ROLE OF CONSERVED RESIDUES FOR RNA BINDING AND CATALYSIS IN TWO BACTERIAL PSEUDOURIDINE SYNTHASES

HOPE M. VIENNEAU
Bachelor of Science, University of Lethbridge, 2018

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

MASTER OF SCIENCE

in

BIOCHEMISTRY

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

ROLE OF CONSERVED RESIDUES FOR RNA BINDING AND CATALYSIS IN TWO BACTERIAL PSEUDOURIDINE SYNTHASES

HOPE M. VIENNEAU

Date of Defense: June 18, 2021

Dr. U. Kothe Thesis Supervisor	Professor	Ph.D.
Dr. H.J. Wieden Thesis Examination Committee Member	Professor	Ph.D.
Dr. T. Patel Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. M. Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

ABSTRACT

Pseudouridine synthases are RNA modification enzymes targeting various forms of RNA in the cell. RluB and RluF are pseudouridine synthases modifying adjacent uridines in *E. coli* ribosomal RNA. These enzymes have a highly conserved RNA-binding loop, distinct between RluB and RluF, proposed to play a role in their specificity. This thesis provides preliminary insight into the importance of the eight-residue binding loops of RluB and RluF for RNA target recognition. I have shown that altering the residues of the binding loop decreases the enzyme's affinity for RNA and effectively abolish the enzyme's pseudouridylation activity. I conclude that the binding loops of RluB and RluF are critical for RNA binding and for positioning the target RNA for catalysis. Better understanding the requirements for specificity and target recognition of RluB and RluF will enhance our knowledge on ribosome formation and may aid in engineering new enzymes to pseudouridylate other functional RNAs.

ACKNOWLEDGEMENTS

First and foremost, I would like to send a heartfelt thank you to my long-time supervisor, Dr. Ute Kothe. She sparked a love of science and research by recruiting me as a young undergraduate student and has pushed me to develop into the scientist I am today. You have mentored me in a personal setting in addition to an academic one, and your endless patience and support have made this possible.

I would also like to thank my committee members, Dr. Hans-Joachim Wieden and Dr. Trushar Patel, for attending my committee meetings to provide advice and ideas on my thesis work throughout my studies.

A special thank you goes out to the Alberta RNA Research and Training Institute at the University of Lethbridge; you are the best work family I could ask for. In particular, I would like to thank Sarah Schultz for taking the time to edit my introduction, Emily Soon for being my original supervisor as an undergraduate student, Damian La Rosa Montes for assistance in mutagenesis and protein purifications, as well as Dr. Julia Guegueniat for providing unwavering support and teaching me a variety of techniques.

Lastly, a sincere thank you to my family and friends, who have listened to me excitedly chatter on about things they didn't understand, who provide endless support when I am stressed out, and who are here to celebrate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

A site	Acceptor Site
β-ME	Beta-mercaptoethanol
BSA	Bovine Serum Albumin
DEAE	Aspartic acid, Glutamic acid, Alanine, Glutamic Acid resin
DC	Decoding center
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DPM	Disintegrations Per Minute
DTT	Dithiothreitol
Dus	Dihydrouridine synthase
E site	Exit site
EDTA	Ethylenediaminetetraacetic acid
GMP	Guanosine monophosphate
hoC	Hydroxycytidine
IDT	Integrated DNA Technologies
iPPase	Inorganic pyrophosphatase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
LB	Lysogeny broth
lncRNA	Long non-coding RNA
lincRNA	Long intergenic non-coding RNA
mRNA	Messenger RNA
miRNA	Micro RNA
misc_RNA	Miscellaneous RNA
mtRNA	Mitochondrial RNA
ncRNA	Noncoding RNA
nt	Nucleotide

NTP	Nucleotide triphosphate
OD	Optical density
P site	Peptidyl site
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	Phenyl-methyl-sulfonyl fluoride
PTC	Peptidyltransferase center
Pus	Pseudouridine synthase
Rlh	Large subunit ribosomal RNA hydroxylation
Rlm	Large subunit ribosomal RNA methylation
Rlu	Large subunit ribosomal RNA pseudouridylation
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal RNA
Rsu	Ribosomal small subunit pseudouridylation
Rsm	Ribosomal small subunit methylation
SAM	S-Adenosyl-L-Methionine
scaRNAs	Small Cajal body-specific RNA
SDS	Sodium dodecyl sulfate
SEC	Size Exclusion Chromatography
snoRNA	Small nucleolar RNA
snoRNP	Small nucleolar ribonucleoproteins
SOC	Super Optimal broth with Catabolite repression
SPOUT	SpoU-TrmD
sRNP	Small ribonucleoprotein
TAE	Tris base, Acetic Acid, EDTA
tRNA	Transfer RNA
Tru	tRNA pseudouridine synthase

UTP	Uridine triphosphate
WT	Wildtype

CHAPTER 1: INTRODUCTION

1.1 General Introduction to rRNA Modifications

Transcription and translation are crucial for cell survival and are highly conserved across all three domains of life. One key requirement of protein synthesis is the production of ribosomes. The prokaryotic ribosome is a complex 70S ribonucleoprotein complex comprised of two subunits: a large 50S subunit, and a small 30S subunit. Eubacterial ribosomes are comprised of three rRNA molecules (23S, 5S, and 16S rRNA) complexed with more than fifty proteins in a molecular volume ratio of two-thirds rRNA to one-third proteins [1, 2]. The large subunit contains the 23S and 5S rRNA with over thirty proteins and catalyzes the formation of peptide bonds in a location known as the peptidyl transferase center (PTC) [3]. The small subunit contains the 16S rRNA in complex with twenty proteins and binds the messenger RNA (mRNA) to facilitate decoding of the nucleotide sequence in the decoding center (DC).

Each subunit is further divided into domains, which help describe the primary and secondary structures of the ribosome (Figure 1.3). The large subunit has six domains, numbered I-VI, with the 5S rRNA adding on a seventh domain. Similarly, the small subunit folds into four domains: the 5′, central, 3′ major, and 3′ minor domains [4]. When the subunits come together, they form the primary binding sites for tRNAs: the acceptor (A) site, peptidyl (P) site, and the exit (E) site [3]. The subunits are held together by inter-subunit bridges formed by RNA-RNA, RNA-protein, and protein-protein contacts [3, 5, 6]. The ribosome is assembled in a highly regulated process known as ribosome biogenesis, which is centered around ribonucleic acid (RNA) maturation and folding.

Ribosome biogenesis is a complex, highly conserved process in bacteria. Biogenesis begins with the transcription of the 5S, 16S, and 23S rRNAs as a single, primary transcript, and the

hierarchy of the maturation process begins prior to completion of transcription [4]. As the primary transcript matures, it is folded into its secondary and tertiary structure and ribosomal proteins begin binding in a cooperative manner [7, 8]. Protein binding has been shown to have a chaperone effect on the RNA, helping to ensure the correct folding of such a large transcript through stabilization, re-orientation, and unwinding of the rRNA [9]. In addition to ribosomal proteins, many other assembly factors are recruited, including various assembly proteins [9, 10], GTPases [11], rRNA modification enzymes, helicases, and other factors [10, 12].

Cleavage of the primary transcript is catalyzed by RNase III, which creates precursor 5S, 16S, and 23S rRNAs. These transcripts are then further processed in a cascade by other RNases, not all of which are known. [4]. This process is regulated according to the cellular environment, therefore, biogenesis is closely linked to growth rate: cells actively dividing will up-regulate ribosome biogenesis, while cells undergoing starvation will down-regulate ribosome formation [9, 12]. Ribosome production and maintenance accounts for the majority of a cell's energy, making ribosome synthesis and processing a key region of research [8, 12]. Many studies have been conducted on determining when in the biogenesis pathway the rRNA modifications are introduced. Typically, such studies involve the isolation of assembly intermediates, and identification of modifications via mass spectrometry [9, 12-16].

rRNA is extensively modified co- and post-transcriptionally *via* base methylation, 2′-O-methylation, uridine isomerization, and other types of chemical alterations [17-19]. These modifications lead to altered steric and electronic effects, ultimately changing the base-pairing, sugar puckers, and base stacking properties of the RNA [19, 20]. The majority of the 23S rRNA modifications occur in the early stages of ribosome biogenesis of the 50S subunit, while the majority of 16S rRNA modifications occur later in the assembly of the 30S subunit [18].

Escherichia coli is the primary model organism of eubacteria [21], and the 36 modifications present in its rRNA have been mapped. Of these, 23 are methylations, 10 are pseudouridines, 1 is a methylated pseudouridine, 1 is a dihydrouridine, and 1 is a hydroxycytidine (Figure 1.1**)** [18]. The 16S rRNA contains 11 of these modifications (10 methylations and 1 pseudouridine), while the 23S rRNA contains the remaining 25 of the 36 modifications (13 methylations, 9 pseudouridines, 1 methylated pseudouridine, 1 dihydrouridine, and 1 hydroxycytidine). Interestingly, the 5S rRNA of *E. coli* contains no known modifications [18]. Many of the rRNA modifications are nearly universally conserved among bacteria, and some conservation extends into eukaryotes and archaea [22, 23]. Such high levels of conservation are indicative of a significant role and potential advantage(s) for organisms containing these rRNA modifications. Only in two cases are severe consequences for the cell observed when the RNA modifying enzymes are deleted: RluD (Ψ1911, Ψ1915, Ψ1917) and RlmE (Um2552) in 23S rRNA (see sections 1.1.5.2 and 1.6.1.1) [24, 25]. Seven 23S rRNA modifications were found to be functionally significant in the reconstruction of active 50S subunits (m^2G2445, D2449, ψ2457, Cm2498, oh^5C2501, m^2A2503, ψ2504,), while none of the 16S rRNA modifications are essential for *in vitro* reconstituting small subunits of the ribosome [18, 26]. Nearly all the enzymes responsible for the rRNA modifications have been identified in bacteria (excluding dihydrouridine), and knockout studies have been performed for the majority of the known enzymes. The functional role(s) of these individual modifications, however, remain poorly understood. Although no significant growth defects of knockout strains are observed under optimal conditions, the absence of even a single rRNA modification leads to a fitness disadvantage and out-competition in a competitive growth environment for many modification enzyme deletions [18]. Thus, it could be argued that these modifications evolved to protect the

bacterial ribosome against various environmental stressors, which may have provided an advantage during natural selection. Many of these rRNA modifications also have widespread effects on the translation capacity of the ribosome, proving their functional value and reason for widespread conservation [19].

rRNA is extensively modified regardless of the organism as these modifications are crucial for protein synthesis. However, the contribution to the ribosome and the number of modifications differs: the more complex an organism is, the more rRNA modifications it has [2]. A common hypothesis is that rRNA structure and function have been fine-tuned and are modulated through these modifications to respond to relevant environmental factors [22]. Information on the conservation of the modifications is conflicting. A potential reason for this discordance arises from the list of organisms used to evaluate conservation: if one source is more extensive and covers a wider range of organisms than the other, it may lower the percent conservation of modification. For example, Baldridge [22] calculated the percent conservation data from MODOMICS by dividing the number of organisms with a given modification at a location by the total number of organisms shown in the RNA sequence alignment [27].

1.2 Methylation

Methylations are a diverse group of modifications found in rRNA and transfer RNA (tRNA) in both pro- and eukaryotes, while in eukaryotes methylations are also present in small nucleolar RNA (snRNA), micro-RNA (miRNA), and long non-coding RNA (lncRNA) [28-30]. In rRNA, methylations are most commonly found in functionally important regions of the ribosome, such as the PTC (Figure 1.3 A) and DC [28, 31, 32]. In *E. coli*, rRNA methylations are present as 2′-O-methylations and base methylations, including a methylated pseudouridine (Figure 1.2).

Classifying methyltransferases into families is difficult as their modification substrates are widespread, acting on both DNA and RNA.

Notably, nearly most of these enzyme families are *S*-adenosyl-L-Methionine (Ado-Met or SAM) dependent, as SAM acts as the methyl donor for modification. However, a small quantity of methyltransferases use N5, N10-methylenetetrahydrofolate (CH_2 = THF) as their methyl donor [33]**.** SAM-dependent methyltransferases have low sequence similarities but share many common structural features. For example, they are composed of a common catalytic domain fold, an α/β fold comparable to a Rossman fold, and a shared sequence motif of the SAM binding site [34]. The SPOUT superfamily is only present in prokaryotes, modifying both tRNA and rRNA. This superfamily shares three core motifs, named motifs 1, 2, and 3 [35]. These motifs are involved in the binding, stabilization, catalysis, and release of the target RNA product. Notably, the function of residues among SPOUT families differs greatly, as not even the catalytic residue is conserved among families [35].

1.2.1 2′-O-methylation

The 2′-O-methylation of the ribose is one of the most common and widespread types of rRNA modification, with many positions being evolutionarily conserved between bacteria and eukaryotes/archaea [29]. While 2′-O-methylations are abundant in eukaryotes, these modifications are relatively rare in prokaryotes. This difference arises from the modification machinery: eukaryotes use box C/D small nucleolar ribonucleoproteins (snoRNPs) to direct methylation, while bacteria require a stand-alone enzyme for each modification site [28, 36]. *E. coli* contains 3 of these modifications in its 23S rRNA: Gm2251, Cm2498, and Um2552, all located in the PTC [28]. These modifications are mediated by the enzymes RlmB [37], RlmE

[38], and RlmM [39], respectively (Table 1.1). The 2′-O-methylations of *E. coli* are generally considered to be ubiquitous in viruses, archaea, eubacteria, yeast, protists, and fungi [28, 40].

Table 1.1 – A Comprehensive summary of the rRNA modifications in the *Escherichia coli*

Enzyme	Modification	Location	Stage of Assembly	% Conser-vation of Mod. *	High Resolution Structure (PDB ID)	Ref.
Methylations:						
RlmA	m^1G745	*23S rRNA*	Early	L	1P91	[41]
RlmB	Gm2251		N.D.	M	1GZ0	[37]
RlmC	m^5U747		N.D.	L	N.D.	[42]
RlmD	m^5U1939		Early	M	1UWV	[42, 43]
RlmE	Um2552		Late	M	1EIZ	[38]
RlmF	m^6A1618		Intermediate	L	N.D.	[44]
RlmG	m^2G1835		Early	L	4DCM	[45]
RlmH	$m^3\Psi1915$		Late	M	1NS5	[46, 47]
RlmI	m^5C1962		Early	L	3C0K	[48]
RlmJ	m^6A2030		Early	L	4BLU	[49]
RlmKL	m^2G2445 and m^7G2069		Early	H	3V8V	[50, 51]
RlmM	Cm2498		Early	L	4ATN	[39]
RlmN	m^2A2503		Early	L	3RF9	[52]
KsgA (RsmA)	$m^6{}_2A1518$ and $m^6{}_2A151$	*16S rRNA*	Late	H	1QYR	[53]
RsmB	m^5C967		Late	M	1J4F	[54]
RsmC	m^2G1207		Late	L	2PJD	[55]
RsmD	m^2G966		Early	H	2FPO	[56]
RsmE	m^3U1498		Late	M	4E8B	[57]
RsmF	m^5C1407		Late	M	2FRX	[58]
RsmG	m^7G527		Late	L	1JSX	[42, 59]
RsmH	$m^4Cm1402$		Late	H	3TKA	[32]
RsmI	Cm1402		Late	H	5HW4	
RsmJ	m^2G1516		Late	L	N.D.	[60]
Pseudouridylations:						
RluA	Ψ746	*23S rRNA*	Early	N.D.	2I82	[17]
RluB	Ψ2605		Early	N.D.	4LAB	[61]
RluC	Ψ955, Ψ2504, Ψ2580		Early	N.D.	1V9K	[62]
RluD	Ψ1911, Ψ1915, Ψ1917		Late	N.D.	1PRZ	[63, 64]

RluE	Ψ2457		Early	N.D.	2OLW	[61]
RluF	Ψ2604		Early	N.D.	2GML	[65]
RsuA	Ψ516	16S rRNA	Intermediate	N.D.	1KSK	[66]
Other Modifications:						
DusB?	D2449	23S rRNA	N.D.	N.D.	N.D.	[67]
RlhA	ho^{5}C2501		N.D.	N.D.	N.D.	[2, 68]

***** H = High conservation (>75%), M = Medium conservation (40-75%), L = low concentration (<40%), N.D. = Not determined. The percent conservation is estimated based on the available rRNA sequences in MODomics; the number of organisms with a modification at a given location divided by the total number of organisms shown in the RNA sequence alignment. [22, 27]

In bacterial and archaeal hyperthermophiles, there is a paralleled tendency in ribosomal RNA to contain increased levels of 2′-O-methylations proportional to the environmental temperature implying a potential role for rRNA stabilization under stress [28]. The 2′-O-methylations in the *E. coli* ribosome are clustered around the functionally relevant domain of the PTC (located at positions 2251, 2498, and 2552), and may increase the number of functioning ribosomes by increasing their stability, thereby influencing the environmental stress response of a cell (Figure 1.3-A).

The nucleotide U2552 is one of the conserved modifications in the ribosomal A site, being methylated in the majority of living organisms (Figure 1.3-A) [18]. Knockout strains of *rlmE* (which catalyzes formation of Um2552) exhibit a general growth phenotype at a variety of temperatures, a three-fold slower heat shock response, an accumulation of ribosome intermediates, and a reduced number of 70S ribosomes [38]. The knockout strain of *rlmE* was also found to have a decreased translational efficiency *in vivo* [18, 25]. RlmE is upregulated four-fold under heat stress, implying it is key in maintaining correct rRNA conformations near the subunit bridges at elevated temperatures [22, 38]. Additionally, Um2552 adopts the same

conformation in the *E.* coli 70S ribosome as it does in the *Haloarcula marismortui* 50S subunit, having the 2′-O-methyl group wedged between the nucleobases U2554 and G2552, which pair with the A-site tRNA [19]. Based on these results, RlmE is proposed to have functional importance in modulating base pairing between 23S rRNA and A-site tRNA.

Bacterial knockout strains of *rlmB* and *rlmM* (lacking Gm2251 and Cm2498) do not have a defined growth phenotype [21, 37]. In a competition assay against the wildtype strain, *rlmB* strains were not outcompeted [37], while *rlmM* strains were [39].

The 16S rRNA of *E. coli* contains a single 2′-O-methylation, m^4Cm1402, which will be discussed under the "Multiple Modifications" subheading (section 1.2.6).

1.2.2 Adenine (A) methylation

The 23S rRNA contains three adenosine base methylations: m^2A2503, m^6A1618, and m^6A2030 (Table 1.1). The modification m^2A2503 is found in the PTC in Domain V of the 23S rRNA at the entrance of the exit tunnel; it is mediated by the methyltransferase RlmN and is the only adenosine methylated at a carbon atom in *E. coli* rRNA [18, 52]. RlmN acts early in ribosome biogenesis, as its modification site is close to the catalytic center of the PTC which is not accessible in the mature 50S subunit (Figure 1.3-A). However, this enzyme has a dual specificity, also catalyzing the m(2)A modification at position 37 in a tRNA [69]. Knockouts of *rlmN* produce an error-prone translational phenotype by increasing the misreading of a UAG stop codon [69]. The PTC is primarily responsible for the formation of peptide bonds in protein synthesis, however, it has also been found to be involved in discriminating between near-cognate tRNAs via the acceptor end of the tRNA [70]. Based on this information, RlmN has been hypothesized to play a crucial role in the proofreading of translation at the PTC, most likely in optimizing ribosomal fidelity via a conformational change in the PTC [69]. Inactivation of *rlmN*

also leads to increased susceptibility to the antibiotics tiamulin, hygromycin A and sparsomycin [71]. Conversely, hypermethylation of A2503 yields antibiotic resistance to antibiotic targeting the PTC, such as those mentioned above. Finally, knockout strains of *rlmN* were out-competed in a competition assay with a wildtype strain, implying a relevant functional role for the modification for bacterial fitness [71]. The methylation alters the chemical reactivity of the residue (Figure 1.2) and may act to maintain the functional conformation of the PTC under stress or affect interactions of nascent regulatory peptides near the exit tunnel.

The methylation m^6A1618 is generated by RlmF in Domain III of 23S rRNA during intermediate stages of ribosome assembly (Table 1.1) [44]. Inactivation of *rlmF* eliminates formation of m^6A1618 and leads to moderate growth retardation, indicating the methylation of A1618 yields a selective growth advantage [44]. Comparatively, the modification m^6A2030 in Domain V in the large ribosomal subunit is generated by RlmJ [49]. RlmJ preferentially modifies de-proteinized 23S rRNA in early ribosome biogenesis (Table 1.1). Knocking out the *rlmJ* gene causes no noticeable effects on cellular fitness when observed under starvation or temperature stressors, although there is a slight growth advantage conveyed to knockout strains when grown in anaerobic conditions [49].

The 16S rRNA of *E. coli* contains two adjacent A methylations in the helix 45 of the minor domain: $m^6{}_2A1518$ and $m^6{}_2A1519$ (Table 1.1) [18]. The neighboring methylations are responsible for the correct packing of Helices 24a, 44, and 45 of 16S rRNA (Figure 1.3 B) [72]. These modifications are formed by the methyltransferase KsgA, which is more commonly referred to as RsmA [53]. RsmA acts late in the process of ribosome biogenesis and has been proposed to signal the end of subunit assembly [73]. As such, RsmA plays a crucial role in ribosome biogenesis: the presence of the enzyme in ribosome assembly has been found to have a

larger impact on cell viability than its methyltransferase activity [22, 74, 75]. RsmA binds to the 30S subunit near the decoding site while in an inactive conformation, hypothetically functioning as a checkpoint in ribosome assembly by blocking incomplete 30S subunits from assembling and beginning translation [22, 53]. When the subunit assembly is complete, the enzyme catalyzes methylation of A1518 and A1519 to release RsmA from the small subunit and allow binding of tRNA, mRNA, the 50S subunit, and other translational factors [73, 76].

1.2.3 Cytosine (C) methylation

The singular cytosine methylation found in the 23S rRNA of *E. coli* is m^5C1962 (Table 1.1) in Domain V of 23S rRNA in the large ribosomal subunit near the exit tunnel and is conserved in bacteria [39]. This methylation is generated by the methyltransferase RlmI, which preferentially modifies de-proteinized 23S rRNA in early stages of ribosome biogenesis [39]. Knockout strains of *rlmI* have a growth disadvantage in both rich and minimal media; due to the phenotype presence independent of nutrient content, the methylation of C1962 is not believed to play a specific role in the starvation stress response of ribosomes [48].

The small ribosomal subunit contains two cytosine methylations (Table 1.1). The first is m^5C1407: a modification in the 3′ minor Domain of 16S rRNA resulting from activity of the methyltransferase RsmF [58]. It is incorporated during intermediate stages of ribosome assembly and preferentially modifies a structured 30S subunit [58]. This is contrary to other 16S rRNA methyltransferases, which seemingly prefer free 16S rRNA as a substrate [18, 58]. Deletion of the *rsmF* gene (and thus the modification) yields growth retardation and reduced cellular fitness in competition assays against cells containing the wildtype enzyme [18, 58].

The second cytosine methylation is m^5C967, which is part of a pair of adjacent methylations in helix 31, in proximity to the tRNA anticodon bound to the P site [54]. m^5C967 is

generated by the methyltransferase RsmB, which preferentially modifies free 16S rRNA as a substrate in early stages of ribosome biogenesis [54, 77].

1.2.4 Guanine (G) methylation

There are eight guanosine methylations in the 70S ribosome: four in the large subunit and four in the small subunit (Table 1.1). m^1G745, found in Domain II of 23S rRNA in the large ribosomal subunit near the exit tunnel, is conserved in all Gram-negative bacteria [41]. This modification is generated by the methyltransferase RlmA, which preferentially modifies rRNA during early stages of ribosome assembly [41, 78]. Deleting the methyltransferase RlmA results in a decreased translation efficiency and peptide chain elongation rate, as well as a growth phenotype under normal conditions [41]. Knockouts of *rlmA* alter the distribution of ribosomal assembly intermediates, increasing the proportions of free 30S and 50S subunits [41]. Based on the data collected by Gustafsson and Persson, the absence of m^1G745 is believed to dissociate loosely coupled 70S ribosomes, thereby suggesting RlmA plays a role in subunit association [22, 41]. It is proposed that RlmA plays an additional role as an rRNA chaperone during assembly of the 50S subunit [78].

The methyltransferase RlmG preferentially acts upon de-proteinized 23S rRNA during early stages of ribosome assembly to generate m^2G1835 [45]. This modification is found in Domain IV of 23S rRNA in the 50S subunit. Knockout strains of *rlmG* have no substantial impact on cell growth under normal conditions but are outcompeted by the wildtype strain in a competition assay, implying a fitness disadvantage of cells lacking the modification [45]. RlmG is differentially expressed under cold, heat, and oxidative stress (Figure 1.1 B): the methyltransferase is upregulated two-fold under cold stress, upregulated two-fold under heat stress, and upregulated four-fold under oxidative stress [79]. The upregulation under temperature

stressors bears similarity to trends observed with RsmJ, suggesting that the enzyme influences subunit association directly. Furthermore, Osterman and colleagues [79] found methylation of G1835 enhances the association of ribosomal subunits, possibly by aiding in the alignment of subunits during association explaining the upregulation of RlmG under stressors.

The methyltransferase RlmKL is responsible for two methylations in free 23S rRNA: m^2G2445 and m^7G2069 in the PTC of Domain V [51]. This enzyme preferentially modifies 23S rRNA in the absence of ribosomal proteins and co-sediments with the 50S assembly intermediate, thus, RlmKL acts early in stages of ribosome biogenesis [50]. Knockouts of *rlmKL* are outcompeted in a competition assay against cells possessing the wildtype enzyme, and a slight growth defect is observed under optimal conditions [51].These knockout strains also have deficits in 50S ribosome subunit assembly (Figure 1.1 B); a phenotype unsurprising when one considers this methyltransferase co-sediments with an assembly intermediate [50].

