradation and translation, we modulated the efficiency of translation initiation by modifying the SD sequences of a *lacZ* reporter mRNA and measured the consequences on protein concentration and mRNA concentration and stability. SD variants with lower complementarity to the 3'end of the 16S RNA were designed. These SD variants were predicted to result in low translation initiation rates, and they indeed decreased β-galactosidase protein level. However, we also found that they decreased *lacZ* mRNA concentration. The lower *lacZ* mRNA concentration was related to a lower transcript stability. These results support the coupling of translation and mRNA degradation, in particular through the protective role of ribosomes.

Following this work, additional experiments were performed to confirm the decreases in *lacZ* mRNA concentration and stability and to provide local variation in concentration and stability all along the

lacZ mRNA molecule (see manuscript in Annex 1). Regarding ribosome protection, additional experiments were also carried out to measure the number of bound ribosomes on *lacZ* mRNAs with SD-wt and SD-3 (see manuscript in Annex 1). These experiments confirmed that the *lacZ* mRNA with SD-3 had a lower number of bound ribosomes than SD-wt in agreement with the predicted lower rate of translation initiation and the lower β-galactosidase protein level.

CONCLUSION AND PERSPECTIVES

Conclusions

The regulation of gene expression is fundamental for the physiological activity of all living organisms, ensuring that they can express their genes in correlation with environmental resources and metabolic requirements. Gene expression regulation is vital for reshaping the transcriptome to better adapt to the new living conditions when the environment is modified. In-depth exploration of gene expression regulation is important to understand this central cellular process but also to be able to develop strategies or molecular tools to better manipulate the physiology and metabolism of cells, or to control the expression of proteins of interest.

In my PhD, I focused my research on mRNAs and the regulation of their expression and stability. The mRNAs are central molecules in the transfer of genetic information from DNA to proteins and are subject, or participate, to different levels of regulation such as transcription, stability and translation. I more particularly focused my work on the role of 5'UTR sequences (untranslated transcribed region) in gene expression control. This study explored the effect of the 5'UTR in regulating gene expression and the contribution of the 5'UTR to bacterial cell adaptation. Two different approaches were developed. In the first part, I developed synthetic 5'UTRs to measure their contribution to regulating mRNA concentration through control of transcription and stability and in regulation of translation. In the second part of my work, I constructed an exhaustive library of native *E. coli* 5'UTR sequences to analyze their role in gene expression regulation when cells must adapt to changing environmental conditions.

In the first part of my work, we developed an approach using synthetic 5'UTR sequences in fusion with a reporter gene (*lacZ*) to analyze their role in regulating gene expression in *E. coli* at the translation, transcription, and mRNA degradation levels. Using the RBS Calculator, a series of 41 synthetic 5'UTRs with different theoretical translation initiation rates (RBS index) were designed and screened. We covered the whole range of RBS index calculated by the software (from 0.1 to 100 000 AI) with several different 5'UTR sequences designed for each RBS index. These synthetic 5'UTRs were fused to the *lacZ* coding sequence under the transcriptional control of the P_{BAD} inducible promoter in a plasmid system. We measured the specific β-galactosidase (*lacZ* encoded protein) activity of the 41 strains developed in this work. The results showed that the level of protein expression was globally correlated with the theoretical translation initiation rate of our synthetic 5'UTRs. However, different levels of protein expression were observed within the group of a similar RBS index, which suggests that the theoretical translation initiation rate cannot be used alone to accurately predict the final protein level. This was expected as the efficiency of translation initiation is not the only parameter controlling

protein expression and the RBS Calculator does not consider the mRNA concentration as a calculation parameter. Indeed, 5'UTRs are known to influence mRNA stability, which in turn can have an impact on mRNA concentration in the cell (Cetnar & Salis, 2021; Hui et al., 2014; Julius & Yuzenkova, 2019; Tuller & Zur, 2015). We measured the *lacZ* mRNA concentration of the 41 strains and found different mRNA concentrations between strains. Altogether, these results show that 5'UTRs impact gene expression not only at the translational level but also at the level of mRNA concentration. The mRNA concentration in the cells results from an equilibrium between transcription and mRNA degradation. Since all constructs were under the control of the same p_{BAD} promoter induced with the same level of inducer (0.001% arabinose), all constructs were expected to be similarly transcribed. Therefore, changes in mRNA concentration could be related to 5'UTR effects on *lacZ* mRNA stability.

To analyze whether differences in *lacZ* mRNA concentration as a function of the 5'UTR were due to differential mRNA stability, we measured the *lacZ* mRNA stability with eight representative 5'UTRs and confirmed that the 5'UTR influences mRNA stability. However, we found that changes in *lacZ* mRNA stability alone could not fully explain the observed changes in mRNA concentration. This result means that some 5'UTRs have a direct effect on transcription. To quantify more precisely how variations in mRNA concentration depend on variation in transcription and/or stability, we used the concept of the degradational regulation coefficient (*ρD*). We demonstrated that 5'UTR-mediated gene expression regulation could be related to changes in transcription, stability, or to a shared control, depending on the 5'UTR. The same results were found when the 5'UTRs were fused to two other reporter genes (*txAbF*, encoding an α-L-arabinofuranosidase and *msfGFP*, encoding a fluorescent protein). Similarly to what was observed with *lacZ*, variations in *txAbF* and *msfGFP* mRNA concentrations can be controlled by changes in transcription, stability, or both, depending on the 5'UTR. Interestingly, the control of a given 5'UTR can be the same for the three reporter genes, or specific to one reporter gene.

In addition, I also showed that modifying the SD sequences of the 5'UTR to modulate translation initiation efficiency of a *lacZ* reporter mRNA had strong consequences on protein concentration and mRNA concentration and stability.

Altogether our results clearly show that the role of 5'UTRs in gene expression control is more complicated than only the regulation of translation initiation. 5'UTRs seem to be at the crossroads of transcription, stability and translation regulation, interacting with all three processes to control gene expression.

The second part of my work focused on the role of native 5'UTRs is gene expression regulation. For this goal, an exhaustive library of 2547 native 5'UTR sequences identified in *E. coli* was designed and constructed. The 5'UTRs were cloned upstream of a fluorescent reporter gene (encoding msfGFP) and under the control of an inducible promoter (P_{BAD}). To explore how these 5'UTRs regulate gene

expression in response to environmental changes, the library was grown in different environmental conditions (temperature and medium composition). After transcription induction of the *msfGFP* gene, the fluorescence was determined by flow cytometry to reflect the level of the reporter protein. A large range (3-log scale) of fluorescence was covered by our library, reflecting the role of the 5'UTRs in modulating gene expression. For each environmental condition, cells were sorted into six windows according to their fluorescence level. Subsequently, the identities of the 5'UTRs present in each window were determined by next generation sequencing. We developed a procedure to identify 5'UTRs that do or do not contribute to msfGFP expression regulation when challenged in different environments. We identified which 5'UTRs can either up-regulate or down-regulate msfGFP expression levels in specific growth conditions. For example, under the effect of temperature up-shift (M9_42°C versus M9_37°C), 19% of the 5'UTRs identified in this condition (248/1311) up-regulated the msfGFP expression and 23% of the 5'UTRs (308/1311) down-regulated the msfGFP expression. This analysis, performed for each condition, led to the determination of lists of 5'UTRs associated with changes in protein expression in response to environmental conditions. Using the Gene Ontology (GO) term enrichment, we linked 5'UTR-mediated gene expression regulation under different growth conditions with the "biological process" category of the 5'UTR-associated genes. The results showed that some genes have part of their expression regulation carried by their 5'UTR sequence, independently of the downstream coding sequence and the upstream promoter. For example, we identified that the 5'UTR of *moaA* is directly involved in upregulated msfGFP expression in response to heat stress. The upregulated expression of *moaA* in response to heat shock can be explained by the presence of a riboswitch in its 5'UTR (Krisko et al., 2014; Regulski et al., 2008). This library provides a new reservoir of constructs to study the regulation of specific genes, considering only the role of 5'UTR in their regulation.

Moreover, by comparing responses obtained in the four different environmental conditions, we identified 85 5'UTRs that lead to high gene expression regardless of the growth temperature and medium composition. These 5'UTRs could be interesting tools to use for robust and environment-insensitive expression of proteins of interest.

Perspectives

Our results confirm that the 5'UTRs are important sequences of the mRNA molecules involved in their lifestyle and can be qualified as "regulatory-hubs" of gene expression in *E. coli*. This work opens questions and perspectives that could be addressed in the future to increase the knowledge on gene expression control and to develop molecular tools to precisely manipulate gene expression in *E. coli*.

We have demonstrated that synthetic 5'UTRs play a key role in *E. coli* in regulating gene expression at the levels of transcription, mRNA degradation and translation. Since all these processes are interconnected, it remains difficult to identify which one the 5'UTR is really acting on, this is particularly true between transcription and stability. We have experimentally determined the mRNA stability and concentration. A limitation of this study is that transcription was estimated indirectly (by measuring changes in mRNA concentration and half-life), but not quantified directly experimentally. Short-term perspective will be to experimentally measure the transcription rate of constructs used in this work. To do this, we will consider two approaches. The first one is to measure the transcription rate *in vitro* for combinations of 5'UTR-reporter. This will give us information on the effect of each 5'UTR on transcription only, without the effect of degradation present *in vivo* on mRNA concentration. A second possibility will be to use a tRNA in transcriptional fusion at the 3'end of the mRNA molecule. As tRNAs are more stable than mRNAs, their concentrations reflect more the transcription level than that of the mRNA part (Iost & Dreyfus, 1995). By comparing the concentration of tRNA between strains, we will be able to more accurately quantify the transcription levels and more precisely determine a potential impact of 5'UTR in transcription itself.

In addition, a more long-term perspective will be to edit precise and general rules of mRNA concentration and protein synthesis regulation by 5'UTRs. This will require to increase the number of synthetic 5'UTRs used in the study. This larger set of 5'UTRs could help to study the two types of regulations identified in our study at the level of translation or mRNA concentration. By selecting constructs with similar mRNA concentrations but discrepancy in protein expressions, we will be able to study 5'UTRs that mediate translational regulation. Likewise, we could select constructs with 5'UTR-mediated regulation of mRNA concentration controlled by the level of its stability or its transcription. In-depth investigation of the sequence characteristics (e.g. folding, length, presence of sequence motifs for binding of regulatory proteins and sRNAs) of these selected 5'UTRs will help us understand in more detail the underlying molecular mechanisms.

The development of the library composed of the 2547 native 5'UTRs led to the identification of 5'UTRs that modify gene expression in response to environmental changes. From this work, we have generated lists of 5'UTRs which up-regulate, down-regulate or do not regulate reporter gene expression. The first short-term perspective will be to characterize in depth the 5'UTRs identified in each expression windows. We will first analyze using bioinformatics, relationships between expression and sequence features of the 5'UTRs. To do this, we will calculate the RBS index of each member of the library and analyze if correlations exist with the expression level. We can expect to find for at least some 5'UTRs a relationship between the expression level and the sequences of the SD sequences suggesting that for these 5'UTRs, the expression is mainly governed by translation initiation. We will analyze potential secondary structures that could modify protein expression. We will examine if these secondary structures could be involved in the regulation observed between the tested environmental conditions (more particularly in responses to temperature shift).

In this work, we only tracked the protein expression level. Another short-term perspective is to determine mRNA concentration and stability for each construct of the library. This will be performed by RNA-seq focused on the 5'UTR sequences. With these data, we will be able for each construct to i) correlate mRNA concentration and protein expression level, ii) measure the role of the degradation in mRNA concentration regulation by determining at the large scale degradational regulation coefficients (ρD) and iii) identify if changes in gene expression in response to environment are related to variations in transcription and/or degradation.

The functional analysis led to the identification of some 5'UTRs responding to environment. Some of them can be interpreted using previously known mechanisms, while others cannot. More detailed conclusions and new discoveries are expected to be drawn by analyzing the sequence characteristics and functional correlation of 5'UTR. We can now select some representative responding 5'UTRs and characterize them further. For example, 5'UTR-msfGFP constructs responding to temperature could be compared to their native counterpart (5'UTR-native gene) to isolate regulation specifically associated to the 5'UTR part of the mRNA. We also identified in this work 85 5'UTRs which never respond to environment. They could be of interest to develop molecular tools for biotechnological applications when constant and environment-independent gene expression is required.

Our library is an easy-to-use tool to study gene expression modifications in response to environment. It would be very interesting to challenge this native 5'UTR library in other growth stress conditions (e.g. oxygen concentration, pH, osmolarity, etc), to provide both 5'UTRs that respond to the imposed stress and 5'UTRs that always lead to high gene expression regardless of the growth conditions. Alternatively, we can also select 5'UTRs that are only sensitive to a specific stimulus. Finally, under a given growth condition, we ranked the native 5'UTRs according to the expression level of the

gene they control. This provides a large repertoire of regulatory sequences that can be used under this given growth condition for custom modulation at the post-transcriptional level of the expression of any gene, or pathway, of interest.

