RPSA AND RIBOSOMAL PROTEIN S1: INVESTIGATING A NON-CANONICAL TRANSLATION INITIATION ELEMENT

JUSTIN ROBERT JAMES VIGAR
Bachelor of Science, University of Lethbridge, 2012

A thesis submitted
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

BIOCHEMISTRY

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

RPSA AND RIBOSOMAL PROTEIN S1: INVESTIGATING A NON-CANONICAL TRANSLATION INITIATION ELEMENT

JUSTIN ROBERT JAMES VIGAR

Date of Defence: July 26, 2019

Dr. Hans-Joachim Wieden Thesis Supervisor	Professor	Ph.D.
Dr. Anthony Russell Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Igor Kovalchuk Thesis Examination Committee Member	Professor	Ph.D.
Dr. Kevin Folta External Examiner University of Florida Gainesville, FL	Professor	Ph.D.
Dr. Trushar Patel Chair, Thesis Examination Committee	Assistant Professor	Ph.D.

This dissertation is dedicated to my parents;

I guess the Hooked on Phonics worked.

ABSTRACT

Translation initiation rates are fine-tuned by altering interactions between the ribosome and the translation initiation region of an mRNA. In bacteria varying levels of RNA structure attenuate these interactions by masking the ribosome-binding site from the small ribosomal subunit. Recent studies have described diverse strategies for recruiting mRNA to the ribosome, and highlighted the contributions of ribosomal protein S1. Here, I provide evidence that the non-canonical initiation mechanism that governs translation of the *rpsA* mRNA, encoding ribosomal protein S1, contains a three-dimensional architecture that is required for efficient translation. Furthermore, S1 plays an essential role during the initiation phase of translation by recruiting mRNA to the ribosome—unfolding structured mRNAs, and allowing for correct start codon positioning on the ribosome. Combining crosslinking immunoprecipitation and high-throughput sequencing approaches revealed the extent of S1's involvement in mRNA recruitment, while also highlighting a broader role as a regulator of many RNA classes.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Graeme Glaister, Jalyce Heller, Dr. Victoria McParland, Dr. Lars Barquist, Dr. Robert Rambo, Nikul Khunti, Fan Mo and Dr. Katsuaki Inoue for assistance with data collection and experiments presented in this thesis. I would also like to thank Dr. Jörg Vogel for initiating the collaboration between ARRTI and IMIB, and for hosting me in his lab; my time there translated into a truly career altering experience. Each time I enter a lab I'm reminded of lessons learned during my time spent under the supervision of Dr. Ute Kothe and Dr. Erik Holmqvist. Both of these scientists were extremely patient, kind, humble, and inspiring; now they serve as role models for how I aim to work in a lab.

Wieden lab members throughout the years have made this experience so much better by providing a wonderful and stimulating work environment. I am now grateful for all of the challenging and stressful lab meetings, which in-turn forced me to become a better scientist—and contributed to the success of my project. Additionally, I'd like to acknowledge my officemate and I^3 cofounder, Luc Roberts, for the many (sometimes) insightful conversations. HJ: the fencepost on which our gate swings. I'm grateful for all the work you do to provide us with the best scientific experience possible. The grants, the edits, your patience and trust—we all appreciate this *so* much.

I would like to thank Alberta Innovates, Graduate Studies at the University of Lethbridge, the University of Lethbridge, and the Alberta RNA Research and Training Institute for funding this work.

My family has always provided me with more love and support than I could ever hope for. For this I am grateful. I'm especially thankful to Braden for introducing Lilly

into our family. Tara, I have countless memories of you making some very rough and stressful days *so* much better—thank you.

Mike, Jordan, Andy, James, Mike, Scott, Brendan, and Carson—it's difficult to articulate exactly how much I appreciate your friendship. I'm incredibly fortunate to have you all in my life. Thank you for taking me along on all of you adventures, and for following me on mine.

PREFACE

The second chapter of this thesis contains a review article authored by myself and Dr. Hans-Joachim Wieden titled *Engineering bacterial translation initiation—Do we have all the tools we need?* (Vigar & Wieden, 2017). In this article we review current strategies used to engineer the initiation rate of ribosomal protein synthesis in bacteria, while highlighting current challenges in the field. The article was published in an Elsevier journal, and as an author I retain the right to include it in my thesis without permission.

Chapter three includes a manuscript written for submission to the journal Molecular Cell, describing an international effort to characterize a unique strategy cells use to initiate ribosomal protein synthesis. The project and experiments were planned by myself and Dr. Hans-Joachim Wieden, with input from Dr. Jörg Vogel and Dr. Erik Holmqvist. I performed all plasmid, strain, and oligo design and construction. The sort-seq procedure was performed by myself under the guidance of Dr. Erik Holmqvist in Dr. Jörg Vogel's laboratory at the Institute for Molecular Infection Biology (IMIB) in Würzburg, Germany. Dr. Lars Barquist at the Helmholtz Centre for Infection Research in Würzburg, Germany developed the algorithms used to process the next-generation sequencing data. I carried out the nitrocellulose filter binding, and RBS Calculator analyses. Jalyce Heller at the Alberta RNA Research and Training Institute (ARRTI) in Lethbridge, Canada performed all of the *in vitro* transcription and translation assays. The SHAPE probing experiments were done in collaboration with Dr. Janusz Bujnicki's laboratory at the International Institute of Molecular and Cell Biology (IIMCB) in Warsaw, Poland. I purified *rpsA* TIR variant RNAs, Dr. Astha performed the probing experiments, while both Dr. Astha and myself analyzed the data. I completed the SAXS experiments at the Diamond Synchrotron in Oxford, United Kingdom with the help of the

staff beam line scientists. Half of the RNAs subjected to SAXS were purified by Jalyce Heller, while I purified the remaining half. Jalyce Heller, and Dr. Nithin Chandran (IIMCB) led the analysis of the SAXS data and construction of *ab initio* models, with input from Dr. Trushar Patel (ARRTI), Dr. Janusz Bujnicki, Dr. Hans-Joachim Wieden and myself.

Chapter four contains a review article authored by Jalyce Heller and I. This review examines how the synthesis of S1 is regulated at the transcriptional, post-transcriptional, and translational levels. Additionally, it outlines the multiple functions of S1. Both authors contributed equally to the manuscript. My main contributions towards this chapter were writing the section describing S1's regulation, and editing, commenting and researching all other sections.

Chapter five contains a manuscript outlining our work in charting the landscape of S1's interaction with cellular RNAs. I performed all wet lab experiments in chapter five in collaboration with Dr. Erik Holmqvist while in Dr. Jörg Vogel's laboratory. Dr. Lars Barquist performed an initial bioinformatic analysis of the data, while I—with advice from Dr. Athan Zovoilis—completed the data analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

30S	30 Svedberg prokaryotic ribosomal subunit
50S	50 Svedberg prokaryotic ribosomal subunit
70S	70 Svedberg prokaryotic ribosome
ATP	Adenosine triphosphate
AU	Arbitrary units
Amp	Ampicillin
bp	Base-pair
Cam	Chloramphenicol
CDS	Coding sequence
CTP	Cytidine triphosphate
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
eCFP	Engineered cyan fluorescent protein
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence activated cell sorting
GTP	Guanosine triphosphate
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRES	Internal ribosome entry site
K_d	Equilibrium dissociation constant
LB	Luria broth
mRNA	Messenger RNA
mRFP	Monomeric red fluorescent protein
nt	Nucleotide
NGS	Next generation sequencing
NSD	Normalized spatial discrepancy
NMIA	N-methyl isatoic anhydride
NMR	Nuclear magnetic resonance
OD	Optical density
UTP	Uridine triphosphate
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
PZA	Pyrazamide
POA	Pyrazinoic acid
RNA	Ribonucleic acid
RNAP	RNA polymerase
R-protein	Ribosomal protein
rRNA	Ribosomal RNA
RBS	Ribosome binding site
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
S1	Ribosomal protein S1
SAXS	Small angle X-ray scattering

SEC	Size exclusion chromatography
SD	Shine-Dalgarno
SHAPE	Selective 2' acylation analyzed by primer extension
SDS	Sodium dodecyl sulfate
sRNA	small non-coding ribonucleic acid
TE	Translation efficiency
TIR	Translation initiation region
tRNA	Transfer RNA
Tris	Tris(hydroxymethyl)aminomethane
TX-TL	Transcription-translation
UTR	Untranslated region

Chapter 1

Introduction

1.1 Introduction

Modulating the frequency of translation initiation is a universal strategy for controlling gene expression. Translation initiation in bacteria is controlled by fine-tuning interactions between the translation initiation region (TIR) of an mRNA and the ribosome, typically by varying the strength of the Shine-Dalgarno (SD)-anti-SD interaction and levels of inhibitory RNA structure within the TIR. Natural translation initiation control elements have been repurposed to tune synthetic genetic circuits. However, limitations exist regarding their predictability, portability, and dynamic range. New studies have revealed alternate strategies for mRNA-ribosome recruitment; exploring their mechanisms will afford new approaches for controlling translation initiation.

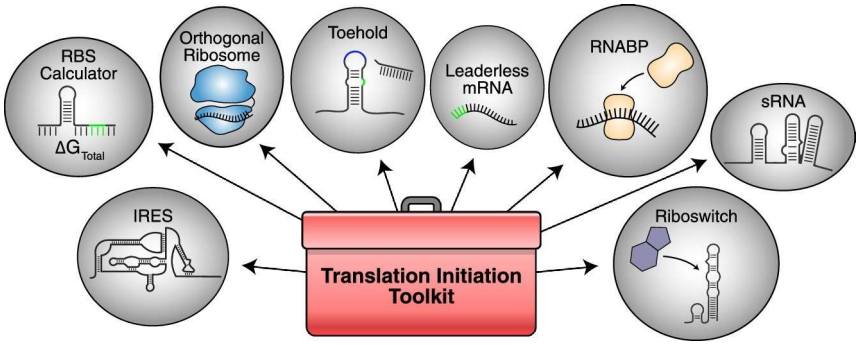


Figure 1.1 Current strategies for engineering translation initiation in bacteria. Natural regulatory elements have been studied and repurposed, expanding the bioengineers "tool box" for controlling translation in bacteria. Elements such as riboswitches, toehold switches, sRNAs binding sites, or RNA binding protein (RNABPs) binding sites can be imbedded in the TIR of mRNAs. Non-canonical initiation elements such as IRESs, leaderless mRNAs, or orthogonal ribosomes can also be used to control cellular behaviour. An accumulation of knowledge about translation has led to the development of predictive models, such as the RBS Calculator, which enable forward and reverse engineering of translation initiation rates *in silico*. Figure reprinted from Vigar & Wieden, 2017.

In this thesis I describe the characterization of a non-canonical translation initiation element: the TIR of the *E. coli rpsA* gene. Although lacking a canonical SD, the

rpsA TIR is one of the most efficient drivers of translation initiation in *E. coli*. Additionally, the typically inhibitory secondary structure within the TIR is somehow necessary for its function. The goal of this work was to uncover the signals within the *rpsA* TIR that contribute to its ability to efficiently initiate translation. To this end a quantitative, dual-reporter *in vivo* screening tool was built, and the sequence-function relationships of the *rpsA* TIR were mapped using large-scale mutagenesis and next-generation sequencing. From this library *rpsA* variants were selected for further structural analyses, allowing their activity levels to be linked to structural elements within the TIR. The activities of the *rpsA* TIRs were tested in a minimal transcription/translation system to investigate function in the absence of additional cellular factors, and to determine the requirement for ribosomal protein S1 (S1). Our integrated *in vivo*, *in vitro*, and *in silico* approach provides new insight into the structure-based mechanism of this non-canonical initiation element and provides a new class of synthetic regulatory RNA devices for engineering biology.

Additionally, RNA-binding proteins modulate translation by interacting with signals within the mRNA. S1 is a global mRNA regulator with a critical role in *E. coli* translation. It influences the bulk of cellular mRNAs, however its mechanism of action is poorly understood. My goal was to uncover S1's RNA binding partners *in vivo*. To this end I performed UV crosslinking combined with high–throughput RNA sequencing to determine the interactome of S1-mRNA binding. Previous reports regarding S1's mRNA binding partners were confirmed, and additional details with respect to its mRNA binding properties are provided. S1 is known to target mRNA, however for the first time S1's role as a small noncoding RNA regulator is described.

Chapter 2

Engineering Bacterial Translation Initiation—Do We Have all the Tools we Need?

Vigar, J. and H. J. Wieden (**2017**). "Engineering bacterial translation initiation - Do we have all the tools we need?"

Biochim Biophys Acta.

2.1 Introduction

The flow of biological information from gene to gene product relies on the coordinated choreography of thousands of biomolecules. These biomolecules transcribe the hereditary genetic information into messenger RNA (mRNA) intermediates that are translated into functional proteins. Protein biosynthesis, or translation, is catalyzed by the ribosome (Stephenson, Thimann, & Zamecnik, 1956) and can be divided into four phases: initiation, elongation, termination, and ribosome recycling. During the rate-determining initiation phase, the small ribosomal subunit (30S) is recruited and correctly positioned on an mRNA translation initiation region (TIR) with the assistance of three initiation factors (IF1, IF2 and IF3) and initiator-tRNA to form a 30S pre-initiation complex (Lockwood, Chakraborty, & Maitra, 1971; Subramanian & Davis, 1970; Wintermeyer & Gualerzi, 1983). The positioning of the aminoacylated initiator-tRNA on the start codon ensures its location in the P-site of the 30S (Shine & Dalgarno, 1974). The large ribosomal subunit (50S) is then recruited to the 30S complex, IF2 hydrolyzes GTP, and the initiation factors are released, resulting in a 70S initiation complex capable of elongation (Hartz, McPheeters, & Gold, 1989; Rodnina et al., 2000). The tRNAs and mRNA transcript are translocated through the ribosome resulting in the codon dependent synthesis (or elongation) of a new polypeptide chain (Nishizuka & Lipmann, 1966). Termination occurs when the ribosome encounters a stop codon in the mRNA and the newly formed polypeptide is released from the ribosome. The ribosome is then recycled, the mRNA is released, and the subunits are dissociated for successive rounds of translation (Hirashima & Kaji, 1970, 1972). Our increased understanding of this process has revealed that diverse and complex strategies of mRNA recruitment to the ribosome for translation have evolved. Translation initiation has consequently become the most highly regulated step of

protein biosynthesis and, in bacteria, relies on characteristics of both the 30S ribosomal subunit and the mRNA (Laursen, Sorensen, Mortensen, & Sperling-Petersen, 2005).

Cells control the rate of ribosome-dependent protein synthesis through several mechanisms during translation initiation. Typically, these involve modulating factors that influence the correct positioning of the mRNA on the 30S initiation complex. Modulating a TIR's capacity to form a complex with the ribosome has been a useful strategy that cells have evolved to respond quickly to changes in their environment, including changes in cellular Mg^{2+} concentration (Cromie, Shi, Latifi, & Groisman, 2006), pH (Nechooshtan, Elgrably-Weiss, Sheaffer, Westhof, & Altuvia, 2009), concentrations of carbon sources and key metabolites (Wang, Lee, Morales, Lim, & Breaker, 2008), and temperature changes (Narberhaus, Waldminghaus, & Chowdhury, 2006). Autogenous negative (Boni, Artamonova, & Dreyfus, 2000) and positive feedback (Mandal et al., 2004) mechanisms have evolved to precisely turn genes on or off as required. Furthermore, TIR accessibility to the 30S can be altered using non-coding small RNAs (sRNAs) (Urban & Vogel, 2007; Waldminghaus, Kortmann, Gesing, & Narberhaus, 2008).

The manipulation and fine-tuning of gene expression is critical for a large number of synthetic biology applications that aim to alter a wide range of cellular behaviours (Mutalik, Guimaraes, Cambray, Lam, et al., 2013). Drawing inspiration from nature, bioengineers have repurposed existing natural translation initiation control mechanisms into useful tools for exerting control over cells. Targeting this critical step in gene expression is appealing, as it allows the tuning of protein biosynthesis over many orders of magnitude by engineering short RNA sequences (George et al., 2015; Mutalik, Guimaraes, Cambray, Lam, et al., 2013; Reeve, Hargest, Gilbert, & Ellis, 2014). A comprehensive toolkit for engineering initiation would enable the user to predictably

control ribosome load on an mRNA. Modulating translation initiation can allow the user to turn off and on translation of a specific mRNA in response to an external signal. Additionally, homeostasis of a gene product can be maintained over time through artificial feedback networks. TIRs that maintain predicted activity when incorporated into different mRNAs in different organisms, under a wide range of different operating conditions will be of great utility. Portable translation regulatory devices will allow bioengineers to work in model organisms before transplanting synthetic genetic circuits into an organism of choice, which may offer challenges such as being in a higher risk group or having a long-generation time. This will reduce the number of design, build, test cycles of gene circuit optimization, as well as ease time-consuming prototyping steps.

Well-characterized gene control elements with the above characteristics are vital when creating engineered microorganisms able to produce new materials, process chemicals, produce energy and food, and help maintain or enhance human health and our environment. Unfortunately, the realization of many of these applications has been hindered by the lack of available tools that allow for effective and scalable biological engineering (Mutalik, Guimaraes, Cambray, Mai, et al., 2013). In this review strategies that have evolved in nature to modulate translation initiation are described, and how these strategies have inspired bioengineers to generate tools to fine-tune and manipulate gene expression via translation initiation are discussed. Additionally, the challenges in the field and future goals in the field are outlined. While most of the topics discussed in this review focus on prokaryotic translation initiation, some concepts are relevant when engineering across domains of life and in cell-free systems.

2.2 The 16S RNA and the Translation Initiation Region

When Shine and Dalgarno identified regions of complementarity between the TIRs of coliphage mRNAs and the 3' end of the 16S rRNA in *E. coli*, an attractive hypothesis regarding mRNA recruitment to ribosomes was realized (Shine & Dalgarno, 1974). The interaction between the mRNA and the 30S was proposed to be mediated solely by purine-rich regions on the 5' end of mRNAs, called the Shine-Dalgarno sequence (SD), and the 3' end of the 16S rRNA, called the anti-Shine-Dalgarno sequence (aSD). Prokaryotic organisms achieve a wide range of translation initiation levels by varying the degree of complementarity of this interaction, coupling the initiation strength to the thermodynamics of the SD:aSD interaction (Schurr, Nadir, & Margalit, 1993).

The realization that modulating the free energy of binding between the SD and the aSD could achieve differential initiation frequencies provided a conceptually straightforward method to rationally design gene expression. Using reverse engineering approaches, modular ribosome binding site (RBS) part libraries have been created, able to modulate translation initiation frequencies of synthetic mRNAs (Carrier & Keasling, 1999). This process requires labour-intensive screening and, in complex systems, requires iterative design. Additionally, RBSs are particularly prone to forming secondary structures with adjacent RNA elements in the TIR, causing some to have 500-fold higher or lower translation initiation levels depending on the context in which they are placed (Salis, Mirsky, & Voigt, 2009). It became apparent that other characteristics of the TIR beyond the SD were responsible for modulating translation initiation.

By systematically changing the distance between the SD and the start codon, the optimal spacing requirements were described (Chen, Bjerknes, Kumar, & Jay, 1994; Ringquist et al., 1992). However, the degree of secondary structure surrounding the SD plays a key role in governing access to the TIR, thereby modulating translation initiation

efficiency (M. H. de Smit & van Duin, 1990; Schauder & McCarthy, 1989; Studer & Joseph, 2006).

2.2.1 The RBS Calculator

A broader understanding of the rules governing translation initiation has allowed bioengineers to develop software compiling thermodynamic properties of translation initiation into predictive models (Chen et al., 1994; M. H. de Smit & van Duin, 1990; Maarten H. de Smit & van Duin, 2003; Duval et al., 2013; Studer & Joseph, 2006; Vellanoweth & Rabinowitz, 1992). These tools have been successfully used to optimize gene expression at the translational level, reducing the number of design, build, test cycles required to construct a functional gene network, reviewed by Reeve et al. (2014). The Ribosome Binding Site (RBS) Calculator is such a predictive tool for designing synthetic RBSs; using an equilibrium statistical model, the calculator determines the strength of a hypothetic RBS (Salis et al., 2009).

To calculate the total free energy of binding between the 30S and the mRNA, energies of several intra- and intermolecular interactions between the components involved in initiation are summed using equation (1) (Figure 2.1). These experimentally established contributors toward initiation frequency (Table 2.1) include energy released upon interaction with the SD and the nine 3'-most nt of the 16S rRNA (aSD), and the

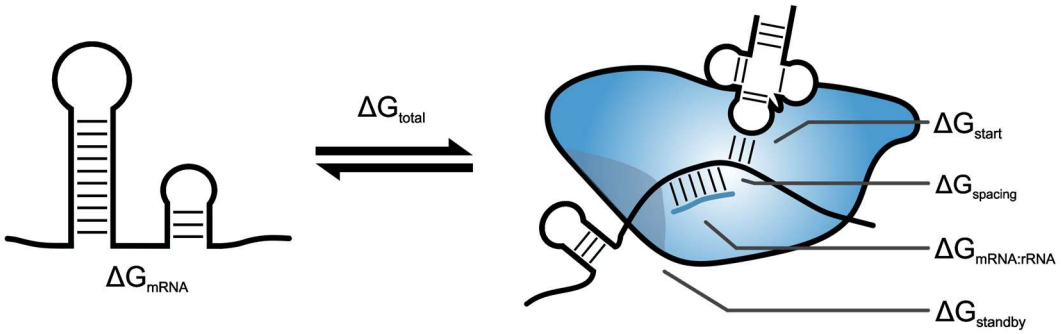


Figure 2.1. Schematic representation of the free energy contributions of mRNA interactions during 30S initiation complex formation, the ribosomal platform is shaded in darker. The RBS Calculator calculates the free energy difference between a free folded mRNA and an mRNA bound to 30S (blue) (Espah Borujeni, Mishler, Wang, Huso, & Salis, 2015). The unfolding of inhibitory mRNA structure in the TIR is represented as ΔG_{mRNA}. The hybridization of the SD to the aSD ($\Delta G_{mRNA:rRNA}$), tRNA hybridization to the start codon (ΔG_{start}), unfolding of standby sites in the mRNA ($\Delta G_{standby}$), and the deviation from optimal spacing between the RBS and the start codon ($\Delta G_{spacing}$) are taken into consideration. The resulting energy change (ΔG_{total}) is then linked to the translation initiation strength using equation (2).

initiator-tRNA hybridization to the start codon. A penalty is included for any deviation from the optimal spacing of 5 nt between the RBS and the start codon that forces the ribosome reposition itself is also included (Chen et al., 1994). Energy needed to unfold inhibitory RNA structures throughout the TIR are incorporated; generally, as structure increases, the initiation frequency decreases (Studer & Joseph, 2006).

$$(1)\ \Delta G_{total} = \Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby} - \Delta G_{mRNA}$$

$$(2)\ r \propto e^{-\beta \Delta G_{total}}$$

Table 2.1. Free energy terms included in various versions of the RBS Calculator: V1.0 /1.1 (●), V2.0 (■), V1.0 Operon Calculator (♦). The energetics of forming and breaking these non-covalent interactions that govern initiation can be fine-tuned and balanced to achieve desired initiation strengths.

Term	Description	Version		
$\Delta G_{mRNA:rRNA}$	SD binding to the aSD	●	■	♦
ΔG_{start}	Initiator-tRNA binding to the start codon	●	■	♦
$\Delta G_{spacing}$	Spacing between the SD and the start codon	●	■	♦
ΔG_{mRNA}	Unfolding inhibitory structure throughout the TIR	●	■	♦
$\Delta G_{standby}$	Unfolding standby site secondary structure	●	■	♦
$\Delta G_{unfolding}$	Unfolding structure surrounding the start codon		■	♦

$\Delta G_{distortion}$	Binding available RNA on standby site		■	♦
$\Delta G_{sliding}$	30S sliding from the standby site to the SD		■	♦
$\Delta G_{coupling}$	Unfolding structures in the preceding CDS			♦
$\Delta G_{noncoupling}$	Unfolding intergenic RNA structure			♦

The translation initiation rate is calculated by an exponential function (2), where *r* is the initiation rate and β is the Boltzmann constant, relating thermodynamic free energies to temperature. This model has proven to be accurate at predicting translation initiation strength within a factor of 2.3 over a range of 100,000-fold. However, many natural mRNAs contain elements upstream of their SD that are not accounted for in this model, causing deviation from the predicted initiation strength.

To address some of these limitations, an improved version incorporated new experimental data and thermodynamic modelling to determine how the 30S accommodates or partially unfolds upstream mRNA secondary structures, selects the correct binding site, and slides across the mRNA (Summarized in Table 2.1) (Espah Borujeni, Channarasappa, & Salis, 2014). This addition aimed to improve accuracy when predicting translation initiation rates of mRNAs containing structured elements upstream of the SD. The capacity of the 30S ribosomal platform (Figure 2.1) to interact with single stranded standby sites on the mRNA was characterized to better predict $\Delta G_{standby}$, described in equation (3). Highly accessible standby sites with long 30S binding regions increase ribosome binding and translation initiation. However, when standby sites have limited accessibility, there is a relationship between unfolding an mRNA structure ($\Delta G_{unfolding}$) and increasing its accessibility to the 30S ($\Delta G_{distortion}$). As a hairpin is unfolded, there are more single stranded regions to interact with the 30S and translation initiation is increased. However, unfolding a hairpin can take a substantial amount of

energy, decreasing translation initiation. Additionally, the 30S will first bind to the standby site that has the lowest binding free energy penalty and slide across the mRNA, starting at the standby site and moving towards the SD, while reorienting the mRNA ($\Delta G_{sliding}$) until a stable pre-initiation complex has formed.

$$(3)\ \Delta G_{standby} = \Delta G_{distorsion} + \Delta G_{unfolding} + \Delta G_{sliding}$$

When equation (3) is used as the $\Delta G_{standby}$ term in equation (2) the resulting model has improved accuracy at predicting translation initiation rates. The *in vivo* translation initiation rates of 136 long structured *E. coli* TIRs were accurately predicted using the RBS Calculator (R^2=0.89) (Espah Borujeni et al., 2014). The RBS Calculator has revolutionized the rational design of translation initiation regions embedded in the 5' UTRs of mRNAs.

Bacteria commonly incorporate several coding sequences, separated by short intergenic TIRs, into a single polycistronic mRNA. These polycistronic mRNAs are used as a means to co-regulate proteins and RNAs that need to be governed by similar regulatory mechanisms (Button & Galan, 2011; Farasat et al., 2014), a concept that proved challenging for the rational design of such constructs using the RBS Calculator. An addition to the calculator included the effects of gene order in an operon on the translation initiation rate, generating the V1.0 Operon Calculator (Tian & Salis, 2015). The translation initiation rate of the second coding sequence in an operon can be more accurately calculated by taking into consideration two additional factors: the upstream coding sequences translation rate and the probability that the ribosome will re-assemble. The latter depends largely on the length and level of structure of the intergenic region between the two coding sequences. The ΔG_{mRNA} term from equation (1) is expanded to

include the unfolding free energies of all inhibitory RNA structures overlapping the first coding sequence (CDS) and all structures that impede other areas from interacting with the ribosome, such as an SD or a standby site. Structured RNA surrounding intergenic regions must be unfolded by the ribosome during the final stages of elongation of the preceding coding sequence, allowing a second ribosome to bind to the intergenic region without having to unfold these RNA structures. Incorporating these factors into the model allowed the design of operons containing 2 or 3 reporter genes with fine-tuned expression ratios by introducing attenuating RNA structures with specific folding energies into intergenic regions (Tian & Salis, 2015).

By manipulating the free energy terms governing mRNA:30S binding, one can predictably modulate the expression level of a wide range of synthetic mRNAs. While the Salis RBS Calculators have been valuable for designing synthetic mRNAs to meet certain specifications and for estimating translation initiation frequencies of existing mRNAs, they don't always generate predictable results. Although our ability to accurately predict translation initiation rates is high, there are still critical details missing in our mechanistic understanding of this process.

When engineering simple genetic circuits, it may be feasible to test several predicted RBSs *in vivo* and choose one that works as predicted. However, as genetic circuits become increasingly complex, it will be critical to have a complete understanding of the molecular mechanisms governing translation initiation to ensure more reliable predictions prior to implementation. Even if we can accurately predict the translation initiation rates using these tools designed, TIRs are static regulators. Once they are implemented, no change is possible; it is a form of non-responsive regulation. However, cells rely on the ability to respond rapidly to external stimuli; they frequently take

advantage of translation initiation as a means to achieve this dynamic regulation (Isaacs et al., 2004).

2.3 Modulating 5' UTR Structure

Prokaryotic cells have incorporated additional signals within mRNAs to achieve tight and rapid control of translation in response to external stimuli. Both cis- and trans-acting riboregulators exploit the structural dynamics of RNA to alter TIR conformation, and therefore its ability to form a competent 30S initiation complex. The importance of these types of regulation strategies is becoming increasingly appreciated as advances in bioinformatic and biochemical techniques uncover numerous distinct classes (Omotajo, Tate, Cho, & Choudhary, 2015; Storz, Vogel, & Wassarman, 2011).

2.3.1 RNA Switches

RNA switches (riboswitches) enable cells to couple RNA-associated cellular processes to internal or external inputs (Figure 2.2A and B). Riboswitches are regulatory RNAs imbedded within the 5' UTRs of some mRNAs (Breaker, 2012). They consist of two domains: a sensor that binds a ligand with high specificity and affinity, and an actuator that regulates gene expression (A. L. Chang, McKeague, Liang, & Smolke, 2014). Each domain can fluctuate between several defined conformations and ligand concentration drives the equilibrium of the sensor toward a specific folded state (Dixon et al., 2010; Grundy & Henkin, 2006; Lynch, Desai, Sajja, & Gallivan, 2007). The state of the sensor domain directs conformational changes in the actuator domain, modulating gene expression (Dixon et al., 2010; Grundy & Henkin, 2006; Lynch et al., 2007).

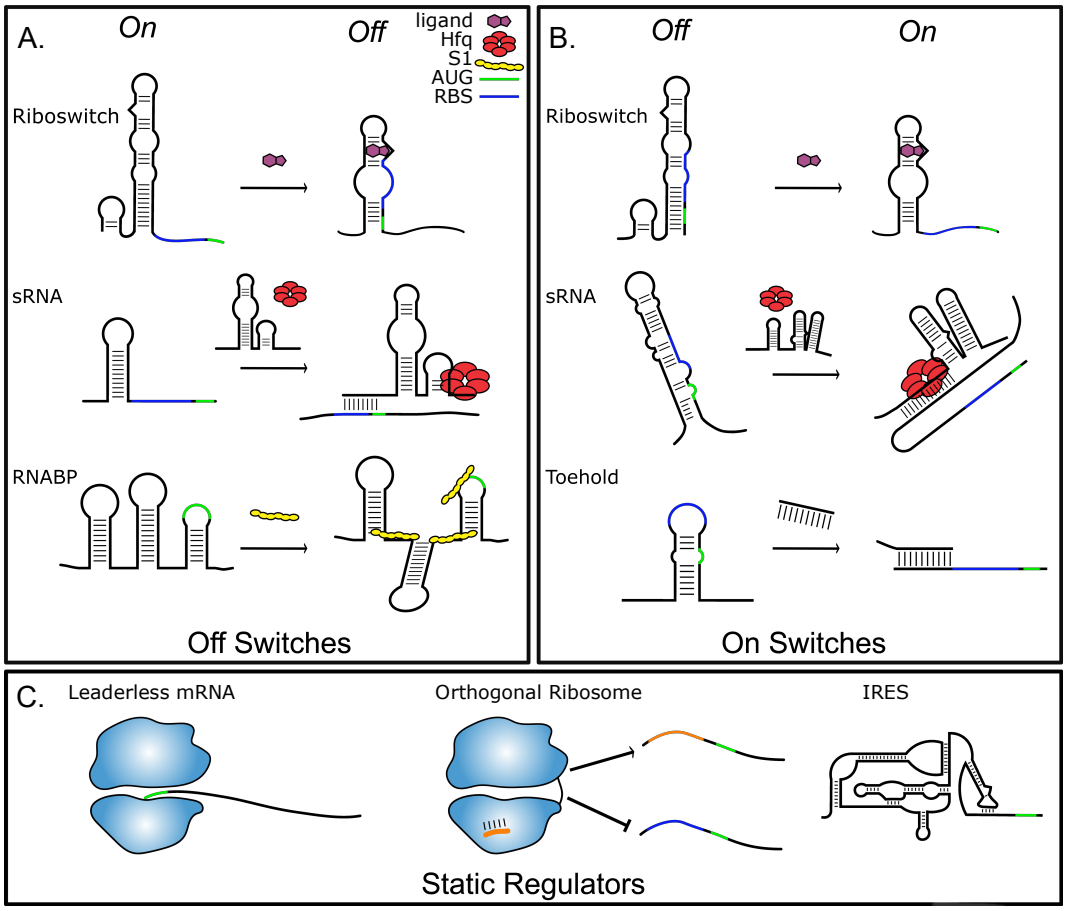


Figure 2.2. Mechanisms for controlling translation initiation in bacteria. RNA sensors and structural elements can be incorporated into genetic circuits enabling a means to up- or down-regulate translation initiation in response to a signal. (A) Riboswitches can be engineered to respond to small molecules, such as, theophylline to down-regulate initiation (S. Topp & Gallivan, 2008); upon incorporation of sRNAs, such as MicC, Hfq mediated annealing of the sRNA and TIR results in down-regulation of translation (Na et al., 2013); the presence of specific RNA binding proteins (RNABPs), such as S1, can result in remodelling of mRNA causing inhibition of translation (Boni, Artamonova, Tzareva, & Dreyfus, 2001). (B) changing the location of an RBS and start codon in some riboswitches, such as the theophylline riboswitch, can result in increased translation initiation in the presence of a ligand (S. Topp & Gallivan, 2008); the Hfq dependent sRNA, DsrA, can be engineered to bind to a target mRNA (Lahiry, Stimple, Wood, & Lease, 2017) while Toehold Switches consisting of a transducer element can sense trigger RNAs independent of Hfq, both resulting in increased accessibility to the RBS (Auslander & Fussenegger, 2014). (C) leaderless mRNAs interact directly with 70S ribosomes to initiation translation (Moll, Grill, Gualerzi, & Blasi, 2002); engineering the 16S rRNA of the 30S ribosomal subunit has been used to generate orthogonal translation systems that specifically translate synthetic orthogonal mRNAs (Orelle et al., 2015); non-canonical ribosome recruitment signals based on RNA structure, as found some viral IRESs, can be used to control translation initiation (Colussi et al., 2015).

Considerable structural and functional diversity exists within the at least 20 classes of riboswitches that have been described thus far (Breaker, 2012; Serganov & Nudler, 2013). They are able to modulate RNA stability, transcription (typically in Gram-

positive bacteria), and translation (typically in Gram-negative bacteria) (Waters & Storz, 2009). In Gram-negative bacteria, the expression platform is most commonly incorporated into the TIR, which is made more or less accessible to the 30S upon ligand binding (Grundy & Henkin, 2004). Microorganisms have used this approach to sense and respond to cellular metabolites such as glucosamine 6-phosphate, lysine, and glycine; coenzymes such as vitamin B-12, SAM (S-adenosyl methionine), thiamine pyrophosphate, flavin mononucleotide; metal ions such as Mg^{2+}; and temperature (Croft, Moulin, Webb, & Smith, 2007; Cromie et al., 2006; Gelfand, Mironov, Jomantas, Kozlov, & Perumov, 1999; Mandal et al., 2004; Narberhaus et al., 2006; Vitreschak, Rodionov, Mironov, & Gelfand, 2003; Wang et al., 2008). Some riboswitches such as the SAM riboswitch have dual functionality and can also regulate in *trans* as a small non-coding RNA (sRNA) (Heppell et al., 2011; Wang et al., 2008).

Bioengineers have taken advantage of the unique features that these riboregulators possess, including their modularity, portability, and rapid mode of action. While bioinformatics techniques have been useful in identifying natural riboswitches, there are limits to the number of ligands detected by these riboswitches. Directed evolution techniques, such as systematic evolution of ligands by exponential enrichment (SELEX), have led to the development of new classes of riboswitches (Wittmann & Suess, 2012). SELEX was first used to generate RNA aptamers able to bind metabolic cofactor mimics and T7 RNA polymerase (Ellington & Szostak, 1990; Tuerk & Gold, 1990). Typically, 10^{14}–10^{15} unique RNA molecules are incubated with an immobilized target of interest *in vitro*. Non-binding species are washed away, enriching species that have a high-affinity interaction with the ligand. These RNAs can then be isolated, amplified, and subjected to further rounds of selection. Several riboswitches, such as the theophylline (Jenison, Gill,

Pardi, & Polisky, 1994) and tetracycline (Berens, Thain, & Schroeder, 2001) riboswitches, have been developed using this technique (Figure 2.2 A and B). Theophylline riboswitches have been engineered to control chemotaxis in *E. coli* (Shana Topp & Gallivan, 2007). They are highly modular, and can be used in different species (S. Topp et al., 2010), such as cyanobacteria (Nakahira, Ogawa, Asano, Oyama, & Tozawa, 2013). They have also been engineered to respond to thiamine pyrophosphate (Muranaka, Sharma, Nomura, & Yokobayashi, 2009), ammeline, azacytosine (Dixon et al., 2010), and temperature changes (Neupert & Bock, 2009). Many useful aptamers have been either discovered or developed using SELEX. However, most of these aptamers cannot modulate an expression platform. There are also limitations in creating aptamers *in vivo* that respond to novel ligands due to solubility, reactivity, or toxicity issues (Espah Borujeni et al., 2015).

To overcome these limitations, a statistical thermodynamic model to both predict translation-regulating riboswitches and generate new riboswitches has been proposed (Espah Borujeni et al., 2015). This *in silico* tool is built largely on principles of the RBS Calculator. The energy required to transition between an unfolded mRNA to a ligand-bound mRNA:ribosome complex is calculated and used to determine initiation strength. The tool requires the user to input the riboswitch mRNA sequence, the three-dimensional structure of the aptamer in its ligand bound state, the protein coding sequence, and the concentrations of both mRNA and ligand. Based on this information, it determines the $\Delta G_{mRNA:ribsome}$, ΔG_{mRNA}, $\Delta G_{standby}$, and ΔG_{start} in both ligand-bound and unbound states to generate a translation initiation rate in the ON and OFF states. To identify a compatible expression platform, the algorithm performs 10^{36} rounds of *in silico* mutations on a randomized effector domain downstream of the aptamer, followed by prediction,

selection and recombination. This approach has successfully yielded 62 new riboswitches, based on six different aptamers, that are able to modulate translation *in vivo* over 380-fold (Espah Borujeni et al., 2015).

To investigate the theoretical limits of synthetic riboswitch-based gene expression, the Salis lab used their thermodynamic model to construct a 'perfect' riboswitch containing hypothetical aptamers (Espah Borujeni et al., 2015). Although the maximal achievable activation is over 1,000-fold when the ligand is abundant, it is limited to 3 to 20-fold under nanomolar ligand concentrations when using an aptamer with either a high or low ligand affinity, respectively. For riboswitches to be a widely used tool, a wide range of ligand binding aptamers must be available to use. The development of this model has greatly expanded the ability to rationally design riboswitches.

2.3.2 sRNAs

sRNAs are typically *trans*-acting RNAs, 50–200 nt in length, that play critical roles in gene expression (A. Zhang et al., 2003). Their dominant mode of action is accomplished by imperfect base pairing to mRNA, modulating mRNA stability or ribosome accessibly (Caron, Lafontaine, & Massé, 2010; Urban & Vogel, 2007). The chaperone protein Hfq facilitates the hybridization of the sRNA to the target mRNA (Valentin‐Hansen, Eriksen, & Udesen, 2004). Upon sRNA binding to the TIR region, the RNA structure changes to either increase or decrease the TIR's initiation strength. In sRNA regulated transcripts, the RBS is buried within structural elements that are modulated by sRNA binding (Figure 2.2A and B). Since these regulators act as functionally active transcripts they can induce rapid changes in gene expression in

response to environmental changes, as opposed to protein regulators that need to be translated first (Shimoni et al., 2007). Additionally, after the stimulus is removed, the RNA is degraded and the cell can quickly revert back to its native state (Shimoni et al., 2007). Hundreds of sRNA targets have been identified and dozens of sRNA networks have been described (Storz et al., 2011). This regulation strategy has been evolutionarily successful, allowing prokaryotic cells to achieve global transcriptome changes by modulating a single sRNA (Modi, Camacho, Kohanski, Walker, & Collins, 2011).

Bioengineers have taken advantage of sRNA-based regulation to build synthetic genetic circuits (Chae, Kim, Choi, Park, & Lee, 2015; Kim, Park, Na, & Lee, 2014; Na et al., 2013; Rodrigo, Landrain, & Jaramillo, 2012; Yoo, Na, & Lee, 2013). Their ability to achieve quick and tight regulation makes them an attractive engineering platform (Kim et al., 2014). Recently the Sang Yup Lee lab has outlined a method for designing synthetic sRNAs (Yoo et al., 2013). Their sRNA system is composed of two parts: a scaffold sequence and a target-binding sequence. The 101 known *E. coli* sRNAs were screened using an *in vivo* reporter system to determine an appropriate scaffold (Hfq binding platform). This enabled them to derive a quantitative relationship between binding energy and repression that can be used to forward engineer sRNAs with different repression response curves. To demonstrate the use of this design strategy, *E. coli* was engineered to produce tyrosine, a commercially relevant amino acid and important precursor to many pharmaceuticals (Na et al., 2013). The group was able to down-regulate many genes in many different *E. coli* strains due to the portability of sRNAs and the universality (among most prokaryotic cells) of sRNA/Hfq-based regulation (Yoo et al., 2013). They were able to identify genes that when down-regulated increased tyrosine production drastically. Interestingly, these genes are essential and would not have been identified using classical

approaches such as genetic knock-out strains (Na et al., 2013). Another recent example that displays the simplicity and power of fine-tuning expression levels using sRNA-based regulation has been demonstrated in the development of an *E. coli* strain with increased 1,3-diaminopropane in production (Chae et al., 2015).

