2016-06-01

Analyses of mRNA Cleavage by RelE and the Role of tRNA Methyltransferase TrmD Using Bacterial Ribosome Profiling

Jae Yeon Hwang
Brigham Young University

Analyses of mRNA Cleavage by RelE and the Role of tRNA Methyltransferase TrmD Using Bacterial Ribosome Profiling

Jae Yeon Hwang

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Allen R. Buskirk, Chair
Barry M. Willardson
William R. McCleary
Joel S. Griffitts
Joshua L. Andersen

Department of Chemistry and Biochemistry

Brigham Young University

June 2016

ABSTRACT

Analyses of mRNA Cleavage by RelE and the Role of tRNA Methyltransferase TrmD Using Bacterial Ribosome Profiling

Jae Yeon Hwang
Department of Chemistry and Biochemistry, BYU
Doctor of Philosophy

Protein synthesis is a fundamental and ultimate process in living cells. Cells possess sophisticated machineries and continuously carry out complex processes. Monitoring protein synthesis in living cells not only inform us about the mechanism of translation but also deepen our insights about all aspects of life. Understanding the structure and mechanism of the ribosome and its associated factors helped us enlarge our knowledge on protein synthesis.

Recently, with the dramatic advances of high-throughput sequencing and bioinformatics, a new technique called ribosome profiling emerged. By retrieving mRNA fragments protected by translating ribosomes, ribosome profiling reveals global ribosome occupancy along mRNAs in living cells, which can inform us with the identity and quantity of proteins being made. Easily adapted to other organisms, ribosome profiling technique is expanding its application in revealing various cellular activities as well as the knowledge on protein synthesis.

Here, we report the mechanism of translating mRNA cleavage by endoribonuclease RelE *in vivo*. RelE is an endoribonuclease that is induced during nutrient deficiency stress and specifically cleaves translating mRNAs upon binding to the ribosomal A site. Overexpression of RelE in living cells causes growth arrest by inhibiting global translation. We monitored RelE activity *in vivo* upon overexpression using ribosome profiling. The data show that RelE actively cuts translating mRNAs whenever the ribosomal A site is accessible, resulting in truncated mRNAs. RelE causes the ribosome complexes to accumulate near the 5' end of genes as the process of ribosome rescue, translation, and cleavage by RelE repeats. RelE cleavage specific sub-codon level ribosome profiling data also represent reading frame in *Escherichia coli* and sequence specificity of RelE cleavage *in vivo*.

We report another ribosome profiling study on a methyltransferase TrmD in *E. coli.* TrmD is known to methylate G37 (the residue at 3' side of anticodon) of some tRNAs and be responsible for codon-anticodon interaction. We constructed a TrmD depletion *E. coli* strain, whose deletion results in lethality of cells. Resulting depletion of m^1G37 in the strain leads to growth arrest. Lack of m^1G37 of some tRNAs whose codons start with C showed frequent frameshift when translating the gene message *in vitro*. By using ribosome profiling, we successfully observed significant difference on translation process when codons interact with anticodons of tRNAs lacking m^1G37. The data reveal slow translation rate or pauses on the tRNAs when missing the appropriate methylation, which corresponds to the previous biochemical data *in vitro*.

Keywords: RelE, endoribonuclease, TrmD, methyltransferase, ribosome profiling, m^1G37

ACKNOWLEDGEMENTS

I had opportunities to experience three different laboratories across the U.S, which is not common for a graduate student. After working two years at BYU, I worked at Scott Strobel's laboratory at Yale for about 10 months, then back to BYU for a year, and again to Rachel Green's laboratory at Johns Hopkins for two years so far. Moving from one place to another with my family slowed down the progress of my research but, with the help from many people I could continue studying.

I would like to first thank my former lab members at BYU. I still remember when I talked to Chris Woolstenhulme for the first time while searching labs to join. He helped me to have an idea what our lab was doing for research and helped me to learn many lab skills later on. Mickey Miller was always nice and helped me to understand better what I was doing in the lab. He was always patient in answering many questions I had.

I express thank you to the lab members at Yale University, who helped me to understand kinetic assays for RelE activity and shared their findings from experiments. I am grateful for their encouragement for me while I was struggling with continuous failing on my experiments. They helped me to realized that I was not the only one who was discouraged with repeated failures by sharing their own experiences. I also appreciate their talents and creativities in approaching problems and carrying experiments.

I would like to thank current lab members at Johns Hopkins for the discussions and for sharing their talents. With their mature citizenship and efficient setup at the lab, I had an excellent environment to focus on the research. I would like to also express thank you to Isao Masuda for the collaboration on TrmD study and for his friendship during the work at Ya-Ming

Hou's lab at Thomas Jefferson University. I also thank Ya-Ming Hou for sharing her knowledge on tRNAs and methylations.

I am grateful for the generosity of Scott Strobel and Rachel Green. They both allowed me to work at their labs. I am honored to be part of their labs, being inspired by their insights and superb attitudes towards science in proximity. Their enthusiasm and knowledge on the field stimulate me to have passion on science as well. Rachel always talks about science without ceasing and her energy and confidence remind me of what I want be.

I also thank for my committee members at BYU for their guidance. I especially appreciate Barry Willardson for his attitude towards science and for teaching ethics in science through a class. He told a group of students in a class, including me, that science is built on the foundations of previous researchers so that we are obligated to share what we obtain. That lesson was one of the most valuable lessons for me during graduate school years. I thank Joel Griffitts for demonstrating his critical thinking in solving problems. I am also thankful for BYU and funding sources that supported me to study during the program. I have been impressed by many great people at BYU for their teachings and care for the students.

Most of all, I deeply thank my advisor and mentor, Allen Buskirk, who provided me with many opportunities to learn. He taught me that science is a really fun thing that I should be excited about. I like his sense of humor and the way he put things together in science. His talent and knowledge on science make him special and that, somewhat, separates me from him, but I always aimed to overcome the gap between mine and his and to catch up the level of his proficiency, which still looks a long way. He has guided and helped me on the course with great patience. I feel like I am just ready to start real science now.

Lastly, I would like to thank my family and friends for their support. I would like to thank Kyu Tae Han and Keum Hee Song, who helped me broaden my views and shape my mind. I extend special thank you to my eldest brother, Jang Yeon for his support since I was young. Without his help, it would not be possible to come to the U.S. and continue to study what I always wanted to. I am so grateful for my parents for their constant love and trust towards me. I love my family, my beautiful wife, Jee Yeon Lee, my daughters, Yiwon and Jiwon, who always give me motivation and strength to continue. There were hard times during the course and they were always with me to encourage me. Without the love and trust from my wife, I would not be able to accomplish this and will not be able to go further.

TABLE OF CONTENTS

List of Tables viii

List of Figures ix

1 Introduction 1

1.1 Bacterial Translation 1

1.1.1 Initiation 2

1.1.2 Elongation Cycle 3

1.1.3 Termination 4

1.1.4 Ribosome Recycling 5

1.1.5 Ribosome Rescue 5

1.2 Ribosome Profiling 8

1.2.1 The Ribosome Profiling Procedure 9

1.2.2 Analysis of Ribosome Profiling Data 10

1.2.3 Power of Ribosome Profiling to Study Protein Synthesis 11

1.2.4 Bacterial Ribosome Profiling: Promises and Obstacles 14

1.3 The Endoribonuclease RelE 17

1.4 tRNA Methylations and TrmD 19

2 mRNA Cleavage by RelE 20

2.1 Introduction 20

2.2 Results and Discussion 23

2.2.1 Ribosomes are Enriched at the 5'-end of Genes in Cells Overexpressing RelE 23

2.2.2 RelE Predominantly Cleaves mRNA after the Second Nucleotide in Empty A sites 28

2.2.3 *in vitro* Digestion with RelE Reveals the Reading Frame 31

2.2.4 RelE Prefers to Cleave after C and before G 34

2.2.5 The Specificity of RelE Interferes with Analyses of Ribosome Pausing 37

2.2.6 Refined RelE-Derived Ribosome Density Better Reflects Reading Frame 39

2.2.7 Concluding Thoughts 41

2.3 Materials and Methods 42
2.3.1 Bacterial Strains and Growth Conditions 42
2.3.2 Preparation of Ribo-seq and RNAseq Libraries 42
2.3.3 Data Analysis 43
3 The Role of tRNA Methyltransferase TrmD 44
3.1 Introduction 44
3.1.1 tRNA Synthesis and Post-Transcriptional Modifications 45
3.2 The TrmD Methyltransferase 47
3.3 Results 49
3.3.1 Achieving Regulation of TrmD, an Essential Protein 49
3.3.2 Depletion of m^1G37 of tRNA Causes Pauses on CCG, CCA, and CGG Codons 53
3.4 Materials and Methods 57
3.4.1 Strains and Construction 57
3.4.2 Cell Growth and Preparation of Ribo-seq Libraries 58
3.4.3 Data Analysis 58
References 59

LIST OF TABLES

Table 1 Codons and their anti-codons with modifications 56

LIST OF FIGURES

Figure 1 Average ribosome occupancy 24
Figure 2 Ribosome enrichment at 5'-end genes of operon 26
Figure 3 Average ribosome density at start- and stop-codon 28
Figure 4 Ribo-seq data at nascent peptide-mediated stalling sites 30
Figure 5 Average ribosome occupancy for a wild-type sample 32
Figure 6 Ribo-seq and RNAseq density at each position 33
Figure 7 Sequence bias at the 3'-end of ribosome footprints 35
Figure 8 Sequence bias at the 3'-end of RNA fragments 36
Figure 9 The ribosome occupancy on codons 37
Figure 10 Correction for better reading frame 39
Figure 11 Programmed frameshifting at the *prfB* gene 40
Figure 12 1-methyl-Guanosine 47
Figure 13 TrmD depletion strain construction 50
Figure 14 TrmD depletion 51
Figure 15 Growth measurements upon TrmD depletion 52
Figure 16 Effect of TrmD depletion on tRNA 53
Figure 17 Ribosome occupancy on codons in control strain 54
Figure 18 Ribosome occupancy on codons in degron strain 55

1 INTRODUCTION

1.1 Bacterial Translation

Ribosomes read the genetic information in mRNA and facilitate peptide-bond formation to make proteins from single amino acids. *Escherichia coli* ribosomes are composed of two subunits termed 50S and 30S because of their sedimentation coefficients. The larger 50S subunit is itself composed of two ribosomal RNAs (23S and 5S) and 34 ribosomal proteins; the active site that links amino acids together is found within the large subunit. The smaller 30S subunit contains the 16S ribosomal RNA (rRNA) and 21 proteins and plays the dominant role in decoding the genetic information in mRNA. The two subunits combine to make 70S ribosome complexes that are active in translation.

Peptide synthesis occurs in the ribosome in a very sophisticated way with the help of key adaptor molecules known as transfer RNAs (tRNAs). tRNAs are first charged with activated amino acids to form aminoacyl-tRNAs (aa-tRNAs). There are 20 different aminoacyl-tRNA synthetases (aaRSs) in *E. coli*[1]. Each aaRS activates the correct amino acid with ATP and then links the amino acid to its cognate tRNA by an ester bond between the carboxyl group of the amino acid and the 3'-terminal adenosine of the tRNA. The aa-tRNAs then enter the ribosome, and when bound to their cognate codons, deliver the appropriate amino acid for the ribosome to incorporate into the growing polypeptide chain. The ribosome has three sites for tRNA binding: the aminoacyl-tRNA or A site, where the correct aa-tRNA is selected, the peptidyl-tRNA or P

site, where the growing nascent chain is attached, and the exit or E site, through which uncharged tRNAs pass as they are released from the ribosome. The protein synthesis process is complex and continuous but it can be conveniently broken down into four discrete steps: initiation, elongation, termination and ribosome recycling[2].

1.1.1 Initiation

At least in most cases, free 50S and 30S ribosomal subunits assemble into 70S complexes at start codons to initiate protein synthesis. Initiation factor 1 (IF1) binds to the A site of the 30S subunit and promotes dissociation of the 30S and 50S subunits. Initiation factor 3 (IF3) binds to the E site of the 30S subunit and further prevents premature association with the 50S subunit[3]. The 30S subunit bound with IFs interacts with the purine-rich region known as the Shine-Dalgarno (SD) sequence on mRNA that is complementary to the 3'-end of the 16S rRNA[4]. The pyrimidine-rich region of 16S rRNA, or the anti-Shine-Dalgarno (aSD) sequence, binds to the SD sequence roughly 7 nt upstream from the start codon[5]. This interaction positions the start codon of the mRNA at the P site of the 30S subunit. In bacteria, start codons are predominantly AUG although GUG is also sometimes used. The fact that ribosomes can be recruited to many internal sites within a polycistronic mRNA differentiates bacterial translation from eukaryotic translation, where ribosomes are recruited by proteins that bind the 5'-cap and then scan and initiate at the first start codon in mRNAs that are, as a rule, monocistronic[6].

Guanosine 5'-triphosphate (GTP)-bound initiation factor 2 (IF2) recognizes the initiator tRNA, *N*-formyl-methionyl-tRNA^{fMet} (fMet-tRNA^{fMet})[3, 7]. IF2 binds in the A site of the 30S subunit and places the initiator tRNA in the P site of the 30S subunit. Binding of initiator tRNA to the start codon at the P site causes conformational changes of the 30S subunit that lead to release of IF3. With the release of IF3, the 50S subunit can bind to the 30S subunit, inducing

GTP hydrolysis by IF2. At that point, IF1 and GDP-bound IF2 are released from the 30S subunit, leaving behind an active 70S initiation complex. With initiator tRNA at its P site and an empty A site, the complex is ready for elongation.

1.1.2 Elongation Cycle

At the start of the elongation cycle, the 70S initiation complex with initiator tRNA bound to its P site accepts the next aa-tRNA into its empty A site[8]. aa-tRNAs are delivered to the ribosome by the abundant GTPase, elongation factor-Tu (EF-Tu)[9]. Upon GTP binding to EF-Tu, EF-Tu arranges its three domains in a compact structure allowing them to bind aa-tRNA[10]. When the ternary complex (EF-Tu, GTP, and aa-tRNA) binds to the A site, it is rejected if the anticodon of the tRNA does not match the codon in the A site[8, 11]. This process repeats until the cognate tRNA comes in[8, 11-12]. Upon binding of the cognate tRNA in the A site, the correct codon-anticodon base pairing causes conformational changes in the ribosome[8]. rRNA residues surrounding the base pairs stabilize tRNA binding to the A site[8, 12].

Conformational changes in the A site upon binding of the correct aa-tRNA result in a distortion of the tRNA, triggering GTP hydrolysis by EF-Tu[10-11]. As the GTP of EF-Tu is hydrolyzed and the γ-phosphate is released, the structure of the EF-Tu is rearranged and releases the tRNA, and EF-Tu dissociates from the A site[10, 13].

GTP hydrolysis occurs faster with cognate tRNAs than with near- or non-cognate tRNAs; this kinetic difference is one basis for selectivity for the correct tRNA[14]. Following GTP hydrolysis, the aminoacyl end of the A-site tRNA swings into the peptidyl-transferase site of the 50S subunit, a movement known as accommodation[15]. Cognate tRNAs are accommodated more quickly than near- or non-cognate tRNAs, as second kinetic selection or proofreading step[14-15].

RNA nucleotides in the peptidyl-transferase center bring the two amino acids together to form a peptide bond[15]. There are two RNA elements in the 50S subunit called the A loop and P loop that have highly conserved sequences and interact with the 3'-ends of aa-tRNAs positioned at A site and P site[15]. Once the C-terminus of growing peptide and the amino group of the A-site bound aa-tRNA are positioned correctly, peptide bond formation occurs intrinsically by a nucleophilic displacement, perhaps involving side chains on the tRNAs themselves but not the rRNA nucleotides in the active site[2, 15].

For translation to continue, the ribosome must move to the next codon on the mRNA. Following peptide-bond formation, the deacylated tRNA is at the 'P/E' hybrid state and the peptidyl-tRNA with two amino acids is at the 'A/P' hybrid state[15-16]. The tRNAs need to be resolved to a new classical state with the deacylated tRNA completely in the E site and the peptidyl-tRNA positioned at the P site[15]. This process is called translocation and is catalyzed by elongation factor G (EF-G), a ribosome-activated GTPase. Hydrolysis of GTP by EF-G helps the 30S subunit move to the right position for the next codon at A site, maintaining reading frame by moving exactly three nucleotides on mRNA[17]. The uncharged tRNA at the E site is released from the ribosome and another aa-tRNA enters the A site[2, 15]. This cycle repeats until a stop codon (UAG, UAA, or UGA) appears at the A site. Translocation is thought to be relatively fast compared to decoding, and bacterial peptides are formed in the ribosome at a rate of about 20 amino acid residues per second[2, 15].

1.1.3 Termination

When the ribosome reaches a stop codon, protein synthesis is terminated. Stop codons are recognized by two release factors, RF1 and RF2[18]. By mimicking tRNA structurally, these factors bind in the A site and interact directly with mRNA[19]. RF1 recognizes the UAA and UAG

stop codons and RF2 recognizes the UAA and UGA stop codons[20]. Release factors hydrolyze the ester bond on peptidyl-tRNA and release the newly synthesized peptide chain from the ribosome, leaving uncharged tRNA in the P site[2, 20]. Both RF1 and RF2 contain a highly conserved tripeptide motif (GGQ motif) that is responsible for the catalysis of peptide release from the peptidyl-tRNA at P site[20-21]. Following hydrolysis of peptidyl tRNA, RF3, a GTPase found in some but not all bacteria, catalyzes the release of RF1 or RF2 at the end of the termination process. GTP hydrolysis then promotes the dissociation of RF3[20-22].

