THE *RPSA* GENE AND ITS PRODUCT RIBOSOMAL PROTEIN S1: INVESTIGATING THE ROLE OF RNA STRUCTURE IN TRANSLATION

JALYCE L. E. HELLER
Bachelor of Science, University of Lethbridge, 2015

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

MASTER OF SCIENCE

in

BIOCHEMISTRY

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

© Jalyce L. E. Heller, 2019

THE *RPSA* GENE AND ITS PRODUCT RIBOSOMAL PROTEIN S1:
INVESTIGATING THE ROLE OF RNA STRUCTURE IN TRANSLATION

JALYCE L. E. HELLER

Date of Defense: July 5, 2019

Dr. Hans-Joachim Wieden Thesis Supervisor	Professor	Ph.D.
Dr. Ute Kothe Thesis Examination Committee Member	Professor	Ph.D.
Dr. Nehal Thakor Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. James Bullard External Examiner University of Texas Rio Grande Valley Austin, Texas	Associate Professor	Ph.D.
Dr. Trushar Patel Chair, Thesis Examination Committee	Assistant Professor	Ph.D.

ABSTRACT

The *rpsA* gene contains a translation initiation region (TIR) which lacks a Shine-Dalgarno element, is highly structured, and which efficiently drives translation in bacteria. The *rpsA* gene is downregulated by its gene product, ribosomal protein S1. To investigate how the structure of the TIR facilitates function both in translation and downregulation, we employed a combined SAXS, SHAPE, and computational approach. The structure-function relationship reveals that the introduction of single nucleotide substitutions in the TIR alters its structure, corresponding to a modulated translation efficiency *in vivo*. The activity of the *rpsA* TIR in a minimal *in vitro* transcription/translation system was investigated to understand the TIR's requirement for additional factors. We observed that S1 is strictly required for its translation, and that S1 specifically recognizes and binds to the TIR when in excess over ribosomes. The structure-driven mechanism of this TIR may represent a previously overlooked strategy of translation initiation in bacteria.

PREFACE

The review article in Chapter 2, "Ribosomal protein S1: Multiple cellular roles," was authored by Jalyce Heller, Justin Vigar, and HJ Wieden and explores how the *rpsA* gene is regulated at the transcriptional and translational levels, as well as the multiple functions of its gene product, ribosomal protein S1. My contribution to the manuscript includes the sections pertaining to S1's multiple roles. Justin Vigar's contribution includes the sections on gene regulation. HJ Wieden contributed to editing the manuscript and providing ideas.

The manuscript in Chapter 3, "The *rpsA* translation initiation region: Biophysical characterization of a structure-driven translation initiation mechanism," was authored by Justin Vigar, Jalyce Heller, Astha, Nithin Chandran, Erik Holmqvist, Lars Barquist, Graeme Glaister, Trushar Patel, Janusz Bujnicki, Jorg Vogel, and HJ Wieden. The manuscript aims to answer the question of how the *rpsA* TIR's unique features, including its strong secondary structure, contribute to its efficiency in driving translation. I purified ribosomes, $30S^{-S1}$ particles, ribosomal protein S1, and T7 RNA polymerase as well as half of the *rpsA* TIR variants for SAXS and SHAPE analysis. I performed RT-qPCR, PURE experiments, SAXS data analysis (with guidance from Dr. Trushar Patel), and fitting of the atomic models provided by the Bujnicki lab into SAXS envelopes. The *rpsA* TIR library was designed, tested, and validated by Justin Vigar. Justin performed these experiments in collaboration with Jorg Vogel, Erik Holmqvist, and Lars Barquist. Justin, along with Graeme Glaister, performed the filter binding assays described in this chapter. Additionally, the preparation of RNA for SAXS and SHAPE experiments was performed by both Justin and myself. SAXS experiments were carried out at the Diamond Light Source Synchrotron in Oxford, United Kingdom. SHAPE experiments were performed by

Astha, and generation of computational models was performed by Astha, Nithin Chandran, and Janusz Bujnicki. HJ Wieden was instrumental in the planning and assistance of these experiments.

ACKNOWLEDGEMENTS

Thank you to my supervisor Dr. HJ Wieden for letting me work in the lab, and to my committee members Dr. Ute Kothe and Dr. Nehal Thakor for their helpful comments on my work. Thank you as well to Dr. Trushar Patel for teaching me how to analyze SAXS data. Special thanks to Justin Vigar for teaching me Organic Chemistry and then recruiting me to the lab. He has been a great mentor for the past three years, and is a good Pokemon Go partner. Also, Rajashekhar Kamalampeta was an awesome supervisor during my undergraduate work. This project was made possible through collaboration with many people, so I would like to thank everyone who was involved in this project – Justin Vigar, Astha, Nithin Chandran, Erik Holmqvist, Lars Barquist, Graeme Glaister, Trushar Patel, Janusz Bujnicki, Jorg Vogel, and HJ Wieden. Additionally, thank you to all of the Wieden, Kothe, and Patel lab members for their assistance in the lab and helpful insights during lab meetings. Special thanks to Fan Mo for keeping the lab up and running, and to Emily Wilton for editing our manuscripts. Finally, I would like to acknowledge my family for their love and support while I have been in school, as well as the friends that I made over the course of the last four years – especially Jimmy.

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 ribosomal subunit
ATP	Adenosine triphosphate
AU	Arbitrary units
Amp	Ampicillin
bp	Base-pair
Cam	Chloramphenicol
CDS	Coding sequence
CTP	Cytidine triphosphate
Ct	Threshold cycle
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
eCFP	Enhanced 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
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
TIR	Translation initiation region
tRNA	Transfer RNA
Tris	Tris(hydroxymethyl)aminomethane
TX-TL	Transcription-translation
UTR	Untranslated region

CHAPTER 1: INTRODUCTION

Protein synthesis (translation) is a fundamental process in all living organisms. It is carried out by the ribosome in four phases: initiation, elongation, termination, and ribosome recycling. Translation initiation is the rate-determining step that involves the loading of a messenger RNA (mRNA) onto the 30S ribosomal subunit, along with three initiation factors and an initiator transfer RNA (tRNA), to form the 30S initiation complex (Laursen, Sorensen, Mortensen, & Sperling-Petersen, 2005). In bacterial initiation, the formation of this complex is tightly regulated by the mRNA's translation initiation region (TIR) and its interaction with the ribosome. In the canonical mechanism of translation initiation, the TIR of an mRNA contains a Shine-Dalgarno (SD) element composed of the consensus sequence GGAGG located approximately 7 nucleotides upstream of the initiation codon (Laursen et al., 2005). Base-pairing between this SD region and the anti-SD region of 16S ribosomal RNA (rRNA) facilitates the correct positioning of the start codon within the P-site of the ribosome. Despite the SD-based mechanism being intensively characterized, it has been shown that a large number of bacterial genes do not contain SD elements (Chang, Halgamuge, & Tang, 2006), raising the question of how non-SD containing mRNAs are able to initiate translation. What molecular mechanisms exist apart from the canonical SD-aSD interaction, and can these interactions be identified in different genes and/or organisms? In order to answer these questions, I have studied the TIR of the *rpsA* gene coding for ribosomal protein S1. Ribosomal protein S1 is an essential protein in proteobacteria and is able to downregulate its expression by recognizing and binding to the *rpsA* TIR (Boni, Artamonova, & Dreyfus, 2000).

Interestingly, the *rpsA* TIR contains only a degenerate SD element (GAAG) which is surrounded by strong secondary structure, a feature that is typically found to be inhibitory

to translation initiation (Cambray, Guimaraes, & Arkin, 2018; de Smit & van Duin, 1990; McPheeters, Christensen, Young, Stormo, & Gold, 1986). Additionally, it has been shown that the 91 nucleotides upstream of the start codon are absolutely required for the function of the TIR both for translational activity and for repression by S1 (Boni, Artamonova, Tzareva, & Dreyfus, 2001). This stands out as another peculiar feature given that a typical ribosome binding site (RBS) is defined as the mRNA covered by the ribosome during translation initiation and typically comprises about 30 nucleotides. It has been shown that the structure of the TIR does not function to bring together a discontinuous SD element (Skorski, Leroy, Fayet, Dreyfus, & Hermann-Le Denmat, 2006), suggesting that the function of the TIR may rely predominantly on structure rather than sequence elements. A mutational analysis of the TIR landscape performed in our lab by Justin Vigar identified single nucleotide substitutions throughout the length of the TIR that alter translational activity. I confirmed via RT-qPCR that these substitutions do not influence the abundance of the variant transcripts. These *rpsA* TIR variants were used for structural analysis to investigate their impact on the 3D shape of the *rpsA* TIR and to correlate structure with function. To this end, we utilized small-angle x-ray scattering (SAXS) and selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) combined with computational modelling. The obtained data supports that the *rpsA* TIR constitutes a bona-fide structure-based initiation element that is capable of efficient translation initiation *in vivo* and in *vitro*. By using a purified and reconstituted coupled transcription/translation system (TX/TL), I also confirmed that this structural element functions independent of additional cellular factors such as RNA chaperone/folding proteins and that it is specifically recognized and repressed by excess free S1. We hypothesize that the structural element within the *rpsA* TIR mediates specific contacts with the conserved core of the ribosome, which may suggest

that direct recruitment of the 70S ribosome via an IRES-like mechanism is possible. To investigate this, I performed binding experiments between the *rpsA* TIR and the 70S ribosome.

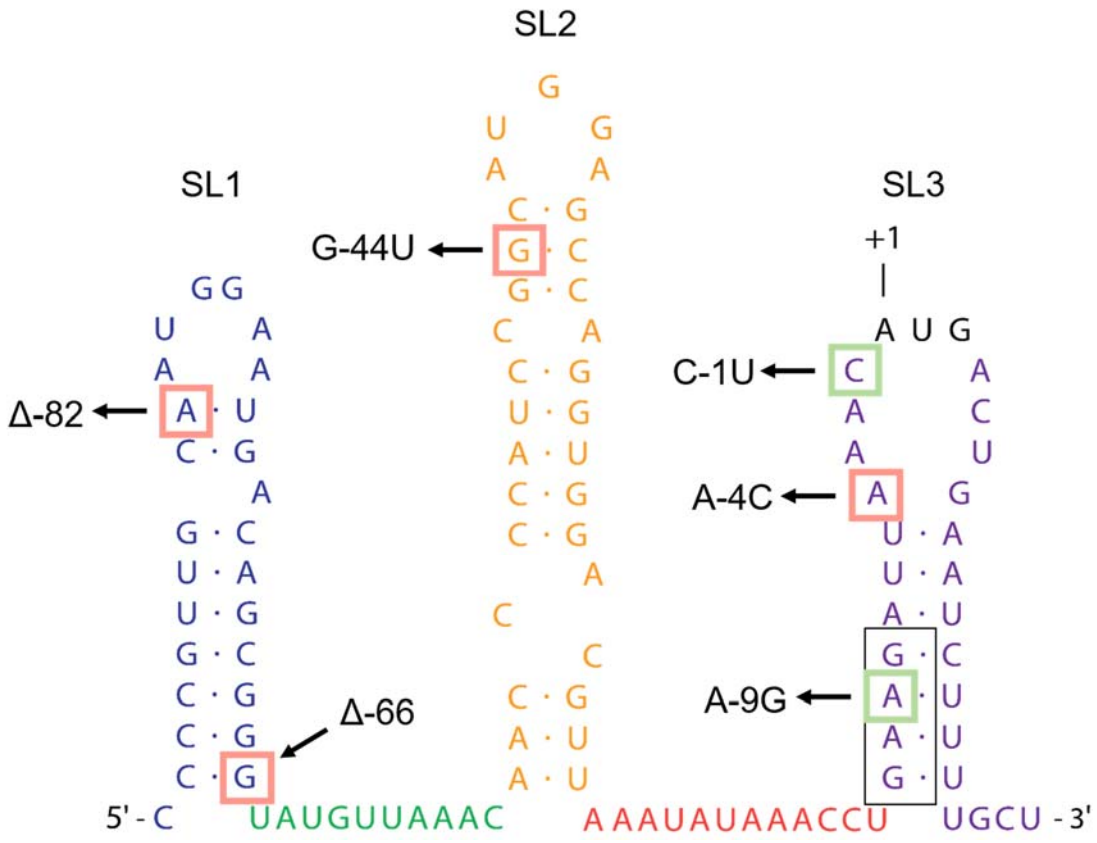


Figure 1.1 Secondary structure of the *rpsA* TIR. Three stem-loops (SL1, SL2, and SL3) are separated by two AU-rich single-stranded regions. The start codon is shown in black, and the degenerate Shine-Dalgarno sequence is boxed. Point mutations and truncations that were analyzed in this study are indicated: green=active, red=inactive. Color scheme is used in chapter 3 for 3D models.

CHAPTER 2: RIBOSOMAL PROTEIN S1: MULTIPLE CELLULAR ROLES

2.1 ABSTRACT

Ribosomal protein S1 (S1) is essential in gram-negative bacteria and plays a crucial role during the initiation phase of translation. S1 recruits mRNA to the initiating ribosome and unfolds structured mRNAs, allowing for their correct positioning on the ribosome. However, S1 also has a number of additional functions in the cell that are less understood. It has been suggested that in addition to translation initiation, S1 may function during elongation. Furthermore, S1 acts as a repressor of its own mRNA, and has been proposed to participate in transcription, as well as during tmRNA translation. S1 also plays an essential role in the Qβ phage replicase and T4 phage RegB endonuclease activity. Despite its essentiality and diverse functionality, the mechanism of action of S1 remains unclear. S1 contains six homologous motifs partitioned into a protein-binding domain and an RNA-binding domain; the level of conservation of the four RNA-binding motifs vary across organisms. In this article we summarize S1's enigmatic role in the cell, discussing how its structure enables it to carry out different functions. Once S1's mechanisms of action are better understood, it will make an attractive target for antibiotic drugs, as well as a key regulator for bioengineers to modulate gene expression.

2.2 INTRODUCTION

Initiation is the rate-determining step of translation and can be carried out through multiple strategies (Figure 2.1). 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 through base pairing interactions with the anti-SD element on the 3′ end of the 16S rRNA (Gold et al., 1981; 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 (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 an SD element (Boni, Isaeva, Musychenko, & Tzareva, 1991; Komarova, Tchufistova, Supina, & Boni, 2002; Sorensen, Fricke, & Pedersen, 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).

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 to utilizing this potential. Although multiple important roles have been identified, our mechanistic understanding functions still lags behind. 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 in order to enable the rational design of novel translation inhibitors.