The methyltransferase RsmC yields the modification m^2G1207 in the 3′ major domain of the small ribosomal subunit during the late stages of ribosome biogenesis, recognizing only 30S particles including ribosomal proteins [54]. G1207 is in the stem of helix 34, which is linked to the decoding site of the ribosome and has been shown to play a role in recognizing termination codons [80]. Characterization of phenotypes arising from the absence of *rsmC* in knockout studies have yet to be conducted. Interestingly, RsmC has been confirmed to have chaperone activity during the biogenesis of the 30S subunit [81].

The methylated m^2G966 is located in helix 31 near the P site of the small ribosomal subunit [18]. This modification is generated by the methyltransferase RsmD [56], and is the corresponding partner of the pair of adjacent methylations (the partner being m^5C967, generated by RsmB) [54].Weitzman *et al.* found binding of ribosomal proteins to rRNA to be inhibitory to

RsmD activity and the methyltransferase had a preference for free 16S rRNA; as such, RsmD was concluded to act early in ribosome biogenesis [77]. Knockouts of *rsmD* are cold sensitive but have no observable growth defects under optimal growth conditions [82]. RsmD deficiency increases translational initiation with AUA codon, and the effect is increased when the S9 tail is deleted in tandem [82]. Similar trends are observed with depletion of RsmB and the S9 tail. As such, the methylations present at G966 and C967 may play a role in stabilizing tRNAs in the P site of the ribosome.

The methyltransferase RsmG generates the m^7G527 modification in the 5′ Domain of the small ribosomal subunit [83]. RsmG preferentially modifies the 30S subunit during the late stages of ribosome assembly [84]. Knockouts of *rsmG* (eliminating m^7G527) result in read-through translational errors [59, 85], as well as growth disadvantages in the presence of streptomycin and low concentrations of chloramphenicol (Figure 1.1 B) [59]. The absence of the enzyme also conveys multiple unexpected growth advantages in *Bacillus subtilis*, such as growth in the presence of EDTA or bicyclomycin, as well as in the absence of a nitrogen source (except ammonium chloride) [83].

The modification m^2G1516 is created by RsmJ and is found in the small subunit 3′ minor domain [60]. RsmJ preferentially modifies an assembled 30S subunit in late stages of ribosome biogenesis. The gene for RsmJ is found in a heat shock operon, strongly implying a role in the environmental stress response [22, 60]. Deletion of *rsmJ* yields a cold-sensitive phenotype and alters the cellular growth rate; however, it does not influence the distribution of ribosome assembly intermediates (Figure 1.1 B) [22, 60]. Based on this, it is likely that RsmJ decreases non-specific interactions between the subunits in conjunction with RlmG and ensures proper subunit alignment through m^2G1516 (Figure 1.1 B) [22].

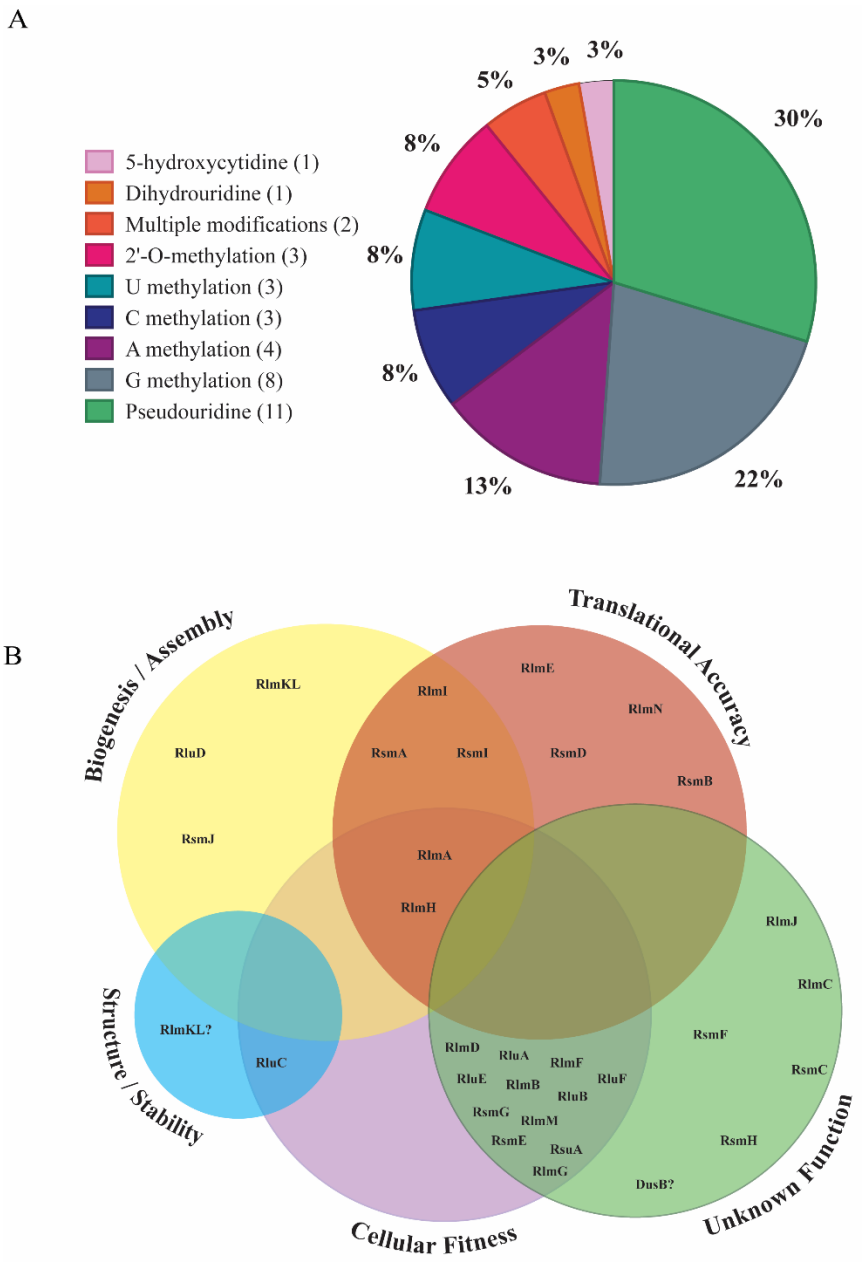


Figure 1.1- Overview of rRNA modifications and their respective functions. (A) Distribution of the 36 rRNA modification types in the *Escherichia coli* 70S ribosome. All modifications to the 23S and 16S rRNA are grouped, and the exact number of a given modification present is listed in brackets following the modification type. If a nucleotide contains more than one modification type, it is considered under "multiple modifications" **(B)** Venn diagram of rRNA modification enzymes in *E. coli* and their respective functions. Modification function/impact is divided into the following categories: structure & stabilization, translational accuracy/initiation, cellular fitness, ribosome biogenesis & assembly, and unknown function. Many modifications result in both a cellular fitness role, in addition to a specific role responsible for the fitness advantage.

1.2.5 Uracil (U) methylation

The methylated uridine, m^5U747, is generated by RlmC [42] in Domain II of the large ribosomal subunit, in a conserved area forming the exit tunnel for the newly synthesized protein [18]. The substrate specificity of this methyltransferase is unknown (Table 1.1); thus, we do not know when in acts in the ribosome biogenesis pathway. Considering the modification is located in a restricted region of the ribosome, one may hypothesize this methyltransferase must act early in ribosome biogenesis. Knockout studies of *rlmC* have shown no growth disadvantage in rich or minimal media [42], but characterization of its role under stress has yet to be investigated.

RlmD is responsible for methylating U1939 to generate m^5U1939 in free 23S rRNA (Table 1.1) [42, 43]. U1939 is in a conserved region of the 23S rRNA of *E. coli*, interacting with the aminoacyl end of tRNAs in the A site of Domain IV [18]. RlmD was found to recognize a conserved 37-nucleotide RNA with the same sequence as a 23S rRNA fragment [86]. An iron-sulfur cluster was discovered within RlmD [87]; the cluster is readily oxidized by ferricyanide and nitric oxide and hypothesized to be readily oxidized under physiological conditions. Oxidation ultimately leads to the decomposition of the cluster and precipitation of RlmD, which Argarwalla [87] has proposed regulates the activity of RlmD. Since methyltransferases do not require a redox step [87], the hypothesis that cells possessing m^5U1939 could readily oxidize the iron to provide a selective advantage under conditions of oxidative stress arose.

The m^3U1498 is present in the 16S rRNA of *E. coli* and is generated by RsmE in the 3′ minor domain of the small subunit [57]. RsmE preferentially acts on an assembled 30S subunit late in the ribosome biogenesis pathway, and residues involved in cofactor and substrate binding are highly conserved across several plant and eubacterial species (Table 1.1) [57, 88]. Knockout studies of *rsmE* have revealed a moderate growth phenotype and that the knockout strain is

outcompeted by the wildtype strain in competition assays [88]. Removal of the methyltransferase, and thus its corresponding modification on U1948 influence the accuracy of initiator tRNA selection [88]. This has been hypothesized to occur to reduce chemical reactivity in the decoding center of the small subunit, where m^3U1498 is located [22].

1.2.6 Multiple Modifications

The methylated pseudouridine, $m^3\Psi1915$, is found in Domain IV in the 50S subunit [46]. This modification is yielded by the methyltransferase RlmH, which is exclusively capable of recognizing associated 70S ribosomes, thereby modifying 23S rRNA late in the ribosome biogenesis process [47]. The methyltransferase makes extensive contacts with both subunits, and methylation via RlmH occurs after pseudouridine formation by RluD [89]. Knockout of *rlmH* results in a slight growth retardation under optimal growing conditions, and knockout strains of *rlmH* are outcompeted by the wildtype in a competition assay [47]. It has been proposed that this methylation functions as a checkpoint to signal the 50S subunit has engaged in translational initiation, as the modification location is close to the decoding center, and the enzyme interacts with the A-site tRNA and translation factors, which have been shown to be involved in ribosome recycling, translation accuracy, initiation, and termination [89, 90].

A dual methylation of cytosine occurs in the 3′ minor domain in the 30S subunit: $m^4Cm1402$ contains methylations both at the N4 and the 2′-O positions of the base [32, 50]. The methylation of the base is universally conserved, and the dual methylation combination is present in nearly all bacteria (Table 1.1) [32, 91]. There are two methyltransferases responsible for these methylations: RsmI initially generates the 2′-O-base methylation of C1402, followed by RsmH, which yields the N4 methylation [32]. Both enzymes are capable of modifying 30S subunits but are unable to modify free 16S rRNA and the 70S ribosome, suggesting a modification pattern in

the intermediate stages of ribosome assembly (after 30S assembly but before 70S association). Interestingly, the efficiency of methylation of the shared target is higher during combined action of both enzymes [32].

Deletion of *rsmH* and *rsmI* results in minor impacts on subunit association and 30S complex formation and is not essential for functioning of reconstituted ribosomes (Figure 1.1 B). However, removal of *rsmH* and *rsmI* does yield severe impacts on tRNA binding to the ribosomal A and P sites and initiation codon selection [18, 92]. These knockout strains exhibit phenotypes with increased non-ATG initiation rates and decreased UGA read-through [22, 32]. Nucleotides 1400-1405 and 1496-1502 of 16S rRNA are conserved in prokaryotes, as they partake in the formation of the P-site of the ribosome. The crystal structure of the *E. coli* ribosome subunits revealed Cm1402 promotes interaction with P-site codon of the mRNA through a stabilized C3′- endo conformation of C1402 [18, 28, 32] .

1.3 Pseudouridylation

Pseudouridine (Ψ) is a 5′-ribosyl isomer of uridine containing a C-C glycosidic bond rather than a canonical N-C glycosidic bond and is often referred to as “the fifth nucleotide” (Figure 1.3) [93]. This modification is conserved in all domains of life and has been positively identified in ribosomal RNA (rRNA), transfer RNA (tRNA), messenger RNAs (mRNAs), microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), mitochondrial tRNAs (Mt-tRNAs), small Cajal Body-specific RNAs (scaRNAs), long intergenic non-coding RNAs (lincRNAs), and miscellaneous RNAs (misc_RNAs) [94, 95]. In prokaryotes, pseudouridine has been confirmed to be present in rRNA and tRNA [93]. Closely related organisms have been found to share specific sets of pseudouridine residues in homologous rRNAs, however, only RluE and RluD

have been shown to generate universally conserved pseudouridines at positions 2457 and 1917 in 23S rRNA [96, 97].

The isomerization of uridine to pseudouridine is catalyzed by a collection of enzymes known as pseudouridine synthases, all of which may be classified into six major families based on sequence alignments: Pus10, RluA, RsuA, TruA, TruB, and TruD [93, 95]. Notably, the Pus10 family of pseudouridine synthases is exclusive to eukaryotes and archaea [93]. Despite the variation in sequence composition, all families contain a similar tertiary structure: bearing a core with a common fold and a conserved active site cleft [98]. This core architecture is unique to pseudouridine synthases [86], and it is hypothesized these enzymes evolved through multiple independent gene duplications of an ancestral pseudouridine synthase [17].

Pseudouridines have an additional hydrogen bond donor group compared to uridine, which is speculated to stabilize RNA structure by providing greater rigidity to the phosphodiester backbone as well as by stabilizing the Ψ-A base pairs through base stacking and intramolecular water bridges with the RNA backbone [99]. More specifically, pseudouridine causes an increased propensity towards the 3′ endo sugar pucker, stabilizing the base. By adopting different sugar puckers, the overall structure and function of RNA is altered; the stability of the RNA helix is influenced as the orientation of the phosphate backbone, nucleobase, and ribonucleotide are altered [100]. This ultimately influences the functional properties of the RNA substrate by manipulating the structure of the RNA through its spatial conformation to provide supplementary stability under stress conditions [94, 99, 101].

There are 11 pseudouridines in the *E. coli* ribosome, of which the corresponding synthases fall into two of the five enzyme families present in bacteria: RluA and RsuA (Table 1.1). The RsuA family contains highly specific synthases, while the RluA family contains synthases that

exhibit regional specificity. The substrate recognition mechanisms for these families are poorly understood, however, neither sequence difference nor enzyme structures provide an explanation [86].

The majority of pseudouridylation target sites are found within functionally important domains, specifically in the 3′ minor domain (the Decoding Center) of 16S rRNA, and Domain V (the Peptidyltransferase Center) of the *E. coli* 23S ribosomal RNA (Figure 1.3-A) [17, 93, 94]. The centrality of these modifications permits them to influence features of ribosomal function [94]. Although the proportion of pseudouridine residues in eukaryotic rRNA is more than 3 times larger than their bacterial counterparts (0.9-1.4% vs 0.03-0.4% of all uridines are pseudouridines, respectively), the additional pseudouridine residues maintain the trend of clustering in the functionally relevant domains [17, 102]. Additionally, both bacteria and eukaryotes employ the same mechanism of pseudouridylation through an invariantly conserved aspartate [103]. Furthermore, the proportion of pseudouridine in bacteria is expected to be decreased compared to eukaryotes, as bacterial organisms employ independent, stand-alone enzymes for each individual substrate. Contrarily, eukaryotes utilize H/ACA guide RNAs with a core set of proteins, called H/ACA snoRNPs to modify rRNA, allowing a versatile range of targets to be pseudouridylated [93, 95]. In addition to introducing pseudouridine into rRNA, H/ACA sRNPs have been found to stabilize telomerase RNA, be involved in mRNA production, rRNA processing, as well as the pseudouridylation of snRNA, mRNA, and other RNAs [104].

As aforementioned, H/ACA snoRNPs are comprised of two main components: the structural guide RNA that determines sequence specificity, and four core proteins (Nhp2, Nop10, Gar1, Cbf5) [104, 105]. The catalytic component of these enzymes is Cbf5, which is a member of the TruB family [104], and it has been found to interact tightly with Nop10 [106]. In most

studied eukaryotes, the guide RNA consists of a hairpin-hinge-hairpin structure, in which the hinge region contains two conserved elements: the H box and ACA box [107]. The H box has a consensus sequence of ANANNA, while the ACA box is relatively conserved, with AUA, AAA, and AGA also being found [104, 107]. Notably, the ACA box is always found three nucleotides upstream of the 3′ end of the guide RNA [107].

The hairpins within the guide RNA contain two single-stranded regions with complementary sequences to the target RNA, commonly referred to as pseudouridylation pockets [107, 108]. Upon binding to the target RNA, both the target uridine and the nucleotide 3` of the target uridine are unpaired within the pseudouridylation pocket. As the target uridine is inserted into the active site of the core protein Cbf5, pseudouridine is formed [107, 108]. It has been previously found that at least eight base-pairs are required between the guide and target RNAs, and that each side of the pseudouridylation pocket must have at least three base pairs [109, 110].

It is worthwhile to note that eukaryotes also employ independent, stand-alone pseudouridine synthases (Pus enzymes) to introduce the fifth nucleotide, and these enzymes are related to their bacterial counterparts [95, 98]. These stand-alone enzymes fall into the TruA, TruB, TruD, RluA, and Pus10 families of pseudouridine synthases in eukaryotes[95]. Notably, some organisms may contain more than one paralog of a given Pus enzyme [95]. Pus enzymes have been observed to introduce pseudouridine into both cytoplasmic and mitochondrial tRNAs, mRNAs, snRNAs, and mitochondrial rRNA, but not cytosolic rRNA [95, 111].

Despite the stark differences between the specificity requirements of pseudouridine synthases in eukaryotes and prokaryotes, there are a number of similarities observed between the two systems. As aforementioned, rRNA modification sites are localized within functionally

relevant domains such as the PTC, DC, and inter-subunit bridges in both bacteria and eukaryotes [17, 102]. Due to the denseness of the modification sites, pseudouridine synthases in both domains (i.e. both H/ACA snoRNPs of eukaryotes and the Pus enzymes of bacteria) act during ribosome biogenesis [18, 112] and are expressed constitutively to ensure modification of target uridines occurs. In eukaryotes, the introduction of pseudouridine during biogenesis is essential [94], while biogenesis may proceed in prokaryotes without the introduction of pseudouridine at these target sites [113]. Additionally, the RNA dependent (H/ACA snoRNPs) and independent (Pus enzyme) mechanisms of pseudouridylation share a common catalytic mechanism [103] and have both been shown to have a secondary function as an RNA chaperone [104, 114-116] and influence ribosomal function [18, 94, 104] (Figure 1.1).

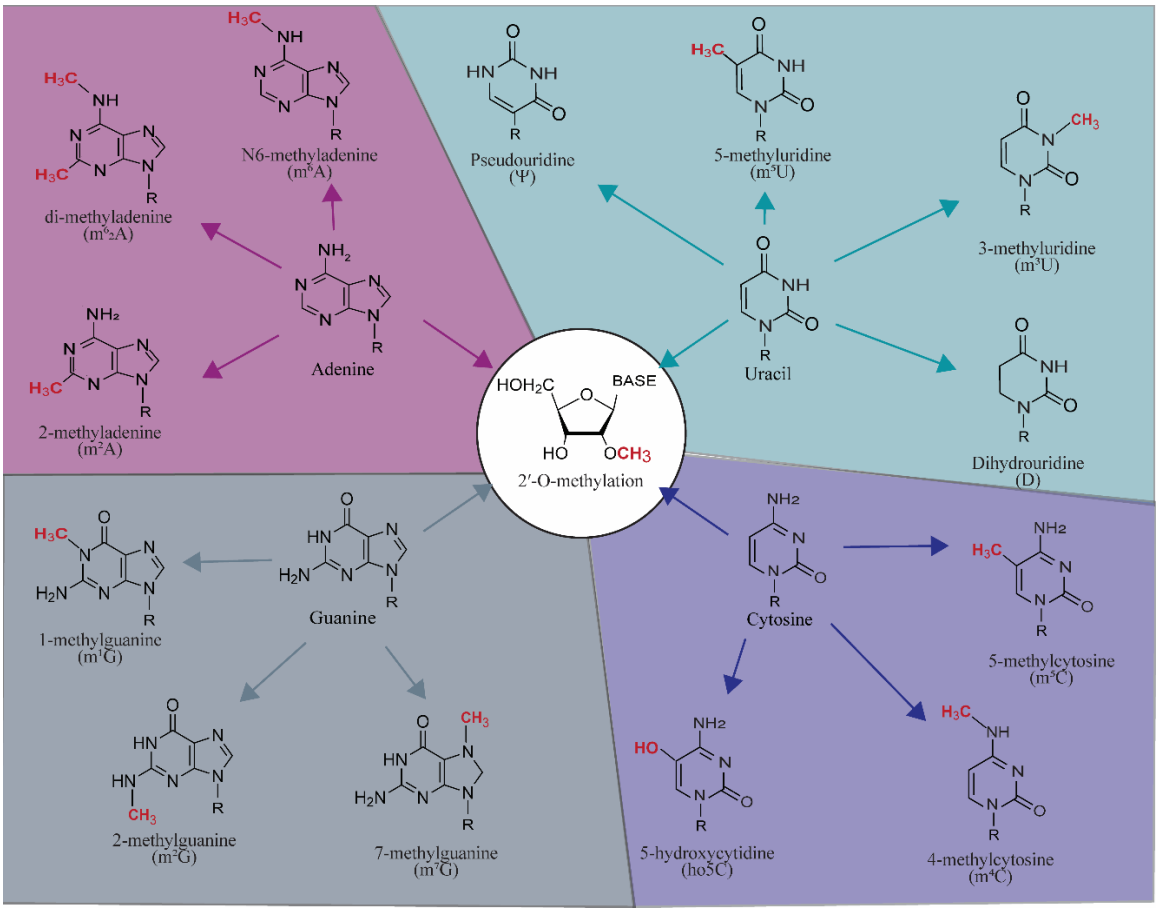


Figure 1.2 - Schematic of ribosomal RNA modifications sorted by nucleoside. A comprehensive overview of the ribosomal RNA modifications is covered, initially divided by nucleotide modification, then subdivided by location and type of modification. Changes from the original nucleotide are highlighted in red. R = ribose sugar.

The PTC of the *E. coli* large ribosomal subunit contains a number of highly conserved rRNA nucleotides that serve as the site of peptide bond formation – a crucial reaction during translation [117]. What is currently known about pseudouridines and their corresponding synthases does not fully explain the degree of conservation or the consequences of pseudouridylation on ribosomal structure and function [94]. It is hypothesized that pseudouridine leads to stabilization of higher-order structure and promotes proper binding between tRNA and the ribosome to influence the speed and accuracy of decoding and proofreading during

translation [118]. The presence of pseudouridine has also been shown to influence ribosome biogenesis and function [119].

1.3.1 RluA family of Ψ synthases

The pseudouridine synthase RluA was the first synthase identified in the RluA family. Characteristic of the RluA family, the RluA synthase has multisite specificity, modifying 23S rRNA as well as tRNA [98]. In 23S rRNA, this enzyme isomerizes U746 in Domain II [66, 120] (Table 1.1). The rRNA modification in *E. coli* is incorporated early in ribosome biogenesis, and the relative conservation of the modification has yet to be determined. There appears to be no global tertiary structural similarity between the rRNA and tRNA substrates, however, parallels in the secondary structure arise: the modification sites are all on the 5′ side of the loop of a simple stem-loop and share the consensus sequence ψUXXAAA [98, 121, 122]. A knockout strain of *rluA* has little to no discernable change in exponential growth rate in minimal or rich media in the temperature range of 24-42°C [66]. Knockout strains are outcompeted in a competition assay against the wild type strain [122]; therefore, one may infer that the presence of pseudouridine 746 in the ribosome confers a selective advantage in a competitive environment (Figure 1.1 B).

RluC also exhibits the multisite specificity typical of the RluA family, modifying three uridines in the 50S subunit to generate Ψ955, Ψ2504, and Ψ2580 [17, 62] . The modification Ψ955 is found in Domain II, while Ψ2504 and Ψ2580 are both located within the PTC in Domain V of 23S rRNA. Ψ2504 is within a single-stranded region of 23S rRNA, and Ψ2580 is at the base of Helix 90 in Domain V. Although a knockout of *rluC* (eliminating the modifications Ψ955, Ψ2504, and Ψ2580) exhibits no discernable effects on the exponential growth rate in rich or minimal media in the temperature ranges of 24-42°C, growth phenotypes are observed in a co-culture competition assay, in which the knockout strain is outcompeted by the wild-type

strain [122]. As concluded with RluA, the pseudouridines Ψ955, Ψ2504, and Ψ2580 in the 50S subunit provide a selective advantage in a competitive environment (Figure 1.1 B).

The pseudouridine synthase RluD generates three pseudouridines in the rRNA of *E.* coli: Ψ1911, Ψ1915, Ψ1917 in 23S rRNA, and this enzyme acts late in the ribosome biogenesis pathway [63, 64]. Interestingly, all three of these modified residues are found in helix 69 of 23S rRNA, which forms a hairpin and inter-subunit bridge essential for ribosome function [123]. The close proximity of these modifications and their occurrence late in the assembly pathway suggests that RluD recognizes the stem-loop structure of helix 69. The modifications by RluD are highly conserved [124], and knockouts of *rluD* reduce growth rate three to five fold [64]. Knockout strains also accumulate ribosome assembly precursors of both the large and subunits, procuring the hypothesis that RluD may have a secondary function as a chaperone [125].

1.3.2 RsuA family of Ψ synthases

RluB is characteristic of the RsuA family having a high target specificity and modifying only Ψ2605 in the PTC of 23S rRNA during the early stages of ribosome biogenesis [61]. This modification is adjacent to the modification target of RluF, but RluB cannot cross-modify the U2604 isomerized by RluF [126, 127]. Knockout studies of *rluB* revealed no significant growth defects [61], however, these strains are outcompeted in a competition assay against the wildtype strain (Vienneau, unpublished).