1.1.4 Ribosome Recycling

After termination, ribosomes need to be dissociated for efficient initiation at another start codon to begin. Ribosome recycling factor (RRF) and EF-G dissociate the ribosome into subunits, releasing mRNA and tRNAs from the ribosome upon GTP hydrolysis[2]. The structure of RRF mimics tRNA in the A site[2, 23]. When the ribosome dissociates, IF1 replaces the deacylated tRNA and the mRNA is released. All translational components are now free for another round of translation[23].

1.1.5 Ribosome Rescue

In bacteria, the translation and transcription of genes are coupled. Since genomic DNA is not separated from the cytoplasm by a nuclear membrane like it is in eukaryotic cells, the whole process of gene expression happens simultaneously and is faster than it is in eukaryotic cells. As soon as an mRNA begins to be synthesized, translation can begin. This can lead to problems when ribosomes translate messages that are defective as when RNA polymerase prematurely terminates prior to reaching the stop codon in the message. Other defective RNAs come from

mutations, chemical damage, and the endonucleolytic cleavage that is part of normal mRNA decay pathways[24].

Truncated mRNAs are deleterious to the cell for at least two reasons: they cause ribosomes to stall at the 3'-end of the message and produce incomplete or premature proteins which might be harmful to the cell. When ribosomes reach the 3'-end of a truncated mRNA, RFs cannot be recruited, since there is no stop codon, and the ribosome-mRNA complex is stably maintained[25]. In the cell, stalled ribosomes can affect the global protein synthesis if not rescued.

Bacterial cells have three mechanisms that can resolve stalled ribosomes. One involves a small, stable RNA known as tmRNA and the other two involve proteins that are alternative rescue factors, ArfA and ArfB[25]. tmRNA rescues the stalled ribosome by acting both as a tRNA and an mRNA[25-26]. tmRNA-mediated ribosome rescue not only dissociates the stalled ribosome for recycling but also tags the incomplete nascent peptide for degradation. ArfA is a small protein that can bind to the A site and recruit RF2, hydrolyzing the nascent peptide and leading to recycling[27]. ArfB has a GGQ domain in its structure as RFs do and facilitates the termination of protein synthesis. Among these three mechanisms, tmRNA is the best characterized[28].

The sophisticated structure of tmRNA makes it act as both tRNA and mRNA. In *E. coli*, tmRNA is 363 nt long. It contains an tRNA-like domain (TLD) and an mRNA-like domain with its own small open reading frame (ORF)[29]. The TLD contains 3'-terminal CCA residues like tRNAs do and has a G3-U357 wobble base pair, which alanyl-tRNA synthetase recognizes and charges with alanine at its 3'-end. With several helices and pseudoknots in its structure, tmRNA is relatively stable against degradation by nucleases. Its partner protein, SmpB, further stabilizes tmRNA[29-30]. The structure of the TLD allows EF-Tu to bind Ala-tmRNA just like other tRNAs do. The GTP-bound EF-Tu-tmRNA-SmpB complex enters the A site of stalled ribosomes and

undergoes peptidyl transfer following GTP hydrolysis by EF-Tu[30]. After the growing peptide is attached to tmRNA, the mRNA-like domain of tmRNA serves as a transcript for the ribosome to continue translation. The ORF encodes the ANDENYALAA peptide and a stop codon at the end[31]. Because the ribosome switches templates, ribosome rescue by tmRNA is referred to as *trans*-translation[32]. The 11 amino acids added by tmRNA target the aborted nascent peptide for degradation. The ClpXP protease system recognizes the hydrophobic C-terminal residues of the tagged peptide and degrades it[31].

While tmRNA and SmpB are found in all bacterial genomes, the ArfA backup system found in some bacteria also facilitate ribosome rescue. It is proposed that ArfA, previously known as YhdL, binds to the ribosomal A site and recruits RF2. Then, RF2 hydrolyzes the peptidyl-tRNA and releases translation complex. It is not known, however, how ArfA recognizes the 3'-end of non-stop mRNA and how it recruits RF2. The deletion of tmRNA and ArfA together are lethal whereas a single deletion of either is viable, which suggests a mutual role of the two mechanisms in ribosome rescue[26].

ArfB, previously known as YaeJ, also rescues stalled ribosomes. In *E. coli* ArfB is a small basic protein with 140 amino acids. As a homolog of RFs, ArfB contains the conserved GGQ motif for hydrolysis of peptidyl-tRNA[28]. Instead of having a specific sequence for interacting with stop codons, ArfB has a positively charged C-terminal tail that interacts with the mRNA channel, selectively binding ribosomes with little or no mRNA downstream of the A-site codon. This is reminiscent of the C-terminal tail of the SmpB protein that provides selectivity for the tmRNA rescue system[28].

1.2 Ribosome Profiling

Insight into protein synthesis can shed light on many aspects of biology; in particular, it is a powerful way to monitor gene expression levels *in vivo*. Previously, researchers have achieved the goal by monitoring transcriptional changes with microarrays or RNAseq. Another approach, proteomics based on mass spectrometry (MS), reveals the identity and quantity of proteins in the cell. However, it is difficult to accurately quantify proteins with low abundance or instability. In addition, MS focuses on the steady-state levels of proteins in the cell rather than the rate of protein synthesis. While steady-state levels are very informative in determining the level of protein activity, they are less helpful in understanding the mechanism of protein synthesis or the translational control of gene expression[33].

Polysome profiling is another method used to monitor protein synthesis *in vivo*[34]. This technique is based on the fact that the number of translating ribosomes bound to mRNAs can tell us the translational status of that transcript. Ribosome-bound mRNAs are collected by ultracentrifugation of cell lysates over sucrose gradients. The identity of heavily translated mRNAs is determined and compared to the background of total mRNA using microarrays or RNAseq[34]. Polysome profiling allowed researchers to move from monitoring transcription levels to monitoring translation levels[34]. However, the approach has some difficulties in separating messages depending on the number of ribosome bound to them. Sucrose gradient fraction collection does not resolve clearly complexes heavier than disomes. The many fractions per sample (based on the number of ribosomes bound to mRNAs) makes it a lot of work to do the following analyses. More importantly, the data generated from polysome profiling cannot give the exact position of ribosomes on mRNAs; rather, they only give some idea of ribosome abundance on certain mRNAs.

To get better information on translation *in vivo*, Ingolia and Weissman developed a powerful new technique called ribosome profiling or Ribo-seq[35]. Simply stated, ribosome profiling improved polysome profiling by isolating short fragments of mRNA corresponding to ribosome footprints, revealing the exact positions of ribosomes on mRNAs. Ribosome profiling provides a genome-wide snapshot of ribosome occupancy on mRNA templates, telling us which proteins are being made and to what extent[35].

The development of ribosome profiling was made possible by advances in sequencing technology. Ribosome profiling involves deep sequencing millions of ribosome-protected short mRNA fragments. With the completion of the human genome project, significant improvements were made in high-throughput DNA sequencing, making the sequencing of hundreds of millions of short reads economically viable. Methods for preparing cDNA libraries have been well developed, including RNA ligation methods and techniques for multiplexing a number of libraries for more efficient and cost effective sample preparation. Another improvement came from easily accessible databases from bioinformatics. Annotated whole genomes provide the standard against which sequencing reads can be mapped or aligned. In addition, bioinformatics has provided many algorithms for processing and analysis of the data generated from high-throughput sequencing. Together, these technologies and methods make possible the ribosome profiling technique that is increasing our understanding of gene regulation and particularly the protein synthesis process.

1.2.1 The Ribosome Profiling Procedure

To collect ribosome-protected mRNA fragments for deep sequencing, grown cells are effectively captured by flash freezing using liquid nitrogen. For lysis, cells are pulverized mechanically in liquid nitrogen in the presence of lysis buffer containing drugs that arrest

ribosomes on mRNA complexes. The pulverized lysate is thawed and centrifuged to pellet cell debris. The lysate is then subject to digestion using a nuclease to degrade naked mRNA, leaving only ribosome-protected mRNA fragments. Ribosome complexes are isolated on sucrose gradients by ultracentrifugation; digestion with the nuclease has reduced the polysomes to single monosomes. Ribosome-protected mRNA fragments are released from the monosomes by treatment with phenol and chloroform followed by RNA precipitation. At this point, the purified RNAs contains rRNA and tRNA fragments as well as mRNA fragments.

To purify ribosome footprints away from other RNA fragments, the total RNA is run on a denaturing acrylamide gel for size selection. RNA corresponding to the size that best represents the footprint of the ribosome are collected. Generally, we isolate mRNA fragments on the gel that are between 10 to 40 nt in length. The strategy to sequence these mRNA fragments is to make a cDNA library. First, a defined RNA linker that contains a primer binding sequence is attached at the 3'-end of each mRNA fragment by ligation with T4 RNA ligase. Linker-ligated fragments are then gel-purified. rRNA fragments are selectively subtracted by hybridization with oligonucleotides complementary to rRNA. Then, the purified mRNA fragments are reverse transcribed to yield cDNAs and these single-stranded cDNAs are circularized by ligation, amplified by a few cycles of PCR, and submitted for 50 bp, single end, Illumina sequencing.

1.2.2 Analysis of Ribosome Profiling Data

Data analysis involves processing the raw sequencing data into genome-wide maps of ribosome occupancy and then further analysis for specific needs. High-throughput sequencing is performed on several samples per lane in a single experiment (a strategy known as 'multiplexing'). Each library is synthesized with a characteristic sequence (usually 6 nt) termed a 'barcode' that distinguishes that library from the others during data analysis. In the first step of

data processing, raw sequences are de-multiplexed by sorting them out according to their barcode sequences. In addition, reads are filtered based on the quality of data as quantified in Phred quality scores, which is a measure of the quality of the identification of the nucleobases, and ambiguous or suspicious reads are excluded. After filtering, the reads are trimmed to remove the constant 3'-linker sequence, leaving only the ribosome footprints. Next, any reads that map to tRNA and rRNA genes are discarded before mapping reads to annotated genomic DNA. Final ribosome occupancy data are represented as densities along the genome sequence; these maps of ribosome occupancy are used for additional analysis using our own computing scripts, which are written mainly using a computer programming language Python.

1.2.3 Power of Ribosome Profiling to Study Protein Synthesis

Although ribosome profiling has mainly been used to monitor gene expression under two or more conditions, using ribosome occupancy to report on protein levels, the method has tremendous potential to shed light on the mechanism of protein synthesis *in vivo*. There have been many insights obtained using ribosome profiling since the method was first introduced by Ingolia and Weissman in 2009. In that first report, several millions of sequencing reads of ribosome-protected mRNAs were obtained in yeast *Saccharomyces cerevisiae*. 28-nt long reads yielded the best information about the position of the ribosome, suggesting that the nuclease digested completely to the 5'- and 3'-boundaries of the ribosome[35]. Sequencing reads that aligned to ORFs revealed ribosome positions on mRNAs with nearly single-nucleotide precision and showed a strong 3-nt periodicity, providing a means of monitoring the ribosome's reading frame[35]. Genome-wide measurements of ribosome occupancy also generated estimates of the levels of protein synthesis, which were found to be better predictors of protein abundance by MS than were measurements of mRNA levels.

In a later paper by Ingolia and Weissman, the ribosome profiling technique was adapted to mammalian cells[36]. Using the drug harringtonine, which specifically inhibits elongating ribosomes near the start codon (but not at later sites in the ORF), new translation was inhibited in cells[36]. Following the addition of harringtonine, ribosomes were allowed to run off for various time periods to generate a series of snapshots and reveal a moving picture of translation *in vivo*[36]. Ribosome footprints successfully revealed a progressive depletion of ribosomes from the 5'- to the 3'-end along mRNAs as the run-off time increased after harringtonine treatment[36]. The data showed that ribosomes translate at a rate of 5.6 amino acids per second, consistent with previous biochemical observations.

In addition to their work on measuring rates of translation with harringtonine, Ingolia and Weissman also detected sites where ribosome occupancy was strongly enriched due to pausing during elongation. Programmed nascent chain-mediated pauses were observed on the Sec61b and Xbp1 genes[36]. More generally, the PPE/D motif was shown to be associated with internal pauses, consistent with the slow translation of polyproline stretches. Finally, they also found that ribosome footprints revealed many alternate ORFs, upstream of canonical ORFs on the same mRNA, many non-canonical translational start sites (especially at the CUG codon), and evidence for translation of putative long, non-coding RNAs (lncRNA)[36].

High-precision ribosome footprints obtained by ribosome profiling make it possible to detect programmed frameshift sites. Coupled with a computational method for detecting transitions between reading frames, ribosome profiling data obtained from human cells revealed where the same genomic segment is translated in more than one reading frame[37]. The 3-nt periodicity of ribosome occupancy shifts if the ribosome has moved into a different reading frame during translation. Ribosome profiling data is a powerful tool for searching for alternative

coding regions in endogenous genes and perhaps in the future in viruses and other mobile genetic elements rich in such phenomena.

Ribosome profiling in *Drosophila melanogaster* expanded our understanding of ribosome readthrough, in which the ribosome continues to translate past a stop codon. Previous studies have shown that this can occur in a regulated fashion. However, it was very difficult to detect readthrough; only a small fraction of ribosomes keeps translating, extending proteins past their predicted C-termini. The presence of ribosome occupancy after stop codons successfully showed that readthrough is far more than previously predicted in *D. melanogaster*[38].

Ribosome profiling helped to clarify the enzymatic activity of Dom34, a yeast protein essential for releasing stalled ribosomes[39]. Dom34 is a homolog of eukaryotic release factor 1 (eRF1) though it lacks a canonical GGQ motif required for peptide release as well as the codon recognition motif required for interacting with stop codons. Given these differences and its role in rescuing stalled ribosomes, Guydosh and Green used ribosome profiling to identify specific targets of Dom34, comparing wild type and *dom34* knock-out strains of yeast[39]. Stalled ribosomes are stabilized by the lack of Dom34; there are higher levels of ribosome occupancy on certain mRNAs at stalling sites in the *dom34* knock-out strain. The fragments protected by stalled ribosomes tend to be shorter (15-18 nt) than normal reads. This is because Dom34 preferentially rescues stalled ribosome at the end of truncated mRNAs, meaning that the mRNA bound to the ribosome is roughly 10 nt shorter at the 3'-end. Furthermore, they found that Dom34 rescues ribosomes that are found in 3' untranslated regions (UTRs) of mRNAs[39]. The codon-independent rescue mechanism of Dom34 nicely supports the previously known biochemistry of Dom34, which lacks conserved codon recognition motif[39].

Ribosome profiling has become a powerful way to understand complex cellular processes in living cells including protein synthesis. By monitoring protein synthesis in real time, ribosome profiling has clarified previous biochemical observations and given ample insights on many aspects. Since 2009, the technique is being more broadly adapted to study various organisms and the data analysis methods are improving immensely.

1.2.4 Bacterial Ribosome Profiling: Promises and Obstacles

Ribosome profiling has also employed to study bacterial translation. The relatively small genome of bacteria and less complex gene regulation would suggest that it should be equally powerful in bacteria as in yeast, where it was first developed. However, there have been several hurdles to its application in bacteria. The first is that RNase I, the nuclease used to digest naked mRNA in the yeast protocol, is inhibited by *E. coli* ribosomes, preventing its use. RNase I exhibits very little sequence specificity in cleavage and degrades RNA cleanly to the boundaries of the ribosome, meaning that the 28 nt fragments observed in yeast yield single-nucleotide resolution and information about the reading frame of the ribosome[40]. In contrast, all published bacterial ribosome profiling studies use micrococcal nuclease (MNase), an enzyme with pronounced sequence specificity[40-41]. The distribution of fragments generated by MNase is very broad, with most reads roughly 15 – 35 nt in length, and no 3 nt periodicity or reading frame is evident. Rather than assigning ribosome occupancy to the 5'-end of reads as was done originally in yeast, Li and Weissman used a center-assignment strategy in which partial occupancy was distributed over the length of the entire read, further blurring the position of the ribosome on the message[41]. The use of MNase in bacterial ribosome profiling has interfered with an accurate determination of the ribosomes position, interfering with analyses of ribosome pausing. In spite

of these limitations, there have been many advances in understanding bacterial translation using ribosome profiling, and there have been on-going improvements on the technique itself.

The first ribosome profiling paper using *E. coli* was reported in 2011 by the Weissman and Bukau groups in a study of the interaction of the nascent peptide chaperone trigger factor (TF)[40]. The authors selectively purified ribosomes bound to TF by cross-linking the nascent peptide to TF to stabilize transient interactions followed by immunoprecipitation with anti-TF antibodies. Ribosome footprints from TF-bound and non-TF-bound ribosomes were then obtained following the regular ribosome profiling procedure[40]. It turns out that trigger factor binds about 120 codons downstream from the translational start site, contradicting previous models that TF interacts as soon as the peptide comes out from the exit tunnel of the 50S subunit *in vitro*. Selective ribosome profiling in *E. coli* by cross-linking the chaperone TF with the newly synthesized polypeptide revealed valuable information of the kinetics of protein synthesis *in vivo*.

Another early ribosome profiling study in *E. coli* claimed that the main source of pausing during elongation in bacteria is Shine-Dalgarno-like sequences within open reading frames[41]. SD motifs accounted for 70% of strong pauses in their dataset—sites where ribosome occupancy were enriched more than 10-fold over the gene average. The authors concluded that elongation is retarded by transient base pairing between SD motifs within ORFs and the aSD sequence in 16S rRNA[41]. However, Mohammad *et al.* contradicted the finding with improved footprint resolution[42]. The precision of the ribosome position is significantly improved by assigning ribosome occupancy using the 3'-end of sequencing reads (3'-assignment) rather than distributing occupancy over the whole read (center-assignment)[42]. This simple adjustment seems to better position the ribosome occupancy along its mRNA transcript on all of known *E. coli*

footprint data so far, lining up the peaks seen at start and stop codons as well as programmed pauses like the one observed in the SecM protein. It seems that MNase does a better job in cleaving mRNAs up to the boundary of the ribosome at the 3'-end site of the bound mRNA, resulting in more tight cleavage at the 3' end but relaxed cleavage at the 5' end of ribosome-protected mRNAs[42].