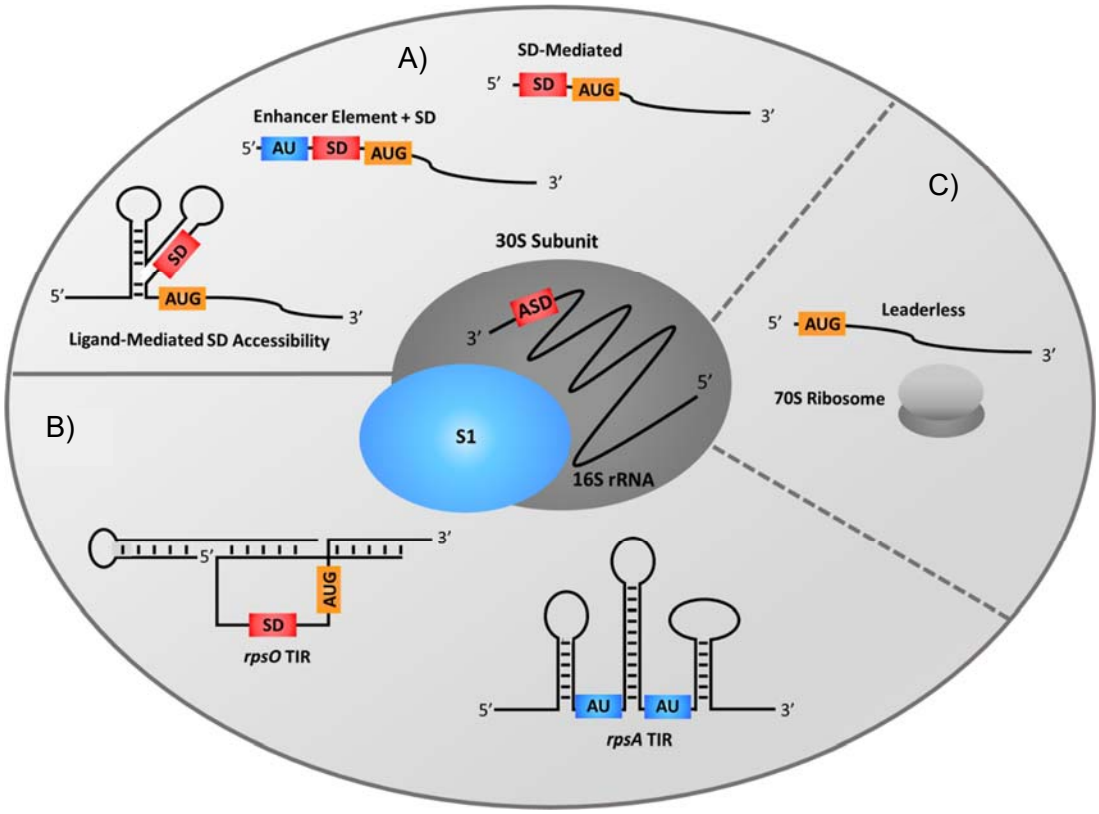


Figure 2.1 Mechanisms of translation initiation control in prokaryotes. A) Shine-Dalgarno mediated: the canonical mechanism of translation initiation in bacteria relies on a sequence upstream of the start codon, called the SD sequence, which interacts with the anti-SD sequence of the 16S rRNA to ensure correct positioning on the ribosome. Bacteria can attenuate gene expression by sequestering the SD sequence in secondary structure that is able to respond to the presence of a ligand. Additionally, mRNAs can contain AU-rich enhancer elements that interact with ribosomal protein S1 to facilitate recruitment to the ribosome. B) Structured translation initiation regions: some mRNAs, such as *rpsA* and *rpsO*, contain highly structured leader sequences that have to be melted by S1 before being translated. C) Leaderless mRNAs do not require ribosomal protein S1 and can be translated directly by 70S ribosomes.

2.3 TRANSCRIPTIONAL REGULATION OF THE *RPSA* GENE

2.3.1 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.

2.3.2 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 (cmkp) promoter is located upstream of *cmk* that drives expression of the *cmk-rpsA* and *cmk-rpsA-ihfB* transcripts.

During exponential growth the *rpsA*p1 and *rpsA*p3 promoters are active. As the cells enter stationary phase there is 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 (Weglenska, Jacob, & Sirko, 1996). During this time the σ^{32} polymerase is able to compete with σ^{70}, activating the P_5 (ihfBp) 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 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*p1 promoter (Marcus, Hassoun, & Nair, 2017). Short single-stranded unstructured faRNAs can form weak, transient complexes with complementary nucleotide sequences (Hui & de Boer, 1987) 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* ihfB expression pattern - in stationary growth phase the reduced demand for translational machinery results in fewer transcripts

being produced from the *rpsA*p1 and *rpsA*p3 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 ihfBp 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).

2.3.3 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 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 tRNA's induce ribosome-associated RelA, and SpoT to synthesize ppGpp. ppGpp and the transcription factor DskA bind to RNAP and inhibit transcription from some promoters to tune ribosome synthesis rates in response to nutrient availability. ppGpp was found to repress transcription from the upstream *rpsA*p1 promoter, but transcription from the *rpsA*p3 promoter was unaffected (Kajitani & Ishihama, 1984). The Gourse lab demonstrated that the *rpsA*p1 promoter is inhibited two-fold by ppGpp and DksA *in vitro*, while the *rpsA*p3 promoter was not (Lemke et al., 2011). Finally, high-throughput transcriptomic approaches have shown that the number of ribosomal protein (r-protein) transcripts decrease during amino acid starvation. (Durfee, Hansen, Zhi, Blattner, & Jin, 2008; Traxler et al., 2008).

S1's regulation on the transcriptional level is partially linked to the amino acid availability and the stringent response. This transcriptional regulation could be a redundant control mechanism, or the multiple levels of regulation could enable separate responses to fast changes in cellular amino acid concentration, while another could be important for slow changes. The transcriptional regulation of S1 is important, however it has long been known that S1's translational regulation is sophisticated.

2.4 TRANSLATIONAL CONTROL OF S1

In 1981 the *rpsA* gene was cloned into a multicopy plasmid enabling the over-expression of S1. Surprisingly, although *rpsA* transcription increased 4-fold, S1 synthesis by the ribosome only doubled (Christiansen & Pedersen, 1981). This highlighted the importance of post-transcriptional regulation for the 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 its gene product S1 itself acting as the effector for this regulation. The regulatory mechanism involves binding of S1to the translation initiation region (TIR) of the *rpsA* mRNA (~95 nt upstream from the start codon) (Skouv, Schnier, Rasmussen, Subramanian, & Pedersen, 1990; Tchufistova, Komarova, & Boni, 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 (Boni et al., 2000; Pedersen et al., 1984). It can fold into three consecutive stem-loops (SL1-SL3) separated by two AU-rich single-stranded regions (Boni et al., 2000; Boni et al., 2001). The two most 5′ stem-loops are very stable, while SL3 containing the start codon within the apical loop is not as strong and contains a weak Shine-Dalgarno (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 the same three-stem-loop structure, further supporting the relevance of its structure. It only has 3 nucleotides that can form a contiguous interaction with the anti-SD sequence on the 16S rRNA (Boni et al., 2000; Pedersen et al., 1984). Despite this high degree of structure and weak SD sequence, the TIR drives the expression of one of the most highly expressed genes in *E. coli*. The phylogenetically conserved secondary structure is required for the high translational activity of the TIR, as well as for the auto-regulation exerted by S1 (Boni et al., 2001; Tchufistova et al., 2003).

There have been several studies investigating the *rpsA* TIR structure and how S1 binds and alters it. The lower part of SL2 is melted by S1 upon binding to regions between the stem-loops. The autogenous regulation is achieved via the competition of free-S1 and ribosome-bound S1 for the TIR. In 1993 the first mutational study was published on the *rpsA* TIR (Vind, Sorensen, Rasmussen, & Pedersen, 1993). This study also looked at a highly active mutation of *rpsA*, A-9G, and saw that S1 is unable to exert autocontrol over it. The apical loops of SL1 and SL2 contain two 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 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 when bound to the ribosome (Mustoe et al., 2018). The specific folding of the *rpsA* TIR may 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). With all this being said 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, when there is insufficient rRNA to bind the proteins. The binding site on their mRNAs mimics their respective rRNA binding sites, ultimately leading to inhibiting 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; 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 to certain U or AU containing single-stranded regions present in mRNAs (Boni et al., 1991; Tzareva et al., 1994). S1 also binds two regions of the QB 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 psuedoknots (Ringquist et al., 1995). Mogridge and Greenblatt (1998) described S1's ability to bind the rrn anti-terminator BoxA (Mogridge & Greenblatt, 1998). It is evident that S1 interacts with a hierarchy of RNA targets. Examining the structure of S1 can provide insight into its range of functions by revealing how it specifically interacts with its RNA substrates.

2.5 DOMAIN ARCHITECTURE AND STRUCTURE OF S1

2.5.1 S1 DOMAINS

S1 is composed of six repeating motifs: M1-M6, as illustrated in figure 2.2a (Salah et al., 2009; Subramanian, 1983). These motifs contain conserved oligonucleotide-binding (OB) fold formed by five β-strands, which are integral components of RNA-binding proteins in all organisms across each domain 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, Rienhardt, Kimura, & Suryanarayana, 1981). This bi-domain structure allows S1 to bind various protein and RNA substrates, enabling its participation in multiple cellular functions.

High-resolution structures of individual S1 domains have been reported. Protein binding motifs 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. 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 have similar structural characteristics. In studies concerning the cellular role of S1, it will be important to consider the contributions of its individual motifs. 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 2.3). 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 may 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).

2.5.2 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). 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 about 250 Å (Labischinski & Subramanian, 1979; Laughrea & Moore, 1977). Although this hydrodynamic data has been collected by multiple groups, no 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 difficult 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 cryogenic electron microscopy (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, Rappsilber, & Reilly, 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 2.2b), 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 4 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.

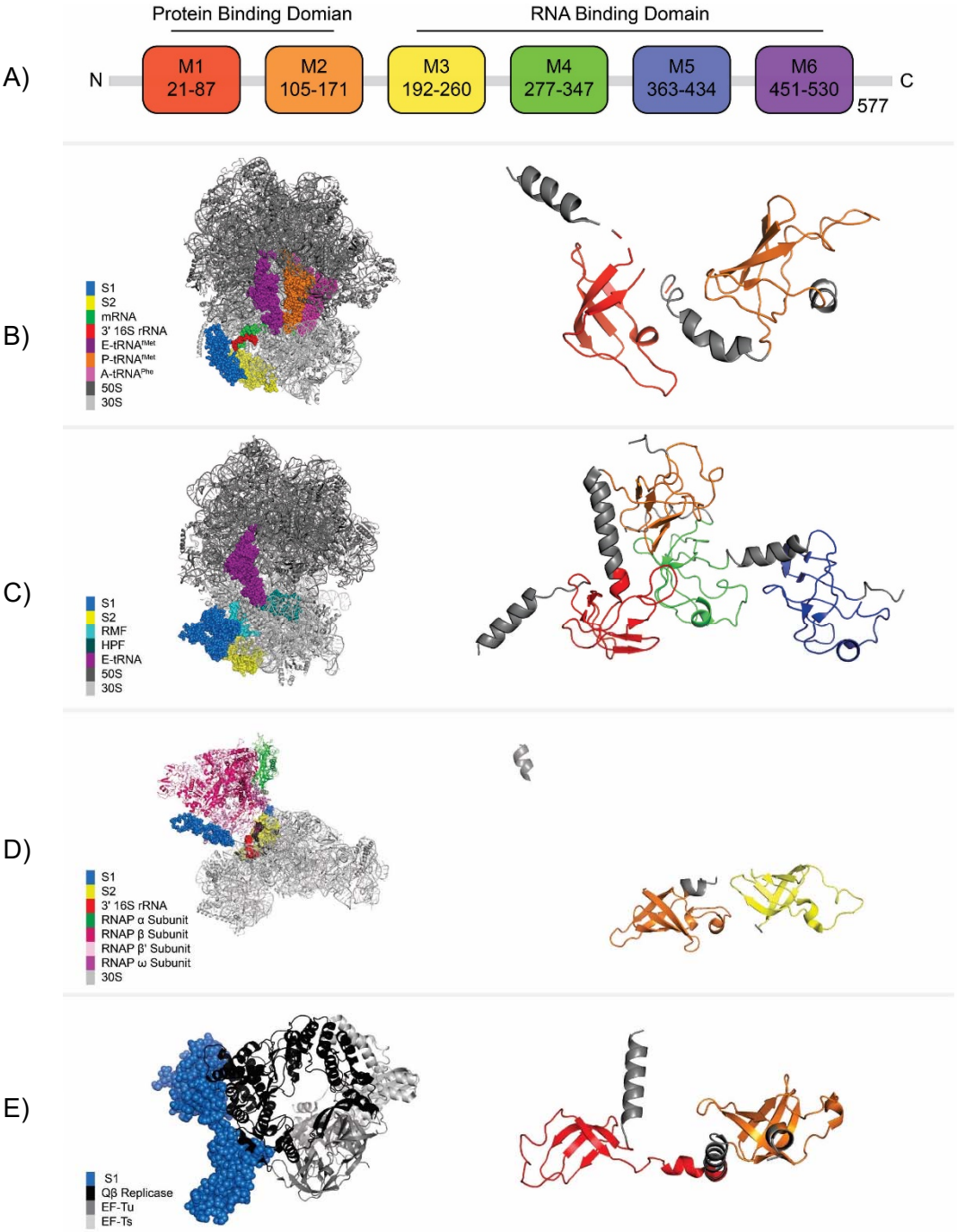


Figure 2.2 High-resolution structures of complexes containing S1. A) Domain architecture. Color scheme is used in the subsequent cartoon representations of S1. B) 70S ribosome (Ensemble cryo-EM, PDB 6BU8). C) Hibernating 70S ribosome (Single-particle cryo-EM, PDB 6H4N). D) RNA Polymerase – 30S Complex (Single-particle cryo-EM, PDB 6AWB). E) Qβ replicase (Crystal structure, PDB 4Q7J). S1 cartoons on the right indicate the protein resolved in the complex.

2.6 FUNCTIONS OF RIBOSOME-BOUND S1

2.6.1 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 remain elusive. Early studies identified the importance of S1 by utilizing translation assays devoid of S1; its absence resulted in a significant decrease in the translation of natural and synthetic mRNAs (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 mRNA's 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 base pairs, prohibiting the strands from reannealing (Qu, Lancaster, Noller, Bustamante, & Tinoco, 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 prohibits accommodation onto the 30S unless S1 is present to unwind its structure (Duval et al., 2013). It is now understood that S1 is absolutely essential 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 increase 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 that do not contain G residues and are therefore unable to strongly interact with the aSD were used to understand the role of S1 in recruiting mRNAs lacking an SD element (Tzareva et al., 1994). In this study S1 binding sites other than U- and AU-tracts, including repeats of CAA, were identified. 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 messages that lack an 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, Roberts, & Rabinowitz, 1992), and leaderless mRNAs. Leaderless mRNAs are believed to initiate directly onto 70S ribosomes, and show no requirement for S1 (Moll, Grill, Grundling, & Blasi, 2002; Moll, Resch, & Blasi, 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? If yes, what is the recognition element? 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 may 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 inadvertent 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 later became understood 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?

2.6.2 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 function 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. However, poly(A)-dependent binding of Lys-tRNA did not require S1, despite its requirement during the translation of a poly(A) mRNA (Linde, Quoc Khanh, Lipecky, & Gassen, 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.

2.6.3 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 requires 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 adopts a much more compact conformation, making it easier to resolve (Figure 2.2c). In this structure, S1 is only 100 Å across compared to the 230 Å measured for S1 on a typical 70S (Labischinski & Subramanian, 1979). 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.

2.7 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.

2.7.1 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 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 level was observed. To investigate the specific role of S1 in this process, the ability of an S1 variant 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, it may be of general interest to investigate the contribution of S1's individual RNA-binding domains for their impact on transcriptional activation, as this may carry over into 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 the coupling of transcription/translation. 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 2.2d) – 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.

2.7.2 S1 AND TMRNA TRANSLATION

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, Roche, & Sauer, 2000). Because S1 co-purifies with and crosslinks to tmRNA, it was believed to somehow be involved in this function (Bordeau & Felden, 2002; Karzai & Sauer, 2001; Wower, Zwieb, Guven, & Wower, 2000). However, a 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.