REFERENCES

Cetnar, D. P., & Salis, H. M. (2021). Systematic Quantification of Sequence and Structural Determinants Controlling mRNA stability in Bacterial Operons. *ACS Synthetic Biology*, *10*(2), 318–332. https://doi.org/10.1021/acssynbio.0c00471

Hui, M. P., Foley, P. L., & Belasco, J. G. (2014). Messenger RNA Degradation in Bacterial Cells. *Annual Review of Genetics*, *48*(1), 537–559. https://doi.org/10.1146/annurev-genet-120213-092340

Iost, I., & Dreyfus, M. (1995). The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO Journal*, *14*(13), 3252–3261. https://doi.org/10.1002/j.1460-2075.1995.tb07328.x

Julius, C., & Yuzenkova, Y. (2019). Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate. *WIREs RNA*, *10*(2), e1512. https://doi.org/10.1002/wrna.1512

Krisko, A., Copic, T., Gabaldón, T., Lehner, B., & Supek, F. (2014). Inferring gene function from evolutionary change in signatures of translation efficiency. *Genome Biology*, *15*(3), R44. https://doi.org/10.1186/gb-2014-15-3-r44

Regulski, E. E., Moy, R. H., Weinberg, Z., Barrick, J. E., Yao, Z., Ruzzo, W. L., & Breaker, R. R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Molecular Microbiology*, *68*(4), 918–932. https://doi.org/10.1111/j.1365-2958.2008.06208.x

Tuller, T., & Zur, H. (2015). Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Research*, *43*(1), 13–28. https://doi.org/10.1093/nar/gku1313

ANNEXES

mRNA is destabilized throughout the molecule when translation is altered while its concentration is locally affected.

Marie-Pierre Duviau[1,#], Fan Chen[1,#], Muriel Cocaign-Bousquet[1,§], Laurence Girbal[1,§] and Sébastien Nouaille[1,§,].

1 :TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

#, § : equally contributors. Corresponding author : sebastien.nouaille@insa-toulouse.fr

Abstract.

The mRNA molecule is located at the crossroads of transcription, translation and mRNA degradation. Many questions currently persist about the coupling of these three processes in *E. coli* and, in particular, how mRNA degradation coordinates with translation and transcription. To characterize the coupling between mRNA degradation and translation while accounting for transcription, we altered translation initiation or elongation of the *lacZ* mRNA and measured the effects on its stability and concentration. A mapping method has been developed to analyze mRNA concentration and stability at the local scale all along the transcript. We demonstrated that a decrease in efficiency of translation initiation leads to a destabilization of the mRNA and a decrease in its concentration, both homogeneously throughout the molecule. Premature termination of translation elongation by insertion of a stop codon provokes the uniform destabilization of the transcript triggered by the ribosome-free portion. This was associated with a drop in local mRNA concentrations downstream of the stop codon, due to the uncoupling of transcription and translation. Altogether, these results demonstrates the protective role of ribosomes on mRNAs and highlight how the three processes of translation, mRNA degradation and transcription are deeply interconnected for quality control implemented by cells to avoid unproductive gene expression.

INTRODUCTION.

The regulation of gene expression is vital for any cell to adjust as much as possible its metabolic capacities to its growing environment. At the confluence of transcription, translation and degradation, mRNA can be considered as a hub of gene expression regulation. Many questions still remain about the coupling of the three processes of transcription, translation and mRNA degradation in *E. coli* and, in particular, questions about how mRNA degradation coordinates with the other processes.

The mRNA degradation is performed in *E. coli* by a large panel of endo- and exo-ribonuleases whose activity and mode of action, at least for the main are rather well characterized from a mechanistic point of view (1). mRNA degradation is generally initiated by an endoribonucleolytic cleavage performed mainly by the essential endoribonuclease RNase E, and more occasionally RNase III. RNase E can form a supramolecular structure called the degradosome by association with the polynucleotide phosphorylase (PNPase), the RNA helicase RhlB and the glycolytic enzyme enolase (Eno) (2,3). The initial cleavage products are then degraded by multiple cycles of attack by RNase E combined with the action of 3'-exoribonucleases and oligoribonuclease to the nucleotide level (4).

The link between mRNA degradation and translation is not so simple as ribosomes act as both positive and negative effectors of transcript stability. At the level of translation initiation, ribosomes can protect transcripts from degradation. Less efficient translation initiation leads to mRNA destabilization by limiting initiation at mutated RBS or decoupling transcription and translation (5,6). These results first demonstrated on a few transcripts were later on confirmed on a large scale using synthetic libraries of RBS and translation initiation regions (7,8). Furthermore, fast ribosome trafficking through the translation initiation region, which reduces ribosome coverage in this region, destabilizes mRNAs (9). The mechanism of transcript protection by ribosomes could be related to ribosomes blocking of the 5' end access to RNase E. In contrast, at the translation elongation level, ribosomes can act as negative effectors of mRNA stability. Slowly elongating or stalled ribosomes are associated with transcript destabilization (5,6). Large-scale studies of correlation between the codon bias and mRNA stability also showed that less optimized codons, leading to slow elongating ribosomes, were associated with less stable RNAs (10,11). A mechanistic clue of a ribosome-mediated degradation came from a reported physical interaction between the degradosome and the 70S ribosomes and polysomes indicating that ribosomes might act as antennae to capture of the RNA degradosome (12). In addition, a recent study shows that the membrane-dependent localization of degradosomes as punta depends on the presence of translationally active polysomes. These sites

were proposed to be involved in the initial step of degradation of actively translated mRNAs (13). When RNase E is delocalized from the membrane, this leads to an increase in ribosome-free mRNA turnover (14). These results open up questions on the degradation and quality control of mRNAs based on their translational state.

Moreover, it is very difficult to understand how the coupling of mRNA degradation and translation relates to transcription. On the one hand, it is known that translation is coupled to transcription at the level of the pioneer ribosome, which is involved in direct and indirect physical associations with RNA polymerase (15-17). Acceleration or deceleration of the pioneer ribosome changes the speed of RNA polymerase (18-20). On the other hand, the coupling of transcription and translation can be limited by the different subcellular locations of transcription and translation: RNA polymerase synthesizes mRNA molecules in the nucleoid while most mRNA molecules in translation are located outside the nucleoid (21). Furthermore, transcription was recently shown to impact the regulation of mRNA degradation and translation. Increasing mRNA concentration by increasing transcription increases translation efficiency and decreases mRNA stability (Nguyen 2022; Nouaille 2017). The relationship between mRNA concentration and translation efficiency or mRNA stability was also observed on a genome-wide scale by correlation analyses (11,22-25). A physical mechanism based on the competition among all mRNAs present in the cell to bind to a finite pool of RNase E (26) or ribosomes was postulated. The interplay between the coupling of translation and transcription with mRNA degradation is not yet fully understood.

To further characterize the coupling between mRNA degradation and translation, which appears to be more complicated than a simple steric barrier of ribosomes against RNase attack, while accounting for transcription, we modulated translation initiation or elongation and measured the effect on mRNA stability and concentration. A novel approach, based on qRT-PCR measurements all along the mRNA molecule, was developed to provide information on local differences in stability and concentration. The decrease in the efficiency of translation initiation destabilized the mRNA homogeneously throughout the molecule, and this global destabilization of the transcript could explain its lower overall concentration. In the case of premature termination of translational elongation, the mRNA was also uniformly destabilized throughout the molecule and this was triggered by the presence of the ribosome-free portion on the mRNA. The decrease in local mRNA concentrations downstream the inserted stop codon was explained by a premature termination of transcription following uncoupling of translation and transcription. These results shed new light on the protective role of ribosomes on mRNAs and highlight how the three processes of translation,

mRNA degradation and transcription are deeply connected in an early quality control developed by the cell to avoid unproductive gene expression.

Materials and methods.

Bacterial growth and induction conditions.

All plasmidic constructions were established in *E. coli* DH5α (New England Biolabs) and transferred into the MG1655 derivative (MET345) in which the *lacZ* chromosomal copy was deleted (25). *E. coli* strains were grown in Luria-Bertani (LB) broth for cloning steps, at 37°C under shaking. For RNA, polysome and protein analyses, cultures were performed in M9 minimal medium supplemented with 3 g/L glucose at 37°C under shaking (27). Ampicillin was used at 100 $\mu g.ml^{-1}$ and chloramphenicol at 20 $\mu g.ml^{-1}$.

Transcription of *lacZ* mRNA under the P_{BAD} promoter was induced as follows: an overnight culture was used to inoculate fresh M9 medium at OD_{600}= 0.1. At OD_{600}= 0.6, arabinose (Sigma) was added at the final concentration of 0.0001% and the culture induced for 30 min.
For determination of the mRNA decay, transcription was arrested by addition of rifampicin at a final concentration of 500 $\mu g.ml^{-1}$. A volume corresponding to 6 mg of cells (dry weight) of culture was collected over time and flash frozen in liquid nitrogen. Six different time points (0, 0.3, 1, 2, 4 and 7 min), including the reference sample (before the addition of rifampicin) were used to analyze mRNA degradation kinetics. For polysome profiling experiments, transcription was induced by adding arabinose (0.0001% [w/v]) at an OD_{600} of 1 for 30 min and then translation was arrested by adding 0.1 mg/mL chloramphenicol.

For incorporation of non-canonical amino-acid, AzF (4-Azido-L-phenylalanine - Iris Biotech GmbH, Germany) was dissolved in 1N NaOH and added to the culture medium at the final concentration of 2 mM. Initial pH of the culture medium was corrected by addition of $1/10^{th}$ volume of HCl at 37%. AzF was added at the beginning of the culture then growth, induction and sampling were treated as described above.

Construction* of *vectors

General DNA manipulation procedures were performed as previously described (28), plasmid DNA was isolated using the QIAprep Miniprep kit (Qiagen). All plasmids containing *lacZ* used in this study were derivatives of the pBAD-lacZ-cmyc-his plasmid (Life Science). pPCR was performed using *Phusion* polymerase (New England Biolabs). Constructs were made by PCR assembly for replacement of SD sequence and deletion of the untranslated regions of the *lacZ* CDS. Amplicons were purified on

gel (Qiagen Gel purification kit), 5'ends were phosphorylated with 15 U of T4 polynucleotide kinase 30 min at 37°C (New England Biolabs) and self-ligated with 30 U of T4 DNA ligase (New England Biolabs) at 16°C overnight. For integration of the stop codon at different positions of the *lacZ* CDS, the vector was amplified with a pair of complementary primers containing the TAG codon as mismatch. After purification, *E. coli* was transformed with the amplicon and clones were selected by the blue/white screening on plates containing X-Gal (Sigma). Deletion of the orthogonal tRNA contained in the pEVOL plasmid was made by amplification of the full vector except the tRNA portion, followed by purification, phosphorylation and self-ligation as described above. The orthogonal tRNA ($tRNA_o$) was integrated in transcriptional fusion with *lacZ* by PCR. Two primers were used, with one part hybridizing to the integration site and one part containing half of the $tRNA_o$. They were used to amplify the whole plasmid and the product was self-ligated. Based on this plasmid containing *lacZ* fused to $tRNA_o$, the version with the non-functional tRNAo was generated by PCR amplification of the full vector without the 3 last bases (CCA) at the end of the tRNAo to generate the *lacZ*-tRNA* construct. All constructs were verified by sequencing (Eurofins). All primers and plasmids used in this study are listed in Table S1.

Polysome profiling.

Polysome profiling experiments were performed as previously described (29). Briefly, cells were harvested after translation arrest, washed twice, and resuspended in lysis buffer (20 mM Tris HCl pH 8, 140 mM KCl, 40 mM MgCl2, 0.5 mM DTT, 100 µg/ml chloramphenicol, 1 mg/ml heparin, 20 mM EGTA, 1% Triton X-100). After mechanical cell disruption with glass beads, mRNA-ribosome complexes were size-separated on a sucrose gradient (10 to 50% (w/v) in polysome gradient buffer (same composition as lysis buffer except for heparin at a final concentration of 0.5 mg/ml and without Triton X-100)) into 24 subfractions. The levels of 16S and 23S rRNAs in each subfraction were calculated using the Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and used to pool the subfractions into seven fractions labelled A to G. Fraction A consisted of free mRNA molecules not undergoing translation, while fractions B to G consisted of mRNA copies in translation bound to increasing numbers of ribosomes (1 bound ribosome in fraction B to around 11 bound ribosomes in fraction G) (29). Protein denaturation and nucleic acid precipitation were performed in the pooled subfractions as previously described (24).

RNA extraction, quality control and cDNA synthesis.

Total RNA was extracted with RNeasy mini extraction kit (Qiagen) according to provider's recommendations using Qiacube (Qiagen). Before cell lysis, each sample was centrifuged for 10 min, at 8,000 rpm, at 4°C, resuspended in 500 µL RLT buffer and transferred into a tube containing 0.1 g of glass beads. Cells were disrupted at 4°C by 3 cycle of 30 sec with a FastPrep-24 instrument (MP Biomedicals). After centrifugation for 10 min, at 13,000 rpm, at 4°C, 350 µL of supernatant was used for purification with Qiacube robot (Quagen). Total RNA was eluted in 50 µL of elution buffer. Additional DNase treatment was applied to remove any residual genomic DNA contamination. Total RNA (50 µg) was treated with 2 U of Turbo-DNase (Ambion) for 15 min at room temperature followed by a RNA clean-up protocol using RNeasy Mini Kit (Qiagen). The absence of significant genomic DNA contamination was checked by qPCR. RNAs were quantified using ND-1000 UV-visible light spectrophotometer (NanoDrop Technologies) and their integrity certified with Bioanalyzer 2100 with the RNA 6000 Nano LabChip kit (Agilent). RNAs were stored at -80°C until use.