With the higher resolution obtained by 3'-end assignment it became clear that the signal observed by Li and Weissman could be split into two separate signals due to two distinct phenomena: first, ribosomes pause with Gly codons in the E site, perhaps due to the nature of the inhibitor in the lysis buffer used to arrest translation. Second, the bona fide SD pauses arise from preferential enrichment of long RNA fragments from the population of ribosome footprints. It turns out that mRNAs that interact with the aSD in 16S rRNA are protected from digestion at the 5'-end and so are longer on average than reads lacking SD motifs. By failing to isolate the entire population of ribosome-protected mRNA fragments, Li and Weissman had artificially enriched for SD containing reads, observing pausing that is not necessarily there *in vivo*.

Ribosome profiling data can be used to quantify the absolute levels of protein synthesis genome-wide. In bacteria, many protein complexes are made of components synthesized from a single operon. In the operon encoding ATP synthase, for example, eight proteins are encoded on a single polycistronic RNA. The components are incorporated into the complex with various stoichiometries: 1:10:2:1:3:1:3:1[43]. Li and Weissman showed that bacteria optimize expression of such complexes by proportional expression, matching the level of expression to the amount required for complex formation[43]. This avoids wasteful synthesis of extra protein that would need to be degraded and raises very interesting questions about how these differences in translational efficiency are generated.

1.3 The Endoribonuclease RelE

In 1983, a mechanism that ensures stable maintenance of plasmids in *E. coli* was reported[44]. When a strain of *E. coli* cell carrying the F plasmid loses it while dividing, the cell without the plasmid is not viable[44]. Only the cells carrying the F plasmid are viable and continued to divide[44]. This is because a segment of the F plasmid plays an important role in maintaining the plasmid in the cell[44]. This segment, designated ccd (coupled cell division)[44], is composed of two parts: one (*ccdB*) that expresses an inhibitor of DNA gyrase[45] and another (*ccdA*) that expresses a protein that inactivates CcdB through forming a protein-protein interaction[44, 45b, 46]. When cells inherit the F plasmid, both CcdA and CcdB are expressed and form complexes that do no harm to the cell[44, 46]. However, if cells loose the plasmid after division, the unstable CcdA[46] gets degraded rapidly, leaving excess CcdB. Once freed, CcdB becomes active and arrests cell growth, leading to death upon further division[45b, 47]. With this mechanism, the plasmid ensures that it is replicated and passed on to all the viable daughter cells[45b, 47]. In this context, the mechanism is called post-segregational killing (PSK)[45b, 48] and the protein module is sometimes called as addiction module[49]. Later, many such protein pairs were found on the *E. coli* chromosome and named as toxin-antitoxin (TA) pairs, since one has cytotoxic activity and the other antagonizes its pair[48].

RelE is a toxin widely found in bacteria and archaea[48]. As an endoribonuclease, RelE is known to selectively cleave translating mRNAs by binding to the ribosomal A site and inducing global translational inhibition[50]. RelE activity is induced under nutritional depletion and is closely related to prokaryotic cell survival and pathogenicity including general stress responses and persister formation[50b]. RelE is one of the most studied toxins, but its activity and mechanism are still not well understood.

Pedersen *et al.* showed codon-specific mRNA cleavage by RelE in their *in vitro* kinetic assays[50a]. They observed that UAG and UAA among stop codons and CAG and UCG among sense codons were the most rapidly cleaved by RelE at the ribosomal A site, normally between the second and third bases[50a]. Neubauer *et al.* solved the structure of 70S ribosomes bound to RelE; their data partly explain why RelE cleaves mRNA in a codon specific manner and suggested a general acid-base catalytic mechanism[51]. They proposed that purines are favored for the second and third base of A-site codon for better stacking with residues from RelE and 16S ribosomal RNA in facilitating RelE activity[51].

Hurley *et al.*, however, observed different RelE cleavage patterns *in vivo*[52]. They found that RelE cut mRNAs at several sites frequently and efficiently within the first 100 codons of coding regions, and rarely cut mRNAs at sites near the 3' end[52]. They did not see any preference for CAG or UCG sense codons or any statistically significant sequence preferences. Moreover, they noticed that some cleavages occurred after the third base of A-site codon while most cleavages occurred between the second and third as previously known. They argued that their observance is more consistent with the fact that RelE causes rapid and comprehensive mRNA degradation and concomitant growth arrest[52].

Recently, we found in our work that RelE has some advantages for use as a nuclease in the ribosome profiling protocol. When added to cell lysates, purified RelE can cleave translating mRNAs at ribosomal A site and provides better information on ribosome position and reading frame than does MNase digestion. However, the sequence preference of RelE activity is problematic and requires further refinement and computational adjustment in analyzing data. This work will be described in Chapter 2.

1.4 tRNA Methylations and TrmD

TrmD is a bacterial-specific *S*-adenosyl-methionine (AdoMet)-dependent tRNA methyl transferase. Methylation of tRNA can affect the accuracy of protein synthesis and the maintenance of reading frame. The N1-methylguanine at position 37 (m^1G37)-tRNA product of TrmD improves the accuracy of protein synthesis on the ribosome. m^1G37-tRNA reduces +1 frameshift (+1FS) errors at slippery mRNA sequences and decoding errors of uridine-5-oxyacetic acid at position 34 (cmo^5U34), a wobble modification that frequently accompanies m^1G37 in natural tRNAs[53]. Loss of TrmD leads to accumulation of +1FS errors leading to pre-mature termination of protein synthesis at out of frame stop codons. *trmD* is an essential gene in *E. coli* and many bacteria and its biochemical mechanism is very different from eukaryotic tRNA methyltransferases, leading to interest in TrmD as a drug target[50b, 54]. We sought to achieve a cellular-level understanding of how inactivation of TrmD reduces synthesis of proteins to determine the reasons for its essentiality. In Chapter 3, we report ribosome profiling studies to observe defects in translation and identify genes whose expression is altered in TrmD deficient cells.

2 mRNA CLEAVAGE BY RelE

2.1 Introduction

Bacteria face enormous selective pressure from their physical and chemical environment and from competing micro-organisms. In response to this pressure, bacteria have evolved mechanisms to rapidly regulate gene expression in response to reduction in the levels of available nutrients, for example, or in response to antibiotics released by other organisms. Another strategy bacteria use to deal with these stresses is to maintain a small fraction of the population in a dormant state that survives environmental insults and then resumes growth when conditions improve[55]. This strategy plays an important role in antibiotic resistance because dormant cells (known as persisters) are not killed, even at high antibiotic concentrations. Shutting down protein synthesis is an essential step for both of these strategies[55]; for example, regulating gene expression in response to nutrient starvation via the classical stringent response and maintaining persister cells in a dormant state as a bet-hedging strategy[56].

The *relE* gene plays a role in stress-response pathways by blocking translation[50b]. The gene was originally discovered in genetic screens involving nutrient starvation and was named for its effect in the stringent response. During amino acid starvation, the concentration of uncharged tRNAs increases; as these tRNAs bind to the ribosome, they activate the synthesis of guanosine tetraphosphate (ppGpp) by the RelA protein[57]. Through its interactions with RNA polymerase, ppGpp leads to stringent inhibition of rRNA synthesis[57]. *E. coli* mutants in which

RelE is more highly expressed quickly shut down translation upon amino acid depletion, stopping the consumption of amino acids and rapidly lowering ppGpp to its pre-starvation levels, allowing rRNA synthesis to resume after a brief delay[58]. The synthesis of rRNA during amino acid starvation is characteristic of the 'relaxed' phenotype observed in mutants of genes in stringent response pathway, hence the name RelE[59].

In addition to its role in gene expression, RelE is also a member of the type II toxin-antitoxin family implicated in persister-cell formation[59a]. Overexpression of the RelE toxin causes a reversible inhibition of cell growth resembling the dormant state characteristic of persister cells[59a]. Growth resumes when RelE is neutralized by overexpression of its binding partner, the RelB anti-toxin. Under rich conditions, RelB is expressed at a slightly higher level, masking RelE activity, but under stress conditions, Lon protease degrades enough of the more labile RelB anti-toxin to induce RelE activity[60]. Even under rich conditions, stochastic activation of RelE and similar toxins is thought to be responsible for inducing a persister-like state in a small fraction of cells in culture[61].

Like at least ten members of the type II toxin-antitoxin family in *E. coli*, RelE exerts its effects by cleaving mRNA and inducing translational arrest[62]. RelE binds in the ribosomal A site and cleaves the RNA after the second nucleotide in the A-site codon[50a]. It is an unusual endonuclease in the sense that it does not cleave RNA outside of this context[50a]; the ribosome provides an environment essential to its catalytic activity. A catalytic mechanism has been proposed based on the X-ray crystal structure of RelE in a 70S ribosome complex and follow-up kinetic studies[51, 63]. Initially, RelE was reported to be highly specific for a select codons (CAG, UCG, and the stop codon UAG) based on its *in vitro* kinetics[50a], but more recent studies suggest that it has quite a broad specificity *in vivo* with only a modest sequence preference[52]. Analyses of

cleavage sites on a handful upon highly-expressed genes identified many cleavage sites with only a modest preference for cleavage before G[50a, 52]. Woychik and co-workers further made the puzzling observation that RelE cleaves primarily at the 5'-end of mRNAs, within about the first 100 codons[52]; the mechanism underlying this polarity is unknown.

Here we report a genome-wide characterization of protein synthesis in *E. coli* upon RelE overexpression. Unlike the elegant method of Woychik and co-workers, who adapted RNAseq to detect cleavage by the MazF toxin[64], the ribosome profiling method we employed reports on the position of ribosomes, allowing us to observe RelE's effect on translation. We find that ribosome density is strongly enriched at the 5'-end of genes and propose a model involving cycles of mRNA cleavage, rescue of stalled ribosomes, and initiation that explains the earlier observation of preferential RelE cleavage in the first 100 codons[52].

Furthermore, we find that RelE can be used to improve the resolution and power of ribosome profiling in bacteria. As originally developed in yeast, the protocol calls for digestion of naked mRNA with RNase I to generate ribosome footprints[35]. As RNase I is inhibited by *E. coli* ribosomes, bacterial ribosome profiling studies have used MNase instead. Unfortunately, MNase is sequence specific, creating strong sequence bias at both the 5'- and 3'-ends of the RNA fragments, enriching or suppressing specific RNA fragments depending on their nucleotide sequence. Additionally, MNase generates a broad distribution of the lengths of the ribosome footprints, unlike the tight 28 nt footprint observed in ribosome profiling studies in yeast[42]. This means that the position of the ribosome cannot be determined with sufficient precision to determine its reading frame[42]. Addition of purified RelE to cell lysates generates ribosome footprints that, for the first time, provide excellent information on the position and reading frame of the ribosome.

2.2 Results and Discussion

2.2.1 Ribosomes are Enriched at the 5'-end of Genes in Cells Overexpressing RelE

To study the effect of RelE on protein synthesis *in vivo*, we overexpressed RelE in wild-type *E. coli* MG1655 cells and analyzed ribosome occupancy genome-wide using ribosome profiling. RelE expression from an *araBAD* promoter was induced for one hour beginning in early log phase. Given the high level of RelE expression (one-quarter of ribosome footprints map to the RelE mRNA) and the long induction period, we observed very pronounced effects on protein synthesis, presumably representing an equilibrium state in which RelE activity has exerted its effect and cell growth has slowed or halted. Although strong overexpression no doubt exaggerates the effects that we observe in comparison to activation of endogenous RelE by physiological stimuli, the resulting data provide a clear picture of the scope and specificity of this endonuclease and its effects on gene expression.

The most striking effect of RelE overexpression is the dramatic enrichment of ribosome occupancy at the 5'-end of genes. In the model of the *manX* gene shown in Figure 1, for example, far more ribosome footprints map to the 5'-end of the gene than the 3'-end (red). In contrast, in the wild-type control (WT1, black), ribosome occupancy is distributed more or less similarly across the *manX* gene. These observations hold true genome wide. In a plot of average ribosome occupancy of about 1000 genes over 1000 nt long that were aligned at the start codon, ribosome coverage remains fairly constant in the wild-type control, whereas in the RelE1 data ribosomes are highly enriched in the first 400 nt and depleted after the first 400 nt. (Note that the vertical spread in the RelE1 data arises from the reading frame, as will be discussed below). In addition, a very strong peak of ribosome density is observed at the start codon in the RelE1 data. These data show that as a result of RelE activity, ribosomes accumulate at the 5'-end of genes

and are depleted at the 3'-end. This suggests that the number of ribosomes completing the synthesis of full-length proteins is strongly reduced, consistent with RelE's known function of inhibiting protein synthesis and arresting cell growth.

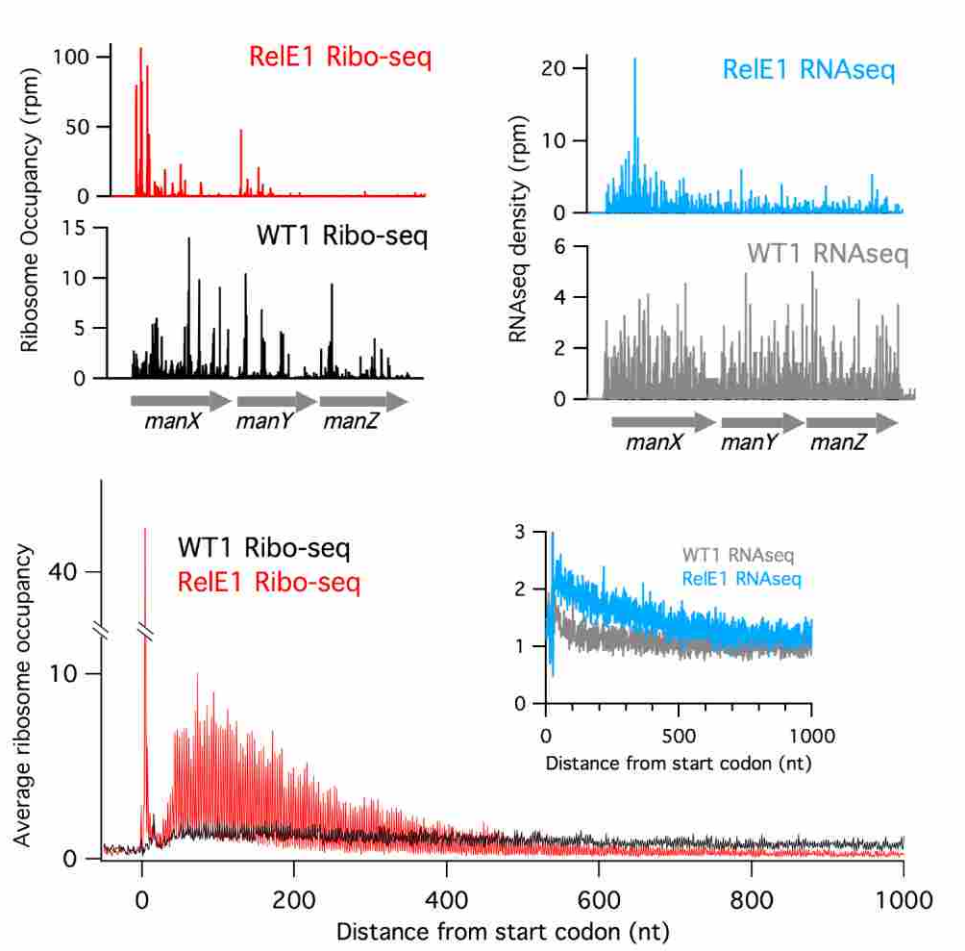

Figure 1 Average ribosome occupancy

Ribo-seq (left) and RNAseq (right) reads at the *manXYZ* operon (top). Average ribosome occupancy at about 1000 genes aligned at their start codons (bottom).

Given that RelE is an endonuclease that cleaves mRNA inside the ribosomal active site[50a], we also looked at steady-state levels of mRNA in RNAseq libraries prepared from the WT1 and RelE1 samples. Here again, we see enriched RNA density at the 5'-end of the *manX*

gene when RelE is overexpressed (Figure 1, top right, blue), whereas the coverage is relatively uniform in the wild-type control (grey). When the RNAseq from many genes is averaged together, a modest two-fold increase in density at the 5'-end is observed in the RelE1 data (Figure 1, bottom, blue). These findings are consistent with the ribosome profiling data in showing a 5' to 3' polarity.

Finally, a careful inspection of gene expression across polycistronic transcripts suggests that ribosome density decays not only across single genes but across operons as well. The *manXYZ* operon, for example, is expressed as a single transcription unit from a single promoter, a fact supported by equal levels of RNA density for the three genes in the WT1 RNAseq dataset (Figure 1, top right). Translation of all three genes is also observed in the WT1 Ribo-seq data (top left, black). In contrast, lower levels of RNA and far lower levels of ribosome footprints are observed for the downstream *manY and manZ* genes compared with *manX*.