2.7.3 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. The factors that influence the translation efficiency of an mRNA may influence its stability due to the fact that mRNAs being translated on polysomes are protected from degradation. 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. The Boni group have studied mRNAs that

contain AU-rich enhancer elements that display improved stability which correlates with improved translation efficiency (Komarova, Tchufistova, Dreyfus, & Boni, 2005). In this scenario, S1 likely increases mRNA stabilization primarily through increasing translation efficiency, but perhaps also by competing with RNase E for binding to an mRNA during translation initiation it can prevent mRNA cleavage.

It has been observed that over-expression of S1 results in hindered cell growth, translation inhibition, as well as stabilization of most cellular mRNAs (Briani et al., 2008; Delvillani, Papiani, Deho, & Briani, 2011), demonstrating that S1's affinity for mRNAs can be detrimental when there is too much free S1 in the cell. 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.

2.8 PHAGE RECRUITMENT OF S1

2.8.1 S1 IS A COMPONENT OF THE QB REPLICASE

The Qβ virus is a positive-sense, single-stranded RNA virus that infects *E. coli*. Its genome is about 4000 nucleotides long and encodes four proteins; one, the RNA-dependent RNA polymerase is responsible for replicating and transcribing the Qβ viral RNA (Blumenthal & Carmichael, 1979). 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, Landers, & Weber, 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 (Takeshita, Yamashita, & Tomita, 2014) (Figure 2.2e). 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).

2.8.2 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. The authors 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.

2.9 S1 IN OTHER ORGANISMS

The gram-negative bacteria are generally classified by their inability to retain crystal violet dye due to the makeup of their cell wall; conversely, gram-positive bacteria

are stained purple. These two general classes of bacteria are different 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 2.3). Recombinant S1 purified from this organism can supplement *E. coli* ribosomes to stimulate translation of an mRNA with 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 do 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 2.3). 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 2.3b) reveals that M4-M6 are variable across S1 homologues. This may suggest an evolutionary path where the redundant RNA-binding motifs were lost, or perhaps never evolved at all. What function do these motifs have that the homologues of the protein could afford to lose? Future studies analyzing these diverged proteins may contribute to our understanding of each domains function.

Mycobacterium tuberculosis is the tuberculosis-causing high-GC content gram-positive bacteria. 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). The hypothesis is that PZA targets the C-terminal extension on *M. tuberculosis* S1 (Figure 2.3b), 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 experiments previously performed and found that PZA does not target either the process of trans-translation, or S1 (Dillon et al., 2017).

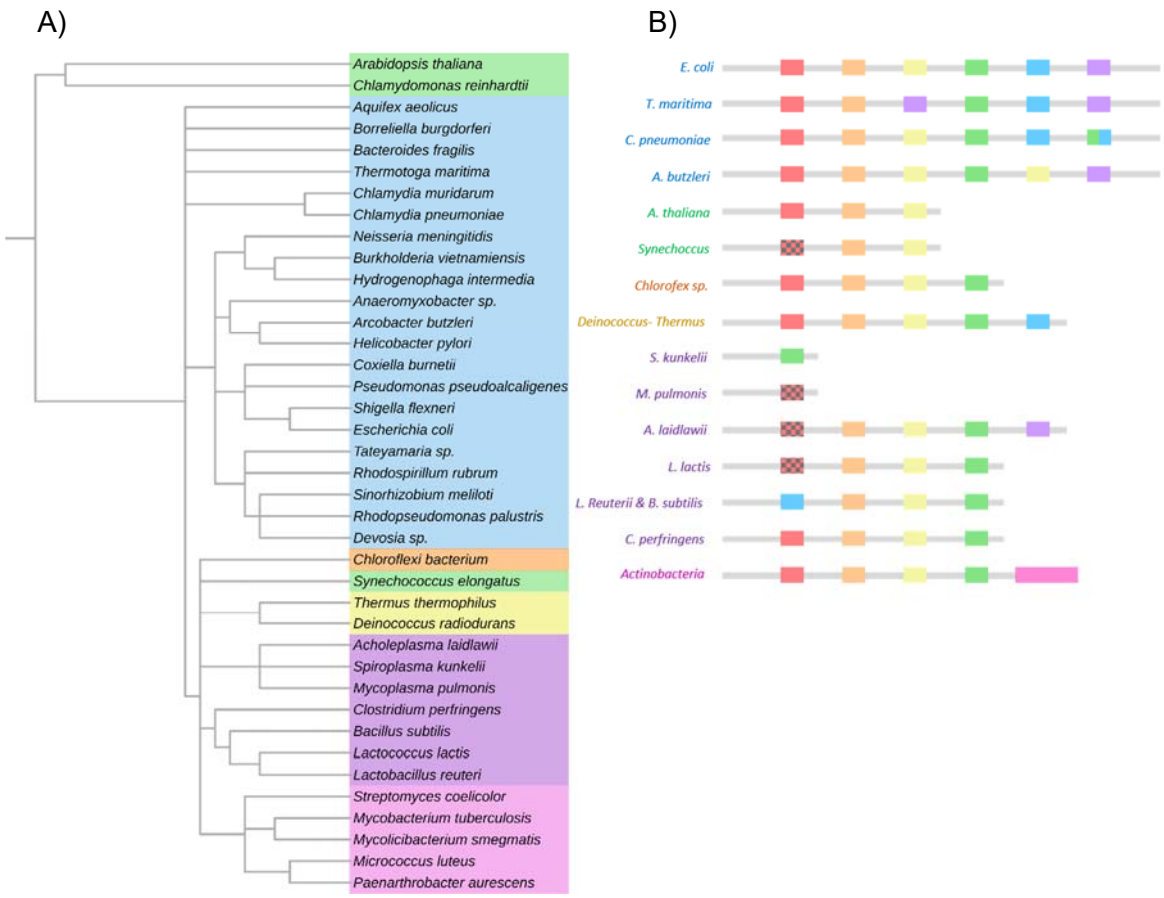


Figure 2.3 Analysis of S1 homologues A) Phylogenetic tree of organisms containing S1. B) Domain architecture of representative S1 homologues. Identification of motifs are as designated in Salah *et al*. 2009. Patterned M1 indicates low conservation as determined in Salah *et al*. 2009. Color scheme: green=eukaryota, blue=gram negative, orange=chloflexi, yellow=deinococcus-thermus, purple=low GC gram positive, pink=high GC gram positive.

Table 2.1 Analysis of S1 Homologues. 49 homologues of S1 were analyzed for sequence similarity/identity using EMBOSS Water Pairwise Sequence Alignment (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*	# of S1 Motifs	Length
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	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

2.10 CONCLUSIONS

Despite the important role that S1 plays in gene expression in *E. coli*, we do not have a complete understanding of how it carries out its important role during translation initiation. Questions that remain include how does S1 recruit 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 may 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). Additionally, S1 is recruited by phages to help carry out processes during phage infection. S1's unique structure allows it to bind proteins and RNAs

simultaneously, however it is not well understood how specific protein/RNAs 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 a 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 3: THE *RPSA* TRANSLATION INITIATION REGION: BIOPHYSICAL CHARACTERIZATION OF A STRUCTURE-DRIVEN TRANSLATION INITIATION MECHANISM

3.1 ABSTRACT

Modulating the efficiency 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 Shine-Dalgarno (SD)-anti-SD complementarity and level of 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. Here we characterize a non-canonical translation initiation element: the TIR of the *E. coli rpsA* gene, which codes for ribosomal protein S1. Although lacking a canonical SD, the *rpsA* TIR is one of the most efficient drivers of translation initiation in *E. coli*. Additionally, the TIR contains a large amount of secondary structure that would typically be considered inhibitory for initiation but in this case is strictly required. We have mapped the sequence-function relationships of the *rpsA* TIR using large-scale mutagenesis and next-generation sequencing. Utilizing this library, we performed structural analysis on select *rpsA* variants and were able to link their activity levels to structural elements within the TIR. We investigated activity of the TIR in a minimal transcription/translation system to show that it functions in the absence of additional cellular factors, and both requires S1 for translation and is specifically repressed by excess 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.

3.2 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 driving translation. Translation initiation efficiency is strongly anti-correlated with a kinetic competition of RBS unfolding versus dissociation of the mRNA from the 30S subunit (Mustoe et al., 2018), while RNA structure acts to attenuate the initiation rate (de Smit & van Duin, 1990; Goodman, Church, & Kosuri, 2013; Salis, Mirsky, & Voigt, 2009). Structured RNA elements surrounding start codons are able to detect changes in cellular Mg2+ concentration (Cromie, Shi, Latifi, & Groisman, 2006), pH (Nechooshtan, Elgrably-Weiss, Sheaffer, Westhof, & Altuvia, 2009), concentrations of carbon sources and metabolites (Wang, Lee, Morales, Lim, & Breaker, 2008), as well as temperature changes (Narberhaus, Waldminghaus, & Chowdhury, 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 strength.

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 may exist (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 structure of the IGR IRES (internal ribosome entry site) in complex with the bacterial ribosome was reported that illustrated an inherent ability to support structure-based initiation in addition to the SD-based initiation mechanism (Colussi et al., 2015). However, no such structure of a bacterial mRNA exists.

The *rpsA* gene coding for ribosomal protein 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 substitution and orthogonal ribosomes provide additional evidence that its translation is not driven solely by an SD-based mechanism (Skorski et al., 2006). Despite its strong secondary structure, large size, and lack of a canonical SD sequence, the *rpsA* 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 nucleotides upstream of the start codon of the *rpsA* mRNA constitutes a minimum segment required to efficiently drive translation (Tchufistova et al., 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 one and two. 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, we performed an in-depth mapping of the *rpsA* TIR activity using high-throughput mutagenesis followed by phenotypic cell sorting and next generation DNA sequencing. We find that mutations in the *rpsA* TIR can achieve varied translation initiation efficiencies, which we utilized to construct a standardized TIR library with different initiation strengths covering three orders of magnitude in expression strength. SHAPE guided near-atomic level structures constrained by *ab initio* models built from solution X-ray scattering enabled us to investigate how the respective point mutations affect the 3-dimensional structure of the *rpsA* TIR. Our results indicate that the 91 nt long TIR of the *rspA* gene constitutes a bona-fide structure-based initiation element, providing the first evidence that a structure based translation initiation mode exists in bacteria.

Table 3.1 Comparison of *rpsA* TIRs from different organisms. Sequences were chosen by selecting 150 basepairs (bp) upstream from the start codon plus 30 bp downstream. Pairwise sequence alignments were performed using EMBOSS Water Pairwise Sequence Alignment (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

3.3 METHODS

3.3.1 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 S3.3.

3.3.2 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 (E0546S, NEB) 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 (CDS) 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 (#200552, 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 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 reaction 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 volume 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 S3.3.

3.3.3 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 (230132, 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 of 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 5000 *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 bandpass filter of 510/80 nm, while mRFP was excited at 561 nm and measured with a splitter of 655 nm and a bandpass filter of 695/40 nm. A scatter gate was set using the forward and side scatter area of *E. coli* cells harboring a non-reporter containing control plasmid. The wild type reporter plasmid was used to set the gate for wild type-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.

3.3.4 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 5000 *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, F530L), 2U Phusion polymerase (Thermo Scientific,

F530L), and a 10 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 (LB to fill in details on custom script). Clustering on Z-scores was calculated on individual libraries (Z-score = (x – mean(X))/sd(X) (# SDs above or below mean)).

3.3.5 PURIFICATION OF 30S, S1, AND 30S^{-S1}

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 and 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 100 µg/mL ampicillin (Amp) or 35 µg/mL Cam, depending on the plasmid used. 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 5000 *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, 17-5166-01) equilibrated with 20 mM Tris-HCl pH 7.5, 40 M 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/$M^{-1}cm^{-1}$. 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 $M^{-1}cm^{-1}$. 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 et al., 1981). Polyuridylic acid-agarose lyophilized powder (Sigma Aldrich, P8563) 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 analysis and mass spectrometry (U of L Mass Spectrometry Facility).

3.3.6 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 5000 *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 gram 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, 1537302) 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 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 4000 *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 volume of 100% glycerol.

3.3.7 RNA PREPARATION AND PURIFICATION

See Supplementary Table S3.2 for plasmids encoding pT7 controlled synthetic *E. coli* mRNAs and Table S3.4 for RNA sequences. The promoter and DNA sequence encoding the RNA of interest was amplified from the plasmid via PCR. The *rpsA*-eCFP mRNA was amplified using primers JVO-059 and JVO-060; the *rpsA*-mRFP mRNA was amplified using JVO-060 and JVO-163; the minimal *rpsA* mRNA was amplified using JVO-060 and JVO-075; the SHAPE RNAs were amplified using JVO-155 and JVO-156. These primers were used to amplify the area of interest in a PCR containing 0.5 μM each

primer, 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, ER1701) 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. The DNA template was subsequently digested with 2 U/mL DNase I (Thermo Scientific, ER1701) 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 seconds, and centrifuged at 5000 *g* for 10 min at 4 °C. The aqueous layer was removed and mixed with an equal volume of chloroform and again vigorously vortexed for 30 s. The mixture was centrifuged at 5000 *g* for 10 min at 4 °C and the aqueous layer removed and this chloroform wash step was repeated. The RNA was then precipitated by adding 0.7 vol. isopropanol and 1/10 volume 3 M NaOAc, followed by incubating for 60 min at -80 °C. The mixture was then centrifuged at 5000 *g* for 30 min at 4 °C. The supernatant was decanted. The resulting pellet was washed in ice-cold ethanol, centrifuged at 5000 *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, 17517501) 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, BS91315) following the manufacturers guidelines, eluting RNA in ultrapure H_2O.

3.3.8 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 pH < 7.0, 95% ethanol), and harvested via centrifugation at 5000 *g* for 15 min, and the pellets shock-frozen in $N2_{(l)}$ 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, BS1361(SK8655)) 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, 89863) overnight at 37 °C and purified using the EZ-10 Spin Column RNA Cleanup and Concentration Kit (Bio Basic, BS91315). 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, and ultimately primer sets JHO-005/JHO-006, and JHO-007/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, 95049-100). 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 min, 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, 95055-100). 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 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 versus 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.

3.3.9 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 16 nM – 0.5 μM 30S or 160 nM – 10.0 μM S1. The entire reaction (50 μL) was applied to a nitrocellulose filter (0.2 μm, 25 mm diameter, GE Healthcare Cat No. 7182-002), 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). K_D's (dissociation constants) and standard deviations were calculated by quantifying the bound

RNA as protein was titrated using the equation y = Bmax * x / (K_D + x), performed in GraphPad Prism. Bmax is the amplitude, or final level of bound S1/30S.

3.3.10 SELECTIVE 2'-HYDROXYL ACYLATION ANALYZED BY PRIMER EXTENSION (SHAPE)

2 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-labeled primer to a total volume 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 labeled 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 labeled 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.

3.3.11 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 (Rg) and forward scattering I(0) values. Further processing in GNOM (D.I. Svergun, 1992) generated pair-distribution function information as well as real-space Rg and maximum particle dimension (Dmax) 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. We chose clusters that were represented by greater than 5/50 members (Figure S3.3), and used these for further analysis.

3.3.12 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, we opted to generate an ensemble of models rather than a single averaged structure, in order 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 S3.4). 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 S3.5) 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 in PyMol.