2.3.3 Toehold Switches

While riboregulators, such as sRNA-based strategies, have advantages, several limitations can be identified. They have a smaller dynamic range for both activation and repression of gene expression when compared to protein regulators (Callura, Cantor, & Collins, 2012; Lutz & Bujard, 1997; Mutalik, Qi, Guimaraes, Lucks, & Arkin, 2012). Due to their nature of sequestering RBSs in helices to prevent ribosome binding, and consequently translation initiation, they have sequence constraints. To facilitate sRNA-mRNA binding, specific host factors (such as Hfq) are required, limiting their ability to act as portable, universal regulators. Toehold switches are rationally engineered synthetic riboregulators that modulate TIR accessibility and their design has been tailored to reduce these limitations (Green, Silver, Collins, & Yin, 2014).

Toehold switches consist of two cognate RNAs: a transducer that encodes the output signal of the system (typically an mRNA) and a trigger RNA that modulates the output signal (Green et al., 2014). Both RNAs contain single-stranded regions that increase the efficiency of their interaction, eliminating the need for chaperones such as Hfq. When the trigger RNA binds to the transducer it unwinds a hairpin, exposing the hidden RBS of the transducer (Figure 2.2B). A noteworthy design feature is that the RBS is located within the loop region of a hairpin, consequently, the trigger RNA binding site is not limited to SD-like elements (Green et al., 2014). Toehold switches can achieve

gene regulation with a dynamic range over 400; they display a strong response when supplied with 5 μM trigger RNA and also able to sense concentrations of trigger RNA as low as 30 nM (Green et al., 2014).

Since the base-pairing between the transducer and trigger RNAs is not facilitated by Hfq, it has increased portability and is able to be used in a wide range of environments and systems. This inherent feature enabled the design of a synthetic genetic construct containing toehold switches that was functional when incorporated into a commercially available, highly purified, reconstituted *in vitro* transcription and translation system (Pardee et al., 2014). The *in vitro* transcription and translation system was successfully embedded into a cellulose matrix of paper able to be freeze-dried. This freeze-dried *in vitro* transcription and translation system maintain near full activity, even when stored at room temperature for over a year (Pardee et al., 2014).

These cell-free synthetic gene circuits have massive applications in the clinic, global health industry, research, and education fields. By engineering transducer elements able to activate gene expression of a reporter protein when introduced to environmentally relevant RNA, a new RNA sensor was developed. The flexible nature of these switches enabled the construction of sensors that are able to detect as low as 30 nM of several strains of the Ebola Virus and Zika Virus RNAs (Pardee et al., 2014; Pardee et al., 2016).

This technology has been used in many applications, such as cell-free storage of genetic circuits at room temperature for extended periods of time (Pardee et al., 2016). Paper-based RNA-regulatory systems can be used to rationally design and test synthetic genetic circuits, eliminating the need for transformations, overnight cultures, and performing labour-intensive *in vivo* fluorescence experiments. While both *cis*- and *trans*-regulators are valuable bioengineering tools, they are governed, and therefore limited, by

the canonical rules of SD mediated translation initiation.

2.4 Non-Canonical Regulation Mechanisms

Some mRNAs do not conform to the canonical rules of translation initiation as discussed above (Malys & McCarthy, 2011). This diversity is becoming increasingly appreciated by bioengineers as variations to the canonical mRNA:30S interactions are being uncovered. Many organisms have evolved alternate initiation mechanisms as a means to regulate gene expression (Boni et al., 2001; Kozak, 2005; Nakagawa, Niimura, Miura, & Gojobori, 2010; J. Zhang & Deutscher, 1992). In proteobacteria on average only 55 % of mRNAs contain SD sequence elements. However, the fraction of SD containing mRNAs varies significantly across prokaryotic genomes; out of 162 prokaryotic genomes analyzed between 11.6 %–90.8 % of the TIRs do not contain recognizable SD sequences, highlighting the overlooked diversity in initiation mechanisms (B. Chang, Halgamuge, & Tang, 2006).

A recent effort to catalogue the entire *E. coli* proteome using ribosome profiling confirmed that many *E. coli* genes do not rely solely on SD mediated initiation, and also highlighted the importance of gene expression regulation at the translation initiation stage (Li, Burkhardt, Gross, & Weissman, 2014). While examining proteins that assemble into multi-protein complexes following translation from a polycistronic mRNA, it was found that the vast majority of these protein components are produced in the correct stoichiometry via tight regulation of translation initiation (rather than being regulated post-translationally) (Li et al., 2014).

Understanding how cells regulate metabolic pathways will ultimately assist bioengineering efforts, including those to construct synthetic metabolic pathways and to

refactor existing ones. The well-characterized L-methionine biosynthesis pathway was examined using this ribosome profiling data. The cellular demand of this pathway was calculated, which allowed catalysis by MetE to be identified as the rate-limiting step (Li et al., 2014). If MetE is over- or under-expressed, cell viability is compromised, and this critical balance is achieved by fine-tuning and optimization at the translational level (Li et al., 2014). These analysis methods can be used by bioengineers to optimize synthetic metabolic pathways and better understand strategies that cells use to fine-tune translation initiation.

Interestingly, there are large discrepancies between the experimentally obtained genome-wide translation initiation rates and the initiation rates predicted by the RBS Calculator (Li et al., 2014). While the RBS Calculator is accurate at predicting translation initiation rates of single genes, its predictions aren't accurate across the entire transcriptome (Li et al., 2014). Although there is a wealth of knowledge describing the mechanisms governing translation initiation, there are apparent gaps in our understanding. Studying these gaps has revealed new modes of regulation that have yet to be repurposed for controlling translation initiation. For example, the contributions from uncharacterized mRNA binding proteins, the influence of structured mRNA (Marzi et al., 2007), or feedback regulation mechanisms (Boni et al., 2000; Nomura, Gourse, & Baughman, 1984) are promising regulatory elements of interest.

2.4.1 Contributions by Ribosomal Protein S1

Ribosomal protein S1 (S1) is an essential protein found in γ-proteobacteria (Sorensen, Fricke, & Pedersen, 1998). It is the largest protein in the 30S ribosomal subunit, spanning over 230Å, and is composed of six homologous RNA-binding domains

(OB-fold domains) that are not functionally equivalent (Duval et al., 2013; Lauber, Rappsilber, & Reilly, 2012). S1 binds the ribosome near the aSD motif of the 16S rRNA, anchoring itself to ribosomal protein S2 (Byrgazov et al., 2015). Among S1's many functions, it is essential for recruiting structured mRNAs to the ribosome (Qu, Lancaster, Noller, Bustamante, & Tinoco, 2012; Ringquist et al., 1995; Sorensen et al., 1998) and melting their structure (Qu et al., 2012).

Some mRNAs, such as *rpsA* and *rpsO*, contain bulky structured regions in their TIRs that act to modulate translation initiation (Boni, Isaeva, Musychenko, & Tzareva, 1991; Del Campo, Bartholomaus, Fedyunin, & Ignatova, 2015; Philippe et al., 1990). Interestingly, although required for the translation of structured mRNAs, S1 is also able to govern the expression of structured TIRs. The regulation of its own mRNA highlights the important role that S1 plays in controlling the expression of structured TIRs (Figure 2.2A). The TIR of the mRNA coding for S1 (*rpsA*) extends 91 nt upstream of the AUG start codon, folds into 3 hairpins, and lacks a canonical Shine-Dalgarno (SD) sequence (Skorski, Leroy, Fayet, Dreyfus, & Hermann-Le Denmat, 2006). Surprisingly, the usually inhibitory structure surrounding the start codon of the *rpsA* mRNA is necessary for its efficient translation (Tchufistova, 2003). When S1 is in molar excess of the 30S, cytosolic S1 can recognize the *rpsA* TIR, modulate its structure, and turn off its ability to be translated (Figure 2.2A) (Boni et al., 2000). Bacterial cells take advantage of feedback mechanisms to maintain precise cellular concentrations of not only S1, but other ribosomal proteins as well (Nomura et al., 1984; Nomura, Yates, Dean, & Post, 1980). While these mechanisms are useful strategies to maintain homeostasis of a protein by having the cellular availability directly modulate its initiation, they are largely unexplored by bioengineers.

Although organisms have evolved structured TIRs to modulate and fine-tune initiation frequencies, there is a lack of understanding regarding the critical role that S1 plays in their recruitment to the ribosome. SD binding has been shown to be important, but not obligatory, for efficiently initiating translation (Li et al., 2014). S1's contribution during translation initiation and its preference for AU-rich sequences are omitted from the RBS Calculator (Komarova, Tchufistova, Dreyfus, & Boni, 2005; Salis et al., 2009). A detailed understanding of S1's contributions to mRNA recruitment to the ribosome will be of great interest for future versions of *in silico* RBS design tools. Engineering synthetic feedback loops, similar to those responsible for the autogenous control of S1, will enable direct sensing of translation products by binding to dedicated structural elements within TIRs. Regulation of translation initiation based on sensing gene products will provide valuable tools for bioengineers, providing control devices that are orthogonal and can be executed in different *in vivo* and *in vitro* systems.

2.4.2 Leaderless mRNAs

All of the initiation mechanisms discussed thus far rely on 30S recruitment to an mRNA TIR upstream of the actual open reading frame. However, there are examples of 70S monosomes initiating translation directly on leaderless mRNAs, starting directly with the adenine of the AUG start codon (Balakin, Skripkin, Shatsky, & Bogdanov, 1992; Jones, Jaskula, & Janssen, 1992; Moll, Grill, Gualerzi, et al., 2002; Udagawa, Shimizu, & Ueda, 2004). There is no evidence of regions within leaderless mRNAs that interact with rRNA in a mechanism analogous to SD mediated initiation (Moll et al., 2001; Resch, Tedin, Grundling, Mundlein, & Blasi, 1996). The only requirement for translation of a leaderless mRNA is the absence of bulky structural elements, as unstructured mRNA

must be threaded through the 70S into the template channel during initiation (Figure 2.2C). This mechanism relies on a 70S:initiator-tRNA complex and is inhibited by IF3, which stimulates dissociation of the ribosome (Grill, Moll, Hasenohrl, Gualerzi, & Blasi, 2001; Huttenhofer & Noller, 1992; Moll, Hirokawa, Kiel, Kaji, & Blasi, 2004; Shean & Gottesman, 1992; Tedin et al., 1999).

Interestingly, unlike the bulk of cellular mRNAs, translation of leaderless mRNAs is more efficient in *E. coli* strains lacking ribosomal protein S1 (Moll et al., 2004; Shean & Gottesman, 1992). Under sub-optimal temperatures, a reduced concentration of cellular S1 which prevents the IF3-dependent destabilization of leaderless mRNA initiation complexes, is observed (Grill, Gualerzi, Londei, & Blasi, 2000). Leaderless mRNAs are then preferentially translated by S1-depleted 70S ribosomal subunits. Alternatively, during S1 over-expression, leaderless mRNAs are not translated (Delvillani, Papiani, Deho, & Briani, 2011).

Through this mechanism, it is possible to ensure the translation of specific subsets of mRNAs under specific conditions, as distinct subpopulations of ribosomes preferentially translate leaderless mRNAs (Moll & Engelberg-Kulka, 2012; Moll, Resch, & Blasi, 1998). Leaderless mRNAs are translated more efficiently under conditions such as carbon source limitations, low temperatures, and during slow growth (Moll & Engelberg-Kulka, 2012; Vesper et al., 2011). This strategy ensures that the translation of a subset of mRNAs continues during times of cellular stress and has been used throughout all domains of life. However, it has not been used by bioengineers. It will be useful to design and build mRNAs that undergo different translation initiation mechanisms favorable during different cellular conditions or during stress response, to enable more complex regulation strategies.

2.4.3 IRESs

In eukaryotes, canonical translation initiation involves the recognition of a covalent modification at the 5'-termini of mRNAs, and at least 9 initiation factors (eIFs) are required to recruit the mRNA to the small ribosomal subunit (40S) and scan for the start codon (Jackson, Hellen, & Pestova, 2010). Viruses and eukaryotic cells have evolved specialized *cis*-acting RNA elements called internal ribosomal entry sites (IRESs) to bypass the rate limiting and highly regulated 5'-cap dependent eukaryotic translation initiation machinery (Hellen & Sarnow, 2001; Jan, 2006). IRESs are able to drive translation with a reduced set of (or no) eIFs and recruit ribosomes through direct interaction with the ribosome (Kieft, 2008; Pestova & Hellen, 2003). The simplest example is the dicistrovirus intergenic region (IGR) IRES, which folds into a tightly compacted structure, part of which mimics a canonical tRNA anticodon:mRNA codon interaction, establishing the correct reading frame in the mRNA upon direct interaction with the ribosome (Figure 2.2C) (Jang, Lo, & Jan, 2009). An interesting feature is that initiation is not required to start at an AUG start codon (Fig. 2I). The core of the ribosome is highly conserved, and so it has been speculated that organisms from all domains of life could take advantage of this initiation mechanism (Melnikov et al., 2012).

The idea that a eukaryotic initiation signal can be portable across domains of life was recently explored by introducing an IGR IRES-controlled reporter system into *E. coli* (Colussi et al., 2015). Remarkably, the IGR IRES was able to drive translation in *E. coli* lysate. Mutagenesis approaches confirmed that structural elements of the IRES were responsible for initiation. The importance of these IRES structural elements was further highlighted using an X-ray structure, proving that the IRES interacted directly with conserved areas of the ribosomal core (specifically the tRNA binding sites). However,

unlike in eukaryotic systems, a start codon is required for IRES activity in *E. coli* (Colussi et al., 2015; Wilson, Pestova, Hellen, & Sarnow, 2000). Furthermore, bacterial IRES-mediated initiation also relies partially on a downstream degenerate SD element. The proposed model is that the IGR IRES acts as a 30S recruitment signal, and the SD sequence is required for correct positioning of the start codon in the P-site of the 30S (Colussi et al., 2015).

In the past, bioengineers have successfully used IRESs to control eukaryotic and viral gene expression (Koh et al., 2013). This method is advantageous, as translation of IRES-controlled mRNAs is reliable, portable, and can be continued under non-optimal cellular conditions, even when canonical translation has been halted. The robust structure that some IRESs adopt may resist context-dependent structural rearrangement resulting in a highly portable and modular tool (Venkatesan & Dasgupta, 2001). To our knowledge the use of these elements in prokaryotic synthetic biology has not been explored. However, the unique features of IRES-mediated translation initiation could result in robust translation initiation tools that could be used in all domains of life, and under non-optimal conditions. Some questions still remain unanswered. Are any prokaryotic initiation factors required? If signals from both RNA structure and SD sequence elements are used, how does the ribosome reposition itself in the mRNA? To this end, it will be important to dissect the mechanisms of 30S recruitment and identify the relevant features to enable their forward-design.

It is common to use *E. coli* as a platform to optimize, test, and build synthetic genetic circuits due to its fast doubling time and ease of cloning before transferring the completed construct to an industrially or medically relevant organism (Smanski et al., 2014). Unfortunately, even in closely related organisms, simple regulators do not behave

similarly. Signals, such as IRESs, that retain function in distantly related organisms may be an efficient way to transfer a synthetic genetic circuit between different hosts. However, a bacterial IRES element (and other synthetic TIRs) can compete with natural mRNAs for ribosomes, causing complex and unknown regulatory problems resulting from competition for finite resources. Insulating the expression of synthetic genetic circuits or partitioning the cellular ribosome pool can address these concerns.

2.4.4 Orthogonal Ribosomes

When a synthetic gene circuit is inserted into a host cell, competition for cellular resources increases (Cardinale, Joachimiak, & Arkin, 2013; Carrera, Rodrigo, Singh, Kirov, & Jaramillo, 2011; Scott, Gunderson, Mateescu, Zhang, & Hwa, 2010; J. Vind, Sorensen, Rasmussen, & Pedersen, 1993). A trade-off takes place and a resource balance must be reached between the host's native cellular functions and the synthetic circuit if impaired cell growth is to be avoided (Brophy & Voigt, 2014). Typically, toxicity is not directly caused by an increased concentration of synthetic biomolecules (Gorochowski, Avcilar-Kucukgoze, Bovenberg, Roubos, & Ignatova, 2016). Negative effects usually arise from reduced ribosome availability (Scott et al., 2010). All translation initiation control mechanisms discussed thus far have focused on the mRNA TIR. However, engineering regions of the ribosome involved in TIR binding is an attractive approach. For example, engineering orthogonal or specialized ribosomes able to specifically translate a particular set of mRNAs at a predictable rate could limit the burden on the host's translation machinery Figure 2.2C).

Early efforts to develop orthogonal *E. coli* ribosomes involved mutating the SD of a synthetic mRNA to an aSD. The mRNA was not translated in *E. coli*; however, when

complementary mutations were introduced in the 16S rRNA, expression was recovered (A. Hui & de Boer, 1987; AS Hui, Eaton, & de Boer, 1988). Unfortunately, cells containing a pool of engineered ribosomes ultimately succumb to cell lysis and death (Jacob, Santer, & Dahlberg, 1987; Lee, Holland-Staley, & Cunningham, 1996; Wood & Peretti, 1991). This approach has several challenges as the ribosome is extremely complex and any changes resulting in faulty activity lead to cell death (Jacob et al., 1987). Interestingly, orthogonal ribosomes are only two-fold disadvantaged over native ribosomes for the translation of bulk cellular mRNAs (Skorski et al., 2006). These findings further highlight that, although the SD:aSD interaction plays a significant role, it is not essential for the synthesis of bulk cellular proteins (Skorski et al., 2006). This also demonstrates the limitations of insulating synthetic gene circuits at the translational level based solely on SD:aSD interactions.

To overcome this, the Chin lab used a two-part selection method to generate completely orthogonal 16S:mRNA pairs (Rackham & Chin, 2005). They built a synthetic genetic circuit that tested a TIR's ability to be actively translated by a natural ribosome, a ribosome containing a mutated 16S rRNA, both, or neither. This strategy enabled a library of TIRs to be enriched for those unable to be actively translated by a natural ribosome, but able to be translated by ribosomes containing mutated 16S rRNAs. Ten of these orthogonal pairs were successfully engineered (Rackham & Chin, 2005). These orthogonal ribosomes acted as necessary stepping-stones to perform more complex mutations.

However, when a completely orthologous 70S ribosome dissociates into its subunits, it will exchange with those of the native ribosome, ultimately leading to a mixed population. To circumvent this, bioengineers from the Mankin lab tethered a 30S subunit

to a 50S. Using available X-ray crystal structures of the *E. coli* ribosome (Brodersen, Clemons, Carter, Wimberly, & Ramakrishnan, 2002), they chose to connect helix 44 on the 30S to helix 101 on the 50S, generating a distinct and orthogonal population of ribosomes (Orelle et al., 2015). The Chin lab used a similar approach to develop a tethered ribosome (Fried, Schmied, Uttamapinant, & Chin, 2015).

In both cases, the tethered orthogonal ribosomes are not as efficient at translating an mRNA as wild-type ribosomes. However, these are a critical platform for carrying out 50S ribosome engineering, such as improving the peptidyl transfer reaction between noncanonical amino acids (Dedkova, Fahmi, Golovine, & Hecht, 2003; Maini et al., 2013). Bioengineers from the Mankin Lab modified regions in the ribosome in areas critical for cell viability, creating ribosomes able to translate a leader peptide that caused the native ribosome to stall (Orelle et al., 2015). A detailed understanding of the TIR-ribosome interactions was necessary for these orthogonal translation systems to be further developed. Increasing our understanding of the mechanisms governing ribosome recruitment to the TIR will enable the development of portable translation systems with increased functionality.

2.5 Conclusions

All organisms rely on precise control over translation to maintain viability. Prokaryotic organisms have evolved cunning tactics to coordinate the vast number of biomolecules responsible for controlling translation. Decades of biochemical research have worked to uncover the complexities governing translation initiation, giving insight into diverse control mechanisms. Of the steps involved in translation, initiation has become the most diverse and ultimately the rate-determining step. These control

mechanisms have been repurposed by bioengineers to exert control over synthetic gene circuits. Approaches as simple as modulating the strength of SD:aSD interactions, and as complex as refactoring riboswitches and engineering orthogonal ribosomes, have been met with great success. Bioengineers have been able to achieve tight and rapid control over translation, spanning several orders of magnitude. Furthermore, the quantification of these regulatory mechanisms has allowed the development of predictive biophysical models, which are used with great accuracy. These tools have been invaluable for designing synthetic mRNAs to conform to certain specifications and to respond to specific stimuli *in silico*, with remarkable accuracy when moved *in vivo*.

Table 2.2. Characteristics of TIR elements

Element	Maximum Expression Range (~fold change)	Dynamic	Reference
RBS Calculator/SD	100,000	no	Espah Borujeni *et al.,* 2014
Riboswitches	383	yes	Espah Borujeni *et al.,* 2014
sRNA	10	yes	Mutalik *et al.,* 2012
Toehold switch	400	yes	Green *et al.,* 2014
IRES	2.5	no	Colussi. *et al.,* 2015
Leaderless mRNA	- - - - - - -	no	- - - - - - -

The role that the three-dimensional structure of an RNA element plays in recruiting the 30S to the TIR and modulating translation initiation is becoming increasingly appreciated. Although much of the ribosome has diversified between species, the ancient core remains similar even between eukaryotic and prokaryotic systems. Taking advantage of initiation mechanisms that rely on interactions with conserved areas of the ribosome core may be an effective way to design universal translation initiation signals. The behavior or strength of an SD element depends largely on the context in which it is placed, as they may form structural elements with adjacent RNA sequences in

the TIR or coding sequence (Cironi, Swinburne, & Silver, 2008; Robinson-Mosher, Shinar, Silver, & Way, 2013). This feature limits the portability of standardized translation initiation elements and creates challenges for the development of well-characterized libraries of genetic parts able to be reused predictably, one of the main approaches towards scalable synthetic biology. Prototype circuits designed and built for operation in a production strain could be first built as a prototype and ported into a model organism for successive rounds of testing, reducing the time to industrial scale-up, if portable TIRs were developed.

The ability to design cell-free synthetic biological systems will be enhanced by adapting translation initiation mechanisms that rely on a limited or reduced set of auxiliary factors. Medically relevant cell-free technologies have already benefited from riboregulators that don't rely on protein factors (Pardee et al., 2016). The independence of toehold switches from the Hfq mediated RNA:RNA hybridization reduced the systems complexity and increased reliability. Exploring non-canonical or alternative mechanisms of translation initiation, such as leaderless mRNAs, will lead to simplified *in vitro* translation systems. More work is required to understand how these non-canonical signals function and to gain an in-depth understanding of the proteins and RNAs involved in governing the translation of structured RNAs. While our ability to predict translation initiation rates has improved, a more detailed understanding of under-explored initiation mechanisms will result in enhanced control and predictability, leading to more diversity in the toolset we have to control cellular behavior. This will be critical for the rational design of synthetic genetic circuits based on the existing features of the translation machinery.

Chapter 3

The *rpsA* Translation Initiation Region: Biophysical Characterization of a Structure-Driven Translation Initiation Mechanism

3.1 Introduction

Over the last century a thorough study of bacterial protein biosynthesis has yielded detailed insight into how microbes tune gene expression at the translation level. This understanding has revealed diverse strategies that cells use to regulate this multi-step process, which can expend 50 % of all cellular energy in a rapidly dividing bacterium (Russell & Cook, 1995). It is critical that this energy intensive process be tightly regulated to ensure proper allocation of cellular resources. Consequently, the initiation of this process has become the rate-determining step of ribosome dependent protein synthesis. While much of the translation process is highly conserved across the domains of life, the mechanistic details of translation initiation have vastly diversified (Laursen et al., 2005). Modulating factors that influence the correct positioning of the mRNA on the small ribosomal subunit during the initial stages of translation is an efficient and effective strategy that cells have evolved to respond quickly to changing cellular conditions.

The initiation rate is related to the thermodynamics of the interaction between the mRNA translation initiation region (TIR) and the 30S, which is fine-tuned by the kinetics of RNA folding in the area surrounding the start codon. Specifically, the energetics of the Shine-Dalgarno (SD) interaction with the 3' end of the 16S rRNA (anti-SD), and the hybridization of initiator-tRNAs to the start codon contributes to initiation strength. The strength of this interaction is tuned by the spacing between the RBS and the start codon, which can force the ribosome to reposition itself. In addition to these base-pairing interactions, additional signals imbedded within the TIR of an mRNA heavily influence its translation initiation rate (Kozak, 2005). Stand-by sites can act as a platform on the mRNA allowing the 30S to first bind non-specifically and transiently, waiting until the SD becomes accessible (Studer & Joseph, 2006).

Sophisticated regulation mechanisms that rely on the structural dynamics of the mRNA are leveraged to modulate access to the TIR. However, to what extent is not fully understood. One common theme is that highly structured TIRs are typically less efficient in facilitating translation. Translation initiation efficiency is strongly anticorrelated with a kinetic competition of RBS unfolding vs. dissociation of the mRNA from the 30S subunit (Mustoe et al., 2018), while RNA structure acts to attenuate the initiation rate (M. H. de Smit & van Duin, 1990; Goodman, Church, & Kosuri, 2013; Salis et al., 2009). Structured RNA elements surrounding start codons are able to detect changes in cellular Mg^{2+} concentration (Cromie et al., 2006), pH (Nechooshtan et al., 2009), concentrations of carbon sources and metabolites (Wang et al., 2008), as well as temperature changes (Narberhaus et al., 2006), enabling rapid changes in gene expression according to environmental conditions. Autogenous negative (Boni et al., 2000) and positive feedback mechanisms (Mandal et al., 2004) have evolved to maintain specific levels of protein as required. This remarkable ability to tune and control translation initiation frequency over several orders of magnitude in response to changing cellular conditions by modulating a relatively short RNA sequence demonstrates the critical role that the structure of the TIR plays in protein biosynthesis. The cumulative knowledge of the mechanisms of protein synthesis has enabled the development of predictive rules for translation initiation rate.

Interestingly, global analysis of the *E. coli* transcriptome has revealed that mRNAs without SD elements are as common as those with, suggesting that other initiation strategies could exist (B. Chang et al., 2006). Central to these sophisticated and diverse regulation strategies are complex and dynamic RNA structures, evident in examples such as ribozymes and riboswitches (Vigar & Wieden, 2017). Typically, RNA structure acts to attenuate translation initiation by occluding the RBS from the mRNA

binding channel. Surprisingly, a cryogenic electron microscopy (cryo-EM) structure of the intergenic region (IGR) internal ribosome entry site (IRES) in complex with the bacterial ribosome has been reported. This illustrated an inherent ability to support structure-based initiation in addition to the SD-based initiation mechanism, however no such structure of a bacterial mRNA has been reported to date (Colussi et al., 2015).

The *rpsA* gene coding for ribosomal protein S1 (S1) is an interesting candidate to investigate non-canonical initiation elements. The highly structured *E. coli rpsA* TIR contains only a degenerate SD-sequence, and studies using single nucleotide (nt) substitution and orthogonal ribosomes provide additional evidence that its translation is not driven solely by a SD-based mechanism (Skorski et al., 2006). Despite its strong secondary structure, large size, and lack of a canonical SD sequence, the TIR is highly efficient in driving translation *in vivo*. The specific fold of the *rpsA* TIR may create an optimal arrangement of sequence elements that interact with the ribosome, similar to how some viral mRNAs operate (Skorski et al., 2006). Previous work has shown that a stretch of 91 nt upstream of the start codon of the *rpsA* mRNA constitutes the minimum segment required to efficiently drive translation (Tchufistova, Komarova, & Boni, 2003). Phylogenetic analysis of *rpsA* TIRs has revealed little sequence similarity between organisms within the ϒ-proteobacteria class (Table 3.1), with the exception of two apical GGA motifs in stem-loops I and II. However, it is hypothesized that all can potentially fold into a similar 3-stem-loop structure, further supporting the importance of this structure (Boni et al., 2001). It remains unclear how and if the structural elements in the *rpsA* TIR contribute to translation initiation.

To investigate the *rpsA* TIR – sequence function relationship, I performed an in-depth mapping of the *rpsA* TIR activity using high-throughput mutagenesis followed by

phenotypic cell sorting and next generation DNA sequencing. Mutations in the *rpsA* TIR resulted in varied translation initiation efficiencies, which were used to construct a standardized TIR library with different initiation strengths covering three orders of magnitude. RNA probing assays guided near-atomic level structures constrained by *ab initio* models built from solution X-ray scattering (SAXS) enabled us to investigate how the respective point mutations affect the three-dimensional structure of the *rpsA* TIR. Our results indicate that the 109 nt long TIR of the *rpsA* gene constitutes a bona fide structure-based initiation element, providing the first evidence that a structure-based translation initiation mode exits in bacteria.

Table 3.1 Comparison of *rpsA* TIRs from different organisms. Sequences were chosen by selecting 150 base-pairs (bp) upstream from the start codon plus 30 bp downstream. Pairwise sequence alignments were performed using EMBOSS Water (Madeira et al., 2019).

Group	Organism	% Identity to *E. coli*
Gram-negative	*Shigella flexneri*	100
	Helicobacter pylori	40
	Aquifex aeolicus	50
	Coxiella burnetii	54
	Borrelia burgdorferi	45
	Chlamydia muridarum	47
	Chlamydia pneumoniae	39
	Neisseria meningitidis	54
High-GC gram-positive	*Mycobacterium tuberculosis*	48
Low-GC gram-positive	*Lactococcus lactis*	46
Deinococcus thermus	*Deinococcus radiodurans*	48
	Thermus thermophilus	44
Eukaryote (plastid)	*Arabidopsis thaliana*	45
	Chlamydamonas reinhardtii	42

3.2 Results

3.2.1 Quantitative High-Throughput Screening Identifies Sequence-Structure-Function Relationships in the *rpsA* TIR

To define the sequence and structural requirements for translation initiation on the *rpsA* mRNA its mutational landscape was surveyed. In order to map the sequence-function relationships of nucleotide variations to their translation efficiency (TE), the *rpsA* TIR was linked to a fluorescent output allowing gene-expression to be measured. To control for compounding factors such as cellular noise (ex. plasmid copy number, cell size, and stage of cell cycle) the reporter system was designed to contain two independently functioning modules: an expression module and noise module allowing for

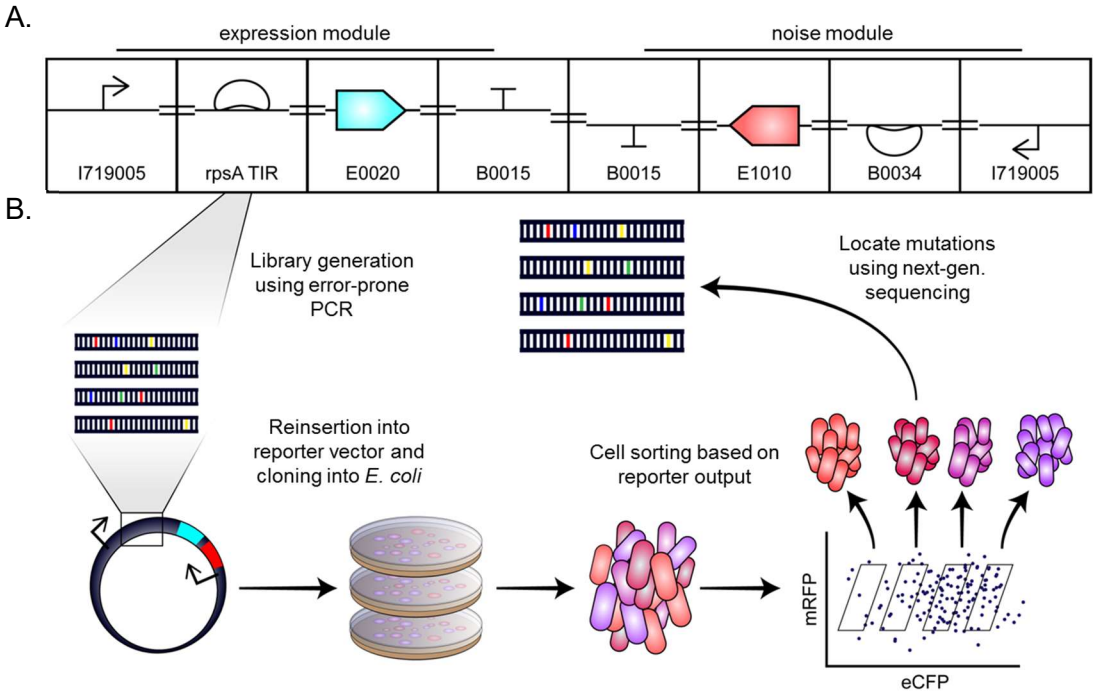


Figure 3.1 Sort-seq strategy for generating a library of translation regulators. (A) Reporter construct used to study the *rpsA* TIR *in vivo*. It consists of two independent modules: an expression module containing the *rpsA* TIR, and a noise module to measure cellular gene expression independent of the *rpsA* TIR. Gene expression from the two modules is normalized by correlating the outputs of the two reporters. (B) General strategy used to generate *rpsA* TIR variants and measure their efficiency *in vivo*.

the accurate *in vivo* quantification of *rpsA* TIR driven protein biosynthesis (Figure 3.1) (Liang, Chang, Kennedy, & Smolke, 2012). A previously identified 109 nt segment of the *rpsA* TIR that includes all signals necessary for efficient translation of the *rpsA* mRNA was inserted into an expression module enabling the quantification of *rpsA* TIR driven translation.

To generate a library of differently expressing *rpsA* TIRs the region upstream of the eCFP CDS was PCR amplified under error-prone conditions. The purified error-prone PCR products were used in a subsequent primer extension reaction, generating reporter plasmids containing mutant *rpsA* TIR controlled expression modules (Figure 3.1B). The resulting plasmids containing the TIR variants were cloned into *E. coli* BL21 GOLD-(DE3), resulting in 120,000 colony forming units (CFUs). Two biologically independent libraries were generated and analyzed via flow-cytometry.

Each of the mutant libraries generated had a dynamic range spanning approximately four orders of magnitude (Figure 3.2B). A fluorescence-activated cell sorting (FACS) based strategy was used to isolate similarly expressing cells among the mutant libraries into four distinctly behaving populations (Figure 3.2C).

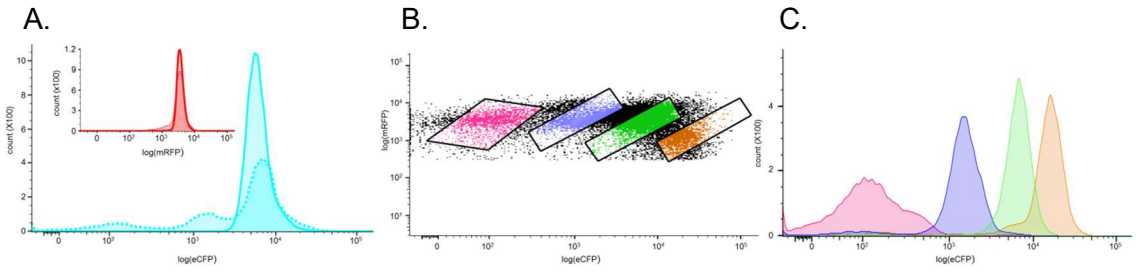


Figure. 3.2 The *rpsA* TIR library resulted in an increase in eCFP expression over four orders of magnitude. (A) Comparing the wt TIR (solid trace) to the mutant library (dotted trace). mRFP expression remains approximately constant in both populations (insert). (B) Cells containing the mutant library were fractionated based on their respective eCFP/mRFP ratio into four distinct populations. (B) The sorted populations were reanalyzed using flow cytometry, the DNA was isolated, and subjected to next-gen. seq.

Coupling this approach with next-generation sequencing (sort-seq) allows for the isolation of variants with defined activities from the population, and to link changes in RNA sequence to a phenotype (Figure 3.1B) (Holmqvist, Reimegard, & Wagner, 2013). Variants isolated from the no-expressing, low-expressing, and high-expressing cells were compared to variants isolated from wt-like cells. Using the wt-like population including wt TIRs and those harboring phenotypically silent variants acted to increase sensitivity. The TIR variants (Q-value < 0.1) are overlaid on the wt *rpsA* secondary structure in Figure 3.3.

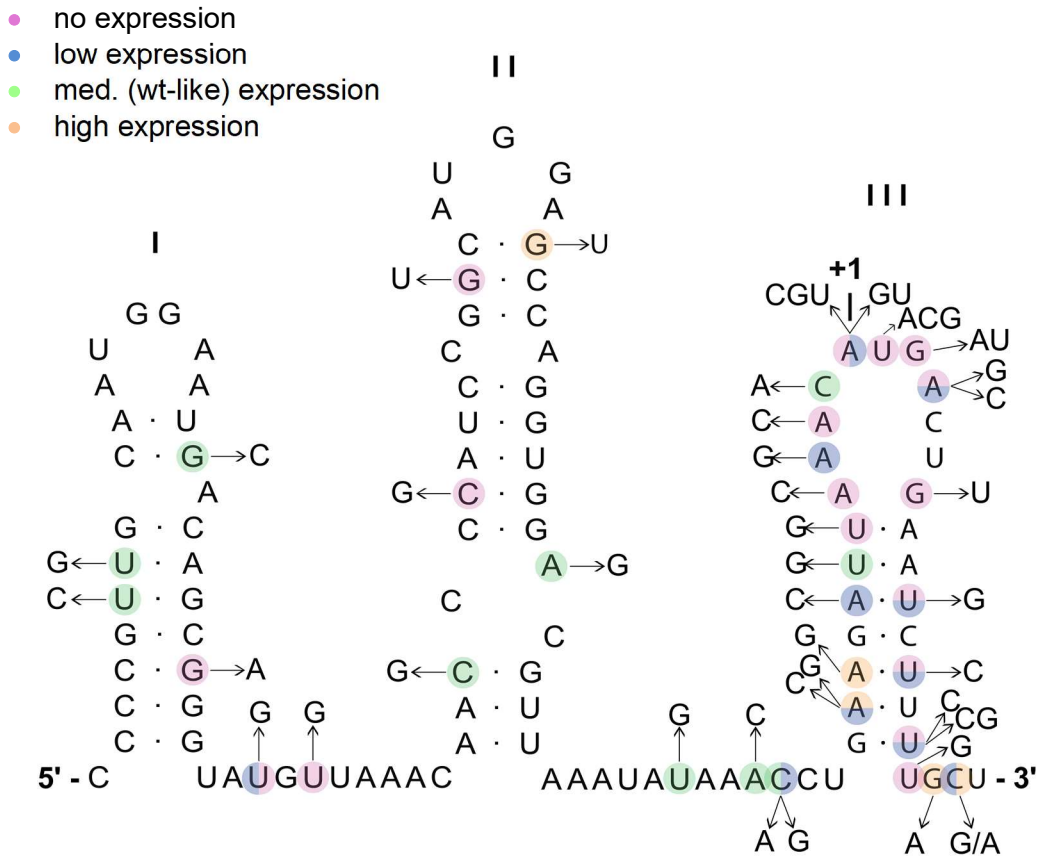


Figure 3.3 High-confidence mutations identified in sort-seq. pipeline mapped onto the *rpsA* TIR. The AUG start codon is set as +1, and the stem-loops are numbered from I–III.

Consistent with the essential role of the start codon for translation initiation, mutating the AUG start codon in the 3' stem-loop abolishes activity. Furthermore, this

confirms that the initiation of translation occurs at this AUG rather than at the downstream AUG of the eCFP CDS. There is also a very slight, but detectable eCFP fluorescence when the AUG start codon is changed to GUG or UUG, but none with a CUC codon. Instances of early stop codon introduction after the start codon, such as at position +7, abolishes eCFP activity. The effect caused by most of the mutations occurring after the start codon can likely be attributed to changes in GC content.. Limiting RNA secondary structure in the first roughly 5–10 codons by reducing GC content is linked to increased translation initiation rates (Bentele, Saffert, Rauscher, Ignatova, & Bluthgen, 2013). There is reduced eCFP fluorescence as GC content increases. Conversely, when GC content decreases (+16 and +17), there is increased eCFP fluorescence. Additionally, the previously reported A-9G variant was located in our high expression pool, consistent with previous studies (J. Vind et al., 1993).

Most mutations associated with a loss of function occur in the two stem-loops located at the 3' end of the TIR, however there are instances of single point mutations further than 70 nt upstream from the start codon that reduce eCFP fluorescence. There are many variants that weaken stems, for example a C to U substitution that could change the G-C bp to a G-U bp at position -37, and a C to G mutation at position -31 that abolishes a G-C bp, which either increase or decrease eCFP fluorescence respectively.

To address the possibility that differential activity for *rpsA* TIR variants could be due to point mutations altering the transcript abundance, RT-qPCR was performed on select variants. The same dual reporter construct used in FACS experiments was used for this assay; the test eCFP gene was analyzed and the opposite-strand mRFP transcript was used as an internal reference. The results show no significant difference in transcript abundance between the variants and wt (Figure S3.3) with the exception of Δ-66 and A-

9G, which show a slight increase and decrease in transcript abundance, respectively. However, since Δ-66 is a low activity variant and A-9G is a high activity variant, these differences in transcript abundance are likely not responsible for their differential activity compared to wt.

3.2.2 *rpsA* Driven Translation Cannot be Predicted Computationally or by 30S Binding Strength.

Algorithms summing the thermodynamic properties that contribute to the interaction between an mRNA and a 30S have been developed, enabling the translation initiation rates of mRNAs to be predicted. These algorithms calculate the total free energy of binding between the 30S and the mRNA, resulting from energies of several intra- and intermolecular interactions between the components involved in initiation. To determine if the fluorescent outputs obtained in the *in vivo* experiments correlate to translation initiation rates predicted *in silico*, V2.1 of the Salis Lab RBS Calculator – which is among the most advanced translation initiation rate calculators – was used (Figure 3.4).

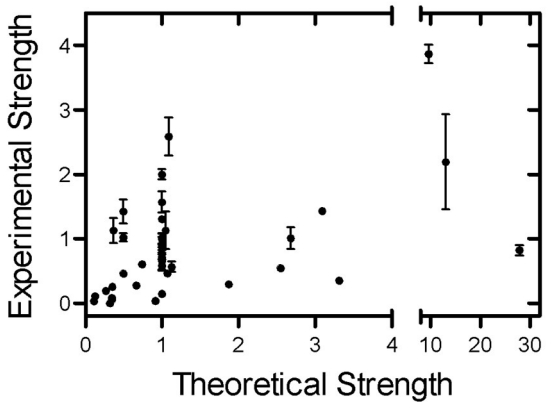


Figure 3.4 *rpsA* driven translation cannot be predicted computationally. Scatter plot comparing the ratio of measure eCFP/mRFP fluorescence to the predicted translation initiation rate. Predicted rates were calculated using version 2.1 of the RBS Calculator.