To quantify this effect, we compared the levels of about 2500 genes in the RelE1 and WT1 samples; these ratios are shown separated by the position of the gene within its operon (Figure 2). The transcriptome of *E. coli* is remarkably complex: many operons overlap and often multiple promoters regulate subsets of genes within an operon. To define single transcription units, we started with roughly 4,000 annotated operons in the RegulonDB database and excluded those in which the genes varied in RNAseq density by more than five-fold (using RNAseq data from the WT1 sample). Like the *manX* gene, which has three-fold more ribosome footprints in the RelE1 sample than in the WT1 sample, most genes that are at the first position in an operon or are monocistronic have more ribosome footprints in the RelE1 sample. In contrast, genes that lie downstream on the operon have fewer ribosome footprints. The *manZ* gene has two-fold

fewer reads in the RelE1 sample than in the WT1 sample (Figure 1). The distribution for genes at the first position versus genes in the fifth position or further downstream is statistically significant (Mann-Whitney P-value of 8.4×10^{-22}). These differences are also observed at the RNA level, although perhaps to a lesser extent. In summary, our data indicate that upon RelE overexpression, ribosomes are enriched not only at the 5'-end of the units of translation (open reading frames) but at the 5'-end of units of transcription (polycistronic mRNAs).

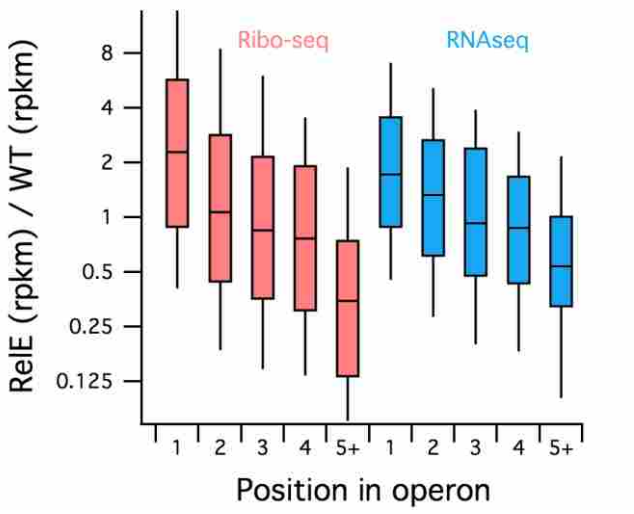

Figure 2 Ribosome enrichment at 5'-end genes of operon
Ribosome occupancy (red) and RNAseq density (blue) are enriched in genes at the 5'-end of operons upon RelE overexpression.

We propose the following model for enrichment of ribosomes at the 5'-end of genes upon RelE expression. As RelE begins to accumulate in the cell, it enters the A site of ribosomes and cleaves mRNA, preventing further cycles of elongation. Stalled ribosomes are rescued by tmRNA-SmpB or by the backup system comprised of the ArfA protein and RF2. Several biochemical studies have shown that tmRNA reacts particularly well with ribosomes following the loss of mRNA downstream of the A site upon RelE cleavage[65]. Both the tmRNA and ArfA systems release the nascent peptide from the stalled ribosome and promote recycling of the subunits. The free subunits are then free to assemble on another transcript at the start codon and

begin elongation again. If RelE cleaves the initiation complex, a strong start codon peak is observed. If a few cycles of elongation take place before cleavage of the mRNA, ribosomes move into the coding sequence away from the start codon. In this manner, cycles of mRNA cleavage, ribosome rescue, and initiation could enrich for ribosomes at the 5'-end of genes.

It is possible that RNA decay mechanisms contribute to this effect as well. Following cleavage by RelE, the 3'-end of the upstream RNA fragment is protected from degradation by exonucleases by the stalled ribosome. The higher ribosome density on the upstream fragment could also prevent degradation by endonucleases such as RNase E. In addition, the upstream fragment has a 5'-triphosphate which is stabilizing. In contrast, the downstream fragment may be less able to recruit ribosomes and will have a 5'-monophosphate that helps to recruit RNase E promoting further endonucleolytic cleavage and RNA decay.

In support of this model, we note that in a previous study of RelE cleavage in highly expressed genes using primer extension, cleavage sites were over-represented at the 5'-end of genes, although no explanation for this observation was provided. This observation is related but subtly different from ours: we observed enrichment of ribosomes at the 5'-end of genes, and they observed high levels of RelE cleavage at the 5'-end of genes. The reloading of ribosomes at start codons followed by a few rounds of elongation prior to additional cleavage events could tie these observations together: RelE is targeted to the 5'-end of genes because that is where the ribosomes accumulate over time.

As further evidence of the importance of rescuing stalled ribosomes, we note that some of the rescue machinery is strongly upregulated in the RelE1 sample. Steady-state levels of the backup rescue factor ArfA are 15-fold higher at the RNA level in the RelE1 sample. The levels of synthesis of the ArfA protein are more than 80-fold increased. ArfA expression is regulated by

tmRNA in an elegant feedback mechanism: the transcript is cleaved by RNase III within the open reading frame, leading to ribosome stalling and production of a truncated protein product. tmRNA rescues these ribosomes stalled during ArfA synthesis and tags the truncated product for proteolysis. If the capacity of tmRNA is exceeded, ArfA begins to accumulate through a poorly understood drop-off mechanism. The fact that ArfA is strongly upregulated in cells treated with RelE indicates that the tmRNA system is overwhelmed and that cells are responding for the need for more ribosome rescue activity by strongly upregulating ArfA expression. Taken together, our findings support a model in which mRNA cleavage by RelE, ribosome recycling, and initiation enrich ribosome occupancy at the 5'-end of genes.

2.2.2 RelE Predominantly Cleaves mRNA after the Second Nucleotide in Empty A sites

The ribosome profiling data obtained from cells overexpressing the RelE endonuclease provide information about cleavage sites genome-wide. In a sense, this information is indirect; we measure ribosome footprints and not the products of RelE cleavage themselves. In preparing

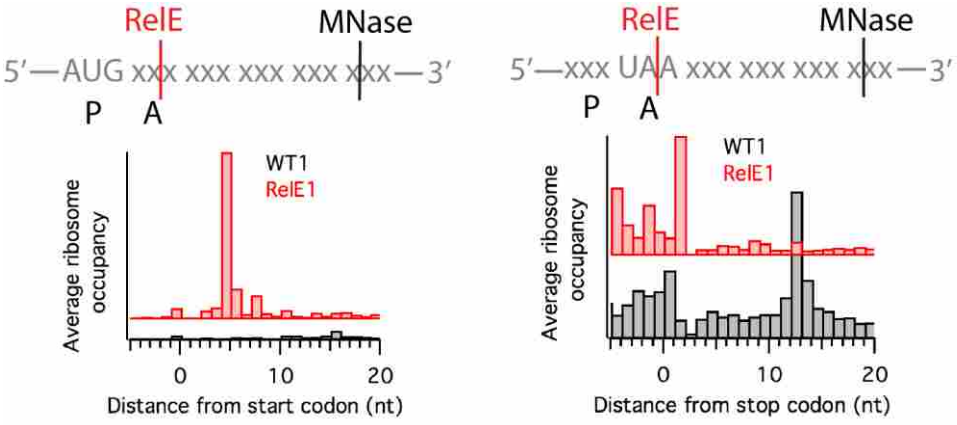

Figure 3 Average ribosome density at start- and stop-codon

Average ribosome density at start (left) and stop codons (right) for Ribo-seq data from wild-type cells or cells overexpressing RelE.

the footprints, another nuclease (MNase) is used to digest mRNA unprotected by ribosomes. Although the MNase treatment could potentially interfere with our ability to detect RelE cleavage events, we find that for the majority (if not all) of the footprints, the 3'-end is generated by RelE cleavage and not MNase. Perhaps this is due to the high concentration of RelE in the cell over the hour-long induction period.

The fact that most ribosome footprints are generated by RelE is clearly evident in the position of the peak at the start codon in plots of average ribosome occupancy (Figure 3, left). During initiation, the AUG start codon is positioned in the ribosomal P site. In the WT1 control sample, MNase digests mRNA back to the 3'-boundary of the ribosome, 15 nt downstream of the first nucleotide of the start codon (black). Although the start codon peak at +15 is small in the WT1 sample, its position matches what we previously observed in many bacterial profiling libraries generated with MNase. The position of the ribosome is most accurately and precisely determined from the 3'-end of footprints in bacterial ribosome profiling; plots of ribosome density reflect the 3'-end of sequencing reads. Furthermore, this position is consistent with the distances observed in toeprinting assays in which reverse transcriptase is blocked by the 3'-boundary of the ribosome and arrests 15-16 nt downstream of the first nucleotide in the P site codon[66]. In contrast, the start codon peak in the RelE1 data is four nt downstream of the first nucleotide of the start codon (red). In other words, the mRNA is cleaved between the second and third nucleotide in the A-site codon within initiation complexes. The fact that the signal is far higher at +4 than at +15 means that most if not all of the ribosome footprints are cleaved by RelE at their 3'-ends.

The same phenomenon is observed at peaks at stop codons in plots of average ribosome occupancy (Figure 3, right). During termination, the stop codons UAG, UGA, or UAA are

positioned in the ribosomal A site. In the WT1 sample, MNase digests back to the 3'-boundary of the ribosome, 12 nt downstream of the first nucleotide in the stop codon (black). In contrast, RelE cleaves after the second nt in the stop codon (red) in the A site. Together, the data at start and stop peaks show that RelE cleaves predominantly after the second nucleotide in the A-site codon and that most ribosome footprints were generated by RelE cleavage at their 3'-ends. This pattern of cleavage is consistent with previous enzymatic, structural, and primer extension studies of RelE activity.

In examining the peaks at other well-characterized ribosome pausing sites, we observed that the ribosomal A site must be empty for RelE cleavage to occur. The SecM and TnaC polypeptides interact with the ribosome to inhibit their own synthesis in response to specific cellular signals[67]. The ribosome stalls at the RAGP sequence in SecM, for example, with the Gly codon in the P site and unreactive Pro-tRNA bound in the A site[67a]. A strong peak 15 nt downstream of the Gly codon in the WT1 sample corresponds to this well-understood pausing event when the samples are treated only with MNase (Figure 4, left). In the RelE1 data, the expected peak after the second nucleotide in the Pro codon is not observed; instead, the strongest

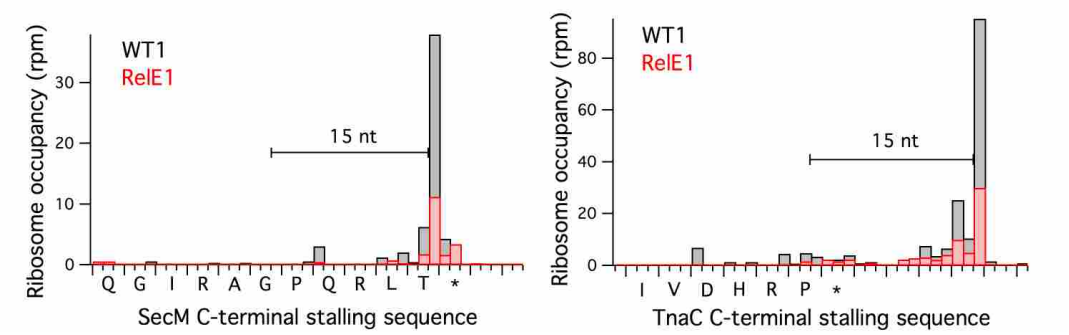

Figure 4 Ribo-seq data at nascent peptide-mediated stalling sites

SecM (left) and TnaC (right).

peak overlaps with the one observed in the WT1 sample 15 nt downstream of the Gly codon. This suggests that RelE is unable to cleave the Pro codon in the A site in stalled SecM complexes, presumably due to the Pro-tRNA trapped and blocking RelE entry. Similarly, ribosome stalling at the C-terminus of the TnaC peptide (Figure 4, right) is observed in the WT1 data with a strong peak 15 nt downstream of the final residue, Pro^{67b}. In the RelE1 data, a peak is expected in the stop codon positioned in the A site, but none is observed, only the same peak observed in the WT1 data. RelE is probably blocked by trapped release factor 2 that is recruited but cannot complete hydrolysis of the nascent peptide. These findings suggest that RelE is capable of competing with aminoacyl-tRNA and release factors for access to the A site during normal elongation and termination, but that factors that are trapped long-term during programmed ribosomal arrest block RelE activity.

2.2.3 *in vitro* Digestion with RelE Reveals the Reading Frame

It has been impossible to detect reading frame in ribosome profiling studies in bacteria and the MNase enzyme is at least partly to blame, leaving one or more nucleotides undigested at the 3'-boundary of the ribosome. Given the precision with which RelE cleaves after the second nucleotide in the A site when expressed *in vivo*, we hypothesized that RelE might prove to be a useful tool in ribosome profiling, generating more reliable 3'-ends and improving the resolution at which the position of the ribosome can be determined. We purified RelE and added it to cell lysates under our regular digestion conditions *in vitro*; MNase was included at the regular concentrations to digest mRNA to the 5'-boundary of the ribosome. Ribosome footprints 10 – 40 nt in length were then cloned and sequenced following our usual protocol.

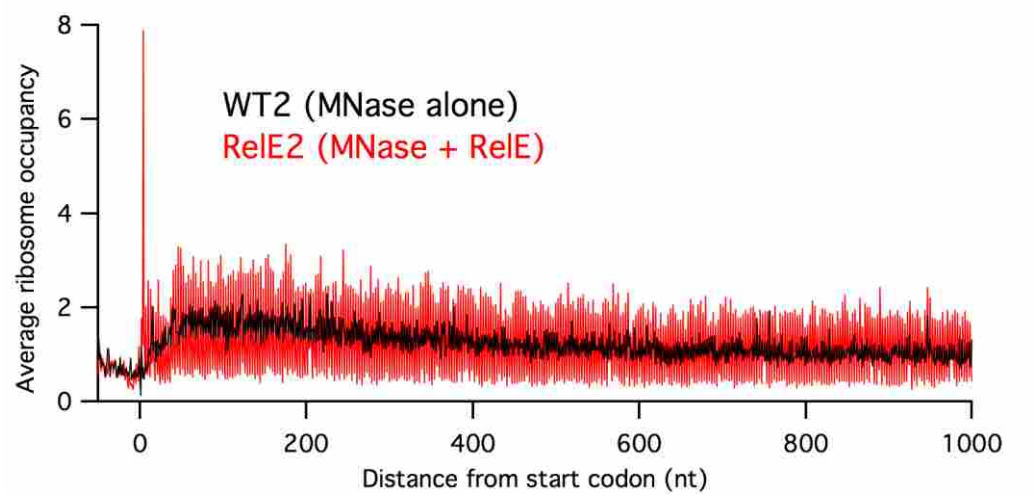

Figure 5 Average ribosome occupancy for a wild-type sample

RNA was digested *in vitro* with MNase alone (black) or with purified RelE protein (red).

Comparison of profiling libraries obtained by expression of RelE *in vivo* (RelE1, Figure 1) with those obtained by *in vitro* RelE digestion (RelE2, Figure 5) reveals many similarities but also an important difference. The activity of RelE *in vitro* can be seen in plots of average ribosome occupancy across many genes aligned at start codons. A strong start codon peak in the RelE-treated sample (RelE2) is found at the +4 and not the +15 site (Figure 5). Furthermore, like the RelE1 data in Figure 1, the vertical spread in the RelE2 signal is substantially larger than that observed in a library prepared from the same biological sample using only MNase (WT2), indicative of 3'-periodicity reflecting reading frame. On the other hand, we do not observe the marked decay of ribosome density at the 3'-end of genes that we observe when RelE is overexpressed *in vivo* (see the data from the RelE1 library in Figure 1). This is because translation is blocked in the cell lysate; elongation is arrested by chloramphenicol and initiation is probably inhibited by dilution of the necessary factors and exhaustion of GTP. This is further

evidence that the accumulation of ribosome occupancy at the 5'-end of genes *in vivo* is due to the dynamics of cleavage, rescue, and initiation and not RelE activity per se.

Returning to the question of reading frame, we calculated the average ribosome density at all three sub-codon positions in open reading frames throughout the genome. The library prepared with MNase (WT2) shows slightly higher density at the first nucleotide (40%) than the other two nucleotides (30% each, Figure 6, top left). Given that this effect is lacking in footprints mapping to 3'-untranslated regions (3'-UTRs) (where the reading frame is defined as the same as the preceding ORF), it may be tempting to attribute this 3 nt periodicity to the ribosome's reading frame (bottom left). We found, however, that RNAseq libraries prepared by digestion of total RNA by low concentrations of MNase yielded the same result—ribosome occupancy is enriched at the first sub-codon position in ORFs but not in 3'-UTRs. The sequence specificity of MNase coupled with the nucleotide bias in ORFs explain this

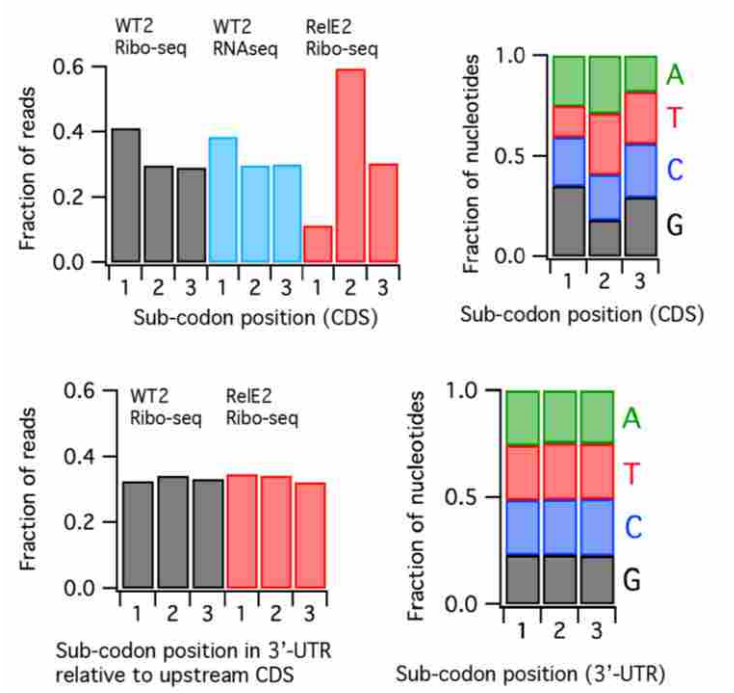

Figure 6 Ribo-seq and RNAseq density at each position

Ribosome density within codons in coding sequences (top) or 3'-untranslated regions (bottom). Right: sequence bias in the genome in the CDS or 3'-UTR.

observation. MNase cleaves preferentially before A and T which are enriched at position two in codons in ORFs but not 3'-UTRs (Figure 6, right). Cleavage before position two makes the 3'-end of footprints align with the first sub-codon position. Given that ribosome occupancy is assigned using the 3'-end of footprints, ribosome density appears higher at the first position (40%) than the other two (30% each).