3.3.13 ACTIVITY OF THE *RPSA* TIR *IN VITRO*

Experiments involving the PURExpress *In vitro* Protein Synthesis kit (New England BioLabs, E6800) 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, E00381), 4.5 µL ultrapure H_2O, and 4.0 nM of plasmid DNA template in a reaction volume 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 (Lmax = 476 nm); mRFP was excited at 584 nm and emission scanned between 599-699 nm (Lmax = 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 (#E3313) and experiments were performed according to the manufacturer's guidelines with some alterations. Single reporter plasmids were used in this assay. Reactions consisted of 10 µL Solution A, 3 µL Factor Mix, 0.5 µL RiboLock (Thermo Scientific, E00381), 5.2 nM template DNA and ultrapure H_2O up to a reaction volume 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.

3.4 RESULTS

3.4.1 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, we surveyed its mutational landscape. In order to map the sequence-function relationships of nucleotide variations to their translation efficiency, we linked the translation initiation activity of the *rpsA* TIR to a fluorescent output allowing us to measure gene-expression. To control for compounding factors such as cellular noise, including plasmid copy number, cell size, and stage of cell cycle, the reporter system was design to contain two independently functioning modules, an expression module and noise module allowing for 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 able to quantify *rpsA* TIR driven translation.

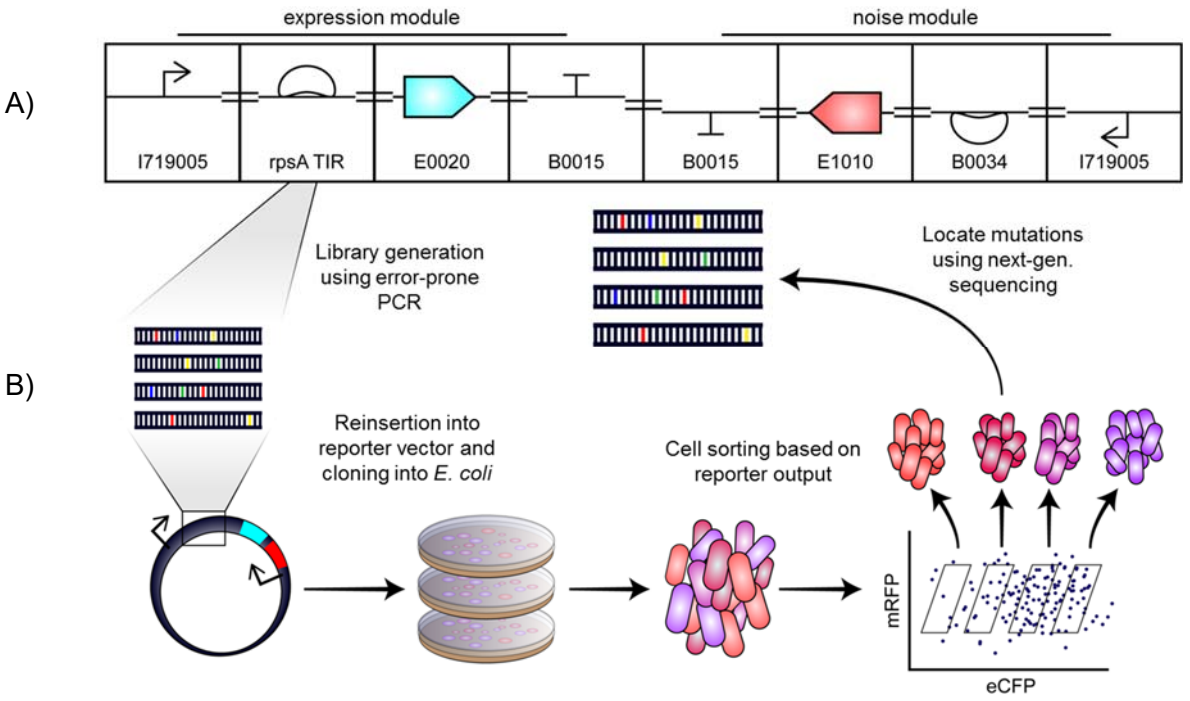


Figure 3.1 Sort-seq strategy for generating a library of translation regulators. A) Reporter construct used to study the *rpsA* TIR *in vivo*. It consists of two independent modules: an expression module containing the *rpsA* TIR, and a noise module to measure cellular gene expression independent of 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*.

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 120,000 colony forming units. 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 3.2). We used a fluorescence-activated cell sorting (FACS) based strategy to isolate similarly expressing cells among the mutant libraries into four distinctly behaving populations (Figure 3.2C).

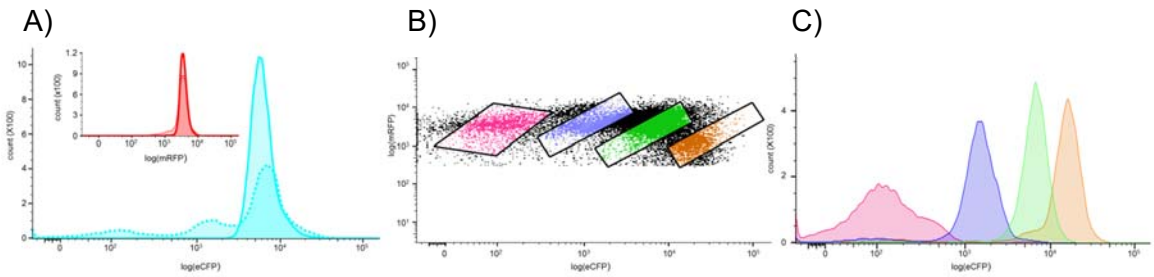


Figure 3.2 The *rpsA* TIR library resulted in an increase in eCFP expression over four order of magnitude. A) Comparing the wt TIR (solid trace) to the mutant library (dotted trace). mRFP expression remains approximately constant in both populations (insert). B) Cells containing the mutant library were fractionated based on their respective eCFP/mRFP ratio into four distinct populations. C) 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 (Qvalue < 0.1) are overlaid on the wt *rpsA* secondary structure in Figure 3.3.

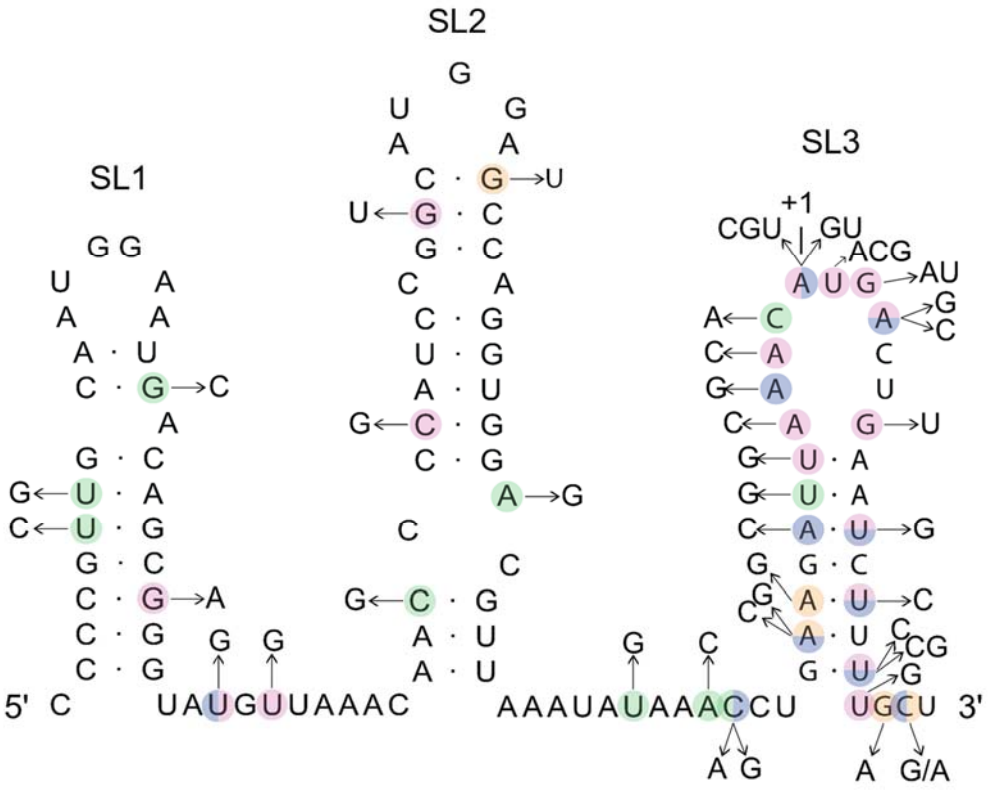


Figure 3.3 High-confidence mutations identified in sort-seq. pipeline mapped onto the *rpsA* TIR. To evaluate significance and validate the pipeline, highly enriched mutations were individually re-introduced into the reporter vector and functionally assayed. Pink=no expression, blue=low expression, green=medium (wt-like) expression, orange=high expression.

Consistent with the essential role of the start codon for translation initiation, mutating the 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 start codon in the eCFP. We also observe a very slight eCFP initiation activity on GUG and UUG start codons, but none at the 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. We observe reduced eCFP fluorescence as GC content increases. Additionally,

when GC content decreases (+16 and +17), we observe increased eCFP. This is consistent with current knowledge that RNA structure in the first roughly 5-10 codons drives this bias (Bentele, Saffert, Rauscher, Ignatova, & Bluthgen, 2013). The previously identified A-9G variant was located in our high expression pool, consistent with previous studies (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 cause weakened stem-loops, for example G-C bp to G-U bp at position -37 and G-C bp to G-G bp at position -31, 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 a select number of variants. The same dual reporter construct used in FACS experiments was utilized for this analysis, where the test eCFP gene was analyzed and the opposite-strand mRFP transcript as an internal reference. The results show no significant difference in transcript abundance between the variants and wt (Figure 3.4) with the exception of Δ-66 and A-9G, which show an 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 cannot explain their differential activity compared to wt.

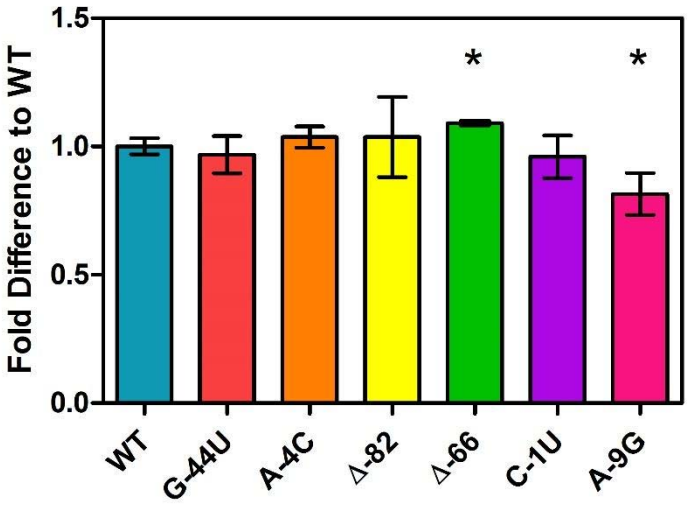


Figure 3.4 Relative abundance of *rpsA* TIR variants compared to wt as determined by RT-qPCR. Fold difference was calculated with the standard curve method, using the mRFP signal as an internal reference for normalization. *Indicates significant difference compared to wt, p-value < 0.05.

3.4.2 *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. We wanted to determine if the fluorescent outputs obtained in our *in vivo* experiments correlate to translation initiation rates predicted *in silico*. We used V2.1 of the Salis Lab RBS calculator, which is among the most advanced translation initiation rate calculators (Figure 3.5). Our results show that the RBS calculator does not always reliably predict our *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.

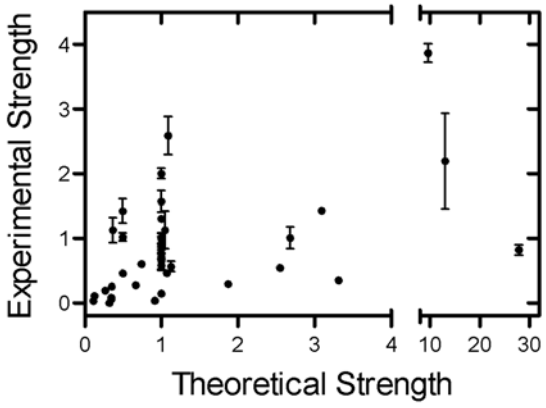


Figure 3.5 *rpsA* driven translation cannot be predicted computationally. Scatter plot comparing the ratio of measure eCFP/mRFP fluorescence to the predicted translation initiation rate. Predicted rates were calculated using version 2.1 of the RBS calculator.

What is contributing to the unpredicted, exceptionally high translation initiation rate of the *rpsA* TIR? We 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 we measured the direct binding of *rpsA* TIR RNAs to the 30S in the absence of the initiator tRNA, before mRNA accommodation.

Several variants isolated from our sort-seq pipeline were analyzed *in vitro* using a filter-binding assay. We selected two low-expressing (A-4C and G-44U) and two high-expressing (C-1U and A-9G) variants. We wanted to investigate how altering the start codon-containing stem-loop affected activity of the TIR; variants C-1U, A-4C, and A-9G are located within this stem-loop. We also wanted to analyze a variant located far from the start codon. A 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). We wanted to investigate G-44U, located in the second stem-loop, to find out why this mutation results in abolished activity.

A correlation would account for all interactions between the 30S and the *rpsA* TIRs *in vitro*, and highlight any interactions currently not included in the thermodynamic models. However, the binding affinities (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 3.6 and Table 3.2). The binding affinity of the *rpsA* TIR variants to 30S is approximately ten-fold higher than their affinity to S1. This suggests that S1 is not the only determinant of binding for the *rpsA* TIR to the ribosome, indicating that the ribosome likely contains additional binding sites for this structured element.

Table 3.2 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_D's were calculated in GraphPad Prism.

rpsA TIR Variant	30S K_D (nM)	S1 K_D (μM)
wildtype	99.4 ± 10.0	2.7 ± 0.3
G-78U	150.0 ± 14.0	----------
A-9G	50.3 ± 7.6	1.6 ± 0.2
G-37U	55.8 ± 12.4	----------
U-63G	37.8 ± 11.3	1.3 ± 0.2
G-44U	117.6 ± 12.6	1.9 ± 0.2
C-1U	41.8 ± 6.4	0.3 ± 0.0
A-4C	35.5 ± 2.2	0.8 ± 0.1

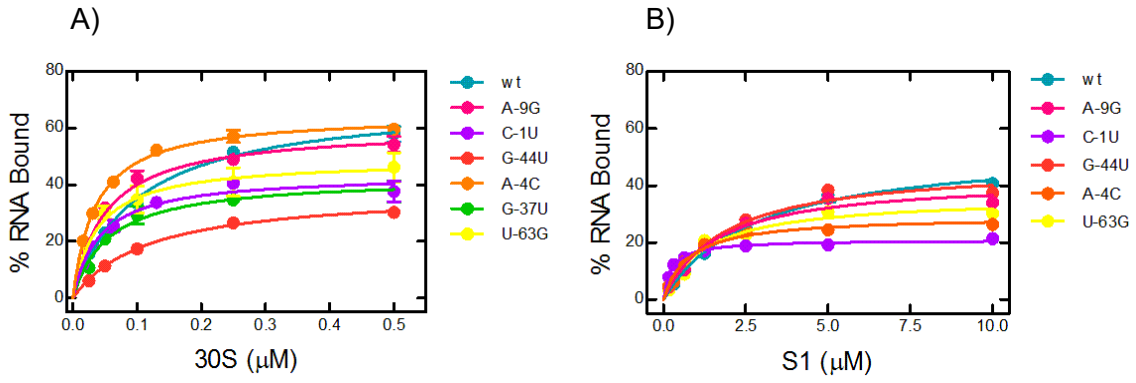


Figure 3.6 *rpsA* driven translation cannot be predicted by 30S or S1 binding strength. Purified ribosomal subunits (30S) (A) and S1 (B) binding to [^{32}P]-labelled *rpsA* TIR variants was quantified by titrating 30S or S1 with RNA. The reaction mixture was subjected to nitrocellulose filtration. The amount of RNA interacting with 30S or S1 was quantified by scintillation counting. Dissociation constants (K_D) for RNA binding were determined using the one-site binding hyperbolic equation in GraphPad Prism. Experiments were performed in triplicate and standard deviation is reported.