The results show that the RBS Calculator poorly predicts the *in vivo* results (correlation coefficient 0.61). This finding indicates that there are mechanistic details governing *rpsA* TIR driven translation initiation not present in the underlying thermodynamic models.

What is contributing to the unpredicted, exceptionally high translation initiation rate of the *rpsA* TIR? I hypothesized that the *rpsA* TIR has evolved to optimally interact with the ribosome to achieve its high level of activity, and that the strength of the interaction between the 30S or S1 and the *rpsA* TIR variants correlate to our *in vivo* fluorescence experiment results. To this end the direct binding of *rpsA* TIR RNAs to the 30S in the absence of the initiator tRNA—before mRNA accommodation—was measured.

Several variants isolated from the sort-seq pipeline were analyzed *in vitro* using a nitrocellulose filter-binding assay, which allowed dissociation constants to be measured. Two low expressing (A-4C and G-44U) and two high expressing (C-1U and A-9G) variants were selected. To investigate how altering the start codon containing stem-loop affected activity of the TIR variants C-1U, A-4C, and A-9G were selected for further analysis. A single point mutation 44 nt upstream of the start codon would not usually be expected to have a large effect on activity since an RBS typically encompasses only about 30 nt around the start codon (Laursen et al., 2005). To find out why some upstream mutations abolish activity, G-44U, located in the second stem-loop was chosen for further study.

A correlation between dissociation constants and *in vivo* fluorescence measurements would suggest all interactions between the 30S and the *rpsA* TIRs *in vitro*

are accounted for. While a discrepancy between the experimental and theoretic results would highlight the presence interactions not included in the current thermodynamic models. The dissociation constants (K_d) obtained for binding between the *E. coli* 30S subunits and S1, and the *rpsA* TIR variants are similar with no correlation between reporter output and binding strength (Figure 3.5 and Table S3.1).

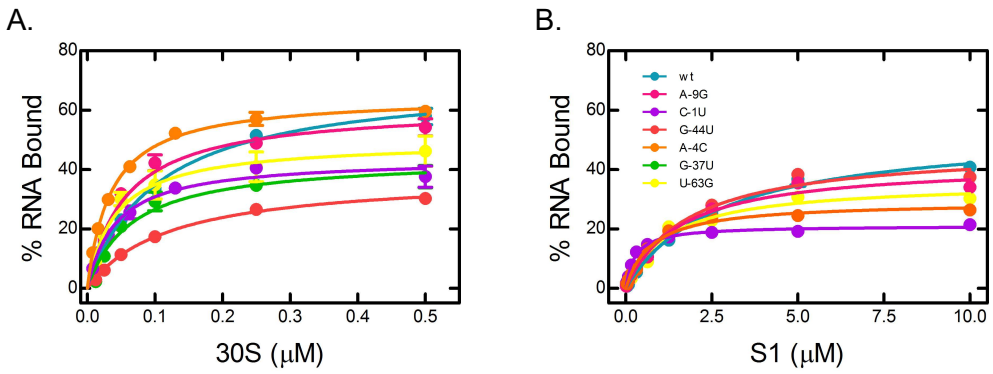


Figure 3.5 *rpsA* driven translation cannot be predicted by 30S binding strength. Purified ribosomal subunits (30S) (A) and S1 (B) binding to [^{32}P]-labelled *rpsA* TIR variants was quantified by titrating 30S or S1 with RNA. The reaction mixture was subjected to nitrocellulose filtration. The amount of RNA interacting with 30S or S1 was quantified by scintillation counting. Dissociation constants (K_d) for RNA binding were determined by fitting a curve using a hyperbolic equation.

3.2.3 Minimal *In Vitro* Reconstituted Protein Expression System is Sufficient for *rpsA*

The high affinity of all sequence variants to the 30S ribosome subunit *in vitro* suggest that other non-ribosomal proteins or non-coding RNAs, not present in our binding assays, might be involved in the efficient translation of the *rpsA* mRNA *in vivo*. To test if additional factors are required for translation initiation on the *rpsA* TIR, *in vitro* protein synthesis assays were performed using a highly purified and reconstituted coupled transcription and translation (TX-TL) system (Shimizu et al., 2001). Our dual reporter plasmid was used to seed the *in vitro* TX-TL reaction, generating an eCFP fluorescence

output analogous to the *in vivo* experiment. Reporter fluorescence expressed in this system show the same relative expression levels as observed in the *in vivo* experiments confirming that *rpsA* translation does not require any additional factors (Figure 3.6A). Additionally, it may be possible that the cellular concentration of S1 is being altered in our *in vivo* experiments by introducing synthetic S1 RNA aptamers. This could have an effect on translation globally; different *rpsA* variants could have different effects on cellular S1 concentrations, leading to an alteration in translation. To address this possibility, the activity of select inactive variants (A-4C and G-44C, Δ66, Δ82) and active variants (C-1U and A-9G) was tested in the TX-TL assay. The *in vitro* TX-TL system does not have the ability to produce additional S1 if it is being titrated off of the ribosome. If there is in fact an increase of ribosomes devoid of S1, there should be a decrease in translation of our noise module (mRFP fluorescence). Additionally, this should be dependent on the translation efficiency of the specific variant in the reaction.

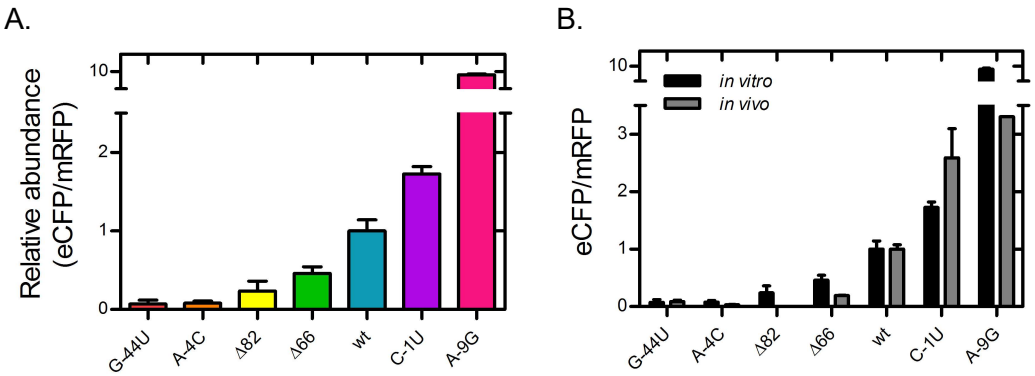


Figure 3.6 *In vitro* activity of *rpsA* TIR variants. (A) DNA expressing eCFP under control of *rpsA* TIRs, and mRFP under control of a strong RBS were analyzed in the PURExpress system using eCFP/mRFP fluorescence to detect output via fluorescence spectroscopy. (B) Activity of *rpsA* TIR variants *in vitro* vs. *in vivo*. Plasmid DNA expressing the dual reporter constructs was used as input for the reaction, and eCFP and mRFP fluorescence was detected. Experiments were performed in triplicate and the standard deviation is shown.

However, the activities of the variants with respect to wt mirror the *in vivo* observations (Figure 3.6B), suggesting that there is not an increase of ribosomes devoid of S1 in the *in vivo* experiments.

S1 has been shown to be essential for the translation of structured mRNAs such as *rpsO*, which contains a pseudoknot structure (Duval et al., 2013). To confirm that the structural element embedded within the *rpsA* TIR still requires unfolding by S1, a TX-TL system with ribosomes lacking S1 was used. The fluorescence output from reactions seeded with single reporter plasmids containing various TIRs (*rpsO*, *rpsA*, and BBa_B0034) previously characterized *in vivo* were monitored (Figure S3.1). BBa_B0034 is a medium to strong RBS with little structure surrounding the start codon, and should be only moderately dependent on S1 for translation. There is a massive reduction in eCFP fluorescence in TX-TL reaction with ribosomes free of S1. In the case of *rpsA*, the eCFP signal is reduced to 3.6 %, *rpsO* is reduced to 12.0 %, and BBa_B0034 shows no measurable signal compared to with S1. The addition of recombinant S1 at a ratio of 1:1 S1 to ribosomes rescues the activity in all three cases (Figure 3.7). These results agree with toe-printing studies showing that 30S initiation complex formation on *rpsA*, as well as SD-containing mRNAs does not occur in the absence of S1. This finding indicates that S1 is required for translation of most leadered mRNAs, even those with relatively strong SD sequence (Boni et al., 2001; Boni et al., 1991; Komarova, Tchufistova, Supina, & Boni, 2002). Other toe-printing studies have shown that RBSs containing strong SD sequences and weak secondary structure are able to be translated in the absence of S1 (Balakin, Bogdanova, & Skripkin, 1992; Duval et al., 2013; Farwell, Roberts, & Rabinowitz, 1992), suggesting that strong base-pairing interactions are able to compensate for lack of S1 on the ribosome. It is possible that the interaction between

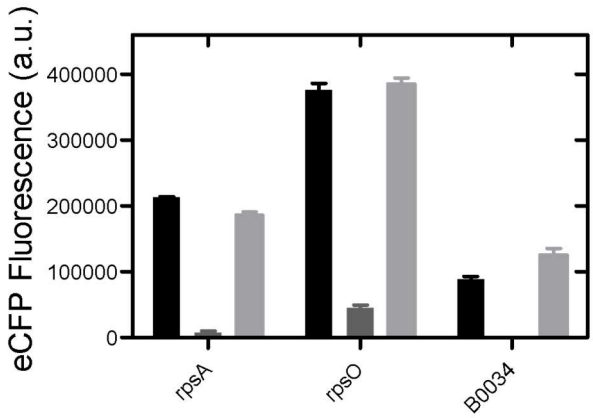


Figure 3.7 Effect of S1 removal on *in vitro* translation. The essentiality of S1 was investigated by monitoring the expression of eCFP translationally controlled by *rpsA*, *rpsO*, and B0034 TIRs in PURE reactions devoid of S1. Black: 30S, Dark grey: $30S^{-S1}$, Light grey: $30S^{-S1}$ + S1. The same experiments were performed in triplicate, and the standard deviation is shown.

BBa_B0034 and the 16S rRNA is not sufficient to overcome any inhibitory structure present in this TIR in the absence of S1. Regardless, these results underline the importance of S1 for the translation of transcripts with different structural elements, including *rpsA*.

Despite its critical importance in initiation, S1 was initially characterized as an interference factor and translational repressor due to the negative effect it had on the initiation of some mRNAs (Jay & Kaempfer, 1974). To better understand this feature and to investigate S1's specificity to the mRNAs in our *in vitro* assay, the TX-TL system was supplemented with recombinant S1 at 5-, 10-, and 15-fold molar excess over ribosomes. At 5-fold excess S1, *rpsA*-eCFP is repressed by 80 % showing a much higher specificity than to BBa_B0034, which is only repressed by 40 % (Figure 3.8). Interestingly, *rpsO* shows no repression but rather a slight increase in activity. At a 10-fold excess of S1, *rpsA* shows almost no activity while BBa_B0034 and *rpsO* are repressed by 60 % and 30 %, respectively. Addition of S1 up to 15-fold excess results in a further reduction in

expression of eCFP driven by BBa_B0034 and *rpsO*, but these never reach the same level of repression as for *rpsA*, indicating that free S1 is highly specific for its own mRNA (Figure 3.8).

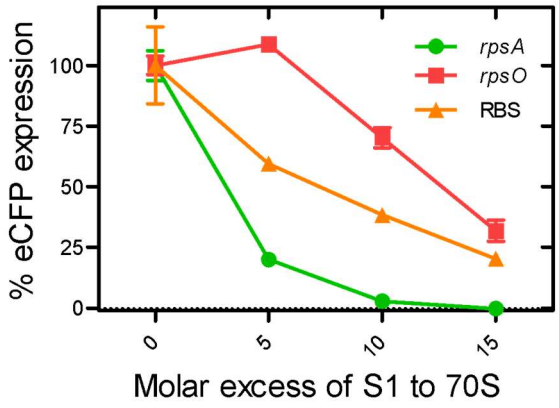


Figure 3.8 Effect of excess S1 on *in vitro* translation. The effect of adding S1 in excess over ribosomes to the PURE *in vitro* transcription/translation system was studied for constructs expressing eCFP under the control of *rpsA*, *rpsO*, and B0034 TIRs. Experiments were performed in triplicate and the standard deviation is shown.

3.2.4 Structural Analysis of *rpsA* TIR Variants via SHAPE and SAXS

Recent high-throughput selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) probing has revealed insight into the dynamic structure of the *rpsA* mRNA, which changes in the presence and absence of proteins (Mustoe et al., 2018). The 5' end of its TIR remains structured in both the presence and absence of protein. However, the area surrounding the start codon is unfolded in the presence of protein. To obtain a detailed understanding of the three-dimensional structure of the *rpsA* TIR variants both *in vitro* and *in silico* structural analyses were performed. Hydrodynamic properties using SAXS, and local nucleotide flexibility determined by SHAPE were used to guide *in silico* structure prediction of the different *rpsA* TIR variants.

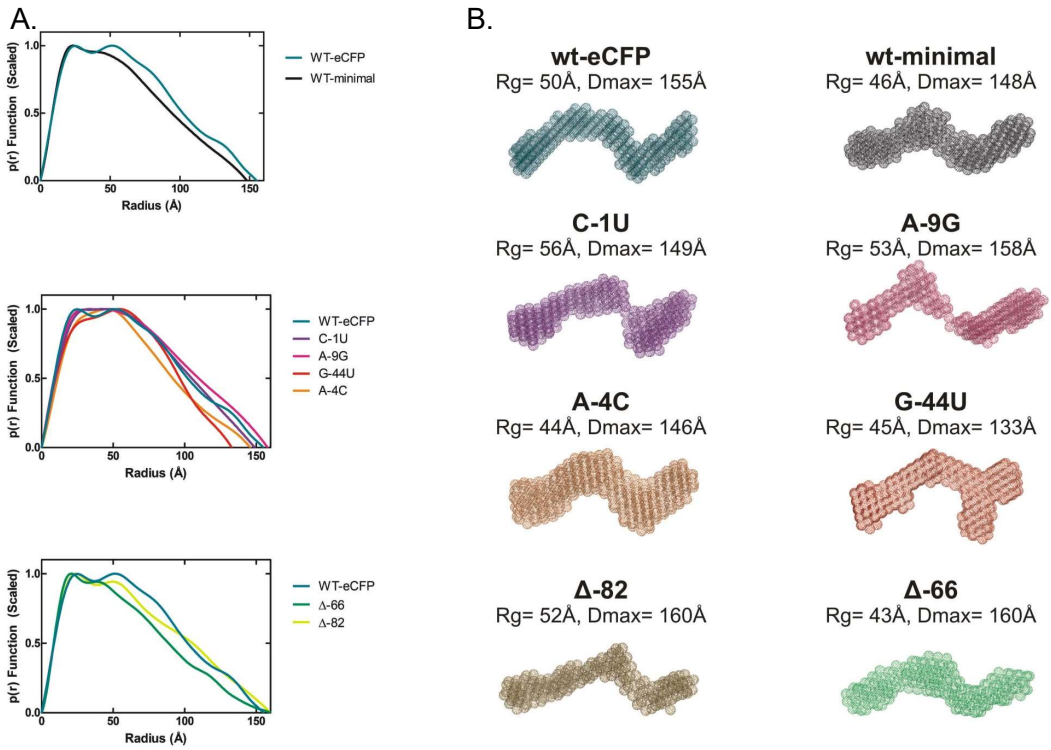


Figure 3.9 Low resolution models of *rpsA* TIR variants. The *rpsA* wt-eCFP and wt-minimal constructs, as well as four point mutants and two truncations were analyzed via SEC-SAXS. (A) Pair-Distribution Plots. The SEC-SAXS data was processed into a merged file in Scatter. GNOM was used to generate the real space data plot. (B) Low resolution *ab initio* models. DAMMIN was used to generate 50 models of each variant which were subsequently averaged in DAMAVER.

The wt and variant *rpsA* TIRs were *in vitro* transcribed and purified (Figure S3.2), for SEC-SAXS. The resulting pair-distribution plots (Figure 3.9A) indicate that the variants share similar elongated conformations but have differences in their overall shape and hydrodynamic properties (Table S3.2). Analysis of the scattering data to determine low-resolution *ab initio* models revealed that the variants each take on a wt-like conformation, with the notable exception of G-44U (Figure 3.9B). This variant has the lowest D_{max} value at 133 Å, suggesting that it is more compact compared to wt.

The truncated variants were constructed to determine directionality of the other variants, but truncating part of the first stem appears to eliminate the native structure

(both variants have larger D_{max} values than wt despite missing part of the sequence), resulting in elongated conformations (Figure 3.9B). This suggests an important role of stem-loop I in maintaining the tertiary structure of the TIR. To gain more insight into the structural dynamics of the *rpsA* TIR variants, SHAPE probing was performed. A plasmid was constructed enabling the *in vitro* transcription of *rpsA* TIRs with 3' and 5' terminal ends to serve as primer binding sites during reverse transcription and sequencing reactions. Following the experiment, the reactivities were normalized and assigned reactivity values: below 0.3 indicated that the nucleotide is constrained; greater than 0.3 and below 0.7 indicates that the nucleotide is likely to be flexible, and greater than 0.7 indicates that the nucleotide is flexible (McGinnis, Duncan, & Weeks, 2009). The results from SHAPE analysis confirm the three-stem-loop secondary structure previously predicted (Boni et al., 2001), which is relatively well conserved among each of the variants, with the exception of the inactive G-44U (Figure 3.10 and Figure S3.5). This variant seems to maintain the most 5' stem-loop but deviates significantly throughout the rest of the structure, particularly in the loop region of the second stem-loop (Figure 3.10). The *ab initio* model of the inactive A-4C variant shows that it assumes a wt-like conformation (Figure 3.9B), despite exhibiting no translation initiation activity *in vivo*. SHAPE reveals that this variant has significant differences in reactivity at some nucleotides, particularly decreased reactivity surrounding the start codon and increased activity within the start codon itself (Figure 3.10). The highly active variants A-9G and C-1U maintain a wt-like conformation as expected, except for stem-loop III of A-9G (Figure 3.10 and Figure S3.5) where it displays increased activity at positions +6 to +10. The C-1U variant exhibits decreased reactivity immediately surrounding the start codon, but a marked increase in activity at position +7.

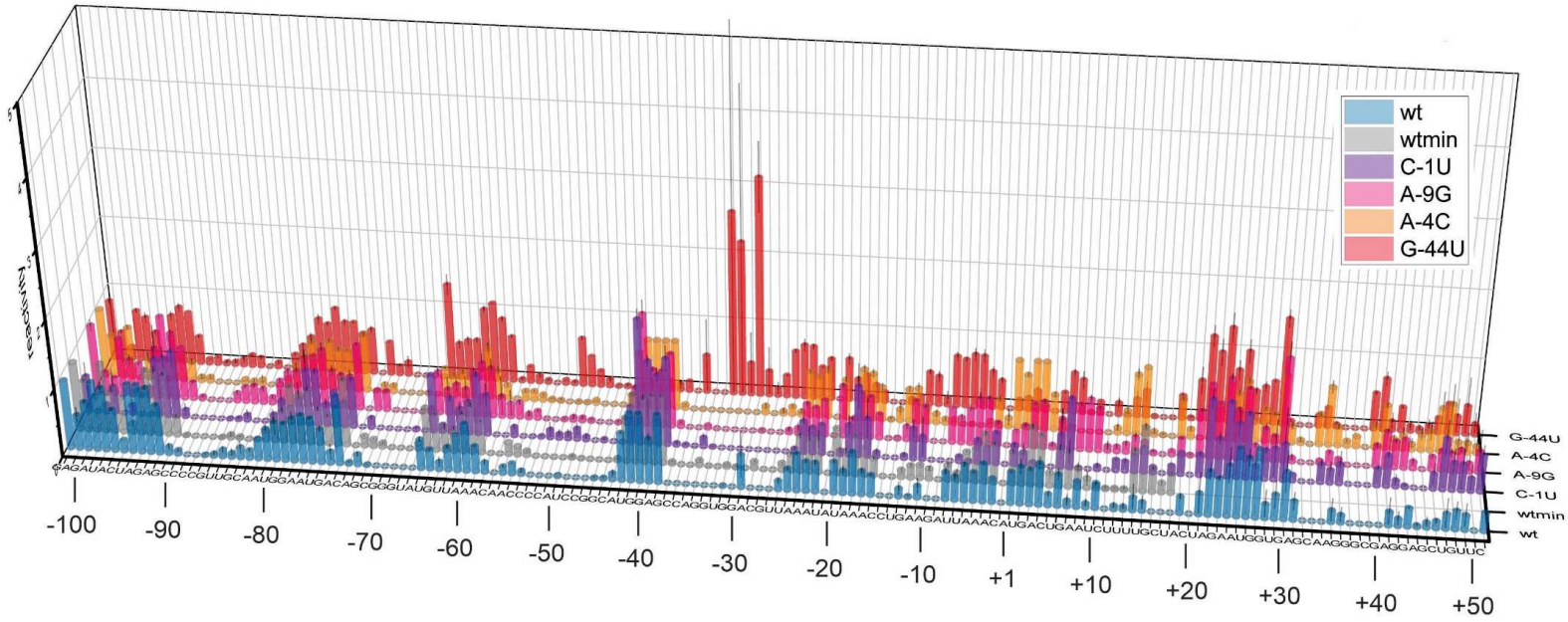
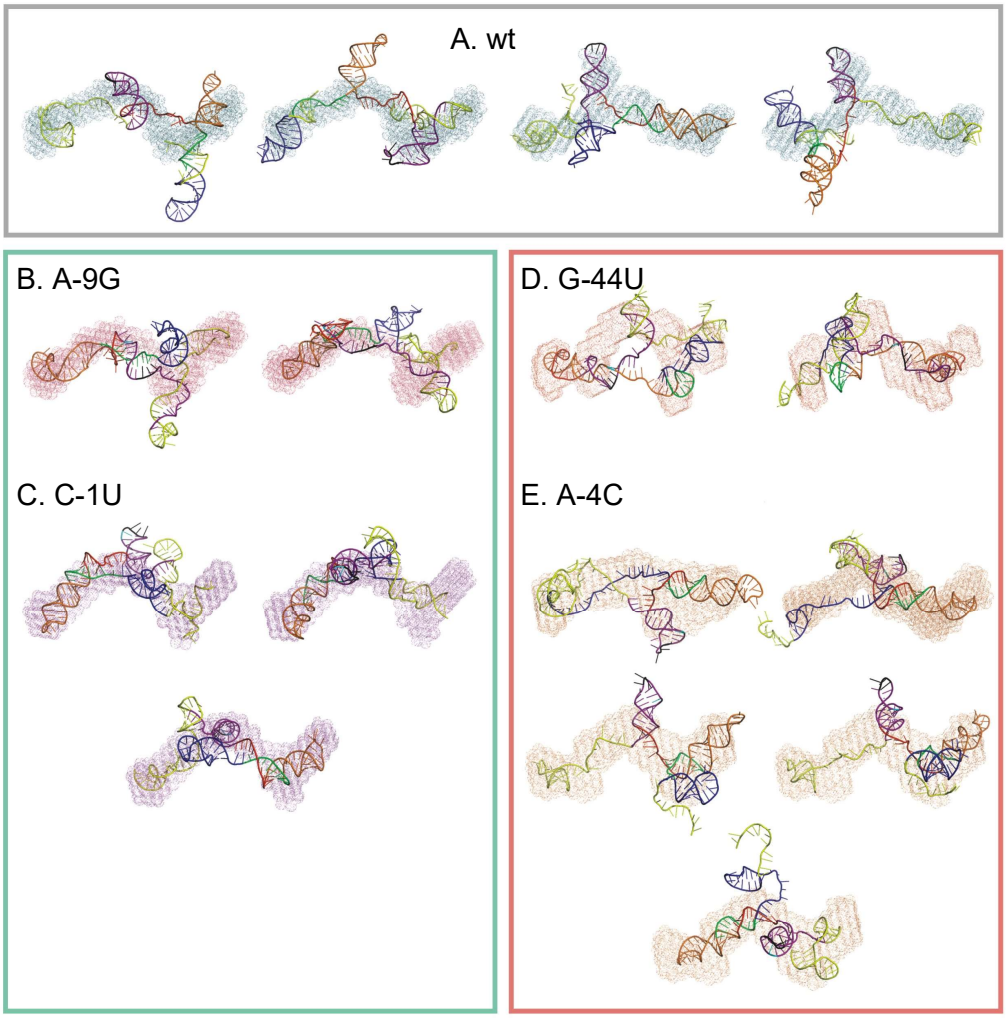


Figure 3.10 SHAPE reactivities of *rpsA* TIR variants. Reactivity values below 0.3 indicate a constrained nucleotide while above 0.7 indicates flexibility of the nucleotide. Nucleotide positions are indicated relative to the start codon at +1



Fig**ure 3.11 Computational models of *rpsA* TIR variants fit into SAXS envelopes.** An ensemble approach was used to generate models of the *rpsA* TIR variants that fit the experimental SAXS and SHAPE data. Models were aligned to SAXS envelopes using SUPCOMB. Each model was fit into multiple clustered SAXS envelopes which were then chosen based on the percent of model inside the envelope. (A) *rpsA* wt: four model ensemble, two SAXS envelopes. (B) *rpsA* A-9G: two model ensemble, one SAXS envelope. (C) *rpsA* C-1U: three model ensemble, one SAXS envelope (D) *rpsA* G-44U: two model ensemble, two SAXS envelopes. (E) *rpsA* A-4C: 5 model ensemble, three SAXS envelopes. Box color indicates activity level: wt = grey, active = green, inactive = red.

Nucleotide reactivities from the SHAPE experiments were used to predict the secondary structures of the variants (Figure S3.5), which were then fed into SimRNA (Boniecki et al., 2016). Multiple models were predicted for each variant, which were then

scored for their fit against the SAXS data, with the best fitting models being chosen as the final ensemble (Figure 3.11 and Figure S3.6).

The *rpsA* TIR is a flexible molecule that adopts different conformations in solution; this is evident when analyzed on a native PAGE as it migrates in multiple bands (Figure S3.2B). Alignment of the atomic models into our SAXS envelopes reveals that flexible portions of the TIR were unable to be resolved through SAXS as they were excluded from the *ab initio* models. The wt clustered envelope (Figure 3.11A, third and fourth models) includes density that may be attributed to stem-loop II, indicating that this clustering method can help to solve different conformations of molecules in solution.

3.3 Discussion

Bacteria use sophisticated regulatory signals to rapidly respond to changes in their environment. These signals are commonly transmitted through RNA regulatory elements, often imbedded in mRNAs, endowing a direct link between translation rate and cellular conditions. Long 5' UTRs with strong RNA secondary structures, and containing multiple regulatory signals are not uncommon in bacteria (B. Chang et al., 2006). However, RNA structure typically acts to attenuate translation by inhibiting ribosome access to the SD. It has been suggested that RNA structure-assisted initiation mechanisms could be a universal strategy used to initiate translation (Colussi et al., 2015). I wanted to determine if any bacterial mRNAs contained three-dimensional architectures with stimulatory effects on translation initiation. This study has provided evidence that the *rpsA* TIR forms critical contacts with the ribosome mediated by its specific three-dimensional shape. This insight culminated from a combined *in vivo*, *in vitro*, and *in silico* experimental approach. First, a library of *rpsA* TIR variants was generated that

enabled the investigation of sequence-function relationships that govern this highly efficient element. The sort-seq procedure then allowed us to isolate phenotypically silent mutations, and key variants for further biophysical interrogation.

Nitrocellulose filter-binding assays were performed to determine the affinities between several *rpsA* variants with different *in vivo* activities and 30S ribosomal subunits or recombinant S1 (Neogy, Chowdhury, & Kerr, 1974; Rio, 2012). An mRNA with a high translation initiation rate could have a higher affinity with S1 or the 30S and *vice versa*. However, if the affinity between a specific TIR and the 30S was significantly high it could prevent the ribosome from sliding along the mRNA preventing the transition into elongation (Komarova et al., 2002). The equilibrium of binding is related to the Gibbs free energy difference between the initial (free RNA) and final states (pre-30S initiation complex-bound). The translation initiation rate should vary with the equilibrium free energy of ΔG unfolding. Surprisingly, there is little difference between the ΔG of binding of the *rpsA* TIR variants, and no correlation between phenotype and binding strength. These variants could have different properties in the kinetics of unfolding the regions surrounding the start codon, while the overall thermodynamics of 30S binding remains unaffected. Additionally, there could also be key nucleotides of the *rpsA* TIR that make physical contacts to the 30S to improve initiation efficiency but do not alter affinity.

Understanding the contributing factors toward an mRNAs initiation rate has led scientists to exert control over synthetic gene circuits by modulating the interactions between the mRNA and 30S (Carrier & Keasling, 1999; Isaacs et al., 2004; Pfleger, Pitera, Smolke, & Keasling, 2006). To streamline this process, biophysical models allowing for more precise forward engineering strategies have been taken. These models have matured into sophisticated RBS Calculators, able to forward and reverse engineer

RBSs with desired initiation rates. Currently, the activity of synthetic mRNAs can be engineered and predicted accurately within a factor of 2.3 over a range of 100,000-fold (Salis et al., 2009). In contrast, our inability to predict the translation rates of some natural mRNAs highlights the complexity of translation initiation and illustrates that out current models do not account for all situations. This is highlighted by the inability to predict the unusually high translation initiation rate of the *rpsA* TIR (Figure 3.4). These algorithms rely on RNA secondary structure predictions, and do not take tertiary structure into consideration. Structural models of these ribosome interacting TIRs will lead to a broadened understanding of their initiation mechanism an enable more accurate prediction of translation initiation rates.

The combined analysis of the *rpsA* TIR has uncovered the principles that govern the structure-based function of the TIR. First, a structural element composed of the first two stem-loops that are required for efficient ribosome recruitment to the *rpsA* TIR was identified. It has been shown previously that a truncation of part or all of stem-loop I abolishes function of this TIR (Boni et al., 2001). Here an inactive variant with a point mutation in stem-loop III that appears to alter the stability of stem-loop I was identified. In the structural ensemble for A-4C, stem-loop I is not always formed (Figure 3.11), as this stem-loop is not restrained in the SHAPE-guided secondary structure prediction (Figure S3.5). This suggests that A-4C exhibits altered dynamics that shift between an intact and disrupted stem-loop I, and may sample the disrupted state often enough to prevent efficient ribosome binding. The second inactive variant studied in detail, G-44U, displays a relatively intact stem-loop I but a disrupted stem-loop II (Figure 3.10 and 3.11). SHAPE probing suggests there are differences in the stem-loop II region of G-44U, as the reactivity of the apical loop nucleotides is severely decreased. The G-44U

structural ensemble confirms that the entire stem-loop is disrupted, and is instead forming interactions with the nucleotides of stem-loop III (Figure 3.11). In fact, stem-loop I appears to be the only element that is maintained in G-44U, as the entire tertiary structure differs significantly from wt, even in our low-resolution *ab initio* models (Figure 3.9). Taken together, the fact that point mutations in the *rpsA* TIR that disrupt stem-loop I or stem-loop II result in loss of translation activity, I propose a model where these stem-loops constitute a structural element that is required for the TIR to optimally interact with the ribosome during initiation complex formation. It is likely that the conserved GGA motifs present in the loops of these stem-loops are important in this interaction as well, although their specific role is yet to be uncovered.

The second structure-based principle that drives the high efficiency of the *rpsA* TIR is the strength of the start codon-containing stem-loop. Stem-loop III is the weakest of the three stems-loops (Boni et al., 2001), presumably because it must be unfolded to allow the start codon to enter ribosomal P-site. Indeed, this stem has been shown to be unfolded when bound to ribosomes (Mustoe et al., 2018). There is a decreased nucleotide reactivity in stem-loop III in the inactive variant A-4C, and increased reactivity of stem-loop III in the highly active variant A-9G (Figure 3.10). Interestingly, the structural ensemble generated for A-9G shows that stem-loop III is not actually formed in this variant, but rather forms weak interactions with other regions of the TIR (Figure 3.11). This suggests that a third stem-loop structure is not absolutely necessary, as long as the start codon is easily accessible. Lastly, the two single-stranded regions of the *rpsA* TIR exhibit dynamic interactions, and can be observed in one of the four wt models, as well as all three of the models in the C-1U ensemble (Figure 3.11). It appears that when these single-stranded regions interact, they elongate stem-loop II.

S1 is known to unwind structured mRNAs prior to their accommodation onto the ribosome (Duval et al., 2013); S1's presence is also required for the translation of its own mRNA (Figure 3.7). In addition to its role in initiating translation, S1 is also involved in its own regulation by recognizing and binding to its own mRNA (Boni et al., 2000). When in excess over ribosomes, S1 is somehow able to specifically recognize the *rpsA* TIR among all other RNAs in the cell (Figure 3.8). However at a certain level of excess, S1 will bind to and repress other messages as well. This is consistent with previous studies that tested initiation complex formation on the *rpsA* and *ssb* mRNAs (an SD-containing mRNA) in the presence of excess S1 (Boni et al., 2001). The inhibitory effect of S1 on other structured mRNAs has not been previously studied. SELEX experiments have reported a high affinity of S1 for pseudoknots (Ringquist et al., 1995), yet S1 does not have an inhibitory effect on *rpsO* at 5-fold excess (Figure 3.8). This could be due to a lack of S1-binding sites, such as the AU-stretches as present in the *rpsA* 5' UTR (Boni et al., 2001). The specific fold of the *rpsA* TIR likely contributes to S1 binding as well – it has been shown that strengthening of stem-loop II relieves autogenous control (Boni et al., 2001). It remains to be seen how S1's interaction with the *rpsA* mRNA differs during translation vs. autoregulation. The concentration of S1 in the cell is approximately 8.4 μM (Wisniewski & Rakus, 2014), but it is unknown how this concentration fluctuates under stress conditions – its autogenous control mechanism may prevent excess S1 concentration from fluctuating significantly enough to affect transcripts other than its own.

Our results suggest that stem-loop III of the *rpsA* TIR plays an important role in start codon accessibility, and that stem-loop I and stem-loop II represent a unique structural element that is absolutely required for the function of the *rpsA* TIR (Figure

3.11). This structural element may provide optimal arrangement of the TIR allowing ribosome recruitment, and an optimally positioning of stem-loop III to be unwound by S1. This element functions as an individual entity, with no requirement for additional cellular factors (Figure 3.6). Future work will require high-resolution structural information on the *rpsA* TIR-ribosome complex in order to identify these important contacts and advance our understanding of how the *rpsA* TIR drives translation initiation in a structure-based manner. This information will allow us to uncover how the structure of the *rpsA* TIR mediates specific nucleotide interactions with the ribosome. Additionally, this effort will provide insights into mechanisms contributing to non-canonical translation initiation in bacteria.

3.4 Methods

3.4.1 Chemicals, Reagents and Oligodeoxyribnucleotides

All chemicals and reagents used in this study were purchased from Sigma-Aldrich, Bio Basic, or Thermo Fischer Scientific unless otherwise specified. Oligonucleotides were purchased from Sigma-Genosys or Integrated DNA Technologies and are listed in Supplementary Table S3.4.

3.4.2 Plasmid Construction and Mutant Library Preparation

The regions between 961 903 bp ↠ 962 011 bp in the *rpsA* gene, and 3 311 674 bp ↞ 3 311 782 bp in the *rpsO* gene of the *E. coli* genome (Riley et al., 2006) were synthesized (GENEWIZ, Inc.) conforming to Biobrick engineering standard RFC_10 (Knight, 2003). Reporter constructs were assembled stepwise from individual Biobrick parts using the Biobrick Assembly Kit (New England Biolabs) and conformed to Biobrick

engineering standard RFC_10 (Knight, 2003) with exception to the junction between the *rpsA* and *rpsO* TIRs and the downstream coding sequence, which conformed to assembly standard 23 (Phillips, 2006). All reporter constructs were ligated into pSB1C3 generating the reporter plasmid outlined in Figure 3.1.

Mutant library plasmids were generated using the GeneMorph II EZClone Domain Mutagenesis Kit (Stratagene). Primers JVO-12273 and JVO-12277, 2.5 ng of template plasmid, and 2.5 units of Mutazyme II DNA polymerase were used to amplify the *rpsA* region of the reporter plasmid in a total volume (vol.) of 50 µL. The polymerase chain reaction (PCR) conditions for the first reaction are as follows: 95 °C for 2 min, and 30 cycles of 30 s at 95 °C, 30 s at 54 °C, and 60 s at 72 °C. The PCR was finalized by 10 min elongation at 72 °C. The PCR product was purified via PCR Purification Kit (QIAGEN). The second reaction (EZClone) contained EZClone enzyme mix, 12.5 ng of the reporter plasmid, 125 ng of the PCR product (megaprimers) from the first reaction, and EZClone solution in a total vol. of 25 µL. The conditions for the second reaction were as follows: 95 °C for 2 min, followed by 25 cycles of 50 s at 95 °C, 50 s at 54 °C, and 8 min at 68 °C. After 5 min on ice, 1 µL of DpnI was added for 2 h at 37 °C to remove the original template plasmid, and then heat inactivated at 98 °C for 5 min.

All site-directed mutagenesis was performed using the QuikChange II site-directed mutagenesis protocol (Agilent Technologies). Mutagenic primers are listed in Supplementary Table S3.4.

3.4.3 Cell Growth, and FACS

LB (Luria Broth) media containing 25 µg/mL chloramphenicol (Cam) was inoculated with an overnight culture of *E. coli* BL21-Gold(DE3) cells (Agilent) (1:100

dilution) and incubated at 37 °C while shaking at 220 rpm. Exponentially growing cells were washed twice with ice-cold ultrapure H_2O. Each pellet resulting from 40 mL of the harvested culture suspension was resuspended in 1.5 mL ice-cold ultrapure H_2O. 100 µL this suspension was mixed with 2 µL 1:10 dilution the EZClone reaction mix. This mixture was transferred to a 0.2 cm electroporation cuvette and 2.5 kV (12.5 kV/cm) was applied resulting in a time constant of ~5.7 ms. Cells were diluted immediately with 1000 µL room temperature SOC media. The cell suspension was incubated for 90 min at 37 °C while shaking at 220 rpm. 250 µL of cell suspension was plated on LB agar plates with 25 µg/mL Cam and incubated overnight at 37 °C.

Colonies on the agar plates were scraped off and pooled. This cell mixture was then used to inoculate cultures of LB media containing 25 µg/mL Cam. These overnight cultures were diluted to an $OD_{600\ nm}$ 0.05 in 20 mL fresh LB media with 25 µg/mL Cam. Cultures were incubated at 37 °C while shaking at 220 rpm. 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to each culture at an $OD_{600\ nm}$ of 0.3 and grown until reaching an $OD_{600\ nm}$ of 1.0. Cells were then harvested by centrifugation at 5,000 *g* for 2 min. The cell pellet was washed in 1 mL ice-cold sterile 1 × PBS before resuspension in 500 µL sterile 1 × PBS. The resulting suspension was further diluted tenfold in 1 × PBS and analyzed using a Becton, Dickinson and Company (BD) FACS Aria III. eCFP was excited at 445 nm and measured with a splitter of 470 nm and band-pass filter of 510/80 nm, while mRFP was excited at 561 nm and measured with a splitter of 655 nm and a band-pass filter of 695/40 nm. A scatter gate was set using the forward and side scatter area of *E. coli* cells harbouring a non-reporter containing control plasmid. The wild type reporter plasmid was used to set the gate for wild type (wt)-like cells, and a previously reported highly active *rpsA* mutant was used to set the gate for the mutations

resulting in high expression (Boni et al., 2001). Two gates corresponding to low eCFP fluorescence were set subjectively. 1×10^6 cells from each population were collected.

Variants to be verified were individually sub-cloned into *E. coli* BL21-Gold(DE3), and fluorescence output was analyzed using a BD FACSAria Fusion using the settings described and growth conditions outlined above. 50,000 events were measured for all mutant validation assays.

3.4.4 Next Generation Sequencing and Bioinformatics Analysis

1×10^6 cells from each population collected during the FACS were mixed with 1×10^7 *E. coli* BL21-Gold(DE3) cells carrying empty pSB1C3. Cell mixtures were centrifuged at 5,000 *g* for 2 min, supernatant was decanted and the pellet was resuspended in 50 µL 0.9 % NaCl. 1.25 µL of this mixture was used in a PCR containing 10 µM dNTPs, 10 × HF Buffer (Thermo Scientific), 2U Phusion polymerase (Thermo Scientific), and a 0.5 nM primer mix containing equimolar amounts of forward primers JVO-12273, JVO-12274, JVO-12275, JVO-12276, and reverse primers JVO-12277, JVO-12278, JVO-12279, JVO-12280. The primer mix adds between 1–4 random nucleotides onto the DNA, this was used as a strategy to increase diversity in the sequencing libraries. The PCR conditions are as follows: 3 min at 98 °C, followed by 25 cycles of 20 s at 98 °C, 20 s at 54 °C, and 1 min at 72 °C, followed by 10 min at 72 °C. PCRs were purified using the QIAGEN PCR clean-up kit and roughly 100 ng of each sample was converted into DNA libraries compatible with next-generation sequencing (NGS) on the MiSeq sequencing platform by Vertis Biotechnologie AG (Freising, Germany).

The resulting DNA libraries were pooled equimolar and subjected to paired-end sequencing (2 × 150 bp). Each library yielded 0.5 to 1 million sequence reads. Read trimming, merging and error correction was performed using SeqPrep. The resulting

sequences were mapped using Bowtie2, and > 99 % of the reads were mapped. A custom script was used to identify and count mutations. Clustering on Z-scores was calculated on individual libraries (Z-score = (x – mean(X))/sd(X) (# SDs above or below mean)).

3.4.5 Purification of 30S, S1, and 30S^{-S1}

Ribosomes and Ribosomal subunits were purified from *E. coli* MRE 600 cells as described in (Rodnina, Fricke, & Wintermeyer, 1994).