The residue U2457 is isomerized by the pseudouridine synthase RluE in the PTC in 23S rRNA of *E. coli* ribosome [61]. This pseudouridine synthase has been shown to act faster than other pseudouridine synthases studied so far, which has been hypothesized to allow it to act early in ribosome biogenesis [96]. This is further supported by the location of the target residue in helix 89, which is in the PTC of Domain V and buried in the center the ribosome. Ψ2457 is the

most conserved pseudouridine in *E. coli* rRNA [96], corresponding to Ψ2826 in the 25S rRNA of *S. cerevisiae*, and Ψ4373 in the 28S rRNA of humans [96]. Knockout studies of *rluE* did not reveal notable growth defects [61], however, these strains were outcompeted in competition assays against the wildtype (Vienneau, unpublished).

RluF generates one of the two adjacent pseudouridine residues in the PTC, forming the pseudouridine adjacent to that catalyzed by RluB [65]. RluF enzyme isomerizes U2604 in 23S rRNA, however, is also capable of modifying U2605 with reduced efficiency [126]. The enzyme acts during the early stages of ribosome biogenesis, and the modification is believed to be conserved [65]. Growth studies with knockout strains of *rluF* revealed no notable phenotypes, although the *rluF* knockout strains are outcompeted in a competition assay against the wildtype (Vienneau, unpublished).

The pseudouridine synthase RsuA is characteristic of its family, having only a single target for modification: U516 in the 5′ Domain of 16S rRNA of the *E. coli* ribosome [128]. This synthase acts during the early to intermediate stages of the assembly of the 30S subunit [17, 129, 130]. Ψ516 is a highly conserved modification [66]. A knockout strain of *rsuA* exhibits little to no discernable effect on the growth rate when tested in rich or minimal medium and under a temperature range of 24-42°C [66]. Notably, this knockout strain is outcompeted by the wildtype strain in a competition assay, indicating that strains containing this pseudouridine synthase have a selective advantage over those without the modification (Figure 1.1 B) [122].

1.4 Other Modifications:

1.4.1 Dihydrouridine

Dihydrouridine is the result of a reduction of uridine into dihydrouridine, yielding a non-planar nucleobase (Figure 1.2) [19]. This modification disrupts base-stacking interactions,

thereby promoting the C2′-endo sugar conformation and allowing for greater conformational flexibility of the RNA in regions where both tertiary interactions and loop formation must be accommodated, such as the PTC [131]. In addition to this disruption, dihydrouridine stabilizes the ribose of the modified base and the 5′ neighboring nucleotide. The methyl group on m^2G2445 adopts a *s*-trans position, occupying a cavity created by the neighboring dihydrouridine at position 2449 [19]. The formation of dihydrouridine represents a biological strategy to enact an effect opposite to methylation or pseudouridylation, which increases the stability of the ribosome through the C3′-endo conformer [131].

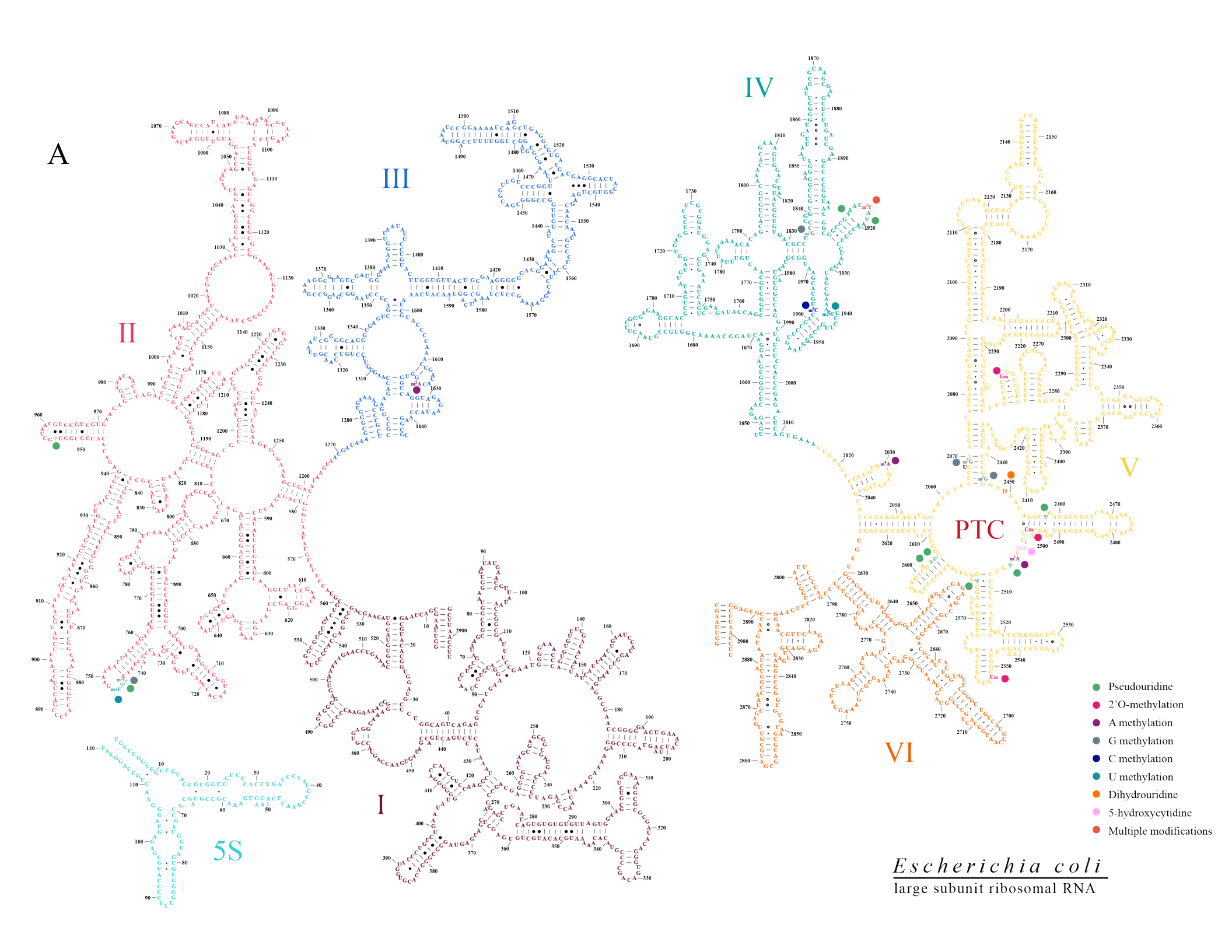
There is a singular dihydrouridine in the *E. coli* ribosome at position 2449 of the large ribosomal subunit 23S rRNA in a single-stranded region of the PTC (Table 1.1). Relatively little is known about this modification or its respective enzyme, as dihydrouridine assays have not been developed for rRNA as they have for tRNA. There are three dihydrouridine synthases known in *E. coli*: DusA, DusB, and DusC [23]. It seems most likely DusB is responsible for this modification, as its specificity requirements are not as well understood as those of DusA and DusC in *E. coli* [23, 132]. Furthermore, DusB has been shown to have multi-site specificity for up to three different sites, as well as to modify bases on opposing sides of an RNA-loop structure in *M. capricolum* [133]. Taken together, this suggests DusB may have specificity requirements flexible enough to target rRNA and generate D2449.

1.4.2 Hydroxycytidine

Hydroxycytidine (hoC) is a novel modification in RNA and has not been found to impact ribosomal RNA structure (Figure 1.2), however, it may be linked to certain types of environmental stress response such as radiation, desiccation, and oxidative stress [2, 134]. There is one hydroxycytidine in the *E. coli* ribosome, found at position C2501 in a single-stranded

region of the PTC (Table 1.1) [2]. The enzyme responsible for this modification is generated by the gene *rlhA* in *E. coli* and binds to an early-stage assembly intermediate of the 50S subunit [68]. The location of this modification in the catalytically active region of 23S rRNA suggests 5-hydroxycytidine may play a role in fine-tuning the peptide-synthesizing activity of the ribosome [2].

Hydroxycytidine is also present in the equivalent position of 23S rRNA in *Deinococcus radiodurans* [2], and RlhA homologs may be found in in γ-proteobacteria, δ-proteobacteria, clostridia, bacteroidia, cyanobacteria, and some species of archaea [68]. The modification of C2501 is universally conserved in all domains of life and has been previously found to be essential in *E. coli* [68]. The degree of modification in 23S rRNA is considerably larger during the stationary phase of growth than during the exponential phase [2, 67]. Interestingly, the formation of ho-5-C2501 is dependent on environmental iron concentration [68] and requires the presence of prephenate (a precursor for aromatic amino acids) [135]. Taken together, these findings suggest that altering the frequency of ho5C2501 could regulate translation associated with metabolic changes in the shikimate pathway [68].


A
I
II
III
IV
V
VI
PTC
5S
Pseudouridine
2'O-methylation
A methylation
G methylation
C methylation
U methylation
Dihydrouridine
5-hydroxycytidine
Multiple modifications
Escherichia coli
large subunit ribosomal RNA

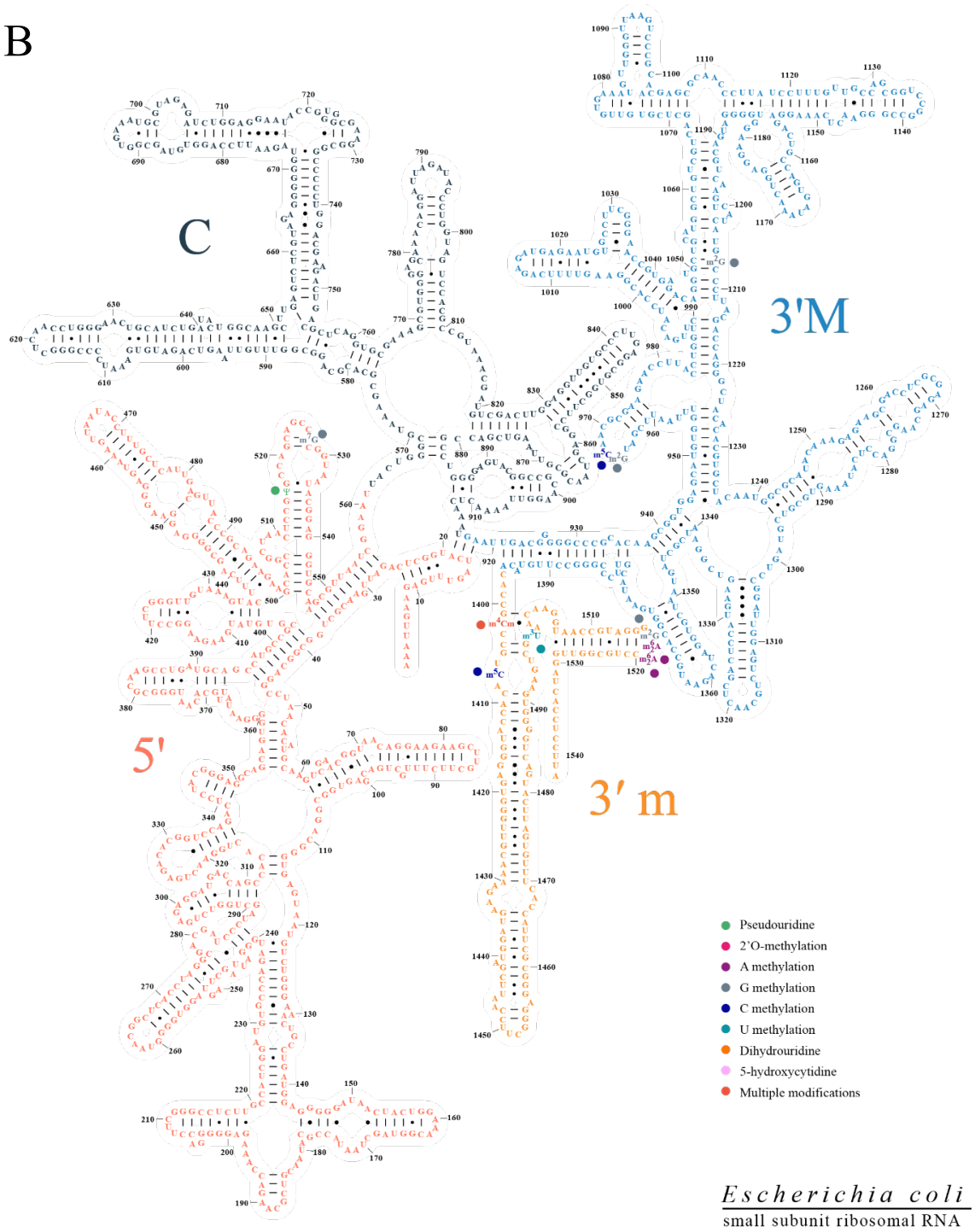


Figure 1.3: The secondary structure of the *Escherichia coli* ribosome. A) The secondary structure of 23S and 5S rRNA, highlighting all modified nucleotides in each domain.* **B)** The secondary structure of 16S rRNA highlighting all modified nucleotides in each domain.*

*Updated and modified image from Center for Molecular Biology of RNA (http://rna.ucsc.edu/rnacenter/ribosome_images.html)

1.5 Conclusion

Ribosomal RNA is extensively modified co- and post-transcriptionally in all organisms. Many of these rRNA modifications are crucial for protein synthesis; however, the contribution of the rRNA modification to ribosome formation and function and the degree of modification differ [2]. These rRNA modifications lead to altered steric and electronic effects, ultimately changing the base-pairing, sugar puckers, and base stacking properties of the RNA [19, 20]. Based on the research performed so far, it is evident that many rRNA modification enzymes are not essential for survival yet have been conserved throughout evolution; many of the rRNA modifications are nearly universally conserved among bacteria, and some conservation extends into eukaryotes and archaea [22, 23]. Such high levels of conservation are indicative of a significant role and potential advantage(s) for organisms containing these modifications. Although the deletion of many rRNA modifying enzymes leads to no significant growth defects under optimal conditions, the absence of even a single rRNA modification often leads to a disadvantage and out-competition in a competitive growth environment for many modification enzymes. Thus, it could be argued that these rRNA modifications evolved to equip the bacterial ribosome towards various environmental stressors, which may have provided a selective advantage during natural selection. The environmental stress response is crucial in bacteria, as environmental conditions change frequently, and these organisms must be able to adapt quickly to survive. By conserving rRNA modifications that confer a selective advantage under stress, bacteria with such rRNA modification enzymes have a selective advantage in a competitive environment. Furthermore, many of these rRNA modifications also have widespread effects on the translation capacity of the ribosome, proving their functional value and reason for widespread conservation [19]. This

would explain why these modifications are highly conserved but not essential in knockout studies.

Elucidating the hierarchy of modification in the *E. coli* ribosome is an expanding area of future research, as certain rRNA modification enzymes may promote proper folding of the RNA, acting as an RNA chaperone during biogenesis. Recent work has clearly demonstrated that certain methyltransferases and pseudouridine synthases have an additional rRNA and tRNA chaperone activity, respectively [114, 136, 137]. If this trend of chaperone activity extends to other modification enzymes, organisms containing these modifications would have an additional selective advantage under environmental stressors and be more likely to survive. These rRNA modifications, therefore, would prove to be evolutionarily advantageous.

1.6 Structure and mechanism of the PTC pseudouridine synthases RluB and RluF

As discussed in section 1.3, all six families of pseudouridine synthases share a common fold with a core β-sheet and conserved residues within the active site [98]. Due to this core architecture being unique to pseudouridine synthases, it is proposed that all pseudouridine synthases share a common catalytic mechanism [86]. Among the conserved active site residues, only the aspartate has been shown to be essential for catalytic activity and is invariantly conserved [138, 139]. Studies have found this essential aspartate serves as a nucleophile to attack the C2′ of the ribose ring and deprotonate it [139]. This eliminates the pyrimidine ring and forms a glycal intermediate, allowing the pyrimidine ring to be repositioned and reattached to form a C-C glycosidic bond [139]. In nearly all pseudouridine synthases, the essential aspartate is the only acid/base group within the vicinity of the base capable of forming the glycal intermediate [139]. However, the crystal structure of RluB has found a tyrosine covalently bound with the C6 of the target base, rather than the water molecule observed with other pseudouridine synthases [127].

Interestingly, this tyrosine is not essential for catalytic activity and does not fit in with the proposed mechanism of the glycal intermediate, leaving open questions about why this tyrosine is conserved and how it relates to the glycal mechanism. Current hypotheses suggest this tyrosine contributes to catalysis by orientating the target base or providing binding stability [127], although these have yet to be resolved.

Notably, RluF has not been found to form this covalent bond, although it does possess an equivalent tyrosine in the active site [126, 127]. This is observed in RluF's closest homolog TruB, in which the tyrosine has been found to be essential for catalysis, and proposed to function as a base for proton abstraction during the final stages of catalysis [126]. Due to the similarity between RluF and TruB, it is possible there is overlap in RNA recognition. This becomes particularly interesting when considering the tRNA target of RluF, as it has been speculated its mechanism mimics TruB: recognizing the anticodon loop based on site-specific isomerization [140].

RluB and RluF modify adjacent uridines in the PTC, separated by ~4 Å [126]. RluB has exclusive specificity for Ψ2605, whereas RluF is not entirely restricted to Ψ2604; it has the ability to also modify Ψ2605 *in vitro* with reduced efficiency [126, 127]. Both pseudouridine synthases have been crystallized with a 21-nt RNA hairpin containing a 5-fluoro derivative at the target uridine to prevent catalysis. Based on these RNA-protein structures, an eight residue RNA-binding loop within each enzyme was identified [127]. The bacterial RluB and RluF enzyme families each have a highly conserved binding-loop that is responsible for binding the RNA substrate to the enzyme [127] and the eight-residue loops differ between RluB and RluF (Figure 1.4 A). Consequently, the hypothesis arises that this loop is what confers such high

specificity between the two enzymes.

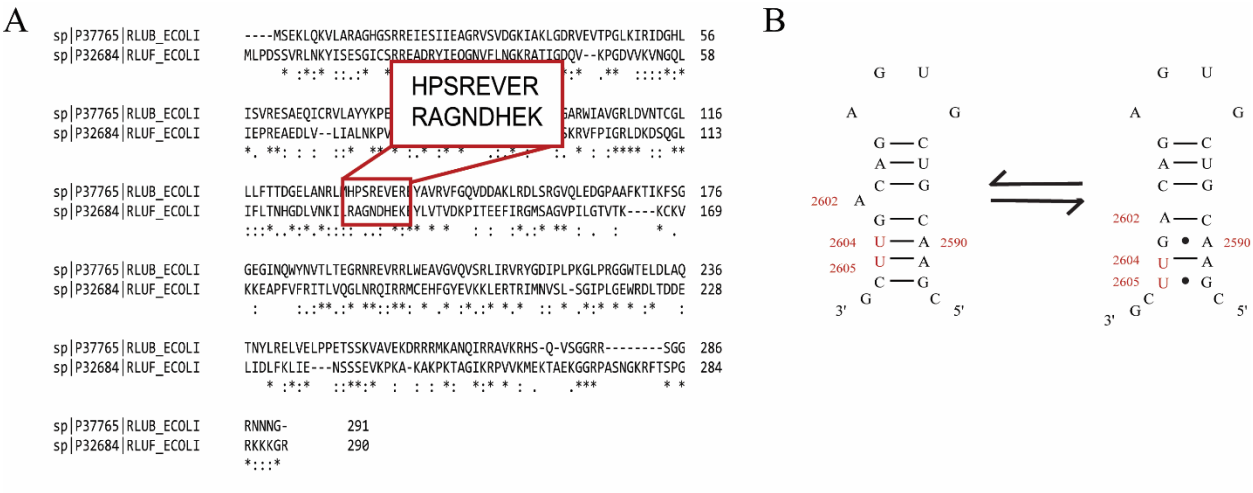


Figure 1.4– Binding of the target RNA stem-loop by RluB and RluF. (A) Eight-residue binding loop in RluB and RluF. Sequence alignment of *E. coli* RluB and RluF. The eight-residue binding loop conserved among RluB and RluF families is indicated. **(B)** Shift in RNA base pairing upon binding RluF as observed in crystal structures. The A2602 bulge observed in the *E. coli* ribosome (left) is recognized by RluB which modifies U2604. The shift in base-pairing to the conformation shown on the right incorporates A2602 into the stem-loop and is partially responsible for U2605 being modified by RluF.

Furthermore, it has been previously found that a difference in binding ability of the target stem-loop arises between RluB and RluF (Figure 1.4 B). The standard A2602 bulge, which is found in the structure of the ribosome, is recognized by RluB, which modifies U2604. However, RluF does not possess the same stabilizing abilities as RluB, and it causes a shift in base-pairing within the target RNA to a conformation which incorporates A2602 into the stem-loop and is partially responsible for U2605 being modified by RluF. When bound to RluB, A2602 is bulged out in the target RNA stem-loop, positioning U2605 in the catalytic pocket [126]. Conversely, when the same target RNA is bound to RluF, A2602 is incorporated into the hairpin, positioning U2604 in the catalytic pocket of RluF. As such, the A2602 bulge in the target 23S rRNA of *E. coli* also confers some specificity between the two enzymes [127].This structural observation

does not fully explain the RNA recognition mechanism: the rRNA rearrangement step would likely not occur until the RNA is bound to the protein, and it does not explain RluF's ability to modify both U2604 and U2605.

Czudnochowski *et al.* [127] described three major conformational changes occurring in RluB upon RNA binding. First, the S4 domain becomes rigid and binds to the major groove of the RNA substrate. Next, two subdomains around the active site cleft undergo a rigid hinge motion to close around the RNA; similar motions are observed in RluF and TruB [126, 141]. Finally, residues 132-134 of the eight-residue binding loop refold in a 3_{10} helix to bind the RNA near the A2602 bulge. This third conformational change becomes the most interesting in the context of this thesis. A 3_{10} helix differs from a canonical α helix in that it has 3 residues per turn, with an angle of 120° between each residue [142]. Comparatively, an α-helix has 3.6 residues per turn, with an angle of 100° between each residue. Thus, 3_{10} helices are more tightly wound, adopting on a longer and thinner shape than conventional helices [142]. Such helices have been previously found to have functional importance, as the R groups of amino acid residues are restricted and form ridges along the helix.

Interestingly, the binding requirements of RluB and RluF differ only from G2597 to A2602 in the rRNA, as these enzymes form congruent hydrogen-bonding interactions from A2590-U2596 [127]. As such, it is logical to assume that the specificity arises from residues in the enzyme that bind to the region of G2597 to A2602. Not surprisingly, this binding occurs with the conserved, eight-residue RNA binding loop among the enzyme families. This region is shifted 3-4 Å in RluF (relative to RluB) as the A2602 bulge is incorporated into the stem-loop [127].

RluF also has a conserved eight-residue RNA binding loop (residues 128-135) analogous to RluB's, although its sequence is distinct from RluB (see results section). These residues are

equivalent to residues 131-138 in RluB, even refolding into a similar helix when RNA is bound [127]. Notably, this helix occurs between α-1/β-4 (RluB) and α-2/β-6 (RluF) secondary structure elements, between which the binding loop is found [127]. However, the differences in the R groups of these side chains in RluF have been found to be incapable of stabilizing the A2602 bulge in the target RNA. While many aspects of the two enzymes are conserved, a notable difference in the apoenzymes has been observed: the secondary structure of the residues following helices α-1/α-2 (the binding loops themselves)[127]. Czudnochowski describes differences in the apo-forms of the RNA binding loops themselves; in RluB, these residues form a loop in the absence of RNA, while RluF forms two additional turns of a helix [127]. Notably, upon RNA binding both RluB and RluF alter the conformation of their binding loops to adopt the same protein fold.

It has been proposed that the stabilization of the A2602 bulge arises from these conserved residues in the RNA-binding loop, and the lack of these stabilizing residues in RluF allow for the stem-loop rearrangement observed [127]. A prime example of this arises in glutamate 135: this residue is conserved among the RluB family and forms a hydrogen bond to the base of the A2602 bulge. The equivalent residue in the RluF family is aspartate and is proposed to be too short to make such a contact. Alian and colleagues hypothesized that the differences in the end of helix α-2 in RluF, primarily arginine 128, are critical for repositioning of A2602 for incorporation into the stem-loop. R128 is hydrogen-bound to the backbone of C2601 (base paired to G2592 in the stem-loop) of the substrate (Figure 1.4), as well as positioned to form a hydrogen bond to the ribose O2′ of A2602 [126]. The hypothesis of R128 re-positioning the RNA is further supported by the high conservation of R128 among RluF families of various species, while a conserved proline exists at the equivalent position in RluB [126]. Taken

together, this combination of specificity and similarity between RluB and RluF procure the questions of how and why these synthases have evolved.

1.7 Objectives

While advancements into the understanding of RNA binding and target selection by RluF and RluB have been made in recent years, the specific roles of each conserved residue in the RNA-binding loops of RluB and RluF are yet to be understood. The high level of conservation among the families of each enzyme implies functional importance. The objective of this thesis is to investigate the role of the conserved residues of the binding loops of RluB and RluF for substrate binding and catalysis. To investigate which structural elements in RluB and RluF are required for target RNA recognition and specificity, several protein variants will be generated, including wildtype enzymes, a catalytically inactive variant, and variants likely to disrupt binding with the target RNA. By altering a single residue in the conserved loop, one may infer the role of said residue in binding and positioning the target RNA for catalysis in catalytic assays, thus gaining insight into the rationale for its conservation throughout evolution.

CHAPTER 2: MATERIALS & METHODS

2.1 Materials

All chemicals were obtained from Fisher Scientific. [5-^{3}H] UTP for *in vitro* transcriptions was purchased from Moravek Biochemicals, chromatography materials from GE Healthcare, and oligonucleotides from Integrated DNA Technologies (IDT).