Synthesis of cDNA was performed on 5 µg of total RNA using SuperScript II reverse transcriptase (Life Technologies) as previously described ((Redon 2002)). The cDNAs was 1:10 serially diluted to identify the dilution that would lead to a cycle threshold (Ct) between 15 and 25 for representative primer pairs targeting *lacZ* cDNA. The appropriate dilution was selected to be quantified with LightCycler 480 II (Roche).

β-galactosidase assay.

3 mg of cells (dry weight) were collected, harvested, washed twice with ice-cold 0.2% KCl, resuspended in 1ml of lysis buffer (15 mM Tris tris 400 mM/ tricarballylate 1, 4.5% Glycerol, 0.9 mM $MgCl_2$, 0.2 mM DTT; pH = 7.2), transferred in screw capped tubes containing 0.1 g glass beads and disrupted with FastPrep-24 instrument (MP Biomedicals) (6 cycles, 6.5 $m.s^{-1}$, 30s with 1 min on ice in between). Supernatant containing soluble proteins was used for immediate quantification. All measurements were carried out on 3 biological with 6 technical replicates for each.

The total protein content of cell extracts was determined by the Bradford method with bovine serum albumin as the protein standard. β-Galactosidase activity was determined by the colorimetric method using O-nitrophenyl-β-D-galactopyranoside (Sigma) as substrate. The activity was determined using the slope of ONP appearance (SpectraMax plus, Molecular Device, 30°C, 420 nm) and expressed as a specific activity ($mmol.min^{-1}.g^{-1}$) using the total protein concentration of the sample. For each quantification, at least 2 independent sample extractions were performed and 6 technical replicates at different dilutions.

Primer design for qRT-PCR and validation

Primers for qPCR were designed using Vector NTI advance v11 (Life Technologies) with a melting temperature of 59-61°C, a length of 20 to 22 bp and 50% to 67% GC content. The reaction efficiency of each pair of primers was tested as a single amplicon on serial dilutions of pBAD-*lacZ-myc*-his as a matrix. The efficiency of validated primer pairs was tested on cDNA serial dilutions and focused around 100%. To quantify *lacZ*, a set of primer pairs was designed (TableS1). In the figures on this study, the amplicons are indicated by a letter. The positions (in nucleotides) of the different amplicons, relative to the *lacZ* +1 transcriptional start site (TSS) are the following : a(113-227), b(701-831), c(861-986), d(1219-1338), e(1516-1617), f(1666-1685), g(2286-2415), h(2523-2615), i(2735-2837), j(3223-3337) and k(3223-3423). As a reminder, the coordinates of the different parts are as follows : 5'UTR (1-33), *lacZ* CDS (34-3269) and $tRNA_o$ (3317-3423) (in nucleotides, referred to the +1TSS). The housekeeping gene *ihfB* (integration host factor β-subunit, (30)) and *bla* (ampicillin resistance gene carried by the plasmid) were used as internal normalization controls.

High throughput real-time quantitative PCR

Low throughput qRT-PCR was performed with LightCycler 480 II (Roche) on 96 well plates (Biorad) with SyberGreen MasterMix (Biorad). High throughput qRT-PCR was carried out using the 96.96 dynamic array™ IFCs and the BioMark™ HD System (Fluidigm Corporation, CA, USA) following the manufacturer's protocol (31) at the Gentiane Plateform (Clermont Ferrand, France). Briefly, the steps were as follows. Fourteen pre-amplification cycles were performed with a pooled primer mixture (0.2 µM). The pre-amplified samples were treated with 8 U of exonuclease I (New England BioLabs), diluted 1:5 with Tris-EDTA buffer and added to a "Sample Mix" consisting of TaqMan® Gene expression Master Mix (Applied Biosystems), DNA Binding Dye Sample Loading Reagent (Fluidigm), EvaGreen® dye (Biotium) plus Tris-EDTA buffer, as recommended. In parallel, each primer pair (20 µM) was added to a "Primer Mix" composed of Assay Loading Reagent (Fluidigm) plus Tris-EDTA buffer, as recommended. An IFC controller was used to prime the fluidics array, then 5 µL of each sample and primer mix were loaded in the appropriate inlets. The loaded chip was transferred to the BioMark™ HD System and qPCR was performed using the following temperature program: 2 min at 50°C, 30 min at 70°C and 10 min at 25°C; followed by a hot start 2 min at 50°C, 2 min; then 10 min activation at 95°C for 35 PCR cycles of 15 s at 95°C for denaturation, and 60 s at 60°C for annealing and elongation. The melting curve analysis consisted of 3 s at 60°C followed by heating to 95°C with a ramp rate of 1°C/3 s. To determine steady state mRNA concentration, each sample was loaded 1-3 times in the array and each primer pair was loaded 2-5 times for technical replicates. Quantification of mRNA decay was

performed in the same manner resulting in 12-30 technical replicates for each biological sample in the degradation kinetics and for each primer pair.

Data analysis and statistical treatment

For quantification using low and high throughput technology, Ct values were determined with automatic baseline detection. For direct comparison, results were expressed as differences (Fold Change) between strains relative to the control strain (containing wt-*lacZ*). The Pfaffl analysis method was applied (32), considering ΔCt ratio between strains exclusively for the same primer pair. Results are expressed as means of fold change with standard deviation of biological and technical replicates.

For determination of local *lacZ* mRNA concentration between strains, Ct values were compared to a normalization range made with an 8-log dilution of pBAD-*lacZ* plasmid and expressed as $\Delta\Delta$Ct after normalization by the ampicillin resistance gene carried by the plasmid to avoid any effect due to putative plasmid copy number changes between strains (32). We verified that same results were obtained with *ihfB* normalization confirming that the plasmid copy number was unmodified between the analyzed strains. For variations of local concentrations along one mRNA molecule (intra-strain), Ct values were compared to the pBAD-*lacZ* plasmid 8-log range and then expressed as Fold Change compared to the values for amplicon “a” located at the 5’ extremity of the mRNA ($\Delta\Delta$Ct). We verified that intrer-strain variations of Ct values for *ihfB* and *bla* mRNAs were not significant.

To determine local mRNA half-life for each technical replicate and for each primer pair, Ct were plotted as a function of time after rifampicin addition (0, 0.3, 1, 2, 4 and 7 min). The mRNA half-life ($t_{1/2}$) was calculated from the degradation rate constant (k) corresponding to the slope of the Ct versus time curve with the relation $t_{1/2} = 1/k$. Only slopes with a R^2>0.85 were considered. The $t_{1/2}$ measured for each repetition and primer pair are expressed as mean and standard deviation.

RESULTS

Reducing the efficiency of translation initiation destabilizes the mRNA evenly throughout the molecule and decreases its overall concentration

To investigate the coupling between translation and mRNA degradation, we first modulated the translation initiation efficiency. We used as model reporter the *lacZ* mRNA coding for the β-galactosidase carried by a plasmid. The wild type version is composed of a 3267 nucleotide (ntd) long coding sequence (CDS) and a 33 ntd 5'UTR sequence containing the close to optimal Shine-Dalgarno (SD) sequence (AGGAGG) to sustain efficient translation initiation as close to the consensual 16S anti-SD sequence (33). To reduce translation initiation efficiency, we recoded the SD sequence to a very divergent version far from the consensual one (TTATAA) (*SD-lacZ*, Fig 1A). We confirmed that the amount of β-galactosidase produced with the inefficient SD was about 10-fold lower than that of the reference SD (Fig 1B). Then we analysed the ribosome repartition on the *wt-lacZ* and *SD-lacZ* mRNAs. As shown in Figure 1C, for *wt-lacZ* with efficient translation initiation, the majority of *lacZ* mRNAs were in the more heavily ribosome-loaded fractions F and G corresponding to molecules with more than 8 bounded ribosomes. Only a small proportion of *wt-lacZ* mRNAs were free of ribosomes (fraction A). These results showed that almost all *wt-lacZ* mRNAs were being translated and with a high number of ribosomes. With an inefficient SD sequence, the proportion of *SD-lacZ* mRNAs loaded with more than 10 ribosomes (fraction G) has significantly decreased in favour of mRNAs less loaded with ribosomes (fractions B, D, E and F) or even without ribosomes (fraction A). Overall, these results showed that reducing the efficiency of translation initiation decreased the proportion of translated *lacZ* mRNAs and, when in translation, they were loaded with fewer ribosomes.

We then measured the results of reducing the efficiency of translation initiation on *lacZ* mRNA stability and concentration. mRNA quantifications are classically performed by Northern blots using a single probe or by qRT-PCR with one amplicon. Both lead to partial information because they target only one part of the analyzed molecule. Here we developed a more comprehensive qRT-PCR approach at the molecule scale to measure local mRNA concentrations and local stabilities throughout the molecule. A set of primer pairs distributed throughout the *lacZ* mRNA molecule, spanning the 5' to 3' end, was designed to map the behaviour of different portions of the mRNA molecule (for amplicon boundary locations, see materials and methods section). Using this approach, we were able to identify both local differences between constructs and between portions of a molecule type in the cell.

For *wt-lacZ* and *SD-lacZ*, we quantified at eight different mRNA locations (Fig 2A) the local half-lives (Fig 2B). For *wt-lacZ,* the local stabilities were rather constant throughout the molecule, with local

half-lives ranging from 2.8 to 2.2 min with an average half-life at 2.4 min (Fig 2C). We can just note a slight increase in local stability at the 5' end of the mRNA. Although local stabilities of *SD-lacZ* mRNA were also constant throughout the molecule, they were all greatly reduced with an average half-life falling to 1 min (Fig 2B, 2C). This shows that reducing translation efficiency strongly destabilized the *lacZ* transcript homogenously throughout the molecule. To test whether the observed decrease in stability of *SD-lacZ* mRNAs was linked to changes in the level of RNA degradation machinery, we quantified the transcript concentrations of a panel of enzymes (RNase E, RNase G, RNase R, PNPase, PAP-I, Eno and Hfq) related to mRNA degradation (Supplementary Fig S1-A). Because their transcript levels were not significantly altered between *wt-lacZ* and *SD-lacZ,* we concluded that the destabilization of *SD-lacZ* mRNAs could not be attributed to changes in the expression of the RNA degradation machinery.

We then determined the effect of reducing the efficiency of translation initiation on the local mRNA concentrations throughout the molecule (Fig 2D). Changes in local mRNA concentrations are expressed as intramolecular fold change relative to 5' end of the molecule (using amplicon "a"). For *wt-lacZ* and *SD-lacZ*, local mRNA concentrations did not vary significantly along the molecules. Averaging the local concentrations over the entire mRNA, the transcript concentration was lower for *SD-lacZ* than for *wt-lacZ* (Fig 2E). The reduction of translation initiation efficiency reduced the average concentration of *SD-lacZ* mRNAs to less than 60% of that of *wt-lacZ.* This decrease in mRNA concentrations is proportional to the destabilization of *SD-lacZ* mRNA (almost half reduction for each). Since mRNA concentration results from a balance between its synthesis via transcription and its degradation, this suggests that transcription is not strongly modified in our condition.

In conclusion, the reduced efficiency of translation initiation resulted in globally less ribosome-loaded, less stable, and concentrated *lacZ* mRNAs, all leading to reduced synthesis of β-galactosidase protein. However, this did not cause locally different stabilities and concentrations throughout the molecule. The nearly constant half-lives measured throughout the molecules, regardless of the translation initiation efficiency applied, demonstrates the absence of accumulation of degradation intermediates.

Premature termination of translation elongation destabilizes the mRNA evenly throughout the molecule and decreases local mRNA concentrations after the stop codon

Our objective was to study the effect of the number of ribosomes in translation on stabilities and concentrations of mRNAs without changing their density on the coding sequence. For this end, we constructed a set of mRNA versions with similar translation initiation but with reduced *lacZ* coding sequence length. By inserting an internal stop codon (UAG amber codon) we generated mRNAs with reduced *lacZ* CDS length to 75%, 50%, 25% and 0.5 % of the initial one (Fig 3A). This resulted in mRNAs with hybrid coverage by ribosomes: a part of the mRNA can be translated up to the inserted stop codon and the rest of the molecule remains free of ribosomes. As expected, truncated (Supplementary Fig S2) and inactive (data not shown) forms of β-galactosidase were produced when expressing mRNAs with reduced *lacZ* coding sequence length. We analysed their ribosome load by polysome profiling experiments. A small proportion of mRNAs was measured in the ribosome-free fraction (Fraction A) regardless of *lacZ* CDS length, showing that most mRNAs were engaged in translation as expected since the efficiency of translation initiation was not altered. More precisely, mRNAs corresponding the full-length and 75%-length *lacZ* CDSs were heavily ribosome loaded, with the majority of mRNAs bound with more than 8 ribosomes per molecule (in fractions F and G). For shorter CDS, the ribosome load was shifted to lighter fractions corresponding to around 4-6 ribosomes for *lacZ-Stop*$_{50\%}$ and 1-3 ribosomes for *lacZ-Stop*$_{25\%}$ and *lacZ-Stop*$_{0.5\%}$ (the two consecutive fractions with the highest amounts were in fractions D-E and in fractions B-C, respectively).