A plasmid containing the S1 CDS with a C-terminal 6×-histidine tag, transcriptionally controlled by a T7 promoter was provided as a gift-in-kind from Dr. Stefano Marzi. This plasmid was transformed into competent *E. coli* BL21-Gold(DE3) cells. Alternatively, S1 purified for *in vitro* reconstitution experiments utilized the *rpsA*-ASKA strain, which contains an N-terminal 6×-histidine tag (Kitagawa et al., 2005). To express S1, cells were grown in LB media supplemented with either 100 μg/mL ampicillin (Amp) or 35 μg/mL Cam depending on if the Marzi or ASKA plasmid was used, respectively. At an $OD_{600\ nm}$ of ~0.6 protein expression was induced by the addition of IPTG to a final concentration of 1 mM. Cells were harvested 3 h after induction by centrifugation at 5,000 *g* for 20 min, flash frozen and stored at -80 °C. Cells were thawed on ice and resuspended in 5 mL/g binding buffer (50 mM Tris-HCl pH 8.0, 40 mM NH_4Cl, 7 mM $MgCl_2$, 7 mM β-mercaptoethanol, 300 mM KCl, 10 mM imidazole, 15 % glycerol, and 0.1 mM phenylmethylsulfonyl fluoride (PMSF)). Cells were lysed by adding lysozyme to 1 mg/mL while slowly stirring at 4 °C. Following a 30 min incubation for 30 min 12.5 mg/g (of cells) sodium deoxycholate was added and the mixture incubated for 60 min at 4 °C while slowly stirring. The resulting lysate was centrifuged for 30 min at 30,000 *g* at 4 °C. Cell lysate was incubated with 500 μL 100

μg/mL RNase A for every 100 mL lysate for 2.5 h at 4 °C. S1 was then dissociated from the ribosome by adding NH_4Cl to a final concentration of 1M and centrifuging at 45,000 *g* for 2 h to remove ribosomes. This supernatant was applied to 5 mL binding buffer equilibrated Ni^{2+} Sepharose resin (GE Healthcare) and incubated at 4 °C for 60 min. This incubation was followed by centrifugation for 2 min at 500 g. The resin was then incubated with 40 mL wash buffer (binding buffer with 20 mM imidazole) for 2 min and centrifuged for 2 min at 500 g. The supernatant was decanted and the wash was repeated 3 times. 4 mL of elution buffer (binding buffer containing 250 mM imidazole) was added to the resin and incubated for 5 min on ice. Samples were then centrifuged for 2 min at 500 *g* and the elution step repeated 10 times. The elutions were pooled, concentrated and applied to a Mono-Q GL column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 7.5, 40 mM NH_4Cl, 60 mM KCl, 1 mM dithiothreitol (DTT), and 10 % glycerol to remove any RNA bound to S1 using an ÄKTAprime Plus chromatography system. The protein was eluted from the column using 20 mM Tris-HCl pH 7.5, 1 M NH_4Cl, 60 mM KCl, 1 mM DTT, and 10 % glycerol. Peak fractions were concentrated and rebuffered (20 mM Tris-HCl pH 7.5, 40 mM NH_4Cl, 60 mM KCl, and 1 mM DTT using a spin-column (Vivaspin 30, GE Healthcare), flash frozen and stored at -80 °C. Protein concentration was determined spectroscopically at 280 nm using molar extinction coefficient of 47 565/M•cm (ExPASy – ProtParam tool). The protein was determined to be free of RNA contamination by analysis on a 12 % 8 M urea-PAGE.

To prepare $30S^{-S1}$ subunits, 30S ribosomal subunits were obtained by dissociating purified 70S ribosomes into subunits. Concentration of 30S was determined spectroscopically using the extinction coefficient 1.37×10^7/M•cm. Ribosomes were diluted tenfold in a high-salt dissociation buffer (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 60

mM KCl, 1 M NH_4Cl, and 1 mM DTT). The mixture was incubated at 37 °C for 10 min to dissociate S1 from the ribosomes. To remove S1 from the solution, poly(U) affinity batch chromatography was used (Duval et al., 2013; Phillips, Pang, Park, Hollis, & Famuyiwa, 1980; Subramanian, Rienhardt, Kimura, & Suryanarayana, 1981). Polyuridylic acid-agarose lyophilized powder (Sigma Aldrich) was hydrated using ultrapure H_2O, and subsequently equilibrated in dissociation buffer. The prepared 30S mixture was added to the resin and incubated at 4 °C for 1 h with gentle inversion. The mixture was centrifuged at 500 *g* for 5 min, and the supernatant containing the $30S^{-S1}$ ribosomes was collected. Three additional washes with dissociation buffer were performed to remove any residual ribosomes from the resin. The $30S^{-S1}$ ribosomes were rebuffered and stored in $TAKM_5$ buffer (50 mM Tris-HCl pH 7.6, 70 mM NH_4Cl, 30 mM KCl, 5 mM $MgCl_2$) via ultracentrifugation with a Sorvall S55-S swinging-bucket rotor ultracentrifuge (Thermo Scientific) at 55,000 rpm, at 4 °C for 24 h. The supernatant was removed and the pellet was resuspended in $TAKM_5$ to a concentration of ~15 μM. The removal of S1 was confirmed via 15 % SDS-PAGE gel electrophoresis and mass spectrometry (University of Lethbridge Mass Spectrometry facility).

3.4.6 Purification of T7 RNA Polymerase

A plasmid containing the gene encoding T7 RNA polymerase (RNAP), transcriptionally controlled by a T7 promoter was provided as a gift-in-kind from Dr. Ute Kothe. This plasmid was transformed into *E. coli* BL21-Gold(DE3) cells. To express T7 RNAP cells were grown in LB media supplemented with 100 μg/mL Amp. At an $OD_{600\ nm}$ of 0.6, protein expression was induced by adding IPTG to a final concentration of 1 mM. Cells were harvested 3 h after induction by centrifugation at 5,000 *g* at 4 °C for 15

min, flash frozen and stored at -80 °C. Cells were thawed on ice and resuspended in 5 mL binding buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF, and 5 % glycerol) per g of cells. Cells were lysed for 30 min by adding lysozyme to a final concentration of 1 mg/mL, while stirring on ice. 12.5 mg sodium deoxycholate per gram of cells was added and incubated for 30 min at 4 °C while slowly stirring. A Branson Sonifer 450 sonicator was used to aid in further cell opening; on ice the cells were pulsed 1 min at intensity level 6, and duty cycle 60 %, 5 times with short pauses in between each pulse. The opened cells were centrifuged for 45 min at 30,000 *g* at 4 °C.

The cleared S30 lysate was applied to Ni^{2+} Sepharose (GE Healthcare) resin equilibrated with binding buffer, and incubated at 4 °C for 1 h, inverting periodically to bind protein to the resin. This incubation was followed by centrifugation for 2 min at 500 g. The resin was then incubated with 40 mL wash buffer (binding buffer with 30 mM imidazole) for 2 min and centrifuged for 2 min at 500 *g* at 4 °C. The supernatant was decanted and the wash was repeated 3 times. 4 mL of elution buffer (binding buffer containing 500 mM imidazole) was added to the resin and incubated for 5 min on ice. Samples were then centrifuged for 2 min at 500 *g* and the elution step repeated 10 times. The elutions were pooled and applied to ~5 mL of Affi-Gel Blue Gel (Bio-Rad Laboratories) resin equilibrated with 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF, and 5 % glycerol. The resin was washed twice with 50 mL of a high-salt buffer (50 mM Tris pH 8.0, 1.0 M NaCl, 5 mM β-mercaptoethanol, and 5 % glycerol) and centrifuged 500 *g* for 5 min at 4 °C. The supernatant was removed, and the resin was washed three times with 50 mL of a low-salt Buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF, and 5 % glycerol). The elutions were pooled to a 1:1 (v/v) ratio of Affi-Gel Blue Gel (Bio-Rad Laboratories)

resin by incubating for 1 h at 4 °C while gently shaking. Samples were centrifuged at 500 *g* for 5 min at 4 °C and the supernatant containing T7 RNAP was retained. The absorbance at 280 nm was measured to determine the concentration of T7 RNAP using the extinction coefficient of 140 260 $M^{-1}cm^{-1}$. The sample was concentrated by centrifugation in a spin-column (Vivaspin 30, GE Healthcare) at 4,000 *g* at 4 °C. The purified protein was rebuffered in storage buffer (40 mM K_2HPO_4 pH 7.5, 2 mM DTT, and 2 mM EDTA) and then mixed with 1 vol. of 100 % glycerol.

3.4.7 RNA Preparation and Purification

See Supplementary Table S3.3 for plasmids encoding pT7 controlled synthetic *E. coli* mRNAs and Table S3.5 for RNA sequences. The promoter and DNA sequence encoding the RNA of interest was amplified from the plasmid via PCR. The region encoding the *rpsA*-eCFP mRNA was amplified using primers JVO-059 and JVO-060; the region encoding the *rpsA*-mRFP mRNA was amplified using JVO-060 and JVO-163; the region encoding the minimal *rpsA* mRNA was amplified using JVO-060 and JVO-075; the region encoding the SHAPE RNAs were amplified using JVO-155 and JVO-156. 0.5 µM of these primers were used to amplify the area of interest in a PCR containing 0.6 ng/µL of template plasmid, 10 mM dNTPs, 1 × HF Buffer (Thermo Scientific), and 0.02 U/µL Phusion polymerase (Thermo Scientific). The polymerase chain reaction (PCR) conditions were: 95 °C for 2 min, 30 cycles of 30 s at 95 °C, 30 s at 54 °C, and 60 s at 72 °C. The reaction was finalized by 10 min at 72 °C. Upon completion the reaction was mixed with 10 U DpnI (Thermo Scientific) and analyzed on 2 % agarose gel stained with ethidium bromide.

Large scale *in vitro* transcription reactions for subsequent SAXS and SHAPE analysis were performed using ~60 µg/mL of template DNA in 40 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, with 2.5 mM ATP, CTP, UTP and GTP each, 5 mM GMP, 0.01 U/µL inorganic pyrophosphatase, and recombinantly purified T7 RNAP for 4 h at 37 °C. GMP was added to prime the reaction, as it is more efficiently incorporated as the initiating nucleotide when compared to GTP (Martin & Coleman, 1989). The DNA template was subsequently digested with 2 U/mL DNase I (Thermo Scientific) for 3 h at 37 °C. A phenol chloroform extraction was performed on the *in vitro* transcription reaction by mixing equal parts Tris-saturated phenol and chloroform. This mixture was vigorously vortexed for 30 s, and centrifuged at 5,000 *g* for 10 min at 4 °C. The aqueous layer was removed and mixed with an equal vol. of chloroform and again vigorously vortexed for 30 s. The mixture was centrifuged at 5,000 *g* for 10 min at 4 °C and the aqueous layer removed; this chloroform wash step was repeated. The RNA was then precipitated by adding 0.7 vols. of isopropanol and 1/10 vols. of 3 M NaOAc, followed by a 60 min incubation at -80 °C. The mixture was then centrifuged at 5,000 *g* for 30 min at 4 °C. The supernatant was decanted. The resulting pellet was washed in ice-cold ethanol, centrifuged at 5,000 *g* for 10 min and the supernatant decanted. The pellet was resuspended in the minimum amount of ultrapure H_2O (ranging from 300 µL to 2 mL depending on sample). The RNA containing solution was loaded with a flow rate of 0.4 mL/min onto a Superdex 200 GL (GE Healthcare) size exclusion column equilibrated with 10 mM Tris-HCl pH 7.5, 100 NaCl, 5 mM $MgCl_2$. The peak fractions were collected, pooled and analyzed for purity on both 15 % 8 M urea and native PAGEs.

Small scale *in vitro* transcription reactions were used to generate RNA for the filter binding assays. This was done the same way as for the large scale reactions, with the difference that the RNA was purified using the EZ-10 Spin Column RNA Cleanup and Concentration Kit (Bio Basic) following the manufacturers guidelines, eluting RNA in ultrapure H_2O.

3.4.8 RT-qPCR

Cells harboring the *rpsA* reporter plasmids were grown in triplicate. 5 mL LB media containing 25 mg/mL Cam was inoculated with an overnight culture (1:100 dilution) and grown at 37 °C, 200 rpm until the $OD_{600\ nm}$ reached 0.3. mRNA production was induced by adding IPTG to 1 mM, and growing for an additional 2 h. The cell cultures were then mixed with 625 µL of stop solution (5 % phenol, 95 % ethanol), and harvested via centrifugation at 5,000 *g* for 15 min, and the pellets shock-frozen in $N_{2(l)}$ and stored at -80 °C for further use.

The frozen cell pellets were thawed on ice, and total RNA was purified using the EZ-10 Spin Column Total RNA Miniprep kit (Bio Basic) following the manufacturer's instructions. The total RNA was eluted in 50 µL ultrapure H_2O, and quantified via BioDrop. The total RNA was treated with 0.1 U/µL DNase I (Thermo Scientific) overnight at 37 °C and purified using the EZ-10 Spin Column RNA Cleanup and Concentration Kit (Bio Basic). The bound RNA was eluted in 50 µL ultrapure H_2O and the integrity analyzed via 1 % MOPS/formaldehyde denaturing agarose gel.

Primers were designed to specifically amplify portions of the eCFP or mRFP CDS. Three sets of primers for both eCFP and mRFP were designed using IDT PrimerQuest® with the nucleotide sequences for eCFP (BBa_E0020) and mRFP

(BBa_E1010) as inputs. Primers were designed to amplify products between 90 and 120 bp, have a GC content of 50–60 %, and a Tm of 50–65 °C. All sets of primers for each CDS were tested; ultimately primer sets JHO-005 and JHO-006, and JHO-007 and JHO-008 for eCFP and mRFP respectively were chosen for further experiments.

Reverse transcription (RT) reactions were carried out using the qScript Flex cDNA kit (Quanta BioSciences). Each 20 µL reaction contained 75 ng total RNA, 1 µM reverse primer, 1 × GSP enhancer, 1 × qScript flex reaction mix, and reverse transcriptase. The mixture was incubated at 65 °C for 5 mins, and then 42 °C before adding the reaction buffer and 1 µL reverse transcriptase. The reaction was then incubated at 42 °C for 45 min, and then 85 °C for 5 min to inactivate the reverse transcriptase. For both mRFP and eCFP reactions minus-RT controls were carried out for each replicate of the wt reporter plasmid (JVS-031) containing cells.

Quantitative PCR (qPCR) was performed using the PerfeCTa SYBR Green SuperMix ROX kit (Quanta Biosciences). Each 10 µL reaction contained, 1 × PerfeCTa SYBR supermix ROX, 0.5 µM forward primer, 0.5 µM reverse primer, and 2.5 µL cDNA generated from 75 ng total RNA. Standard curves were performed for each set of primers (eCFP, and mRFP) using the JVS-031 containing cells, using four ten-fold serial dilutions. Cycling conditions were as follows: 95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 60 °C for 1 min, followed by a melt-curve cycle of 95 °C for 15 s, 60 °C for 1 min, then increasing by 3 °C every 15 s until 95 °C. The qPCR reactions were carried out on a StepOnePlus RT PCR system (Thermo Scientific).

Results were analyzed using the standard curve method of comparative quantification. Ct values for the target gene (eCFP) samples were compared between the wt and variants, and normalized to the reference gene (mRFP) values. Variations in

amplification efficiency between eCFP and mRFP primers were adjusted for using the efficiency values derived from the standard curves ($E = 10^{-1/slope} - 1$). For statistical analysis, T-tests were carried out in Microsoft Excel 2016 using a two-tailed test and two sample equal variance.

Fold Difference Calculation, wt vs. G-44U:

$$\Delta CteCFP = Ct(wt) - Ct(G44U)$$

$$\Delta CtmRFP = Ct(wt) - Ct(G44U)$$

$$\text{Fold Difference} = EeCFP^{\Delta CteCFP} / EmRFP^{\Delta CtmRFP}$$

Where ΔCteCFP is the difference in threshold cycle (Ct) between the wt-eCFP and G44U-eCFP samples and ΔCtmRFP is the difference in threshold cycle between the wt-mRFP and G44U-mRFP samples.

3.4.9 Nitrocellulose Filter Binding

Protein/ribosomes and RNA were incubated at 37 °C for 10 min to allow the RNA to bind the protein in buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$). 50 nM of RNA was titrated against 12.5 nM – 2.0 μM 30S or 20.0 nM – 10.0 nM S1. The entire reaction (50 μL) was applied to a nitrocellulose filter (0.2 μm, 25 mm diameter, GE Healthcare), which was then washed with 1 mL ice-cold reaction buffer. Filters were dissolved in 10 mL EcoLite scintillation cocktail (EcoLite (+), MP Biomedical). The amount of RNA retained on the filter via protein binding was quantified by scintillation counting (Perkin Elmer Tri Carb 2800TR liquid scintillation analyzer).

Dissociation constants were calculated by quantifying the bound RNA as protein was titrated using the following equation:

$$y = Bmax * x / (K_d + x)$$

Bmax is the amplitude, or final level of bound S1/30S. The K_d and standard deviation were determined by fitting in GraphPad Prism version 5.02 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

3.4.10 Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE)

Two pmol of RNA in 6 µL of 5 mM Tris-HCl pH 8 and 1mM EDTA was heated at 95 °C for 2 min, and cooled for 2 min on ice. 3 µL of 3.3 × folding buffer (333 mM HEPES pH 8.0, 333 mM NaCl, 66.6–99 mM $MgCl_2$) was then added to the RNA and incubated at 37 °C for 15 min. The folded RNA was treated with 1 µL 30–130 mM NMIA (dissolved in anhydrous DMSO), depending on the length of RNA, and incubated at 37 °C for 45 min (Wilkinson et al., 2006). Control reactions were performed using the same method, however 1 µL DMSO was added instead of NMIA. The RNA was subsequently precipitated with ethanol, and resuspended in 5 mM Tris-HCl pH 8 and 1 mM EDTA containing 5 pmol of a VIC-labelled primer to a total vol. of 13 µL. Primers were annealed by heating at 65 °C for 5 min, and at 37 °C for 1 min. Extension buffer (5 mM DTT, 0.5 mM each dNTP, 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, and 100 units of SuperScript III Reverse Transcriptase (Invitrogen)) was added, and the reactions mixture incubated for 5 min at 37 °C, 20 min at 52 °C, and 5 min at 60 °C. Reactions were precipitated with 2 µL of 5 M NaCl and ethanol. Dideoxy sequencing

ladders were generated by primer extension, using unmodified RNA and primer labelled with NED in the presence of ddCTP (Duncan and Weeks, 2008). NMIA, control, and sequencing reactions were recovered by ethanol precipitation; the pellets were dried and resuspended in 6 μL deionized formamide and combined.

Fluorescently labelled DNA was resolved on an Applied Biosystems 3130 capillary electrophoresis instrument. Raw capillary electrophoresis traces were processed using QuSHAPE software (Karabiber, McGinnis, Favorov, & Weeks, 2013). Integrated intensities were normalized by dividing the data set by the average of the 8 % most reactive nucleotides, after first excluding the top 2 % reactivities. With this normalization, the mean cleavage intensity of the most highly reactive nucleotides becomes 1.0.

3.4.11 Small Angle X-ray Scattering (SAXS)

SEC-SAXS data was collected at beamline 21, Diamond Light Source (Didcot, UK). 50 μL of the RNA samples (2 mg/mL) were transferred to a 96-well plate and placed in the beamline HPLC robotics chamber at 4 °C. 40 μL of the samples were injected onto a Shodex KW402.5-4F column equilibrated with 1 × RNA SAXS buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$) at 0.160 mL/min. 40 μL of 10 mg/mL BSA in RNA buffer was injected as an experimental control before all other samples. The eluent from the column was analyzed using SAXS. The beamline operated at a camera length of 4.014 m with a wavelength of 1 Å, and was configured to measure a scattering (s) range of 0.0032 to 0.38 Å−1 ($s = 4\pi\sin\theta/\lambda$, where θ is the scattering angle). The scattering data was processed as described in (Mrozowich, McLennan, Overduin, & Patel, 2018) with some modifications. Initial processing was carried out in Scatter version 3.0 (Förster, Apostol, & Bras, 2010) to perform background subtraction and merging of

the data under the SEC peak. Guinier analysis was performed in PRIMUS (Konarev, 2003) to obtain radius of gyration (r_G) and forward scattering I(0) values. Further processing in GNOM (D.I. Svergun, 1992) generated pair-distribution function information as well as real-space r_G and maximum particle dimension (D_{max}) values. The resulting output files were input into DAMMIN (D. I. Svergun, 1999) for *ab initio* modelling. Simulations were performed in slow-mode with no symmetry enforced (P1). Fifty models were predicted for each molecule, and were either averaged using the program DAMAVER (Volkov & I. Svergun, 2003) or clustered using DAMCLUST (Petoukhov et al., 2012). DAMCLUST was performed for the *rpsA* TIR variants (excluding truncation variants) to investigate possible alternate conformations. The output of this analysis is a representative model for each cluster. Clusters that were represented by greater than 5/50 members (Figure S3.4) were chosen and used these for further analysis.

3.4.12 Generation of Computational Models

From SHAPE analysis, many of the nucleotides scored between 0.3 and 0.7, suggesting that the *rpsA* TIR wt and variants are quite flexible and disordered (Figure 3.11). Due to this disordered nature, an ensemble of models rather than a single averaged structure was generated to display multiple conformations that each satisfy the experimental data. To generate initial models, secondary structures were predicted from SHAPE reactivities using RNAStructure (Reuter & Mathews, 2010) (Figure S3.5). These secondary structures were then used as inputs for modelling in SimRNA (Boniecki et al., 2016). The trajectories from SimRNA were converted into all-atom models (decoys), on which CRYSOL was run to calculate chi-square values against the SAXS curves. These

decoys were filtered against a minimal secondary structure to determine which models were to be used in further steps. The decoys with lowest chi-square values were further filtered for the best SimRNA energy. GAJOE was used to pick ensembles, and the selected models were optimized in QRNAS (Stasiewicz, Mukherjee, Nithin, & Bujnicki, 2019). These atomic models (Figure S3.6) were aligned to their respective SAXS envelopes using SUPCOMB. Clustered envelopes were used for final figures. In cases where a variant has multiple clusters, the cluster was chosen for each model based on the percentage of model that fit inside the envelope. Figures were generated using PyMol 2.1.1, Schrödinger, Inc.

3.4.13 Activity of the *rpsA* TIR *In Vitro*

Experiments involving the PURExpress *In vitro* Protein Synthesis kit (New England Biolabs) were performed according to the manufacturer's guidelines with some alterations. Reactions were seeded with the wt reporter vector and six variants. Reactions included 10 µL Solution A, 7.5 µL Solution B, 0.5 µL RiboLock (Thermo Fisher), 4.5 µL ultrapure H_2O, and 4.0 nM of plasmid DNA template in a reaction vol. of 25 µL. The reactions were incubated at 37 °C for 16 h; this incubation is longer than the suggested 2–4 h which ensured complete maturation of the fluorescent reporter proteins. A negative control with no DNA added was included with each set of experiments.

After 16 h the reactions were placed on ice, and then diluted in 125 µL of cold $TAKM_7$ buffer (50 mM Tris-HCl pH 7.6, 70 mM NH_4Cl, 30 mM KCl, 7 mM $MgCl_2$) for analysis by fluorescence spectroscopy. A QuantaMaster Fluorimeter (Photon Technology International (Canada) Inc) was used to analyze the production of fluorescent proteins in each sample using the following parameters: eCFP was excited at 439 nm and emission

scanned between 454–554 nm (λ_{max} = 476 nm); mRFP was excited at 584 nm and emission scanned between 599–699 nm (λ_{max} = 607 nm). Emission scans were recorded for each sample and the negative control, which was subtracted from each spectrum.

The PURExpress ΔRibosome *In vitro* Protein Synthesis kit was purchased from New England Biolabs and experiments were performed according to the manufacturer's guidelines with some alterations. Single reporter plasmids were used in this assay (JVS-113, JVS-166, JVS-169). Reactions consisted of 10 µL Solution A, 3 µL Factor Mix, 0.5 µL RiboLock (Thermo Scientific), 5.2 nM template DNA and ultrapure H_2O up to a reaction vol. of 25 µL. The reactions were supplemented with 60 pmol of purified 30S or $30S^{-S1}$ subunits and 60 pmol of purified 50S subunits. The reactions were incubated at 37 °C for 16 h and eCFP expression was analyzed via fluorescence spectroscopy as described above. To test whether translation activity could be rescued, reactions were supplemented with a 1:1 ratio of S1 to ribosomes.

3.5 Supplemental Information

3.5.1 Supplemental Figures

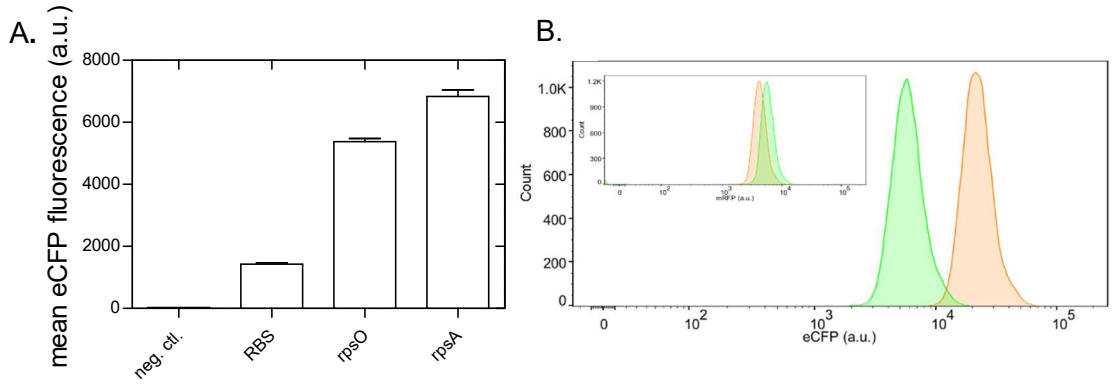


Figure S3.1 Differential protein expression can resolved using our dual reporter plasmid. (A) eCFP expression driven translationally by various TIRs. Fluorescence was measured using flow-cytometry and the geometric mean plotted. Three biological replicates were analyzed. (B) The dual reporter plasmid containing the wt *rpsA* TIR (green trace) and a high expressing variant (orange trace) were analyzed using flow cytometry. The mRFP expression from the noise module is in the insert.

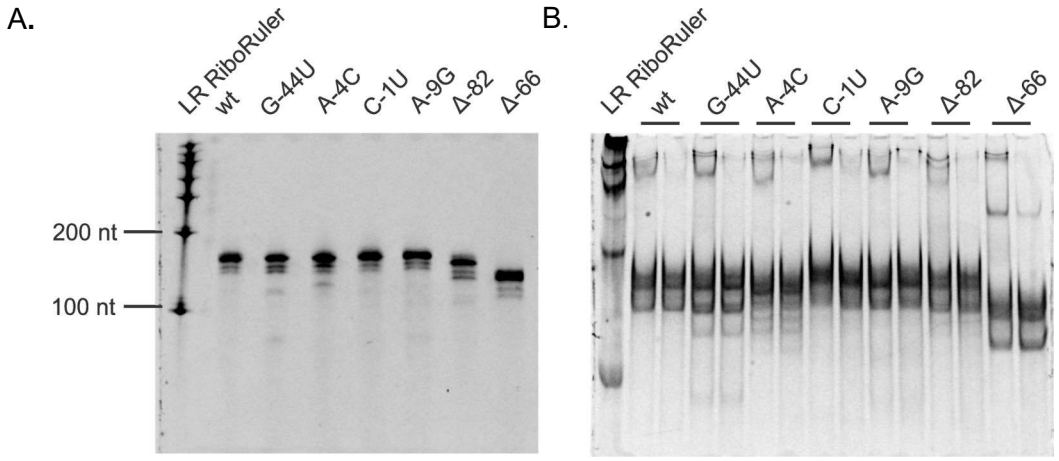


Figure S3.2 *In vitro* transcribed *rpsA* TIR variants. (A) RNA was *in vitro* transcribed using T7 RNA polymerase overnight at 37 °C and purified via spin column. A 200 ng sample of RNA was loaded onto an 8 % Urea PAGE ran for 25 min at 200 V, stained with ethidium bromide. (B) *rpsA* TIR variants before and after folding. The purified RNA was folded by heating at 95 °C and cooling to room temperature. A 200 ng sample of RNA before and after folding was loaded onto an 8 % Native PAGE and ran for 1 h at 150 V, stained with ethidium bromide.

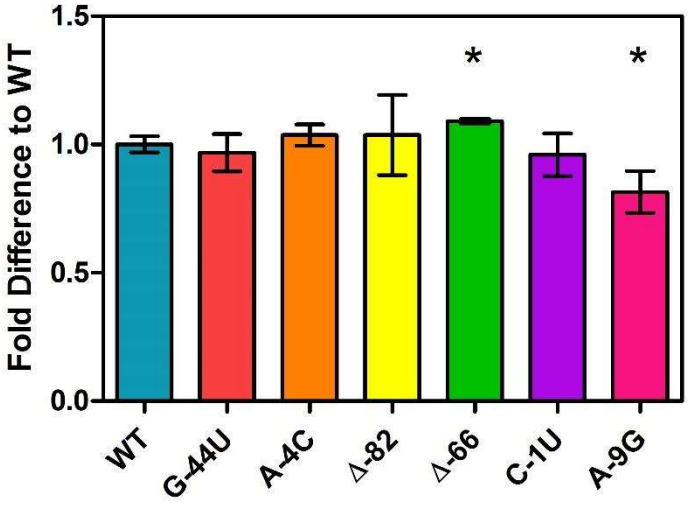


Figure S3.3 Relative abundance of *rpsA* TIR variants compared to wt as determined by RT-qPCR. Fold difference was calculated with the standard curve method, using the mRFP signal as an internal reference for normalization. * Indicates significant difference compared to wt, p-value < 0.05.

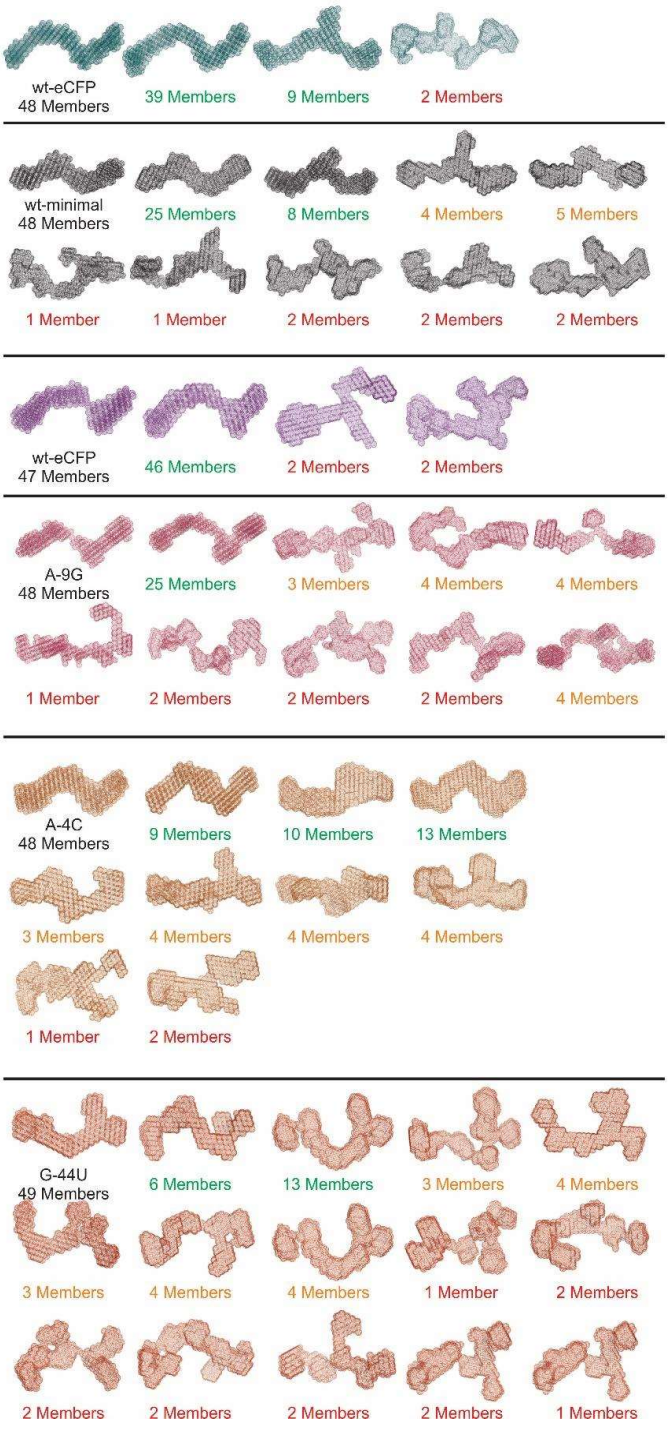


Figure S3.4 Clustering of *rpsA* SAXS envelopes. 50 models for each *rpsA* variant were averaged using DAMCLUST. Clusters with more than 5/50 members were chosen as alternative solutions to the scattering data. The DAMAVER averaged model for each variant is shown first.

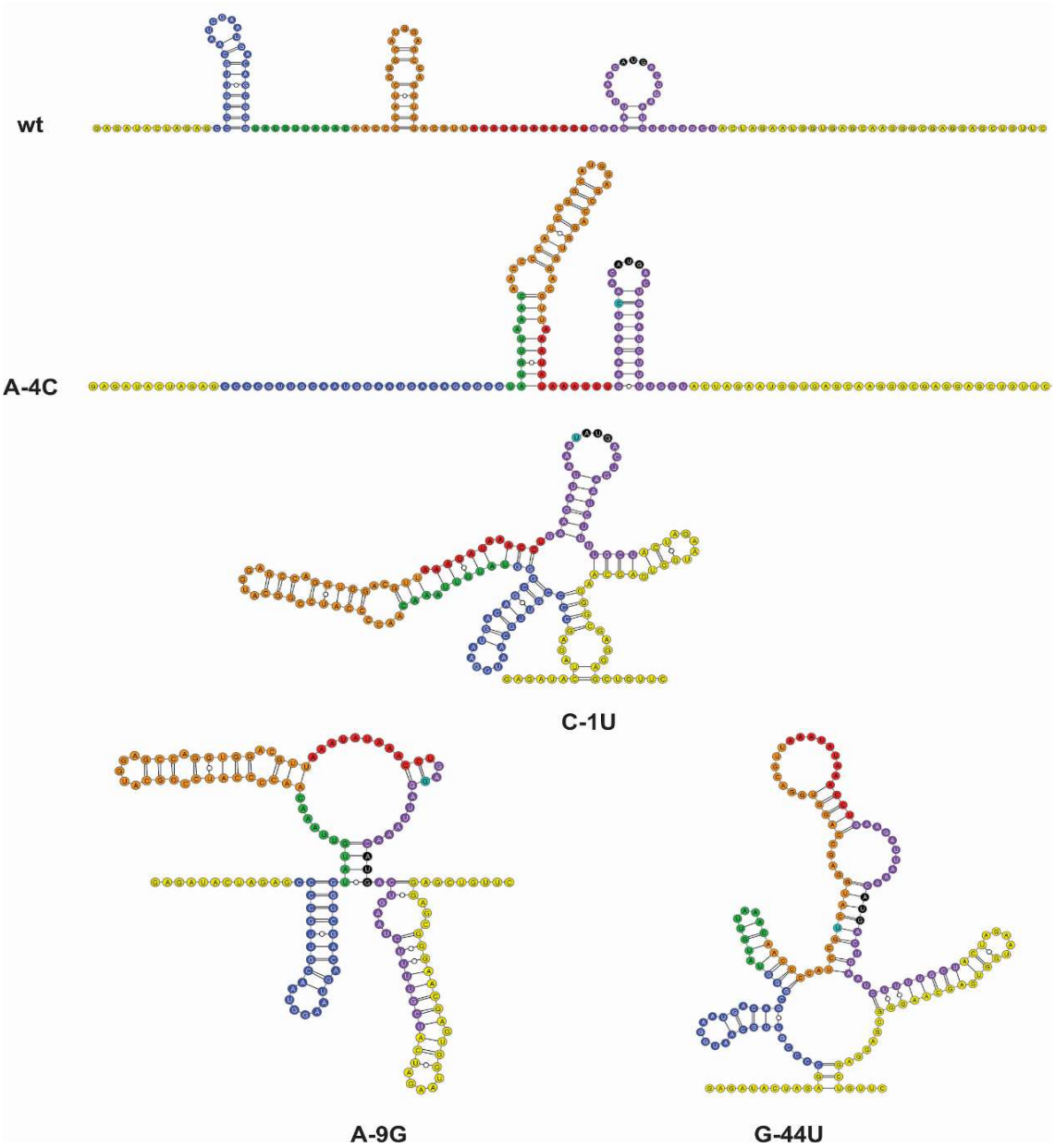


Figure S3.5 Secondary structures predicted using SHAPE reactivities. Secondary structures were predicted using RNAStructure and used as restraints for SimRNA modelling. SimRNA enforces the base pairing indicated, but does not prevent new base-pairs being formed. Figures were made in VARNA. Stem-loops and single-stranded regions are colored according to the wt. Yellow indicates scar sequence.

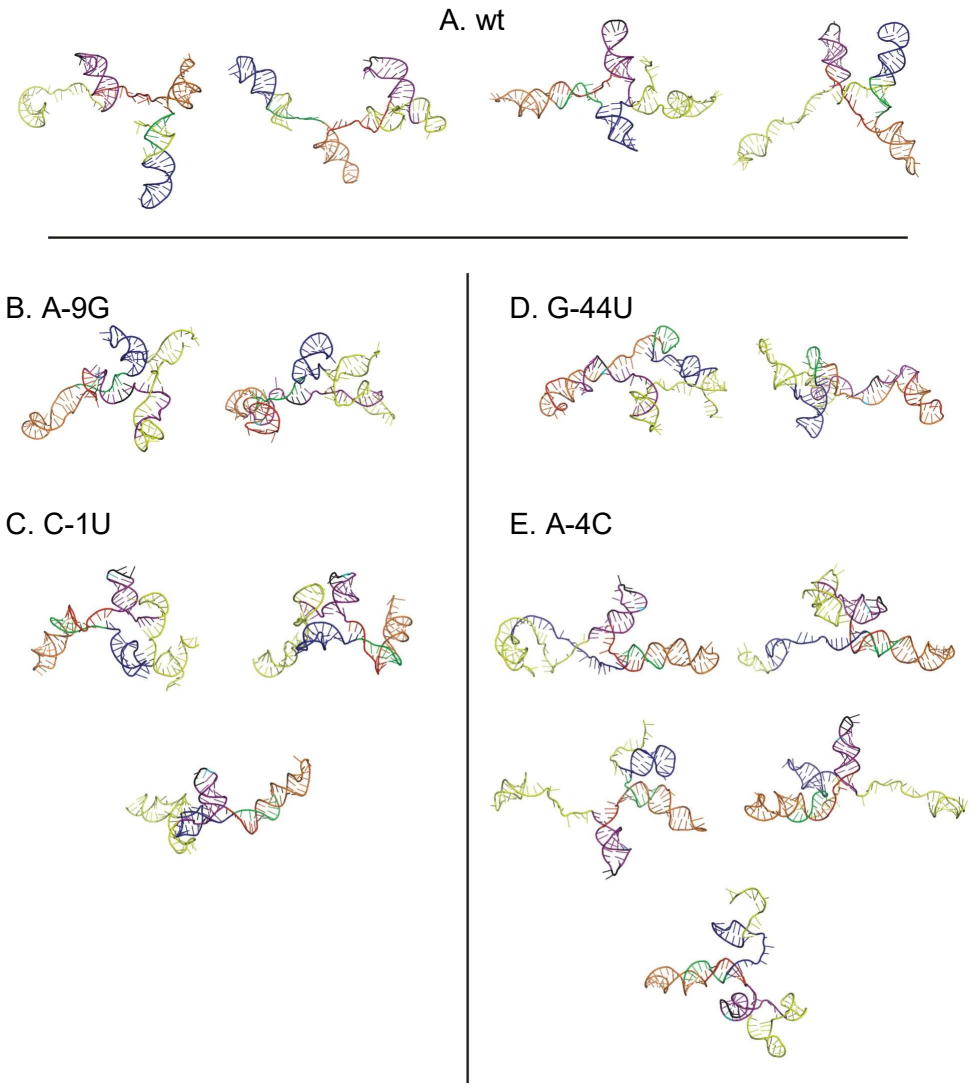


Figure S3.6 Computational models of *rpsA* TIR variants. An ensemble approach was used to generate models of the *rpsA* TIR variants that fit the experimental SAXS and SHAPE data. (A) *rpsA* wt: four model ensemble. (B) *rpsA* A-9G: two model ensemble. (C) *rpsA* C-1U: three model ensemble. (D) *rpsA* G-44U: two model ensemble. (E) *rpsA* A-4C: five model ensemble. Models are colored according to scheme in Figure 1.1, with scar sequences in yellow and point mutations in teal.