2.2 *DNA, RNA, and Protein Electrophoresis*

All DNA samples were analyzed on a 1% Agarose gel in TAE buffer at 120 V for 1.5 hours and stained in Ethidium Bromide. All RNA samples were separated on a 10% Urea PAGE at 300 V for 1 hour, stained in Syber Green, and imaged on an Amersham Typhoon. All protein samples were analyzed on a 12% SDS-PAGE at 180 V for 1 hour. Gels were stained by boiling in 50 mL dehydration buffer (50% ethanol, 10% glacial acetic acid) and incubating 20 minutes on shaking platform. Gels were then transferred into 50 mL stain buffer (5% Ethanol, 7.5% glacial acetic acid, 0.25% Coomassie blue), brought to a boil, and incubated until bands visible. PAGEs were imaged on Amersham Typhoon.

2.3 *Site-directed mutagenesis of RluB and RluF*

Primers for each variant of RluB and RluF can be found in Table 2.1. Plasmids for RluB (pCA24N_RluB) and RluF (pCA24N_RluF) wildtypes were obtained from the ASKA collection [143] and purified using a BioBasic EZ-10 Spin Column Plasmid DNA Miniprep kit. The appropriate plasmid was then used in subsequent mutagenesis reactions.

Each reaction consisted of 1x Q5 reaction buffer (New England Biolabs), 1 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer, 3-5 ng/µL DNA template (either pCA24N_RluB or pCA24N_RluF), and 0.01 U/µL Q5 High Fidelity DNA Polymerase (New England Biolabs). The PCR was carried out with the following PCR temperature profile: initial denaturation at 98ºC for

30 seconds, 20 cycles of 98ºC for 30 seconds, 72ºC for 30 seconds, and 72ºC for 3 minutes. Final extension occurred at 72ºC for 5 minutes prior to holding at 4ºC. Each sample was incubated with 0.04 U DpnI (Thermo Scientific) for 3 hours at 37ºC, and 5 µL of the reaction was analyzed on an agarose gel.

When generating the loop-variants for both RluB and RluF, an additional ligation step was performed following PCR and prior to transformation to anneal both fragments. Following DpnI digestion, samples were incubated with 2 U T4 DNA ligase (New England Biolabs), 1x T4 DNA ligase buffer (New England Biolabs), and 50 ng of amplified vector DNA (either pCA24N_RluB_in_RluF or pCA24N_RluF_in_RluB) at room temperature for 2 hours. Next, the T4 ligase was inactivated by heating samples at 65°C for 10 minutes, and samples chilled on ice prior to transformation.

Table 2.1- Primers for Mutagenesis of RluB and RluF genes**.** The nucleotides altered with respect to the wild-type sequence are bolded.

VARIANT	SENSE PRIMER	ANTISENSE PRIMER
RluB D110N	3′ - CGCTTG**AAC**GTTAATACCTGTG GTCTGC– 5′	3′ – GTATTAAC**GTT**CAAGCGAC CCACGGC – 5′
RluF D107N	3′ - GCCGCCTG**AAT**AAAGACTCCC AGGG - 5′	3′ – GTCTTT**ATT**CAGGCGGCCA ATCGGGAAC – 5’
RluB H131A	5′ - CGTTTAATG**GCC**CCAAGCCGT GAAGTTGAACG - 3′	5′ - CGGCTTGG**GGC**CATTAAA CGGTTTGC - 3′
RluB E135D	5′ - CCCAAGCCGT**GAC**GTTGAACG TGAATATGC - 3′	5′ - CACGTTCAAC**GTC**ACGGCT TGGGTGCATTAAACG - 3′
RluF R128A	5′ - GAATAAGATCCTG**GCT**GCTGG CAATGATC - 3′	5′ - CCAGC**AGC**CAGGATCTTAT TCACCAGATCG - 3′
RluF-in-B	5′ - /5Phos/ **GAT CAT GAG AAA** GAATATGCCGTGCGTGTATTTG G - 3′	5′ - **ATT GCC AGC ACG** CATTAAACGGTTTGC - 3′

RluB-in-F	5′ - /5Phos/ **GAA GTT GAA CGT** GAGTATCTGGTGACGGTC GATAAACCG - 3′	5′ -**ACG GCT TGG GTG** CAGGATCTTATTCA CAG ATCGCCG - 3′

2.4 Transformation

0.5-3 µL (10 pg – 100 ng) of mutagenesis PCR product was added to 20 µL of competent *E. coli* AG1 (ME5305) cells and incubated on ice for 20 minutes. Cells were then heat shocked at 42ºC for 45 seconds and placed back on ice for 2 minutes. 250 µL of SOC media was added, and transformation mixture was incubated at 37ºC for 45 minutes. The transformation mixture was plated onto LB Agar with chloramphenicol (25 mg/mL) and grown at 37ºC overnight. Following plasmid minipreps (BioBasic), sequences were verified via Sanger sequencing (Genewiz).

2.5 Overexpression of RluB and RluF proteins

RluF and RluB wildtype and variants were grown in *E. coli* AG1 (ME5305) cells at 37ºC in LB media with 25 mg/mL chloramphenicol to an OD_{600} of 0.6-0.8, and expression was induced with 1 mM Isopropyl β- d-1-thiogalactopyranoside (IPTG). A sample equivalent to OD_{600} 1.0 was taken at 0, 1, 2, and 4 hours post induction. Samples were re-suspended in 0.1 M Tris-HCl with 8 M Urea and analyzed on a 12% SDS-PAGE at 180 V for 1 hour. After 4 hours of growth, cells were pelleted at 5,000 x g for 15 minutes. Cell pellets were shock frozen in liquid nitrogen and stored at -80°C until purification.

The RluF wildtype and variant strains were grown in *E. coli* AG1 (ME5305) cells at 37ºC in LB media with chloramphenicol (25 mg/mL) to an OD_{600} of ~0.8, and expression was induced with 0.25 mM Isopropyl β- d-1-thiogalactopyranoside (IPTG). These samples were then grown overnight at 18°C for ~15 hours. A sample equivalent to OD_{600} 1.0 was taken at 0 hours and overnight post induction. All further steps are identical to those described above.

2.6 Purification of RluB and RluF proteins

Cell pellets were resuspended in 5 mL/g of Buffer A (20 mM Tris-HCl pH 8.1, 400 mM KCl, 5% glycerol, 1 mM β-Mercaptoethanol, 0.5 mM phenyl-methyl-sulfonyl fluoride (PMSF), 30 mM imidazole) and incubated with 1 mg/mL lysozyme for 30 minutes. Sodium deoxycholate was added to 12.5 mg/g of cell pellet, and the lysate was further incubated on ice for 30 minutes. Subsequently, cells were opened via sonication (Branson Sonifier 450) in ten 1-minute cycles at intensity level 6, 60% duty cycle, and the cell lysate centrifuged at 30,000 x g at 4ºC for 30 minutes in a JA-25.5 rotor. Lysate was applied to 5 mL of Ni-Sepharose (50% slurry, GE Healthcare) and incubated for 1 hour with inversion. Subsequently, the slurry was centrifuged at 500 x g for 5 minutes at 4ºC, supernatant decanted, and resin washed 5 times with 3 resin volumes each of Buffer A. Proteins were eluted 8 times with 90% resin volume Buffer B (20 mM Tris-HCl pH 8.1, 400 mM KCl, 5% glycerol, 1 mM beta-mercaptoethanol (β-Me), 500 mM imidazole) by incubating for 5 minutes on ice and then centrifuging at 500 x g for 5 minutes at 4ºC. Elutions were pooled and concentrated to a final volume of 1-2 mL using a Vivaspin MWCO 10,000 Da by centrifuging at 4,000 x g.

The pooled elutions were further purified using a Superdex 75 Size Exclusion Chromatography column (GE Healthcare) pre-equilibrated with Buffer C (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 20% glycerol, 0.5 mM EDTA), and A_{280} was monitored. 3 mL fractions were collected, and sample fractions were analyzed by SDS-PAGE. Protein-containing fractions were pooled and concentrated to a final volume of 1-2 mL using a Vivaspin MWCO 10,000 Da by centrifuging at 4,000 x g. Samples were aliquoted and shock frozen to be stored at -80ºC.

Protein concentrations were determined via concentration gels with a BSA standard and analyzed with ImageJ. Percent purity of the band of interest (marked with an asterisk) as well as

contaminant bands were determined using equation 2.0, and the purity of each lane was averaged for each enzyme. Gel background was determined by calculating the intensity of a region of the gel without a band in the same area as the band selection, and then subtracting the given intensity of the background from each band intensity (individual or grouped).

Equation 2.0: Percent Purity (%) $= \frac{\text{intensity of band of interest} - \text{gel background}}{\text{intensity of all bands} - \text{gel background}} * 100$

2.7 *Preparation of radio-labelled RNA substrates via in vitro transcription*

The template for transcription was generated via PCR using the pCW1 plasmid encoding 23S rRNA as template [96] and the primers:

5′ -GACCGAACTGTCTCACGACGTTCTAAACCC- 3′ and

5′ -GCTAATACGACTCACTATAGGGATAACAGGCTGATACCGCCC- 3′. The PCR product was purified (BioBasic EZ-10 Spin Column DNA Cleanup Miniprep Kit). *In vitro* transcriptions were prepared to a final volume of 1 mL with 1x Transcription buffer (40 mM Tris HCl pH 7.5, 15 mM $MgCl_2$ 2 mM spermidine, 10 mM NaCl), 10 mM DT, 3 mM each of ATP, CTP, GTP, 0.1 mM [5-^{3}H] UTP (Moravek), 5 mM GMP, 0.01 U/ µL inorganic pyrophosphatase (iPPase), 0.3 µM T7 RNA Polymerase, and 1 ng/µL DNA template. Reactions were incubated at 37ºC for 4 hours, taking a sample every hour. Each sample was incubated with 1 U/µL DNaseI at 37ºC for one hour, centrifuged at maximal speed for 10 minutes at 4ºC, and stored at -20ºC. Samples were analyzed by 10% Urea PAGE.

2.8 Purification of radio-labelled RNA substrates via DEAE anion exchange chromatography

Soluble supernatants of *in vitro* transcription reactions were diluted four-fold in Buffer D (50 mM Na_3PO_4 pH 6.5, 0.2 mM EDTA) and applied to 2 mL of diethyl-aminoethyl (DEAE) slurry

pre-packed in a GE Healthcare PD10 gravity flow column. The column was washed 5 times with 1 resin volume of Buffer E (50 mM Na_3PO_4 pH 6.5, 250 mM NaCl, 0.2 mM EDTA), eluted 3 times with 1 resin volume of Buffer F (50 mM Na_3PO_4 pH 6.5, 600 mM NaCl, 0.2 mM EDTA), and then with 1 resin volume of Buffer G (50 mM Na_3PO_4 pH 6.5, 1 M NaCl, 0.2 mM EDTA). All samples were counted in the scintillation counter. Elutions were pooled and two-fold diluted in MilliQ water and subjected to ethanol precipitation. The pellet was isolated and air dried before resuspending in 300 µL of MilliQ water. RNA concentration was determined via absorbance at 260 nm, and activity was determined by counting 2 µL of resuspended RNA directly and using Equation 2.1.

Equation 2.1: $C = \frac{\varepsilon d}{A}$

where ε is the molar extinction coefficient and equals 204100 M^{-1} cm^{-1}, and d is the path length of the cuvette, equal to 1 cm.

2.9 *Tritium release assays*

To fold the substrate RNA, 0.3 µM tritium-labelled RNA was heated in $TAKEM_4$ buffer at 65ºC for 5 minutes and cooled at room temperature for 10 minutes. For each enzyme variant, a reaction mixture was prepared with 1 µM of enzyme, 0.3 µM tritium-labelled RNA substrate, and $TAKEM_4$ buffer (50 mM Tris HCl pH 7.5, 70 mM NH_4Cl, 30 mM KCl, 1 mM EDTA, 4 mM $MgCl_2$) to a final volume of 385 µL. Reactions were started via addition of enzyme to a final concentration of 1 µM, incubated at 37ºC, and quenched at different time points by adding sample to 1 mL 5% Norit A in 0.1 M HCl [144]. Quenched samples were incubated for 1 hour at room temperature, centrifuged at 13,000 x g for 2 minutes, and 900 µL of supernatant transferred to 0.5 mL 5% Norit A in 0.1 M HCl. Samples were centrifuged at 13,000 x g for 2 minutes, and 1 mL of supernatant was filtered through glass wool. The solution was centrifuged at 13,000 x g

for 2 minutes, and 900 µL of the cleared supernatant was added to 4 mL of scintillation cocktail, vortexed, and radioactivity determined by liquid scintillation counting. In addition, 5µL of the remaining reaction solution was counted directly. Direct counts (in DPM) were related to the pmole of pseudouridine formed with Equation 2.2.

$$\text{Equation 2.2: } P_{\psi} = \frac{DPM_{\text{time point}} * F_{dil}}{\frac{DPM_{direct}}{V_{direct}} * \frac{U}{RNA} * [RNA] * V_{\text{time point}}} * 100$$

where P_{ψ} is the picomoles of pseudouridine formed, $DPM_{\text{time point}}$ is the direct counts measured at the particular data point in the time course, F_{dil} is the dilution factor of each data point (found to be $0.01\bar{5}$), DPM_{direct} is the total direct count measured for each reaction tube, V_{direct} is the volume counted to obtain the direct counts (in µL), U/RNA is the total number of uridines present in the RNA substrate, [RNA] is the final concentration of RNA in the reaction (µM), vol of time point is the sample volume taken at each data point during the time course (in µL).

The amount of pseudouridine formed (in picomoles) was plotted against the time, and each graph was analyzed by fitting with a single-exponential function (Equation 2.3). All data was replicated at least three times to generate averages.

$$\text{Equation 2.3: } Y = Y_{max} + Amp * e^{-k\psi t}$$

where $k\psi$ is apparent rate of pseudouridine formation (k_{app}), Y_{max} is the maximal amount of pseudouridines formed, and Amp is the amplitude of the curve.

The amplitude of experiments containing no enzyme were averaged to use as a standard for noise of tritium exchange with water.

2.10 *Tritium Release Assays with pre-modified RNA Substrates*

Each reaction was prepared according to the scheme above, with the initiation of a reaction occurring via addition of one wildtype enzyme to a final concentration of 1 μM. This reaction was then incubated at 37ºC for 30 minutes before adding the second wildtype enzyme to a final concentration of 1 μM. Analyses of these reactions proceeded as described above.

2.11 *Nitrocellulose Filter Binding Assays*

Tritium-labelled RNA was premixed to a final concentration of 300 nM in TAKEM$_4$ buffer, heated to 65ºC for 5 minutes, and cooled for 10 minutes to room temperature. 60 nM (3 picomoles) of RNA was added to increasing concentrations of each enzyme variant (0 μM to 15 μM) in TAKEM$_4$ buffer and incubated for 10 minutes each at 37 ºC. Reactions were filtered through a GE Healthcare 0.2 μm pore and 25 mm diameter nitrocellulose filter soaked in TAKEM$_4$ buffer and washed with 1 mL of cold TAKEM$_4$ buffer prior to transfer into a scintillation vial. Nitrocellulose membranes were dissolved for 30 minutes in 10 mL of EcoLite scintillation cocktail (MP Biomedical). The amount of RNA bound to each enzyme, i.e. retained on the filter, was determined by scintillation counting [144]. Each binding curve was fitted to a quadratic equation using GraphPad Prism 5 to obtain the diss42ociation constant (K_D) using equation 2.4:

Equation 2.4: $P_{bound} = \text{Amp x } \frac{(K_D+[RNA]+[protein])^2}{2} - \sqrt{\frac{(kD+[RNA]+[protein])^2}{4} - [protein]x\ RNA}$

where P_{bound} is the % of bound RNA, Amp is the maximal level of bound RNA, [RNA] is the RNA concentration, [protein] the protein concentration, and K_D is the dissociation constant. Averages and standard deviations are reported of at least three titration curves.

To establish a standard for tritium exchange with water, binding amplitudes of reactions containing 0 μM of enzyme were averaged and used as a measure for noise.

CHAPTER 3: RESULTS

3.1 Mutagenesis & Purification of RluB and RluF Variants

As described in detail in the introduction, the RluB and RluF enzymes each have a highly conserved binding-loop that is responsible for binding the RNA substrate to the enzyme [127], and the eight-residue loops differ between RluB and RluF (Figure 1.4 A). Consequently, the hypothesis arises that this eight-residue binding loop confers the distinct RNA specificity of the two enzymes. Furthermore, it has been previously found that the A2602 bulge in the target 23S rRNA of *E. coli* also confers some specificity between the two enzymes [127]. The standard A2602 bulge (Figure 1.4 B) is recognized by RluB, which modifies U2604. However, RluF does not possess stabilize the A2602 bulge, instead causing a shift in base-pairing within the target RNA to a conformation which incorporates A2602 into the stem-loop and is partially responsible for U2605 being modified by RluF.

To investigate which structural elements in RluB and RluF are required for target RNA recognition and specificity, several protein variants were generated (Figure 3.1). In addition to the wildtype enzymes, a catalytically inactive variant was generated for both RluB and RluF by substituting the catalytic aspartate residue with asparagine (RluB D110N and RluF D107N). This substitution is expected to maintain the basic structure of the active site and remove the catalytic ability as shown previously for other pseudouridine synthases [96, 114, 144]. In doing this, I will generate a reliable standard to compare the binding variants against, in which a variant may bind to the target RNA with lower affinity to the catalytically inactive variant, but not undergo catalysis of the target uridine. Additional binding variants for RluB were designed based on the likelihood to disrupt binding with the A2602 bulge of the binding loop (RluB E135D and H131A), as well as a full “loop switch” variant which replaces the 8-residue binding loop from

one pseudouridine synthase with the opposing 8-residues (RluF-in-RluB). Likewise, binding variants for RluF were designed to disrupt binding with the RNA backbone at arginine 128 (RluF R128A), as well as a full “loop switch” variant (RluB-in-RluF).

The binding variant RluB E135D was designed with the intention to maintain the similarity of structure and available chemical groups (as glutamate and aspartate differ only by one carbon group). In all other single binding variants, the original amino acid residue was substituted with alanine, which is small, stable, and chemically inert. Alanine allows the protein to maintain secondary structure, as it has a high tendency to form helices and has been found to induce structure [145]. This makes alanine a more suitable choice for substitutions than glycine, which is lacking the β-carbon; this yields high conformational freedom and a low likelihood of conserved structure of the protein of interest [145]. By substituting with alanine, all side chain atoms beyond the β-carbon on the R-group are removed, thus, the effects of an individual side chain may be inferred from modification and RNA binding assays.

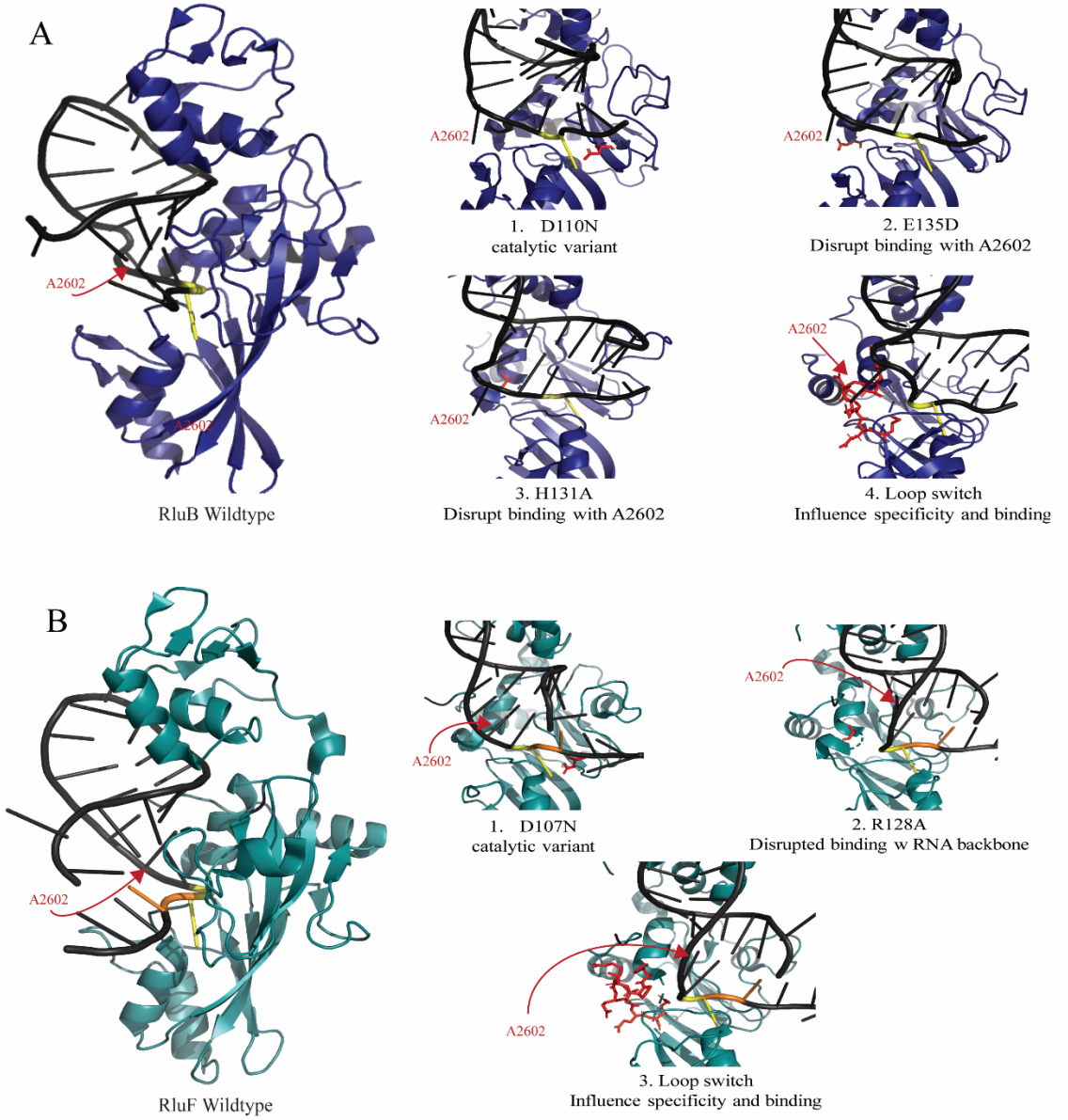


Figure 3.1– Proposed enzyme variants for RluB and RluF. (A) Blue: RluB enzyme variants with target RNA stem-loop (from left to right). Wildtype enzyme, catalytic variant (D110N), A2602 bond disruptions (E135D and H131A), and loop switch containing the 8-residue loop from RluF. **(B)** Green: RluF enzyme variants (from left to right). Wildtype enzyme, catalytic variant (D107N), RNA backbone bond disruption (R128A), and loop switch containing the 8-residue loop from RluB. Red = substituted residues among variants, black = target RNA stem-loop, yellow base = target uridine (U2605 in the case of RluB and U2604 in the case of RluF), orange base = U2605. A2602 can be observed flipped out of the stem loop in RluB variants. Structures were prepared using PYMOL software from original RluB (PDB ID: 4LGT) and RluF (PDB ID: 3DH3) structures. Each image was rotated to provide the best view of the substituted residue with the target RNA.

To analyze the impact of these substitutions in the RNA binding loop, plasmids encoding the mutated *rluB* and *rluF* genes were transformed into *E. coli* ME5305 cells; the proteins were overexpressed and purified via Ni-Sepharose and Size Exclusion Chromatography (SEC) for use in later RNA binding and activity assays. Concentrations were determined (in μM) via concentration gels compared to a Bovine serum albumin (BSA) standard. For comparative purposes, wildtype proteins were also generated. Figure 3.2 depicts a sample protein preparation of the variant RluF R128A. RluB and RluF are 32.7 and 32.5 kDa in size, respectively, thus one would expect a band at this size in the overexpression and purification of the wildtype and variant proteins. Panel A shows the overnight overexpression, in which one can clearly see an increase in intensity of a band around 35 kDa. A band of similar size is observed in each of the eight elutions from the Ni-Sepharose resin (Figure 3.2 B**)**, and in the SEC elution (volumes of 200-280 mL) (Figure 3.2 C-D). These bands occur at the expected size of the enzyme and in high concentration, suggesting the purification was successful.

Panels A and C of Figure 3.2 depict contamination around 66.2 kDa. These bands are most likely dimers of the purified enzyme, as the enzyme is quite concentrated, and the bands occur around the expected molecular weight of a dimer (expected at 64 kDa). The presence of dimers could potentially impact the accuracy of catalytic activity assays, thus, the percent purity based on single band intensities of the correct size were determined. For RluF R128A, I determined the lowest band to have a concentration of 489 ± 92 μM, with all other enzyme variants also producing final concentrations within the 100-1000 μM range (Table 3.1). In summary, all nine variants were produced to a high concentration and reasonable purity. To confirm substitutions only altered the site of binding and each variant maintains the overall protein fold of the wildtype, C/D Spectrometry could be performed.