3.4.3 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 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, we performed *in vitro* protein synthesis assays using the commercially available PURExpress® system based on highly purified and reconstituted coupled transcription and translation (TX-TL) components (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 3.7A). Additionally, it may be possible that we are altering the cellular concentration of S1 in our *in vivo* experiments by introducing synthetic S1 RNA aptamers in the cell. 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, we tested the activity of select inactive variants (A-4C and G-44C, Δ66, Δ82) and active variants (C-1U and A-9G) 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 that is in fact the case, we should see a decrease in translation of our noise module (mRFP fluorescence). Additionally, this should be dependent on the translation efficiency of the specific variant we have in the reaction. By introducing different mutations into the *rpsA* TIR and putting them in the cell we are altering the level of eCFP expression. The activities of the variants with respect to wt mirror our *in vivo* observations (Figure 3.7B), suggesting

that we are not affecting S1 concentration/binding by introducing variant S1 aptamers into the cell in our *in vivo* experiments.

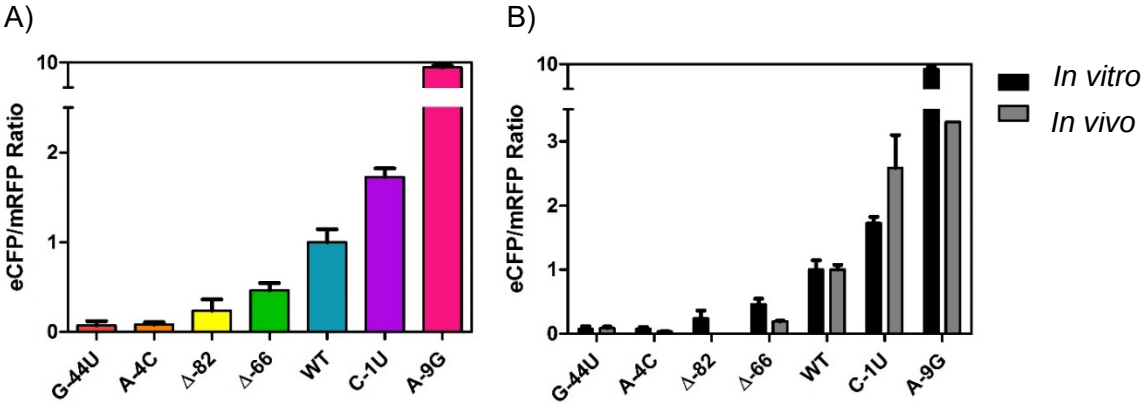


Figure 3.7 *In vitro* activity of *rpsA* TIR variants. A) DNA expressing eCFP under control of *rpsA* TIRs, and mRFP under control of a strong RBS were analyzed in the PURExpress system using eCFP/mRFP fluorescence to detect output via fluorescence spectroscopy. B) Activity of *rpsA* TIR variants *in vitro* versus *in vivo*. Results from PURExpress versus FACS are compared for each *rpsA* TIR variant. Experiments were performed in triplicate and the standard deviation is shown.

Ribosomal protein 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, we utilized a TX-TL system lacking ribosomes that we could supplement with ribosomes from which S1 had been removed. We monitored fluorescence output from reactions seeded with plasmids containing various TIRs that we had previously characterized *in vivo* (Figure S3.1), *rpsO*, *rpsA*, and BBa_B0034, as single reporters expressing eCFP only. BBa_B0034 is a medium to strong RBS with little structure surrounding the start codon; we hypothesized this to be only moderately dependent on S1 for translation. We see 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 B0034 shows no measurable signal. The addition of recombinant S1 at a ratio of 1:1 S1 to ribosomes rescues the activity in all three cases (Figure 3.8). These results agree with toeprinting studies showing that 30S initiation complex formation on *rpsA* as well as SD-containing mRNAs does not occur in the absence of S1, indicating 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 et al., 2002). Other toeprinting studies have shown that RBS's 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 et al., 1992), suggesting that a strong base-pairing interactions are able to compensate for lack of S1 on the ribosome. It is possible that the interaction between 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, we supplemented the TX-TL system 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. At 10-fold excess 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).

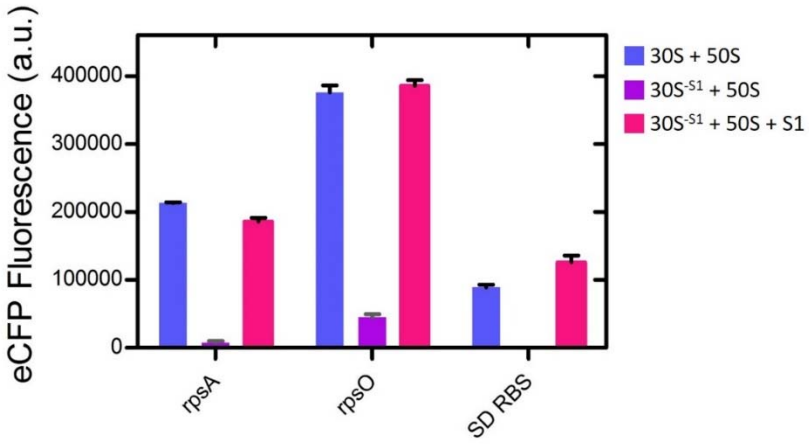


Figure 3.8 Effect of S1 removal on *in vitro* translation assay. The essentiality of S1 was investigated by monitoring the expression of eCFP translationally controlled by *rpsA*, *rpsO*, and an SD-containing RBS (B0034). Experiments were performed in triplicate and the standard deviation is shown.

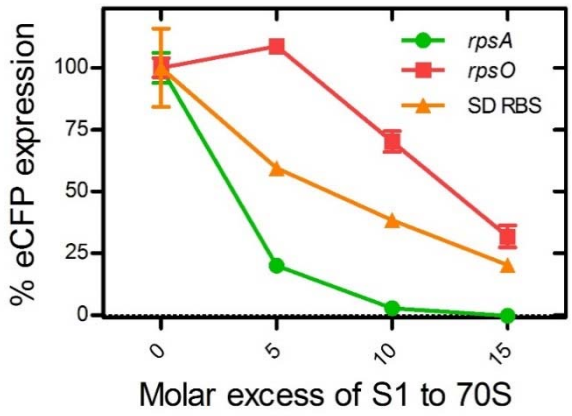


Figure 3.9 Effect of excess S1 on *in vitro* translation assay. The effect of adding S1 in excess over ribosomes to the PURE *in vitro* transcription/translation system was studied for constructs expressing eCFP under the control of *rpsA*, *rpsO*, and an SD-containing RBS (B0034). Experiments were performed in triplicate and the standard deviation is shown.

3.4.4 STRUCTURAL ANALYSIS OF *RPSA* TIR VARIANTS VIA SHAPE AND SAXS

Recent high-throughput SHAPE probing has revealed the dynamic structure of the *rpsA* mRNA, which changes in the presence and absence of protein/ribosomes (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/ribosomes. To obtain a detailed understanding of the three dimensional structure of the *rpsA* TIR variants we performed both *in vitro* and *in silico* structural analysis. We obtained *in vitro* hydrodynamic properties using small angle X-ray scattering (SAXS), and local nucleotide flexibility determined by selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) to guide *in silico* structure prediction of the different *rpsA* TIR variants

The wt and variant *rpsA* TIRs were *in vitro* transcribed and purified (Figure S3.2) for SEC-SAXS. We included the four point mutants and two truncation variants, as well as a minimal version of the wt construct lacking the 3′ scar sequence for comparison to the wt containing both scar sequences (Table S3.4). The resulting pair-distribution plots (Figure 3.10A) indicate that the variants share similar elongated conformations but have differences in their overall shape and hydrodynamic properties (Table S3.1). Analysis of the scattering data to determine low-resolution *ab initio* models reveals that wt-minimal, C-1U, A-9G, A-4C, Δ-66, and Δ-82 each take on a wt-like conformation, with the notable exception of G-44U (Figure 3.10B). This variant has the lowest Dmax value at 133 Å, suggesting that it is more compact relative to wt.

The truncated variants Δ-66 and Δ-82, which include partial or full deletion of stem-loop 1, were intended to determine directionality of the other variants, but truncating

part of the first stem appears to eliminate the native structure (both variants have larger Dmax values than wt despite missing part of the sequence), resulting in elongated conformations that are barely distinguishable from wt in our low-resolution models (Figure 3.10B). This suggests an important role of the first stem-loop in maintaining the tertiary structure of the TIR.

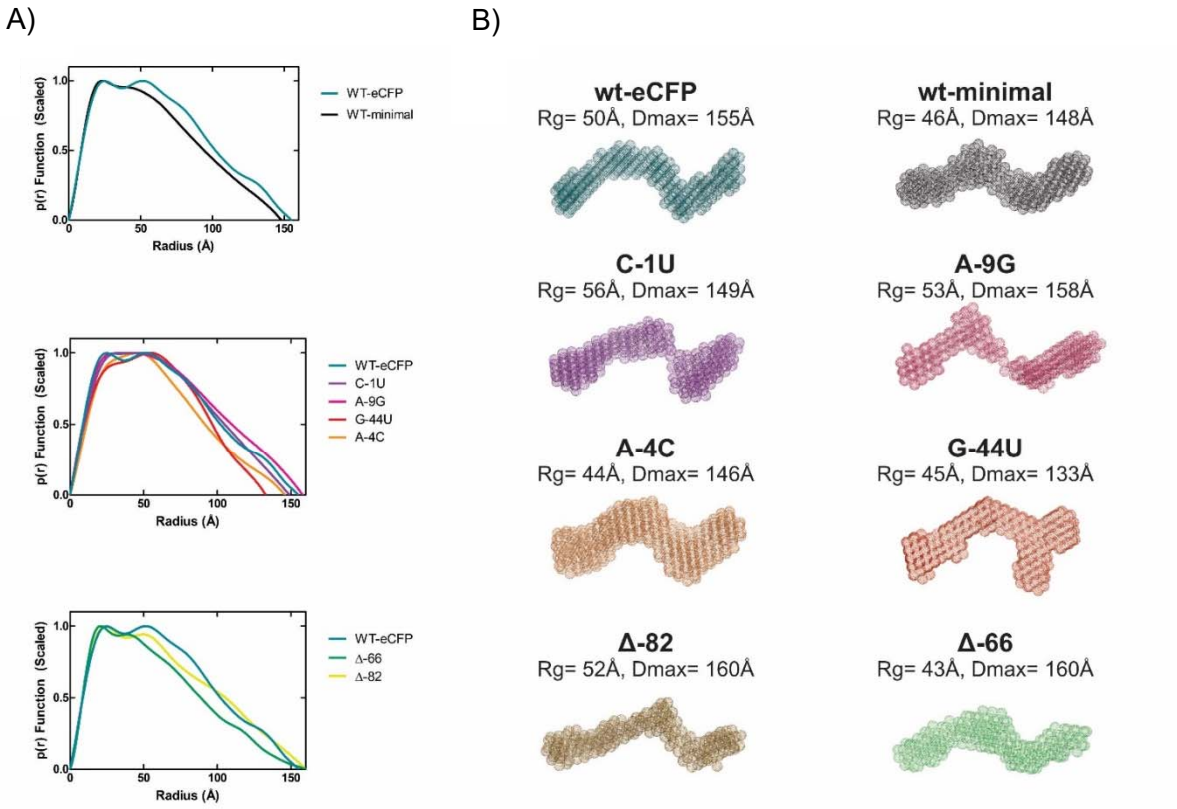


Figure 3.10 Low resolution models of *rpsA* TIR variants. The *rpsA* wt-eCFP and wt-minimal constructs, as well as four point mutants and two truncations were analyzed via SEC-SAXS. A) Pair-Distribution Plots. The SEC-SAXS data was processed into a merged file in Scatter. GNOM was used to generate the 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.