In contrast, the libraries prepared by RelE digestion *in vitro* show the strongest density at sub-codon position two (Figure 6, top left), corresponding to cleavage before the third nucleotide in the A site codon, as observed at both start and stop codons in average ribosome density plots (Figures 3 and 4). Roughly 60% of footprints map to the second sub-codon position; 30% map to position three while only about 10% map to position one. The large difference between the density at the first two positions explains the large vertical spread in the RelE2 data observed in Figure 5 and provides the first reliable information about ribosome reading frame in bacterial ribosome profiling data. Footprints from the 3'-UTR show no evidence of reading frame, consistent with the expectation that ribosomes found there are not synthesizing protein or moving in 3 nt steps (Figure 6, bottom left).

2.2.4 RelE Prefers to Cleave after C and before G

By analyzing the 3'-ends of ribosome footprints, we are able to observe the specificity of the MNase and RelE enzymes. As others have shown, MNase preferentially cleaves before the nucleotides A and T; in our data we observe that A and T are strongly enriched downstream of the cleavage site, both +1 nt and to a lesser extent +2 nt as well (Figure 7, left). In contrast, A and T are underrepresented upstream of the cleavage site, both −1 nt and to a lesser extent −2 nt. The sequence preferences of RelE are very different from those of MNase: C is preferred at the −1 position while G is selected against (Figure 7, right). Following the cleavage site, G is

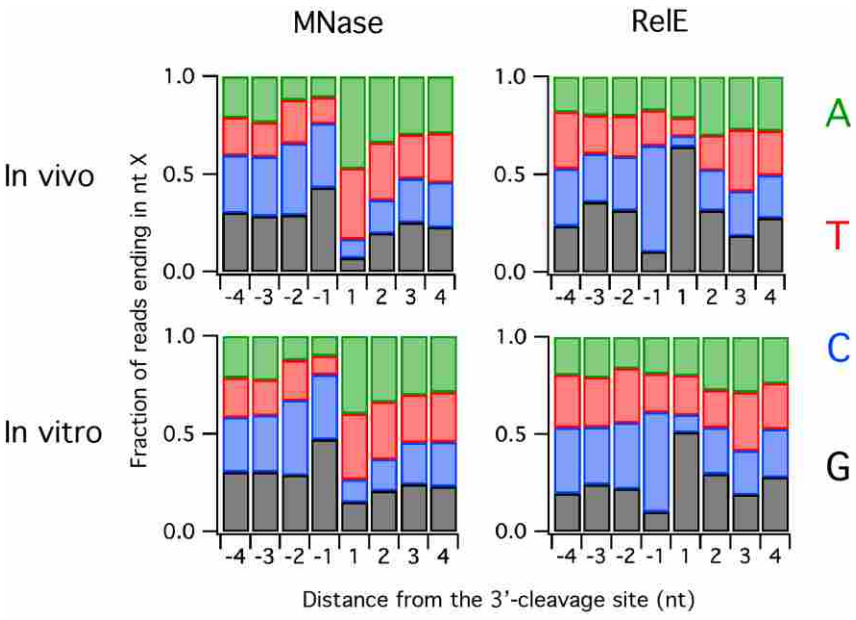

Figure 7 Sequence bias at the 3'-end of ribosome footprints

Different cleavage patterns due to cleavage activity by RelE both *in vivo* and *in vitro*. All samples were also digested with MNase *in vitro*.

strongly preferred and C is selected against. There appears to be a slight avoidance of A and T at both positions. These tendencies hold true whether RelE digestion occurs *in vitro* or *in vivo*. In contrast with what we observed with MNase, the sequence two or more nucleotides away from cleavage site exert a small effect.

Although initial reports of RelE activity described the endonuclease as highly specific for a few codons, additional studies revealed a more relaxed specificity. The preference for G after the cleavage site was anticipated by the kinetic studies of Ehrenberg and co-workers, who found that k_{cat}/K_M values for RelE cleavage were markedly higher for codons ending in G[50a]. The presence of G at the third position of the codon had the strongest effect in explaining their kinetic data. Woychik and co-workers also reported a strong preference for cleavage before G in five highly expressed genes in *E. coli*[52]. Finally, the X-ray crystal structure of RelE bound in the A

site of 70S ribosomes offers a possible explanation of this preference: the G in the third position of the A site codon stacks on the base of C1054, a conserved 16S rRNA nucleotide in the decoding center. The structure suggests that direct contact of RelE residues with the G nucleotide are also possible.

In contrast to the good agreement of our data with these earlier studies regarding the specificity of the downstream nucleotide at the cleavage site, our observations of a marked preference upstream of the cleavage site are unexpected. There is little evidence in the literature of a strong bias for C and against G at the −1 position, raising the possibility that this effect is an artifact of ribosome profiling, arising from biases in cloning ribosome footprints. We find, however, that RNAseq libraries show no such bias when fragments were prepared by alkaline hydrolysis and then cloned and sequenced exactly as the profiling libraries. This finding rules out the possibility that the preference for C is the result of cloning bias and strongly implicates RelE cleavage itself as the cause, given that the method of RNA digestion is the only difference between the two protocols.

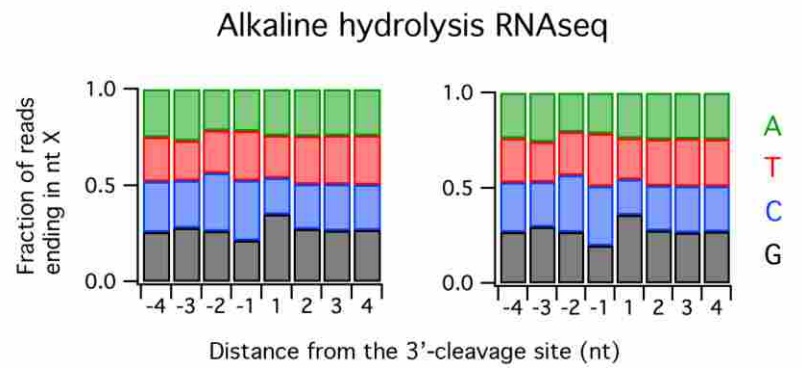

Figure 8 Sequence bias at the 3'-end of RNA fragments

in two libraries. The fragmentation was done by alkaline hydrolysis.

2.2.5 The Specificity of RelE Interferes with Analyses of Ribosome Pausing

Although the generation of ribosome footprints with RelE allows us to determine the ribosome position and reading frame with high resolution, the use of RelE incurs certain disadvantages as well. Both MNase and RelE are sequence-specific nucleases, a property that creates bias at the ends of ribosome footprints. As shown in Figure 9 (top), plots of average ribosome occupancy at sense codons are noisy when the 3'-end of the reads line up with the codon of interest (that is, the distance from the codon is close to zero). This noise arises from enrichment of certain sequences at the 3'-end of reads at the specific nucleotides in the codon being averaged. For example, there is a strong peak at 0 in the GTA codon (green) because cleavage after G and before T is optimal given the specificity of MNase. In contrast, there is a strong peak at 1 for the CCA codon (blue) because cleavage after position 2 is preferred. Although the sequence bias at

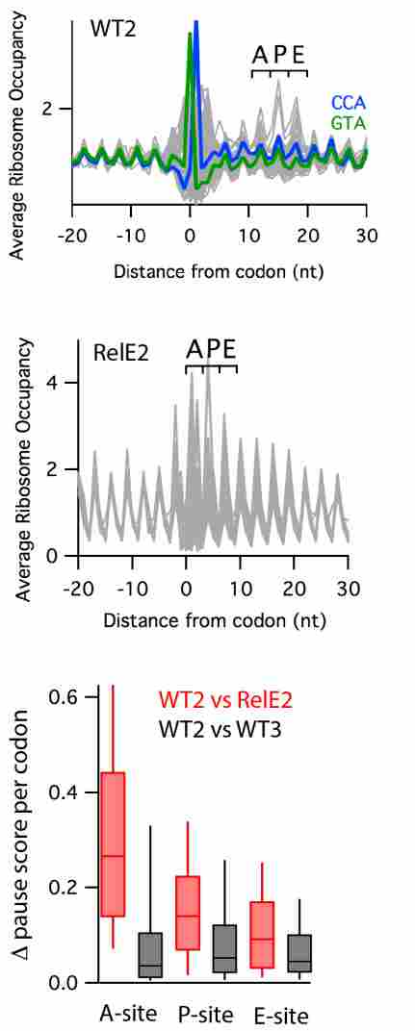

Figure 9 The ribosome occupancy on codons

Top: Average ribosome occupancy at all 61 sense codons. The position corresponding to the codon in the A, P, or E site is indicated. Middle: The same plot with data from the RelE *in vitro* digest. Bottom: the differences in pause score at all 61 sense codons between ribosome profiling libraries.

the 3'-end of reads skews the data, in practice this noise can be ignored because the noise that this creates is at the 3'-boundary of the ribosome, far away from functional sites such as the A, P, and E sites for tRNA-binding. When the 3'-end of reads are 15 nt downstream of codons of interest, positioning the CCA and GTA codons in the P site, for example, noise from cloning bias does not interfere.

In contrast, the pronounced sequence specificity of RelE creates noise that obscures the signal right at functional sites in the ribosome. Because RelE cleaves mRNA in the ribosomal A site, the noise generated by bias at the 3'-end exactly overlaps the A site, making it difficult to draw conclusions about the dwell time of ribosomes on specific codons during elongation or termination (Figure 9, middle). We computed pause scores for three ribosome profiling libraries; two biological replicates in which the footprints were generated by MNase (WT2 and WT3) and the third library from the same biological sample as the WT2 library but using both RelE and MNase (RelE2). We observed relatively small differences in the pause scores between the WT2 and WT3 libraries, in spite of the fact that they were derived from independent biological samples (Figure 9, bottom). This holds true for all three tRNA-binding sites. There were significant differences, however, between the levels of pausing in the WT2 and RelE2 data, even though they were derived from the same biological sample. The differences were more pronounced in the A site than the P or E sites. We conclude that *in vitro* digestion by RelE introduces noise that complicates the analysis of ribosome pausing and argue that the MNase data are more likely to represent an accurate view of pausing *in vivo* given that the cloning bias in the MNase libraries is farther away from the sites of interest.

2.2.6 Refined RelE-Derived Ribosome Density Better Reflects Reading Frame

The sequence specificity of RelE alters the pattern of cleavage in predictable ways; by taking this into account, we can obtain even better information regarding reading frame. Looking in more detail at average ribosome occupancy at sense codons positioned in the A site, we find that most codons are cleaved after the second nucleotide, as expected from the analyses of start

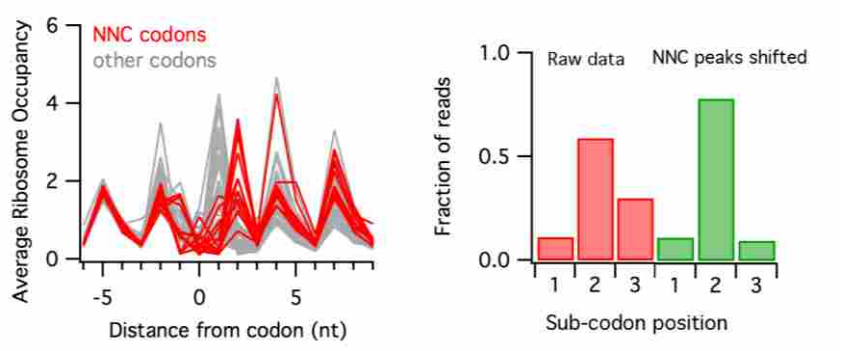

Figure 10 Correction for better reading frame

Left: average ribosome occupancy on all sense codons, shifted from Figure 9 so that the A site codon starts at zero. Right: reading frame when the reads on NNC are shifted from position 3 to position 2.

and stop codons above. In Figure 10 (left), the ribosome occupancy has been shifted such that the A-site codon starts at zero; the peak at 1 observed in most codons (grey) corresponds to cleavage after the second nucleotide. A subset of codons has higher density at position 2 corresponding to cleavage after the third nucleotide in the codon (Figure 10, red): all of these codons end in C. Given the specificity depicted in Figure 7, where cleavage after C is preferred and cleavage before C is strongly inhibited, it makes sense that NNC codons are cleaved after the third position, between the A site codon and the codon downstream. To compensate for this cleavage bias, we shifted the ribosome density at all NNC codons from the third nucleotide to the second. This improved the signal at the second position from about 60% to about 80%, with 10% remaining at position one and 10% at position three (Figure 10, right). This suggests that two-

thirds of reads that align to the third sub-codon position arise from NNC codons. Given what we know about sequence specificity a more sophisticated algorithm could be employed in the future

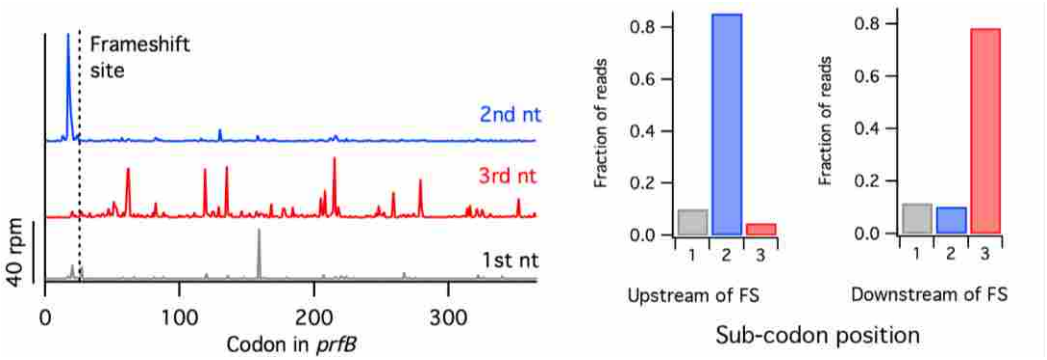

Figure 11 Programmed frameshifting at the *prfB* gene

Left: ribosome occupancy split into three component parts, by sub-codon position. Given RelE's preference for cleaving after the second nt, the blue data represent in-frame translation upstream of the frameshift site and the red data represent in-frame translation downstream (in the +1 frame).

to realign ribosome density to optimize reading frame.

Even by simply shifting density on NNC codons from the third to second position, reading frame is significantly improved. With this level of resolution, frameshifting events can be detected in bacterial ribosome profiling data, as can be seen in analysis of the programmed frameshift on the *prfB* gene encoding RF2[68]. When RF2 levels are limiting, ribosomes pause at a stop codon at the 28th codon in the gene and then shift into the +1 frame. Following the frameshift, ribosomes complete the synthesis of the RF2 protein which is encoded in the new reading frame. When we split the ribosome occupancy signal into three components, with reads that map to the first, second, or third sub-codon position, this frameshifting behavior is evident in models of the *prfB* gene. Upstream of the frameshift site, the majority of ribosome footprints are cleaved after the second sub-codon position, as expected. But downstream of the frameshifting event, most footprints are cleaved after the third position, consistent with a +1 frameshift. Very few reads map to the first sub-codon position, either before or after the frameshift site.

Quantitation of the reading frame upstream and downstream of the programmed frameshift site further supports this conclusion. Going forward, the high degree of three nucleotide periodicity arising from translational reading frame can be used to search for other programmed frameshifting sites in bacteria, as has been done in human ribosome profiling studies[37] .

2.2.7 Concluding Thoughts

Our ribosome profiling analyses of RelE activity *in vivo* reveal dynamic cycles of cleavage, ribosome rescue, and initiation that enrich ribosomes at the 5'-end of genes. Although it was previously known that RelE-cleaved mRNAs are good targets for the tmRNA rescue system, our data highlight the importance of the tmRNA and ArfA systems in the cellular response to and recovery from RelE activity. How cells exit the dormant state with high RelE activity is not yet clear, but it seems likely that ribosome rescue is essential for this to occur. Antibiotics are beginning to be developed to target the tmRNA rescue pathway, raising the possibility of targeting persister cells due to their reliance on toxins to cleave mRNA and block translation.

The use of RelE as a nuclease for ribosome profiling suffers from the same problem of sequence specificity as the enzyme normally used, MNase. This makes the enzyme less than ideal for generating libraries where careful measurement of ribosome pausing is the goal. However, RelE cleaves the 3'-end of fragments more precisely, allowing us to observe the ribosome's reading frame for the first time in bacteria.

In the future, it may be possible to generate ribosome profiling libraries using RelE alone. In our *in vitro* digests, we also added MNase to cleave the mRNA back to the 5'-end of the ribosome generating a ribosome footprint. With RelE alone, cleavage would only occur in the A site, generating fragments that are far longer than the 10 – 40 mer ribosome footprints that we

work with now. In theory, these fragments could be cloned without size selection. The distance between ribosomes on messages could then be determined by sequencing both ends using 50 bp paired Illumina sequencing. This might give insights into ribosome stacking at strong pausing sequences and a picture of ribosome distribution on single messages, since the sequence between two ribosomes would remain intact.