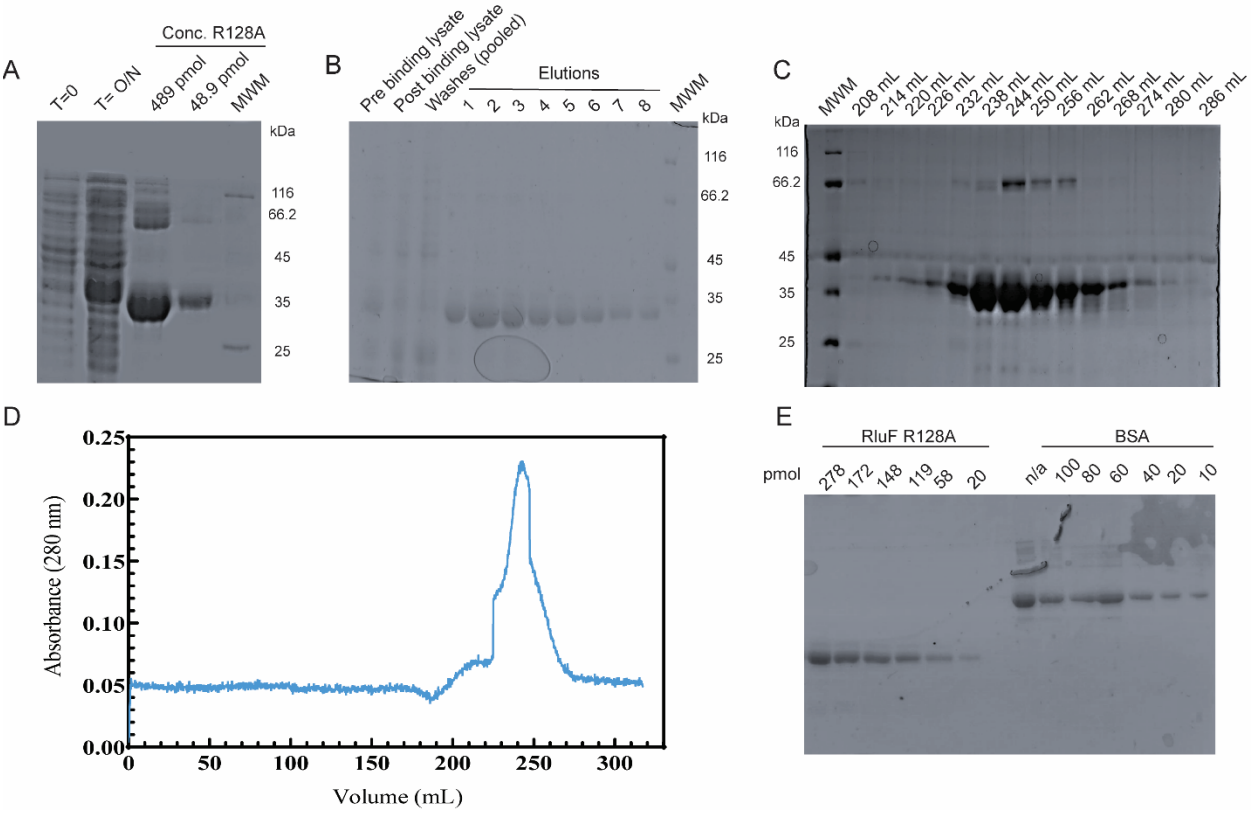


Figure 3.2 – Sample protein preparation of RluF R128A. **(A)** Overexpression of protein in *E. coli* ME5305 cells at 18 ºC analyzed by 12% SDS-PAGE. Lanes (from left to right): T=0 is the sample taken prior to over-expression. T = O/N is the sample taken after expression was induced and cells were allowed to grow overnight. Conc R128A are diluted samples after elutions from Ni-Sepharose were pooled and concentrated. MWM = Molecular Weight Marker. **(B)** SDS-PAGE of Nickel Sepharose purification of R128A variant in batch. **(C)** SDS-PAGE of protein-containing fractions of size exclusion chromatography (Superdex 75). **(D)** Size Exclusion Chromatography (Superdex 75) chromatogram. Blue = A280 of collected fractions as eluted from the column. **(E)** SDS-PAGE to measure the concentration of RluF R128A compared to a BSA standard. Protein concentration was determined to be 489 ± 92 µM using ImageJ software.

Table 3.1 – Protein preparation summary of RluB and RluF variants**.**

ENZYME	CONCENTRATION (μ M)	RELATIVE PURITY (# OF BANDS)	PURITY TOP BAND (%)	PURITY CONTAMINANT BAND (%)
RluB wildtype	713 ± 13	2	97 ± 4	1 ± 1
RluB D110N	101 ± 46	1	88 ± 4	N/A
RluB E135D	493 ± 140	2	68 ± 3	28 ± 6
RluB H131A	1139 ± 310	3	46 ± 3	22 ± 3 35 ± 2
RluF in RluB	1063 ± 366	3	83 ± 7	21 ± 3
RluF Wildtype	148 ± 25	1	96 ± 4	N/A
RluF D107N	171 ± 25	2	84 ± 3	11 ± 4
RluF R128A	489 ± 92	2	91 ± 5	7 ± 1
RluB in RluF	413 ± 61	1	92 ± 3	N/A

3.2 Purification of RNA Substrates

To reliably analyze pseudouridylation activity of the RluB and RluF binding variants, I generated and purified a radiolabeled RNA substrate to use in activity assays. This substrate (PTC 195) is a fragment of the 23S rRNA of *E. coli* representing the peptidyl transferase center (PTC) (Figure 3.3-A). The substrate is theorized to be large enough to be representative of the 23S rRNA *in vivo*, yet small and stable enough to be readily prepared and radio-labelled *in vitro*. Furthermore, this substrate has previously been used with success with the bacterial pseudouridine synthase RluE [96]. The template for the 195nt PTC RNA was PCR amplified and analyzed by DNA-PAGE (Figure 3.3-B**)**. The expected size of template is 195 nt, and a single

band is observed just below 200 nt at the expected size in the three PCR reactions. Following *in vitro* transcription, the RNA was purified by DEAE gravity flow chromatography to remove abortive transcripts, unincorporated NTPs, and proteins such as RNA polymerase, RNase inhibitor, and iPPase. Figure 3.3 C-D show representative gels of the *in vitro* transcription reaction and DEAE purification of the RNA substrate PTC 195, respectively. The RNA appears to be intact and at the correct size, as a single band occurs just below the size indicator of snR34, which is 203 nt in length. RNA of the proper size (~195 nt) can be observed eluting from the DEAE column with 600 mM NaCl . These elutions were pooled, diluted 4-fold, and precipitated. Following purification, the concentration was determined using UV absorbance and the specific activity determined by scintillation counting of RNA.

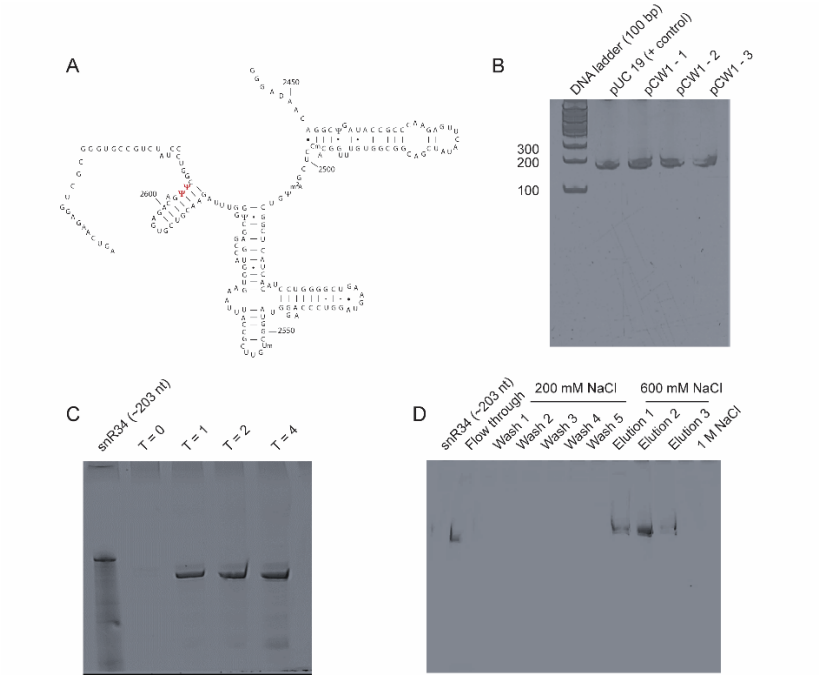


Figure 3.3 – Radiolabeled RNA substrate preparation. (A) Secondary structure of 195 nt RNA substrate derived from *E. coli* 23S rRNA Peptidyl Transferase Centre. Target uridines U2604 and U2605 are shown bolded and in red. (B) PCR preparation of template for *in vitro* transcription from the pCW1 plasmid. (C) Urea-PAGE of the *in vitro* transcription of PTC 195 RNA substrate. T is representative of the time in hours of the *in vitro* transcription at which the sample was taken. (D) Urea-PAGE of samples from the DEAE purification of radiolabeled RNA substrate via gravity flow.

3.3 Experimental Evaluation of Pseudouridylation by RluB Variants

3.3.1 Basic characterization of RluB wildtype, catalytically inactive variant, and negative controls

To test the relative pseudouridylation activity of enzyme variants and compare them to wildtype enzyme activity, single turnover tritium release assays were utilized, in which the concentration of enzyme in solution exceeds the concentration of the RNA substrate. During the modification of uridine to pseudouridine, the C1′-N1 glycosidic bond is cleaved, and the base is rearranged in three-dimensional space to form a C1′-C5 glycosidic bond (Figure 3.4**).**

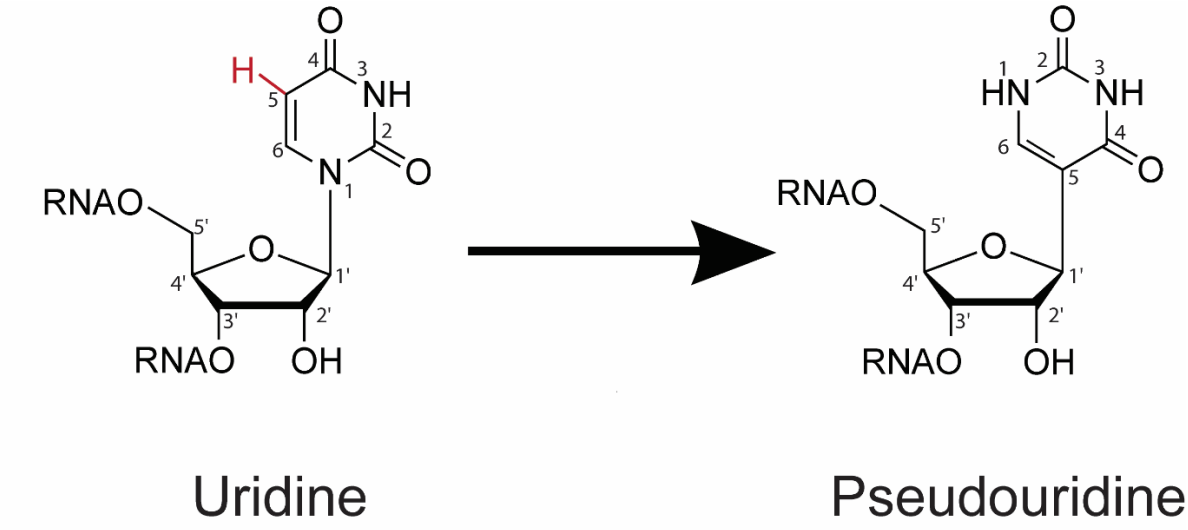


Figure 3.4 – Conversion of a uridine nucleoside to a pseudouridine nucleoside. The N-C glycosidic bond present between the uracil base and the ribose is interchanged with a C-C glycosidic bond in pseudouridine. This creates a free imino proton released from C5 (indicated in red in uridine) which is exploited in tritium release assays.

When the target RNA is transcribed in the presence of [C5-^{3}H-] UTP, one can monitor the release of tritium into the supernatant as the ^{3}H-C5 bond is cleaved and infer a rate of pseudouridylation (see Methods). Each enzyme was tested in biological triplicate for activity against the PTC-195 target RNA, which was generated by *in vitro* transcription using [C5-^{3}H] UTP. The results were averaged and plotted against the time for each experiment (Figure 3.5).

When observing the modification of uridine to pseudouridine via tritium release assays, the wildtype enzyme RluB produced an average of 3.78 ± 0.43 pmol of pseudouridine during a five-minute time-course with an apparent rate of 0.18 pmol/min (Figure 3.5 A). RluB is known to have high specificity for U2605 of the PTC center of *E. coli* 23S rRNA [61, 127]. As such, a modification end level of ~100% is expected. Given that there were 3 pmol of RNA present in each reaction, this yields approximately 126% modification. Based on this information, RluB is active, enabling it to serve as a reliable standard for comparing all enzyme variants.

To investigate RNA binding independently of catalysis, I made use of the nitrocellulose filter binding assay. Through this, I could investigate if an enzyme variant disrupted only the ability of the enzyme to modify its target substrate, or also the ability of the enzyme to bind its target substrate, which would indirectly impact its ability to modify the target uridine. Furthermore, by comparing the variant binding affinities to their respective wildtype enzyme, one may infer the relative importance of a residue. The RluB wildtype enzyme has a K_D of 1.02 ± 0.29 μM for binding the PTC 195 RNA (Table 3.2). This dissociation constant is within a typical range for stand-alone pseudouridine synthases, as those enzymes tested have affinities in the micromolar range. TruB yields a K_D of 2.4 ± 0.3 μM [114], while RluE has a K_D of 0.6 ± 0.2 μM [96].

In addition to analyzing the enzyme variant strains, a baseline for experimental noise had to be established. This was accomplished via a no-enzyme control, which consistently yields a background noise level of 0.14 ± 0.05 pmol of tritium release (Figure 3.5 F). Similarly, in all nitrocellulose filter binding assays, a reaction with 0 μM of t protein was included to ensure a negative control was present. By establishing such baselines, I can be sure that my results are due to the presence of the enzyme itself, and not due to some extraneous factor.

The catalytically inactive variant of RluB, RluB D110N, yields 0.20 ± 0.04 pmol of tritium release after five minutes of incubation (Figure 3.5 B). The trendline is relatively flat over the five-minute time-course, and the picomoles formed are quite comparable to the no enzyme control of 0.14 ± 0.05 pmol. From this, I conclude that the catalytically inactive variant is incapable of modifying the target RNA as predicted. To ensure this inability to modify was due to the mutagenesis and not from an inability to bind the substrate, I determined the binding affinity of RluB D110N for RNA via nitrocellulose filter binding (Figure 3.6 B). RluB D110N has comparable binding to the RNA substrate as the wildtype, exhibiting K_{DS} of 0.85 and 0.42 μM in the two trials conducted (Table 3.2). Notably, D110N reached amplitudes of 7.7 and 6.3 pmol, which are comparable to the wildtype amplitude of 6.7 ± 1.5 pmol. From this, I was able to conclude that the catalytically inactive RluB variant is both capable of binding the target RNA, and incapable of modifying the target uridine, U2605. With these standards in place, I could now begin investigating how each variant generated influenced the RNA binding and modification activity of RluB.

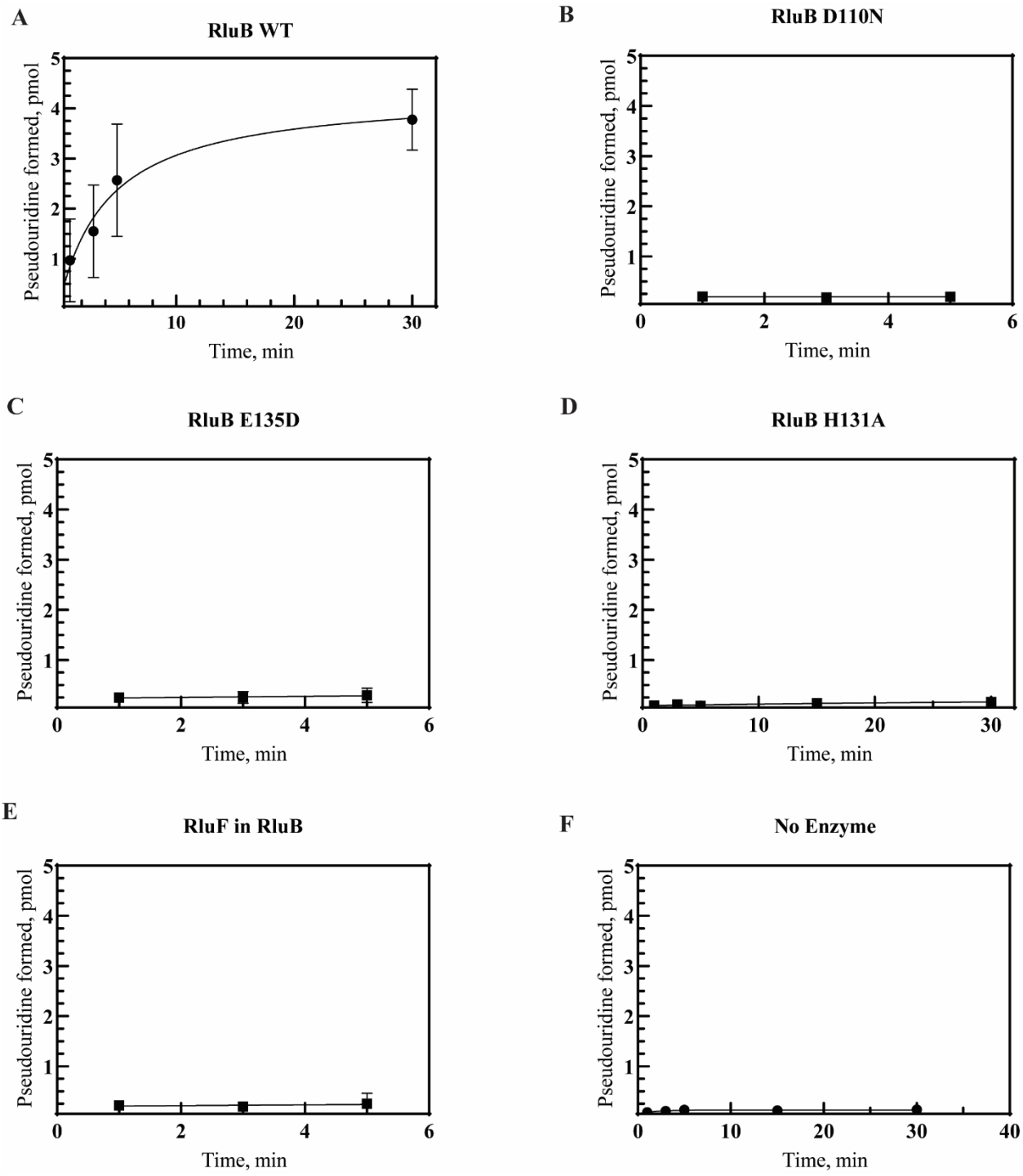


Figure 3.5 – Pseudouridylation activity of RluB wildtype and enzyme variants with PTC 195 substrate. (A-F) Tritium release assay with respective enzyme variants from RluB examining the pseudouridylation activity towards *in vitro* transcribed PTC 195 substrate. Reactions were performed under single turnover conditions (1000 nM enzyme, 300 nM substrate RNA), and graphs depict an average of 3 reactions.

3.3.2 The impact of residue substitutions on the function of RluB

Interestingly, when analyzing the other enzyme variants of RluB, single residue substitutions within the eight-residue RNA binding loop can produce a substantial impact on the enzyme's ability to modify its substrate RNA. For example, when observing RluB E135D, the pmol of uridine converted to pseudouridine in five minutes decreased to 0.29 ± 0.12 pmol of pseudouridine formed compared to the wildtype average of 3.78 ± 0.43 pmol (Figure 3.5 C). The overall trend of modification over time for this enzyme variant is flat, closer resembling the no enzyme control than the exponential trends of an active enzyme. This observation, coupled with the low quantity of pseudouridine formed, is enough to conclude the RluB E135D enzyme variant is inactive. Further investigation into the ability of the variant to bind the target RNA in the nitrocellulose filter binding assay revealed a K_D of 2.76 ± 0.73 μM and an amplitude of 8.6 ± 0.9 pmol of RNA bound (Figure 3.6 C and Table 3.2). This is a two-fold reduction in binding affinity when compared to the wildtype enzyme. Due to the presence of contaminants in the protein preparations of RluB E135D (panel C), caution must be used when drawing conclusions from assays involving these protein preparations. The contaminants may influence binding (and thus pseudouridylation) of the target RNA, skewing the data observed in the assays. Further purification and repeat experiments with these enzymes are required to prove the reduction in binding. Given the data collected, my thesis supports my initial hypothesis: substituting E135 influences RluB's capacity to bind to the target RNA, thereby negating its ability to introduce a pseudouridine into the RNA.

When investigating the role of histidine 131, RluB H131A produces 0.16 ± 0.05 pmol of tritium release over a 30-minute time course (Figure 3.5 D). This is a significant decrease in pseudouridine formed compared to the wildtype, which is to be expected; H131A is a more

drastic modification than a single carbon addition or removal. The trendline for the pmol of pseudouridine formed over time by RluB H131A are almost identical to that observed with the no enzyme control (Figure 3.5-F). As such, this enzyme may be deemed to be inactive. When testing the variant RluB H131A in nitrocellulose filter binding assays, the binding affinity for RNA was expected to decrease. This trend was observed in the data, as H131A demonstrates a K_D of 5.01 ± 1.61 μM (Figure 3.6 D and Table 3.2). This is a five-fold reduction in affinity from the wildtype, which is consistent with the pattern observed with RluB E135D, exhibiting a two-fold decrease in binding affinity and a negation of the enzyme's ability to modify its target. As aforementioned in the discussion of RluB E135D, caution must be used when drawing conclusions from these assays with RluB H131A, as there were contaminants in the protein preparation of (panel D). These contaminants may influence binding (and thus pseudouridylation) of the target RNA, skewing the data observed in the assays. Further purification and repeat experiments with these enzymes are required to prove the reduction in binding. Given the data collected, it is important to note the more drastic substitution of H131A had a larger impact on the binding affinity of RluB, confirming my hypothesis of histidine 131 being influential in RluB's ability to bind the target RNA.

When designing the RNA recognition loop variant by swapping the sequence from RluF and placing it into RluB, the possibility of altered specificity arose. If this were to happen, RluB would then be able to modify both U2604 and U2605. However, the experimental data reveal that the RluF-in-RluB variant is inactive, yielding an end level of 0.25 ± 0.17 pmol of tritium release over a five-minute time-course (Figure 3.5 E). Furthermore, the time-course of this enzyme variant reveals once again a flat trendline, consistent with an inactive enzyme. Based on the level of pseudouridine formed and the flatness of the trendline, this enzyme variant is deemed

inactive. When testing the RluF-in-RluB variant in nitrocellulose filter binding a K_D of 1.21 ± 0.97 μM is observed (Figure 3.6 E and Table 3.2). This K_D is comparable to wildtype binding; however, the amplitude of binding is only 1.6 ± 1.0 pmol, which is comparable to the no enzyme control of 1.2 ± 0.8 pmol. Furthermore, the standard deviations of both the amplitude and determined K_D for the loop variant are quite large, indicating there may not be binding to the RNA substrate at all. Taken together, I can conclude the loop variant of RluF in RluB must be inactive and incapable of binding the substrate.

To summate, my data demonstrate the immense impact a single amino acid substitution has on the enzymatic function of RluB. Changing any of the two chosen residues in the eight-residue RNA binding loop of the enzyme effectively abolishes catalytic activity: each of these changes yield end levels of pseudouridine formation comparable to the no enzyme control. Furthermore, my data have provided a relative importance of these residues in binding the target RNA, thereby providing insight to RluB's specificity requirements.

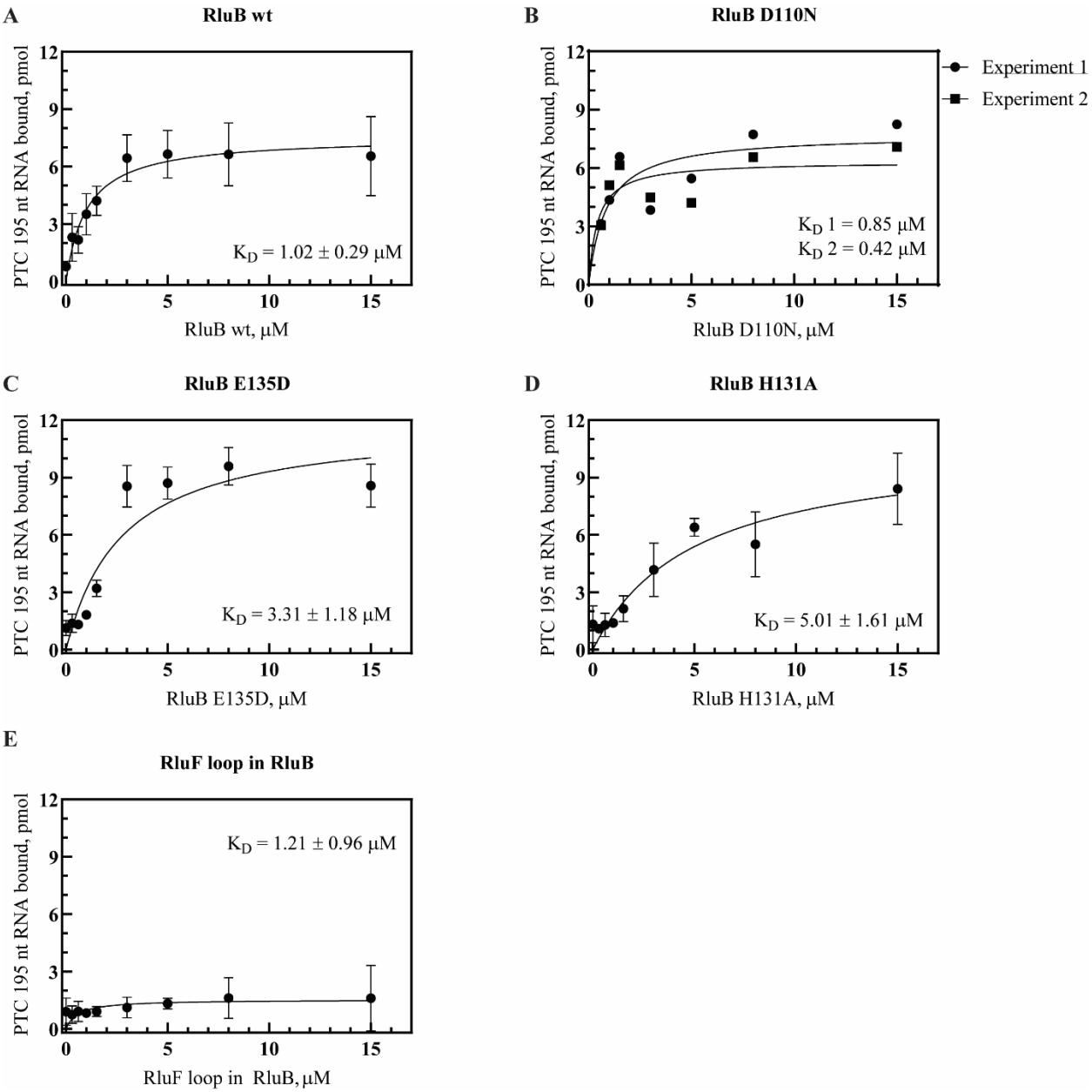


Figure 3.6 – Nitrocellulose filter binding analysis of RluB wildtype and enzyme variants with PTC 195 Substrate. Nitrocellulose filter binding assays were performed using 3 pmol of ^{3}H-labelled PTC 195 with enzyme concentrations spanning from 0-15 µM. Different enzyme variants of RluB were tested **(A-E).** The dissociation constants for each enzyme are listed in Table 3.2. Negative controls without protein yield 1.15 ± 0.75 pmol RNA remaining bound on the nitrocellulose filter.