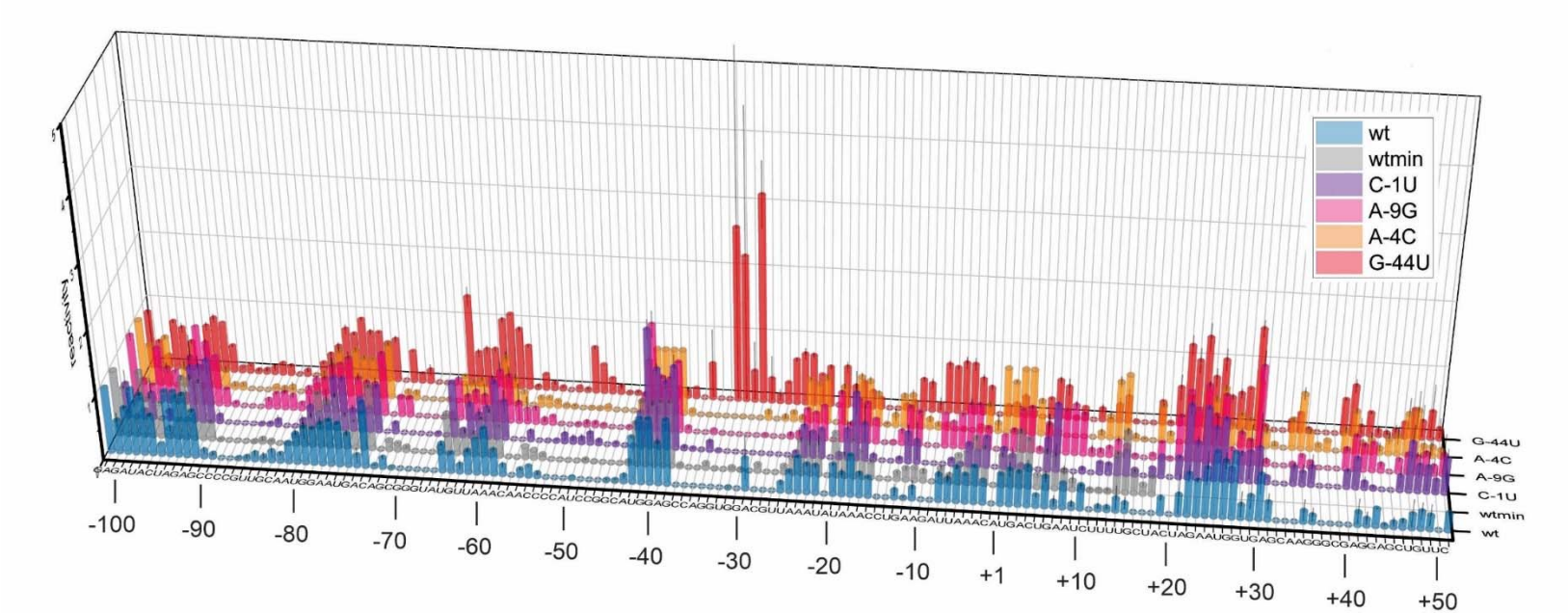


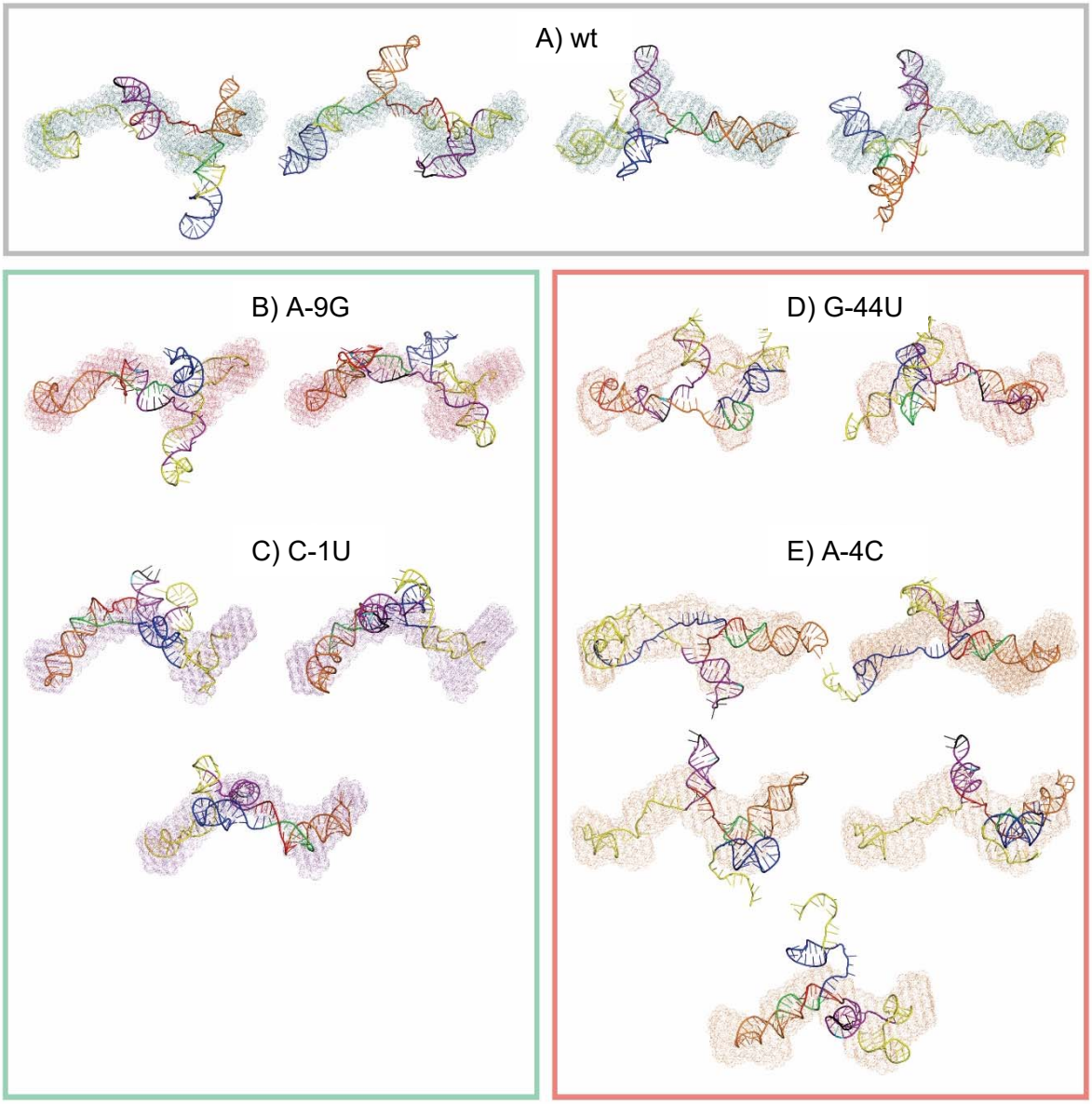
Figure 3.11 SHAPE reactivities of *rpsA* TIR variants. Reactivity values below 0.3 indicate a constrained nucleotide while above 0.7 indicates flexibility of the nucleotide. Nucleotide positions are indicated relative to the start codon at +1.

To gain a higher resolution picture of the structural dynamics of the *rpsA* TIR variants, we performed SHAPE probing analysis. 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 indicates 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 3.11 and Figure S3.4). 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 3.11). The *ab initio* model of the inactive A-4C variant shows that it assumes a wt-like conformation (Figure 3.10B), despite it exhibiting no 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 3.11). The highly active variants A-9G and C-1U maintain a wt-like conformation as expected, except for SL3 of A-9G (Figure 3.11 and Figure S3.4) 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.

We wanted to confirm that the *rpsA* TIR wt maintains its structure even when the extended scar sequence nucleotides at the 3′ end are removed. To do this, we utilized our minimal version of the *rpsA* TIR that lacks the 3′ scar sequence (Table S3.4). The

reactivities are consistent among these RNAs, except for some differences after the start codon (Figure 3.11), suggesting that the three-stem-loop structure is maintained.

Nucleotide reactivities from the SHAPE experiments were used to predict the secondary structures of the variants (Figure S3.4), 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 S3.5). We observe that the *rpsA* TIR is a flexible molecule that adopts different conformations in solution, as it gives multiple bands when run on a native gel (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 SL2, indicating that this clustering method can help to solve different conformations of molecules in solution.



Fig**ure 3.12 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. 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. Models are colored according to scheme in Figure 1.1: SL1=blue, SL2=orange, SL3=purple, with scar sequences in yellow and point mutations in teal.

3.5 DISCUSSION

Bacteria use sophisticated regulatory signals to rapidly respond to changes in their environment. These signals are commonly transmitted through RNA regulatory elements, often embedded in mRNAs, endowing a direct link between translation rate and cellular conditions. Long 5′ UTRs with strong RNA secondary structures and containing multiple regulatory functions are not uncommon in bacteria (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 may be a universal strategy (Colussi et al., 2015). We wanted to determine if any bacterial mRNAs contained three-dimensional architectures with stimulatory effects on translation initiation. In this study we have provided evidence that the *rpsA* TIR forms critical contacts with the ribosome mediated by its specific three-dimensional fold. This insight culminated from a combined *in vivo*, *in vitro*, and *in silico* experimental approach. First, we generated a library of *rpsA* TIR variants which 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 variants that were further studied. Key variants were isolated and subjected to 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 *visa versa*. However, if the affinity between a specific TIR and the 30S was significantly high it may prevent the ribosome from sliding along the mRNA, preventing the transition into

elongation (Komarova et al., 2002). The equilibrium constant 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, we observed that 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 or other unidentified biomolecules 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, we can engineer and predict the activity of synthetic mRNAs 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 in the inability to predict the unusually high translation initiation rate of the *rpsA* TIR (Figure 3.5). 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 SL1 abolishes the function of the TIR (Boni et al., 2001). Here, an inactive variant with a point mutation in SL3 that appears to alter the stability of SL1 was identified. In the structural ensemble for A-4C, SL1 is not always formed (Figure 3.12), as this stem-loop is not restrained in the SHAPE-guided secondary structure prediction (Figure S3.4). This suggests that A-4C exhibits altered dynamics that shift between an intact and disrupted SL1, and increased sampling of the disrupted state may prevent ribosome binding. The second inactive variant studied, G-44U, displays a mainly intact SL1, but a disrupted SL2 (Figure 3.12). SHAPE probing suggests there are differences in the SL2 region of G-44U, as the reactivity of the apical loop nucleotides is severely decreased (Figure 3.11). The G-44U structural ensemble confirms that the entire stem-loop is disrupted, and is instead forming interactions with the nucleotides of SL3 (Figure 3.12). In fact, SL1 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 3.10). This difference is highlighted by the fact that alignment of all *rpsA* TIR *ab initio* models in DAMAVER results in the removal of G-44U due to its dissimilarity, in that its normalized spatial discrepancy (NSD) is greater than the mean NSD + 2*St.Dev. Taken together, the fact that point mutations in the *rpsA* TIR that disrupt SL1 or SL2 result in loss of translation activity, we 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. SL3 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). We observe decreased nucleotide reactivity in SL3 in the inactive variant A-4C, and increased reactivity of SL3 in the highly active variant A-9G (Figure 3.11). Interestingly, the structural ensemble generated for A-9G shows that SL3 is not actually formed in this variant, but rather forms weak interactions with other regions of the TIR (Figure 3.12). This suggests that a third stem-loop structure is not absolutely necessary, as long as the start codon is easily accessible.

Lastly, we observe that the two single-stranded regions of the *rpsA* TIR have a dynamic interaction. This is observed in one of the four wt models, as well as all three of the models in the C-1U ensemble (Figure 3.12). It appears that when these single-stranded regions interact, they elongate SL2 and result in a more compact conformation of the TIR. Because these regions have been shown to constitute binding sites for ribosomal protein S1, perhaps this compact conformation is less amenable to S1 binding.

Ribosomal protein S1 is known to unwind structured mRNAs prior to their accommodation onto the ribosome (Duval et al., 2013), and we have confirmed that this is the case for its own mRNA (Figure 3.8). 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 3.9). 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 *rpsA* and *ssb* (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 3.9). This may be due to a lack of S1-binding sites, such as the AU-stretches present in the *rpsA* 5′ UTR (Boni et al., 2001). The specific fold of the *rpsA* TIR likely contributes to recognition by S1 as well; it has been shown that strengthening of SL2 relieves autogenous control (Boni et al., 2001). It remains to be seen how S1's interaction with the *rpsA* mRNA differs during translation versus 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 SL3 of the *rpsA* TIR plays an important role in start codon accessibility, and that SL1 and SL2 represent a unique structural element that is absolutely required for the function of the *rpsA* TIR (Figure 3.12). This structural element may provide optimal arrangement of the TIR to recruit it to the ribosome, where it is then unwound by ribosomal protein S1. This element functions as an individual entity, with no requirement for additional cellular factors (Figure 3.7). 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 is useful to further explore how the

structure of the *rpsA* TIR mediates specific nucleotide interactions with the ribosome. Additionally, this project will provide insights into mechanisms contributing to non-canonical translation initiation in bacteria.

3.6 SUPPLEMENTAL INFORMATION

3.6.1 SUPPLEMENTAL FIGURES

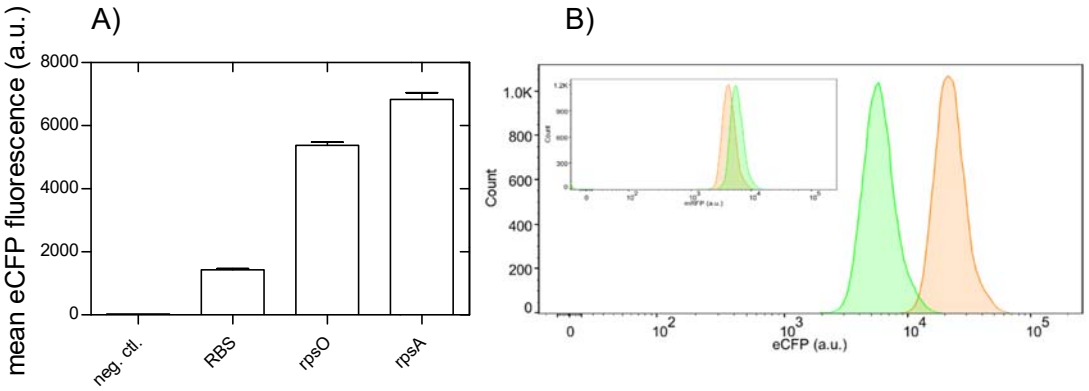


Figure S3.1 Differential protein expression can be resolved using our dual reporter plasmid. A) eCFP expression driven translationally by various TIRs. Fluorescence was measured using flow-cytometry and the geometric mean plotted. Three biological replicates were analyzed. B) The dual reporter plasmid 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.

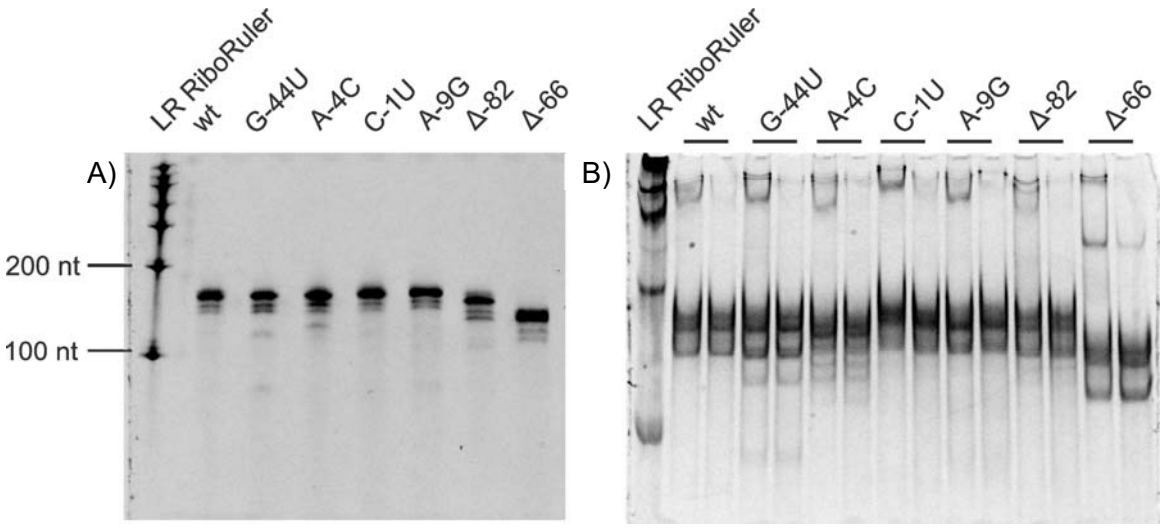


Figure S3.2 *In vitro* transcribed *rpsA* TIR variants. A) RNA was *in vitro* transcribed using T7 RNA polymerase overnight at 37˚C and purified via spin column. A 200 ng sample of RNA was loaded onto an 8% Urea PAGE ran for 25 minutes at 200V, stained with ethidium bromide. B) *rpsA* TIR variants before 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 150V, stained with ethidium bromide.