2.3 Materials and Methods

2.3.1 Bacterial Strains and Growth Conditions

E. coli K-12 strain MG1655 was used as a wild-type strain. The RelE overexpression strain was constructed by transformation with a plasmid, pJC203, which contains an araBAD promoter. Cell cultures were grown at 37°C in either LB media or MOPS media supplemented with 1% glucose, all 20 amino acids, and other nutrients (Teknova) and RelE overexpression was induced by adding arabinose (final 0.2%) at an $O.D_{600}$ of 0.2-0.3. Cell cultures were grown further either for 1h or 20 min before harvesting.

2.3.2 Preparation of Ribo-seq and RNAseq Libraries

Ribo-seq libraries were generated as described previously[66] with some modifications. Ribosome-protected fragments were size-selected between 20 – 40 nt in length initially, but 10 – 40 nt in the latter libraries in an attempt to include short fragments generated by RelE cleavage activity. For the library digested by RelE *in vitro*, 1 nmol of purified RelE was also added when 25 AU of RNA in the lysate was digested with 3000 U of MNase (Roche) for 1 h at 25 °C.

RNAseq libraries were also size-selected between 10 – 40 nt in length in the latter preparation to reveal 3'-end sequences, which was not available due to longer reads size selection with the length between 40 – 60 nt.

2.3.3 Data Analysis

Data analysis obtained from Illumina sequencing were described previously[66].

3 THE ROLE OF tRNA METHYLTRANSFERASE TrmD

This work was a collaboration with Isao Masuda of Ya-Ming Hou's lab at Thomas Jefferson University. Dr. Masuda created and characterized the TrmD depletion system and his work is described here for completeness. My contribution was to assist him in optimizing the growth conditions, generate the ribosome profiling libraries, and analyze and interpret the profiling sequencing data.

3.1 Introduction

tRNAs are adaptor molecules that deliver amino acids to the ribosome for protein synthesis. To ensure the fidelity of translation, each tRNA must carry its cognate amino acid and decode mRNA using the correct match between its anticodon and the codon in mRNA. tRNAs undergo numerous chemical modifications after their transcription; an average *E. coli* tRNA contains 77 nucleotides and have about 8 of them modified. These modifications play an important role in the fidelity of both aminoacylation and decoding.

Most tRNA modifications are located near the anticodon[69]. One of the purposes of modification is to fine tune the anticodon so that a single tRNA can read multiple, degenerate codons. There are 47 different tRNAs produced from 86 genes in *E. coli*; these tRNAs are able to decode all 61 sense codons. tRNA modifications promote this flexibility in codon recognition while at the same time maintaining fidelity. This is a difficult but critical balance to achieve: the

tRNAs must be allowed a certain amount of broadening of specificity but only under the right circumstances.

tRNA modifications also play a critical role in maintaining the reading frame during translation[46]. One particularly important enzyme for preventing frameshifting is TrmD, a bacterial enzyme responsible for methylation of N1 of G37 of several tRNAs[70]. Unlike the case with other modifications, deficiency of m^1G37 leads to cell death in many bacteria including *E. coli*[70b, 71]. TrmD is very different structurally and mechanistically from its eukaryotic homolog, Trm5[72]. Studying *E. coli* TrmD can give us a framework for understanding methyltransferases, the role of modifying enzymes in tRNA stability and activity, and insight into a promising antibiotic target. Here we report our efforts to understand the role of TrmD in living cells using ribosome profiling.

3.1.1 tRNA Synthesis and Post-Transcriptional Modifications

In *E. coli*, tRNAs are transcribed as long precursors[73]. Various RNases process the precursor tRNAs (pre-tRNAs) to mature, functional tRNAs by cutting excess residues at the 5'- and 3'-ends[74]. For most pre-tRNAs, RNase E removes nucleotides at the 3'-end by an endonucleolytic cleavage, leaving only a short 3'-trailer[75]. Then, RNase P interacts with the CCA motif and removes nucleotides at the 5'-end of the pre-tRNA with a single endonucleolytic cleavage; this creates the mature 5'-end of the tRNA. Finally, exonucleases such as RNase T, PH, II, and D further trim the remaining 3'-trailer nucleotides up to the mature CCA motif at the 3'-end of the tRNA[73].

All tRNAs have L-shaped tertiary structures formed by base pairing and coaxial stacking of the helices. Each tRNA is slightly different but shares common secondary structural features[76]. When depicted in two-dimensions, tRNA have a cloverleaf structure consisting of an

acceptor stem, D (dihydrouridine) loop, anticodon loop, and T loop. The acceptor stem contains the conserved 3'-terminal CCA motif where the amino acid is covalently linked[77]. The structure of the anticodon loop presents the anticodon in a way that allows it to form the correct base pairs during decoding[77]. The other structures contribute to interactions with the translation machinery, including the ribosome, aminoacyl-tRNA synthetases (aaRSs), and EF-Tu.

Following transcription, tRNAs are modified at several sites to enhance their cellular functions, promoting their structure and stability, optimizing translational fidelity and accuracy, and maintaining the correct reading frame. The first modification identified was pseudouridine (Ψ), as the fifth nucleotide, which is an isomer of uridine[78]. Ψ is present in the T loop that is recognized by aaRS's and EF-Tu. Likewise, the dihydrouridine modification is found in the D loop[79]. With two extra hydrogens added to its pyrimidine ring, dihydrouridine is no longer planar or aromatic, disturbing base stacking, and destabilizes the structure, conferring flexibility[79].

Most critical modifications of tRNA, however, occur in the anticodon loop. The first two positions in a given codon base pair (in classical Watson-Crick fashion) with the second and third positions of the anticodon, nucleotides 35 and 36 in the tRNA. The third nucleotide in the codon in the 'wobble' position, however, can participate in a number of different pairing geometries in its interaction with position 34 of the tRNA. Modification of position 34 can broaden base pairing[80]: deaminating A34 to inosine, for example, in $tRNA^{Arg}$ in bacteria allows for pairing with A, C, and U at the wobble position. Alternatively, modifications of U34 to 5-methoxycarbonyl-methyl-2-thouridine (mcm^5s^2U34) and the 5-methylation of cytosine 34 (m^5C34) allow a single tRNA isoacceptor to read multiple codons[81]. The modifications promote accurate codon-anticodon interactions, optimizing translation of the genetic code.

tRNA modifications also play a critical role in maintaining the reading frame of translation[80, 82]. In most cases, frameshifting is far more detrimental than other decoding errors. Missense errors that replace one amino acid for another often have little or no effect on the stability or activity of the protein product. On the other hand, most shifts in reading frame lead to pre-mature termination at a stop codon in the new reading frame, yielding truncated and inactive polypeptides. Pseudouridine and 2-thiouridylation (s^2U_{34}) play roles in frame maintenance in addition to their other functions[78a, 83]. One of the best characterized modifications involved in frame maintenance is the methylation of G37, just downstream of the anticodon.

3.2 The TrmD Methyltransferase

The m^1G37 modification involves a very ancient pathway; proteins catalyzing G37 methylation are found in all three domains of life[84]. m^1G37 is found in most tRNAs that read codons starting with C. It is thought that methylation of G37 may block base pairing between this nucleotide and the mRNA that would promote decoding of a quadruplet instead of a triplet (Figure 12)[84]. Codons starting with C are read by tRNAs with G36; the combination of G36 and unmethylated G37 is thought to be problematic. Other mechanisms involving structural changes in the tRNA or faulty translocation on the ribosome may also play a role.

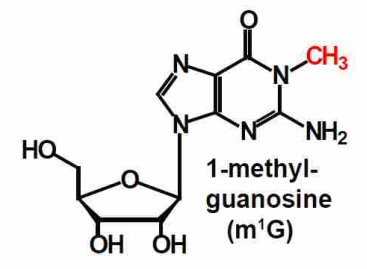

Figure 12 1-methyl-Guanosine

There are many enzymes that facilitate tRNA methylation. Some enzymes add methyl groups to carbon or nitrogen atoms in the bases, while others methylate oxygens in the ribose sugars. In general, methyltransferases catalyze the transfer of a methyl group (CH_3) from the donor to a substrate molecule. In most cases, the methyl donor is *S*-adenosyl-L-methionine

(AdoMet)[85]. AdoMet-dependent methylation often involves SN2 substitution reactions in which the substrate attacks on the reactive methyl group in AdoMet, yielding a single stereoisomeric product.

In bacteria, methylation of G37 is accomplished by the TrmD enzyme, which is composed of 255 amino acids (28.4 kD) and acts as a homodimer. The active site is formed by residues from each of the monomers[86]. The structure contains a unique trefoil knot deep within each of the homodimers whose function is not fully understood. When AdoMet binds to the structure, it causes conformational changes. The adenosine moiety of AdoMet is surrounded by a loop and AdoMet is in a bent conformation[85].

It has been proposed that the TrmD homodimer requires the anticodon loop of tRNA to be precisely positioned near AdoMet for the methylation of G37 since bound AdoMet is located in a deep, seemingly inaccessible pocket near the center of TrmD[85]. The conserved phenylalanine residue is proposed to provide critical stacking interactions required for the final positioning of G37[86]. It has been proposed that an aspartate of TrmD acts as a general base while other conserved residues provide G37 specificity for the reaction allowing methylation of the N1 position of G via a deprotonation, followed by the transfer of the methyl group of AdoMet[85].

TrmD binds to tRNAs non-specifically searching for its recognition element, the dinucleotide G36-G37[86]. Once TrmD recognizes the two consecutive G residues at the 3' end of the anticodon, AdoMet-dependent methylation occurs[86]. tRNA is positioned in the cleft and, if the identity element G36pG37 is present, G36pG37 might be 'flipped' into the dual pocket catalytic center for subsequent methylation by bound AdoMet[85-86]. If a G is not found at position 36, then a stable complex of G37 near AdoMet cannot form and catalysis does not proceed[86].

The bacterial TrmD enzyme is substantially different from its eukaryotic homolog, Trm5, raising the possibility of targeting TrmD with antibiotics[87]. The two enzymes show no detectable similarity in sequence or structure. While TrmD is an obligate dimer, Trm5 is a monomer[88]. TrmD binds AdoMet with the rare trefoil knot, Trm5 binds AdoMet in the open space of a dinucleotide-fold. As noted above, TrmD binds AdoMet in a bent conformation, where the adenosine and methionine moieties are bent toward each other by 90° whereas Trm5 binds AdoMet in the more common straight conformation, with an 180° angle between the two moieties[86, 89]. The fact that these enzymes have different active site geometries and both bind small molecules makes them attractive drug targets, and pharmaceutical companies are working on developing small molecule inhibitors of TrmD[84]. Loss of TrmD activity is lethal in *E. coli*; this is unusual for tRNA modification enzymes, most of which can be deleted with little or no detectable phenotype. Although increased levels of frameshifting are observed when TrmD and m^1G37 are lost, it is not entirely clear what the mechanism of cell death is. To obtain a better understanding of protein synthesis in living cells upon depletion of G37 methylation, we performed ribosome profiling.

3.3 Results

3.3.1 Achieving Regulation of TrmD, an Essential Protein

An inducible TrmD-depletion *E. coli* strain was constructed to observe defects in translation and identify genes whose expression is altered when m^1G37 of tRNA is depleted in cells. Since m^1G37 is an essential modification of several tRNAs and deletion of *trmD* is lethal, we constructed a strain in which TrmD can be conditionally depleted, starting with the G78 strain[90]. In G78, proteolysis of certain target proteins can be regulated by controlling the

expression of the peptide recognition protein, ClpX, which binds to hydrophobic C-terminal peptide tags, and Clp, a protease. The ClpXP complex degrades proteins tagged by tmRNA as well as many other proteins with C-termini recognized by the complex[91]. In the G78 strain, the ClpXP genes are regulated by an arabinose inducible promoter, allowing us to regulate its expression by changing media conditions[90]. We also express TrmD from a constitutive promoter with the tmRNA tag translationally fused to its C terminus; this effectively targets the protein for degradation when ClpXP is expressed upon induction with arabinose.

To construct the TrmD-depeletion strain, the *trmD* gene, which is expressed with another three genes on an operon, is modified at its 3'-end to contain a peptide tag (Figure 13). This tag contains a sequence that targets it for degradation by ClpXP (YALAA) as well as His_6 and FLAG epitopes (Figure 13)[90]. This is called the degron strain (with the Deg-tag); the control strain also contains the epitope tags but not the YALAA degradation signal (the Cont-tag). The trmD mutations were first established by λ red recombination in *E. coli* strain SM1405[90]. Then, the *trmD* alleles with the added degradation tag or control tag are

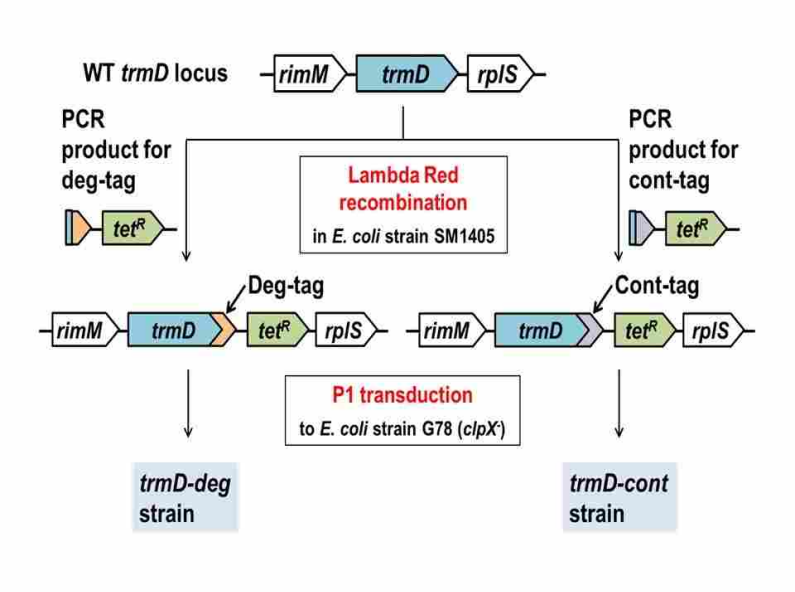

Figure 13 TrmD depletion strain construction

verified by PCR, and then the alleles were transduced using P1 phage in the G78 strain in which ClpXP can be regulated, generating the trmD-deg and trmD-cont strains (Figure 13).

The two strains were tested by growing on agar plates (Figure 14, top). When cells were plated on media containing glucose, both strains formed colonies (data not shown). However, when cells were grown on the plate with 0.2% arabinose added, there were significant differences in colony formation. The degron strain (trmD-deg) formed fewer colonies than the control strain, suggesting that the tagged TrmD was successfully depleted upon ClpXP overexpression, arresting cell growth. In addition, steady state levels of TrmD were visualized by western blot (Figure 14, bottom). Cells were first grown in LB media (with trace amount of glucose) to suppress ClpXP expression. Upon induction with arabinose (final concentration of 0.2%), the cells were collected and subjected to western blot assay with various time points. The western blot shows that TrmD levels drop significantly in the cell between 30 – 40 min after ClpXP overexpression.

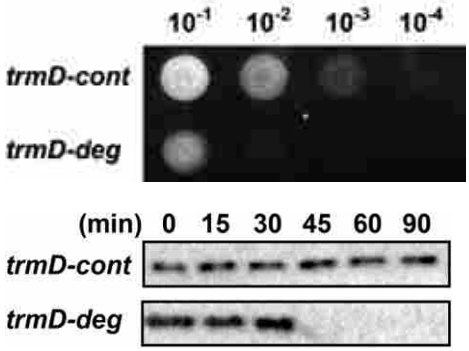

Figure 14 TrmD depletion

Cell growths on plate (Top) and western blot for tagged proteins (Bottom).

Having confirmed that colonies fail to grow when TrmD is depleted, and that the depletion kinetics are fairly rapid, growth curves were obtained to determine the optimal time after induction for harvesting the cultures for ribosome profiling. Surprisingly, however, the growth curves of degron and control strains did not show much differences (data not shown). We hypothesized that the slow decay of residual methylated tRNAs, produced prior to TrmD

depletion, left enough tRNA for the cells to continue to function and grow relatively normally over short time frame.

By diluting the culture twice back to an OD_{600} of 0.1, we found that differences in growth rates between the degron and control strains could be observed after about 4 h of induction of the ClpXP protease (Figure 15). Without the dilutions, the cultures reached saturation at nearly the same rates. With the dilutions, however, differences could be observed when the cultures were still growing in exponential phase. It was important not to reach saturation because starvation for certain amino acids ensues during stationary phase, creating a signal of pausing at those codons in the ribosome profiling data that would interfere with our analyses of the effects of depletion of trmD (Figure 15).

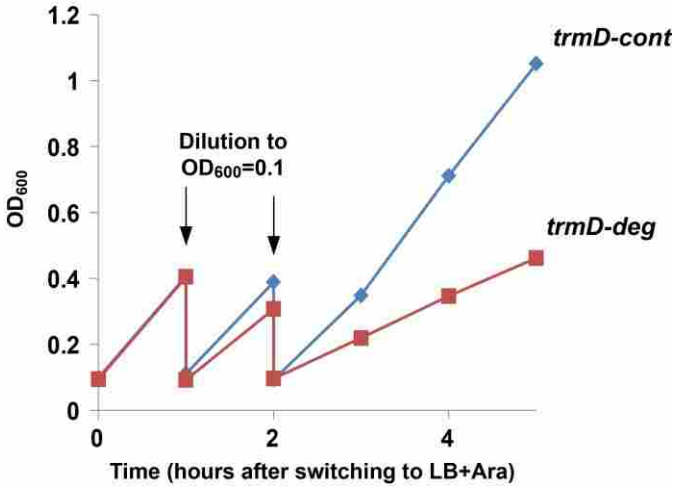

Figure 15 Growth measurements upon TrmD depletion

Cell growth in LB media. At Time 0, exponentially growing cells were diluted in fresh media with 0.2% arabinose for ClpXP induction. Additional dilutions of cell culture media to fresh media after 1h and 2h.