3.5.2 Supplemental Tables

Table S3.1. Binding affinities for the interaction between *rpsA* TIR variants and 30S/S1. RNA was incubated with ribosomes/protein for 10 min at 37 °C and binding was analyzed via nitrocellulose filtrations. K_ds were calculated in GraphPad Prism.

rpsA TIR Variant	30S K_d (nM)	S1 K_d (μM)
wildtype	100 ± 10	2.7 ± 0.3
G-78U	150 ± 14	----------
A-9G	56 ± 8.0	1.6 ± 0.1
G-37U	65 ± 14	----------
U-63G	42 ± 10	1.3 ± 0.2
G-44U	120 ± 12	1.8 ± 0.2
C-1U	42 ± 5.0	0.27 ± 0.02
A-4C	35 ± 2.0	0.76 ± 0.09

Table S3.2 *rpsA* TIR variants SEC-SAXS parameters. SAXS analysis was performed at Diamond Light Source Synchrotron. r_G= radius of gyration. D_{max} = maximum dimension. I(0)= forward scattering intensity. NSD= normalized spatial discrepancy.

		wt-eCFP	wt-min	A-9G	C-1U	A-4C	G-44U	Δ 66	Δ 82
Data Collection		HPLC-SAXS	HPLC-SAXS	HPLC-SAXS	HPLC-SAXS	HPLC-SAXS	HPLC-SAXS	HPLC-SAXS	HPLC-SAXS
Guinier	r_G (Å)	49.57	45.86	52.89	56.25	44.02	45.00	42.58	51.59
	error	3.36	3.42	2.45	0.82	3.80	7.06	1.88	5.10
	q. r_G	0.38–1.29	0.32–1.29	0.36–1.29	0.33–1.29	0.33–1.29	0.37–1.28	0.51–1.29	0.35–1.30
	I(0)	0.0077	0.0260	0.021	0.022	0.0062	0.0053	0.0052	0.0019
	error I(0)	5.2×10^{-5}	9.0×10^{-5}	1.7×10^{-4}	1.2×10^{-4}	2.3×10^{-5}	5.0×10^{-5}	3.1×10^{-5}	2.6×10^{-5}
	Points used	14–81	10–105	7–87	1–76	14–93	16–90	9–76	11–78
GNOM	r_G (Å)	49.82	46.84	51.89	49.04	46.31	44.86	46.33	50.78
	error r_G	0.3695	0.1370	0.2438	0.1077	0.2057	0.2719	0.3074	0.4459
	I(0)	0.7668×10^{-2}	0.2558×10^{-1}	0.2009×10^{-1}	0.1871×10^{-1}	0.6924×10^{-1}	0.5236×10^{-2}	0.5340×10^{-2}	0.1876×10^{-2}
	error I(0)	0.5063×10^{-4}	0.7242×10^{-4}	0.1076×10^{-3}	0.7232×10^{-4}	0.3623×10^{-4}	0.4436×10^{-4}	0.2378×10^{-4}	0.1899×10^{-4}
	D_{max}	155	148	158	149	146	133	160	160
	Points used	23–814	20–1300	7–1500	58–967	22–1115	37–721	22–1100	20–900
Ab initio modelling		DAMMIN	DAMMIN	DAMMIN	DAMMIN	DAMMIN	DAMMIN	DAMMIN	DAMMIN
	Models calculated	50	50	50	50	50	50	50	50
	X^2	1.211	1.530	1.001	0.849	1.319	1.089	1.317	1.337
	NSD	0.974	0.909	1.050	1.063	0.903	1.062	0.826	0.971
	error NSD	0.038	0.030	0.029	0.028	0.022	0.036	0.027	0.025
M_w (kDa)	Sequence	50.03	39.31	50.04	47.34	50.00	49.99	39.64	44.66
	Volume	47.2	43.3	48.9	47.9	66.1	46.4	38.7	40.5

Table S3.3 List of all *E. coli* strains and plasmids used in this study. All strains were generated in this study unless otherwise stated. TIR = 109 nt *rpsA* TIR unless otherwise stated. Genotypes of strains used: *E. coli* BL21 GOLD (DE3): *E. coli* B F– *ompT hsdS*($r_B^-m_B^-$) *dcm*$^+$ *Tetr gal λ*(DE3) *endA Hte*; *E. coli* DH5α: F– *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)*U169, hsdR17($r_K^-m_K^+$), λ– ; *E. coli* NEBα: *fhuA2 (argF-lacZ)*U169 *phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*.

Strain number	Strain	Plasmid	Marker	Description of hosted plasmid
JVS-001	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	G-26C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-011	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-9G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-019	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	empty vector
JVS-023	*E. coli* BL21 GOLD (DE3)	pUC57	KanR	Reverse (pT7-RBS-mRFP-DT)
JVS-030	*E. coli* BL21 GOLD (DE3)	- - - - - - - - - - - - - - -	- - - - - - - - - - -	wt cells - Agilent
JVS-031	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	wt pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-034	*E. coli* DH5α	pET-28(a)	AmpR	*rpsA* Good 2 - gift in kind from Stefano Marzi
JVS-045	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	G-78U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-046	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	G-68A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-049	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-50G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-050	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-47G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-051	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	G-44U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-052	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-42U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-053	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-42C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-054	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	G-37U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-055	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-25G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-056	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-24G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-059	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-4C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-060	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-2C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-061	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-3G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-063	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-1T pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)

JVS-064	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A+1G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-065	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A+1T pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-066	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C+5A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-073	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-61G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-074	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-3C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-075	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-63G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-082	*E. coli* DH5α	- - - - - - - - - - - -	- - - - - - - - - -	wild type cells
JVS-083	*E. coli* DH5α	pSB1C3	CmR	G-78U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-084	*E. coli* DH5α	pSB1C3	CmR	G-68A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-085	*E. coli* DH5α	pSB1C3	CmR	U-63G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-086	*E. coli* DH5α	pSB1C3	CmR	U-61G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-087	*E. coli* DH5α	pSB1C3	CmR	C-50G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-088	*E. coli* DH5α	pSB1C3	CmR	C-47G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-089	*E. coli* DH5α	pSB1C3	CmR	G-44U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-090	*E. coli* DH5α	pSB1C3	CmR	A-42U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-091	*E. coli* DH5α	pSB1C3	CmR	A-42C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-092	*E. coli* DH5α	pSB1C3	CmR	G-37U pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-093	*E. coli* DH5α	pSB1C3	CmR	U-25G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-094	*E. coli* DH5α	pSB1C3	CmR	U-24G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-095	*E. coli* DH5α	pSB1C3	CmR	U-24C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-097	*E. coli* DH5α	pSB1C3	CmR	A-4C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-098	*E. coli* DH5α	pSB1C3	CmR	A-3G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-099	*E. coli* DH5α	pSB1C3	CmR	A-3C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-100	*E. coli* DH5α	pSB1C3	CmR	A-2C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-101	*E. coli* DH5α	pSB1C3	CmR	C-1T pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)

JVS-102	*E. coli* DH5α	pSB1C3	CmR	A+1G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-103	*E. coli* DH5α	pSB1C3	CmR	A+1T pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-104	*E. coli* DH5α	pSB1C3	CmR	C+5A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-109	*E. coli* BL21 GOLD (DE3)		AmpR	IPTG inducible T7 polymerase
JVS-113	*E. coli* DH5α	pSB1A2	AmpR	pT7-RBS-eCFP-DT
JVS-127	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-10G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-128	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-36A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-129	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-6G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-130	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-13G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-131	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-75G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-132	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A+4G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-133	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-52A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-134	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	A-7C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-135	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	G-29C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-136	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	C-1A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-137	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U-5G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-138	*E. coli* BL21 GOLD (DE3)	pSB1C3	CmR	U+14C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT) #1
JVS-139	*E. coli* DH5α	pJET1.2 Blunt	AmpR	minimal 109 nt *rpsA* TIR in SHAPE cassette
JVS-140	*E. coli* DH5α	pJET1.2 Blunt	AmpR	C-1T *rpsA* TIR w/ eCFP CDS in SHAPE cassette
JVS-141	*E. coli* DH5α	pJET1.2 Blunt	AmpR	G-44U *rpsA* TIR w/ eCFP CDS in SHAPE cassette
JVS-142	*E. coli* DH5α	pJET1.2 Blunt	AmpR	A-4C *rpsA* TIR w/ eCFP CDS in SHAPE cassette
JVS-143	*E. coli* DH5α	pJET1.2 Blunt	AmpR	A-9G *rpsA* TIR w/ eCFP CDS in SHAPE cassette
JVS-144	*E. coli* DH5α	pJET1.2 Blunt	AmpR	wt *rpsA* TIR w/ eCFP CDS in SHAPE cassette
JVS-146	*E. coli* DH5α	pSB1C3	CmR	A-10C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-147	*E. coli* DH5α	pSB1C3	CmR	C-14G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)

JVS-148	*E. coli* DH5α	pSB1C3	CmR	T+12C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-149	*E. coli* DH5α	pSB1C3	CmR	G-8C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-150	*E. coli* DH5α	pSB1C3	CmR	G-44A pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-151	*E. coli* DH5α	pSB1C3	CmR	G-44C pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-155	*E. coli* DH5α	pSB1C3	CmR	A-10G pT7-TIR-eCFP-DT-Reverse (pT7-RBS-mRFP-DT)
JVS-161	*E. coli* DH5α	pJET1.2 Blunt	AmpR	minimal 109 nt *rpsA* TIR in SHAPE cassette
JVS-166	*E. coli* NEBα	pSB1C3	CmR	pT7-*rpsA*_TIR-eCFP-DT
JVS-169	*E. coli* NEBα	pSB1C3	CmR	pT7-rpsO_TIR-eCFP-DT

Table S3.4 List of all DNA oligonucleotides used in this study. DNA sequences are given in 5' to 3' direction; m denotes a methylation.

Name	Sequence (5' > 3')	Description
JVO-001	GAGCCCCGTTGCAATGTAATGACAGCGGG	site directed mutagenesis: G-78U sense
JVO-002	CCCGCTGTCATTACATTGCAACGGGGCTC	site directed mutagenesis: G-78U antisense
JVO-003	GCAATGGAATGACAGCAGGTATGTTAAACAACCCC	site directed mutagenesis: -68A sense
JVO-004	GGGGTTGTTTAACATACCTGCTGTCATTCCATTGC	site directed mutagenesis: G-68A antisense
JVO-005	GGAATGACAGCGGGTAGGTTAAACAACCCCATCCGGC	site directed mutagenesis: U-63G sense
JVO-006	GCCGGATGGGGTTGTTTAACCTACCCGCTGTCATTCC	site directed mutagenesis: U-63G antisense
JVO-007	GGAATGACAGCGGGTATGGTAAACAACCCCATCCGGC	site directed mutagenesis: U-61G sense
JVO-008	GCCGGATGGGGTTGTTTACCATACCCGCTGTCATTCC	site directed mutagenesis: U-61G antisense
JVO-009	GGTATGTTAAACAACCCGATCCGGCATGGAGCCAGG	site directed mutagenesis: C-50G sense
JVO-010	CCTGGCTCCATGCCGGATCGGGTTGTTTAACATACC	site directed mutagenesis: C-50G antisense
JVO-011	GGTATGTTAAACAACCCCATGCGGCATGGAGCCAGG	site directed mutagenesis: C-47G sense
JVO-012	CCTGGCTCCATGCCGCATGGGGTTGTTTAACATACC	site directed mutagenesis: C-47G antisense
JVO-013	GTTAAACAACCCCATCCGTCATGGAGCCAGGTGG	site directed mutagenesis: G-44U sense
JVO-014	CCACCTGGCTCCATGACGGATGGGGTTGTTTAAC	site directed mutagenesis: G-44U antisense
JVO-015	GTTAAACAACCCCATCCGGCTTGGAGCCAGGTGGACG	site directed mutagenesis: A-42U sense
JVO-016	CGTCCACCTGGCTCCAAGCCGGATGGGGTTGTTTAAC	site directed mutagenesis: A-42U antisense
JVO-017	GTTAAACAACCCCATCCGGCCTGGAGCCAGGTGGACG	site directed mutagenesis: A-42C sense
JVO-018	CGTCCACCTGGCTCCAGGCCGGATGGGGTTGTTTAAC	site directed mutagenesis: A-42C antisense
JVO-019	CCCCATCCGGCATGGATCCAGGTGGACGTTAAATATAAAC	site directed mutagenesis: G-37U sense
JVO-020	GTTTATATTTAACGTCCACCTGGATCCATGCCGGATGGGG	site directed mutagenesis: G-37U antisense
JVO-021	GGAGCCAGGTGGACGGTAAATATAAACCTGAAGATTAAAC	site directed mutagenesis: U-25G sense
JVO-022	GTTTAATCTTCAGGTTTATATTTACCGTCCACCTGGCTCC	site directed mutagenesis: U-25G antisense
JVO-023	GGAGCCAGGTGGACGTGAAATATAAACCTGAAGATTAAAC	site directed mutagenesis: U-24C sense

JVO-024	GTTTAATCTTCAGGTTTATATTTCACGTCCACCTGGCTCC	site directed mutagenesis: U-24C antisense
JVO-025	GGACGTTAAATATAAACCTGCAGATTAAACATGACTGAATC	site directed mutagenesis: A-10C sense
JVO-026	GATTCAGTCATGTTTAATCTGCAGGTTTATATTTAACGTCC	site directed mutagenesis: A-10C antisense
JVO-029	CGTTAAATATAAACCTGAAGATTCAACATGACTGAATCTTTTGC	site directed mutagenesis: A-4C sense
JVO-030	GCAAAAGATTCAGTCATGTTGAATCTTCAGGTTTATATTTAACG	site directed mutagenesis: A-4C antisense
JVO-031	CGTTAAATATAAACCTGAAGATTAACCATGACTGAATCTTTTGC	site directed mutagenesis: A-2C sense
JVO-032	GCAAAAGATTCAGTCATGGTTAATCTTCAGGTTTATATTTAACG	site directed mutagenesis: A-2C antisense
JVO-033	CGTTAAATATAAACCTGAAGATTAGACATGACTGAATCTTTTGC	site directed mutagenesis: A-3G sense
JVO-034	GCAAAAGATTCAGTCATGTCTAATCTTCAGGTTTATATTTAACG	site directed mutagenesis: A-3G antisense
JVO-035	CGTTAAATATAAACCTGAAGATTACACATGACTGAATCTTTTGC	site directed mutagenesis: A-3C sense
JVO-036	GCAAAAGATTCAGTCATGTGTAATCTTCAGGTTTATATTTAACG	site directed mutagenesis: A-3C antisense
JVO-037	CCTGAAGATTAAATATGACTGAATCTTTTGCTactagAatgg	site directed mutagenesis: C-1T sense
JVO-038	ccatTctagtAGCAAAAGATTCAGTCATATTTAATCTTCAGG	site directed mutagenesis: C-1T antisense
JVO-039	CCTGAAGATTAAACGTGACTGAATCTTTTGCTactagAatgg	site directed mutagenesis: A+1G sense
JVO-040	ccatTctagtAGCAAAAGATTCAGTCACGTTTAATCTTCAGG	site directed mutagenesis: A+1G antisense
JVO-041	CCTGAAGATTAAACTTGACTGAATCTTTTGCTactagAatgg	site directed mutagenesis: A+1T sense
JVO-042	ccatTctagtAGCAAAAGATTCAGTCAAGTTTAATCTTCAGG	site directed mutagenesis: A+1T antisense
JVO-043	CCTGAAGATTAAACATGAATGAATCTTTTGCTactagAatgg	site directed mutagenesis: C+5A sense
JVO-044	ccatTctagtAGCAAAAGATTCATTCATGTTTAATCTTCAGG	site directed mutagenesis: C+5A antisense
JVO-059	mGmAACAGCTCCTCGCCCTTGC	amplifying transcription template: TIR Universal Suffix (binds within eCFP)
JVO-060	CTAGAGTAATACGACTCACTATAGGG	amplifying transcription template: T7 Promter
JVO-075	mAmGCAAAAGATTCAGTCATGTTTAATC	amplifying transcription template: *rpsA* TIR - (amplifies minimal RNAs used in SAXS and filter binding)
JVO-088	CGTCTTCGGAGGAAGCCATTCTAGAAGCGGCCGCGAATTC	site directed mutagenesis: BBa_E1010 - converting from RFC_10 to RFC_23 forward
JVO-089	GAATTCGCGGCCGCTTCTAGAATGGCTTCCTCCGAAGACG	site directed mutagenesis: BBa_E1010 - converting from

		RFC_10 to RFC_23 reverse
JVO-092	CTCGCCCTTGCTCACCATTCTAGTAGCAAAAGATTCAGTC	site directed mutagenesis: BBa_E1010 - converting from RFC_10 to RFC_23 reverse
JVO-093	GACTGAATCTTTTGCTACTAGAATGGTGAGCAAGGGCGAG	site directed mutagenesis: BBa_E0020 - converting from RFC_10 to RFC_23 reverse
JVO-094	GACTGAATCTTTTGCTACTAGACTGGTGAGCAAGGGCGAG	site directed mutagenesis: BBa_E0020 - converting from RFC_10 to RFC_23 reverse
JVO-096	GCCAGGTGGACGTTAAATATAAACCTGAGGATTAAACATGACTGAATC	site directed mutagenesis: A-9G sense
JVO-097	GATTCAGTCATGTTTAATCCTCAGGTTTATATTTAACGTCCACCTGGC	site directed mutagenesis: A-9G antisense
JVO-098	CCGGCATGGAGCCAGGTGGACCTTAAATATAAACCTGAAG	site directed mutagenesis: G-26C sense
JVO-099	CTTCAGGTTTATATTTAAGGTCCACCTGGCTCCATGCCGG	site directed mutagenesis: G-26C antisense
JVO-102	CTCTAGTATCTCCCTATAGTGAGTCG	generating *rpsA* TIR truncation mutants in pT7 controlled biobrick constructs: upstream
JVO-103	AATGGAATGACAGCGGGTATG	generating *rpsA* TIR truncation mutants in pT7 controlled biobrick constructs: -82 downstream
JVO-104	GTATGTTAAACAACCCCATCCGG	generating *rpsA* TIR truncation mutants in pT7 controlled biobrick constructs: -66 downstream
JVO-113	GGA CGT TAA ATA TAA ACC TGA ACA TTA AAC ATG ACT GAA TC	site directed mutagenesis: G-8C sense
JVO-114	GAT TCA GTC ATG TTT AAT GTT CAG GTT TAT ATT TAA CGT CC	site directed mutagenesis: G-8C antisense
JVO-115	CCT GAA GAT TAA ACA TGA CTG AAT CCT TTG CTA CTA GAA TGG	site directed mutagenesis: U+12C sense
JVO-116	CCA TTC TAG TAG CAA AGG ATT CAG TCA TGT TTA ATC TTC AGG	site directed mutagenesis: U+12C antisense
JVO-117	GGA CGT TAA ATA TAA ACC TGG AGA TTA AAC ATG ACT GAA TC	site directed mutagenesis: A-10G sense
JVO-118	GAT TCA GTC ATG TTT AAT CTC CAG GTT TAT ATT TAA CGT CC	site directed mutagenesis: A-10G antisense
JVO-119	GGA CGT TAA ATA TAA ACC TGA AGG TTA AAC ATG ACT GAA TC	site directed mutagenesis: A-7G sense
JVO-120	GAT TCA GTC ATG TTT AAC CTT CAG GTT TAT ATT TAA CGT CC	site directed mutagenesis: A-7G antisense
JVO-121	CAG CGG GTA TGT TAA ACA ACA CCA TCC GGC ATG GAG CC	site directed mutagenesis: C-52A sense
JVO-122	GGC TCC ATG CCG GAT GGT GTT GTT TAA CAT ACC CGC TG	site directed mutagenesis: C-52A antisense
JVO-123	GGA CGT TAA ATA TAA ACC TGA AGC TTA AAC ATG ACT GAA TC	site directed mutagenesis: A-7C sense
JVO-124	GAT TCA GTC ATG TTT AAG CTT CAG GTT TAT ATT TAA CGT CC	site directed mutagenesis: A-7C antisense

JVO-125	CCT GAA GAT TAA ACA TGA CTG AAC CTT TTG CTA CTA GAA TGG	site directed mutagenesis: U+10G sense
JVO-126	CCA TTC TAG TAG CAA AAG GTT CAG TCA TGT TTA ATC TTC AGG	site directed mutagenesis: U+10G antisense
JVO-127	GGT GGA CGT TAA ATA TAA AGC TGA AGA TTA AAC ATG ACT G	site directed mutagenesis: C-14G sense
JVO-128	CAG TCA TGT TTA ATC TTC AGC TTT ATA TTT AAC GTC CAC C	site directed mutagenesis: C-14G antisense
JVO-129	CCG GCA TGG AGC CAG GTG CAC GTT AAA TAT AAA CCT G	site directed mutagenesis: G-29C sense
JVO-130	CAG GTT TAT ATT TAA CGT GCA CCT GGC TCC ATG CCG G	site directed mutagenesis: G-29C antisense
JVO-131	CCC ATC CGG CAT GGA GAC AGG TGG ACG TTA AAT ATA AAC C	site directed mutagenesis: C-36A sense
JVO-132	GGT TTA TAT TTA ACG TCC ACC TGT CTC CAT GCC GGA TGG G	site directed mutagenesis: C-36A antisense
JVO-133	GGT GGA CGT TAA ATA TAA AAC TGA AGA TTA AAC ATG ACT G	site directed mutagenesis: C-14A sense
JVO-134	CAG TCA TGT TTA ATC TTC AGT TTT ATA TTT AAC GTC CAC C	site directed mutagenesis: C-14A antisense
JVO-135	GGA CGT TAA ATA TAA ACC TGA AGA GTA AAC ATG ACT GAA TC	site directed mutagenesis: U-6G sense
JVO-136	GAT TCA GTC ATG TTT ACT CTT CAG GTT TAT ATT TAA CGT CC	site directed mutagenesis: U-6G antisense
JVO-137	GGT GGA CGT TAA ATA TAA ACG TGA AGA TTA AAC ATG ACT G	site directed mutagenesis: C-13G sense
JVO-138	CAG TCA TGT TTA ATC TTC ACG TTT ATA TTT AAC GTC CAC C	site directed mutagenesis: C-13G antisense
JVO-139	CCT GAA GAT TAA AAA TGA CTG AAT CTT TTG CTA CTA GAA TGG	site directed mutagenesis: C-1A sense
JVO-140	CCA TTC TAG TAG CAA AAG ATT CAG TCA TTT TTA ATC TTC AGG	site directed mutagenesis: C-1A antisense
JVO-141	CCC CGT TGC AAT GGA AGG ACA GCG GGT ATG TTA AAC	site directed mutagenesis: U-75G sense
JVO-142	GTT TAA CAT ACC CGC TGT CCT TCC ATT GCA ACG GGG	site directed mutagenesis: U-75G antisense
JVO-143	CCT GAA GAT TAA ACA TGA CTU AAT CTT TTG CTA CTA GAA TGG	site directed mutagenesis: G+7U sense
JVO-144	CCA TTC TAG TAG CAA AAG ATT AAG TCA TGT TTA ATC TTC AGG	site directed mutagenesis: G+7U antisense
JVO-145	GGA CGT TAA ATA TAA ACC TGA AGA TGA AAC ATG ACT GAA TCT TTT GC	site directed mutagenesis: U-5G sense
JVO-146	GCA AAA GAT TCA GTC ATG TTT CAT CTT CAG GTT TAT ATT TAA CGT CC	site directed mutagenesis: U-5G antisense
JVO-147	CCT GAA GAT TAA ACA TGG CTG AAT CTT TTG CTA CTA GAA TGG	site directed mutagenesis: A+4G sense
JVO-148	CCA TTC TAG TAG CAA AAG ATT CAG CCA TGT TTA ATC TTC AGG	site directed mutagenesis: A+4G antisense

JVO-149	CAT GAC TGA ATC TTC TGC TAC TAG AAT GGT GAG C	site directed mutagenesis: U+14C sense
JVO-150	GCT CAC CAT TCT AGT AGC AGA AGA TTC AGT CAT	site directed mutagenesis: U+14C antisense
JVO-155	TAATACGACTCACTATAGGGCCTTCG	*rpsA* SHAPE construct forward primer
JVO-156	mGmAACCGGACCGAAGCCCG	*rpsA* SHAPE construct reverse primer
JHO-005	GCTGAAGGGCATCGACTTC	RT-qPCR - eCFP forward
JHO-006	CTTGTCGGCGGTGATATAGAC	RT-qPCR - eCFP reverse
JHO-007	CTCCCACAACGAAGACTACAC	RT-qPCR - mRFP forward
JHO-008	GCGATCTACACTAGCACTATCAG	RT-qPCR - mRFP reverse

Table S3.5 List of RNAs used in *in vitro* experiments. RNA sequences are given in 5' to 3' direction.

Name	DNA template	Primers	Size of RNA	Experiment	Sequence (5' > 3')
***rpsA* TIR**	JVS-031	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
minimal *rpsA* TIR	JVS-031	JVO-060 JVO-075	121	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCU
A-9G *rpsA* TIR	JVS-011	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAGGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
G-44U *rpsA* TIR	JVS-051	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGUCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
C-1U *rpsA* TIR	JVS-063	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAAUAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
A-4C *rpsA* TIR	JVS-059	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUCAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
Δ-82 *rpsA TIR*	JVS-125	JVO-060 JVO-059	145	SAXS	GAGAUACUAGAGAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
Δ-66 *rpsA TIR*	JVS-126	JVO-060 JVO-059	129	SAXS	GAGAUACUAGAGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC
***rpsA* TIR**	JVS-144	JVO-060 JVO-059	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGCCGGAUCCAAA

					UCGGGCUUCGGUCCGGUUCUCGAG
minimal *rpsA* TIR	JVS-139	JVO-155 JVO-156	183	SHAPE probing	GGCCTTCGGGCCAAGAGATACTAGAGCCCCGTTGCAATGGAATGACAGCG GGTATGTTAAACAACCCCATCCGGCATGGAGCCAGGTGGACGTTAAATAT AAACCTGAAGATTAAACATGACTGAATCTTTTGCTTCGATCCGGTTCGCCG GATCCAAATCGGGCTTCGGTCCGGTTCTCGAG
A-9G *rpsA* TIR	JVS-143	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAG CGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAA AUAUAAACCUGAGGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUG GUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGCCGGAUCCAAA UCGGGCUUCGGUCCGGUUCUCGAG
G-44U *rpsA* TIR	JVS-141	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAG CGGGUAUGUUAAACAACCCCAUCCGUCAUGGAGCCAGGUGGACGUUAA AUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUG GUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGCCGGAUCCAAA UCGGGCUUCGGUCCGGUUCUCGAG
C-1U *rpsA* TIR	JVS-140	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAG CGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAA AUAUAAACCUGAAGAUUAAAUAUGACUGAAUCUUUUGCUACUAGAAUG GUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGCCGGAUCCAAA UCGGGCUUCGGUCCGGUUCUCGAG
A-4C *rpsA* TIR	JVS-142	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAG CGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAA AUAUAAACCUGAAGAUUCAACAUGACUGAAUCUUUUGCUACUAGAAUG GUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGCCGGAUCCAAA UCGGGCUUCGGUCCGGUUCUCGAG

3.5.3 Supplemental Discussion

Sort Seq

The expression module contains a reporter gene (eCFP) translationally controlled by the *rpsA* TIR. Changes in the eCFP output can be correlated to translation initiation. On the same plasmid a noise module was incorporated to measure cellular gene expression independent of the expression module. The noise module contains a reporter gene (mRFP) translationally controlled by a Shine-Dalgarno sequence (BBa_B0034). To reduce strain competition during library construction and propagation, each module was placed under the control of a T7 RNA polymerase promoter.

Reporter output resulting from the expression module was related to the noise module expression levels to characterize the activity of each *rpsA* TIR mutant. By correlating the reporter output between the two modules, cells with low activity were isolated from those that have lost or mutated the plasmid. The frequency of differentially expressing cells within the original input mutant libraries were significantly enriched in all gated populations following FACS. To determine the efficiency of the cell sorting, 10,000 cells from the isolated populations were re-analyzed using flow-cytometry (Figure 3C). The FACS procedure enriched cells with specific phenotypes as follows: high expressing were enriched from 7.6 % in the mutant pool to 72.8 % in the high expressing cells, wildtype (wt) like cells were enriched 50.7 % to 92.0 %, low-expressing cells were enriched from 10.1 % to 82.6 %, and cells with no detectable eCFP expression was 5.9 % to 56.3 %.

The *rpsA* region from the isolated cells was sequenced using NGS to analyze the diverse reporter output, therefore determining the relationship between TIR sequence and translation initiation activity (Figure 4A). A significant proportion of sequence reads in

each sample consisted of the wt sequence, which has been attributed to experimental noise. 25–30 % contained a single mutation, roughly 10 % had two or more mutations, and less than 5 % had three or more mutations, which is consistent with similar studies (Holmqvist et al., 2013). The mutation frequency achieved allowed for a survey of single point mutations. The sequence reads from each mutant pool were normalized against the sequence reads from the wt-like pool. To evaluate the accuracy of our method and to create a library of highly characterized standard parts, several mutations were re-introduced into the TIR and functionally assayed using flow-cytometry. The results generally correlate well with those from the high-throughput mutagenesis strategy.

First the basal activity of the expression module was analyzed based on eCFP output using flow-cytometry. Our data indicates that the *rpsA* TIR is capable of driving translation of eCFP (Figure 2A). eCFP output resulting from the *rpsA* TIR was compared with other TIRs: the *rpsO* TIR, and BBa_B0034 (Figure 2A). BBa_B0034 is a commonly used strong SD sequence; the *rpsO* TIR controls the expression of the highly expressed ribosomal protein S15. The *rpsA* TIR achieves highest reporter output, significantly higher than BBa_B0034 despite having less 16S complementarity. The basal fluorescence level of reporter plasmid containing both the expression module and noise module was tested using flow-cytometry (Figure 2B). In addition to the wt plasmid a previously characterized *rpsA* variant was incorporated into the expression module, that confirmed that differences in expression could be resolved using this strategy (Figure 2B) (Rasmussen et al., 1993). By correlating outputs between the two modules, this population was able to be resolved from the cells harboring the non-mutated construct (Figure 2B). A plasmid where the expression module encoded mRFP and the noise module encoded eCFP was built. When the same variant was incorporated into the *rpsA*

region in this plasmid a similar response to mRFP fluorescence (~ 4-fold increase) as in our original reporter plasmid was observed (Figure S2). Its ability to drive expression predictably when placed under several different contexts highlights a potential use as modular RNA control device with broad applicability.

SHAPE Analysis of *rpsA* TIR Variants

Our original RNAs include extended coding sequence nucleotides; to confirm that the *rpsA* TIR wt maintains its structure upon removal of these areas a minimal version of the *rpsA* TIR lacking the 3' scar sequence was built. The reactivities are consistent among these RNAs, except for some differences after the start codon (Figure 3.10). These two mRNAs are more reactive at the 3' end. Importantly, all three have similar reactivities suggesting that the three-stem-loop structure is maintained.

Chapter 4

Ribosomal Protein S1 – Multiple Cellular Roles

4.1 Introduction

Translation initiation is the rate-determining step of protein biosynthesis and can be regulated through multiple strategies (Chapter 2). In bacteria the small ribosomal subunit (30S) initiation complex is assembled by mRNA binding to the 30S along with three initiation factors and an initiator tRNA (Laursen et al., 2005). The canonical mechanism of bacterial initiation involves a motif upstream of the start codon, called the Shine-Dalgarno (SD) element. The SD facilitates ribosome binding and correct positioning of the start codon in the P-site of the 30S via a base pairing interaction with the anti-SD element on the 3' end of the 16S rRNA (Gold et al., 1981; A. Hui & de Boer, 1987; Shine & Dalgarno, 1974). However, it has been shown that a surprising number of bacterial genes (between 10 and 90 %, depending on the species) do not contain an SD element (B. Chang et al., 2006). This suggests that other mechanisms must exist that recruit and correctly position mRNAs onto the 30S. Ribosomal protein S1, in part, bridges this gap as is able to recognize and bind mRNA during translation initiation, regardless of the presence or absence of a SD element (Boni et al., 1991; Komarova et al., 2002; Sorensen et al., 1998; Tzareva, Makhno, & Boni, 1994). This function is essential for cell viability, and required for the translation of most natural mRNAs in *Escherichia coli* (*E. coli*) and other gram-negative bacteria (Sorensen et al., 1998).

Besides translation initiation S1 is also involved in multiple other roles. Non-ribosome bound— or free S1—is able to bind its own mRNA and, as a consequence, down regulate its own expression (Boni et al., 2000). S1 is suggested to be involved in transcriptional cycling in *E. coli* (Sukhodolets, Garges, & Adhya, 2006), and to have a role in tmRNA translation (McGinness & Sauer, 2004)—although this function is currently under debate (Qi, Shimizu, & Ueda, 2007). Ribosomal protein S1 has also been

implicated in processes involving viral proteins, including Qβ RNA replication (Vasilyev, Kutlubaeva, Ugarov, Chetverina, & Chetverin, 2013) and T4 RegB activation (Ruckman, Ringquist, Brody, & Gold, 1994). Interestingly, S1 is only loosely bound to the 30S and purified 70S ribosome often contain less than stoichiometric amounts of S1 (Voynow & Kurland, 1971).

As an essential protein in gram-negative bacteria, S1 presents as an attractive potential drug target. However, a full understanding of its structure and function(s) will be critical. Although multiple important roles have been identified, our mechanistic understanding of its functions is lacking. Structural studies of S1 are difficult and obtaining high-resolution structural information of free S1, as well as bound to its many functional substrates should be one of the main aims of future research into S1, enabling the rational design of novel translation inhibitors.

4.2 Transcriptional Regulation of the *rpsA* Gene

4.2.1 S1: From Gene to Gene Product and Back

The fascinating features of ribosomal protein S1 extend from the proteins elaborate six-domain architecture, to its multilayered network of transcriptional and translation control signals that ensure its correct cellular abundance. Early studies hinted at an interesting regulation strategy different from typical ribosomal proteins, and responsive to amino acid availability (Pedersen, Bloch, Reeh, & Neidhardt, 1978; Reeh, Pedersen, & Friesen, 1976). Even after several decades of studies investigating how S1 production is regulated, we are still parsing the complex signals and regulatory elements that enable its tight response to changing cellular conditions.

4.2.2 Operon Architecture

The gene encoding S1 (*rpsA)* shares an operon with two other genes. It is flanked upstream by the *cmk* gene, which encodes cytidylate kinase, and downstream by the *ihfB* gene which encodes the β-subunit of the integration host factor. The operon contains at least four different transcriptional units, and the specific mRNA species expressed at a given time is dynamic and depends on several cellular factors. Pedersen *et al.,* first mapped the transcription start sites for *rpsA* in 1984, and were able to determine a region within the *cmk* gene between 240 and 113 base pairs (bp) upstream of the S1 start codon that was necessary for transcription. A more in-depth investigation of this region revealed four distinct transcription products: P_1, P_2, P_3, and P_4 (Pedersen, Skouv, Kajitani, & Ishihama, 1984). It was proposed that RNAs derived from the P_2 and P_4 promoters were simply degradation products resulting from P_1 (*rpsA*p_1) and P_3 (*rpsA*p_3) driven transcription, and that the two main promoters driving *rpsA* transcription are P_1 and P_3 (Pedersen et al., 1984). In addition to these promoters, a P_0 (*cmk*p) promoter is located upstream of *cmk* that drives expression of the *cmk-rpsA* and *cmk-rpsA-ihfB* transcripts.

During exponential growth the *rpsA*p_1 and *rpsA*p_3 promoters are active. As the cells enter stationary phase a sharp decrease in the levels of long mRNAs containing cotranscribed *rpsA* and *ihfB* genes, and an increase in the monocistronic species of the *ihfB* gene can be observed (Weglenska, Jacob, & Sirko, 1996). During this time the σ^{32} polymerase is able to compete with σ^{70}, activating the P_5 (*ihfB*p) promoter, resulting in a 5–10 fold increase in IHF levels (Ditto, Roberts, & Weisberg, 1994).

The resulting mRNAs from this operon terminate after the 3' end of the *ihfB* gene, however products that terminate in the intercistronic region between *rpsA* and *ihfB* have also been observed. This region can potentially fold into two alternate secondary

structures, resembling an interrupted palindrome followed by a uridine rich region that can serve as an inefficient Rho-independent terminator (d'Aubenton Carafa, Brody, & Thermes, 1990; Weglenska et al., 1996). This regulation mechanism requires further investigation; however, the respective transcription stop site has been observed in recent large-scale transcriptomic 3'-end mapping studies (Dar & Sorek, 2018).

The intergenic terminating stem-loop structures have been proposed to interact with functional prematurely terminated RNA fragments (faRNA) produced from abortive transcription cycling on the *rpsA*p$_1$ promoter (Marcus, Hassoun, & Nair, 2017). Short single-stranded unstructured faRNAs can form weak, transient complexes with complementary nucleotide sequences (M. P. Hui, Foley, & Belasco, 2014), and have been predicted to result in anti-termination and co-transcription of *rpsA* and *ihfB* (Marcus et al., 2017). This hypothesis is consistent with the observed *in vivo ihf*B expression pattern—in stationary growth phase the reduced demand for translational machinery results in fewer transcripts being produced from the *rpsA*p$_1$ and *rpsA*p$_3$ promoters. This in turn would lead to a lower concentration of abortive fragments, and reduced binding to the *rpsA* terminator. In this case *cmk-rpsA* and *rpsA* transcription would terminate before the *ihfB*p promoter. During exponential growth phase as *rpsA* transcription rates increase, more abortive fragments would lead to a higher chance of anti-termination causing the number of *ihfB* transcripts produced from *rpsA* promoters to increase during exponential phase (Marcus et al., 2017).

4.2.3 Transcriptional Response to Amino Acid Levels

In 1984 the promoter selectivity of *E. coli* RNA polymerase (RNAP) in the presence and absence of the alarmone guanosine tetraphosphate (ppGpp) was

investigated (Kajitani & Ishihama, 1984). ppGpp (and the intermediate pppGpp) is known to regulate the stringent response in bacteria (Traxler et al., 2008). During stringent response two main regulatory events happen: (1) rRNA and ribosomal protein synthesis slows down, and (2) the biosynthesis of various metabolites is halted. These processes are mediated directly or indirectly by the regulatory nucleotides ppGpp and pppGpp. Briefly, uncharged tRNAs induce ribosome-associated RelA, and SpoT to synthesize ppGpp. ppGpp and the transcription factor DskA bind to RNAP and inhibit transcription from most promoters of ribosomal protein operons to tune ribosome synthesis rates in response to nutrient availability. ppGpp was found to repress transcription from the upstream *rpsA*p$_1$ promoter, but transcription from the *rpsA*p$_3$ promoter was unaffected (Kajitani & Ishihama, 1984). The Gourse lab demonstrated that the *rpsA*p$_1$ promoter is inhibited two-fold by ppGpp and DksA *in vitro*, while the *rpsA*p$_3$ promoter was not (Lemke et al., 2011). Finally, high-throughput transcriptomic approaches have shown that the number of ribosomal protein (r-protein) transcripts decreases during amino acid starvation (Durfee, Hansen, Zhi, Blattner, & Jin, 2008; Traxler et al., 2008).

S1's regulation on the translational level is partially linked to the amino acid availability and the stringent response. The varied layers of regulatory elements within this operon could enable separate responses to fast changes in cellular amino acid concentration, while another could be important for slow changes. This transcriptional regulation of S1 is important, however it has long been known that S1's translational regulation is sophisticated.

4.3 Translational Control of S1

In the early 80s the *rpsA* gene was cloned into a multi-copy plasmid enabling the over expression of S1. Surprisingly, although *rpsA* transcription increased 40-fold, S1 synthesis by the ribosome only doubled (Christiansen & Pedersen, 1981). This highlighted the importance of post-transcriptional regulation for cellular S1 levels, hinting at a mechanism different from those reported for other r-proteins (Pedersen et al., 1978; Reeh et al., 1976). The *rpsA* gene is under negative feedback control with the gene product itself acting as the effector for this regulation. The regulatory mechanism involves binding of S1 to the translation initiation region (TIR) of the *rpsA* mRNA (~95 nt upstream from the start codon) (Skouv, Schnier, Rasmussen, Subramanian, & Pedersen, 1990; Tchufistova et al., 2003). To understand this interaction, it is important to recognize the unique features of both the RNA and S1.

The *rpsA* TIR has several unusual features; it can fold into three consecutive stem-loops separated by two A/U-rich single-stranded regions (Boni et al., 2000; Boni et al., 2001). The most 5' stem-loops are very stable, while the third stem-loop containing the start codon within the apical loop is not as strong and contains a weak SD sequence. Phylogenetic analysis of *rpsA* TIRs has revealed little sequence similarity between organisms within ϒ-proteobacteria (Boni et al., 2001). However, it is hypothesized that all can potentially fold into a similar three-stem-loop structure, further supporting the relevance of its structure (Boni et al., 2001). The TIR of this mRNA has only three nucleotides that can form a contiguous interaction with the anti-SD sequence on the 16S rRNA. Despite this high degree of secondary structure and weak SD sequence, this TIR drives the expression of one of the most highly expressed genes in *E. coli*. The phylogenetically conserved architecture is not only required for its high translational activity, but also the auto-regulation mechanism (Boni et al., 2001; Tchufistova et al.,

2003). There have been several studies investigating this TIR, and how S1 binds to and alters it structure. The lower part of stem-loop II is melted by S1 upon binding to regions between the stem-loops (Figure 4.1). The autogenous regulation is achieved via the competition of free-S1 and ribosome-bound S1 for binding to the TIR. In 1993 the first mutational study was published on the *rpsA* TIR (Jesper Vind, Sørensen, Rasmussen, & Pedersen, 1993).

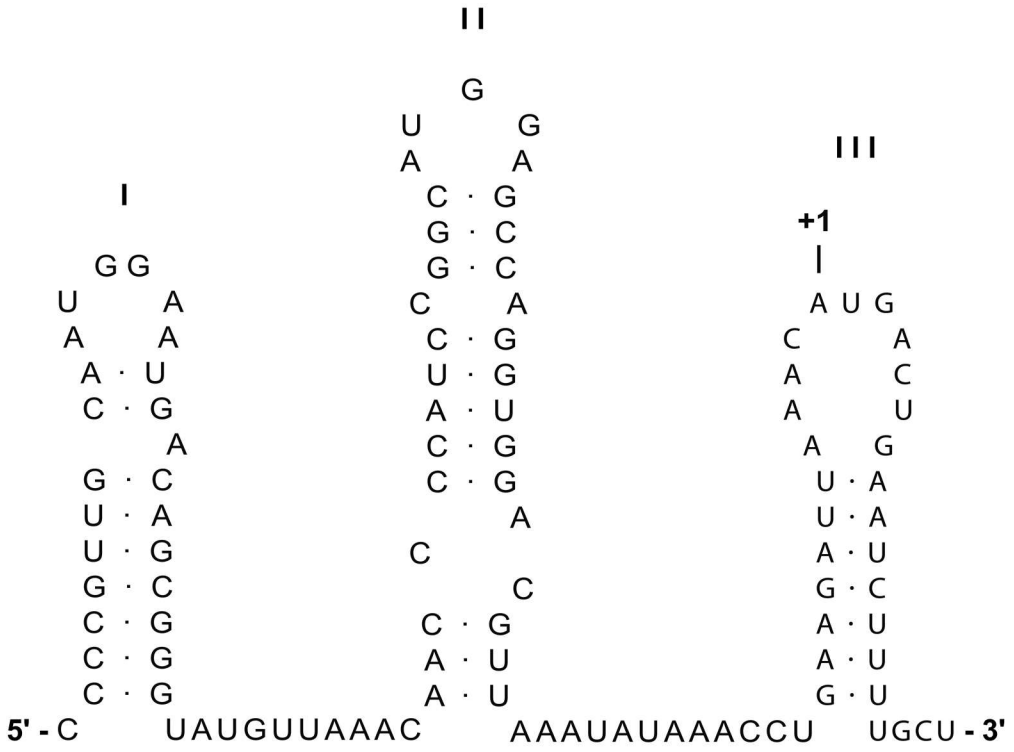


Figure 4.1. The *rpsA* TIR governing the translation of S1 is highly efficient, despite its high degree of secondary structure and containing a weak SD element. The TIR is proposed to fold into the stem-loops (I, II, and III) separated by A/U rich single stranded regions. +1 denotes the start codon.