Table 3.2 - Affinity between RluB enzyme variants and the PTC 195 radiolabeled RNA substrate.

RLUB VARIANT	DISSOCIATION CONSTANT (μM)	BINDING AMPLITUDE (PMOL)
RluB WT	1.0 ± 0.3	6.7 ± 1.5
D110N	0.85* 0.42*	7.7 6.3
E135D	3.3 ± 1.2	8.6 ± 0.9
H131A	5.0 ± 1.6	7.6 ± 1.9
RluF in RluB	1.2 ± 1.0	1.6 ± 1.0
No Enzyme	-	1.2 ± 0.8

***** Individually determined K_Ds, as only two replicates were obtained prior to submission of this thesis.

3.4 Experimental Evaluation of Pseudouridylation of RluF Variants

3.4.1 Standards of RluF with the wildtype enzyme, catalytically inactive variants, and negative controls

The same enzyme activity and RNA binding assays established for RluB were repeated with RluF. When observing the modification of uridine to pseudouridine via tritium release assays, the wildtype enzyme RluF produces 6.08 ± 2.17 pmol of pseudouridine over a five-minute time course with an apparent rate of 0.23 pmol/min (Figure 3.7 A). Given that there were 3 pmol of RNA present in each reaction, this yields a 203% increase in modification levels detected. While RluB is known to have high specificity for only U2605, RluF is also known to have dual specificity; this enzyme shows a preference for U2604 but is also capable of modifying U2605 with reduced efficiency [127]. Bearing this in mind, the end level of 200% modification may be illustrating the modification of the two adjacent uridines at position 2604 and 2605 to pseudouridine. To test this theory, tritium release assays were later performed with

RluF and pre-modified RNA (see section 3.4 for results). Based on the reported levels of modification, I conclude the RluF wildtype enzyme is, indeed, active.

In nitrocellulose filter binding, the RluF wildtype has a K_D of 0.12 ± 0.10 μM for the PTC RNA and an amplitude of 3.3 ± 1.2 pmol (Figure 3.8 A and Table 3.3). As aforementioned, this is within the observed range of binding affinities for pseudouridine synthases, with TruB exhibiting a K_D of 2.4 ± 0.3 μM [114] , and RluE exhibiting a K_D of 0.6 ± 0.2 μM [96]. As with RluB, it is reasonable to assume that RluF is active and produce detectable levels of pseudouridylation in tritium release assays and capable of binding the PTC 195 substrate, serving as a reliable standard to compare the enzyme variants against.

As done with RluB, a baseline for noise was established with a no enzyme control in tritium release assays. By establishing such baselines, I could be sure that my results were due to the presence of the enzyme itself, and not due to some extraneous factor.

When investigating the modification ability of the catalytically inactive enzyme variant, RluF D107N, I observed end modification levels of 0.22 ± 0.05 pmol (Figure 3.7 B). When comparing this to the no enzyme control of 0.14 ± 0.05 pmol of tritium release, in combination with the flat trendline of the modification time-course, one can conclude that this RluB D017N variant is unable to produce pseudouridine in the target RNA. In the nitrocellulose filter binding assay, RluF D107N displays similar binding to the radio-labelled substrate as the wildtype enzyme, revealing a K_D of 0.12 ± 0.08 μM (Figure 3.8 D and Table 3.3). The amplitude of the binding curve is 3.8 ± 0.5 pmol, which is quite comparable to the wildtype amplitude of 3.3 ± 1.2 pmol. The data illustrates that the enzyme is capable of binding, but not modifying, the target RNA in a manner almost identical to the wildtype enzyme. As such, I concluded that RluF D107N was, in fact, catalytically inactive and a reliable standard to compare any enzyme variant against.

3.4.2 The impact of substitutions on the function of RluF

When analyzing the other enzyme variants of RluF, residue substitutions within the eight-residue RNA binding loop produce a considerable impact on the enzyme's ability to modify its substrate RNA. For example, RluF R128A yields 0.40 ± 0.10 pmol of pseudouridine formation over a five-minute time course (Figure 3.7 C). While the modification level after five minutes is higher than the no enzyme control of 0.14 ± 0.05 pmol, the overall reduction from the wildtype production level is substantial. Furthermore, this enzyme variant exhibits a flat trendline over the time-course, suggesting it is an inactive variant instead of a partially active one. When testing the binding variant RluF R128A in nitrocellulose filter binding, the assay revealed the K_D of the R128A variant to be 0.30 ± 0.14 μM (Figure 3.8 D and Table 3.3). This is a comparable affinity to the wildtype enzyme when considering the standard deviation, suggesting that the enzyme is capable of binding the target RNA. Furthermore, this variant has an amplitude of 5.4 ± 1.7 pmol RNA. When compared to the wildtype amplitude of 3.3 ± 1.2 pmol and no enzyme control of 1.2 ± 0.8 pmol of pseudouridine, it is clear that the R128A variant is binding the target RNA with a similar affinity to the wildtype. This does not confirm my initial hypothesis of arginine 128 playing a role in RluF's capacity to bind the target RNA.

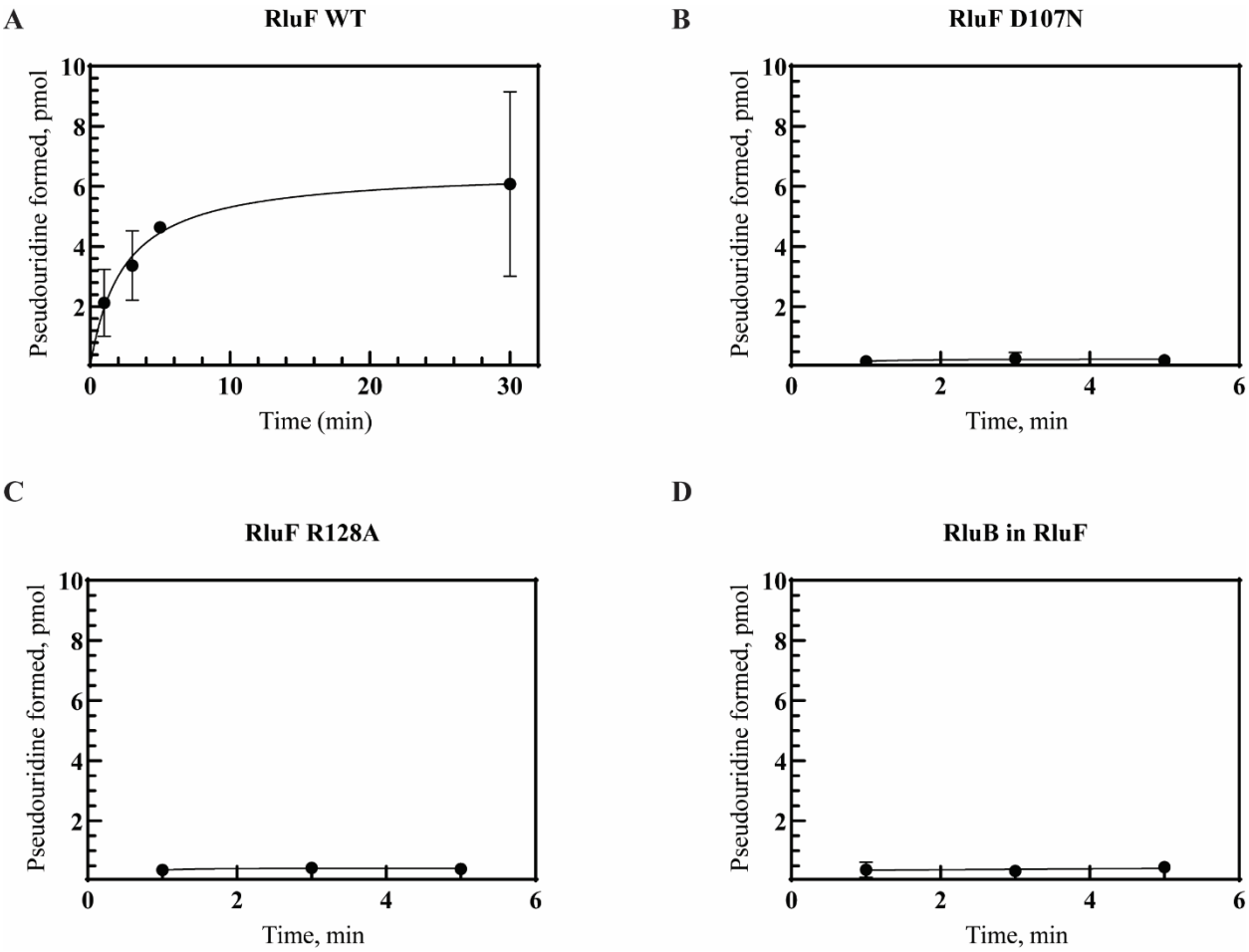


Figure 3.7 – Pseudouridylation activity of RluF wildtype and enzyme variants with PTC 195 substrate. (A-F) Tritium release assay with *in vitro* transcribed PTC 195 substrate examining the pseudouridylation activity of respective enzyme variants from RluF. Reactions were performed under single turnover conditions (1000 nM enzyme, 300 nM substrate RNA) and shown are averages of 3 experiments. Each reaction sample has 3 pmol of radiolabeled substrate RNA present.

When analyzing the RNA modification by the RNA binding loop variant RluB-in-RluF, one observes a decrease in modification from the wildtype to 0.46 ± 0.14 pmol over the five-minute time course (Figure 3.7 D). Like R128A, the relative flatness of the trendline is crucial in determining whether the enzyme is inactive or partially active. The standard deviation on each time point is quite large when considering the magnitude of the numbers, which is likely contributing to the slight upward trend. By considering these in tandem, I can conclude the

binding loop variant is indeed, unable to produce pseudouridine in the target RNA. In nitrocellulose filter binding, RluB-in-RluF yields a K_D of 0.98 ± 0.44 μM with an amplitude of 4.4 ± 0.8 pmol of pseudouridine (Table 3.8 D and Table 3.3). When comparing the K_D of the loop variant to the RluF wildtype, there is an eight-fold reduction, indicating a reduced affinity for the RNA substrate. The amplitude is within the expected range of the wildtype enzyme's; thus, I can conclude that the RluF loop variant is inactive, but capable of binding the target RNA with reduced affinity.

In summary, my data demonstrate the immense impact a single amino acid substitution has on the enzymatic function of RluF. Changing the R128 residue in the eight-residue RNA binding loop of the enzyme effectively abolishes RNA modification activity, as these changes yield end levels of pseudouridine formation comparable to the no enzyme control and display flat trendlines over the time-courses of pseudouridine formation. Interestingly, some RluF variants are still capable of binding their substrate RNA with amplitudes comparable to the wildtype enzyme.

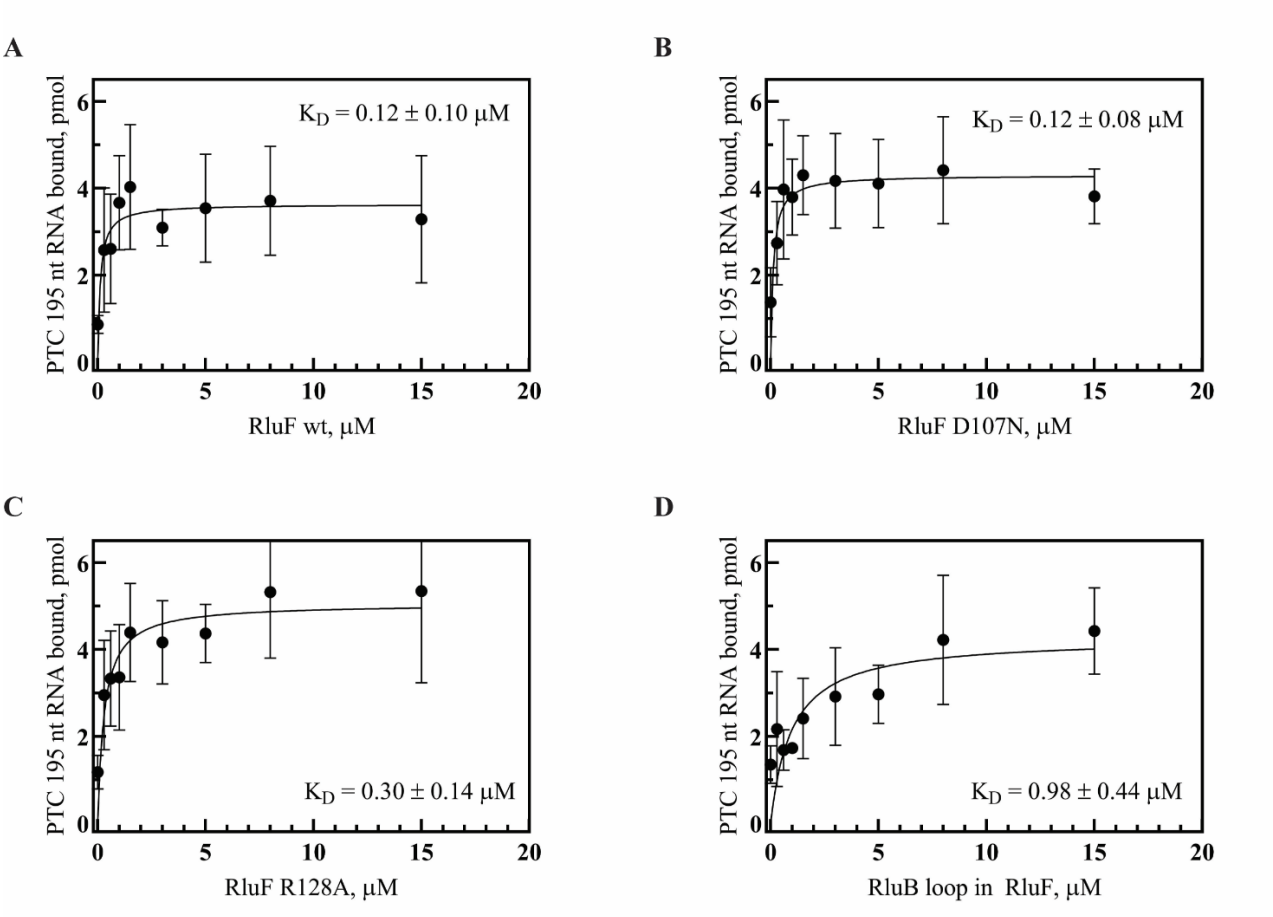


Figure 3.8 – Nitrocellulose filter binding analysis of RluF wildtype and enzyme variants with PTC 195 Substrate. Nitrocellulose filter binding assays were performed using 3 pmol of ^{3}H-labelled PTC 195 with enzyme concentrations spanning from 0-15 μM. Different enzyme variants of RluB were tested **(A-D).** The dissociation constants for each enzyme are listed in Table 3.3. Negative controls yield 1.15 ± 0.75 pmol of RNA bound.

Table 3.3 - Affinity between RluF enzyme variants and the PTC 195 radio labelled RNA substrate.

RluF Variant	Dissociation constant (μM)	binding amplitude (pmol)
RluF WT	0.12 ± 0.10	3.3 ± 1.2
D107N	0.12 ± 0.08	3.8 ± 0.5
R128A	0.30 ± 0.14	5.3 ± 1.7
RluB in RluF	0.98 ± 0.44	4.4 ± 0.8
No Enzyme	-	1.2 ± 0.8

3.5 Investigation of a potential hierarchy of modification

Considering that RluB and RluF modify adjacent uridines in the PTC separated by ~4 Å [126], it is logical to assume they cannot modify simultaneously. As such, the potential for a preferred order of modification arises. As mentioned, RluB is specific for Ψ2605, whereas RluF is not entirely restricted to Ψ2604, harboring the ability to also modify Ψ2605 *in vitro* [126, 127]. I propose that RluB must act first in the hierarchy of modification, because otherwise RluF would be sufficient to modify both target uridines and RluB might have been lost in evolution.

As a first preliminary step towards investigating this likelihood, I conducted tritium release assays with pre-modified RNA. This was accomplished by incubating my target RNA with either RluB or RluF for thirty minutes, and then adding 1 μM of the opposite enzyme and taking a time course for an additional thirty minutes. The addition of one of the wildtype enzymes prior to the other should modify the target uridine(s) in the PTC195 substrate before the time-course begins. In the case of RluB being added first, only U2605 should be modified, leaving U2604 to be modified with the addition of RluF. If RluF is added to the reaction first, both U2604 and U2605 should be modified over the thirty-minute incubation. As such, there should not be an observable

change in the end levels of pseudouridylation after the addition of RluB, as both target uridines should have been modified already. Any deviation from this would be indicative of other uridines being modified and would have to be investigated further. As such, this experiment will indirectly verify that only U2604 and U2605 are being pseudouridylated; tritium release assays are not nucleotide specific, thus, the detection of tritium in solution could indicate the modification of any of the forty-six uridines present in the PTC 195 substrate.

When RluF was incubated in the reaction mixture prior to the addition of RluB, an initial pseudouridylation level of 4.9 ± 1.3 pmol was recorded after thirty minutes (Figure 3.8 A). Given there are three picomoles of RNA in the reaction, this yields approximately 160% modification. Upon adding RluB, this pseudouridylation level remained very constant across the first five minutes, increasing only at the thirty-minute time point to 6.6 ± 1.8 pmol of pseudouridine (~220%). Considering the error and the relative consistency in pseudouridylation levels detected, I can conclude that my hypothesis was correct, and both target uridines were modified prior to the addition of RluB.

When RluB was incubated in the reaction mixture prior to the addition of RluF, an initial pseudouridylation level of 7.48 ± 2.20 pmol was recorded after the thirty-minute incubation (~250% modification) (Figure 3.8 B). Upon addition of RluF, this remains relatively constant over the first five minutes of the time-course, finally increasing to a pseudouridylation level of 10.95 ± 3.92 pmol at thirty-minutes (~365% modification). While the pseudouridylation level are not quite doubled upon addition of RluF, there is a distinct increase (Figure 3.8 D). Taken together, this data illustrates my hypothesis was correct; RluB modified U2605 prior to the reaction time-course, and RluF subsequently modified U2604 with slower kinetics. This is further emphasized by comparing the two time-courses on the same graph; one can clearly

observe the flatness of this case when compared to the doubling of end levels when RluF is added after incubation with RluB.

To further evaluate my data and support my conclusions, I conducted an unpaired student's t-test with the averages of each time-point taken for both time courses (Figure 3.8 D). The P-value is 0.0357; by conventional criteria, my data are considered to be statistically significant. To summate, I have shown that RluB and RluF are modifying the target U2604 and U2605 in the 195 nucleotide RNA substrate and not any of the other existing uridines.

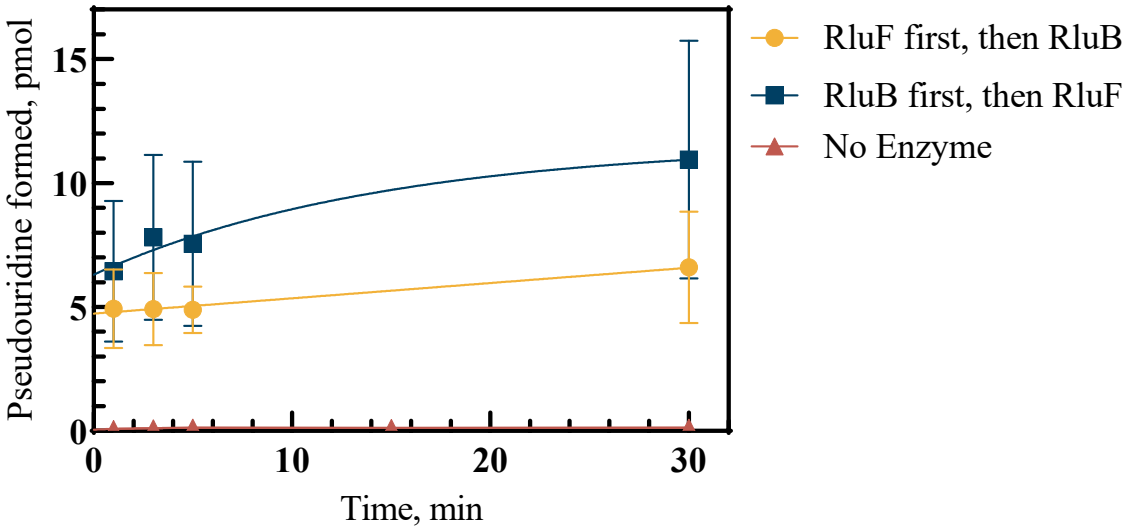


Figure 3.9 – Pseudouridylation activity of RluB and RluF with pre-modified RNA substrate. Tritium release assay of wildtype enzymes RluB and RluF examining the pseudouridylation activity of pre-modified *in vitro* transcribed PTC 195 substrate. Reactions were performed under single turnover conditions (1000 nM enzyme, 300 nM substrate RNA) and shown are averages of 3 experiments. Each reaction sample has 3 pmol of radiolabeled substrate RNA present. **Yellow circles** = RluF incubated with the RNA substrate for thirty-minutes prior to the addition of RluB **Blue squares** = RluB incubated with the RNA substrate for thirty-minutes prior to the addition of RluF **Red triangles** = No enzyme control.

CHAPTER 4: DISCUSSION

4.1 Optimizations, Benefits, and Disadvantages of this study

To investigate the required structural elements for RluB and RluF's function, nine proteins and variants thereof were purified. Throughout the purification process, a number of optimizations had to be completed, while some steps still require further optimization in the future. One of the most difficult complications I had was the production of RluF; the first overexpression yielded an inactive enzyme, and further attempts at overexpression failed. I re-transformed the plasmid containing RluF into *ME5305 E. coli* cells, and after various test expressions, I was successful in overexpressing the wildtype enzyme in large scale. A crucial change to this expression protocol was to grow the cells to an OD_{600} of ~0.8, induce with 0.25 mM IPTG, and then express the cells overnight (~15 hours) at 18°C. The benefits of these changes are four-fold: a larger quantity of cells to produce protein, a decrease in the metabolic drag of the cells, an increase in protein solubility, and circumvent the potential toxicity of RluF overexpression as further explained below.

Firstly, by allowing the cells to grow to a higher optical density, there are a higher number of viable cells to produce the protein of interest. Having more cells present to produce protein post-induction should increase the final yield of protein from the cells after lysis [146]. Furthermore, it has been found that cells later in the exponential phase of growth experience a reduced metabolic cost of producing excess protein [146, 147]. As such, these cells are more likely to produce an abundance of the protein of interest. Secondly, IPTG reduces the growth rate of *E. coli* cells by inducing the lac operon and activating protein expression. Rather than energy being allocated toward cellular growth, it is instead directed to the production of the protein of interest placed in the operon [147-149]. A lower concentration of IPTG will reduce the 'metabolic drag' on the cells and allow them to produce the protein more slowly [146]. Thirdly, expressing for a

longer period of time and at a reduced temperature promotes proper folding of the protein. As temperature decreases, so does the rate of cellular functions; therefore, at a lower temperature, translation speed is reduced [150]. By slowing down translation, there are more opportunities for chaperones to bind to and assist improperly folded proteins. Furthermore, low temperature and longer induction will reduce the production of inclusion bodies [151, 152]. Solubility and protein folding are directly related; a properly folded protein will have increased solubility, and thus, will be more likely to be overexpressed in its active form.

Lastly, there arises the potential of RluF overexpression becoming toxic to *E. coli* cells. As observed in nitrocellulose filter binding studies, RluF displays a high affinity for its rRNA target (0.12 ± 0.10 μM). At high concentrations of RluF, maturing rRNA may be bound tightly by RluF, thereby inhibiting rRNA release, folding and protein binding with an overall negative effect on ribosome biogenesis. By overexpressing RluF with a lower IPTG concentration, cells are permitted to grow in lower concentrations of RluF, and thus fewer maturing rRNAs will be bound and sequestered. Therefore, the toxic effect of RluF overexpression can be circumvented by simply reducing the concentration of RluF present in the rapidly dividing cells. Furthermore, should this high affinity be observed between RluF and its tRNA target, the possibility of non-specific binding and sequestering of tRNAs also arises. Should this occur, the pool of available tRNAs (whether they be $tRNA^{Tyr}$ alone or a combination of non-specific tRNA binding by RluF), may be greatly reduced, slowing (or altogether stopping) translation and protein production. Without the ability to produce essential proteins cells cannot survive. In a similar fashion to rRNA binding, this potential issue can be bypassed by simply reducing the concentration of RluF in the *E. coli* cells producing the protein, effectively reducing the amount of tRNA bound to RluF.