We then monitored the level of m^1G37 over the 4 h induction time course. A 14-nt long primer complementary to $tRNA^{Leu/CAG}$ was used for extension by reverse transcriptase (Figure 16, left). Methylation blocks extension, generating a 15 nt product. In the absence of methylation

at G37, a 53 nt product is generated. In the degron strain, a gradual depletion m^1G37 methylated tRNA was observed (Figure 16, right). At the same time, newly synthesized tRNA is

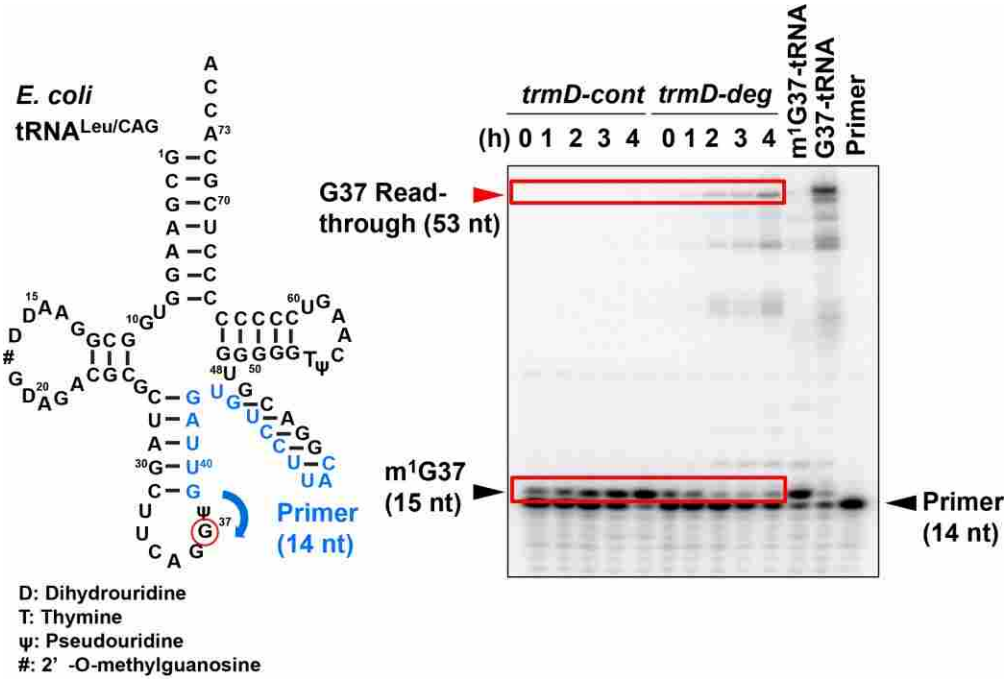

Figure 16 Effect of TrmD depletion on tRNA

Primer extension assay using tRNAs as templates for verifying methylation depletion on tRNA.

unmethylated, as evidenced by the gradual accumulation of the 53 nt product over time in the degron strain (Figure 16, right). No unmethylated tRNA was evident in the control strain.

3.3.2 Depletion of m^1G37 of tRNA Causes Pauses on CCG, CCA, and CGG Codons

Ribosome profiling libraries were generated from cultures of the control and degron strains after 4 h of induction. We analyzed the data for pauses at specific codons that are read by tRNAs normally containing m^1G37. To detect pauses, we averaged the ribosome occupancy surrounding instances of the 61 sense codons. Plots of average ribosome occupancy for the control strain, for example, show very strong pauses when Ser codons are found in the ribosomal A, P, and E sites (Figure 17). In these plots (Figure 17 and 18), the signal at 0 corresponds to the

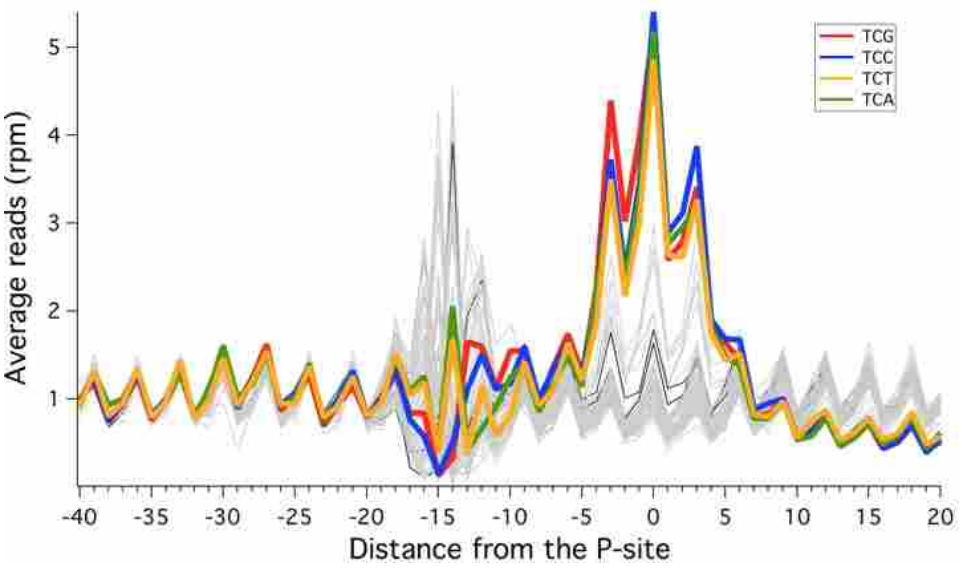

Figure 17 Ribosome occupancy on codons in control strain

High level of ribosome occupancy on some codons (colored) compared to other codons (grey).

codon is positioned in the P site, in the A site at -3, and in the E site at +3. (Note that the noise between -10 and -20 corresponds to sequence bias at the 3'-end of the reads and can be ignored). The Ser codons TCG, TCC, TCT, and TCA have the strong signals in the control strain, shown in color (Figure 17). Strong pauses on serine codons are a signature of growth in LB media that is frequently observed in *E. coli* ribosome profiling. Presumably this occurs because levels of Ser within the cell are limiting for protein synthesis.

In the degron strain (Figure 18), on the other hand, the usual Ser pauses are not observed and other pauses have become rate limiting. Ribosome occupancy is strongly enriched at CCG, CCA, and CGG codons. This result, seemingly representing the effect of depletion of tRNAs containing m^1G37, makes sense because all three of these codons are decoded by tRNAs with m^1G37. Perhaps ribosomes pause at these codons because the concentration of tRNA is reduced or because unmethylated tRNA behaves poorly in decoding or translocation.

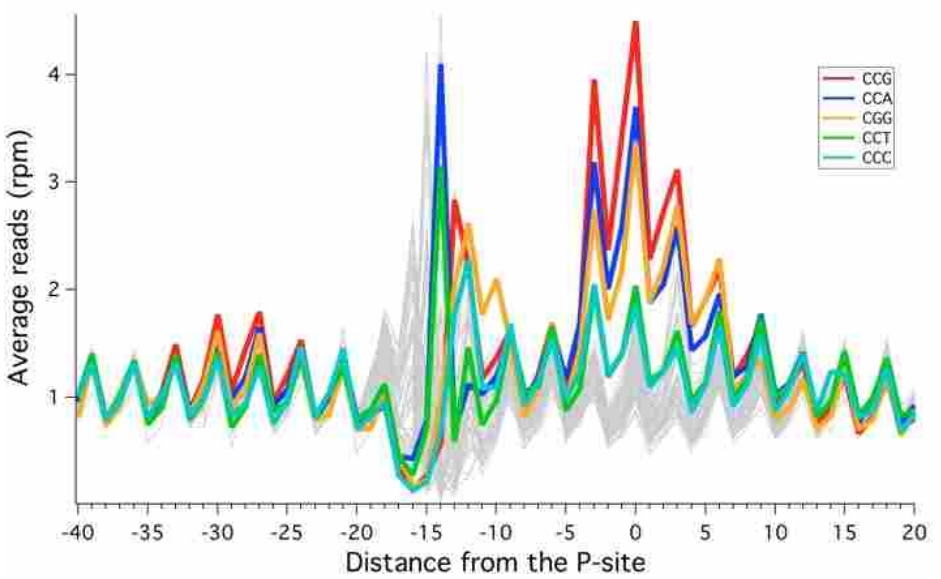

Figure 18 Ribosome occupancy on codons in degron strain

High level of ribosome occupancy on some codons (colored) compared to other codons (grey).

What is surprising is that only these three codons have strong pauses in the profiling data from the degron strain, when seven or perhaps nine codons are predicted to be decoded by tRNAs containing m^1G37. Table 1 lists the codons that start with C and are decoded by tRNA isoacceptors that are candidate substrates of TrmD. Note that the CAN codons encoding His and Gln are not shown in the table; the corresponding tRNAs contain A37 rather than G37 and are therefore not substrates of TrmD; pauses are not observed on these codons. The same thing is true for three of the CGN Arg codons: no pauses are observed for the CGU, CGC, and CGA, codons that are decoded by a tRNA containing an ICG anticodon with A37 and not G37 and so not a TrmD substrate. In contrast, the CGG Arg codon whose cognate tRNA has m^1G37 modification shows very strong pauses, as noted above. This is perhaps the simplest case.

Surprisingly, the CUN leucine codons do not show pausing, even though they are read by tRNAs thought to contain m^1G37. There is some controversy in the literature over whether *E. coli* $tRNA^{Leu/GAG}$ (green) is actually methylated; this tRNA decodes the CUC and CUU codons. Further work will be required to ascertain the methylation status of G37 of this tRNA, but our data may argue against it. Another possibility is that because both the GAG and CAG anticodon isoacceptors for $tRNA^{Leu}$ have pseudouridine at position 38, they may be able to perform correct base pairing even when m^1G37 is depleted. Only these two tRNAs have pseudouridine at position 38 among the isoacceptors in Table 1.

The pausing on Pro codons is likewise surprising. All three $tRNA^{Pro}$ isoacceptors are bona fide substrates for TrmD; why are strong pauses observed only at CCG and CCA? One thought is that the GGG isoacceptor that reads the other two codons (CCC and CCU) can maintain some activity even in the absence of methylation. The possibilities for frameshifting are increased dramatically with this isoacceptor (which has the four nt anticodon GGGG in the absence of methylation). Perhaps pausing on the CCC and CCU codons leads to rapid frameshifting so that strong pauses are not observed.

	Codon	Anti-codon	
Leucine	CUC	$GAGm^1G37$	
	CUA		
	CUG		$CAGm^1G37$
	CUU		
Proline	CCU	$GGGm^1G37$	$UGGm^1G37$
	CCC		
	CCA		
	CCG	$CGGm^1G37$	
Arginine	CGU	ICG	
	CGC		
	CGA		
	CGG		$CCGm^1G37$

Table 1 Codons and their anti-codons with modifications

Codons with most ribosome occupancy (red) and isoacceptors with m^1G37 modification.

3.4 Materials and Methods

3.4.1 Strains and Construction

The figure 13 shows the scheme of constructing the trmD-deg and trmD-cont cells. First, the degron- and control-tag followed by tetracycline marker were amplified from plasmids created by Carr *et al*, provided by Dr. Sean Moore[90].

The primers contain homologous extensions at 5' ends to the trmD-flanking regions (Forward: CGGAACACGCACAACAGCAACATAAACATGATGGGATGGCGGGTGGCTCC GACTACAAGG, Reverse: ATAATTTAATCTCTTATCCTGGGTAAACTGATATCTCGGGG GCTTAGGTCGAGGTGGCCC).

PCR products were electroporated into *E. coli* recombinogenic strain SM1405[90] and homologous recombination was confirmed by colony-PCR using primers targeting 5'-terminal and 3'-flanking regions of trmD (Forward: ATGTGGATTGGCATAATTAGCCTGTTT CC, Reverse: GAATTCCGGTTACGAATAGCGATAACCACGCC).

trmD-deg SM1405 and trmD-cont SM1405 were used as a donor for P1 phage lysate preparation, and the tagged chromosomal trmD loci were transferred by P1 transduction to the recipient *E. coli* strain G78, in which the chromosomal *clpX* gene is knocked out. After tetracycline marker selection, genotype was confirmed again by colony-PCR.

As described[90], trmD-deg G78 strain was then transformed with a library of pClpPX which contains random mutations at the promoter region of the *clpPX* genes, and transformant clones were screened out to pick up the one with highest efficiency of degradation, as proven also by the rapid degradation at protein level. This specific plasmid clone that confers high degradation efficiency was extracted from trmD-deg G78 and was transformed into trmD-cont G78 strain. This pair of trmD-cont G78 and trmD-deg G78 is used for all the experiments.

3.4.2 Cell Growth and Preparation of Ribo-seq Libraries

Each cell from degron- and control-strain was grown in LB media supplemented with 0.2% glucose, 0.2% glycerol and an appropriated antibiotic at 37° C for overnight. The next day, fresh 10-mL LB media supplemented with 0.002% glucose, 5 mM serine and an appropriate antibiotic were inoculated with overnight cell culture media with 100-fold dilution, followed by incubation for 1.5-2 h shaking at 37° C.

By measuring the $O.D_{600}$ to reach up to 0.3-0.4, cell cultures were diluted aiming $O.D_{600} \approx 0.1$ in fresh 300 mL LB media supplemented with 0.2% arabinose, 5 mM serine and an appropriate antibiotic, followed by incubation for 1 h shaking at 37° C. Then, by measuring the $O.D_{600}$, cell cultures were further diluted in 300-mL LB media supplemented with 0.2% arabinose, 5 mM serine and an appropriate antibiotic and this was repeated once more after another hour. Finally, when the $O.D_{600}$ reached up to 0.3, cells were harvested by filtering cell culture using 49 μM filter paper and the cell pellet was flash frozen in liquid nitrogen.

Ribo-seq libraries were generated as described previously[66] with some modifications. Ribosome-protected fragments were size-selected between 10 – 40 nt. Cell lysates were digested with 3000 U of MNase (Roche) for 1 h at 25 °C.

3.4.3 Data Analysis

Data analysis obtained from Illumina sequencing were described previously[66].

REFERENCES

1. Ibba, M.; Soll, D., Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev* **2004,** *18* (7), 731-8.

2. Ramakrishnan, V., Ribosome structure and the mechanism of translation. *Cell* **2002,** *108* (4), 557-72.

3. Boelens, R.; Gualerzi, C. O., Structure and function of bacterial initiation factors. *Curr Protein Pept Sci* **2002,** *3* (1), 107-19.

4. Steitz, J. A.; Jakes, K., How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. *Proc Natl Acad Sci U S A* **1975,** *72* (12), 4734-8.

5. (a) Shine, J.; Dalgarno, L., 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* **1974,** *71* (4), 1342-6; (b) Marzi, S.; Myasnikov, A. G.; Serganov, A.; Ehresmann, C.; Romby, P.; Yusupov, M.; Klaholz, B. P., Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell* **2007,** *130* (6), 1019-31.

6. Shatkin, A. J., Capping of eucaryotic mRNAs. *Cell* **1976,** *9* (4 PT 2), 645-53.

7. Carter, A. P.; Clemons, W. M., Jr.; Brodersen, D. E.; Morgan-Warren, R. J.; Hartsch, T.; Wimberly, B. T.; Ramakrishnan, V., Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **2001,** *291* (5503), 498-501.

8. Ogle, J. M.; Carter, A. P.; Ramakrishnan, V., Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci* **2003,** *28* (5), 259-266.

9. Diaconu, M.; Kothe, U.; Schlunzen, F.; Fischer, N.; Harms, J. M.; Tonevitsky, A. G.; Stark, H.; Rodnina, M. V.; Wahl, M. C., Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **2005,** *121* (7), 991-1004.

10. Berchtold, H.; Reshetnikova, L.; Reiser, C. O.; Schirmer, N. K.; Sprinzl, M.; Hilgenfeld, R., Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* **1993,** *365* (6442), 126-32.

11. Nissen, P.; Kjeldgaard, M.; Thirup, S.; Clark, B. F.; Nyborg, J., The ternary complex of aminoacylated tRNA and EF-Tu-GTP. Recognition of a bond and a fold. *Biochimie* **1996,** *78* (11-12), 921-33.

12. Ogle, J. M.; Brodersen, D. E.; Clemons, W. M.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V., Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **2001,** *292* (5518), 897-902.

13. Voorhees, R. M.; Schmeing, T. M.; Kelley, A. C.; Ramakrishnan, V., The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **2010,** *330* (6005), 835-8.

14. Blanchard, S. C.; Gonzalez, R. L.; Kim, H. D.; Chu, S.; Puglisi, J. D., tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol* **2004,** *11* (10), 1008-14.

15. Voorhees, R. M.; Ramakrishnan, V., Structural basis of the translational elongation cycle. *Annu Rev Biochem* **2013,** *82*, 203-36.

16. (a) Li, W.; Frank, J., Transfer RNA in the hybrid P/E state: correlating molecular dynamics simulations with cryo-EM data. *Proc Natl Acad Sci U S A* **2007,** *104* (42), 16540-5; (b) Flanagan, J. F. t.; Namy, O.; Brierley, I.; Gilbert, R. J., Direct observation of distinct A/P hybrid-state tRNAs in translocating ribosomes. *Structure* **2010,** *18* (2), 257-64.