This study also looked at a variant with strengthened SD sequence (A to G mutation 9 nt upstream of the AUG) and found that S1 is unable to regulate this mRNA. The apical loops of the two most 5' stem-loops contain conserved GGA motifs that are essential for its high translation activity (Boni et al., 2001). It has been speculated that the

structure of the TIR brings these conserved motifs in close proximity in order to form a discontinuous SD sequence. Ultimately, this hypothesis was disproven using specialized ribosomes (Skorski et al., 2006). The GGAs in the *rpsA* TIR were mutated to CCUs and the translation activity was monitored *in vivo*. These mutations did reduce translation of this mRNA, however introducing specialized ribosomes with a compensatory mutation in the aSD (CCUCCU to a GGAGG) could not rectify the decrease in translation. The non-canonical SD in the third stem was shown to be dispensable for activity, highlighting that the *rpsA* TIR does not primarily rely on the SD-aSD interaction for translation initiation (Boni et al., 2001; Skorski et al., 2006). In *in vivo* SHAPE assays this area seems to be unfolded (Mustoe et al., 2018). The specific folding of the *rpsA* TIR could generate an optimal spatial arrangement of sequence elements that interact with the 30S (Boni et al., 2001), similar to how some viral IRES containing mRNAs operate (Skorski et al., 2006). Against this background it is still unclear how the structural elements in the *rpsA* TIR contribute to translation initiation.

Although feedback loops and autoregulation mechanisms are common among r-proteins, the mechanism employed by S1 is unique. Typically, r-proteins dock onto the 23S or 16S rRNA during ribosome biogenesis. The binding site on their mRNAs mimics their respective rRNA binding sites, ultimately leading to inhibition of their translation. In contrast, S1 does not bind to rRNA and must identify its own mRNA using a different signal. Furthermore, S1 acts as a global regulator of RNA, while acting as a highly specific translational repressor of its own mRNA (Boni et al., 2000; Skouv et al., 1990; Jesper Vind et al., 1993).

Structurally, S1 belongs to the OB-fold family of proteins that are highly specific for single stranded nucleic acids (Draper & Reynaldo, 1999), binding polyU, polyC,

polyA, as well as heterogenous RNAs (Subramanian, 1983; Suryanarayana & Subramanian, 1984), and has no strict sequence specificity. Various S1 footprinting studies have revealed that S1 binds specifically within certain U or A/U containing single-stranded regions present in mRNAs (Boni et al., 1991; Tzareva et al., 1994). S1 also binds two regions of the Qβ phage RNA (Boni et al., 1991; Goelz & Steitz, 1977; Miranda et al., 1997). SELEX experiments from the Gold Lab have shown that both free S1 and S1-bound 30S can bind RNA aptamers with very high affinity, specifically those with GGA containing pseudoknots (Ringquist et al., 1995). Mogridge and Greenblatt (1998) described S1's ability to bind the rrn anti-terminator BoxA; in doing so they showed that different BoxA mutations had different affinities to S1 (Mogridge & Greenblatt, 1998). It is evident that S1 has a hierarchy of RNA targets to bind. Examining the structure of S1 will provide insight into its broad range of functions by revealing the specific interactions with its RNA substrates.

4.4 Domain Architecture and Structure of S1

4.4.1 S1 Domains

S1 is composed of six repeating motifs: M1–M6 illustrated in Figure 4.2A (Salah et al., 2009; Subramanian, 1983). These motifs contain conserved oligonucleotide-binding (OB) folds formed by five β-strands, which are integral components of RNA-binding proteins in organisms across all domains of life (Bycroft, Hubbard, Proctor, Freund, & Murzin, 1997). Sequence analysis of S1's six OB motifs revealed that certain β-strands are less conserved throughout the protein than others: strands β1, β2, and β4 are all relatively conserved throughout M1–M6, while β3 and β5 are less conserved in M1 and M2 (Salah et al., 2009). The evolutionary divergence of M1 and M2 can likely be

ascribed to its diverse protein-binding functions, which enable S1 the ability to bind to the ribosome and the Qβ replicase, while M4–M6 make up S1's RNA binding domain (Byrgazov, Manoharadas, Kaberdina, Vesper, & Moll, 2012; Giorginis & Subramanian, 1980; Subramanian et al., 1981). This bi-domain structure allows S1 to bind various protein and RNA substrates, enabling its participation in multiple cellular processes.

High-resolution structures of individual S1 domains have been reported using different experimental techniques, including nuclear magnetic resonance (NMR) spectroscopy. Protein binding domains M1 and M2 were solved using NMR (Giraud, Crechet, Uzan, Bontems, & Sizun, 2015). While M2 has the canonical fold and topology expected for an OB-fold domain, M1's structure has diverged. Additional NMR studies of M1 suggests that this motif only has four of the usual five β-strands (Byrgazov et al., 2015). The three-dimensional structure of two of the RNA-binding domains (M4 and M6) have also been solved using NMR (Salah et al., 2009), and reveal similar arrangements. In future studies concerning the cellular role of S1, it will be important to consider the contributions of each individual motif. Only M1–M3 are essential for cell growth and mRNA binding/30S initiation complex formation (Cifuentes-Goches, Hernandez-Ancheyta, Guarneros, Oviedo, & Hernandez-Sanchez, 2019; Duval et al., 2013), and there exist S1 homologues that contain only these three motifs (Figure 4.3). Various groups have aimed to determine why S1 has retained the last three motifs if they are not essential. It has been shown that M6 is dispensable for S1's role in translation initiation, but could be essential for autoregulation (Boni et al., 2000). This motif is also dispensable for S1's role in Qβ replication (Guerrier-Takada, Subramanian, & Cole, 1983). The group of Francois Bontems has analyzed the modularity of S1's RNA-binding motifs with regards to S1's function during RegB activation, revealing that the four C-terminal motifs

behave cooperatively and that their essentiality is substrate-dependent (Bisaglia, Laalami, Uzan, & Bontems, 2003), with recognition of different RNAs stemming from the adaptability of the M3–M5 module (Aliprandi et al., 2008).

4.4.2 Structural Analysis and Location of S1 on the Ribosome

S1 is the largest protein of the *E. coli* ribosome, consisting of 667 amino acids, with a molecular weight of 61 kDa (Giri & Subramanian, 1977; Kimura, Foulaki, Subramanian, & Wittmann-Liebold, 1982). Due to its large size and flexibility it is difficult to obtain high-resolution structural information. In fact, crystallization of *E. coli* and *Thermus thermophilus* ribosomes requires the removal of S1 (Pioletti et al., 2001; Schluenzen et al., 2000; Schuwirth et al., 2005; Wimberly et al., 2000). Low-resolution hydrodynamic analyses revealed that S1 is an elongated protein both on and off the ribosome, with a maximum length of over 250 Å (Labischinski & Subramanian, 1979; Laughrea & Moore, 1977). Although this hydrodynamic data has been collected by multiple groups, no low-resolution model of the protein has been published, likely due to the lack of computational tools available during the time of the study.

Obtaining a high-resolution structure of S1 off the ribosome may be too challenging with current techniques, however some structural studies of ribosome-bound S1 have elucidated its binding site and interaction partners on the 30S. Initial structural evidence obtained through cryo-EM, and cross-linking combined with mass spectrometry experiments placed S1 at the junction of the head, platform, and main body of the 30S. This would enable S1 to be situated in the correct position to interact with mRNA sequences upstream of the SD element (Lauber et al., 2012; Sengupta, Agrawal, & Frank, 2001). S1 adopts an elongated conformation and exhibits extreme flexibility even while

bound to the ribosome (Lauber et al., 2012). The dynamic nature of S1 can be explained by a higher-resolution structure focusing on S1's interaction with S2, revealing a short 86 amino acid segment of S1's N-terminal domain folding into a helical element that is connected by a hinge region to M1 (Byrgazov et al., 2015; Loveland & Korostelev, 2018). To-date the most interesting snap-shot of S1 on the ribosome was obtained by the Korostolev group through ensemble cryo-EM (Loveland & Korostelev, 2018). In this structure the protein-binding domain is well-resolved (Figure 4.2b), and highlights the dynamic nature of S1's protein-binding domain. In the complex M1 is able to adopt multiple positions that differ by up to 50 Å, while M2 binds in close proximity to the 5' end of the mRNA. Unfortunately, the resolution is still not sufficient to assign domain identity or any details of the interaction between S1's RNA-binding domains with the respective mRNA present in the complex, which only contains four nucleotides upstream of an SD element. What can be concluded from this however, is that several of their maps reveal smaller globular density at the mRNA exit channel consistent with one of S1's four C-terminal domains (Loveland & Korostelev, 2018). Future studies should aim to elucidate the elusive RNA-binding domain of S1 while interacting with various mRNA transcripts to identify the mechanism of RNA recruitment and binding.

4.5 Functions of Ribosome-Bound S1

4.5.1 S1's Essential Role in Translation Initiation

The prokaryotic ribosome contains 52 r-proteins that accompany the 16S, 5S, and 23S rRNAs to form the mature 70S ribosome. S1 is one of the most highly studied r-proteins and has a well-established role in translation initiation, however many details of its structure and cellular mechanism still remain elusive. Early studies underlined the

importance of S1 by demonstrating a significant decrease in the translation of both natural and synthetic mRNAs by ribosomes devoid of S1 (Tal, Aviram, Kanarek, & Weiss, 1972; Van Dieijen, Van Der Laken, Van Knippenberg, & Van Duin, 1975; Van Duin & Kurland, 1970). This information, combined with the observation that S1 binds in close proximity to the 3' end of the 16S rRNA (Dahlberg, 1974; Sillers & Moore, 1981), and has a strong affinity for RNA (Draper, Pratt, & von Hippel, 1977; Kalapos, Paulus, & Sarkar, 1997; Subramanian et al., 1981) led to the hypothesis that S1 aids in accommodating mRNA binding to the 30S to form the 30S initiation complex. The current model for S1's function explains the elongated bi-domain architecture of the protein, where the N-terminal domain anchors S1 to the ribosome via interaction with S2 (Boni, Zlatkin, & Budowsky, 1982; Byrgazov et al., 2012) and the flexible C-terminal domain extends away from the ribosome to capture mRNAs for translation (Suryanarayana & Subramanian, 1983). The above model describes S1 as effectively increasing the affinity of the ribosome for mRNA, suggesting that mRNA can be translated in the absence of S1, albeit at a lower rate. This effect has in fact been demonstrated with poly(U) and poly(A) mRNAs (Van Dieijen et al., 1975), yet there are some mRNAs that are not translated in the absence of S1, such as the structured MS2 phage mRNA. When MS2 is treated with formaldehyde to relieve its structure, S1 becomes dispensable (van Dieijen, van Knippenberg, & van Duin, 1976); this early evidence suggested that S1 increases an mRNAs accessibility to the 16S rRNA by removing inhibitory secondary structure, acting as a helicase.

S1's helicase activity has been demonstrated in multiple studies (Bear et al., 1976; Kolb, Hermoso, Thomas, & Szer, 1977; Szer, Hermoso, & Boublik, 1976; Thomas, Boublik, Szer, & Subramanian, 1979). Optical tweezer experiments have been used to

analyze the mechanism of RNA unwinding by S1, revealing that S1 melts RNA structure by binding to single stranded regions that are transiently formed during thermal breathing of RNA bps, preventing the strands from reannealing (Qu et al., 2012). This behavior has been shown to occur on the ribosome, and is essential for the translation of mRNAs containing a high degree of secondary structure within their ribosome binding site (RBS). Specifically, the RBS of *rpsO* contains a pseudoknot structure that prevents accommodation onto the 30S unless S1 is present to unwind its structure (Duval et al., 2013). It is now understood that S1 is absolutely critical for the translation of structured mRNAs in *E. coli*.

In addition to its essential role in the translation of structured mRNAs, S1 can also regulate translation efficiency by recognizing enhancer elements on mRNAs. These are often U- or AU-rich regions upstream of the SD element that increases the affinity of the transcript for S1 (Boni et al., 1991; Komarova et al., 2002; Tzareva et al., 1994). These AU-tracts upstream of an SD element have been shown to result in a significant increase in the translation of a reporter gene, regardless of SD element strength (Komarova et al., 2002). This in turn suggests that S1's affinity for such enhancer elements contributes significantly to translation initiation. More recent evidence suggests that in addition to upstream AU-rich regions, S1 also recognizes these regions downstream of the start codon in position 4–9 (Cifuentes-Goches et al., 2019). Interestingly, S1 is able to enhance translation of mRNAs lacking both SD and AU elements. Plant viral transcripts devoid of G residues, and unable to form continuous pairing with the aSD were used to understand the role of S1 in recruiting mRNAs lacking a SD element (Tzareva et al., 1994). In this study S1 binding sites other than the typical U- and AU-tracts were identified, including

CAA repeats. The ability of S1 to recognize diverse RNA sequences present in the 5' UTRs of mRNA transcripts partly explains the ability of *E. coli* ribosomes to translate

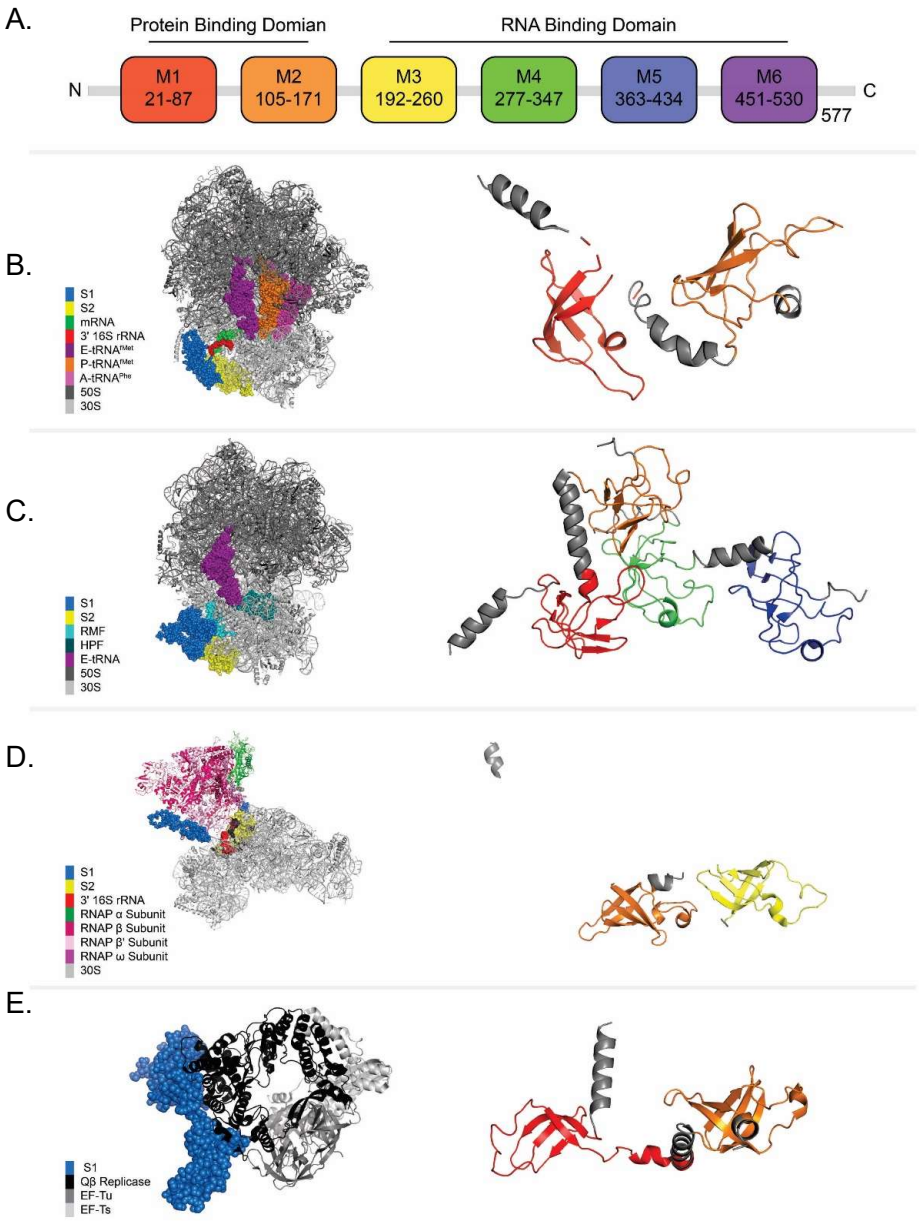


Figure 4.2 High-resolution structures of complexes containing S1. (A) Domain architecture. Color scheme is used in the subsequent cartoon representations of S1. (B) 70S ribosome (Ensemble cryo-EM, PDB 6BU8). (C) Hibernating 70S ribosome (Single-particle cryo-EM, PDB 6H4N). (D) RNA Polymerase – 30S Complex (Single-particle cryo-EM, PDB 6AWB). (E) Qβ replicase (Crystal structure, PDB 4Q7J).

messages that lack a SD element, as opposed to ribosomes from gram-positive bacteria, which lack S1 and rely more heavily on SD elements to initiate translation (Isono & Isono, 1976).

The only classes of mRNA transcripts that do not require S1 are those with a strong SD element (Farwell et al., 1992), and leaderless mRNAs. Leaderless mRNAs are proposed to initiate directly onto 70S ribosomes, and show no requirement for S1 (Moll, Grill, Grundling, & Blasi, 2002; Moll et al., 1998; Tedin, Resch, & Blasi, 1997). Although S1's essentiality during initiation is well established, many questions remain unanswered: Is S1 able to preferentially recognize specific mRNAs under different cellular conditions, allowing for translational control during stress? And if yes, what is the recognition element? Additionally, it is unknown whether S1's protein binding domain plays additional roles besides serving as an anchor to the ribosome. There is a possibility that additional contacts are made with ribosome-bound mRNA that may not be apparent from *in vitro* studies currently in the literature. A high-resolution structure of S1 bound to the ribosome in complex with different mRNAs could answer these questions. Additionally, does S1's activity as a translational repressor have any physiologically relevant roles (other than for *rpsA*) or is it an in adverted consequence of its ability to bind mRNA? S1 was initially studied under the name interference factor i, named for its activity as a translation repressor (Jay & Kaempfer, 1974). It has only later become evident that S1 primarily contributes to translation initiation and is only inhibitory when in excess (Van Dieijen et al., 1975). How often does this role as a repressor come into play in the cell?

4.5.2 Evidence for a Role During Translation Elongation

S1 is present in near-stoichiometric amounts on polysomes, indicating that it does not dissociate after the initiation step of protein synthesis but remains associated with actively translating ribosomes (van Knippenberg, Hooykaas, & van Duin, 1974). This raises the question of a possible role during elongation, however very few studies have addressed this question. One study reports that S1 is required for poly(U)-dependent binding of Phe-tRNA to the ribosome using nitrocellulose filtration assays (Linde, Quoc Khanh, Lipecky, & Gassen, 1979). However, poly(A)-dependent binding of Lys-tRNA did not require S1, despite its requirement during the translation of a poly(A) mRNA (Linde et al., 1979). The authors of this study hypothesized that S1 has a defined function in the elongation cycle, stabilizing codon-anticodon interactions, mimicking its role in initiation where it stabilizes mRNA interactions with 16S rRNA. Additionally, S1's effect on the misreading of poly(U) and poly(dT) transcripts was analyzed to investigate S1's involvement in tRNA selection during elongation (Potapov & Subramanian, 1992). It was shown that S1 confers decreased misreading on poly(U) transcripts in *in vitro* translation assays. These studies on polynucleotide transcripts suggest a role for S1 in elongation as a mediator of codon-anticodon interactions. An *in vivo* study analyzed the effect of reduced cellular S1 concentration, providing further evidence for S1's role during elongation. The elongation rate of two mRNAs (encoding DnaK and EF-G) decreased in the absence of S1 (Sorensen et al., 1998). This role is not essential, as some classes of mRNAs are translated in the absence of S1 as discussed above. Additional studies are required to understand the specifics of S1's role during elongation.

4.5.3 S1 During Stress: 100S Formation

Recently, the cryo-EM structure of the hibernating 100S ribosome revealed an inactive state of S1 (Beckert et al., 2018). The 100S ribosome is formed under stress conditions to lower the rate of translation in the cell and is comprised of two 70S ribosomes bound by their 30S. Formation of the 100S ribosome required the action of the two proteins HFP and RMF (Ueta et al., 2008; Wada, Yamazaki, Fujita, & Ishihama, 1990). Because of S1's large size and flexibility, it is difficult to visualize using current ensemble high-resolution methods, particularly its flexible C-terminal RNA-binding domain. However, in the 100S complex S1 seems to adopt a much more compact conformation, making it easier to resolve (Figure 4.2C). In this structure, S1 is only 100 Å across compared to the 230 Å measured on the 70S ribosome (Beckert et al., 2018). This compact conformation can be attributed to the C-terminal RNA binding domain folding back onto the ribosome rather than extending away as it usually does to capture mRNAs for translation. It appears that the M6 domain actually extends from one 70S ribosome to the other in the 100S complex, where it interacts with ribosomal protein S10. Furthermore, the authors observe that this conformation of S1 appears to be stabilized by M4 directly interacting with both RMF and the 3' end of 16S rRNA, suggesting that S1 is an important player in the inactivation of *E. coli* ribosomes under stress.

4.6 Functions of Non-Ribosome Bound S1

S1 is the most loosely associated r-protein of the *E. coli* ribosome, with a stoichiometry lower than 1:1 in purified ribosomes (Voynow & Kurland, 1971). Presumably, S1 has multiple functions in the cell that require it to be free to interact with multiple substrates and complexes. S1's complex domain architecture lends to its non-ribosomal functions, as it is often used as an RNA-recognition protein that improves the

affinity and/or activity of other proteins that interact with RNA (Hajnsdorf & Boni, 2012).

4.6.1 S1 Activates Transcriptional Cycling

S1 aids in transcription in *E. coli* by stimulating the activity of a DNA-dependent RNA polymerase (Sukhodolets & Garges, 2003; Sukhodolets et al., 2006). This activity was first proposed by Sukhodolets *et al*., when they observed that S1 co-purifies with the *E. coli* RNA polymerase in stoichiometric amounts. Subsequent interaction experiments demonstrated that this binding is specific (Sukhodolets & Garges, 2003). Additionally, purified S1 was added to *in vitro* transcription reactions, and a significant increase in transcript levels was observed (Sukhodolets & Garges, 2003). To investigate the specific role of S1 in this process, the ability of an S1 mutant lacking two of its C-terminal RNA-binding motifs to activate transcriptional cycling was tested and found to be inactive (Sukhodolets et al., 2006). Interestingly, the truncated protein was still able to bind to mRNA, suggesting a more complex mechanism than simple RNA sequestration. The current model for S1's involvement in transcriptional cycling is that the cooperative interaction of S1's RNA-binding motifs enables the release of mRNA from RNAP to allow for continuous reinitiation of transcription.

Although this particular function of S1 is understudied, investigating the contribution of each individual RNA-binding domain during transcriptional activation will help elucidate S1's other cellular functions. Could various domains be involved in sensing particular mRNAs during transcriptional cycling, as opposed to simply releasing the transcript?

In addition to transcriptional cycling, S1 appears to have a role in co-transcription/translation (Demo et al., 2017). A single particle cryo-EM complex of RNA polymerase bound to the 30S subunit shows that the RNA exit tunnel of the polymerase lies next to the mRNA binding site of the ribosome with S1 positioned around this opening (Demo et al., 2017) (Figure 4.2D) – the authors suggest that S1 could function to direct the mRNA coming from the polymerase through to the mRNA binding site, assisting in positioning the mRNA in the proper position on the ribosome.

4.6.2 S1 and tmRNA Translation: Essential or Not Essential?

The tmRNA-mediated trans-translation system in bacteria is a ribosomal rescue system that bypasses the normal translation cycle. A tmRNA molecule mimics both a tRNA and mRNA and is able to release the growing peptide from stalled ribosomes and attach a degradation tag to the translated polypeptide (Karzai et al., 2000). S1 co-purifies with and crosslinks to tmRNA, leading to the belief that it is involved in this function (Bordeau and Felden, 2002; Karzai and Sauer, 2001; Wower et al., 2000). However, its critical role in trans-translation has been called into question (Qi et al., 2007). Even if it is not essential in this process, it is possible that S1 has a stimulatory effect.

4.6.3 The Relationship Between S1 and mRNA Stability

The processes of translation and mRNA stability/decay have a strong influence on each other, yet the specifics of this interconnection are not well understood. A highly translated mRNA in a polysome is protected from degradation; therefore, there is a direct link between factors that influence the translation efficiency of an mRNA and its stability. In bacteria, mRNA decay pathways often begin with an initial cleavage by RNase E at the

5' UTR of the mRNA, specifically in AU-rich motifs. AU-rich enhancer element-containing mRNAs are more stable in the cell, resulting in an improved translation efficiency (Komarova et al., 2005). In this scenario, S1 likely increases mRNA stabilization primarily through increasing translation efficiency, but perhaps also by competing with RNase E for mRNA binding sites during translation initiation, preventing mRNA cleavage. However, in general an excess of free-S1 has an inhibitory effect on cell growth and global translation levels, and stabilizes most cellular mRNAs (Briani et al., 2008; Delvillani et al., 2011). When in excess over ribosomes, free-S1 can bind to an mRNA's 5' UTR and prevent its association with the ribosome, while also preventing cleavage by RNase E. S1's autoregulation mechanism likely prevents its concentration from increasing high enough to effect mRNAs other than *rpsA*, but this has not been specifically studied. Perhaps under stress conditions, stabilization by S1 could play a role in halting translation but increasing stability of specific genes.

4.7 Phage Recruitment of S1

4.7.1 S1 is a Component of the Qβ Replicase

The Qβ virus is a positive-sense, single-stranded RNA virus that infects *E. coli*. Its genome is about 4,000 nucleotides long and encodes four proteins; one, the RNA-dependent RNA polymerase is responsible for replicating and transcribing the Qβ viral RNA (Blumenthal, Landers, & Weber, 1972). The Qβ replicase complex is composed of the virus-encoded polymerase (β subunit), as well as the host derived factors: EF-Tu, EF-Ts, and S1 (Blumenthal et al., 1972; Kamen, 1970; Kondo, Gallerani, & Weissmann, 1970; Wahba et al., 1974). The core complex is made of the β subunit, EF-Tu, and EF-Ts, and is active for most templates *in vitro*. The addition of S1 to this complex is essential

for synthesis of the negative-strand RNA from the positive strand Qβ RNA, yet is not required for the synthesis of positive-strand RNA from negative-strand template (Blumenthal et al., 1972).

S1's RNA-binding function is thought to facilitate the specific recognition of positive-strand Qβ template RNA. Biochemical studies have shown that S1 is able to recognize and bind two internal sites of the Qβ positive-strand, and the initiation of replication strongly depends on S1's interaction with one of these sites (Miranda et al., 1997). The crystal structure of the Qβ replicase containing M1–M3 reveals that S1's two N-terminal protein-binding motifs are responsible for anchoring to the β subunit of the complex, while M3 interacts with an RNA fragment derived from an internal region of the Qβ RNA (Figure 4.2E) (Takeshita, Yamashita, & Tomita, 2014). Thus, it has generally been accepted that S1 facilitates the initiation step of Qβ positive-strand RNA replication. However, this model was called in to question by Vasilyev *et al*., who dispute S1's involvement in the initiation step and suggest that S1 is instead involved in the termination of replication (Vasilyev et al., 2013). If the latter is true, this function is reminiscent of S1's proposed function during transcription in *E. coli*, where it facilitates the release of the RNA transcript (Sukhodolets et al., 2006).

4.7.2 S1 Activates RegB Activity

Regulation of cellular mRNA levels through degradation by endo- and exoribonucleases is an essential process of gene regulation. One endoribonuclease, the T4 phage-encoded RegB, specifically cleaves at the sequence GGAG in early phage mRNAs (Sanson, Hu, Troitskayadagger, Mathy, & Uzan, 2000). RegB has a low affinity for its RNA substrates and requires S1 as a cofactor to efficiently cleave its target (Lebars, Hu,

Lallemand, Uzan, & Bontems, 2001; Ruckman et al., 1994). However, it is not clear how S1 facilitates activation of RegB. It has been shown that S1 does not directly interact with RegB (Uzan, 2001), and likely uses its RNA-binding domain to interact with RegB RNA targets. How it activates degradation of these products remains unclear; does S1 enhance affinity for RegB to its target RNA, or does it alter RNA substrates in order to increase RegB cleavage activity? Durand and coworkers have identified an 11-nucleotide consensus sequence required for S1 activation of RegB (Durand et al., 2006). This sequence begins with GGA, which can be recognized by RegB alone, and is followed by a number of motifs that are recognized specifically by S1. Durand and co-workers then propose a model where S1 enhances RegB activity by recognizing and binding this consensus sequence, inducing a conformational change that exposes the GGA motif to RegB for efficient cleavage. There is a puzzling discrepancy between the large size of S1's RNA-binding domain and the small RNA target sequence. This has spurred the question of whether all four RNA-binding domains are required for efficient binding to the target sequence, and how these domains are organized spatially. NMR and SAXS analysis of different fragments of the S1 RNA-binding domain bound to target RNAs revealed that M4 and M5 are associated with each other to form a continuous RNA-binding surface, while the M3 and M4 domains alternate between an interacting/non-interacting state (Aliprandi et al., 2008). The authors of this study propose that the ability of S1 to bind different target RNAs is due to the adaptability of a common binding surface made up of M3, M4, and M5.

4.8 S1 in Other Organisms

The gram-negative and gram-positive bacteria are generally classified by the makeup of their cell wall. However, these two classes of bacteria differ not only in their cell walls, but also in their mechanisms of gene expression. Gram-negative bacteria require S1 for translation of most mRNAs, whereas gram-positive bacteria generally lack S1. The presence of S1 affords gram-negative bacteria the ability to translate a wider variety of messages that don't necessarily contain a consensus SD element (Farwell et al., 1992), whereas organisms that lack S1 generally require a strong SD-aSD interaction (Band & Henner, 1984; Farwell et al., 1992; Vellanoweth & Rabinowitz, 1992). Some groups of gram-positive bacteria, such as the high-GC content organism *Micrococcus luteus*, have an S1 homolog (Figure 4.3). Recombinant S1 purified from this organism can supplement *E. coli* ribosomes to stimulate translation of an mRNA with a weak SD element, albeit to a lesser degree than purified S1 from *E. coli*, suggesting that S1 from *M. luteus* performs a function similar to *E. coli* S1 (Farwell et al., 1992). Interestingly, S1 from either of these organisms could not stimulate translation by ribosomes from *B. subtilis,* a low-GC content gram-positive organism. These comparative studies on S1 homologs from gram-negative, high-GC gram-positive, and low-GC gram-positive organisms are important for determining the evolutionary history of S1 and of gene expression in general. However, studies on these homologs are very sparse and the available information is sometimes contradictory. For example, low-GC gram-positive bacteria are generally thought to not contain an S1 homolog (Higo, Otaka, & Osawa, 1982; Isono & Isono, 1976), yet phylogenetic analysis reveals S1 homologs in several of these organisms, including *B. subtilis* (Salah et al., 2009). It is possible that the copy of S1 gene present in these organisms does not perform the same function as in *E. coli*, but more biochemical analysis is required to answer this question.

S1 homologues are also present in chloroplasts (Figure 4.3). Despite very low sequence identity (Table 2.1), these homologues seem to function similarly to bacterial S1; they associate with the 30S subunit and bind preferentially U-rich sequences (Merendino, Falciatore, & Rochaix, 2003). Analysis of the S1 domain architecture (Figure 4.3B) reveals that the M4–M6 domains are variable across S1 homologues. This could suggest an evolutionary path where the redundant RNA-binding motifs were lost or gained over time. What functionality does the loss or gain of a certain motif confer to the respective S1 homologues? Future studies analyzing these diverged proteins will contribute to our understanding of each domains function.

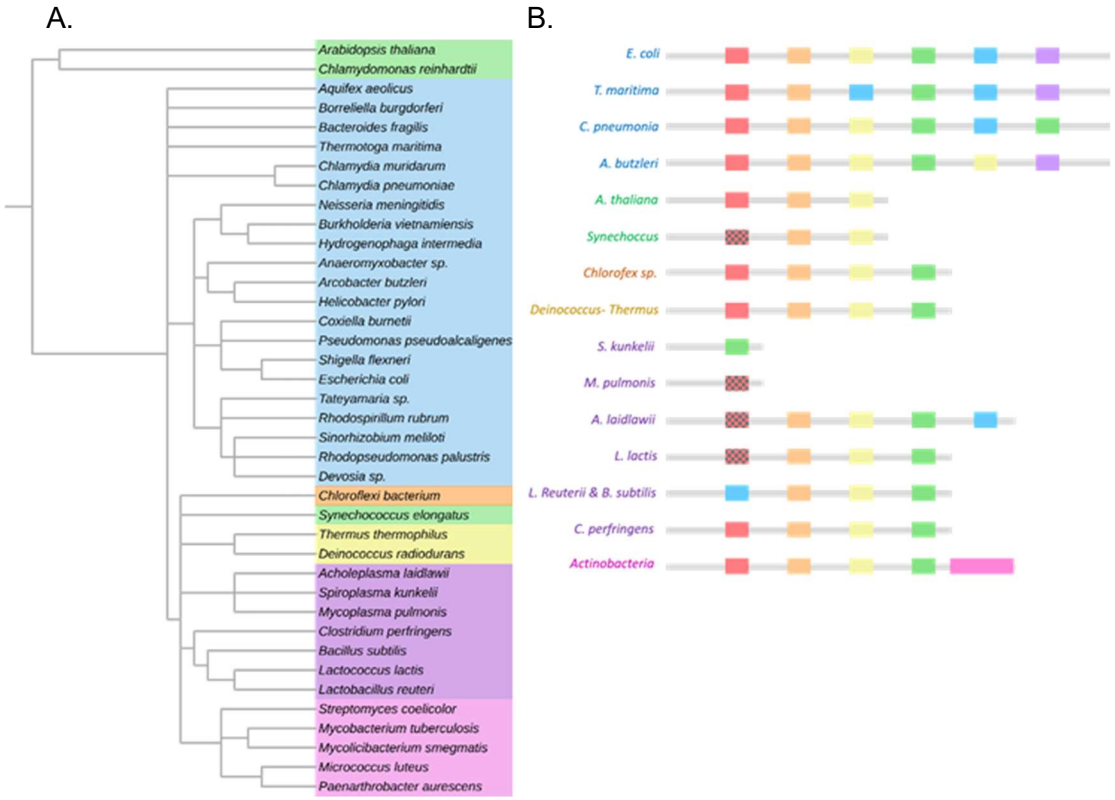


Figure 4.3 Analysis of S1 homologues in different organisms. (A) Phylogenetic tree of organisms containing S1. (B) Domain architecture of representative S1 homologues. Identification of motifs as 1–6 are as designated as in Salah et al. 2009, including the low conservation of particular M1s denoted by shaded pattern. Color scheme: green = eukaryota, blue = gram negative, orange = *chloflexi*, yellow = *Deinococcus thermus*, purple = low GC gram positive, pink = high GC gram positive.

Mycobacterium tuberculosis is the tuberculosis-causing high-GC content gram-positive bacterium. It has been proposed that the S1 homolog found in this organism is the target of the pro-drug pyrazamide (PZA), which is hydrolyzed to pyrazinoic acid (POA) by pyrazinamidase in the cell and shortens tuberculosis treatment time from 1 year to 6 months (Mitchison, 1985). It has been hypothesized that PZA targets the C-terminal extension on *M. tuberculosis* S1 (Figure 4.3B), and inhibits its trans-translational activity (Fan et al., 2017; Shi et al., 2011; Yang et al., 2015). As discussed above, S1's role in trans-translation in *E. coli* has been disputed, and there is no definitive evidence that it carries this function in *M. tuberculosis*. Additionally, there is now convincing evidence from the laboratory of Anthony Baughn that S1 is not the target of PZA. This group replicated and expanded on previous experiments and found that PZA does not target either the process of trans-translation, or S1 (Dillon, Peterson, Feaga, Keiler, & Baughn, 2017).

Table 4.1 Analysis of S1 Homologues. 49 homologues of S1 were analyzed for sequence similarity/identity using EMBOSS water (Madeira et al., 2019). Color scheme: green = eukaryota, blue = gram negative, orange = *chloflexi*, yellow = *Deinococcus thermus*, purple = low GC gram positive, pink = high GC gram positive.

Organism	UniProt Code	Similarity to *E. coli* %	Identity to *E. coli* %	Number of S1 Motifs	Length (aa)
Arabidopsis thaliana (chloroplast)	Q93VC7	46.9	27.9	3	416
Chlamydomonas reinhardtii (chloroplast)	Q70DX8	50.5	29.3	3	436
Synechococcus sp. (cyanobacteria)	P46228	52.1	28.6	3	307
Aquifex aeolicus	O67462	51.0	32.4	6	535
Borrelia burgdorferi	O51153	53.1	30.0	6	551
Bacteroides fragilis	Q5LGN8	58.8	36.5	6	597
Thermotoga maritima	G4FFI0	52.3	30.1	6	543
Chlamydia muridarum	P38016	65.3	44.4	6	570
Chlamydia pneumoniae	Q9Z8M3	65.5	44.0	6	580

Neisseria meningitidis	Q9JZ44	81.7	65.7	6	561
Burkholderia vietnamiensis	A0A103HP97	80.1	63.9	6	570
Hydrogenophaga intermedia	A0A1L1PH09	79.2	63.8	6	561
Anaeromyxobacter	A0A0D6QP89	66.9	47.9	6	568
Arcobacter butzleri	A8EWE3	56.4	32.8	6	550
Helicobacter pylori	P56008	55.5	30.7	6	556
Coxiella burnetii	Q83E09	80.3	64.0	6	551
Pseudomonas pseudoalcaligenes	A0A0F3GDV1	85.9	76.3	6	560
Shigella flexneri	P0AG70	100	100	6	557
Escherichia coli	P0AG67	-	-	6	557
Tateyamaria sp.	A0A0B4CG38	71.3	51.6	6	558
Rhodospirillum rubrum	Q2RXP7	70.9	50.8	6	573
Sinorhizobium melitoti	P14129	69.1	47.8	6	568
Rhodopseudomonas palustris	Q6NDP1	69.7	48.2	6	565
Devosia sp.	A0A0K0U1W8	70.4	48.7	6	550
Chloroflexi bacterium OLB15	A0A136KZW8	46.6	28.8	4	468
Thermus thermophilus	Q83YV9	57.0	37.9	5	536
Deinococcus radiodurans	Q9RSY6	62.0	40.6	5	629
Acholaplasma laidlawii	A9NGK8	50.1	27.1	5	470
Spiroplasma kunkelii	Q6XZ00	56.5	37.0	1 (M4)	111
Mycoplasma pulmonis	Q98R80	53.2	26.0	1 (M1)	111
Clostridium perfringens	Q8XIH0	60.3	35.9	4	378
Bacillus subtilis	P38494	56.3	36.3	4	382
Lactococcus lactis	Q9CHA0	54.5	34.7	4	408
Lactobacillus reuterii	R9WL11	60.0	34.4	4	416
Streptomyces coelicolor	Q9S2K5	56.3	38.2	4 + CTD	502
Mycobacterium tuberculosis	P9WH43	59.5	41.9	4 + CTD	481
Mycobacterium smegmatis	A0QYY6	56.9	38.2	4 + CTD	479
Micrococcus luteus	A0A031GKJ0	54.7	37.4	4 + CTD	485
Arthrobacter aurescens	A1R6F0	54.5	37.4	4 + CTD	491

4.9 Conclusions

Despite the important role that S1 plays for gene expression in *E. coli*, there is not a complete understanding of how it carries out its important role during translation

initiation. Questions that remain include how S1 recruits mRNA to the ribosome, and whether it is bound or not to the 30S during this role. Additionally, does S1 interact differently with its own mRNA while initiating its translation, given that it also has the ability to autoregulate its own expression? To answer these questions, it is critical that high-resolution structures of S1 bound to the ribosome, and in complex with the *rpsA* TIR be obtained to identify the specific interactions made, and to compare these to S1's interaction with other mRNAs on and off the ribosome.

Other functions of S1, such as a possible role in elongation, have not been extensively studied. However, the fact that S1 does not dissociate from the ribosome after completing its role in initiation suggests that it could play a part in elongation as well (van Knippenberg et al., 1974). It would be interesting to determine if S1 is making contacts with the mRNA during translation, and whether it has specificity with regards to certain mRNA sequences. Perhaps a high-resolution structure of a translating ribosome containing S1 could answer these questions.

Apart from its role as a ribosome-associated protein, S1 performs multiple functions by itself, or in complex with other non-ribosomal proteins. S1 has been shown to activate transcriptional cycling by releasing mRNA from RNAP (Sukhodolets et al., 2006), and it may be involved in tmRNA translation, although there is debate as to whether this function is relevant *in vivo* (Qi et al., 2007), S1 is recruited by phages to help carry out processes during infection (Miranda et al., 1997), and also activates the T4 phage-encoded RegB endoribonuclease, allowing RegB to cleave mRNA leader sequences (Ruckman et al., 1994). S1's unique structure allows it to bind proteins and RNAs simultaneously, however it is not well understood how specific protein/RNA are identified as substrates by S1, while also remaining a relatively general RNA binding

protein. It seems likely that S1's many RNA-binding motifs (M4–M6) help recognize and bind specific RNA substrates and may work cooperatively. Learning more about the specific role of S1's motifs will help us better understand S1's mechanism of action, and to assess its potential as target for drug development. Additionally, a better understanding of S1's mechanism will allow synthetic biologists to better control translation in bacteria, which is necessary for the ability to accurately fine-tune gene expression.

Chapter 5

The RNA Interactome of Global Regulator Ribosomal Protein S1

5.1 Introduction

RNA plays important structural, regulatory, and enzymatic roles in all organisms (Isaacs et al., 2004). Prokaryotic organisms exploit RNA as an efficient platform to modulate cellular behavior and regulate gene expression. RNA-binding proteins are powerful tools for exerting control over RNA-mediated regulatory processes, as they can alter networks of deeply connected interactions. These proteins add an additional layer of regulation, modulating RNA structure and availability throughout all stages of its cellular lifecycle. RNA-binding proteins govern RNA levels by enhancing or attenuating RNA stability, and commonly interact with the untranslated regions (UTRs) of mRNAs, or non-coding RNAs, to alter translation rates. Current efforts focus on understanding and mapping these protein-RNA interactions, which can either interact promiscuously with many different RNAs, specifically with a small population of RNAs, or a combination of both.