Next, these enzymes had to be purified for use in later biochemical assays. During the purification process, RluB E135D and H131A proved difficult, as multiple bands were present after SEC. Furthermore, after concentrating the protein, these bands persist. RluB E135D showed two bands present in the final gel, and H131A showed three bands present in the final gel. Due to the clustering of the bands around 35 kDa, I can conclude that these bands are not the result of dimer and/or trimer formation. Thus, two likely possibilities arise: I have co-purified contaminants, or I have a truncated/degraded version of the protein present just below the 35 kDa mark.

Regarding the former case of co-purification, the most common, general *E. coli* co-contaminants are Elongation Factor Thermal unstable (EF-Tu), GroEL/ES, and DnaK [153-155]. The sizes of these proteins are 43 kDa (Uniprot P0CE47), 57 and 10 kDa (Uniprot P0A6F5 AND P0AF59), and 69 kDa (Uniprot P0A6Y8), respectively. Due to the difference in size of these common co-contaminants from the band size near 35 kDa, it is unlikely that these are the cause of the extra bands. Furthermore, the SEC should have removed these contaminants during this purification step. Consequently, it becomes more likely that the secondary band is a result of a truncated version of my protein. Since the histidine tag is on the N-terminus of the protein, these truncated versions would still bind to the nickel sepharose column and be present during SEC. Due to the closeness in size of these bands on my gel (notably less than 10 kDa), it is likely that the size exclusion column would not separate these. This could be optimized in future studies by evoking the optimizations made with RluF: grow the cells to an OD_{600} of 0.8 instead of 0.6 before inducing, induce with a lower concentration of IPTG, and express overnight at 18°C. Ultimately, this should reduce the cellular stress placed on the expression cells and minimize premature termination, hopefully eliminating the purity problems I have observed in my study.

Finally, optimization of my target RNA substrate was required. Obtaining a high specificity (DPM/pmol uridine) of my target RNA proved to be difficult, although the reasons behind this are unclear. Since the specific activity is critical for the analysis of both tritium release and nitrocellulose filter binding assays, a higher specific activity facilitates obtaining a detectable signal in these assays. I overcame this issue by increasing the RNA concentration in each reaction from 100 nM to 300 nM, effectively tripling the observed signal in the assays. In doing this, I maintained single-turnover conditions by having notably more enzyme than substrate (1000 nM vs 300 nM) and increased my signal to ensure I was not measuring in the background noise of radioactive counts.

To summate, I optimized the synthesis and purification of the biomolecules required for the *in vitro* characterization of my enzymes of interest. In doing this, future studies of RluB and RluF will be efficient. Additionally, these optimizations can be applied to other expression systems, particularly those of other *E. coli* RNA modification enzymes.

4.2 Characterization of the RNA Binding Loops of RluB and RluF

Czudnochowski *et al.* [127] described three major conformational changes occurring in RluB upon RNA binding. First, the S4 domain becomes rigid and binds to the major groove of the RNA substrate. Next, two subdomains around the active site cleft undergo a rigid hinge motion to close around the RNA (similar to motions that are observed in RluF and TruB) [126, 141]. Finally, residues 132-134 of the eight-residue binding loop refold in a 3_{10} helix to bind the RNA near the A2602 bulge. This third conformational change becomes the most interesting in the context of this thesis. Such 3_{10} helices have been previously found to have functional importance, as the R groups of amino acid residues are restricted and form ridges along the helix [142].

Interestingly, the RNA binding requirements of RluB and RluF differ only in how they contact the RNA from G2597 to A2602, as these enzymes form corresponding hydrogen-bonding interactions with A2590 to U2596 [127]. As such, it is logical to assume that the specificity arises from residues in the enzyme that bind to the region of G2597 to A2602. Not surprisingly, this interaction involves the conserved, eight-residue RNA binding loop among the enzyme families. This region is shifted 3-4 Å in RluF (relative to RluB) as the A2602 bulge is incorporated into the stem-loop [127].

Evaluating the combined data of a variant's impact on RNA binding and modification abilities is essential to reach an accurate depiction of a residue's influence on pseudouridylation. If a variant is incapable of binding its substrate, obviously it cannot modify the same substrate. Similarly, if a variant binds the RNA too tightly, a low modification level may be observed; if product release is the rate limiting step, then a high RNA affinity will yield slow modification. These are not the only two possibilities: perhaps an enzyme variant yields an enzyme capable of binding the substrate RNA, but its ability to properly orient the target base for modification is impaired. Similarly, an enzyme may be able to bind the target RNA but be unable to flip the target uridine base out of the RNA loop for modification. This thesis does not cover the last possibility but does address all former possibilities and relate them to the change in structure and interactions observed with the 3_{10} helix and the A2602 bulge. Interactions of the RNA with the S4 domain and hinge motions are not addressed here, as the enzyme binding variants generated are unlikely to impact these processes.

4.2.1 Evaluation of binding variants on activity of RluB

Czudnochowski [127] found that RluB stabilizes the A2602 bulge with a collection of hydrogen bonds, primarily from histidine (H) 131, serine (S) 133, glutamate (E) 135, and

arginine (R) 138. Several water molecules were also found to participate in this stabilization to expand the hydrogen-bonding network and increase the stabilization effects [127]. I substituted residues E135 and H131 of the eight-residue binding loop, which flank either side of the 3_{10} helix observed to re-fold upon RNA binding by RluB [127]. Since these variants maintain the catalytic aspartate, it is logical to assume the changes made to these residues will influence only binding and orientation of the RNA substrate for catalysis, and not catalysis itself.

Glutamate 135 is the only residue of the binding loop in direct hydrogen-bonding contact with A2602 (PDB 4LGT); a hydrogen bond forms between the carboxyl group of glutamate and the amino group of C6 in adenine [156], ultimately stabilizing the base of the bulge [127]. As such, substituting this residue is expected to disrupt hydrogen bonding with the bulge, the influence of which will provide insight to the importance of stabilizing this bulge for the enzyme's RNA binding specificity. Glutamate 135 is conserved among the RluB family, and the equivalent residue in the RluF family is aspartate which is proposed to be too short to make the hydrogen bonding contact with A2602. The sole difference between the residues arises from the additional methylene group on the glutamic acid side chain. Based on the data collected, losing the contact between glutamate's carboxyl group and the RNA decreases the binding affinity two-fold (Table 3.2). However, based on the drastic reduction in activity for this a single-residue substitution, one may assume the interaction with glutamic acid is crucial for positioning the target RNA in the correct orientation for catalysis. Given that glutamate 135 has been shown to participate in direct hydrogen bonding with the A2602 bulge [127], it is reasonable to assume this hydrogen bonding plays a considerable role in positioning the target uridine for catalysis. Thus, maintaining the hydrogen bond with the A2602 bulge and thereby stabilizing this orientation, is crucial for modification of the target uridine.

Histidine 131 flanks the opposite side of 3_{10} helix than glutamate 135 in RluB [127]. It also partakes in the stabilization of the A2602 bulge in the target RNA. H131A is a more drastic modification than a single carbon addition or removal; histidine and alanine are quite different amino acids. Not only is there a removal of the aromatic ring of the histidine, but histidine is also a polar, charged amino acid, while alanine is non-polar and uncharged. Furthermore, alanine has a high tendency to form helices and has been found to induce structure [145]. As such, this variant was expected to have a notable decrease in RNA binding, which is exactly what was observed (Figure 3.6 and Table 3.2).

When considering the notable reduction in catalysis and the five-fold decreased binding affinity for RNA of this variant, it is likely that the reduction in modification arises from a combination of the enzyme's inability to bind to its substrate tight enough for modification to occur and/or to properly orient the RNA within the catalytic pocket. As discussed with E135D, it is likely that substituting residues near the binding site of A2602 alters the orientation of the target RNA while bound to the enzyme. Histidine 131 binds to the bottom of the A2602 bulge, effectively stabilizing the RNA structure within the active site. Based on the reduction of RNA binding and the removal of the charge, one may infer that the positive charge on His131 plays an important role in binding the negatively charged RNA substrate to RluB and stabilizing the A2602 bulge via hydrogen bonding. Furthermore, this bulge has been previously found to pack against the loop that connects α1–β4, establishing contact with proline 132 and serine 133 of the conserved RNA binding loop [127]. Perhaps the introduction of alanine alters the structure of this loop enough to disrupt the Van der Waals contacts with P132 and S133, especially because alanine is known to induce helical formations. When considering the induction of structure in the presence of alanine, it becomes possible that histidine has been conserved in this location

throughout RluB evolution to allow formation of the 3_{10} helix. Since alanine induces structure through the formation of helices, it is reasonable to assume the presence of alanine at the begging of the 3_{10} helix prevents the structural rearrangement from occurring. Thus, as RluB undergoes this conformational change, its substrate becomes correctly oriented and stabilized to allow for catalysis. Without the formation of this 3_{10} helix, the substrate cannot be correctly positioned for catalysis. Consequently, the role is histidine 131 may be multifold: (i) this residue could allow for formation of the 3_{10} helix, thereby correctly positioning the RNA for catalysis, (ii) this residue may stabilize the bulge formed at A2602 through hydrogen bonding to ensure U2605 is flipped into the active site.

When designing the RNA recognition loop variant by swapping the sequence from RluF and placing it into RluB, the possibility of altered specificity arose. However, the experimental data reveal that this variant is inactive, and incapable of binding the RNA substrate. A possible explanation for this is that the RluF binding loop residues are not in the correct conformation in the context of the RluB enzyme.

Pseudouridylation has been found to be uniformly slow in bacterial tRNA pseudouridine synthases [144]; while the mechanistic steps for modification include binding, base flipping into the active site, catalysis, base flipping back into the RNA structure, and RNA release, catalysis has been found to be the rate-limiting step. Thus, for pseudouridylation to occur, the enzyme and substrate must stably bind, or the RNA may dissociate from the enzyme before catalysis has occurred. In this scenario, the enzyme would still bind to its target, but pseudouridine would not be detected in tritium release assays. Notably, the tritium release assay has lower concentrations of enzyme than in nitrocellulose filter binding and lower concentrations than the determined K_D for RNA binding. Thus, in the current tritium release assay conditions, the enzymes will bind

RNA only very inefficiently, but binding could be improved by increasing the enzyme concentrations to 10 μM. At 10 μM, the enzyme concentration is four times the K_D, and the RluB variants should bind the RNA allowing these variants to possibly modify the RNA and introduce pseudouridine at the target site.

To summate, combining my data with that of Czudnochowski's reveals that E135 and H131 in RluB are critical for the binding and stabilization of the RNA bulge. Furthermore, manipulating these residues may abolish formation/stabilization of the 3_{10} helix, which changes the active conformation of the enzyme. Without this, RluB may lose the ability to orient its target for catalysis.

4.2.2 Evaluation of binding variants on activity of RluF

RluF also has a conserved eight-residue RNA binding loop (residues 128-135) analogous to RluB's, although its sequence is distinct from RluB (Figure 3.1). These residues are equivalent to residues 131-138 in RluB, even achieving the same protein fold [126, 127]. However, the differences in the side chains in RluF have been found to be incapable of stabilizing the A2602 bulge in the target RNA.

RluF reliably illustrated 200% RNA modification (6 pmol of pseudouridine formation) over a 30-minute time-course in tritium release assays (Figure 3.7 A). It is worthwhile to note that RluF has been previously observed to modify U2605 with reduced efficiency compared to U2604 [126]. This trend is consistent in my data, as one can only see 200% modification after 30 minutes of incubation. From this, I hypothesize that within the first five minutes, only U2604 is modified, as it is the preferential and efficient target of the enzyme. After this, the enzyme likely proceeds to modify U2605, albeit at a reduced rate. This explains why it takes an additional 25 minutes to reach double the levels of pseudouridine. My hypothesis is further proven in the

tritium release assays with pre-modified RNA (see section 4.2.4), in which RluF can modify both target uridines prior to addition of RluB (Figure 3.9).

The enzyme variant RluF R128A was designed to disrupt binding with the target RNA backbone; this residue has been shown to be in contact with the backbone of C2601-G2592 base pair (Figure 3.1), as well as positioned to form a hydrogen bond to the ribose O2′ of A2602 [126]. Based on the nitrocellulose filter binding assays, the K_D does not change significantly, thus there is not a strongly reduced affinity of the enzyme for the substrate RNA. Binding the RNA backbone is not expected to provide much specificity for RluF, thus, it seems logical that there is not a notable change in binding affinity. However, it is curious that this lack of change in binding affinity does not result in modification activity in the tritium release assay. From this, three likely scenarios arise: the enzyme requires very tight binding for uniformly slow catalysis to occur and the flip out the target base; R128 plays a role in positioning the RNA at the correct angle for modification through its hydrogen bonding; or a combination of the two effects. I hypothesize a combination of the two effects is most likely, as a single residue substitution should not be enough to completely abolish binding (which is true based on my data), and proximity and orientation are well known factors for enzyme catalysis. Furthermore, larger RNA substrates are more likely to be influenced by point mutations; large substrates will bind along a larger surface of an enzyme than a small substrate. As such, small changes further away from the active site interaction can influence the positioning of a substrate [114]. Furthermore, R128 is proposed to be critical for repositioning of A2602 for incorporation into the stem-loop [126].

A similar helix to that observed in RluB forms after the α-1 helix in RluF, and the A2062 bulge lies beneath this [127]. Arginine 128 lies at the beginning of the helix formed, thus, a substitution with alanine could influence the formation of this structure. Since alanine induces

structure through the formation of helices, it is reasonable to assume the presence of alanine at the beginning of the helix prevents the rearrangement observed in Czudnochowski [127] from occurring. Thus, as RluF undergoes this conformational change, its substrate becomes correctly oriented and stabilized to allow for catalysis. Without the formation of this helix, the substrate cannot undergo rearrangement to be correctly positioned for catalysis. Therefore, I propose the role of arginine 128 to be two-fold: (i) this residue may allow for formation of the helix, thereby correctly positioning the RNA for catalysis and ensuring room for the stem-loop rearrangement (ii) this residue could also stabilize the backbone of the nucleotides adjacent to A2602 via hydrogen bonding to ensure U2604 is flipped into the active site.

When analyzing the modification of the RNA binding loop variant RluB in RluF, one observes a decrease in modification to 0.46 ± 0.14 pmol. Due to the closeness of this modification level to noise, coupled with the relative flatness of the trendline, one may conclude this variant is inactive. When comparing the K_D of the loop variant to the RluF wildtype there is an eight-fold increase, indicating a notable reduction in binding affinity. The data suggests the binding affinity is not strong enough for RNA binding to occur under the conditions of the pseudouridylation assay which should be repeated at high enzyme concentrations above the K_D.

In brief, the data shown in this thesis illustrates the importance of R128 in positioning and stabilizing the target RNA in proximity to the active site. When this residue is substituted, the RluF enzyme loses the ability to orient its target base for catalysis, perhaps due to the inability to form the helical structure observed upon RNA binding, which Czudnochowski [127] has shown changes the active conformation of the enzyme.

4.2.3 *Comparing the requirements of RluB and RluF*

Similar trends to the RluB results were observed with RluF binding variants: a single amino acid substitution within the RNA binding loop drastically reduced the enzyme's modification levels. This reduction is assumed to arise from an inability to properly orient the target RNA for catalysis in the case of RluB and RluF variants. Interestingly, the RluF enzyme variants did not exhibit the same trend as the RluB variants in the nitrocellulose filter binding assays; a single amino acid substitution did not impact binding to nearly the same degree. This is most clearly illustrated when comparing the RluB E135D/H131A binding variants to the RluF R128A variant. The RluB variants illustrated a two and five-fold increase in K_D with a single residue substitution, while R128A did not exhibit a notable increase in K_D. This trend is likely due to the relative importance of the E135 and H131 residues of RluB in stabilizing the A2602 bulge; a more important role in stabilizing the bound RNA is reflected in a more drastic change in K_D upon substituting an interacting residue. Since RluF does not stabilize the A2602 bulge and binding to the RNA backbone does not confer specificity, it makes sense that R128A did not have an influence on biding affinity.

Another notable difference that arises is in the K_Ds of the wildtype enzymes: RluF binds its substrate nine times more tightly than RluB. Enzymes bind RNA in a specific conformation in an environment of optimized stabilization residues [157, 158]. While binding the RNA substrate to the enzyme will be favorable overall, there is an additional energy barrier between free RNA and bound RNA when these RNAs have different conformations [159]. The RNA conformational change may be unfavorable requiring energy; however, the overall free energy state of the complex of enzyme and rearranged RNA will still be favorable compared to free enzyme and free RNA. As such, the enzyme may use a portion of the binding energy for the

conformational change of RNA to occur [158]. In the context of RluB and RluF, the relative importance of binding energy makes sense; as the secondary structure of the RNA changes upon binding to RluF, the energy input for RNA rearrangement will increase, and a larger allocation of binding energy must be given to achieve the conformational change of the RNA bound. Since RluB stabilizes the A2602 bulge as found in free RNA, while simultaneously flipping out the target uridine for modification, there is less rearrangement of RNA required than with RluF (which induces a stem-loop rearrangement in addition to base-flipping for catalysis). Thus, it makes sense that with a larger conformational change in the substrate RNA, RluF has developed tighter affinity for the RNA to be able to induce the conformational change.

4.2.4 *Modification Sites of RluB and RluF*

Considering that RluB and RluF modify adjacent uridines in the PTC separated by ~4 Å [126], it is logical to assume they cannot modify RNA simultaneously. As such, the potential for a preferred order of modification arises. As mentioned, RluB is specific for Ψ2605, whereas RluF is not entirely restricted to Ψ2604, harboring the ability to also modify Ψ2605 *in vitro* [126, 127]. Furthermore, the tritium release assays used to assess catalytic activity of the enzymes are not specific for the target uridines; the modification of any of the existing uridines within the target RNA will result in a detectable signal from the assay. As such, it is important to confirm that only U2604 and U2605 are being modified during these time-courses.

When RluF is added to the reaction first, both U2604 and U2605 should be modified over the thirty-minute incubation (Figure 3.9). As such, the time-course should begin at 6 picomoles of pseudouridine and there should not be an observable change in the levels of pseudouridylation after the addition of RluB, as both target uridines should have been modified already. Any deviation from this would be indicative of other uridines being modified by RluB and would

have to be investigated further. The expected trend and values for pseudouridine produced are observed in the data (Figure 3.9); however, while the trendline is relatively flat, a slight increase in pseudouridine formation is still observed. This is likely because RluF modifies U2605 with reduced efficiency, and there are still some uridines left to be modified after the addition of RluB. Thus, when RluB is added, these remaining uridines are modified, releasing tritium to be detected. It is most important to focus on the overall trendline of the time-course, however, which is primarily flat.

In the case of RluB being added first, only U2605 should be modified, leaving U2604 to be modified with the addition of RluF. Consequently, one would expect the trendline to begin at 3 picomoles of pseudouridine and increase to 6 picomoles over the thirty-minute time course. This trend is observed in the data, as there is a notable increase in pseudouridine detection following the addition of RluF into the reaction (Figure 3.9). These observations support the previous reports that RluB is specific for U2604, and RluF is specific for U2605. Notably, the level of pseudouridine detected in this time-course is quite high and do not correspond to those predicted; six picomoles of pseudouridine formed by RluB corresponds to 200% RNA modification (Figure 3.7 A). One possible explanation of this observation could be that RluB has a second target within PTC 195, modifying both of these uridines prior to the addition of RluF. However, such levels of pseudouridine formation are not observed in tritium release assays with only RluB (reaches 3.77 pmol of pseudouridine formed over a thirty-minute time-course) (Figure 3.5), nor are they observed in the experiment in which RluB is added post incubation with RluF. Alternatively, the observation of apparent 200% RNA modification after incubation with RluB could be an artifact of the experiment, which may be partially explained by the lack of a 0-minute data point. The first data collected for pseudouridylation occurs at 1-minute post addition

of RluF. When observing the pseudouridine formed by the RluF wildtype alone, 0.97 ± 0.83 pmol of pseudouridine is formed after 1 minute. Thus, it is conceivable that RluB generated approximately 3 pmol of pseudouridines during the 30 minute incubation and that RluF produced another 1 pmol of pseudouridine in the 1 minute incubation period after addition of RluF , resulting in an overall level of at least 4 pmol similar to the observed level (Figure 3.9). This will need to be confirmed with replicate experiments of the enzymes with pre-modified RNA in which a 0-minute data point is collected **before** the addition of the opposing enzyme.

Further error in the pre-modified RNA experiments likely arose from the differences in affinity for rRNA between RluB and RluF; RluF was found to bind the target RNA with a significantly higher affinity than RluB (0.12 ± 0.10 μM for RluF vs 1.0 ± 0.3 μM for RluB). In the assay, there are 1 μM of both RluB and RluB present in the reaction along with 0.3 μM of RNA. As such, RluF can be expected to bind the majority of the RNA and potentially sequester the target RNA from being modified by RluB. This can be confirmed by performing the experiment with 2 μM of RluB (double the original concentrations of enzyme) to ensure equal binding between the enzymes and the target RNA. Further experiments could also be conducted in which the first enzyme is removed (e.g. by phenol-chloroform extraction) prior to the addition of the second enzyme to the pre-modified rRNA to remove steric competition and the potential for RluF to sequester the target RNA from RluB [160].

To summate, this experiment indirectly verifies that only U2604 and U2605 are being pseudouridylated; tritium release assays are not nucleotide specific, thus, the detection of tritium in solution could indicate the modification of any of the forty-six uridines present in the PTC 195 substrate. Furthermore, this assay confirms that RluF is capable of modifying U2605 with reduced efficiency.

4.3 Physiological Function of RluB and RluF

Bacteria are renowned for their ability to live and function in a diverse range of environments: extreme temperatures, high salt concentrations, little to no oxygen, nutrient source(s), competition, etc. [161]. As the conditions change rapidly, so too must the bacteria, most commonly achieved through transcriptional and translational regulation of protein expression. Many of the rRNA modifications in bacteria ensure the ribosome remains functional under the changing stress conditions, thereby promoting cell survival.

Unlike known stress-response proteins [162-166], RluB and RluF are thought to be produced constitutively. In particular, these proteins are expressed under standard, non-stress lab conditions, although they may not directly contribute to *E. coli* fitness under these ideal growth conditions. Due to their modification sites being buried deep within the PTC of the ribosome, these enzymes must be expressed during periods of ribosome biogenesis to reach the target modification sites at U2604 and U2605. Since it is almost inevitable *E. coli* cells will encounter an environmental stressor during their lifetime, ensuring ribosome stability and protection is crucial. RluB and RluF have been identified to play a role under minimal media stress conditions (Vienneau, unpublished); these enzymes play a role in stabilizing the ribosome when nutrients are depleted, such as in late-exponential and stationary phase after cells have been rapidly dividing and depleting nutrients in the environment. Consequently, *E. coli* cells express RluB and RluF under non-stress conditions during ribosome biogenesis to ensure the rRNA of the ribosome is pre-emptively modified to protect the ribosome under the expected later stress conditions of depleted nutrients when ribosome biogenesis is halted, and cells rely on previously produced ribosomes.

This hypothesis could potentially be investigated by fluorescently tagging RluB or RluF with GFP or mCherry to investigate the stage at which RluB and RluF are produced during

bacterial growth. Plasmids containing a fluorescent reporter fusion with *rluB* or *rluF* could be constructed and transformed, and the fluorescent levels monitored during growth in LB and M9 minimal medium. Suitable controls of constitutively expressed proteins such as *rssA* (encodes for 16S ribosomal RNA) [161] and a minimal media stress protein such as *stpA* (encodes for a DNA binding protein) [167] would also need to be evaluated. One would expect levels of the fluorescent reporters of RluB and RluF to increase proportionately to the rate of cell growth and the levels should not change between the two medias, comparable to the positive control, as RluB and RluF should be expressed constitutively and introduce pseudouridine pre-emptively into the ribosome. Contrarily, the minimal media stress condition control should have low levels of production of the fluorescent reporter (as *stpA* is produced transiently in rich medium [167]), and increased production proportionate to the rate of cell growth when grown in M9 minimal medium.

To investigate the effect of Ψ2604 (RluF) and Ψ2605 (RluB) on translation under nutrient depletion, plasmids containing a lacZ promoter and *rluB/rluF* knockout strains (wildtype, single knockout, and double knockout) could be constructed to monitor the production of β-galactosidase under IPTG presence in M9 minimal growth medium. By monitoring the production of β-galactosidase with and without the presence of pseudouridine in rRNA at a given position, one could infer the role of pseudouridine at these positions during translation.

4.4 Future directions:

The work in this thesis displays an expansion on the characterization of the pseudouridine synthases, RluB and RluF. An enticing next avenue for the characterization of these enzymes would be to investigate how these variants influence the base flipping of the A2602 bulge via fluorescent binding assays. Czudnochowski previously reported activity with RluB and a 21-nt ^{3}H-UTP-labelled hairpin [127], while Alian detected activity with RluF and a 22-nt ^{3}H-UTP-

labelled hairpin [126]. As such, it is a reasonable assumption that a small hairpin will bind to both enzymes and provide reportable fluorescence, e.g. when using a 3′ fluorescently labelled hairpin. The fluorophore is expected induce a fluorescent change upon binding of the protein to the RNA due to a change in the fluorophore environment which should be located close to the protein surface thus, providing insight to the influence of each RNA binding loop variant. Alternatively, one could use an RNA hairpin labelled with 2-aminopurine RNA at the A2602 bulge. 2-aminopurine alters fluorescence upon a shift in base stacking and bears structural similarity to adenine [168, 169]. Thus, by placing 2-aminopurine at the A2602 bulge one will be able to compare the base flipping of A2602 upon binding of the RNA to a pseudouridine synthase. This may be exploited for the study of target specificity and A2602 base flipping with RluB and RluF, and thus used to infer which element(s) of the binding variants are essential for target selection. When A2602 exists as a bulge in the stem-loop, it is un-paired and should not cause a strong fluorescent change in substrates labelled with 2-aminopurine when binding to RluB. Conversely, when A2602 exists in the stem loop and is base-paired, there will be a fluorescent shift in substrates labelled with 2-aminopurine when binding to RluF. By exploiting these known mechanisms, further insight to the mechanism of selection of these enzymes may be gained.