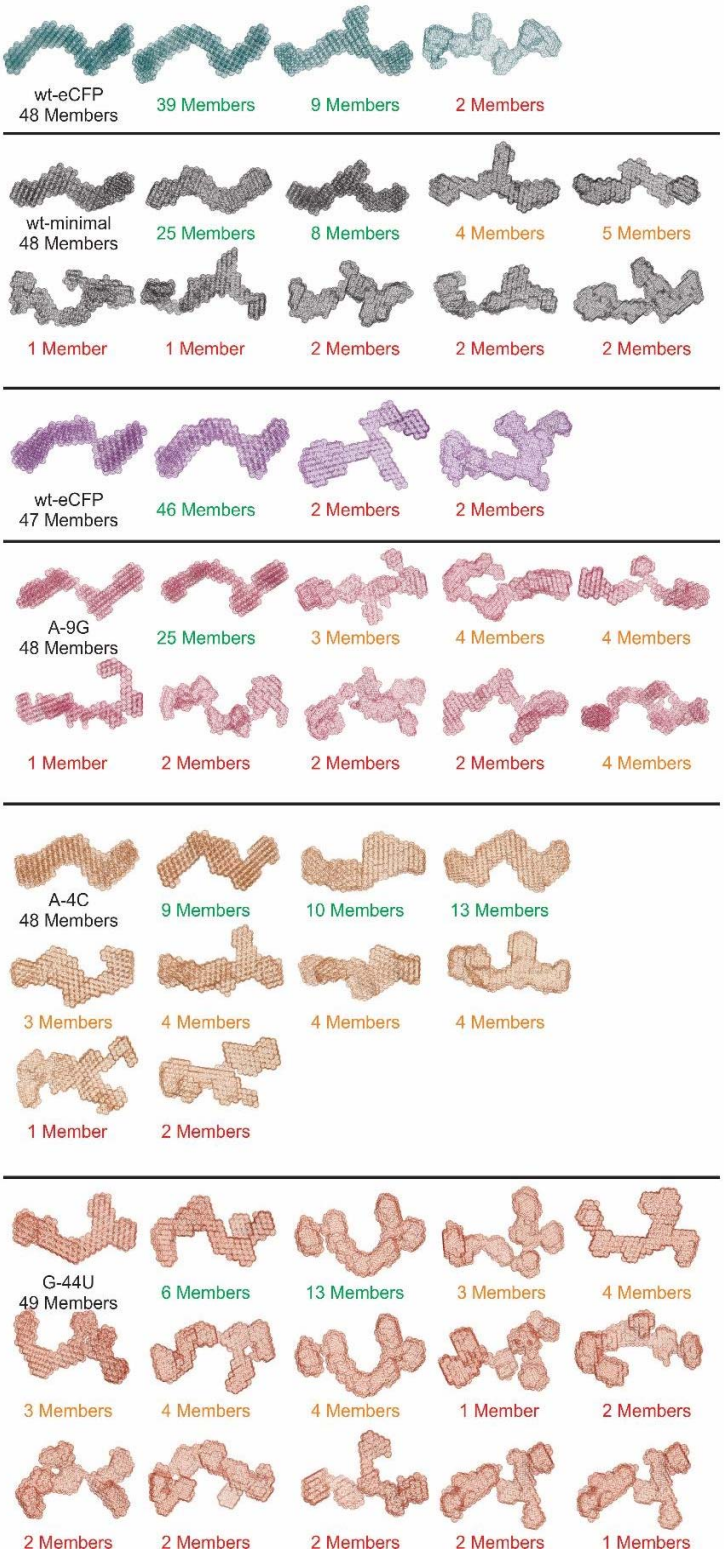


Figure S3.3 Clustering of *rpsA* SAXS envelopes. 50 models for each *rpsA* variant were averaged using DAMCLUST. Clusters with more than 5/50 members were chosen as alternative solutions to the scattering data. The DAMAVER averaged model for each variant is shown first.

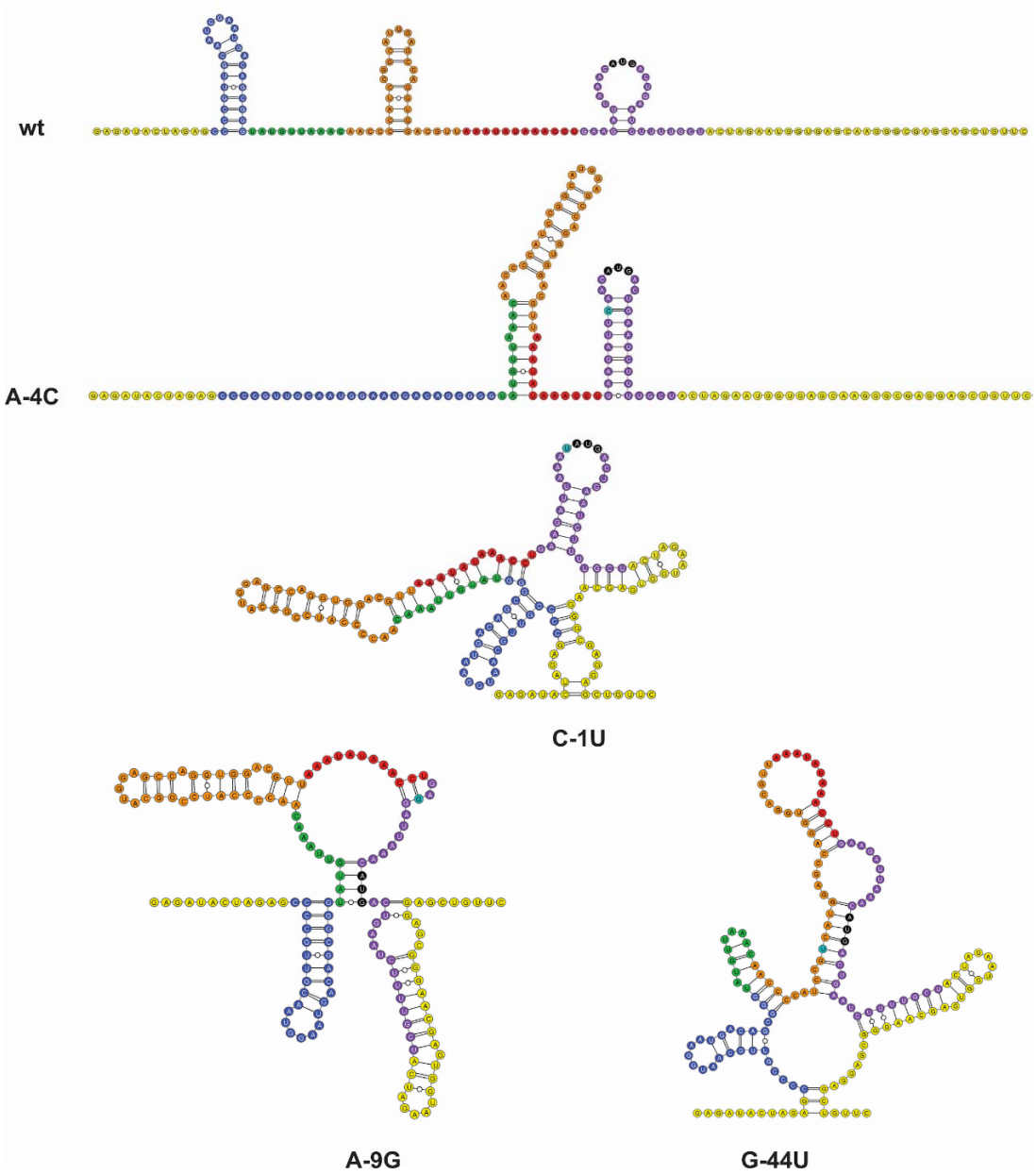


Figure S3.4 Secondary structures predicted using SHAPE reactivities. Secondary structures were predicted using RNAStructure and used as restraints for SimRNA modelling. SimRNA enforces the base-pairing indicated, but does not prevent new base-pairs being formed. Figures were made in VARNA. Stem-loops and single-stranded regions are colored according to the wt. Yellow indicates scar sequence.

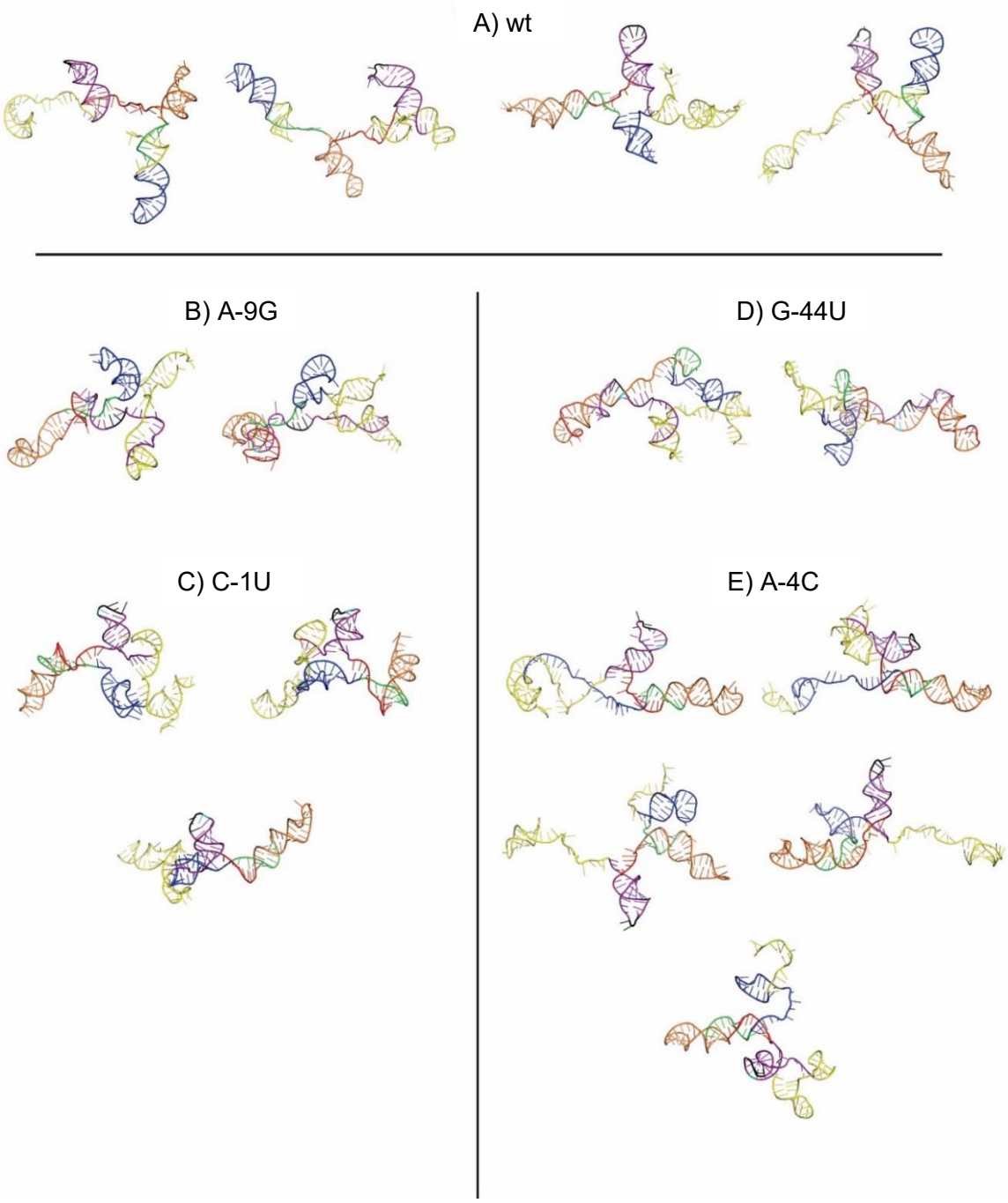


Figure S3.5 Computational models of *rpsA* TIR variants. An ensemble approach was used to generate models of the *rpsA* TIR variants that fit the experimental SAXS and SHAPE data. a) *rpsA* wt: four model ensemble. b) *rpsA* A-9G: two model ensemble. c) *rpsA* C-1U: three model ensemble. d) *rpsA* G-44U: two model ensemble. e) *rpsA* A-4C: 5 model ensemble.Models are colored according to scheme in Figure 1.1, with scar sequences in yellow and point mutations in teal.

3.6.2 SUPPLEMENTAL TABLES

Table S3.1 *rpsA* TIR variants SEC-SAXS parameters. SAXS analysis was performed at Diamond Light Source Synchrotron. Rg= radius of gyration. Dmax= 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
	$X^{\wedge 2}$	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 S3.2 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(rB–mB–) 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(rK–mK+), 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 S3.3 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 S3.4 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	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUA AACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACC UGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGC AAGGGCGAGGAGCUGUUC
minimal *rpsA* TIR	JVS-031	JVO-060 JVO-075	121	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUA AACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACC UGAAGAUUAAACAUGACUGAAUCUUUUGCU
A-9G *rpsA* TIR	JVS-011	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUA AACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACC UGAGGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGC AAGGGCGAGGAGCUGUUC
G-44U *rpsA* TIR	JVS-051	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUA AACAACCCCAUCCGUCAUGGAGCCAGGUGGACGUUAAAUAUAAACC UGAAGAUUAAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGC AAGGGCGAGGAGCUGUUC
C-1U *rpsA* TIR	JVS-063	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUA AACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACC UGAAGAUUAAAUAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGC AAGGGCGAGGAGCUGUUC
A-4C *rpsA* TIR	JVS-059	JVO-060 JVO-059	154	filter binding, SAXS	GAGAUACUAGAGCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUA AACAACCCCAUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACC UGAAGAUUCAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGC AAGGGCGAGGAGCUGUUC
Δ-82 *rpsA TIR*	JVS-125	JVO-060 JVO-059	145	SAXS	GAGAUACUAGAGAAUGGAAUGACAGCGGGUAUGUUAAACAACCCC AUCCGGCAUGGAGCCAGGUGGACGUUAAAUAUAAACCUGAAGAUU AAACAUGACUGAAUCUUUUGCUACUAGAAUGGUGAGCAAGGGCGA GGAGCUGUUC
Δ-66 *rpsA TIR*	JVS-126	JVO-060 JVO-059	129	SAXS	GAGAUACUAGAGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCA GGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCU UUUGCUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC

***rpsA* TIR**	JVS-144	JVO-060 JVO-059	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGAC AGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACG UUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUAC UAGAAUGGUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGC CGGAUCCAAAUCGGGCUUCGGUCCGGUUCUCGAG
minimal *rpsA* TIR	JVS-139	JVO-155 JVO-156	183	SHAPE probing	GGCCTTCGGGCCAAGAGATACTAGAGCCCCGTTGCAATGGAATGACA GCGGGTATGTTAAACAACCCCATCCGGCATGGAGCCAGGTGGACGTT AAATATAAACCTGAAGATTAAACATGACTGAATCTTTTGCTTCGATCC GGTTCGCCGGATCCAAATCGGGCTTCGGTCCGGTTCTCGAG
A-9G *rpsA* TIR	JVS-143	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGAC AGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACG UUAAAUAUAAACCUGAGGAUUAAACAUGACUGAAUCUUUUGCUAC UAGAAUGGUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGC CGGAUCCAAAUCGGGCUUCGGUCCGGUUCUCGAG
G-44U *rpsA* TIR	JVS-141	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGAC AGCGGGUAUGUUAAACAACCCCAUCCGUCAUGGAGCCAGGUGGACG UUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUGCUAC UAGAAUGGUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGC CGGAUCCAAAUCGGGCUUCGGUCCGGUUCUCGAG
C-1U *rpsA* TIR	JVS-140	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGAC AGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACG UUAAAUAUAAACCUGAAGAUUAAAUAUGACUGAAUCUUUUGCUAC UAGAAUGGUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGC CGGAUCCAAAUCGGGCUUCGGUCCGGUUCUCGAG
A-4C *rpsA* TIR	JVS-142	JVO-155 JVO-156	216	SHAPE probing	GGCCUUCGGGCCAAGAGAUACUAGAGCCCCGUUGCAAUGGAAUGAC AGCGGGUAUGUUAAACAACCCCAUCCGGCAUGGAGCCAGGUGGACG UUAAAUAUAAACCUGAAGAUUCAACAUGACUGAAUCUUUUGCUAC UAGAAUGGUGAGCAAGGGCGAGGAGCUGUUCUCGAUCCGGUUCGC CGGAUCCAAAUCGGGCUUCGGUCCGGUUCUCGAG

CHAPTER 4: INTERACTION BETWEEN *RPSA* TIR AND 70S RIBOSOME

4.1 INTRODUCTION

Due to its large size and important secondary structure, the *rpsA* TIR has been compared to viral IRESs, the structured RNA elements found in eukaryotes and viruses. IRESs are able to recruit ribosomes onto internal mRNA sites to initiate translation in a 5′ cap-independent manner. Considering the conserved structural features contained inside the ribosome, it is possible that an IRES-like signal may exist in all domains of life (Roberts & Wieden, 2018). The *rpsA* TIR represents a candidate for this cross-kingdom translation initiation signal as we have shown that its function relies on its tertiary fold, which we believe may mediate specific contacts with the conserved ribosomal core in order to facilitate translation initiation.