An important member of this class of regulators is ribosomal protein S1 (S1). S1 is an essential protein found in gram-negative bacteria (Tchufistova et al., 2003) with diverse roles involving translation, transcription, trans-translation, mRNA degradation, polyadenylation, and viral RNA synthesis (Delvillani et al., 2011; Saguy, Gillet, Skorski, Hermann-Le Denmat, & Felden, 2007; Sorensen et al., 1998; Usui, Ichihashi, Kazuta, Matsuura, & Yomo, 2014). It is composed of six homologous OB-fold domains (72–75 amino acid residues each) that are not functionally equivalent (Duval et al., 2013; Lauber et al., 2012). These functionally diverse domains give S1 both protein- and RNA-binding characteristics. Two N-terminal domains give S1 the ability to form protein-protein interactions with the ribosome near the anti-SD motif of the 16S rRNA, anchoring itself to ribosomal protein S2 (Byrgazov et al., 2015; Lauber et al., 2012). The four C-terminal

domains harbor RNA-binding activity, giving it the ability to interact with a broad set of RNAs including poly(U), poly(C), and poly(A). Its ability to interact with the 30S via protein-protein interactions coupled with its RNA-binding activity has made it an essential part of the translation machinery in γ-proteobacteria such as *E. coli*. S1 is essential for recruiting and accommodating structured mRNAs to the 30S and melting their respective structure (Qu et al., 2012). However, it also plays a critical role in the translation of most cellular mRNAs (Sorensen et al., 1998). S1 possesses a higher affinity for U or A/U regions of mRNAs (Boni et al., 1991), which can act as translational enhancers when located in the translation initiation region (J. Zhang & Deutscher, 1992). The affinity of S1 for these enhancer regions has been shown to positively correlate with translation efficiency (Hook-Barnard, Brickman, & McIntosh, 2007; Komarova et al., 2005; Komarova et al., 2002). Directed evolution studies have also been performed on S1, resulting in an increased ability of GC rich mRNAs to be translated by *E. coli* ribosomes (Bernstein, Bulter, Shen, & Liao, 2007). Despite S1's broad range of mRNA substrates, its binding is highly specific to its own mRNA. The translation initiation region (TIR) of the *rpsA* gene, coding for S1, extends far upstream of the AUG start codon and folds into a secondary structure composed of three stem-loops, and surprisingly lacks a canonical Shine-Dalgarno (SD) sequence (Skorski et al., 2006). When S1 is in excess over ribosomes, the unbound S1 can recognize the *rpsA* TIR, modulate its structure, and reduce its ability to be translated effectively down-regulating the production of S1 (Boni et al., 2001).

Although its importance as a translational regulator can't be understated, relatively little is known about additional possible functional and regulatory roles of S1. Additionally, little is known about the critical role S1 plays during initiation and how it

"chooses" which mRNA to recruit. S1 is highly structurally flexible, and is the ribosomal protein most loosely associated with the 30S; consequently, it is absent from all high-resolution X-ray structures of ribosomes. A lack of understanding of S1's role in mRNA recruitment during translation initiation has hindered the development of more accurate prediction of mRNA translation rates and bioengineering tools such as the RBS Calculator (Komarova et al., 2005; Salis et al., 2009). Up until now mapping S1's extensive network of RNA-binding partners has been limited to bioinformatic predictions (Boni et al., 2001), SELEX studies (Ringquist et al., 1995), and a few in-depth biochemical studies on specific mRNAs (Duval et al., 2013). An increased understanding of this mRNA regulator will enhance both our ability to design (forward engineer) synthetic TIRs accurately, and to predict (reverse engineer) the translation initiation rates of natural TIRs.

In the past immunoprecipitation techniques followed by cDNA sequencing have been successfully used to characterize several RNA-binding proteins, however these techniques sometimes generate high numbers of false-positives and do not determine the exact location of the protein-binding site (Moore & Silver, 2008). Recent advances incorporating ultraviolet light (UV) induced RNA-protein crosslinking and RNA-sequencing allows for stringent purification steps reducing false-positives, while mistakes during cDNA preparation generates protein binding maps up to single-nucleotide resolution (Holmqvist et al., 2016; Licatalosi et al., 2008). Here, I combine crosslinking immunoprecipitation and high-throughput RNA sequencing (HITS-CLIP) to generate a high-resolution map of the RNA interactome of S1. The high-resolution RNA-binding map provides important insight into S1's RNA binding specificity. Additionally, the obtained data demonstrates that S1 does not only interact with mRNAs, but also with

small non-coding RNAs such as GcvB, indicating a previously overlooked layer of regulation.

5.2 Materials and Methods

5.2.1 Crosslinking and Immunoprecipitation

Crosslinking was performed as described in (Holmqvist et al., 2016). *E. coli* MG1655 expressing the chromosomally 3×-FLAG tagged S1 were grown in LB media at 37 °C, 220 rpm to an $OD_{600\ nm}$ of ~0.5 in triplicate. The cell suspension was split and irradiated one at a time in a 22 × 22 cm tray with UV-C light at 800 mJ/cm^2. Cells were pelleted in 50 mL fractions by centrifugation for 30 min at 3,000 *g* at 4 °C and flash frozen in liquid N_2. The cell pellets were thawed on ice and resuspended in 800 µL NP-T buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 0.05 % Tween 20) and mixed with 1 mL glass beads (0.1 mm radius). The cell suspension was mixed with 100 µL lysis mix (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 0.05 % Tween 20, 0.5 mg/mL lysozyme, 0.1 U/µL DNase I (Thermo Scientific)). Cells were lysed by shaking at 30 Hz for 10 min. The cell suspension was then centrifuged for 15 min at 16,000 *g* at 4 °C, and the lysate transferred to new tubes and centrifuged again. The cleared lysates were mixed with 1 mL NP-T buffer with 8 M urea and incubated for 5 min at 65 °C while shaking at 900 rpm. The mix was then diluted 10 X in ice-cold NP-T buffer and incubated with 15 µL anti-FLAG magnetic beads (Sigma) pre-washed and equilibrated with NP-T buffer, and incubated while rotating for 75 min at 4 °C. The bead suspension was then centrifuged at 800 *g* for 1 min and resuspended in 1 mL NP-T buffer. The supernatant was decanted and the beads washed twice with a high-salt buffer (50 mM NaH_2PO_4 pH 8.0, 1 M NaCl, 0.05 % Tween20) and twice with NP-T buffer. Beads were resuspended in 100 µL NP-T buffer

containing 1 mM $MgCl_2$ and 12.5 U benzonase nuclease (Sigma Aldrich) and incubated for 10 min at 37 °C while shaking at 800 rpm, and incubated on ice for 2 min. The beads were then washed twice with both high-salt buffer and CIP buffer (50 mM Tris–HCl pH 7.4, 10 mM $MgCl_2$, 100 mM NaCl), and resuspended in 100 µL CIP buffer containing 10 U calf intestinal alkaline phosphatase (New England Biolabs) and incubated for 30 min at 37 °C while shaking at 800 rpm. The beads were then washed once with high-salt buffer and twice with PNK buffer (50 mM Tris HCl pH 7.4, 10 mM $MgCl_2$, 0.1 mM spermidine). The beads were then resuspended in 100 µL PNK buffer containing 10 U of T4 polynucleotide kinase (Thermo Scientific) and 10 µ a γ-^{32}P-ATP and incubated for 15 min at 37 °C. The beads were then washed twice with NP-T buffer, resuspended in 30 µL protein loading buffer (0.3 M Tris–HCl pH 6.8, 0.05 % bromophenol blue, 10 % glycerol, 7 % DTT) and incubated for 5 min at 95 °C. The supernatant was separated from the magnetic beads using a DynaMag-2 (Thermo Scientific) separator. The samples were loaded on a 12 % SDS polyacrylamide gel and run at 340 mA for 2 h. The RNA–protein complexes were transferred to a nitrocellulose membrane for 90 min with 340 mA at 4 °C, and exposed to a phosphor screen overnight. The autoradiogram was used as a template to cut out the labelled RNA–protein complexes from the membrane. The membrane pieces were diced into smaller pieces, and mixed with 400 µL PK buffer (50 mM Tris HCl pH 7.4, 75 mM NaCl, 6 mM EDTA, 1 % SDS) with 10 U of SUPERaseIN (Life Technologies) and 1 mg/mL proteinase K (Thermo Scientific) and incubated for 1 h at 37 °C while shaking at 1,000 rpm in LoBind tubes (Eppendorf). 100 µL 9 M urea was then added to the membrane suspension and the incubation was continued for one additional hour. The supernatant from this mix was added to 450 µL phenol:chloroform:isoamyl alcohol in a phase-lock tube (QuantaBio) and incubated for 5

min at 30 °C while shaking at 1,000 rpm. The aqueous phase was isolated via centrifugation for 12 min at 16,000 *g* at 4 °C and the RNA precipitated using 3 vols. of ice-cold ethanol, 1/10 vols. of 3 M NaOAc pH 5.2, and 1 µL of GlycoBlue (Life Technologies) in LoBind tubes (Eppendorf). The precipitate was pelleted by centrifugation at 16,000 *g* for 30 min at 4 °C, washed with 80 % ethanol, and the centrifuge repeated. The supernatant was decanted and the pellet dried at room temperature, and resuspended in 12.5 µL nuclease-free water (New England Biolabs) by shaking at 65 °C for 2 min while shaking at 100 rpm.

5.2.2 Library Preparation and RNA-Sequencing

The isolated RNA samples were converted to cDNA libraries for Illumina sequencing using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs) according to the manufacturer's instructions. 2.5 µL of the purified RNA (or sterile water as negative control) was mixed with 0.5 µL 3' SR Adaptor (diluted 1:10) and incubated for 2 min at 70 °C and incubated on ice. 5 µL of the 3' ligation reaction buffer and 1.5 µL 3' ligation enzyme mix were added to the samples and incubated for 60 min at 25 °C. 0.25 µL SR RT primer and 2.5 µL nuclease-free water were added to the reaction and incubated for 5 min at 75 °C, 15 min at 37 °C, and 15 min at 25 °C. A 1:1 mix of 5' SR adaptor and nuclease free water was denatured by incubating for 2 min at 70 °C and 30 min on ice. 0.5 µL of the denatured SR adaptor and 0.5 µL 10x ligation reaction buffer, and 1.25 µL ligation enzyme mix was added to the reaction and incubated for 60 min at 25°C. The RNA was reverse transcribed by adding 4 µL first strand synthesis reaction buffer, 0.5 µL murine RNase inhibitor, and 0.5 M MulV Reverse Transcriptase and incubated at 50 °C for 60 min. The enzyme was heat inactivated by a

15 min, 70 °C incubation. The resulting cDNA was PCR amplified by mixing 5 µL cDNA, 12.5 µL LongAmp Taq 2× PCR master mix, 0.625 µL SR primer, and 6.25 µL nuclease-free water and incubating at: 94 °C for 30 s, followed by 18 cycles of 94 °C for 15 s, 62 °C for 30 s, and 70 °C for 15 s. The PCR products were purified using spin-columns (QIAGEN), eluted in 10 µL nuclease-free water, and loaded on 6 % polyacrylamide gels with 7 M urea. Gels were stained with SYBRGold (Life Technologies), and fragments between 50 and 150 bp were excised from the gels. Elution of DNA fragments was performed in 500 µL DNA elution buffer (New England Biolabs) at 16 °C overnight in a thermomixer at 1,000 rpm followed by EtOH precipitation. Pellets were resuspended in 10 µL nuclease-free water. 2 µL gel-purified DNA was mixed with 25 µL LongAmp Taq 2× PCR master mix, 2 µL each of primer (P5) and (P7), and 19 µL nuclease-free water and amplified using the following program: 94 °C for 30 s, and 6 cycles of 94 °C for 15 s, 60 °C for 30 s, 65 °C for 15 s. PCRs were purified on spin-columns (QIAGEN) and eluted in 15 µL nuclease-free water.

5.2.3 Sequencing and Bioinformatic Analysis

Nine cDNA libraries were pooled equimolar and subjected to paired-end sequencing (2 × 75 bp) on an Illumina MiSeq. To assure high sequence quality, files containing the Illumina paired-end reads in FASTQ format were trimmed independently from each other with a Phred score cutoff of 20 by the program fastq_quality_trimmer from FASTX toolkit version 0.0.13. In the same step, after quality trimming 3'-adapters were trimmed using Cutadapt version 1.7.1 (Martin, 2011) and reads without any remaining bases were discarded. The PCR duplicates were collapsed using FASTUniq (H. Xu et al., 2012). The remaining reads were mapped to the reference genome using

READemption with an 80 % accuracy cutoff and a minimum read length of 10 (Forstner, Vogel, & Sharma, 2014). The uniquely mapped reads were then collapsed. For all analyses related to annotated genomic features such as CDSs, tRNAs, and rRNAs, gene annotations from NCBI were used.

Areas of enriched sequence reads were identified by user SICER which uses a scanning with a sliding window of fixed width (24 nucleotides) on the sequences mapped to the reference genome (S. Xu, Grullon, Ge, & Peng, 2014). All windows with significant enrichment were identified, and merged to neighbouring significant windows (within 24 nts). The resulting peaks were used in further bioinformatics analyses.

5.3 Results & Discussion

5.3.1 Cross Linking Immunoprecipitation of S1

For genome-wide detection of S1 binding targets *in vivo* CLIP-seq was employed (Figure 5.1) (Holmqvist et al., 2016). First *E. coli* MG1655 cells expressing a chromosomally FLAG-tagged S1 were cultured in LB medium into exponential growth phase. Cells were harvested during this phase because *E. coli* cells will contain a high number of actively translating ribosomes. This allows us to capture S1 while recruiting mRNA to the ribosome. Due to the essentiality of the *rpsA* gene in *E. coli* I tested to see if the engineered strain had any growth defects associated with the addition of the FLAG tag to the C-terminus of the protein. The strain containing a FLAG-tagged chromosomal version of S1 grows similarly to the wild type strain, indicating that S1 is still able to perform its essential roles during translation (Figure 5.2A.) S1-RNA complexes were immunoprecipitated in cell-lysates with a monoclonal anti-FLAG antibody. These complexes were then eluted from the antibody, separated by denaturing SDS–PAGE, and

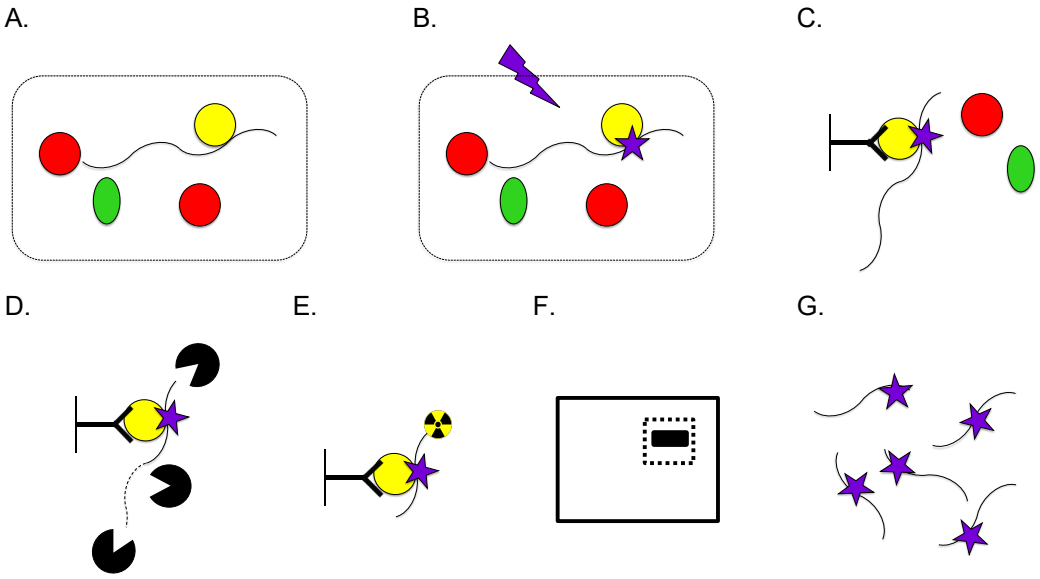


Figure 5.1. CLIP-seq strategy used to identify global binding patterns of RNA binding proteins: (A). cells harbouring an affinity tagged protein of interest (S1) are grown; (B). UV light is applied to the cells to induce protein-RNA crosslink formation; (C).immunoprecipitation of tagged-S1 covalently linked to RNA, followed by a stringent washing to remove contaminants; (D). RNase digestion of bound RNA leaving behind the protected segment; (E). radiolabelling of RNA; (F). SDS-PAGE purification, transfer to membrane, and band excision; (G). RNA isolation and next generation sequencing. Figure adapted from Holmqvist et al., 2016.

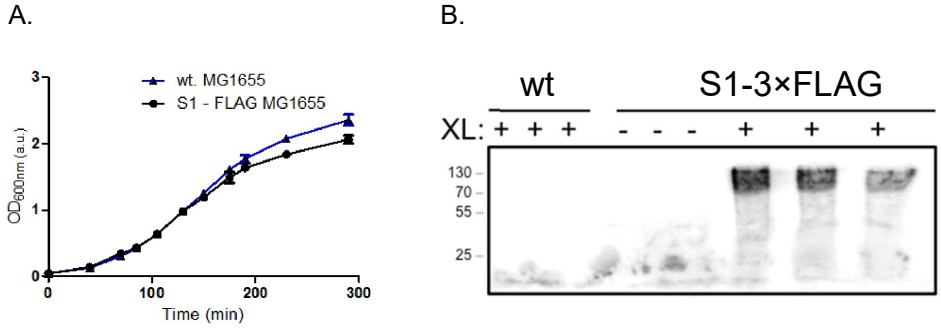


Figure 5.2. CLIP-seq depletes non-cross-linked RNA binding partners. (A). The growth characteristics of *E. coli* MG1655 wt and a variant with a chromosomal modification on the **N-terminal of** the *rpsA* gene (*rpsA-3×flag*) was measured in LB media at 37 °C, 220 rpm. Chromosomal modification does not result in growth defect. (B). Phosphor-image of radiolabelled RNA in complex with S1. The samples result from wt cells exposed to UV light, S1-3×FLAG cells with no UV light exposure, and S1-3×FLAG cells exposed to UV light. XL – cross-linked**.**

transferred to a membrane. A clear radioactive signal corresponding to radiolabelled RNA

was detected in tagged and cross-linked samples, and absent in all other samples illustrating that uncross-linked RNAs were successfully depleted (Figure 5.2B).

5.3.2 Next Generation Sequencing Data Analysis

The resulting cDNA was sequenced using Illumina paired-end sequencing. The RNA-seq data was filtered to ensure high sequence quality. The paired-end reads with a

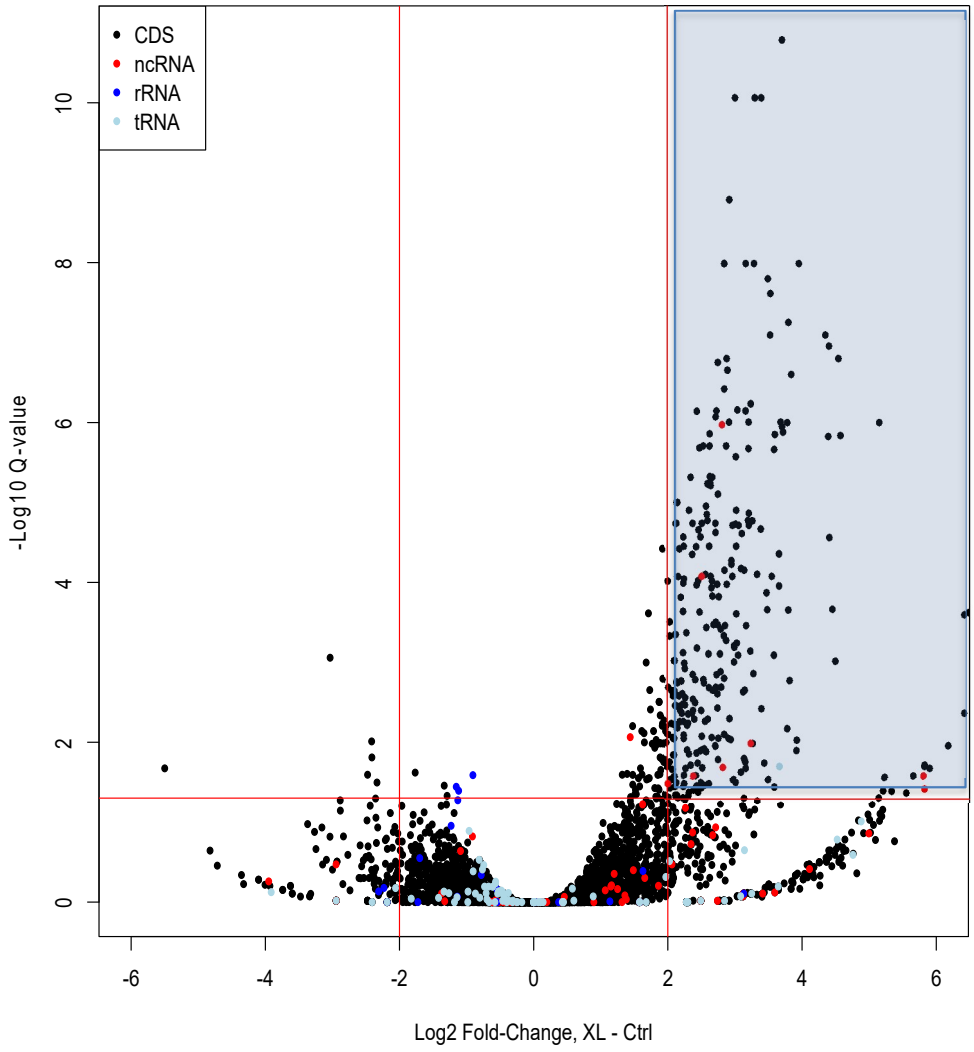


Figure 5.3. Volcano plot of CLIP-seq reads maps the transcriptome wide binding partners of S1. X-axis indicates the fold change in enrichment between cross-linked and non-cross-linked samples y-axis indicates the statistical significance (q-values). There is a clear enrichment in specific RNAs (shaded area).

phred score below 20 were to be discarded, however all sequences were above this quality threshold. The 3' Illumina adaptors were trimmed, and reads without any remaining bases were discarded. The PCR duplicates were collapsed to avoid any biases introduced during library amplification. The remaining reads were mapped to the reference genome using READemption, and reads shorter than 10 nucleotides were discarded (Forstner et al., 2014). The uniquely mapped reads were then collapsed. At each step throughout the filtering process there was a fraction of reads that were discarded (with exception to the phred score filtering).

The mapped data provided a view of S1's RNA binding partners over the entire transcriptome (Figure 5.3). While most of the sequence reads mapped to mRNAs as expected, additional classes of RNAs are represented in our sequencing data (Figure 5.4).

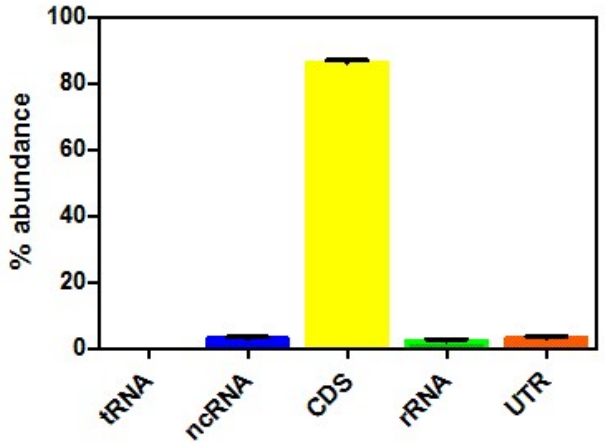


Figure 5.4. Distribution of filtered S1-CLIP reads. Transcript classes reveal diverse RNA binding partners.

Roughly 90 % of the sequence reads map to mRNAs, with ~85 % mapping to protein coding sequences and < 5 % to UTRs. It was previously known that S1 does not dissociate after the initiation step of protein synthesis, but remains ribosome bound throughout the entire translation cycle (van Knippenberg et al., 1974). It has been

proposed that S1's role in translation elongation is to act as a mediator of codon-anticodon interactions (Sorensen et al., 1998). Interestingly, patterns of S1 binding were observed throughout the entire CDS of many genes; this is clear when looking at the *rpsA* transcript (Figure 5.5).

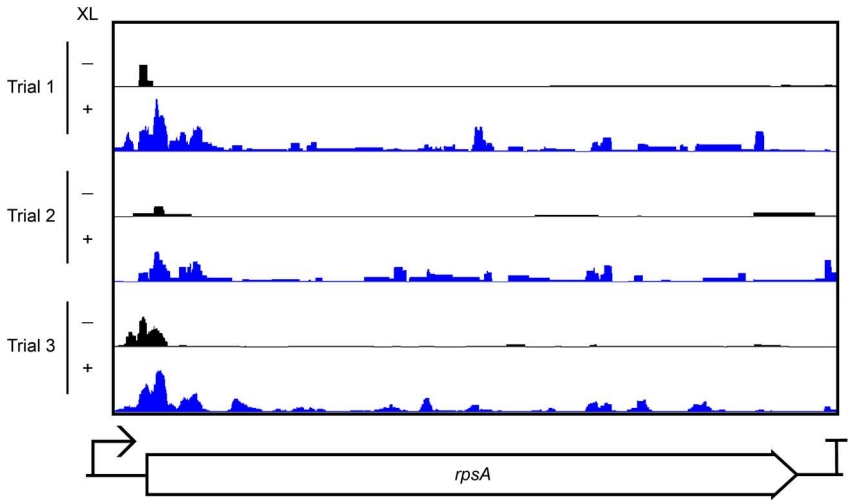


Figure 5.5. During exponential growth phase S1 binds throughout the *rpsA* mRNA. Illumina-Solexa sequencing of S1-bound RNA identified by CLIP-seq. Resulting hits from biological triplicates of cross-linked (blue) and non-cross-linked background control (black) are plotted over a schematic of the *rpsA* gene.

One of the most peculiar features of S1 is the regulatory role it plays in its own synthesis. Ribosome-bound S1 must recruit the *rpsA* mRNA for translation, however free-S1 binds to and represses the translation of its own mRNA. To determine the nucleotides of the *rpsA* TIR that interact with ribosome-bound S1 *in vivo* the reads mapping to the *rpsA* TIR were examined. The highest number of reads map to the most 3' stem-loop, containing the start codon. This observation is consistent with recent chemical probing data that illustrate that this region of the RNA is unfolded *in vivo* (Mustoe et al., 2018). The data clearly shows that S1 is interacting with these nucleotides, and likely is responsible for this unfolding. Interesting, S1 does not interact with the most 5' stem-loop in its translation initiation region (TIR) (Figure 5.6). This binding pattern leaves the

question: what is the purpose of this region of the mRNA, and what it is interacting with to generate the observed high translation initiation strength? The 5' stem- loop is absolutely essential for the efficient translation of this mRNA, however it is too far upstream from the start codon to contribute directly to interactions with the anti-Shine-Dalgarno (aSD) sequence on the 30S subunit (Boni et al., 2000; Skorski et al., 2006). Upstream, unstructured translation enhancer elements are A/U-rich regions known to stimulate translation initiation. However, the 5' region of the *rpsA* TIR differs from these elements as it has an additional G/C rich region able to form two stem-loops *in vivo*.

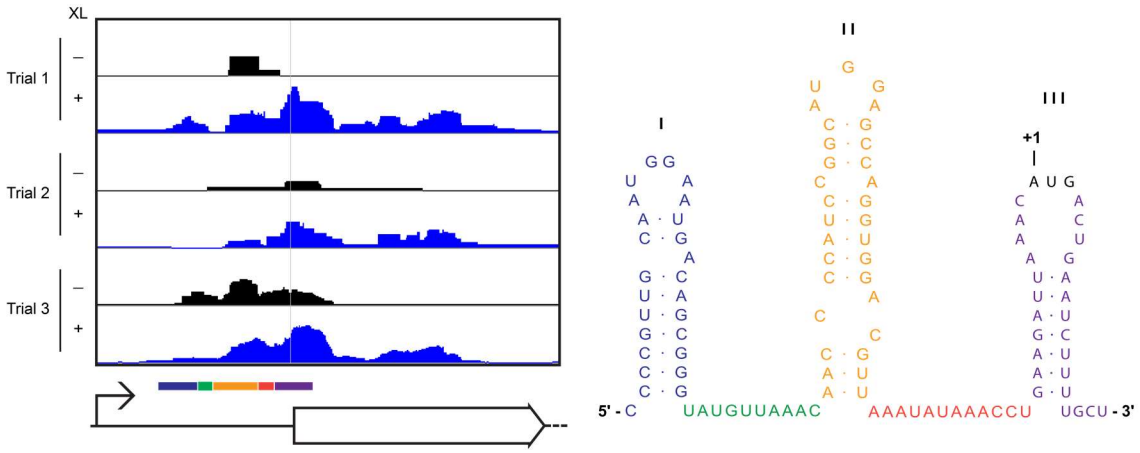


Figure 5.6. During exponential growth phase S1 binds stem-loop II and three of the *rpsA* TIR. Illumina-Solexa sequencing of S1-bound RNA identified by CLIP-seq. Resulting hits from biological triplicates of cross-linked (blue) and non-cross-linked background control (black) are plotted over a schematic of the *rpsA* gene. Coloured bar corresponds to the secondary structure of the *rpsA* TIR on the right.

Although it is unclear how the structured region of this RNA contributes to initiation, these findings provide evidence that this TIR contains a structured RNA translation enhancer element. This region could act to correctly position the area around the start codon for optimal interaction with S1, the aSD, and/or the platform of 30S ribosomal subunit. A high-resolution structure of this RNA on the 30S ribosomal subunit

or the 70S ribosome will be critical in uncovering the complete mechanism of this non-canonical initiation mechanism.

One of our most interesting findings is that ~3 % of reads map to non-coding RNA loci (Figure 5.4). In fact, one of the most highly enriched RNAs in the entire experiment is the widely conserved Hfq-dependent small noncoding RNA (sRNA) GcvB. Enterobacteria have sophisticated regulatory cascades controlling amino acid biosynthesis, and GcvB acts as global regulator responsible for regulating both amino acid transport and metabolism (McArthur, Pulvermacher, & Stauffer, 2006; Sharma, Darfeuille, Plantinga, & Vogel, 2007; Silveira et al., 2010). It regulates over 1 % of genes

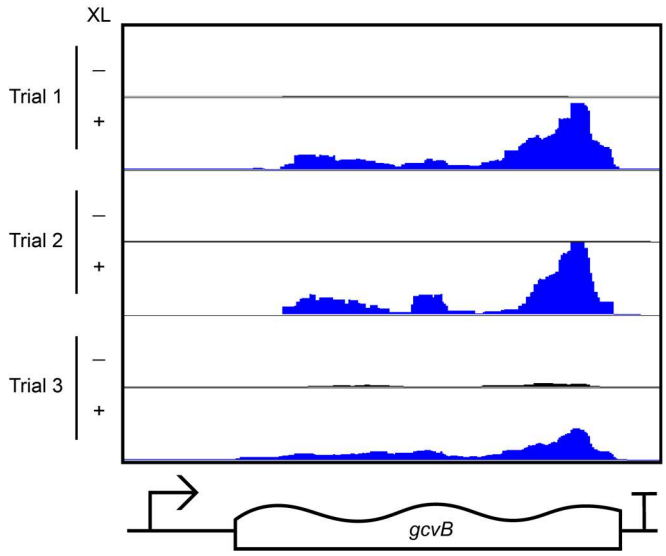


Figure 5.7. During exponential growth phase S1 binds the 3' end of the GcvB sRNA. Illumina-Solexa sequencing of S1-bound RNA identified by CLIP-seq. Resulting hits from biological triplicates of cross-linked (blue) and non-cross-linked background control (black) are plotted over a schematic of the *gcvB* gene.

in *E. coli* and is highly expressed when nutrients are abundant, as in the experimental conditions (Sharma et al., 2011; Vanderpool, 2011). When the *gcvB* locus was examined, sequence reads were found to be enriched over the 3' end. Interestingly, this mirrors the

binding site for Hfq (Holmqvist et al., 2016). While a more in-depth study is required to determine S1's role in the regulatory network of GcvB, it may prove to be an interesting strategy used by the ribosome to directly sense amino acid availability in the cell. S1 could also be protecting this RNA from degradation by RNases, or it could simply be an artifact of its interaction with mRNAs that GcvB is interacting with.

5.4 Conclusions

In this work the entire RNA interactome of S1 was mapped *in vivo* using HITS-CLIP. This important mRNA regulator is shown to be a global regulator, interacting with many different classes of RNA. These maps of S1 binding sites confirm many of its known binding partners at high-resolution. Interestingly, an increased insight into the regulatory mechanisms contributing to the translation of the *rpsA* mRNA is gained. S1 interacts with this mRNA in the region surrounding the start codon, and likely contributes to unfolding RNA structure in this area. Finally, it is clearly shown that S1 interacts with several sRNAs including GcvB—a master regulator of amino acid bioavailability.

Chapter 6

Conclusions

6.1 Concluding Remarks

In this thesis I have described how decades of studying gene expression has uncovered diverse regulation strategies used by living organisms, and discussed how these findings have inspired bioengineers to develop useful synthetic biology-based technologies. As our understanding of the underlying mechanisms of gene expression increases the ability in which forward engineering tools can be developed improved. While tools such as the RBS Calculator for engineering translation initiation, have shown great utility in synthetic gene design, predicting the translation initiation rates of native RNAs remains a challenge. Studying diverse native genes will help undercover the regulatory signals that govern and controlling gene expression.

I combined several different *in vivo*, *in vitro*, and *in silico* approaches to uncover the non-canonical initiation mechanism that governs the translation of S1. This mRNA contains a three-dimensional structure that is critical for efficient translation initiation, similar to how viral IRESs operate. Additionally, the importance of S1 in the translation of the *rpsA* TIR is described. A high-throughput CLIP assay identified where S1 interacts with the *rpsA* mRNA *in vitro*, while emphasizing S1's role as an RNA regulator. S1 interacts with areas throughout entire transcripts, not only in TIRs of structured RNAs. Interestingly, while S1 is a critical regulator of mRNAs, HITS-CLIP identifies binding partners in several different classes of RNAs, including sRNAs.

Obtaining a high-resolution image of the *rpsA* TIR in complex with S1 on the ribosome is critical for determining all the contributing factors to this initiation mechanism. The S1 HITS-CLIP has revealed many new targets that highlight a more sophisticated regulatory role in gene expression than previously thought. A more in-depth biochemical analysis of the relationship between S1 and its binding partners will be

required to fully understand its role as an global regulator of cellular behaviour. Doing so will uncover fundamental principles of gene expression, which will lead to improved accuracy when predicting gene expression in other organisms, while improving forward engineering approaches to controlling gene expression in synthetic systems.

References

Aliprandi, P., Sizun, C., Perez, J., Mareuil, F., Caputo, S., Leroy, J. L., . . . Bontems, F. (2008). S1 ribosomal protein functions in translation initiation and ribonuclease RegB activation are mediated by similar RNA-protein interactions: an NMR and SAXS analysis. *J Biol Chem, 283*(19), 13289-13301.

Auslander, S., & Fussenegger, M. (2014). Synthetic biology: Toehold gene switches make big footprints. *Nature, 516*(7531), 333-334.

Balakin, A. G., Bogdanova, S. L., & Skripkin, E. A. (1992). mRNA containing an extended Shine-Dalgarno sequence is translated independently of ribosomal protein S1. *Biochem Int, 27*(1), 117-129.

Balakin, A. G., Skripkin, E. A., Shatsky, I. N., & Bogdanov, A. A. (1992). Unusual ribosome binding properties of mRNA encoding bacteriophage lambda repressor. *Nucleic Acids Res, 20*(3), 563-571.

Band, L., & Henner, D. J. (1984). *Bacillus subtilis* requires a "stringent" Shine-Dalgarno region for gene expression. *DNA, 3*(1), 17-21.

Bear, D. G., Ng, R., Van Derveer, D., Johnson, N. P., Thomas, G., Schleich, T., & Noller, H. F. (1976). Alteration of polynucleotide secondary structure by ribosomal protein S1. *Proc Natl Acad Sci U S A, 73*(6), 1824-1828.

Beckert, B., Turk, M., Czech, A., Berninghausen, O., Beckmann, R., Ignatova, Z., . . . Wilson, D. N. (2018). Structure of a hibernating 100S ribosome reveals an inactive conformation of the ribosomal protein S1. *Nat Microbiol, 3*(10), 1115-1121.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Bluthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. *Mol Syst Biol, 9*, 675.

Berens, C., Thain, A., & Schroeder, R. (2001). A tetracycline-binding RNA aptamer. *Bioorg Med Chem, 9*(10), 2549-2556.

Bernstein, J. R., Bulter, T., Shen, C. R., & Liao, J. C. (2007). Directed evolution of ribosomal protein S1 for enhanced translational efficiency of high GC *Rhodopseudomonas palustris* DNA in *Escherichia coli. J Biol Chem, 282*(26), 18929-18936.

Bisaglia, M., Laalami, S., Uzan, M., & Bontems, F. (2003). Activation of the RegB endoribonuclease by the S1 ribosomal protein is due to cooperation between the S1 four C-terminal modules in a substrate-dependant manner. *J Biol Chem, 278*(17), 15261-15271.

Blumenthal, T., Landers, T. A., & Weber, K. (1972). Bacteriophage Q replicase contains the protein biosynthesis elongation factors EF Tu and EF Ts. *Proc Natl Acad Sci U S A, 69*(5), 1313-1317.

Boni, I. V., Artamonova, V. S., & Dreyfus, M. (2000). The last RNA-binding repeat of the *Escherichia coli* ribosomal protein S1 is specifically involved in autogenous control. *J Bacteriol, 182*(20), 5872-5879.

Boni, I. V., Artamonova, V. S., Tzareva, N. V., & Dreyfus, M. (2001). Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. *EMBO J, 20*(15), 4222-4232.

Boni, I. V., Isaeva, D. M., Musychenko, M. L., & Tzareva, N. V. (1991). Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res, 19*(1), 155-162.

Boni, I. V., Zlatkin, I. V., & Budowsky, E. I. (1982). Ribosomal protein S1 associates with *Escherichia coli* ribosomal 30-S subunit by means of protein-protein interactions. *Eur J Biochem, 121*(2), 371-376.

Boniecki, M. J., Lach, G., Dawson, W. K., Tomala, K., Lukasz, P., Soltysinski, T., . . . Bujnicki, J. M. (2016). SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res, 44*(7), e63.

Breaker, R. R. (2012). Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol, 4*(2).

Briani, F., Curti, S., Rossi, F., Carzaniga, T., Mauri, P., & Deho, G. (2008). Polynucleotide phosphorylase hinders mRNA degradation upon ribosomal protein S1 overexpression in *Escherichia coli*. *RNA, 14*(11), 2417-2429.

Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Wimberly, B. T., & Ramakrishnan, V. (2002). Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. *J Mol Biol, 316*(3), 725-768.

Brophy, J. A., & Voigt, C. A. (2014). Principles of genetic circuit design. *Nat Methods, 11*(5), 508-520.

Button, J. E., & Galan, J. E. (2011). Regulation of chaperone/effector complex synthesis in a bacterial type III secretion system. *Mol Microbiol, 81*(6), 1474-1483.

Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell, 88*(2), 235-242.

Byrgazov, K., Grishkovskaya, I., Arenz, S., Coudevylle, N., Temmel, H., Wilson, D. N., . . . Moll, I. (2015). Structural basis for the interaction of protein S1 with the *Escherichia coli* ribosome. *Nucleic Acids Res, 43*(1), 661-673.

Byrgazov, K., Manoharadas, S., Kaberdina, A. C., Vesper, O., & Moll, I. (2012). Direct interaction of the N-terminal domain of ribosomal protein S1 with protein S2 in *Escherichia coli*. *PLoS One, 7*(3), e32702.

Callura, J. M., Cantor, C. R., & Collins, J. J. (2012). Genetic switchboard for synthetic biology applications. *Proc Natl Acad Sci U S A, 109*(15), 5850-5855.

Cardinale, S., Joachimiak, M. P., & Arkin, A. P. (2013). Effects of genetic variation on the *E. coli* host-circuit interface. *Cell reports, 4*(2), 231-237.

Caron, M.-P., Lafontaine, D. A., & Massé, E. (2010). Small RNA-mediated regulation at the level of transcript stability. *RNA Biology, 7*(2), 140-144.

Carrera, J., Rodrigo, G., Singh, V., Kirov, B., & Jaramillo, A. (2011). Empirical model and *in vivo* characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnol J, 6*(7), 773-783.

Carrier, T. A., & Keasling, J. D. (1999). Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol Prog, 15*(1), 58-64.

Chae, T. U., Kim, W. J., Choi, S., Park, S. J., & Lee, S. Y. (2015). Metabolic engineering of *Escherichia coli* for the production of 1,3-diaminopropane, a three carbon diamine. *Scientific Reports, 5*, 13040.

Chang, A. L., McKeague, M., Liang, J. C., & Smolke, C. D. (2014). Kinetic and equilibrium binding characterization of aptamers to small molecules using a label-free, sensitive, and scalable platform. *Anal Chem, 86*(7), 3273-3278.

Chang, B., Halgamuge, S., & Tang, S. L. (2006). Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene, 373*, 90-99.

Chen, H., Bjerknes, M., Kumar, R., & Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res, 22*(23), 4953-4957.

Christiansen, L., & Pedersen, S. (1981). Cloning, restriction endonuclease mapping and post-transcriptional regulation of *rpsA*, the structural gene for ribosomal protein S1. *Mol Gen Genet, 181*(4), 548-551.

Cifuentes-Goches, J. C., Hernandez-Ancheyta, L., Guarneros, G., Oviedo, N., & Hernandez-Sanchez, J. (2019). Domains two and three of *Escherichia coli*

ribosomal S1 protein confers 30S subunits a high affinity for downstream A/U-rich mRNAs. *J Biochem*.

Cironi, P., Swinburne, I. A., & Silver, P. A. (2008). Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. *J Biol Chem, 283*(13), 8469-8476.

Colussi, T. M., Costantino, D. A., Zhu, J., Donohue, J. P., Korostelev, A. A., Jaafar, Z. A., . . . Kieft, J. S. (2015). Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature, 519*, 110-113.

Croft, M. T., Moulin, M., Webb, M. E., & Smith, A. G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci U S A, 104*(52), 20770-20775.