The local half-lives of the different *lacZ-Stop* mRNAs were measured using primers distributed along the molecule and on either side of the inserted stop codon (Fig 4A). The *wt-lacZ* data of the full-length *lacZ* CDS in Figure 4B correspond to the *wt-lacZ* data presented in Figure 2B. When the length of *lacZ* CDS is reduced, all the transcripts were destabilized homogenously throughout the molecule (Fig 4B) resulting in average half-lives on the order of one minute, regardless of the size of the CDS reduction (Fig 4C). This means that the local stabilities were similar in the ribosome-covered and ribosome-free parts of the mRNA molecule. Furthermore, no major differences in local stabilities were observed between molecules with a large translated portion (*lacZ* Stop$_{75\%}$) or virtually untranslated version (*lacZ* Stop$_{0.5\%}$). The lack of increase in destabilization with increasing length of untranslated portion suggests that only the presence of an untranslated portion of the mRNA governs the destabilisation of the transcript, independent of its length. Again, we verified that this destabilization was not related to decreased expression of genes related to RNA degradation (Supplementary Fig 1-B). The constancy of measured local half-lives suggests that also in the case of premature translation elongation termination no degradation intermediates accumulated into the cells.

Then we analysed the effect of premature termination of translation elongation on local concentrations of mRNAs with reduced *lacZ* CDS length to 75%, 50% and 25% relative to the full-length CDS in the *wt-lacZ* (Fig 4D). Changes in local mRNA concentrations are expressed as intramolecular fold change relative to 5' end of the molecule (using amplicon "a"). For the *wt-lacZ* mRNA, the local concentrations only slightly decreased from the 5' to the native stop codon and then sharply at the level of the native stop codon. Similar biphasic decreasing behaviour of local concentrations upstream and downstream of the stop codon was observed in the case of premature termination of translation elongation. In the case of internal stop codons, the sharp decrease in local concentrations of the untranslated parts resulted in a basal level of about 5% relative to the 5'end. As we described above that local half-lives were similar upstream and downstream of the internal stop codon (Fig 4B), the decrease in local concentrations after the stop codon cannot be attributed to a different stability between translated and untranslated parts of the mRNA.

To quantify the impacts of the inserted stop codon on the *lacZ* mRNA concentration between strains, we compared 5'end mRNA concentration between the full-length *lacZ* CDS and the 75%, 50% and 25% reduced forms (Fig 4E). Reducing *lacZ* CDS length decreased the 5'end mRNA concentration into the cells, the more the CDS length was reduced the more the concentration decreased. This reduction in mRNA concentration at the 5'end may be at least partially related to the destabilization of the molecule described above (Fig 4B).

The untranslated mRNA portion is responsible of the destabilization of the whole mRNA molecule

We have shown in the previous section that the presence of an untranslated portion destabilizes the whole *lacZ* mRNA and this destabilization is not proportional to the length of the untranslated portion. To better understand the role of the untranslated portion in mRNA destabilization, we removed the untranslated portion to generate a new set of *lacZ* mRNAs (Fig 5A). For all new constructs, 5'UTR and 3'UTR were identical and only the length of the *lacZ* CDS was reduced to 75%, 50% and 25% from the *wt-lacZ,* hereafter referred to as $lacZ_{75\%}\Delta$, $lacZ_{50\%}\Delta$ and $lacZ_{25\%}\Delta$, respectively.

We measured the local half-lives of the portion of the mRNA corresponding to the remaining *lacZ* CDS region. All mRNAs showed a similar constant local stability profile, independent of the length of the remaining CDS region and close to the profile of the full-length *lacZ* CDS of the *wt-lacZ* strain (Fig 5B). This means that removal of the untranslated portion of the CDS increased the overall mRNA half-lives from 1 minute (Fig 4C) to 2-2.6 min (Fig 5C), restoring stability to the full-length CDS of the *wt-lacZ* strain (Fig 5C). We can conclude that the untranslated part of the CDS was the main determinant for triggering transcript destabilization. The restoration of stability was not due to changes in the

expression of genes involved in the degradation machinery since they showed similar expression in all strains (Supp Fig 1C).

In terms of mRNA concentration, a slight decrease in local concentrations in the portion of mRNA corresponding to the *lacZ* CDS region were obtained for all constructs (Fig 5D). These decreasing patterns were similar to those observed upstream of the stop codon when the untranslated portion remained (Fig 4D). However, after removing the untranslated portion we found increases in mRNA concentration measured at the 5'end (compare Fig 4E and Fig 5E).

Altogether, these results show that removing the untranslated portion of the *lacZ* CDS region restores the stability the mRNA and its 5'end concentration to that of the full-length *lacZ* CDS. Then we can conclude that after premature termination of translation elongation, it is the presence of an untranslated portion (whatever its length) on the mRNA that destabilizes the whole mRNA molecule leading to a decrease in its 5'end concentration. However, removal of the untranslated portion of the *lacZ* CDS region did not abolish the decrease in local concentrations upstream of the inserted stop codon. Because local stabilities were constant along the mRNA, we can speculate that changes of local concentrations observed here might instead be related to changes in transcription. Changes in transcription could also explain, when the untranslated portion of the *lacZ* CDS region is present, the drop in local concentrations downstream of the inserted stop codon while destabilization is similarly homogeneous (Fig 4B and 4D).

Drop in local mRNA concentrations downstream of the inserted stop codon is due to transcription arrest

Transcription and translation in bacteria are coupled, with translation initiation beginning before transcription is complete (34,35). This coupling occurs through an interaction between the pioneer ribosome and RNA polymerase. The interaction can be indirect, mediated by the NusG protein (17,36) or direct in a form of a one-to-one complex, at least *in vitro* (15,16). This coupling was proposed to allow the rate of translation elongation by the pioneer ribosome to dictate the rate of RNA polymerase transcription elongation (19) and also to prevent premature Rho-mediated transcription arrest (36). To test whether uncoupling transcription and translation that occurs shortly after the stop codon could lead to transcription elongation arrest and explain the drop in downstream mRNA concentrations, we developed a system to functionally restore translation of the inserted stop codon into the *lacZ* CDS in order to couple transcription and translation again.

The pEVOL system has the ability to genetically encode non-canonical amino acids (ncAA) directly in *E. coli* cells in response to nonsense codon (37). This system is based on the use of an orthogonal amino-acid tRNA synthetase ($aaRS_o$) and an orthogonal tRNA ($tRNA_o$). In the presence of ncAA, the $aaRS_o$ acylates the $tRNA_o$ with ncAA which is in turn used by ribosomes to decode amber stop codon to restore its translation instead of stopping it. We hijacked this system to restore translation of the *lacZ*-$Stop_{50\%}$ mRNA and measured the consequences on its local concentrations and stabilities. The $tRNA_o$ was inserted in transcriptional fusion with *lacZ-*$Stop_{50\%}$ mRNA to give *lacZ-*$Stop_{50\%}$*-*$tRNA_o$ and combined with a pEVOL plasmid derivative in which the $tRNA_o$ was deleted. The different constructs used and the principle of developed approach are depicted in Fig 6A. Briefly, all strains expressed the $aaRS_o$ and were cultivated in growth medium containing ncAA (AzF). For *lacZ-*$Stop_{50\%}$*-*$tRNA_o$, the $tRNA_o$ is transcribed, maturated, acylated with AzF by $aaRS_o$ and used by ribosomes to decode the stop codon. Once decoded, translation continues to produce full-length active β-galactosidase containing AzF. We developed a non-functional version (called tRNA*) in which the tri-nucleotide CCA motif of the tRNA end was removed. In *lacZ-*$Stop_{50\%}$*-tRNA*,* the tRNA* cannot be acylated, translation cannot be restored and a truncated inactive β-galactosidase is produced. Two control constructs were used, the *wt-lacZ* in transcriptional fusion with $tRNA_o$ (*wt-lacZ-*$tRNA_o$) as positive control and the *lacZ-*$Stop_{50\%}$ without $tRNA_o$ as negative.

We have measured the changes in local mRNA concentrations throughout the molecule for all the constructs (Fig 6B). Translation at the stop codon in *lacZ-*$Stop_{50\%}$*-*$tRNA_o$ using our orthogonal system resulted in constant local concentrations of the mRNA throughout the molecule. In contrast, a drop in local concentrations was observed downstream the stop codon in *lacZ-*$Stop_{50\%}$ in which translation stops at the stop codon, or in *lacZ-*$Stop_{50\%}$*-tRNA** in which a non-functional tRNA* prevents the restoration of orthogonal translation. The drop in local concentrations downstream of the stop codon was thus restored when translation elongation was not present. This clearly shows that the drop in local concentrations of the mRNA after the stop codon was due to the translation elongation arrest. Regarding local concentrations upstream of the stop codon, it is more difficult to conclude because we did not observe in the presence of functional orthogonal translation and even in the *wt-lacZ-*$tRNA_o$ positive control where native translation occurs the slight decrease seen previously (Fig 4D).

We compared the transcription level of three full-length transcripts, *wt-lacZ-*$tRNA_o$ (active native translation), *lacZ-*$Stop_{50\%}$*-*$tRNA_o$ (functional orthogonal translation) and *lacZ-*$Stop_{50\%}$*-tRNA** (non-functional translation) (Fig 6C). We used the amplicon "j" (Fig 6B) encompassing the last 77 nucleotides of the *lacZ* CDS and the $tRNA_o$ to measure the transcript concentration of the 3'end of the *lacZ* CDS and pre-tRNA before maturation. Compared with native translation, only 10 % of full-

length *lacZ-Stop$_{50\%}$-tRNA** was detected (Fig 6C) showing that the majority of transcription events were aborted when translation is stopped, probably due to uncoupling transcription and translation by the pioneer ribosome. A residual concentration of 10% of full-length mRNA also showed that uncoupling translation and transcription did not result in complete arrest of transcription and that some RNA polymerases are able to continue transcription even without the association with ribosomes. When translation was restored at the stop codon with our orthogonal system, the concentration of full-length *lacZ-Stop$_{50\%}$-tRNA$_o$* mRNAs was 60% of the concentration of full-length *wt-lacZ-tRNA$_o$* mRNAs with native translation. The lower full-length transcript concentration with the orthogonal translation than the native translation may be the result of the competition of tRNA$_o$ with endogenous tRNA for ribosome binding at the amber codon. Weaker ribosome binding on *lacZ-Stop$_{50\%}$-tRNA$_o$* mRNAs was confirmed at the protein level (Fig 6D). Only one-third of β-galactosidase activity was produced by orthogonal translation of *lacZ-Stop$_{50\%}$-tRNAo* mRNAs compared with native translation of *wt-lacZ-tRNA$_o$*. We confirmed that no active β-galactosidase was produced when native and orthogonal translations were not functional on the stop codon for *lacZ-Stop$_{50\%}$-tRNA**.

Altogether, we demonstrated that recovery of translation elongation at the inserted stop codon restores transcription elongation. The uncoupling of transcription and translation at the inserted stop codon was responsible for the drop in local transcript concentrations observed downstream of the stop codon, confirming the major function of pioneer ribosome in maintaining transcription elongation.

DISCUSSION.

In our overall goal to study the role of mRNA degradation in interaction with other cellular processes to regulate gene expression, we examined here the impact of translation on mRNA stability and concentration. We developed an approach to measure local stabilities and local concentrations all along the molecule. To our knowledge, measurements of local stabilities have never been included in studies of mRNA degradation. This intramolecular mapping provides more detailed information about the transcript behaviour when facing translational perturbations. Both the reduction in the efficiency of translation initiation and the introduction of premature termination of translation elongation resulted in a homogenous destabilization of the transcript throughout the molecule. The homogeneous stabilities indicate that we did not measure the accumulation of specific degradation products. This is in agreement with Northern blot quantifications of *lacZ* mRNA stability that also showed no accumulation of intermediates or smears (38,39). Several hypotheses can be proposed to explain the lack of observed accumulation of decay intermediates. The most plausible explanation

could be that the degradation fragments did not accumulate because they were degraded too rapidly. After the initial internal endonucleolytic cleavage, two degradation products are generated. The fragment with newly generated 5’ end can be efficiently processed by RNase E due to its increased activity towards 5’-monophosphorylated ends (40) while the newly generated 3' end can be rapidly processed by 3' exoribonucleases such as PNPase and RNase II due to the absence of protective secondary structure usually present at the 3’end of the full-length transcript (41). This explanation is based on the model that the initial internal cleavage is the limiting step in the degradation process followed by extremely rapid decay of intermediates (42). A second explanation to explain the lack of accumulation of decay intermediates could be the presence of a large set of heterogeneous fragments covering the whole mRNA molecule, generated by initial cleavages randomly distributed throughout the mRNA. In agreement with this hypothesis, Herzel et al (43) recently showed that a large fraction of cellular RNAs is composed of decay fragments with 3' ends widely distributed throughout internal positions in the transcript. In this case, this implies that the stabilities of the decay fragments are not higher than the stability of the full-length transcript. Finally, we cannot exclude that particular fragments may accumulate without being quantified as non-targeted by our qRT-PCR quantifications. We consider the latter explanation unlikely as we quantified 8 different mRNA portions widely distributed all along the molecule.