The *rpsA* TIR and selected variants were previously shown to interact with the 30S ribosomal subunit with nanomolar affinity, although no correlation between variant activity and binding strength was observed (Figure 3.6 and Table 3.2). Based on the striking similarities to IRES elements that interact directly with the 70S ribosome (Colussi et al., 2015), it is possible that the TIR follows a similar mechanism. Therefore, I analyzed the interaction between the TIR variants and purified 70S ribosomes.

4.2 METHODS

4.2.1 NITROCELLULOSE FILTER BINDING

RpsA TIR variants were prepared as described in section 3.4.7. The nitrocellulose filter binding method was performed as described in section 3.4.9 with some alterations. RNA and 70S ribosomes were incubated at 37˚C for 15 min in Binding Buffer (20 mM tris-HCl pH 7.6, 5 mM $MgCl_2$, 100 mM NH_4Cl). Reactions contained RNA at a final concentration of 0.5 nM, titrated against 5 nM to 1280 nM ribosomes. K_Ds and standard

deviations were calculated by quantifying the bound RNA as ribosomes were titrated using the equation y=Bmax*x/(K_D+x), which was performed in GraphPad Prism.

4.3 RESULTS AND DISCUSSION

The wt *rpsA* TIR binds to the 70S ribosome with a K_D of 17.5 ± 3.3 nM (Figure 4.1). This indicates that the TIR binds to the 70S with approximately 5-fold higher affinity than the 30S subunit (Table 4.1). As seen with the 30S and S1 filter binding experiments described in section 3.5.2, there does not appear to be a correlation between binding affinity to the 70S and initiation strength as each of the variants, both active and inactive, have K_D's higher than the wt (Figure 4.1). It might be that the inactive variants G-44U and A-4C are able to bind to the ribosome but are unable to initiate translation due to an altered structure that does not facilitate the correct contacts with the ribosomal core in the case of G-44U. A-4C on the other hand may bind to the ribosome but is unable to unwind the strengthened third stem-loop to allow access to the start codon.

Although these results represent initial evidence that the *rpsA* TIR might directly recruit the 70S ribosome, its high affinity may be due to the increased RNA interaction surface of the 70S compared to the 30S. Further experiments will be required to determine the physiological significance of this interaction.

Table 4.1 Binding affinities for the interaction between *rpsA* TIR variants and 70S ribosomes. RNA was incubated with ribosomes/protein for 15 min at 37˚C and binding was analyzed via nitrocellulose filtrations. K_Ds and standard deviation were calculated in GraphPad Prism.

rpsA TIR Variant	70S K_D (nM)
WT	17.5 ± 3.3
A-4C	45.3 ± 3.9
A-9G	43.6 ± 3.7
G-44U	43.2 ± 8.2
C-1U	85.2 ± 10.4

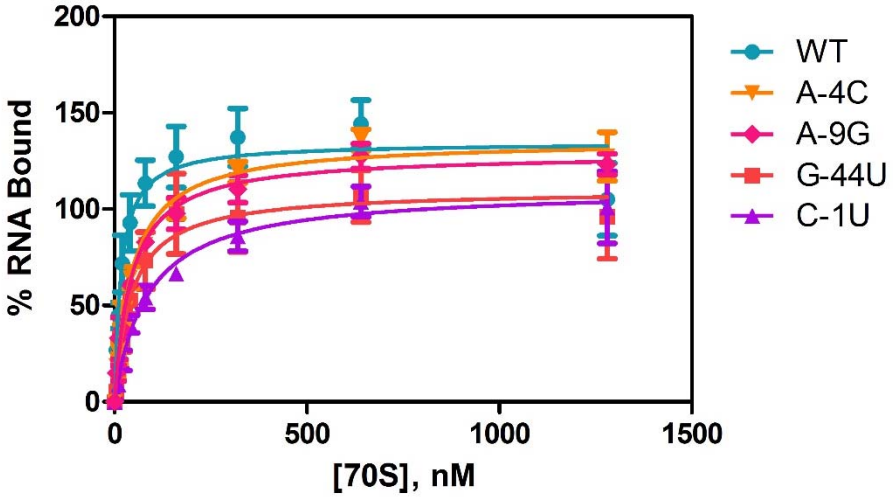


Figure 4.1 Interaction of the *rpsA* TIR and 70S ribosome. Filter binding was performed to assess binding of the wt-eCFP *rpsA* TIR to the 70S ribosome. Experiments were performed in triplicate and fit using the one-site binding hyperbolic equation in GraphPad Prism.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

Here I performed *in vitro* assays and structural analysis on the *rpsA* TIR to elucidate the role that structure plays in its function during translation and regulation by S1. First, I have shown that introducing point mutations into the *rpsA* TIR does not influence abundance of the variant mRNAs in the cell (Figure 3.4). Through *in vitro* TX/TL experiments, I have shown that the *rpsA* TIR does not require additional cellular factors for its function, confirming that its high activity is an intrinsic property of the TIR (Figure 3.7). Additionally, the *rpsA* TIR requires ribosomal protein S1 for its activity, and is specifically repressed by S1 (Figures 3.8, and 3.9). Through our structural analysis, we have shown that the structural basis for the function of the *rpsA* TIR may involve 1) dynamic conformational switching from an open to a closed state, 2) the presence of SL1 and SL2, and 3) the ability of SL3 to unfold in order to allow the ribosome to access the start codon (Figure 3.11). In particular, we have shown that introducing a single point mutations 44 nucleotides upstream of the start codon (G-44U) results in abolished activity as well as structure. Finally, I show that the *rpsA* TIR displays a high affinity for the 70S ribosome (Figure 4.1). These results strongly suggest that the mechanism employed by the *rpsA* TIR relies on a specific tertiary fold. A structure-based mechanism such as this is a novel characteristic in bacterial translation, and may contain similar characteristics to viral and eukaryotic IRESs.

To confirm that the G-44U *rpsA* variant that we have characterized does indeed show abolished activity due its altered structure, it would be beneficial to characterize a compensatory mutation whereby the opposite nucleotide is mutated to an adenine, which would restore the base-pair at that position and presumably rescue normal activity. Similar compensatory mutations could be performed with our other variants as well. Additionally,

future studies will aim to identify the contacts made between the *rpsA* TIR and the ribosome in order to determine the role of SL1 and SL2 in ribosome recruitment. Although S1 is very large and flexible, and therefore difficult to observe using structural analysis techniques, it may be stabilized upon binding to the *rpsA* TIR allowing its visualization as part of this complex using cryo-EM.

By gaining a better understanding of the structure-function relationship of ribosomal protein S1 and the *rpsA* TIR, I contributed information that will be critical in utilizing these elements in fields such as synthetic biology and antibiotic discovery. The *rpsA* TIR constitutes a non-canonical translation initiation element that can display a large range of gene expression strength simply by changing a single nucleotide, making it a candidate for fine-tuning gene expression at the translation initiation level. Additionally, understanding the role of ribosomal protein S1 during translation initiation will be important to the further development of RBS calculators that are used for synthetic biology and bioengineering. Finally, as an essential protein in gram-negative bacteria (and that does not have a homolog in humans), ribosomal protein S1 represents a potential target for the development of new antibiotics.

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. doi:10.1074/jbc.M707111200

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.

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. doi:10.1038/s41564-018-0237-0

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. doi:10.1038/msb.2013.32

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. doi:10.1074/jbc.M212731200

Blumenthal, T., & Carmichael, G. G. (1979). RNA replication: function and structure of Qbeta-replicase. *Annu Rev Biochem, 48*, 525-548. doi:10.1146/annurev.bi.48.070179.002521

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. doi:10.1093/emboj/20.15.4222

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. doi:10.1093/nar/gkv1479

Bordeau, V., & Felden, B. (2002). Ribosomal protein S1 induces a conformational change of tmRNA; more than one protein S1 per molecule of tmRNA. *Biochimie, 84*(8), 723-729.

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. doi:10.1261/rna.1123908

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. doi:10.1093/nar/gku1314

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. doi:10.1371/journal.pone.0032702

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat Biotechnol, 36*(10), 1005-1015. doi:10.1038/nbt.4238

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. doi:10.1021/bp9801143

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. doi:10.1016/j.gene.2006.01.033

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*. doi:10.1093/jb/mvz006

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*(7541), 110-113. doi:10.1038/nature14219

Cromie, M. J., Shi, Y., Latifi, T., & Groisman, E. A. (2006). An RNA sensor for intracellular Mg(2+). *Cell, 125*(1), 71-84. doi:10.1016/j.cell.2006.01.043

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 30 S 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. doi:10.1093/nar/gky274

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. doi:10.1073/pnas.87.19.7668

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. doi:10.1093/nar/gkr417

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*. doi:10.7554/eLife.28560

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. doi:10.1128/jb.176.12.3738-3748.1994

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. doi:10.1093/nar/27.2.381

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. doi:10.1093/nar/gkl911

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. doi:10.1128/jb.01092-07

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. doi:10.1371/journal.pbio.1001731

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. doi:10.1016/j.bbrc.2017.04.048

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 J Appl Cryst, 43, 639-646.

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. doi:10.1007/s12104-014-9554-2

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. doi:10.1146/annurev.mi.35.100181.002053

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. doi:10.1126/science.1241934

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.

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.

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. doi:10.1093/nar/gkt267

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.

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. doi:10.1038/nbt986

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

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.

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. doi:10.1261/rna.036327.112

Karzai, A. W., Roche, E. D., & Sauer, R. T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol, 7*(6), 449-455. doi:10.1038/75843

Karzai, A. W., & Sauer, R. T. (2001). Protein factors associated with the SsrA.SmpB tagging and ribosome rescue complex. *Proc Natl Acad Sci U S A, 98*(6), 3040-3044. doi:10.1073/pnas.051628298

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. doi:10.1093/dnares/dsi012

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. doi:10.1128/jb.187.4.1344-1349.2005

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. doi:10.1016/j.gene.2005.06.037

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.

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. doi:10.1074/mcp.M112.019562

Laughrea, M., & Moore, P. B. (1977). Physical properties of ribosomal protein S1 and its interaction with the 30 S 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. doi:10.1128/mmbr.69.1.101-123.2005

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. doi:10.1074/jbc.M010680200

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. doi:10.1073/pnas.1019383108

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. doi:10.1093/nar/gks636

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.

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. doi:10.1016/j.ymeth.2017.12.004

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. doi:10.1093/nar/gkz268

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. doi:10.1126/science.1100829

Marcus, J. I., Hassoun, S., & Nair, N. U. (2017). Computational prediction of functional abortive RNA in *E. coli. Genomics, 109*(3-4), 196-203. doi:10.1016/j.ygeno.2017.03.003

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. doi:10.1073/pnas.0405521101

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. doi:10.1016/s0076-6879(09)68004-6

McPheeters, D. S., Christensen, A., Young, E. T., Stormo, G., & Gold, L. (1986). Translational regulation of expression of the bacteriophage T4 lysozyme gene. *Nucleic Acids Res, 14*(14), 5813-5826. doi:10.1093/nar/14.14.5813

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. doi:10.1006/jmbi.1997.0939

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

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., 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., 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.

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). doi:10.3791/58538

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. doi:10.1016/j.cell.2018.02.034

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

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., & Altuvia, S. (2009). A pH-responsive riboregulator. *Genes Dev, 23*(22), 2650-2662. doi:10.1101/gad.552209

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.

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.

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. doi:10.1107/s0021889812007662

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. doi:10.1038/nbt1226

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. doi:10.1080/00327488008061715

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. doi:10.1093/emboj/20.8.1829

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. doi:10.1016/j.jmb.2007.02.068

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. doi:10.1073/pnas.1208950109

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.

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics, 11*, 129. doi:10.1186/1471-2105-11-129

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. doi:10.1093/nar/gkj405

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.

Rio, D. C. (2012). Filter-binding assay for analysis of RNA-protein interactions. *Cold Spring Harb Protoc, 2012*(10), 1078-1081. doi:10.1101/pdb.prot071449

Roberts, L., & Wieden, H. J. (2018). Viruses, IRESs, and a universal translation initiation mechanism. *Biotechnol Genet Eng Rev, 34*(1), 60-75. doi:10.1080/02648725.2018.1471567

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.

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.

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. doi:10.1093/nar/gkp547

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. doi:10.1038/nbt.1568

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. doi:10.1006/jmbi.2000.3626

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.

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. doi:10.1126/science.1117230

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. doi:10.1073/pnas.211266898

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. doi:10.1126/science.1208813

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. doi:10.1038/90802

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 30 S ribosomal subunit of *Escherichia coli*. *J Mol Biol, 153*(3), 761-780.

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. doi:10.1128/jb.00591-06

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.

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. doi:10.1006/jmbi.1998.1909

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. doi:10.1186/s12900-019-0103-1

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

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

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. doi:10.1021/bi020638i

Sukhodolets, M. V., Garges, S., & Adhya, S. (2006). Ribosomal protein S1 promotes transcriptional cycling. *RNA, 12*(8), 1505-1513. doi:10.1261/rna.2321606

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. . *App. Cryst, 25*, 495-503.

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical J, 76*(6), 2879-2886. doi:10.1016/S0006-3495(99)77443-6

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. doi:10.1093/nar/gku745

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., 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., 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.

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. doi:10.1111/j.1365-2958.2008.06229.x

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.

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. doi:10.1093/jb/mvm243

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

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.

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. doi:10.1038/ncomms2807

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.

Vigar, J. R., & Wieden, H. J. (2017). Engineering bacterial translation initiation - Do we have all the tools we need? *Biochim Biophys Acta*. doi:10.1016/j.bbagen.2017.03.008

Vind, J., Sorensen, 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. *J Mol Biol, 231*(3), 678-688. doi:10.1006/jmbi.1993.1319

Volkov, V., & I. Svergun, D. (2003). Uniqueness of ab initio shape determination in small-angle scattering *J Appl Cryst*, 36, 860-864.

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 30 S ribosomal protein S1 of *Escherichia coli*. Evidence for the identity of the two proteins. *J Biol Chem, 249*(10), 3314-3316.

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. doi:10.1016/j.molcel.2008.01.012

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

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. doi:10.1038/35030006

Wisniewski, J. R., & Rakus, D. (2014). Quantitative analysis of the *Escherichia coli* proteome. *Data Brief, 1*, 7-11. doi:10.1016/j.dib.2014.08.004

Wower, I. K., Zwieb, C. W., Guven, S. A., & Wower, J. (2000). Binding and cross-linking of tmRNA to ribosomal protein S1, on and off the *Escherichia coli* ribosome. *Embo J, 19*(23), 6612-6621. doi:10.1093/emboj/19.23.6612

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. doi:10.1111/mmi.12892