17. Dorner, S.; Brunelle, J. L.; Sharma, D.; Green, R., The hybrid state of tRNA binding is an authentic translation elongation intermediate. *Nat Struct Mol Biol* **2006,** *13* (3), 234-41.

18. Klaholz, B. P.; Pape, T.; Zavialov, A. V.; Myasnikov, A. G.; Orlova, E. V.; Vestergaard, B.; Ehrenberg, M.; van Heel, M., Structure of the Escherichia coli ribosomal termination complex with release factor 2. *Nature* **2003,** *421* (6918), 90-4.

19. Ito, K.; Ebihara, K.; Uno, M.; Nakamura, Y., Conserved motifs in prokaryotic and eukaryotic polypeptide release factors: tRNA-protein mimicry hypothesis. *P Natl Acad Sci USA* **1996,** *93* (11), 5443-5448.

20. Korostelev, A. A., Structural aspects of translation termination on the ribosome. *Rna-a Publication of the Rna Society* **2011,** *17* (8), 1409-1421.

21. Mora, L.; Heurgue-Hamard, V.; Champ, S.; Ehrenberg, M.; Kisselev, L. L.; Buckingham, R. H., The essential role of the invariant GGQ motif in the function and stability in vivo of bacterial release factors RF1 and RF2. *Mol Microbiol* **2003,** *47* (1), 267-75.

22. Freistroffer, D. V.; Pavlov, M. Y.; MacDougall, J.; Buckingham, R. H.; Ehrenberg, M., Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *Embo J* **1997,** *16* (13), 4126-33.

23. Pavlov, M. Y.; Freistroffer, D. V.; MacDougall, J.; Buckingham, R. H.; Ehrenberg, M., Fast recycling of Escherichia coli ribosomes requires both ribosome recycling factor (RRF) and release factor RF3. *Embo J* **1997,** *16* (13), 4134-41.

24. Doma, M. K.; Parker, R., Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **2006,** *440* (7083), 561-4.

25. Keiler, K. C., Mechanisms of ribosome rescue in bacteria. *Nat Rev Microbiol* **2015,** *13* (5), 285-297.

26. Pech, M.; Nierhaus, K. H., Three mechanisms in Escherichia coli rescue ribosomes stalled on non-stop mRNAs: one of them requires release factor 2. *Mol Microbiol* **2012,** *86* (1), 6-9.

27. Chadani, Y.; Ono, K.; Ozawa, S.; Takahashi, Y.; Takai, K.; Nanamiya, H.; Tozawa, Y.; Kutsukake, K.; Abo, T., Ribosome rescue by Escherichia coli ArfA (YhdL) in the absence of trans-translation system. *Mol Microbiol* **2010,** *78* (4), 796-808.

28. Chadani, Y.; Ono, K.; Kutsukake, K.; Abo, T., Escherichia coli YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. *Mol Microbiol* **2011,** *80* (3), 772-85.

29. Metzinger, L.; Hallier, M.; Felden, B., The highest affinity binding site of small protein B on transfer messenger RNA is outside the tRNA domain. *RNA* **2008,** *14* (9), 1761-72.

30. Miller, M. R.; Healey, D. W.; Robison, S. G.; Dewey, J. D.; Buskirk, A. R., The role of upstream sequences in selecting the reading frame on tmRNA. *BMC Biol* **2008,** *6*, 29.

31. Karzai, A. W.; Roche, E. D.; Sauer, R. T., The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol* **2000,** *7* (6), 449-55.

32. Keiler, K. C., Biology of trans-translation. *Annu Rev Microbiol* **2008,** *62*, 133-51.

33. Zechel, D. L.; Konermann, L.; Withers, S. G.; Douglas, D. J., Pre-steady state kinetic analysis of an enzymatic reaction monitored by time-resolved electrospray ionization mass spectrometry. *Biochemistry-Us* **1998,** *37* (21), 7664-9.

34. Arava, Y.; Wang, Y.; Storey, J. D.; Liu, C. L.; Brown, P. O.; Herschlag, D., Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **2003,** *100* (7), 3889-94.

35. Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R. S.; Weissman, J. S., Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **2009,** *324* (5924), 218-223.

36. Ingolia, N. T.; Lareau, L. F.; Weissman, J. S., Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **2011,** *147* (4), 789-802.

37. Michel, A. M.; Choudhury, K. R.; Firth, A. E.; Ingolia, N. T.; Atkins, J. F.; Baranov, P. V., Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* **2012,** *22* (11), 2219-2229.

38. Dunn, J. G.; Foo, C. K.; Belletier, N. G.; Gavis, E. R.; Weissman, J. S., Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *Elife* **2013,** *2*, e01179.

39. Guydosh, N. R.; Green, R., Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **2014,** *156* (5), 950-62.

40. Oh, E.; Becker, A. H.; Sandikci, A.; Huber, D.; Chaba, R.; Gloge, F.; Nichols, R. J.; Typas, A.; Gross, C. A.; Kramer, G.; Weissman, J. S.; Bukau, B., Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **2011,** *147* (6), 1295-308.

41. Li, G. W.; Oh, E.; Weissman, J. S., The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **2012,** *484* (7395), 538-U172.

42. Mohammad, F.; Woolstenhulme, C. J.; Green, R.; Buskirk, A. R., Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Rep* **2016,** *14* (4), 686-94.

43. Li, G. W.; Burkhardt, D.; Gross, C.; Weissman, J. S., Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **2014,** *157* (3), 624-35.

44. Ogura, T.; Hiraga, S., Mini-F Plasmid Genes That Couple Host-Cell Division to Plasmid Proliferation. *P Natl Acad Sci-Biol* **1983,** *80* (15), 4784-4788.

45. (a) Bernard, P.; Kezdy, K. E.; Vanmelderen, L.; Steyaert, J.; Wyns, L.; Pato, M. L.; Higgins, P. N.; Couturier, M., The F-Plasmid Ccdb Protein Induces Efficient Atp-Dependent DNA Cleavage by Gyrase. *J Mol Biol* **1993,** *234* (3), 534-541; (b) Bernard, P.; Couturier, M., Cell Killing by the F-Plasmid Ccdb Protein Involves Poisoning of DNA-Topoisomerase-Ii Complexes. *J Mol Biol* **1992,** *226* (3), 735-745.

46. Tam, J. E.; Kline, B. C., The F-Plasmid Ccd Autorepressor Is a Complex of Ccda and Ccdb Proteins. *Mol Gen Genet* **1989,** *219* (1-2), 26-32.

47. Jaffe, A.; Ogura, T.; Hiraga, S., Effects of the Ccd Function of the F-Plasmid on Bacterial-Growth. *J Bacteriol* **1985,** *163* (3), 841-849.

48. Gerdes, K., Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J Bacteriol* **2000,** *182* (3), 561-572.

49. Lehnherr, H.; Maguin, E.; Jafri, S.; Yarmolinsky, M. B., Plasmid Addiction Genes of Bacteriophage-P1 - Doc, Which Causes Cell-Death on Curing of Prophage, and Phd, Which Prevents Host Death When Prophage Is Retained. *J Mol Biol* **1993,** *233* (3), 414-428.

50. (a) Pedersen, K.; Zavialov, A. V.; Pavlov, M. Y.; Elf, J.; Gerdes, K.; Ehrenberg, M., The bacterial toxin RelE displays codon-specific cleavage of rnRNAs in the ribosomal A site. *Cell* **2003,** *112* (1), 131-140; (b) Christensen, S. K.; Mikkelsen, M.; Pedersen, K.; Gerdes, K., ReIE, a global inhibitor of translation, is activated during nutritional stress. *P Natl Acad Sci USA* **2001,** *98* (25), 14328-14333.

51. Neubauer, C.; Gao, Y. G.; Andersen, K. R.; Dunham, C. M.; Kelley, A. C.; Hentschel, J.; Gerdes, K.; Ramakrishnan, V.; Brodersen, D. E., The Structural Basis for mRNA Recognition and Cleavage by the Ribosome-Dependent Endonuclease ReIE. *Cell* **2009,** *139* (6), 1084-1095.

52. Hurley, J. M.; Cruz, J. W.; Ouyang, M.; Woychik, N. A., Bacterial Toxin RelE Mediates Frequent Codon-independent mRNA Cleavage from the 5 ' End of Coding Regions in Vivo. *J Biol Chem* **2011,** *286* (17), 14770-14778.

53. Gamper, H. B.; Masuda, I.; Frenkel-Morgenstern, M.; Hou, Y. M., The UGG Isoacceptor of tRNAPro Is Naturally Prone to Frameshifts. *Int J Mol Sci* **2015,** *16* (7), 14866-83.

54. Brule, H.; Elliott, M.; Redlak, M.; Zehner, Z. E.; Holmes, W. M., Isolation and characterization of the human tRNA-(N1G37) methyltransferase (TRM5) and comparison to the Escherichia coli TrmD protein. *Biochemistry-Us* **2004,** *43* (28), 9243-55.

55. Lewis, K., Persister cells, dormancy and infectious disease. *Nat Rev Microbiol* **2007,** *5* (1), 48-56.

56. Veening, J. W.; Smits, W. K.; Kuipers, O. P., Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Annu Rev Microbiol* **2008,** *62*, 193-210.

57. Hauryliuk, V.; Atkinson, G. C.; Murakami, K. S.; Tenson, T.; Gerdes, K., Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* **2015,** *13* (5), 298-309.

58. Christensen, S. K.; Maenhaut-Michel, G.; Mine, N.; Gottesman, S.; Gerdes, K.; Van Melderen, L., Overproduction of the Lon protease triggers inhibition of translation in Escherichia coli: involvement of the yefM-yoeB toxin-antitoxin system. *Mol Microbiol* **2004,** *51* (6), 1705-1717.

59. (a) Gotfredsen, M.; Gerdes, K., The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family. *Mol Microbiol* **1998,** *29* (4), 1065-76; (b) Christensen, S. K.; Gerdes, K., Delayed-relaxed response explained by hyperactivation of RelE. *Mol Microbiol* **2004,** *53* (2), 587-597.

60. Brzozowska, I.; Zielenkiewicz, U., Regulation of toxin-antitoxin systems by proteolysis. *Plasmid* **2013,** *70* (1), 33-41.

61. Gerdes, K.; Maisonneuve, E., Bacterial persistence and toxin-antitoxin loci. *Annu Rev Microbiol* **2012,** *66*, 103-23.

62. Maisonneuve, E.; Shakespeare, L. J.; Jorgensen, M. G.; Gerdes, K., Bacterial persistence by RNA endonucleases. *Proc Natl Acad Sci U S A* **2011,** *108* (32), 13206-11.

63. Griffin, M. A.; Davis, J. H.; Strobel, S. A., Bacterial Toxin RelE: A Highly Efficient Ribonuclease with Exquisite Substrate Specificity Using Atypical Catalytic Residues. *Biochemistry-Us* **2013,** *52* (48), 8633-8642.

64. Schifano, J. M.; Vvedenskaya, I. O.; Knoblauch, J. G.; Ouyang, M.; Nickels, B. E.; Woychik, N. A., An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nat Commun* **2014,** *5*.

65. Christensen, S. K.; Gerdes, K., RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol Microbiol* **2003,** *48* (5), 1389-400.

66. Woolstenhulme, C. J.; Guydosh, N. R.; Green, R.; Buskirk, A. R., High-Precision Analysis of Translational Pausing by Ribosome Profiling in Bacteria Lacking EFP. *Cell Rep* **2015,** *11* (1), 13-21.

67. (a) Nakatogawa, H.; Ito, K., The ribosomal exit tunnel functions as a discriminating gate. *Cell* **2002,** *108* (5), 629-636; (b) Gong, F.; Yanofsky, C., Instruction of translating ribosome by nascent peptide. *Science* **2002,** *297* (5588), 1864-1867.

68. Craigen, W. J.; Caskey, C. T., Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* **1986,** *322* (6076), 273-5.

69. Qian, Q.; Bjork, G. R., Structural requirements for the formation of 1-methylguanosine in vivo in tRNA(GGG)(Pro) of Salmonella typhimurium. *J Mol Biol* **1997,** *266* (2), 283-296.

70. (a) Holmes, W. M.; Andraos-Selim, C.; Redlak, M., tRNA-m1G methyltransferase interactions: touching bases with structure. *Biochimie* **1995,** *77* (1-2), 62-5; (b) Bjork, G. R.; Wikstrom, P. M.; Bystrom, A. S., Prevention of Translational Frameshifting by the Modified Nucleoside 1-Methylguanosine. *Science* **1989,** *244* (4907), 986-989; (c) Bystrom, A. S.; Bjork, G. R., Chromosomal location and cloning of the gene (trmD) responsible for the synthesis of tRNA (m1G) methyltransferase in Escherichia coli K-12. *Mol Gen Genet* **1982,** *188* (3), 440-6.

71. Bystrom, A. S.; Bjork, G. R., The structural gene (trmD) for the tRNA(m1G)methyltransferase is part of a four polypeptide operon in Escherichia coli K-12. *Mol Gen Genet* **1982,** *188* (3), 447-54.

72. Christian, T.; Gamper, H.; Hou, Y. M., Conservation of structure and mechanism by Trm5 enzymes. *Rna-a Publication of the Rna Society* **2013,** *19* (9), 1192-1199.

73. Li, Z. W.; Deutscher, M. P., Maturation pathways for E-coli tRNA precursors: A random multienzyme process in vivo. *Cell* **1996,** *86* (3), 503-512.

74. Frank, D. N.; Pace, N. R., Ribonuclease P: Unity and diversity in a tRNA processing ribozyme. *Annual Review of Biochemistry* **1998,** *67*, 153-180.

75. Li, Z. W.; Deutscher, M. P., RNase E plays an essential role in the maturation of Escherichia coli tRNA precursors. *Rna-a Publication of the Rna Society* **2002,** *8* (1), 97-109.

76. Perona, J. J.; Rould, M. A.; Steitz, T. A., Structural basis for transfer RNA aminoacylation by Escherichia coli glutaminyl-tRNA synthetase. *Biochemistry-Us* **1993,** *32* (34), 8758-71.

77. Rould, M. A.; Perona, J. J.; Steitz, T. A., Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **1991,** *352* (6332), 213-8.

78. (a) Charette, M.; Gray, M. W., Pseudouridine in RNA: What, where, how, and why. *Iubmb Life* **2000,** *49* (5), 341-351; (b) Davis, F. F.; Allen, F. W., Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem* **1957,** *227* (2), 907-15.

79. Dalluge, J. J.; Hashizume, T.; Sopchik, A. E.; McCloskey, J. A.; Davis, D. R., Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Res* **1996,** *24* (6), 1073-1079.

80. Urbonavicius, J.; Qian, Q.; Durand, J. M.; Hagervall, T. G.; Bjork, G. R., Improvement of reading frame maintenance is a common function for several tRNA modifications. *Embo J* **2001,** *20* (17), 4863-73.

81. Nasvall, S. J.; Chen, P.; Bjork, G. R., The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo. *RNA* **2004,** *10* (10), 1662-73.

82. Gamper, H. B.; Masuda, I.; Frenkel-Morgenstern, M.; Hou, Y. M., Maintenance of protein synthesis reading frame by EF-P and m(1)G37-tRNA. *Nat Commun* **2015,** *6*, 7226.

83. Bregeon, D.; Colot, V.; Radman, M.; Taddei, F., Translational misreading: a tRNA modification counteracts a+2 ribosomal frameshift. *Gene Dev* **2001,** *15* (17), 2295-2306.

84. Christian, T.; Evilia, C.; Williams, S.; Hou, Y. M., Distinct origins of tRNA(m1G37) methyltransferase. *J Mol Biol* **2004,** *339* (4), 707-19.

85. Ito, T.; Masuda, I.; Yoshida, K.; Goto-Ito, S.; Sekine, S.; Suh, S. W.; Hou, Y. M.; Yokoyama, S., Structural basis for methyl-donor-dependent and sequence-specific binding to tRNA substrates by knotted methyltransferase TrmD. *Proc Natl Acad Sci U S A* **2015,** *112* (31), E4197-205.

86. Elkins, P. A.; Watts, J. M.; Zalacain, M.; van Thiel, A.; Vitazka, P. R.; Redlak, M.; Andraos-Selim, C.; Rastinejad, F.; Holmes, W. M., Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J Mol Biol* **2003,** *333* (5), 931-949.

87. O'Dwyer, K.; Watts, J. M.; Biswas, S.; Ambrad, J.; Barber, M.; Brule, H.; Petit, C.; Holmes, D. J.; Zalacain, M.; Holmes, W. M., Characterization of Streptococcus pneumoniae TrmD, a tRNA methyltransferase essential for growth. *J Bacteriol* **2004,** *186* (8), 2346-54.

88. Christian, T.; Hou, Y. M., Distinct determinants of tRNA recognition by the TrmD and Trm5 methyl transferases. *J Mol Biol* **2007,** *373* (3), 623-32.

89. Goto-Ito, S.; Ito, T.; Ishii, R.; Muto, Y.; Bessho, Y.; Yokoyama, S., Crystal structure of archaeal tRNA(m(1)G37)methyltransferase aTrm5. *Proteins* **2008,** *72* (4), 1274-89.

90. Carr, A. C.; Taylor, K. L.; Osborne, M. S.; Belous, B. T.; Myerson, J. P.; Moore, S. D., Rapid depletion of target proteins allows identification of coincident physiological responses. *J Bacteriol* **2012,** *194* (21), 5932-40.

91. Gottesman, S.; Roche, E.; Zhou, Y.; Sauer, R. T., The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **1998,** *12* (9), 1338-47.