The role of ribosomes in linking translation and mRNA degradation was addressed more specifically in our study. Reducing translation initiation efficiency decreases the number of bound ribosomes and the stability of the entire mRNA molecule. These results confirm the protective role of ribosomes when translation initiation is efficient probably by preventing 5’ end-dependent degradation by RNase E (5,6) and extend ribosome protection to the whole molecule. In the case of premature termination of translation elongation, we have shown that the presence of an untranslated portion, regardless of its length, destabilizes the whole mRNA molecule. This also supports the protective role of ribosomes and, in this case, probably through steric hindrance of ribosomes against the action of RNases. The literature reports cases of transcript destabilization by ribosomes when ribosomes are stalled on the transcript or when they elongate at low rate (5,6). We did not measure here a negative effect of ribosomes on transcript stability since the same half-life was measured for the ribosome-covered translated portion and the ribosome-free untranslated portion of the mRNA. This was expected because insertion of an internal stop codon arrests translation and induces ribosome drop off but does not generate stalled ribosomes or affect the rate of upstream ribosome elongation.

When transcription is considered, it appears that inefficient translation initiation leading to destabilization of the whole mRNA did not strongly influence transcription. Premature termination of translation elongation also destabilized the whole mRNA but, in turn, also affected transcription.

Insertion of an internal stop codon resulted in premature downstream termination of transcription elongation and a sharp decrease in mRNA concentration. This is probably due the lack of coupling of transcription with translation via the interaction between the pioneer ribosome and the RNA polymerase downstream of the inserted stop codon (15-17). However, the coupling between transcription and translation is not an all-or-nothing relationship. A minority (~10%) of transcription elongation events leading to full-length transcripts were effective even when transcription and translation were uncoupled (Fig 6C). On the other hand, the coupling did not seem to ensure perfect full-length transcription. We measured a slight decrease in local mRNA concentrations upstream of the inserted stop codon while local stabilities were not affected (Fig 4D and Fig 4B). This suggests that even when translation and transcription are coupled, there may be a fraction of transcription events that do not synthesize the full-length transcript. This appears to be independent of the presence or absence of an untranslated region (Fig 4D and Fig 5D). This premature termination of transcription could generate heterogeneity in the transcript population. Further investigation is required to explore this hypothesis.

The coupling of the three processes of transcription, translation and mRNA degradation is essential for the proper functioning of the cell. This coupling would constitute a first line of translation quality control. In eukaryotes, the first round of translation by the pioneer ribosome serves as an mRNA quality control, as defects in maturation steps can lead to the emergence of alternative stop codons generating potentially toxic truncated proteins (44,45). These premature stop codons recognized by the pioneer ribosome lead to mRNA degradation by a mechanism called Nonsense Mediated Decay. In bacteria, transcription is not an error-free event, with an estimated frequency of ribonucleotide misincorporation of $5x10^{-5}$ to $5x10^{-6}$ per nucleotide in *E. coli* (46,47). For ~4% of cases, the misincorporation results in appearance of stop codons (46) and the production of defective proteins that can impair cellular integrity and fitness. This can be particularly deleterious for poorly transcribed genes for which only a few mRNA molecules are present in the cell to ensure translation. In this case, it is particularly necessary for the cell to remove this non-productive translation template. As shown in this study, as soon as a defect in coupling appears, cells respond by rapidly degrading the corrupted mRNA.

In conclusion, this study provides new evidence on how degradation of an mRNA molecule is in close relationship with its translational state. When the ribosome density is lower or in presence of a ribosome-free portion, the mRNA molecule is homogeneously destabilized. This global destabilization contributed to lowering its concentration while it was the changes in transcription

that explained the local differential concentrations along the mRNA molecule. Therefore, it is the interplay of the triptych transcription-translation- mRNA degradation that actively governs quality control and mRNA concentration. This “ménage à trois” is a central coordinated mechanism developed by bacteria to shape gene expression.

Conflict of interest.

The authors declare no conflict of interest.

Acknowledgments.

The authors thank M. Audonnet, P Lopez, A Emile for technical supports and AJ Carpousis for comments on the manuscript.

Authors contributions.

SN et LG conceived the study. FC performed constructions and analysis related to translation initiation modification. MPD performed part of stability analysis and all the polysome profiling experiments. SN performed all other experiments. SN, LG and MCB equally participated to the manuscript conception and writing.

References

1. Mohanty, B.K. and Kushner, S.R. (2016) Regulation of mRNA Decay in Bacteria. *Annual review of microbiology*, **70**, 25-44.
2. Callaghan, A.J., Aurikko, J.P., Ilag, L.L., Günter Grossmann, J., Chandran, V., Kühnel, K., Poljak, L., Carpousis, A.J., Robinson, C.V., Symmons, M.F. *et al.* (2004) Studies of the RNA degradosome-organizing domain of the Escherichia coli ribonuclease RNase E. *Journal of molecular biology*, **340**, 965-979.
3. Carpousis, A.J. (2007) The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. *Annual review of microbiology*, **61**, 71-87.
4. Anderson, K.L. and Dunman, P.M. (2009) Messenger RNA Turnover Processes in Escherichia coli, Bacillus subtilis, and Emerging Studies in Staphylococcus aureus. *International Journal of Microbiology*, **2009**, 525491.
5. Deana, A. and Belasco, J.G. (2005) Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes & development*, **19**, 2526-2533.
6. Dreyfus, M. (2009) Killer and protective ribosomes. *Progress in molecular biology and translational science*, **85**, 423-466.
7. Cambray, G., Guimaraes, J.C. and Arkin, A.P. (2018) Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nature biotechnology*, **36**, 1005-1015.
8. Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D. and Church, G.M. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 14024-14029.
9. Pedersen, S., Terkelsen, T.B., Eriksen, M., Hauge, M.K., Lund, C.C., Sneppen, K. and Mitarai, N. (2019) Fast Translation within the First 45 Codons Decreases mRNA Stability and Increases Premature Transcription Termination in E. coli. *Journal of molecular biology*, **431**, 1088-1097.
10. Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.H., Su, M., Luff, J., Valecha, M., Everett, J.K., Acton, T.B. *et al.* (2016) Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, **529**, 358-363.
11. Esquerré, T., Moisan, A., Chiapello, H., Arike, L., Vilu, R., Gaspin, C., Cocaign-Bousquet, M. and Girbal, L. (2015) Genome-wide investigation of mRNA lifetime determinants in Escherichia coli cells cultured at different growth rates. *BMC genomics*, **16**, 275.
12. Tsai, Y.C., Du, D., Domínguez-Malfavón, L., Dimastrogiovanni, D., Cross, J., Callaghan, A.J., García-Mena, J. and Luisi, B.F. (2012) Recognition of the 70S ribosome and polysome by the RNA degradosome in Escherichia coli. *Nucleic acids research*, **40**, 10417-10431.
13. Hamouche, L., Poljak, L. and Carpousis, A.J. (2021) Polyribosome-Dependent Clustering of Membrane-Anchored RNA Degradosomes To Form Sites of mRNA Degradation in Escherichia coli. *mBio*, **12**, e0193221.
14. Hadjeras, L., Poljak, L., Bouvier, M., Morin-Ogier, Q., Canal, I., Cocaign-Bousquet, M., Girbal, L. and Carpousis, A.J. (2019) Detachment of the RNA degradosome from the inner membrane of Escherichia coli results in a global slowdown of mRNA degradation, proteolysis of RNase E and increased turnover of ribosome-free transcripts. *Molecular microbiology*, **111**, 1715-1731.
15. Fan, H., Conn, A.B., Williams, P.B., Diggs, S., Hahm, J., Gamper, H.B., Jr., Hou, Y.M., O'Leary, S.E., Wang, Y. and Blaha, G.M. (2017) Transcription-translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic acids research*, **45**, 11043-11055.

16. Kohler, R., Mooney, R.A., Mills, D.J., Landick, R. and Cramer, P. (2017) Architecture of a transcribing-translating expressome. *Science (New York, N.Y.)*, **356**, 194-197.
17. Saxena, S., Myka, K.K., Washburn, R., Costantino, N., Court, D.L. and Gottesman, M.E. (2018) Escherichia coli transcription factor NusG binds to 70S ribosomes. *Molecular microbiology*, **108**, 495-504.
18. Iyer, S., Le, D., Park, B.R. and Kim, M. (2018) Distinct mechanisms coordinate transcription and translation under carbon and nitrogen starvation in Escherichia coli. *Nature microbiology*, **3**, 741-748.
19. Proshkin, S., Rahmouni, A.R., Mironov, A. and Nudler, E. (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science (New York, N.Y.)*, **328**, 504-508.
20. Vogel, U. and Jensen, K.F. (1994) The RNA chain elongation rate in Escherichia coli depends on the growth rate. *Journal of bacteriology*, **176**, 2807-2813.
21. Castellana, M., Hsin-Jung Li, S. and Wingreen, N.S. (2016) Spatial organization of bacterial transcription and translation. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 9286-9291.
22. Choe, D., Lee, J.H., Yoo, M., Hwang, S., Sung, B.H., Cho, S., Palsson, B., Kim, S.C. and Cho, B.-K. (2019) Adaptive laboratory evolution of a genome-reduced Escherichia coli. *Nature communications*, **10**, 935.
23. Dressaire, C., Pobre, V., Laguerre, S., Girbal, L., Arraiano, C.M. and Cocaign-Bousquet, M. (2018) PNPase is involved in the coordination of mRNA degradation and expression in stationary phase cells of Escherichia coli. *BMC genomics*, **19**, 848.
24. Nguyen, H.L., Duviau, M.P., Cocaign-Bousquet, M., Nouaille, S. and Girbal, L. (2019) Multiplexing polysome profiling experiments to study translation in Escherichia coli. *PloS one*, **14**, e0212297.
25. Nouaille, S., Mondeil, S., Finoux, A.L., Moulis, C., Girbal, L. and Cocaign-Bousquet, M. (2017) The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic acids research*, **45**, 11711-11724.
26. Etienne, T.A., Cocaign-Bousquet, M. and Ropers, D. (2020) Competitive effects in bacterial mRNA decay. *Journal of Theoretical Biology*, **504**, 110333.
27. Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A.J., Girbal, L. and Cocaign-Bousquet, M. (2014) Dual role of transcription and transcript stability in the regulation of gene expression in Escherichia coli cells cultured on glucose at different growth rates. *Nucleic acids research*, **42**, 2460-2472.
28. Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989) *Molecular cloning: a laboratory manual*, NY.
29. Nguyen, H.L., Duviau, M.P., Laguerre, S., Nouaille, S., Cocaign-Bousquet, M. and Girbal, L. (2022) Synergistic Regulation of Transcription and Translation in Escherichia coli Revealed by Codirectional Increases in mRNA Concentration and Translation Efficiency. *Microbiology spectrum*, **10**, e0204121.
30. Morin, M., Ropers, D., Letisse, F., Laguerre, S., Portais, J.C., Cocaign-Bousquet, M. and Enjalbert, B. (2016) The post-transcriptional regulatory system CSR controls the balance of metabolic pools in upper glycolysis of Escherichia coli. *Molecular microbiology*, **100**, 686-700.
31. Spurgeon, S.L., Jones, R.C. and Ramakrishnan, R. (2008) High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PloS one*, **3**, e1662.
32. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, **29**, e45.
33. Shine, J. and Dalgarno, L. (1974) The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences*, **71**, 1342-1346.
34. Miller, O.L., Jr., Hamkalo, B.A. and Thomas, C.A., Jr. (1970) Visualization of bacterial genes in action. *Science (New York, N.Y.)*, **169**, 392-395.

35. Stent, G.S. (1964) THE OPERON: ON ITS THIRD ANNIVERSARY. MODULATION OF TRANSFER RNA SPECIES CAN PROVIDE A WORKABLE MODEL OF AN OPERATOR-LESS OPERON. *Science (New York, N.Y.)*, **144**, 816-820.
36. Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E. and Rösch, P. (2010) A NusE:NusG complex links transcription and translation. *Science (New York, N.Y.)*, **328**, 501-504.
37. Young, T.S., Ahmad, I., Yin, J.A. and Schultz, P.G. (2010) An enhanced system for unnatural amino acid mutagenesis in E. coli. *Journal of molecular biology*, **395**, 361-374.
38. Iost, I. and Dreyfus, M. (1995) The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO journal*, **14**, 3252-3261.
39. Komarova, A.V., Tchufistova, L.S., Dreyfus, M. and Boni, I.V. (2005) AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in Escherichia coli. *Journal of bacteriology*, **187**, 1344-1349.
40. Mackie, G.A. (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature*, **395**, 720-723.
41. Andrade, J.M., Pobre, V., Silva, I.J., Domingues, S. and Arraiano, C.M. (2009) The role of 3'-5' exoribonucleases in RNA degradation. *Progress in molecular biology and translational science*, **85**, 187-229.
42. Carpousis, A.J., Luisi, B.F. and McDowall, K.J. (2009) Endonucleolytic initiation of mRNA decay in Escherichia coli. *Progress in molecular biology and translational science*, **85**, 91-135.
43. Herzel, L., Stanley, J.A., Yao, C.-C. and Li, G.-W. (2022) Ubiquitous mRNA decay fragments in E. coli redefine the functional transcriptome. *bioRxiv*, 2021.2011.2018.469045.
44. Maquat, L.E., Tarn, W.Y. and Isken, O. (2010) The pioneer round of translation: features and functions. *Cell*, **142**, 368-374.
45. McGlincy, N.J. and Smith, C.W. (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in biochemical sciences*, **33**, 385-393.
46. Li, W. and Lynch, M. (2020) Universally high transcript error rates in bacteria. *eLife*, **9**, e54898.
47. Traverse, C.C. and Ochman, H. (2016) Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proceedings of the National Academy of Sciences*, **113**, 3311-3316.