Cromie, M. J., Shi, Y., Latifi, T., & Groisman, E. A. (2006). An RNA sensor for intracellular Mg^{2+}. *Cell, 125*(1), 71-84.

d'Aubenton Carafa, Y., Brody, E., & Thermes, C. (1990). Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J Mol Biol, 216*(4), 835-858.

Dahlberg, A. E. (1974). Two forms of the 30S ribosomal subunit of *Escherichia coli*. *J Biol Chem, 249*(23), 7673-7678.

Dar, D., & Sorek, R. (2018). High-resolution RNA 3'-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res, 46*(13), 6797-6805.

de Smit, M. H., & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci U S A, 87*(19), 7668-7672.

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.

Dedkova, L. M., Fahmi, N. E., Golovine, S. Y., & Hecht, S. M. (2003). Enhanced D-amino acid incorporation into protein by modified ribosomes. *J Am Chem Soc, 125*(22), 6616-6617.

Del Campo, C., Bartholomaus, A., Fedyunin, I., & Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet, 11*(10), e1005613.

Delvillani, F., Papiani, G., Deho, G., & Briani, F. (2011). S1 ribosomal protein and the interplay between translation and mRNA decay. *Nucleic Acids Res, 39*(17), 7702-7715.

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A. B., Diaz-Avalos, R., . . . Korostelev, A. A. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife, 6*.

Dillon, N. A., Peterson, N. D., Feaga, H. A., Keiler, K. C., & Baughn, A. D. (2017). Anti-tubercular Activity of Pyrazinamide is Independent of trans-Translation and RpsA. *Sci Rep, 7*(1), 6135.

Ditto, M. D., Roberts, D., & Weisberg, R. A. (1994). Growth phase variation of integration host factor level in *Escherichia coli*. *J Bacteriol, 176*(12), 3738-3748.

Dixon, N., Duncan, J. N., Geerlings, T., Dunstan, M. S., McCarthy, J. E., Leys, D., & Micklefield, J. (2010). Reengineering orthogonally selective riboswitches. *Proc Natl Acad Sci U S A, 107*(7), 2830-2835.

Draper, D. E., Pratt, C. W., & von Hippel, P. H. (1977). *Escherichia coli* ribosomal protein S1 has two polynucleotide binding sites. *Proc Natl Acad Sci U S A, 74*(11), 4786-4790.

Draper, D. E., & Reynaldo, L. P. (1999). RNA binding strategies of ribosomal proteins. *Nucleic Acids Res, 27*(2), 381-388.

Durand, S., Richard, G., Bisaglia, M., Laalami, S., Bontems, F., & Uzan, M. (2006). Activation of RegB endoribonuclease by S1 ribosomal protein requires an 11 nt conserved sequence. *Nucleic Acids Res, 34*(22), 6549-6560.

Durfee, T., Hansen, A. M., Zhi, H., Blattner, F. R., & Jin, D. J. (2008). Transcription profiling of the stringent response in *Escherichia coli*. *J Bacteriol, 190*(3), 1084-1096.

Duval, M., Korepanov, A., Fuchsbauer, O., Fechter, P., Haller, A., Fabbretti, A., . . . Marzi, S. (2013). *Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol, 11*(12), e1001731.

Ellington, A. D., & Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature, 346*(6287), 818-822.

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 Res, 42*(4), 2646-2659.

Espah Borujeni, A., Mishler, D. M., Wang, J., Huso, W., & Salis, H. M. (2015). Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Res*.

Fan, Y., Dai, Y., Hou, M., Wang, H., Yao, H., Guo, C., . . . Liao, X. (2017). Structural basis for ribosome protein S1 interaction with RNA in trans-translation of *Mycobacterium tuberculosis*. *Biochem Biophys Res Commun, 487*(2), 268-273.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., & Salis, H. M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol Syst Biol, 10*, 731.

Farwell, M. A., Roberts, M. W., & Rabinowitz, J. C. (1992). The effect of ribosomal protein S1 from *Escherichia coli* and *Micrococcus luteus* on protein synthesis *in vitro* by *E. coli* and *Bacillus subtilis*. *Mol Microbiol, 6*(22), 3375-3383.

Förster, S., Apostol, L., & Bras, W. (2010). *Scatter: Software for the analysis of nano- and mesoscale small-angle scattering* (Vol. 43).

Forstner, K. U., Vogel, J., & Sharma, C. M. (2014). READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics, 30*(23), 3421-3423.

Fried, S. D., Schmied, W. H., Uttamapinant, C., & Chin, J. W. (2015). Ribosome Subunit Stapling for Orthogonal Translation in *E. coli*. *Angew Chem Int Ed Engl, 54*(43), 12791-12794.

Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I., & Perumov, D. A. (1999). A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends in Genetics, 15*(11), 439-442.

George, K. W., Thompson, M. G., Kang, A., Baidoo, E., Wang, G., Chan, L. J. G., . . . Lee, T. S. (2015). Metabolic engineering for the high-yield production of isoprenoid-based C5 alcohols in *E. coli*. *Scientific Reports, 5,* 11128.

Giorginis, S., & Subramanian, A. R. (1980). The major ribosome binding site of *Escherichia coli* ribosomal protein S1 is located in its N-terminal segment. *J Mol Biol, 141*(4), 393-408.

Giraud, P., Crechet, J. B., Uzan, M., Bontems, F., & Sizun, C. (2015). Resonance assignment of the ribosome binding domain of *E. coli* ribosomal protein S1. *Biomol NMR Assign, 9*(1), 107-111.

Giri, L., & Subramanian, A. R. (1977). Hydrodynamic properties of protein S1 from *Escherichia coli* ribosome. *FEBS Lett, 81*(1), 199-203.

Goelz, S., & Steitz, J. A. (1977). *Escherichia coli* ribosomal protein S1 recognizes two sites in bacteriophage Qbeta RNA. *J Biol Chem, 252*(15), 5177-5179.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S., & Stormo, G. (1981). Translational initiation in prokaryotes. *Annu Rev Microbiol, 35*, 365-403.

Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science, 342*(6157), 475-479.

Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A., Roubos, J. A., & Ignatova, Z. (2016). A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synth Biol. 5*, 710-720.

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: *de-novo*-designed regulators of gene expression. *Cell, 159*(4), 925-939.

Grill, S., Gualerzi, C. O., Londei, P., & Blasi, U. (2000). Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J, 19*(15), 4101-4110.

Grill, S., Moll, I., Hasenohrl, D., Gualerzi, C. O., & Blasi, U. (2001). Modulation of ribosomal recruitment to 5'-terminal start codons by translation initiation factors IF2 and IF3. *FEBS Lett, 495*(3), 167-171.

Grundy, F. J., & Henkin, T. M. (2004). Regulation of gene expression by effectors that bind to RNA. *Curr Opin Microbiol, 7*(2), 126-131.

Grundy, F. J., & Henkin, T. M. (2006). From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol, 41*(6), 329-338.

Guerrier-Takada, C., Subramanian, A. R., & Cole, P. E. (1983). The activity of discrete fragments of ribosomal protein S1 in Q beta replicase function. *J Biol Chem, 258*(22), 13649-13652.

Hajnsdorf, E., & Boni, I. V. (2012). Multiple activities of RNA-binding proteins S1 and Hfq. *Biochimie, 94*(7), 1544-1553.

Hartz, D., McPheeters, D. S., & Gold, L. (1989). Selection of the initiator tRNA by *Escherichia coli* initiation factors. *Genes Dev, 3*(12a), 1899-1912.

Hellen, C. U., & Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev, 15*(13), 1593-1612.

Heppell, B., Blouin, S., Dussault, A. M., Mulhbacher, J., Ennifar, E., Penedo, J. C., & Lafontaine, D. A. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat Chem Biol, 7*(6), 384-392.

Higo, K., Otaka, E., & Osawa, S. (1982). Purification and characterization of 30S ribosomal proteins from *Bacillus subtilis*: correlation to *Escherichia coli* 30S proteins. *Mol Gen Genet, 185*(2), 239-244.

Hirashima, A., & Kaji, A. (1970). Factor dependent breakdown of polysomes. *Biochemical and biophysical research communications, 41*(4), 877-883.

Hirashima, A., & Kaji, A. (1972). Factor-dependent release of ribosomes from messenger RNA: requirement for two heat-stable factors. *Journal of Molecular Biology, 65*(1), 43-58.

Holmqvist, E., Reimegard, J., & Wagner, E. G. (2013). Massive functional mapping of a 5'-UTR by saturation mutagenesis, phenotypic sorting and deep sequencing. *Nucleic Acids Res, 41*(12), e122.

Holmqvist, E., Wright, P. R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., . . . Vogel, J. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J, 35*(9), 991-1011.

Hook-Barnard, I. G., Brickman, T. J., & McIntosh, M. A. (2007). Identification of an AU-rich translational enhancer within the *Escherichia coli* fepB leader RNA. *J Bacteriol, 189*(11), 4028-4037.

Hui, A., & de Boer, H. A. (1987). Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc Natl Acad Sci U S A, 84*(14), 4762-4766.

Hui, A., Eaton, D. H., & de Boer, H. A. (1988). Mutagenesis at the mRNA decoding site in the 16S ribosomal RNA using the specialized ribosome system in Escherichia coli. *The EMBO journal, 7*(13), 4383.

Hui, M. P., Foley, P. L., & Belasco, J. G. (2014). Messenger RNA degradation in bacterial cells. *Annu Rev Genet, 48*, 537-559.

Huttenhofer, A., & Noller, H. F. (1992). Hydroxyl radical cleavage of tRNA in the ribosomal P site. *Proc Natl Acad Sci U S A, 89*(17), 7851-7855.

Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., & Collins, J. J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol, 22*(7), 841-847.

Isono, K., & Isono, S. (1976). Lack of ribosomal protein S1 in *Bacillus stearothermophilus*. *Proc Natl Acad Sci U S A, 73*(3), 767-770.

Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol, 11*(2), 113-127.

Jacob, W. F., Santer, M., & Dahlberg, A. E. (1987). A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc Natl Acad Sci U S A, 84*(14), 4757-4761.

Jan, E. (2006). Divergent IRES elements in invertebrates. *Virus Res, 119*(1), 16-28.

Jang, C. J., Lo, M. C., & Jan, E. (2009). Conserved element of the dicistrovirus IGR IRES that mimics an E-site tRNA/ribosome interaction mediates multiple functions. *J Mol Biol, 387*(1), 42-58.

Jay, G., & Kaempfer, R. (1974). Host interference with viral gene expression: mode of action of bacterial factor i. *J Mol Biol, 82*(2), 193-212.

Jenison, R. D., Gill, S. C., Pardi, A., & Polisky, B. (1994). High-resolution molecular discrimination by RNA. *Science, 263*(5152), 1425-1429.

Jones, R. L., 3rd, Jaskula, J. C., & Janssen, G. R. (1992). *In vivo* translational start site selection on leaderless mRNA transcribed from the *Streptomyces fradiae aph* gene. *J Bacteriol, 174*(14), 4753-4760.

Kajitani, M., & Ishihama, A. (1984). Promoter selectivity of *Escherichia coli* RNA polymerase. Differential stringent control of the multiple promoters from ribosomal RNA and protein operons. *J Biol Chem, 259*(3), 1951-1957.

Kalapos, M. P., Paulus, H., & Sarkar, N. (1997). Identification of ribosomal protein S1 as a poly(A) binding protein in *Escherichia coli*. *Biochimie, 79*(8), 493-502.

Kamen, R. (1970). Characterization of the subunits of Q-beta replicase. *Nature, 228*(5271), 527-533.

Karabiber, F., McGinnis, J. L., Favorov, O. V., & Weeks, K. M. (2013). QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA, 19*(1), 63-73.

Kieft, J. S. (2008). Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci, 33*(6), 274-283.

Kim, B., Park, H., Na, D., & Lee, S. Y. (2014). Metabolic engineering of *Escherichia coli* for the production of phenol from glucose. *Biotechnol J, 9*(5), 621-629.

Kimura, M., Foulaki, K., Subramanian, A. R., & Wittmann-Liebold, B. (1982). Primary structure of *Escherichia coli* ribosomal protein S1 and features of its functional domains. *Eur J Biochem, 123*(1), 37-53.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., & Mori, H. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res, 12*(5), 291-299.

Koh, E. Y., Ho, S. C., Mariati, Song, Z., Bi, X., Bardor, M., & Yang, Y. (2013). An internal ribosome entry site (IRES) mutant library for tuning expression level of multiple genes in mammalian cells. *PLoS One, 8*(12), e82100.

Kolb, A., Hermoso, J. M., Thomas, J. O., & Szer, W. (1977). Nucleic acid helix-unwinding properties of ribosomal protein S1 and the role of S1 in mRNA binding to ribosomes. *Proc Natl Acad Sci U S A, 74*(6), 2379-2383.

Komarova, A. V., Tchufistova, L. S., Dreyfus, M., & Boni, I. V. (2005). AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J Bacteriol, 187*(4), 1344-1349.

Komarova, A. V., Tchufistova, L. S., Supina, E. V., & Boni, I. V. (2002). Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. *RNA, 8*(9), 1137-1147.

Konarev, P. V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr., 36*, 1277-1282.

Kondo, M., Gallerani, R., & Weissmann, C. (1970). Subunit structure of Q-beta replicase. *Nature, 228*(5271), 525-527.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene, 361*, 13-37.

Labischinski, H., & Subramanian, A. R. (1979). Protein S1 from *Escherichia coli* ribosomes: an improved isolation procedure and shape determination by small-angle X-ray scattering. *Eur J Biochem, 95*(2), 359-366.

Lahiry, A., Stimple, S. D., Wood, D. W., & Lease, R. A. (2017). Retargeting a Dual-Acting sRNA for Multiple mRNA Transcript Regulation. *ACS Synth Biol, 6,* 648-658.

Lauber, M. A., Rappsilber, J., & Reilly, J. P. (2012). Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. *Mol Cell Proteomics, 11*(12), 1965-1976.

Laughrea, M., & Moore, P. B. (1977). Physical properties of ribosomal protein S1 and its interaction with the 30S ribosomal subunit of *Escherichia coli*. *J Mol Biol, 112*(3), 399-421.

Laursen, B. S., Sorensen, H. P., Mortensen, K. K., & Sperling-Petersen, H. U. (2005). Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev, 69*(1), 101-123.

Lebars, I., Hu, R. M., Lallemand, J. Y., Uzan, M., & Bontems, F. (2001). Role of the substrate conformation and of the S1 protein in the cleavage efficiency of the T4 endoribonuclease RegB. *J Biol Chem, 276*(16), 13264-13272.

Lee, K., Holland-Staley, C. A., & Cunningham, P. R. (1996). Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *RNA, 2*(12), 1270-1285.

Lemke, J. J., Sanchez-Vazquez, P., Burgos, H. L., Hedberg, G., Ross, W., & Gourse, R. L. (2011). Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proc Natl Acad Sci U S A, 108*(14), 5712-5717.

Li, G. W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell, 157*(3), 624-635.

Liang, J. C., Chang, A. L., Kennedy, A. B., & Smolke, C. D. (2012). A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. *Nucleic Acids Res, 40*(20), e154.

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., . . . Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature, 456*(7221), 464-469.

Linde, R., Quoc Khanh, N., Lipecky, R., & Gassen, H. G. (1979). On the function of the ribosomal protein S1 in the elongation cycle of bacterial protein synthesis. *Eur J Biochem, 93*(3), 565-572.

Lockwood, A. H., Chakraborty, P. R., & Maitra, U. (1971). A complex between initiation factor IF2, guanosine triphosphate, and fMet-tRNA: an intermediate in initiation complex formation. *Proc Natl Acad Sci U S A, 68*(12), 3122-3126.

Loveland, A. B., & Korostelev, A. A. (2018). Structural dynamics of protein S1 on the 70S ribosome visualized by ensemble cryo-EM. *Methods, 137*, 55-66.

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

Lynch, S. A., Desai, S. K., Sajja, H. K., & Gallivan, J. P. (2007). A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem Biol, 14*(2), 173-184.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res, 47*(W1), W636-W641.

Maini, R., Nguyen, D. T., Chen, S., Dedkova, L. M., Chowdhury, S. R., Alcala-Torano, R., & Hecht, S. M. (2013). Incorporation of beta-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorg Med Chem, 21*(5), 1088-1096.

Malys, N., & McCarthy, J. E. (2011). Translation initiation: variations in the mechanism can be anticipated. *Cell Mol Life Sci, 68*(6), 991-1003.

Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., & Breaker, R. R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science, 306*(5694), 275-279.

Marcus, J. I., Hassoun, S., & Nair, N. U. (2017). Computational prediction of functional abortive RNA in *E. coli*. *Genomics, 109*(3-4), 196-203.

Martin, C. T., & Coleman, J. E. (1989). T7 RNA polymerase does not interact with the 5'-phosphate of the initiating nucleotide. *Biochemistry, 28*(7), 2760-2762.

Marzi, S., Myasnikov, A. G., Serganov, A., Ehresmann, C., Romby, P., Yusupov, M., & Klaholz, B. P. (2007). Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell, 130*(6), 1019-1031.

McArthur, S. D., Pulvermacher, S. C., & Stauffer, G. V. (2006). The *Yersinia pestis gcvB* gene encodes two small regulatory RNA molecules. *BMC Microbiol, 6*, 52.

McGinness, K. E., & Sauer, R. T. (2004). Ribosomal protein S1 binds mRNA and tmRNA similarly but plays distinct roles in translation of these molecules. *Proc Natl Acad Sci U S A, 101*(37), 13454-13459.

McGinnis, J. L., Duncan, C. D., & Weeks, K. M. (2009). High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol, 468*, 67-89.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., & Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol, 19*(6), 560-567.

Merendino, L., Falciatore, A., & Rochaix, J. D. (2003). Expression and RNA binding properties of the chloroplast ribosomal protein S1 from *Chlamydomonas reinhardtii*. *Plant Mol Biol, 53*(3), 371-382.

Miranda, G., Schuppli, D., Barrera, I., Hausherr, C., Sogo, J. M., & Weber, H. (1997). Recognition of bacteriophage Qbeta plus strand RNA as a template by Qbeta replicase: role of RNA interactions mediated by ribosomal proteins S1 and host factor. *J Mol Biol, 267*(5), 1089-1103.

Mitchison, D. A. (1985). The action of antituberculosis drugs in short-course chemotherapy. *Tubercle, 66*(3), 219-225.

Modi, S. R., Camacho, D. M., Kohanski, M. A., Walker, G. C., & Collins, J. J. (2011). Functional characterization of bacterial sRNAs using a network biology approach. *Proc Natl Acad Sci U S A, 108*(37), 15522-15527.

Mogridge, J., & Greenblatt, J. (1998). Specific binding of *Escherichia coli* ribosomal protein S1 to boxA transcriptional antiterminator RNA. *J Bacteriol, 180*(8), 2248-2252.

Moll, I., & Engelberg-Kulka, H. (2012). Selective translation during stress in *Escherichia coli*. *Trends Biochem Sci, 37*(11), 493-498.

Moll, I., Grill, S., Grundling, A., & Blasi, U. (2002). Effects of ribosomal proteins S1, S2 and the DeaD/CsdA DEAD-box helicase on translation of leaderless and canonical mRNAs in *Escherichia coli*. *Mol Microbiol, 44*(5), 1387-1396.

Moll, I., Grill, S., Gualerzi, C. O., & Blasi, U. (2002). Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol Microbiol, 43*(1), 239-246.

Moll, I., Hirokawa, G., Kiel, M. C., Kaji, A., & Blasi, U. (2004). Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res, 32*(11), 3354-3363.

Moll, I., Huber, M., Grill, S., Sairafi, P., Mueller, F., Brimacombe, R., . . . Blasi, U. (2001). Evidence against an Interaction between the mRNA downstream box and 16S rRNA in translation initiation. *J Bacteriol, 183*(11), 3499-3505.

Moll, I., Resch, A., & Blasi, U. (1998). Discrimination of 5'-terminal start codons by translation initiation factor 3 is mediated by ribosomal protein S1. *FEBS Lett, 436*(2), 213-217.

Moore, M. J., & Silver, P. A. (2008). Global analysis of mRNA splicing. *RNA, 14*(2), 197-203.

Mrozowich, T., McLennan, S., Overduin, M., & Patel, T. R. (2018). Structural Studies of Macromolecules in Solution using Small Angle X-Ray Scattering. *J Vis Exp* (141).

Muranaka, N., Sharma, V., Nomura, Y., & Yokobayashi, Y. (2009). An efficient platform for genetic selection and screening of gene switches in *Escherichia coli*. *Nucleic Acids Res, 37*(5), e39.

Mustoe, A. M., Busan, S., Rice, G. M., Hajdin, C. E., Peterson, B. K., Ruda, V. M., . . . Weeks, K. M. (2018). Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell, 173*(1), 181-195 e118.

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., . . . Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods, 10*(4), 354-360.

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q. A., Christoffersen, M. J., Martin, L., . . . Arkin, A. P. (2013). Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat Methods, 10*(4), 347-353.

Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., & Arkin, A. P. (2012). Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol, 8*(5), 447-454.

Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., & Lee, S. Y. (2013). Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat Biotechnol, 31*(2), 170-174.

Nakagawa, S., Niimura, Y., Miura, K., & Gojobori, T. (2010). Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci U S A, 107*(14), 6382-6387.

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., & Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Cell Physiol, 54*(10), 1724-1735.

Narberhaus, F., Waldminghaus, T., & Chowdhury, S. (2006). RNA thermometers. *FEMS Microbiol Rev, 30*(1), 3-16.

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., & Altuvia, S. (2009). A pH-responsive riboregulator. *Genes Dev, 23*(22), 2650-2662.

Neogy, R. K., Chowdhury, K., & Kerr, I. (1974). Nitrocellulose filter retention method for studying drug-nucleic acid interactions. *Biochim Biophys Acta, 374*(1), 96-107.

Neupert, J., & Bock, R. (2009). Designing and using synthetic RNA thermometers for temperature-controlled gene expression in bacteria. *Nat Protoc, 4*(9), 1262-1273.

Nishizuka, Y., & Lipmann, F. (1966). Comparison of guanosine triphosphate split and polypeptide synthesis with a purified *E. coli* system. *Proc Natl Acad Sci U S A, 55*(1), 212-219.

Nomura, M., Gourse, R., & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem, 53*, 75-117.

Nomura, M., Yates, J. L., Dean, D., & Post, L. E. (1980). Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein MRNA. *Proc Natl Acad Sci U S A, 77*(12), 7084-7088.

Omotajo, D., Tate, T., Cho, H., & Choudhary, M. (2015). Distribution and diversity of ribosome binding sites in prokaryotic genomes. *BMC Genomics, 16*(1), 604.

Orelle, C., Carlson, E. D., Szal, T., Florin, T., Jewett, M. C., & Mankin, A. S. (2015). Protein synthesis by ribosomes with tethered subunits. *Nature*. *524*, 119-124

Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., & Collins, J. J. (2014). Paper-based synthetic gene networks. *Cell, 159*(4), 940-954.

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., . . . Collins, J. J. (2016). Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell, 165*(5), 1255-1266.

Pedersen, S., Bloch, P. L., Reeh, S., & Neidhardt, F. C. (1978). Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell, 14*(1), 179-190.

Pedersen, S., Skouv, J., Kajitani, M., & Ishihama, A. (1984). Transcriptional organization of the *rpsA* operon of *Escherichia coli*. *Mol Gen Genet, 196*(1), 135-140.

Pestova, T. V., & Hellen, C. U. T. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes & Development, 17*(2), 181-186.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., . . . Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr, 45*(Pt 2), 342-350.

Pfleger, B. F., Pitera, D. J., Smolke, C. D., & Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol, 24*(8), 1027-1032.

Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B., & Ehresmann, C. (1990). Target site of *Escherichia coli* ribosomal protein S15 on its messenger RNA. Conformation and interaction with the protein. *J Mol Biol, 211*(2), 415-426.

Phillips, L. A., Pang, R. H., Park, J. J., Hollis, V. W., Jr., & Famuyiwa, F. (1980). Poly(U)-agarose affinity chromatography: specific, sensitivity selectivity, and affinity of binding. *Prep Biochem, 10*(1), 11-26.

Pioletti, M., Schlunzen, F., Harms, J., Zarivach, R., Gluhmann, M., Avila, H., . . . Franceschi, F. (2001). Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J, 20*(8), 1829-1839.

Potapov, A. P., & Subramanian, A. R. (1992). Effect of *E. coli* ribosomal protein S1 on the fidelity of the translational elongation step: reading and misreading of poly(U) and poly(dT). *Biochem Int, 27*(4), 745-753.

Qi, H., Shimizu, Y., & Ueda, T. (2007). Ribosomal protein S1 is not essential for the trans-translation machinery. *J Mol Biol, 368*(3), 845-852.

Qu, X., Lancaster, L., Noller, H. F., Bustamante, C., & Tinoco, I., Jr. (2012). Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proc Natl Acad Sci U S A, 109*(36), 14458-14463.

Rackham, O., & Chin, J. W. (2005). A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol, 1*(3), 159-166.

Reeh, S., Pedersen, S., & Friesen, J. D. (1976). Biosynthetic regulation of individual proteins in relA+ and relA strains of *Escherichia coli* during amino acid starvation. *Mol Gen Genet, 149*(3), 279-289.

Reeve, B., Hargest, T., Gilbert, C., & Ellis, T. (2014). Predicting translation initiation rates for designing synthetic biology. *Front Bioeng Biotechnol, 2*, 1.

Resch, A., Tedin, K., Grundling, A., Mundlein, A., & Blasi, U. (1996). Downstream box-anti-downstream box interactions are dispensable for translation initiation of leaderless mRNAs. *EMBO J, 15*(17), 4740-4748.

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics, 11*, 129.

Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., . . . Wanner, B. L. (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res, 34*(1), 1-9.

Ringquist, S., Jones, T., Snyder, E. E., Gibson, T., Boni, I., & Gold, L. (1995). High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry, 34*(11), 3640-3648.

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. *Mol Microbiol, 6*(9), 1219-1229.

Rio, D. C. (2012). Filter-binding assay for analysis of RNA-protein interactions. *Cold Spring Harb Protoc, 2012*(10), 1078-1081.

Robinson-Mosher, A., Shinar, T., Silver, P. A., & Way, J. (2013). Dynamics simulations for engineering macromolecular interactions. *Chaos, 23*(2), 025110.

Rodnina, M. V., Fricke, R., & Wintermeyer, W. (1994). Transient conformational states of aminoacyl-tRNA during ribosome binding catalyzed by elongation factor Tu. *Biochemistry, 33*(40), 12267-12275.

Rodnina, M. V., Stark, H., Savelsbergh, A., Wieden, H. J., Mohr, D., Matassova, N. B., . . . Wintermeyer, W. (2000). GTPases mechanisms and functions of translation factors on the ribosome. *Biol Chem, 381*(5-6), 377-387.

Rodrigo, G., Landrain, T. E., & Jaramillo, A. (2012). *De novo* automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proc Natl Acad Sci U S A, 109*(38), 15271-15276.

Ruckman, J., Ringquist, S., Brody, E., & Gold, L. (1994). The bacteriophage T4 regB ribonuclease. Stimulation of the purified enzyme by ribosomal protein S1. *J Biol Chem, 269*(43), 26655-26662.

Russell, J. B., & Cook, G. M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev, 59*(1), 48-62.

Saguy, M., Gillet, R., Skorski, P., Hermann-Le Denmat, S., & Felden, B. (2007). Ribosomal protein S1 influences trans-translation *in vitro* and *in vivo*. *Nucleic Acids Res, 35*(7), 2368-2376.

Salah, P., Bisaglia, M., Aliprandi, P., Uzan, M., Sizun, C., & Bontems, F. (2009). Probing the relationship between Gram-negative and Gram-positive S1 proteins by sequence analysis. *Nucleic Acids Res, 37*(16), 5578-5588.

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol, 27*(10), 946-950.

Sanson, B., Hu, R. M., Troitskayadagger, E., Mathy, N., & Uzan, M. (2000). Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. *J Mol Biol, 297*(5), 1063-1074.

Schauder, B., & McCarthy, J. E. (1989). The role of bases upstream of the Shine-Dalgarno region and in the coding sequence in the control of gene expression in *Escherichia coli*: translation and stability of mRNAs in vivo. *Gene, 78*(1), 59-72.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., . . . Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell, 102*(5), 615-623.

Schurr, T., Nadir, E., & Margalit, H. (1993). Identification and characterization of *E.coli* ribosomal binding sites by free energy computation. *Nucleic Acids Res, 21*(17), 4019-4023.

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., & Cate, J. H. (2005). Structures of the bacterial ribosome at 3.5 A resolution. *Science, 310*(5749), 827-834.

Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., & Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. *Science, 330*(6007), 1099-1102.

Sengupta, J., Agrawal, R. K., & Frank, J. (2001). Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. *Proc Natl Acad Sci U S A, 98*(21), 11991-11996.

Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell, 152*(1-2), 17-24.

Sharma, C. M., Darfeuille, F., Plantinga, T. H., & Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev, 21*(21), 2804-2817.

Sharma, C. M., Papenfort, K., Pernitzsch, S. R., Mollenkopf, H. J., Hinton, J. C., & Vogel, J. (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol, 81*(5), 1144-1165.

Shean, C. S., & Gottesman, M. E. (1992). Translation of the prophage λ cI transcript. *Cell, 70*(3), 513-522.

Shi, W., Zhang, X., Jiang, X., Yuan, H., Lee, J. S., Barry, C. E., 3rd, . . . Zhang, Y. (2011). Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. *Science, 333*(6049), 1630-1632.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nat Biotechnol, 19*(8), 751-755.

Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., & Margalit, H. (2007). Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol, 3*, 138.

Shine, J., & Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A, 71*(4), 1342-1346.

Sillers, I. Y., & Moore, P. B. (1981). Position of protein S1 in the 30S ribosomal subunit of *Escherichia coli*. *J Mol Biol, 153*(3), 761-780.

Silveira, A. C., Robertson, K. L., Lin, B., Wang, Z., Vora, G. J., Vasconcelos, A. T., & Thompson, F. L. (2010). Identification of non-coding RNAs in environmental *vibrios*. *Microbiology, 156*(Pt 8), 2452-2458.

Skorski, P., Leroy, P., Fayet, O., Dreyfus, M., & Hermann-Le Denmat, S. (2006). The highly efficient translation initiation region from the *Escherichia coli rpsA* gene lacks a shine-dalgarno element. *J Bacteriol, 188*(17), 6277-6285.

Skouv, J., Schnier, J., Rasmussen, M. D., Subramanian, A. R., & Pedersen, S. (1990). Ribosomal protein S1 of *Escherichia coli* is the effector for the regulation of its own synthesis. *J Biol Chem, 265*(28), 17044-17049.

Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., L, B. A. W., Giannoukos, G., . . . Voigt, C. A. (2014). Functional optimization of gene clusters by combinatorial design and assembly. *Nat Biotechnol, 32*(12), 1241-1249.

Sorensen, 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 vivo*. *J Mol Biol, 280*(4), 561-569.

Stasiewicz, J., Mukherjee, S., Nithin, C., & Bujnicki, J. M. (2019). QRNAS: software tool for refinement of nucleic acid structures. *BMC Struct Biol, 19*(1), 5.

Stephenson, M. L., Thimann, K. V., & Zamecnik, P. C. (1956). Incorporation of C14-amino acids into proteins of leaf disks and cell-free fractions of tobacco leaves. *Archives of Biochemistry and Biophysics, 65*(1), 194-209.

Storz, G., Vogel, J., & Wassarman, K. M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell, 43*(6), 880-891.

Studer, S. M., & Joseph, S. (2006). Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol Cell, 22*(1), 105-115.

Subramanian, A. R. (1983). Structure and functions of ribosomal protein S1. *Prog Nucleic Acid Res Mol Biol, 28*, 101-142.

Subramanian, A. R., & Davis, B. D. (1970). Activity of initiation factor F3 in dissociating *Escherichia coli* ribosomes. *Nature, 228*(5278), 1273-1275.

Subramanian, A. R., Rienhardt, P., Kimura, M., & Suryanarayana, T. (1981). Fragments of ribosomal protein S1 and its mutant form m1-S1. Localization of nucleic-acid-binding domain in the middle region of S1. *Eur J Biochem, 119*(2), 245-249.

Sukhodolets, M. V., & Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry, 42*(26), 8022-8034.

Sukhodolets, M. V., Garges, S., & Adhya, S. (2006). Ribosomal protein S1 promotes transcriptional cycling. *RNA, 12*(8), 1505-1513.

Suryanarayana, T., & Subramanian, A. R. (1983). An essential function of ribosomal protein S1 in messenger ribonucleic acid translation. *Biochemistry, 22*(11), 2715-2719.

Suryanarayana, T., & Subramanian, A. R. (1984). Function of the repeating homologous sequences in nucleic acid binding domain of ribosomal protein S1. *Biochemistry, 23*(6), 1047-1051.

Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr., 25*, 495-503.

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal, 76*(6), 2879-2886.

Szer, W., Hermoso, J. M., & Boublik, M. (1976). Destabilization of the secondary structure of RNA by ribosomal protein S1 from *Escherichia coli*. *Biochem Biophys Res Commun, 70*(3), 957-964.

Takeshita, D., Yamashita, S., & Tomita, K. (2014). Molecular insights into replication initiation by Qbeta replicase using ribosomal protein S1. *Nucleic Acids Res, 42*(16), 10809-10822.

Tal, M., Aviram, M., Kanarek, A., & Weiss, A. (1972). Polyuridylic acid binding and translating by *Escherichia coli* ribosomes: stimulation by protein I, inhibition by aurintricarboxylic acid. *Biochim Biophys Acta, 281*(3), 381-392.

Tchufistova, L. S. (2003). A key role for the mRNA leader structure in translational control of ribosomal protein S1 synthesis in -proteobacteria. *Nucleic Acids Research, 31*(23), 6996-7002. doi:10.1093/nar/gkg883

Tchufistova, L. S., Komarova, A. V., & Boni, I. V. (2003). A key role for the mRNA leader structure in translational control of ribosomal protein S1 synthesis in gamma-proteobacteria. *Nucleic Acids Res, 31*(23), 6996-7002.

Tedin, K., Moll, I., Grill, S., Resch, A., Graschopf, A., Gualerzi, C. O., & Blasi, U. (1999). Translation initiation factor 3 antagonizes authentic start codon selection on leaderless mRNAs. *Mol Microbiol, 31*(1), 67-77.

Tedin, K., Resch, A., & Blasi, U. (1997). Requirements for ribosomal protein S1 for translation initiation of mRNAs with and without a 5' leader sequence. *Mol Microbiol, 25*(1), 189-199.

Thomas, J. O., Boublik, M., Szer, W., & Subramanian, A. R. (1979). Nucleic acid binding and unfolding properties of ribosomal protein S1 and the derivatives S1-F1 and m1-S1. *Eur J Biochem, 102*(1), 309-314.

Tian, T., & Salis, H. M. (2015). A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Res, 43*(14), 7137-7151.

Topp, S., & Gallivan, J. P. (2007). Guiding Bacteria with Small Molecules and RNA. *Journal of the American Chemical Society, 129*(21), 6807-6811.

Topp, S., & Gallivan, J. P. (2008). Riboswitches in unexpected places--a synthetic riboswitch in a protein coding region. *RNA, 14*(12), 2498-2503.

Topp, S., Reynoso, C. M., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., . . . Gallivan, J. P. (2010). Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl Environ Microbiol, 76*(23), 7881-7884.

Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., Hightower, G. A., Smith, J. T., & Conway, T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol, 68*(5), 1128-1148.

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science, 249*(4968), 505-510.

Tzareva, N. V., Makhno, V. I., & Boni, I. V. (1994). Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions. *FEBS Lett, 337*(2), 189-194.

Udagawa, T., Shimizu, Y., & Ueda, T. (2004). Evidence for the translation initiation of leaderless mRNAs by the intact 70S ribosome without its dissociation into subunits in eubacteria. *J Biol Chem, 279*(10), 8539-8546.

Ueta, M., Ohniwa, R. L., Yoshida, H., Maki, Y., Wada, C., & Wada, A. (2008). Role of HPF (hibernation promoting factor) in translational activity in *Escherichia coli*. *J Biochem, 143*(3), 425-433.

Urban, J. H., & Vogel, J. (2007). Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res, 35*(3), 1018-1037.

Usui, K., Ichihashi, N., Kazuta, Y., Matsuura, T., & Yomo, T. (2014). Effects of ribosomes on the kinetics of Qbeta replication. *FEBS Lett, 588*(1), 117-123.

Uzan, M. (2001). Bacteriophage T4 RegB endoribonuclease. *Methods Enzymol, 342*, 467-480.

Valentin-Hansen, P., Eriksen, M., & Udesen, C. (2004). MicroReview: The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Molecular microbiology, 51*(6), 1525-1533.

Van Dieijen, G., Van Der Laken, C. J., Van Knippenberg, P. H., & Van Duin, J. (1975). Function of *Escherichia coli* ribosomal protein S1 in translation of natural and synthetic messenger RNA. *J Mol Biol, 93*(3), 351-366.

van Dieijen, G., van Knippenberg, P. H., & van Duin, J. (1976). The specific role of ribosomal protein S1 in the recognition of native phage RNA. *Eur J Biochem, 64*(2), 511-518.

Van Duin, J., & Kurland, C. G. (1970). Functional heterogeneity of the 30S ribosomal subunit of *E. coli*. *Mol Gen Genet, 109*(2), 169-176.

van Knippenberg, P. H., Hooykaas, P. J., & van Duin, J. (1974). The stoichiometry of *E. coli* 30S ribosomal protein S1 on *in vivo* and *in vitro* polyribosomes. *FEBS Lett, 41*(2), 323-326.

Vanderpool, C. K. (2011). Combined experimental and computational strategies define an expansive regulon for GcvB small RNA. *Mol Microbiol, 81*(5), 1129-1132.

Vasilyev, N. N., Kutlubaeva, Z. S., Ugarov, V. I., Chetverina, H. V., & Chetverin, A. B. (2013). Ribosomal protein S1 functions as a termination factor in RNA synthesis by Qbeta phage replicase. *Nat Commun, 4*, 1781.

Vellanoweth, R. L., & Rabinowitz, J. C. (1992). The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli in vivo*. *Mol Microbiol, 6*(9), 1105-1114.

Venkatesan, A., & Dasgupta, A. (2001). Novel fluorescence-based screen to identify small synthetic internal ribosome entry site elements. *Mol Cell Biol, 21*(8), 2826-2837.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H., & Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell, 147*(1), 147-157.

Vigar, J., & Wieden, H. J. (2017). Engineering bacterial translation initiation - Do we have all the tools we need? *Biochim Biophys Acta*. *1861(11 Pt B),* 3060-3069.

Vind, J., Sørensen, M. A., Rasmussen, M. D., & Pedersen, S. (1993). Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: expression from reporter genes does not always reflect functional mRNA levels. *Journal of Molecular Biology, 231*(3), 678-688.

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., & Gelfand, M. S. (2003). Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA, 9*(9), 1084-1097.

Volkov, V., & I. Svergun, D. (2003). *Uniqueness of ab initio shape determination in small-angle scattering* (Vol. 36).

Voynow, P., & Kurland, C. G. (1971). Stoichiometry of the 30S ribosomal proteins of *Escherichia coli*. *Biochemistry, 10*(3), 517-524.

Wada, A., Yamazaki, Y., Fujita, N., & Ishihama, A. (1990). Structure and probable genetic location of a "ribosome modulation factor" associated with 100S ribosomes in stationary-phase *Escherichia coli* cells. *Proc Natl Acad Sci U S A, 87*(7), 2657-2661.

Wahba, A. J., Miller, M. J., Niveleau, A., Landers, T. A., Carmichael, G. G., Weber, K., . . . Slobin, L. I. (1974). Subunit I of Qbeta replicase and 30S ribosomal protein S1 of *Escherichia coli*. Evidence for the identity of the two proteins. *J Biol Chem, 249*(10), 3314-3316.

Waldminghaus, T., Kortmann, J., Gesing, S., & Narberhaus, F. (2008). Generation of synthetic RNA-based thermosensors. *Biol Chem, 389*(10), 1319-1326.

Wang, J. X., Lee, E. R., Morales, D. R., Lim, J., & Breaker, R. R. (2008). Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell, 29*(6), 691-702.

Waters, L. S., & Storz, G. (2009). Regulatory RNAs in bacteria. *Cell, 136*(4), 615-628.

Weglenska, A., Jacob, B., & Sirko, A. (1996). Transcriptional pattern of *Escherichia coli ihfB* (himD) gene expression. *Gene, 181*(1-2), 85-88.

Wilson, J. E., Pestova, T. V., Hellen, C. U., & Sarnow, P. (2000). Initiation of protein synthesis from the A site of the ribosome. *Cell, 102*(4), 511-520.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., . . . Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature, 407*(6802), 327-339.

Wintermeyer, W., & Gualerzi, C. (1983). Effect of *Escherichia coli* initiation factors on the kinetics of N-Acphe-tRNAPhe binding to 30S ribosomal subunits. A fluorescence stopped-flow study. *Biochemistry, 22*(3), 690-694.

Wisniewski, J. R., & Rakus, D. (2014). Quantitative analysis of the *Escherichia coli* proteome. *Data Brief, 1*, 7-11.

Wittmann, A., & Suess, B. (2012). Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett, 586*(15), 2076-2083.

Wood, T. K., & Peretti, S. W. (1991). Construction of a specialized ribosome vector or cloned gene expression in *E. coli*. *Biotechnol Bioeng, 38*(8), 891-906.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., . . . Chen, S. (2012). FastUniq: a fast *de novo* duplicates removal tool for paired short reads. *PLoS One, 7*(12), e52249.

Xu, S., Grullon, S., Ge, K., & Peng, W. (2014). Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods Mol Biol, 1150*, 97-111.

Yang, J., Liu, Y., Bi, J., Cai, Q., Liao, X., Li, W., . . . Lin, D. (2015). Structural basis for targeting the ribosomal protein S1 of *Mycobacterium tuberculosis* by pyrazinamide. *Mol Microbiol, 95*(5), 791-803.

Yoo, S. M., Na, D., & Lee, S. Y. (2013). Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*. *Nat Protoc, 8*(9), 1694-1707.

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., & Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol, 50*(4), 1111-1124.

Zhang, J., & Deutscher, M. P. (1992). A uridine-rich sequence required for translation of prokaryotic mRNA. *Proc Natl Acad Sci U S A, 89*(7), 2605-2609.