Figure 1.

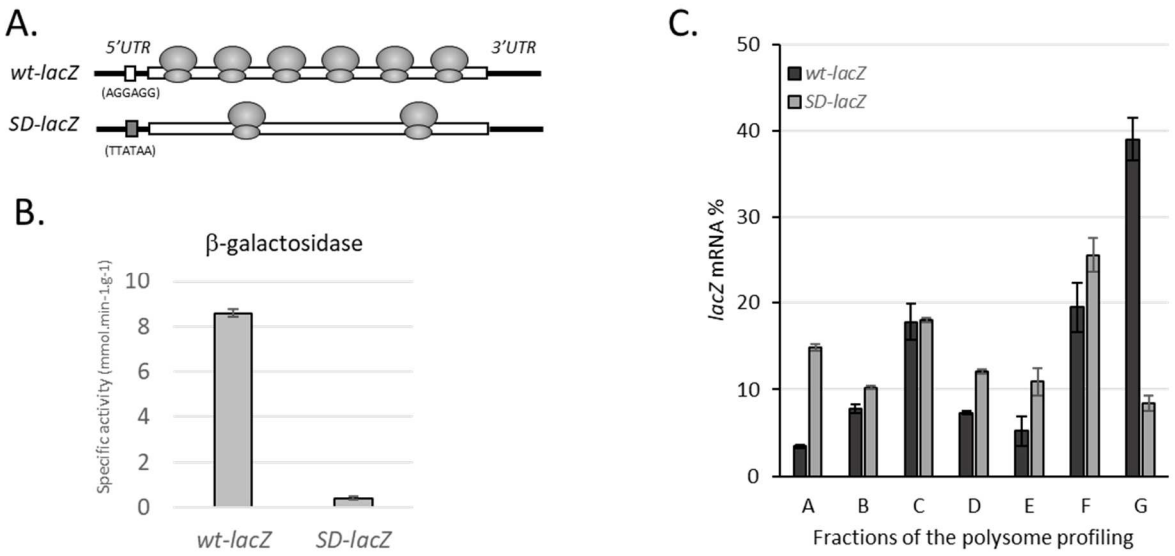


Figure 1.

A: Schematic representation of *lacZ* mRNA loaded with different number of ribosomes due to different translation initiation efficiency. *wt-lacZ*, reference mRNA with efficient translation efficiency, *SD-lacZ*, mRNA with reduced translation initiation efficiency. The two Shine-Dalgarno (SD) sequences are symbolized and their sequence indicated.

B: b-galactosidase activity measured for *wt-lacZ* and *SD-lacZ* containing strains. Error bars represent standard deviation (n=9, 3 biological and 3 technical replicates).

C: Ribosome repartition on *wt-lacZ* (dark grey) and *SD-lacZ* (light grey) mRNAs. Fraction A consists of free mRNA molecules not bound to the ribosome, while fractions B to G consist of mRNA molecules in translation bound to increasing numbers of ribosomes (according to ((Nguyen 2022)): B: 1.1 ± 0.1, C: 2.8 ± 0.3; D: 4.7 ± 0.4; E: 5.9 ± 0.5; F: 8.2 ± 0.6 and G: 11.4 ± 0.4 bound ribosomes per mRNA molecules). The percentage of *lacZ* mRNA in a fraction represents the amount of that mRNA in that fraction relative to the sum of the amounts measured in all fractions. Error bars represent standard deviation of two technical replicates.

Figure 2.

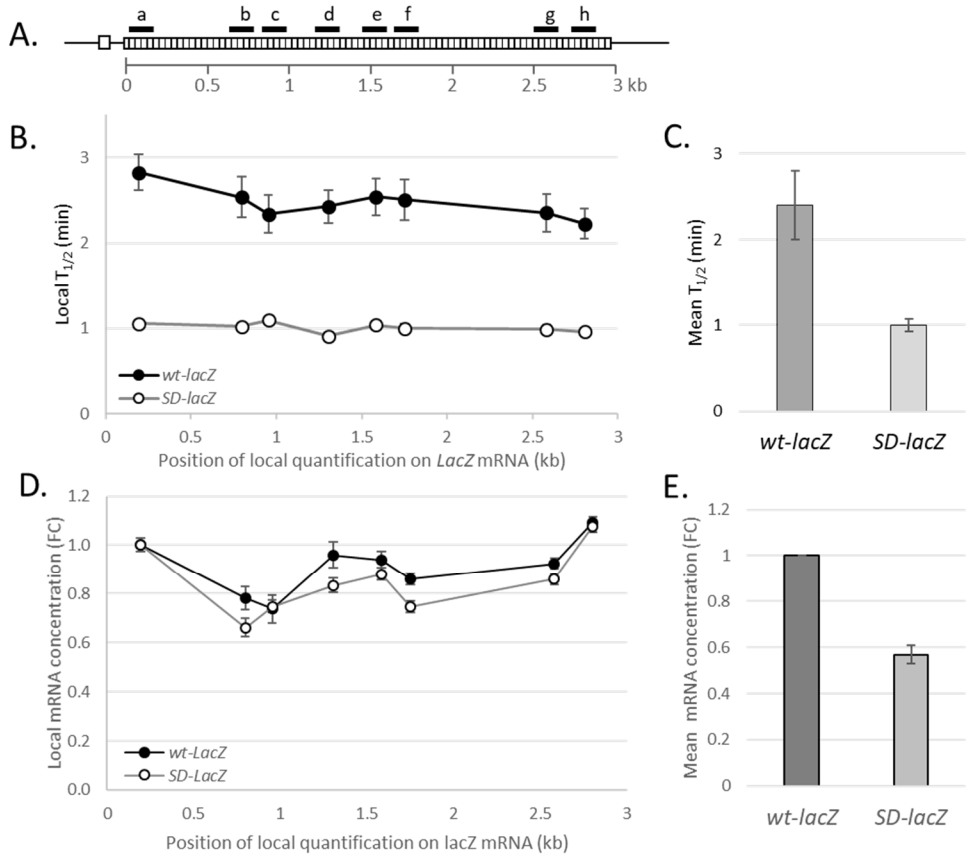


Figure 2.

A. Schematic representation of *lacZ* mRNA with the CDS part (stripped rectangle). Black marks with letters indicate the amplicons used for qRT-PCR quantification, with their position on the mRNA relative to the first codon of the CDS (scale in kb). See Materials and methods for precise coordinates. White square at the 5'end represents SD sequence (AGGAGG for *wt-lacZ*, TAATTA for *SD-lacZ*).

B. Local *lacZ* mRNA stabilities. Local half lives (T1/2) were determined at different positions throughout the mRNA molecules using different primer pairs (qPCR amplicon localizations are shown on the above scheme). Error bars denote standard deviation (n=5 to 10).

C. Mean half-lives of the whole *lacZ* mRNA molecules for *wt-lacZ* and *SD-lacZ*. Error bars represent standard deviation (n=40 to 80).

D. Local *lacZ* mRNA concentrations. The local *lacZ* mRNA concentrations were determined at different positions of the molecule using different primer pairs by qRT-PCR. Expressed as fold

change (FC) of concentrations relative to the concentration determined for *wt-lacZ* by the amplicon “a”. Error bars denote standard deviation (n=5 to 10).

E. Mean mRNA concentration of the whole *lacZ* mRNA molecules. *SD-lacZ* concentration expressed as fold change (FC) relative to *wt-lacZ*. Error bars denote standard deviation (n=8).

Figure 3.

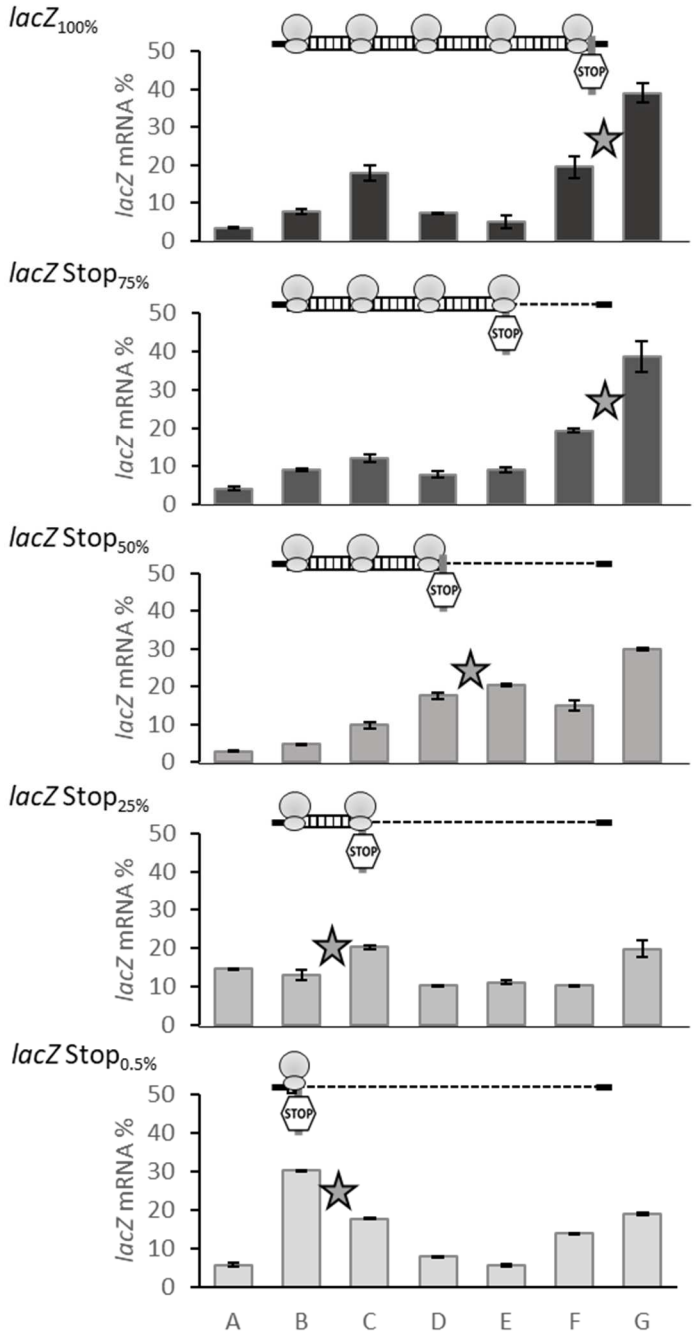


Figure 3.

Ribosome repartition on *lacZ* mRNAs with length of CDS from 100% to 0.5%. mRNA-ribosome complexes were fractionated on sucrose gradient, *lacZ* mRNA was then purified and quantified in each fraction. Fraction A consists of free mRNA molecules not bound to the ribosome, while fractions B to G consist of mRNA molecules in translation bound to increasing numbers of ribosomes (according to Nguyen 2022: B: 1.1 ± 0.1, C: 2.8 ± 0.3; D: 4.7 ± 0.4; E: 5.9 ± 0.5; F: 8.2 ± 0.6 and G: 11.4 ± 0.4 bound ribosomes per mRNA molecules). The percentage of *lacZ* mRNA in a fraction represents

the amount of that mRNA in that fraction relative to the sum of the amounts measured in all fractions. Error bars represent standard deviation of two technical replicates. Construct present in each strain is schematized above each graph with the CDS part (stripped rectangle) and the non-translated part after the inserted stop codon (dashed line). Translating ribosomes are symbolized. The large star indicates the two consecutive fractions with the highest amounts of *lacZ* mRNAs.

Figure 4.

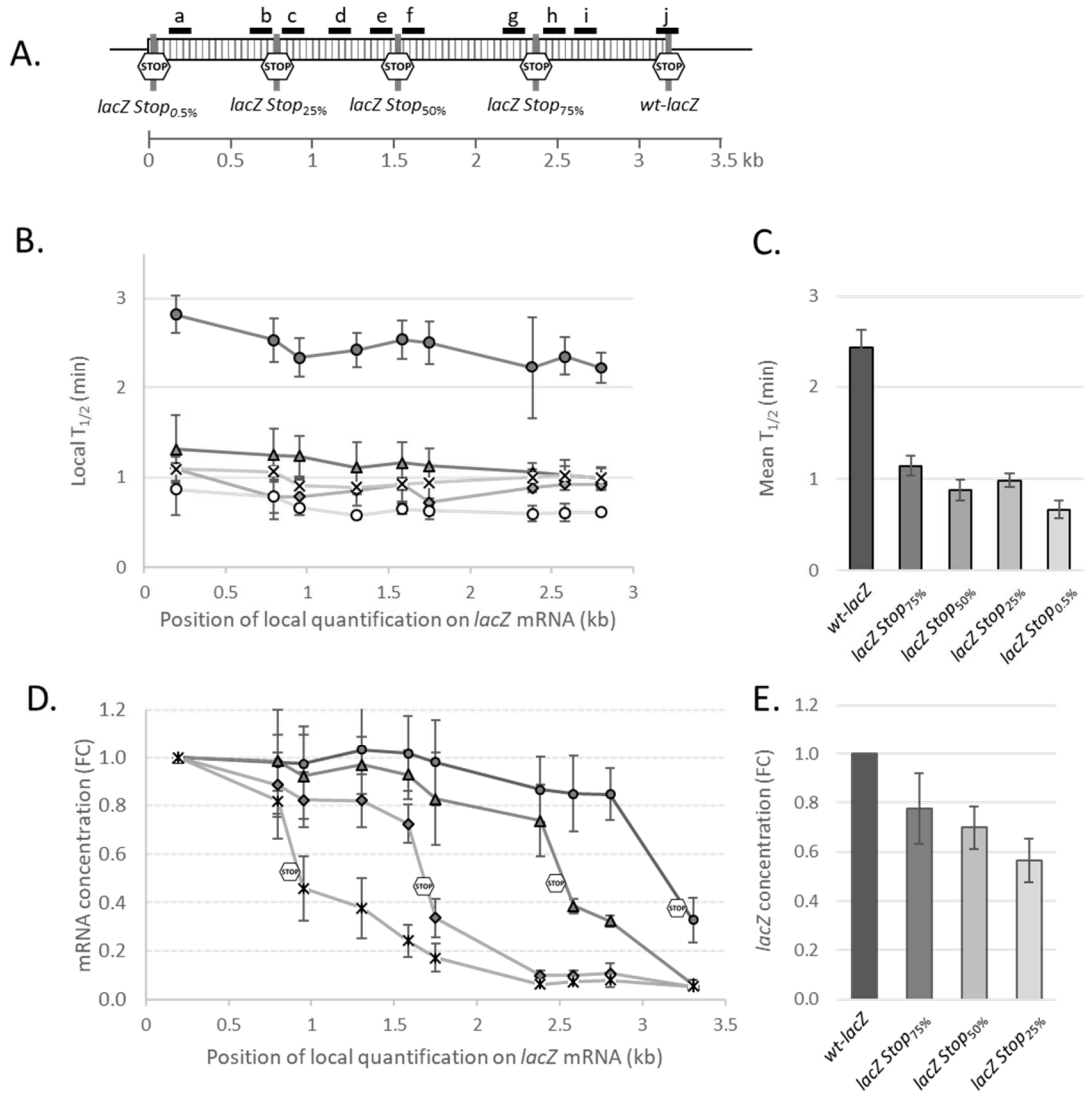


Figure 4.

A. Schematic representation of analyzed *lacZ* mRNAs. Different positions of the stop codon (STOP logo) for the different versions of *lacZ* mRNA are represented as well as the proportion of remaining CDS relative to the *wt-lacZ*. Black marks with letters indicate the amplicons used for qRT-PCR quantification, and their position on the mRNA relative to the inserted stop codon. The scale corresponds to *lacZ* mRNA (in kb) relative to the start codon. See Materials and methods for precise coordinates.

B. Local *lacZ* mRNA stabilities. Local half lives (T1/2) were determined for full length *lacZ* CDS (black circles) and *lacZ* CDS corresponding to 75 % (black triangles), 50% (black diamonds), 25% (crosses) and 0.5% (open diamonds) of the original coding sequence. Quantification

positions are aligned with amplicon locations shown in the upper scheme. Error bars are standard deviation (n=10)

C. Mean half-lives of the whole *lacZ* mRNA molecules for *wt-lacZ* and *SD-lacZ*. Error bars represent standard deviation (n=9)

D. Local *lacZ* mRNA concentration. The local *lacZ* mRNA concentrations were determined for full length *lacZ* CDS (black circles) and *lacZ* CDS corresponding to 75 % (black triangles), 50% (black diamonds), 25% (crosses) and 0.5% (open diamonds) of the original coding sequence. Quantification positions are aligned with amplicon locations shown in the upper scheme. Expressed as fold change relative to the concentration determined with amplicon "a". Error bars represent standard deviation (n=5 to 10). STOP logos symbolize the locations of the stop codon position along the mRNAs.

E. mRNA concentration of *lacZ* mRNAs. Concentrations were determined using amplicon "a" and expressed as fold change (FC) relative to *wt-lacZ*. Error bars denote standard deviation (n=12 to 20).

Figure 5.

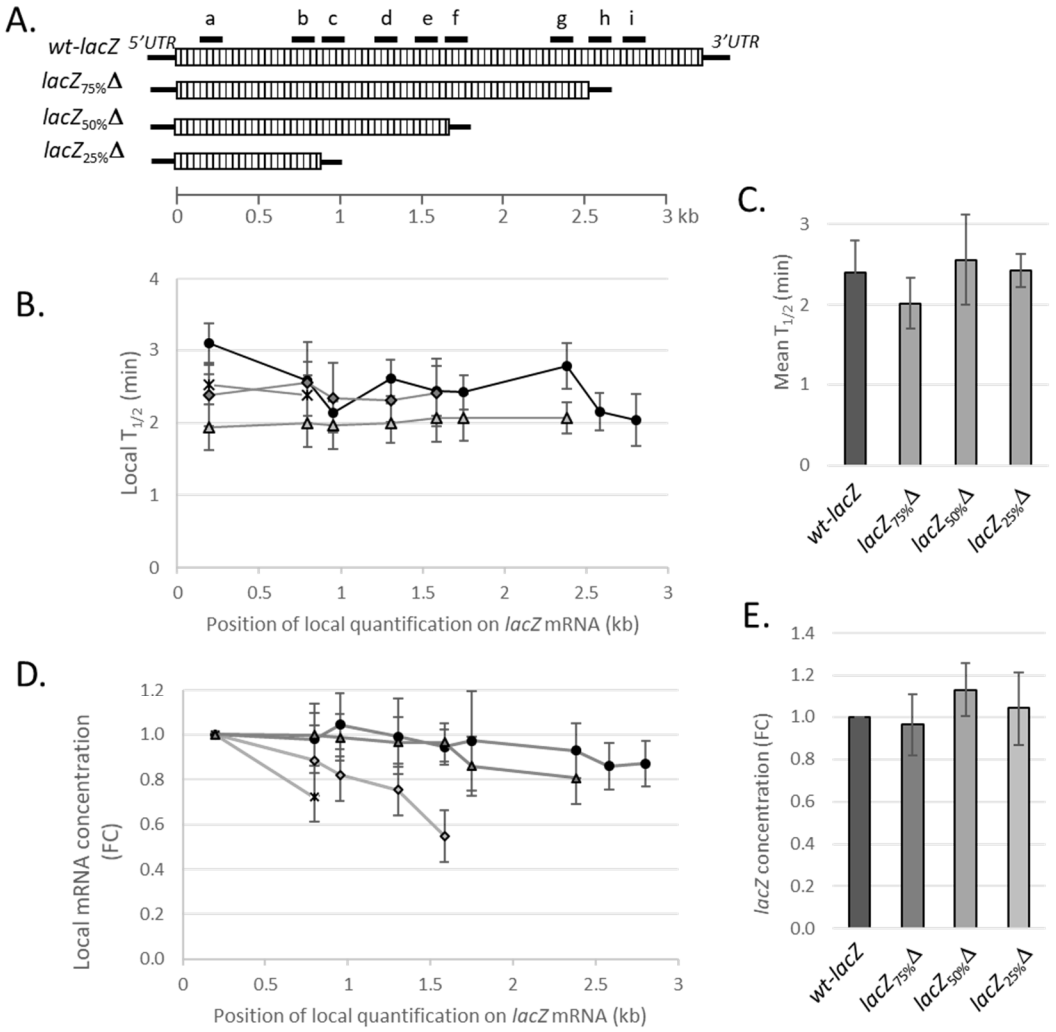


Figure 5.

A. Schematic representation of *lacZ* mRNAs with variable CDS length. Black marks with letters indicate the amplicons used for qRT-PCR quantification, with their relative position on the mRNA. The scale corresponds to *lacZ* mRNA (in kb) relative to the start codon. See Materials and methods for precise coordinates.

B. Local *lacZ* mRNA stabilities. Local half lives ($T_{1/2}$) were determined using primer pairs located along the molecule of *wt-lacZ (*circles), *lacZ*$_{75\%}$D (triangles), *lacZ*$_{50\%}$D (diamonds) and *lacZ*$_{55\%}$D (crosses). Error bars denote standard deviation (n=4 to 10).

C. Mean half-lives of each *lacZ* mRNA. Values are the average of all quantifications with each pair of primer for the corresponding *lacZ* variant. Error bars represent standard deviation (n=8 to 45)

D. Local mRNA concentrations. The *lacZ* mRNA local concentrations were determined for full-length *wt-lacZ* CDS (black circles) and versions after removing the untranslated portion*:* *lacZ*$_{75\ \%}$D (black triangles), *lacZ*$_{50\ \%}$D (black diamonds) and *lacZ*$_{25\ \%}$D (crosses). Quantification positions are aligned with amplicon locations shown in the upper scheme. Expressed as fold change compared with the concentration determined with amplicon "a". Error bars represent standard deviation (n=5 to 10).

E. Relative *lacZ* mRNA concentration between strains. mRNA concentration was determined using the "a" amplicon and expressed as fold change relative to *wt-lacZ*. Error bars represent standard deviation (n=6 to 18).

Figure 6.

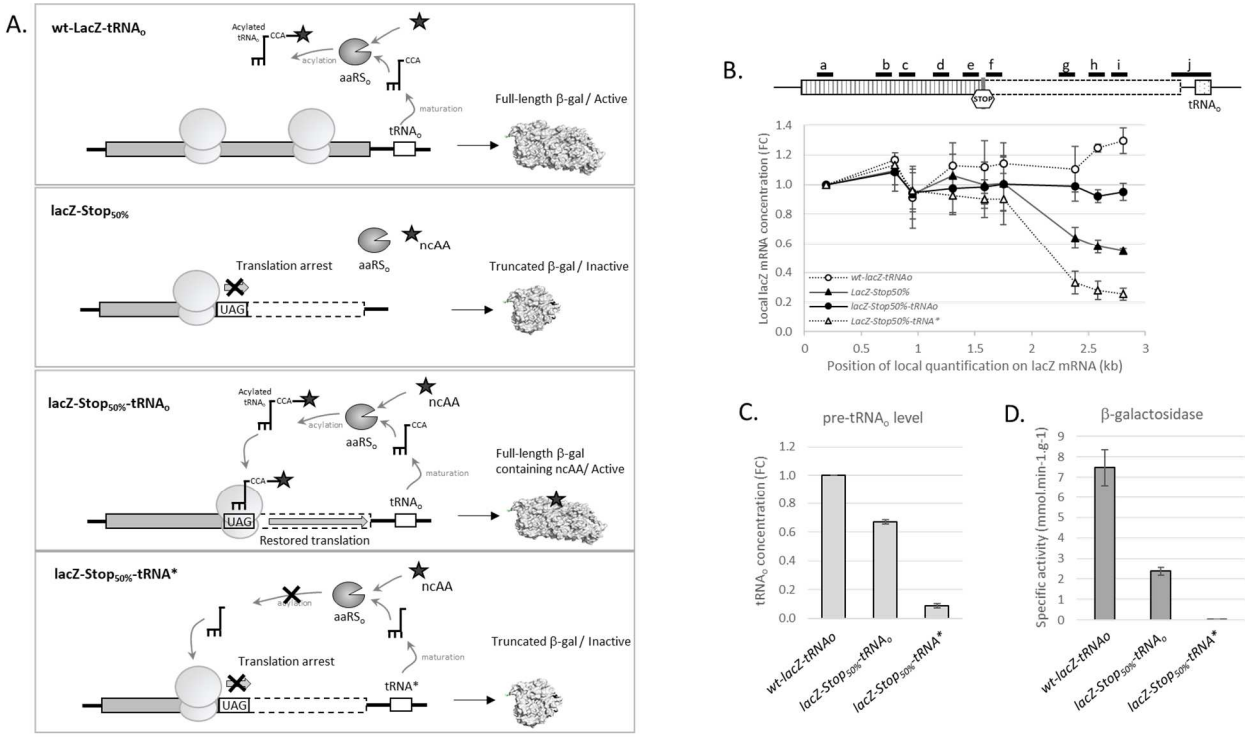


Figure 6.

A. Schematic representation of the mRNAs used in the study. In the upper panel (*wt-lacZ-tRNA*$_o$) the full-length CDS of *lacZ* (grey) is in transcriptional fusion with the orthogonal functional tRNA$_o$ (white rectangle). Ribosomes translate the full-length active protein. In the second panel (*lacZ*-Stop$_{50\%}$), the CDS is reduced by half by insertion of a stop codon. Ribosomes translate a truncated and inactive form of the protein. In the third panel representing lacZ-Stop$_{50\%}$-tRNA$_o$, orthogonal tRNA$_o$ is matured and acylated with a non natural amino acid (AzF, star) by an orthogonal tRNA synthetase (aaRS$_o$). Ribosomes can load tRNA$_o$ to decode the amber codon and continue translation to produce a full-length active protein containing the ncAA. The fourth panel represents the *lacZ* mRNA with amber codon and fused with an inactive version of tRNA (tRNA*) due to removal of the acylation motif (CCA) at the 3'end of the mature tRNA. tRNA* cannot be acylated by the aaRS$_o$ and cannot be used by the ribosome to decode the amber codon. A truncated and inactive form of protein is produced.

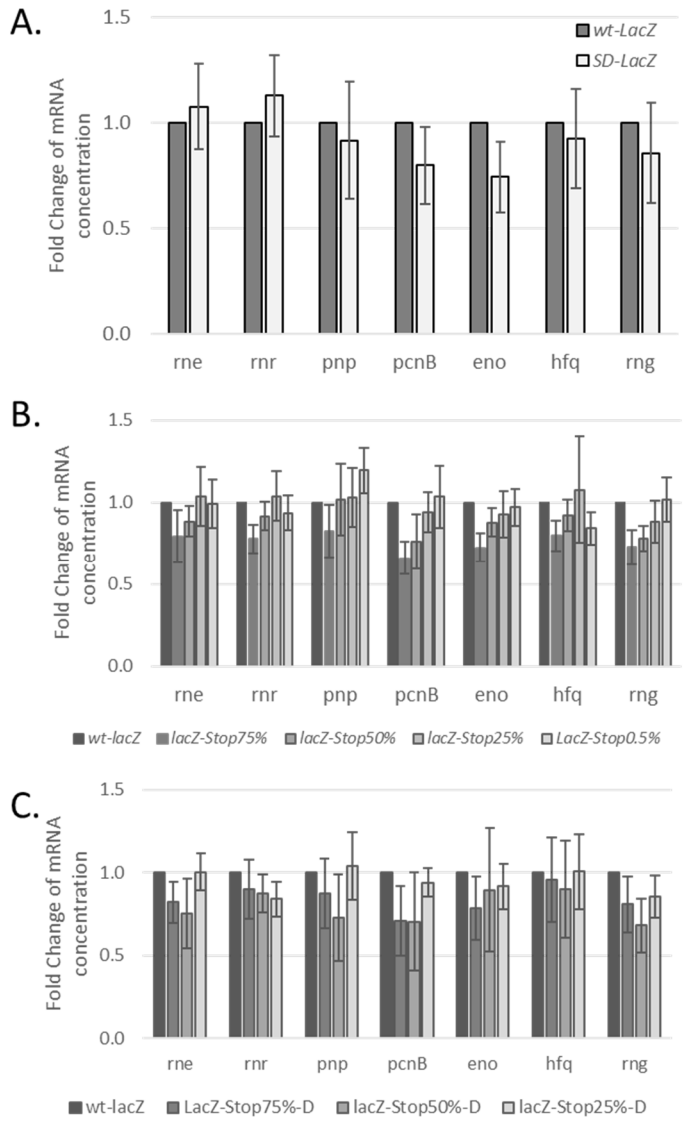
B. Schematic representation of *lacZ* – tRNA with positioning of qPCR amplicons (a to j). Local *lacZ* mRNA concentrations were measured for full-length *lacZ* fused to functional tRNA$_o$ (*wt-lacZ-tRNA*$_o$, open circles), *lacZ* with an amber codon without tRNA (*lacZ-Stop*$_{50\%}$, black triangles), *lacZ* with an amber codon and with a functional tRNA$_o$ (*lacZ-Stop*$_{50\%}$*-tRNA*$_o$, black

circles) and *lacZ* with an amber codon and a non-functional tRNA (*lacZ-Stop$_{50\%}$-tRNA*,* open triangles). Quantification positions are relative to their position along the mRNA (in kb). Concentration is expressed as fold change compared with the quantification of amplicon "a" . Error bars represent standard deviation (n= 5 to 10).

C. Quantification of pre-tRNA. To quantify the transcription level of tRNA before its maturation, we used a pair of primers spanning from the end of the *lacZ* CDS to the end of the tRNA (amplicon "j" on the above scheme). Concentration is expressed as fold change compared with *wt-lacZ-tRNA$_o$*. Error bars represent standard deviation (n=5)

D. Quantification of the specific activity of b-galactosidase produced by the constructs. Error bars represent standard deviation (n=6).

Supplementary Figure S1.



Supplementary Figure S1.

Expression of genes related to RNA metabolism

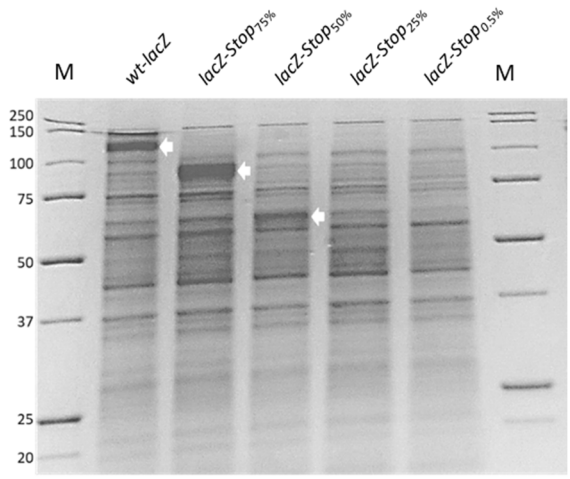
Transcript concentration of RNase E (*rne*), RNase R (*rnr*), Polynucleotide phosphorylase (*pnp*), Poly(A)polymerase (*pcnB*), Enolase (*eno*), RNA-binding protein Hfq (*hfq*) and RNase G (*rng*) were quantified and expressed as fold change compared to the *wt-lacZ*.

A. *wt-lacZ* (dark grey) and *SD-lacZ* with reduced translation initiation efficiency (light grey). Concentration expressed as fold change relative to the *wt-lacZ*. Error bars represent standard deviation (n=4).

B. : *lacZ* CDS with an inserted stop codon at different positions of the CDS leading to untranslated portions of the molecule. Color/construct correspondences are depicted in the figure. Error bars represent standard deviation (n=3).

C. : Same constructs than in B but the untranslated mRNA portions were removed (named “D”). Color/construction correspondences are depicted in the figure. Error bars represent standard deviation (n=3).

Supplementary Figure S2.



Supplementary Figure S2.

b-galactosidase expression. Total proteins of strains containing *wt-lacZ* (123 kDa) and *lacZ* with inserted stop codon at 75% (93 kDa), 50% (61 kDa) and 25% (30 kDa) of the CDS were separated using SDS-PAGE and revealed with coomassie blue staining. Arrows indicate the truncated proteins. The truncated b-galactosidases produced by *lacZ-Stop*$_{25\%}$ and *lacZ-Stop*$_{0.5\%}$ were not detectable due to a low intensity and a small size, respectively. M: protein marker with molecular sizes (kDA) indicated.

Supplementary Tables

Primers used for cloning		
Primer name	used for:	5'-3' sequence
1097-SD-F	SD replacement	TTATAAAATTAACCATGGATCCACTAGTAACGGCC
1099-SD-R	SD replacement	GTTAGCCCAAAAAACGGGTATGGAGAAACA
719-TAG-bGal5-Fw	Stop insertion for lacZ-$Stop_{0.5\%}$	ATGGATCCACTAGTATAGGCCGCCAGTGTGCTGG
720-TAG-bGal5-Rv	Stop insertion for lacZ-$Stop_{0.5\%}$	CCAGCACACTGGCGGCCTATACTAGTGGATCCAT
721-TAG-bGal263-Fw	Stop insertion for lacZ-$Stop_{25\%}$	GCGGCGAGTTGCGTGACTAGCTACGGGTAACAGTTTC
722-TAG-bGal263-Rv	Stop insertion for lacZ-$Stop_{25\%}$	GAAACTGTTACCCGTAGCTAGTCACGCAACTCGCCGC
723-TAG-bGal542-Fw	Stop insertion for lacZ-$Stop_{50\%}$	CTTTCGCTACCTGGATAGACGCGCCCGCTGATC
724-TAG-bGal542-Rv	Stop insertion for lacZ-$Stop_{50\%}$	GATCAGCGGGCGCGTCTATCCAGGTAGCGAAAG
725-TAG-bGal819-Fw	Stop insertion for lacZ-$Stop_{75\%}$	CATTGACCCTAACGCCTAGGTCGAACGCTGGAAGG
726-TAG-bGal819-Rv	Stop insertion for lacZ-$Stop_{75\%}$	CCTTCCAGCGTTCGACCTAGGCGTTAGGGTCAATG
912-tRNA-ortho5'	insertion $tRNA_o$	AATCCGCATGGCAGGGGTTCAAATCCCCTCCGCCGGAATTCGAAGCGGATGAGAGAAGATTTTCAGCC
913-tRNA-ortho3'	insertion $tRNA_o$	TAGAGTCCGCCGTTCTGCCCTGCTGAACTACCGCCGGCCAAATGCAGATCTTCGAACAAAACAGCCAAGCT
996-termtRNAortho	$tRNA_o$ deletion	AATTCGAAAAGCCTGCTCAACGAGCAGGC
998-termtRNAorthoII	$tRNA_o$ deletion	GAGCAGCTCAGGGTCGAATTTGCTTTCG
1102-CCA-tRNAortho	tRNA*	AATTCGAAGCGGATGAGAGAAGAT

Primers used for RT-qPCR		
Primers name	Comments	5'-3' sequence
361-LacZ / 362-LacZ	*lacZ* amplicon "a"	GTCGTGACTGGGAAAACCCTGG / AACTGTTGGGAAGGGCGATCG
741-LacZ / 742-LacZ	*LacZ* amplicon "b"	TCCGTGACGTCTCGTTGCTGC / TCACGCAACTCGCCGCACAT
743-LacZ / 744-LacZ	*LacZ* amplicon "c"	GGGTGAAACGCAGGTCGCCA / TCGGCGCTCCACAGTTTCGG
363-LacZ / 364-LacZ	*LacZ* amplicon "d"	AACAACTTTAACGCCGTGCGCT / CACCATGCCGTGGGTTTCAATA
745-LacZ / 746-LacZ	*LacZ* amplicon "e"	TCCCGCCCGGTGCAGTATGA / AGCCGGGAAGGGCTGGTCTT
747-LacZ / 748-LacZ	*LacZ* amplicon "f"	CGCCCGCTGATCCTTTGCGA / CAGTCCCAGACGAAGCCGCC
749-LacZ / 750-LacZ	*LacZ* amplicon "g"	CCCGCATCTGACCACCAGCG / CAGCGGCGTCAGCAGTTGTT
751-LacZ / 752-LacZ	*LacZ* amplicon "h"	CCAGGCCGAAGCAGCGTTGT / TTCCCCTGATGCTGCCACGC
365-LacZ / 366-LacZ	*LacZ* amplicon "i"	CAGCTGGCGCAGGTAGCAGAG / GGCAGATCCCAGCGGTCAAA
753-LacZ / 911-tRNAorthoRv	*lacZ-tRNA* amplicon "j"	TCGAAGCTTGGGCCCGAACAA / TCCGGCGGAGGGGATTTGAAC
520-ihfB-F / 521-ihfB-R	*ihfB* chromosomal normalisation	GCCAAGACGGTTGAAGATGC / CAAAGAGAAACTGCCGAAACC
1119-AmpR-F / 1120-AmpR-R	*bla* plasmidic normalisation	ATTCAACATTTCCGTGTCGCCC / TGTAACCCACTCGTGCACCCAA
671-RnaseE-F / 672-RnaseE-R	quantification *rne* transcript	CACCAGAACAGCAAGAAGAG / GACGTACAATTGGATAGCGG
673-RnaseR-F / 674-RnaseR-R	quantification *rnr* transcript	GAAGCGATTACCTCTTTCCG / AGATATTTAATGGCGCGGTG
675-PNPase-F / 676-PNPase-R	quantification *pnp* transcript	AAAGAGATCATGCAGGTTGC / CCGCCTTTACCGATAACATC
683-pcnB-F / 684-pcnB-R	quantification *pcnB* transcript	TCCCGCAAAGATATCAGTGA / GCAACTTCGATAATCTCCGG
685-Eno-F / 686-Eno-R	quantification *eno* transcript	GTCGTGAAATCATCGACTCC / CAGCTTTGGTTACGCCTTTA
715-Hfq-F / 716-Hfq-R	quantification *hfq* transcript	GGGCAAATCGAGTCTTTTGA / TGGTAGTTACTGCTGGTACC
681-RnaseG-F / 682-RnaseG-R	quantification *rng* transcript	GTTTATCATCCGTACCGCAG / ATGTACTCCGAGGTGAACTC