A secondary future avenue of study arises in studying how the variants of RluF influence its target RNA selection in tRNA; RluF is known to also modify U35 of $tRNA^{Tyr}$ [140]. Investigating the influence of these binding residues in RluF for binding and modifying the tRNA substrate would confirm if the same RNA binding loop is used between different targets and will provide more insight to the design of specificity of this enzyme.

When confirming the modification sites of RluB and RluF, the potential for a hierarchy of modification arose. For RluB to be able to modify U2605 it should, theoretically, act prior to RluF to prevent both target uridines being modified by RluF. To confirm this hypothesis, further investigation into how pre-existing modifications impact the binding of these enzymes must be completed to gain insight to the preferential binding and catalysis of RluB and RluF, as reported in Schultz and Kothe [160]. Expanding this study of the hierarchy of modification to the other pseudouridine synthases targeting the PTC, as well as other modification within the PTC, could yield interesting perspectives into the order of rRNA modification during ribosome biogenesis.

The four pseudouridine synthases responsible for modification of the PTC of the *E. coli* 23S rRNA are RluB, RluC, RluD, and RluF. While not essential, knockouts of these enzymes are outcompeted in *in vivo* competition assays, suggesting they provide a selective advantage under conditions mimicking natural selection. Other modification enzymes such as TruB and TrmA have been previously shown to have a secondary function as a tRNA chaperone [114, 137]. As such, it would be interesting to investigate the potential of these four rRNA-modifying enzymes to function as a chaperones under cellular stress. These rRNA modifications occur at various stages during ribosome biogenesis and likely have a preferential order of modification. Furthermore, these modifications are clustered in stem-loop structures, which may require re-folding and chaperone activity. This function has been primarily observed in pseudouridine synthases targeting tRNAs; accordingly, it could be most interesting to investigate RluF, which has the additional RNA target in $tRNA^{Tyr}$. Moreover, RluF and TruB are close homologs [65], since TruB has been found to have chaperone activity, it would be most interesting to study RluF in the context of being an RNA chaperone.

4.5 Conclusions

I successfully completed my objective of investigating the role of the conserved residues of the binding loops of RluB and RluF for substrate binding and catalysis. As this thesis demonstrates, hydrogen bonding to stabilize the A2602 bulge through E135 and H131 are crucial for RluB's ability to orient the target uridine for catalysis; without hydrogen bonding and the conformational shift of residues 132-135 into a 3_{10} helix, the RNA cannot be properly oriented for catalysis. Furthermore, R128 of RluF is shown to play a role in the stabilization of the target RNA near the active site. Without the hydrogen bonding of R128 to the target RNA backbone near the A2602 bulge, RluF loses the ability to orient its target base for catalysis. Finally, this thesis indirectly verifies previous reports that only U2604 and U2605 are being pseudouridylated [61]; tritium release assays are not nucleotide specific, thus, the detection of tritium in solution could indicate the modification of any of the forty-six uridines present in the PTC 195 substrate. Furthermore, this assay suggests that RluF may modify U2605 with a reduced speed compared to pseudouridylating U2604.

The work in this thesis provides preliminary insight into which elements of the eight-residue binding loops of RluB and RluF are important for RNA target recognition. The residues studied are critical in positioning the target RNA for catalysis; altering the residues of the binding loop chosen decreases the enzyme's affinity for its substrate and effectively abolish each enzyme's ability to introduce a pseudouridine at the target site. Better understanding the requirements for specificity and target recognition of RluB and RluF may aid in engineering new enzymes to pseudouridylate other functional targets within cells and influence gene expression.

BIBLIOGRAPHY:

1. Gabashvili, I.S., et al., *Solution structure of the E. coli 70S ribosome at 11.5 A resolution.* Cell, 2000. **100**(5): p. 537-49.
2. Havelund, J.F., et al., *Identification of 5-hydroxycytidine at position 2501 concludes characterization of modified nucleotides in E. coli 23S rRNA.* J Mol Biol, 2011. **411**(3): p. 529-36.
3. Liu, Q. and K. Fredrick, *Intersubunit Bridges of the Bacterial Ribosome.* J Mol Biol, 2016. **428**(10 Pt B): p. 2146-64.
4. Kaczanowska, M. and M. Ryden-Aulin, *Ribosome biogenesis and the translation process in Escherichia coli.* Microbiol Mol Biol Rev, 2007. **71**(3): p. 477-94.
5. Frank, J., et al., *Three-dimensional reconstruction of the 70S Escherichia coli ribosome in ice: the distribution of ribosomal RNA.* J Cell Biol, 1991. **115**(3): p. 597-605.
6. Yusupov, M.M., et al., *Crystal structure of the ribosome at 5.5 A resolution.* Science, 2001. **292**(5518): p. 883-96.
7. Rodgers, M.L. and S.A. Woodson, *Transcription Increases the Cooperativity of Ribonucleoprotein Assembly.* Cell, 2019. **179**(6): p. 1370-1381 e12.
8. Woodson, S.A., *RNA folding and ribosome assembly.* Curr Opin Chem Biol, 2008. **12**(6): p. 667-73.
9. Woodson, S.A., *RNA folding pathways and the self-assembly of ribosomes.* Acc Chem Res, 2011. **44**(12): p. 1312-9.
10. Shajani, Z., M.T. Sykes, and J.R. Williamson, *Assembly of bacterial ribosomes.* Annu Rev Biochem, 2011. **80**: p. 501-26.
11. Britton, R.A., *Role of GTPases in bacterial ribosome assembly.* Annu Rev Microbiol, 2009. **63**: p. 155-76.
12. Davis, J.H. and J.R. Williamson, *Structure and dynamics of bacterial ribosome biogenesis.* Philos Trans R Soc Lond B Biol Sci, 2017. **372**(1716).
13. Abeysirigunawardena, S.C., et al., *Evolution of protein-coupled RNA dynamics during hierarchical assembly of ribosomal complexes.* Nat Commun, 2017. **8**(1): p. 492.
14. Kim, H., et al., *Protein-guided RNA dynamics during early ribosome assembly.* Nature, 2014. **506**(7488): p. 334-8.
15. Kirpekar, F., S. Douthwaite, and P. Roepstorff, *Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry.* RNA, 2000. **6**(2): p. 296-306.
16. Wetzel, C. and P.A. Limbach, *Mass spectrometry of modified RNAs: recent developments.* Analyst, 2016. **141**(1): p. 16-23.
17. Charette, M. and M.W. Gray, *Pseudouridine in RNA: what, where, how, and why.* IUBMB Life, 2000. **49**(5): p. 341-51.
18. Sergeeva, O.V., A.A. Bogdanov, and P.V. Sergiev, *What do we know about ribosomal RNA methylation in Escherichia coli?* Biochimie, 2015. **117**: p. 110-8.
19. Noeske, J., et al., *High-resolution structure of the Escherichia coli ribosome.* Nat Struct Mol Biol, 2015. **22**(4): p. 336-41.
20. Agris, P.F., *The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function.* Prog Nucleic Acid Res Mol Biol, 1996. **53**: p. 79-129.
21. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol Syst Biol, 2006. **2**: p. 2006 0008.

22. Baldridge, K.C. and L.M. Contreras, *Functional implications of ribosomal RNA methylation in response to environmental stress.* Crit Rev Biochem Mol Biol, 2014. **49**(1): p. 69-89.
23. McCown, P.J., et al., *Naturally occurring modified ribonucleosides.* Wiley Interdiscip Rev RNA, 2020: p. e1595.
24. Gutgsell, N.S., M.P. Deutscher, and J. Ofengand, *The pseudouridine synthase RluD is required for normal ribosome assembly and function in Escherichia coli.* RNA, 2005. **11**(7): p. 1141-52.
25. Caldas, T., et al., *Translational defects of Escherichia coli mutants deficient in the Um(2552) 23S ribosomal RNA methyltransferase RrmJ/FTSJ.* Biochem Biophys Res Commun, 2000. **271**(3): p. 714-8.
26. Green, R. and H.F. Noller, *In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function.* RNA, 1996. **2**(10): p. 1011-21.
27. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2017 update.* Nucleic Acids Res, 2018. **46**(D1): p. D303-D307.
28. Ayadi, L., et al., *RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions.* Biochim Biophys Acta Gene Regul Mech, 2018.
29. Sloan, K.E., et al., *Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function.* RNA Biol, 2017. **14**(9): p. 1138-1152.
30. Romano, G., et al., *RNA Methylation in ncRNA: Classes, Detection, and Molecular Associations.* Front Genet, 2018. **9**: p. 243.
31. Decatur, W.A. and M.J. Fournier, *rRNA modifications and ribosome function.* Trends Biochem Sci, 2002. **27**(7): p. 344-51.
32. Kimura, S. and T. Suzuki, *Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA.* Nucleic Acids Res, 2010. **38**(4): p. 1341-52.
33. Hamdane, D., H. Grosjean, and M. Fontecave, *Flavin-Dependent Methylation of RNAs: Complex Chemistry for a Simple Modification.* J Mol Biol, 2016. **428**(24 Pt B): p. 4867-4881.
34. Blanco, S. and M. Frye, *Role of RNA methyltransferases in tissue renewal and pathology.* Curr Opin Cell Biol, 2014. **31**: p. 1-7.
35. Watanabe, K., et al., *Roles of conserved amino acid sequence motifs in the SpoU (TrmH) RNA methyltransferase family.* J Biol Chem, 2005. **280**(11): p. 10368-77.
36. Ye, K., et al., *Structural organization of box C/D RNA-guided RNA methyltransferase.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13808-13.
37. Lovgren, J.M. and P.M. Wikstrom, *The rlmB gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in Escherichia coli.* J Bacteriol, 2001. **183**(23): p. 6957-60.
38. Caldas, T., et al., *The FtsJ/RrmJ heat shock protein of Escherichia coli is a 23 S ribosomal RNA methyltransferase.* J Biol Chem, 2000. **275**(22): p. 16414-9.
39. Purta, E., et al., *YgdE is the 2'-O-ribose methyltransferase RlmM specific for nucleotide C2498 in bacterial 23S rRNA.* Mol Microbiol, 2009. **72**(5): p. 1147-58.
40. Rana, A.K. and S. Ankri, *Reviving the RNA World: An Insight into the Appearance of RNA Methyltransferases.* Front Genet, 2016. **7**: p. 99.

41. Gustafsson, C. and B.C. Persson, *Identification of the rrmA gene encoding the 23S rRNA m1G745 methyltransferase in Escherichia coli and characterization of an m1G745-deficient mutant.* J Bacteriol, 1998. **180**(2): p. 359-65.
42. Madsen, C.T., et al., *Identifying the methyltransferases for m(5)U747 and m(5)U1939 in 23S rRNA using MALDI mass spectrometry.* Nucleic Acids Res, 2003. **31**(16): p. 4738-46.
43. Agarwalla, S., et al., *Characterization of the 23 S ribosomal RNA m5U1939 methyltransferase from Escherichia coli.* J Biol Chem, 2002. **277**(11): p. 8835-40.
44. Sergiev, P.V., et al., *The ybiN gene of Escherichia coli encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel.* J Mol Biol, 2008. **375**(1): p. 291-300.
45. Sergiev, P.V., et al., *Identification of Escherichia coli m2G methyltransferases: II. The ygjO gene encodes a methyltransferase specific for G1835 of the 23 S rRNA.* J Mol Biol, 2006. **364**(1): p. 26-31.
46. Ero, R., et al., *Identification of pseudouridine methyltransferase in Escherichia coli.* RNA, 2008. **14**(10): p. 2223-33.
47. Purta, E., et al., *YbeA is the m3Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA.* RNA, 2008. **14**(10): p. 2234-44.
48. Purta, E., et al., *YccW is the m5C methyltransferase specific for 23S rRNA nucleotide 1962.* J Mol Biol, 2008. **383**(3): p. 641-51.
49. Golovina, A.Y., et al., *The last rRNA methyltransferase of E. coli revealed: the yhiR gene encodes adenine-N6 methyltransferase specific for modification of A2030 of 23S ribosomal RNA.* RNA, 2012. **18**(9): p. 1725-34.
50. Kimura, S., et al., *Base methylations in the double-stranded RNA by a fused methyltransferase bearing unwinding activity.* Nucleic Acids Res, 2012. **40**(9): p. 4071-85.
51. Lesnyak, D.V., et al., *Identification of Escherichia coli m2G methyltransferases: I. the ycbY gene encodes a methyltransferase specific for G2445 of the 23 S rRNA.* J Mol Biol, 2006. **364**(1): p. 20-5.
52. Yan, F., et al., *RlmN and Cfr are radical SAM enzymes involved in methylation of ribosomal RNA.* J Am Chem Soc, 2010. **132**(11): p. 3953-64.
53. Desai, P.M. and J.P. Rife, *The adenosine dimethyltransferase KsgA recognizes a specific conformational state of the 30S ribosomal subunit.* Arch Biochem Biophys, 2006. **449**(1-2): p. 57-63.
54. Tscherne, J.S., et al., *Purification, cloning, and characterization of the 16S RNA m5C967 methyltransferase from Escherichia coli.* Biochemistry, 1999. **38**(6): p. 1884-92.
55. Tscherne, J.S., et al., *Purification, cloning, and characterization of the 16 S RNA m2G1207 methyltransferase from Escherichia coli.* J Biol Chem, 1999. **274**(2): p. 924-9.
56. Lesnyak, D.V., et al., *Methyltransferase that modifies guanine 966 of the 16 S rRNA: functional identification and tertiary structure.* J Biol Chem, 2007. **282**(8): p. 5880-7.
57. Basturea, G.N. and M.P. Deutscher, *Substrate specificity and properties of the Escherichia coli 16S rRNA methyltransferase, RsmE.* RNA, 2007. **13**(11): p. 1969-76.
58. Andersen, N.M. and S. Douthwaite, *YebU is a m5C methyltransferase specific for 16 S rRNA nucleotide 1407.* J Mol Biol, 2006. **359**(3): p. 777-86.

59. Nishimura, K., et al., *Mutations in rsmG, encoding a 16S rRNA methyltransferase, result in low-level streptomycin resistance and antibiotic overproduction in Streptomyces coelicolor A3(2).* J Bacteriol, 2007. **189**(10): p. 3876-83.
60. Basturea, G.N., et al., *YhiQ is RsmJ, the methyltransferase responsible for methylation of G1516 in 16S rRNA of E. coli.* J Mol Biol, 2012. **415**(1): p. 16-21.
61. Del Campo, M., Y. Kaya, and J. Ofengand, *Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli.* RNA, 2001. **7**(11): p. 1603-15.
62. Conrad, J., et al., *The rluC gene of Escherichia coli codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA.* J Biol Chem, 1998. **273**(29): p. 18562-6.
63. Wrzesinski, J., et al., *Isolation and properties of Escherichia coli 23S-RNA pseudouridine 1911, 1915, 1917 synthase (RluD).* IUBMB Life, 2000. **50**(1): p. 33-7.
64. Raychaudhuri, S., et al., *A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of Escherichia coli.* RNA, 1998. **4**(11): p. 1407-17.
65. Sunita, S., et al., *Domain organization and crystal structure of the catalytic domain of E.coli RluF, a pseudouridine synthase that acts on 23S rRNA.* J Mol Biol, 2006. **359**(4): p. 998-1009.
66. Conrad, J., et al., *16S ribosomal RNA pseudouridine synthase RsuA of Escherichia coli: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases.* RNA, 1999. **5**(6): p. 751-63.
67. Andersen, T.E., B.T. Porse, and F. Kirpekar, *A novel partial modification at C2501 in Escherichia coli 23S ribosomal RNA.* RNA, 2004. **10**(6): p. 907-13.
68. Kimura, S., et al., *Biogenesis and iron-dependency of ribosomal RNA hydroxylation.* Nucleic Acids Res, 2017. **45**(22): p. 12974-12986.
69. Benitez-Paez, A., M. Villarroya, and M.E. Armengod, *The Escherichia coli RlmN methyltransferase is a dual-specificity enzyme that modifies both rRNA and tRNA and controls translational accuracy.* RNA, 2012. **18**(10): p. 1783-95.
70. Jenner, L., et al., *Structural rearrangements of the ribosome at the tRNA proofreading step.* Nat Struct Mol Biol, 2010. **17**(9): p. 1072-8.
71. Toh, S.M., et al., *The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA.* RNA, 2008. **14**(1): p. 98-106.
72. Demirci, H., et al., *Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function.* RNA, 2010. **16**(12): p. 2319-24.
73. Xu, Z., et al., *A conserved rRNA methyltransferase regulates ribosome biogenesis.* Nat Struct Mol Biol, 2008. **15**(5): p. 534-6.
74. Connolly, K., J.P. Rife, and G. Culver, *Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA.* Mol Microbiol, 2008. **70**(5): p. 1062-75.
75. Pulicherla, N., et al., *Structural and functional divergence within the Dim1/KsgA family of rRNA methyltransferases.* J Mol Biol, 2009. **391**(5): p. 884-93.
76. Mangat, C.S. and E.D. Brown, *Ribosome biogenesis; the KsgA protein throws a methyl-mediated switch in ribosome assembly.* Mol Microbiol, 2008. **70**(5): p. 1051-3.
77. Weitzmann, C., et al., *A paradigm for local conformational control of function in the ribosome: binding of ribosomal protein S19 to Escherichia coli 16S rRNA in the presence*

of S7 is required for methylation of m2G966 and blocks methylation of m5C967 by their respective methyltransferases. Nucleic Acids Res, 1991. **19**(25): p. 7089-95.
78. Liu, M., G.W. Novotny, and S. Douthwaite, *Methylation of 23S rRNA nucleotide G745 is a secondary function of the RlmAI methyltransferase.* RNA, 2004. **10**(11): p. 1713-20.
79. Osterman, I.A., et al., *Methylated 23S rRNA nucleotide m2G1835 of Escherichia coli ribosome facilitates subunit association.* Biochimie, 2011. **93**(4): p. 725-9.
80. Moine, H. and A.E. Dahlberg, *Mutations in helix 34 of Escherichia coli 16 S ribosomal RNA have multiple effects on ribosome function and synthesis.* J Mol Biol, 1994. **243**(3): p. 402-12.
81. Gc, K., et al., *Ribosomal RNA Methyltransferase RsmC Moonlights as an RNA Chaperone.* Chembiochem, 2020. **21**(13): p. 1885-1892.
82. Arora, S., et al., *Role of the ribosomal P-site elements of m(2)G966, m(5)C967, and the S9 C-terminal tail in maintenance of the reading frame during translational elongation in Escherichia coli.* J Bacteriol, 2013. **195**(16): p. 3524-30.
83. Nishimura, K., et al., *Identification of the RsmG methyltransferase target as 16S rRNA nucleotide G527 and characterization of Bacillus subtilis rsmG mutants.* J Bacteriol, 2007. **189**(16): p. 6068-73.
84. Okamoto, S., et al., *Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria.* Mol Microbiol, 2007. **63**(4): p. 1096-106.
85. Benitez-Paez, A., M. Villarroya, and M.E. Armengod, *Regulation of expression and catalytic activity of Escherichia coli RsmG methyltransferase.* RNA, 2012. **18**(4): p. 795-806.
86. Hur, S., R.M. Stroud, and J. Finer-Moore, *Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective.* J Biol Chem, 2006. **281**(51): p. 38969-73.
87. Agarwalla, S., R.M. Stroud, and B.J. Gaffney, *Redox reactions of the iron-sulfur cluster in a ribosomal RNA methyltransferase, RumA: optical and EPR studies.* J Biol Chem, 2004. **279**(33): p. 34123-9.
88. Basturea, G.N., K.E. Rudd, and M.P. Deutscher, *Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family.* RNA, 2006. **12**(3): p. 426-34.
89. Ero, R., et al., *Specificity and kinetics of 23S rRNA modification enzymes RlmH and RluD.* RNA, 2010. **16**(11): p. 2075-84.
90. O'Connor, M. and A.E. Dahlberg, *The involvement of two distinct regions of 23 S ribosomal RNA in tRNA selection.* J Mol Biol, 1995. **254**(5): p. 838-47.
91. Cannone, J.J., et al., *The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs.* BMC Bioinformatics, 2002. **3**: p. 2.
92. Cunningham, P.R., et al., *The absence of modified nucleotides affects both in vitro assembly and in vitro function of the 30S ribosomal subunit of Escherichia coli.* Biochimie, 1991. **73**(6): p. 789-96.
93. Spenkuch, F., Y. Motorin, and M. Helm, *Pseudouridine: still mysterious, but never a fake (uridine)!* RNA Biol, 2014. **11**(12): p. 1540-54.
94. Penzo, M. and L. Montanaro, *Turning Uridines around: Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function.* Biomolecules, 2018. **8**(2).

95. Rintala-Dempsey, A.C. and U. Kothe, *Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression?* RNA Biol, 2017. **14**(9): p. 1185-1196.
96. Tillault, A.S., et al., *Molecular Determinants for 23S rRNA Recognition and Modification by the E. coli Pseudouridine Synthase RluE.* J Mol Biol, 2018. **430**(9): p. 1284-1294.
97. Ofengand, J., *Ribosomal RNA pseudouridines and pseudouridine synthases.* FEBS Lett, 2002. **514**(1): p. 17-25.
98. Hamma, T. and A.R. Ferre-D'Amare, *Pseudouridine synthases.* Chem Biol, 2006. **13**(11): p. 1125-35.
99. Davis, D.R., *Stabilization of RNA stacking by pseudouridine.* Nucleic Acids Res, 1995. **23**(24): p. 5020-6.
100. Huang, M., et al., *Improvement of DNA and RNA Sugar Pucker Profiles from Semiempirical Quantum Methods.* J Chem Theory Comput, 2014. **10**(4): p. 1538-1545.
101. Penzo, M., et al., *RNA Pseudouridylation in Physiology and Medicine: For Better and for Worse.* Genes (Basel), 2017. **8**(11).
102. Ofengand, J. and A. Bakin, *Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts.* J Mol Biol, 1997. **266**(2): p. 246-68.
103. De Zoysa, M.D. and Y.T. Yu, *Posttranscriptional RNA Pseudouridylation.* Enzymes, 2017. **41**: p. 151-167.
104. Czekay, D.P. and U. Kothe, *H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes.* Front Microbiol, 2021. **12**: p. 654370.
105. Watanabe, Y. and M.W. Gray, *Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria.* Nucleic Acids Res, 2000. **28**(12): p. 2342-52.
106. Hamma, T., et al., *The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs.* Nat Struct Mol Biol, 2005. **12**(12): p. 1101-7.
107. Ganot, P., M. Caizergues-Ferrer, and T. Kiss, *The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation.* Genes Dev, 1997. **11**(7): p. 941-56.
108. Liang, B., et al., *Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex.* Nat Struct Mol Biol, 2007. **14**(12): p. 1189-95.
109. De Zoysa, M.D., et al., *Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation.* RNA, 2018. **24**(8): p. 1106-1117.
110. Kelly, E.K., D.P. Czekay, and U. Kothe, *Base-pairing interactions between substrate RNA and H/ACA guide RNA modulate the kinetics of pseudouridylation, but not the affinity of substrate binding by H/ACA small nucleolar ribonucleoproteins.* RNA, 2019. **25**(10): p. 1393-1404.
111. Carlile, T.M., et al., *Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells.* Nature, 2014. **515**(7525): p. 143-6.
112. Maden, B.E., *The numerous modified nucleotides in eukaryotic ribosomal RNA.* Prog Nucleic Acid Res Mol Biol, 1990. **39**: p. 241-303.
113. O'Connor, M., M. Leppik, and J. Remme, *Pseudouridine-Free Escherichia coli Ribosomes.* J Bacteriol, 2018. **200**(4).

114. Keffer-Wilkes, L.C., G.R. Veerareddygari, and U. Kothe, *RNA modification enzyme TruB is a tRNA chaperone.* Proc Natl Acad Sci U S A, 2016. **113**(50): p. 14306-14311.
115. Watkins, N.J. and M.T. Bohnsack, *The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA.* Wiley Interdiscip Rev RNA, 2012. **3**(3): p. 397-414.
116. Wolin, S.L., *Two for the price of one: RNA modification enzymes as chaperones.* Proc Natl Acad Sci U S A, 2016. **113**(50): p. 14176-14178.
117. Polacek, N. and A.S. Mankin, *The ribosomal peptidyl transferase center: structure, function, evolution, inhibition.* Crit Rev Biochem Mol Biol, 2005. **40**(5): p. 285-311.
118. Fernandez, I.S., et al., *Unusual base pairing during the decoding of a stop codon by the ribosome.* Nature, 2013. **500**(7460): p. 107-10.
119. Ge, J. and Y.T. Yu, *RNA pseudouridylation: new insights into an old modification.* Trends Biochem Sci, 2013. **38**(4): p. 210-8.
120. Wrzesinski, J., et al., *A dual-specificity pseudouridine synthase: an Escherichia coli synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe).* RNA, 1995. **1**(4): p. 437-48.
121. Hoang, C., et al., *Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure.* Mol Cell, 2006. **24**(4): p. 535-45.
122. Raychaudhuri, S., et al., *Functional effect of deletion and mutation of the Escherichia coli ribosomal RNA and tRNA pseudouridine synthase RluA.* J Biol Chem, 1999. **274**(27): p. 18880-6.
123. Kulik, M., et al., *Helix 69 of Escherichia coli 23S ribosomal RNA as a peptide nucleic acid target.* Biochimie, 2017. **138**: p. 32-42.
124. Vaidyanathan, P.P., M.P. Deutscher, and A. Malhotra, *RluD, a highly conserved pseudouridine synthase, modifies 50S subunits more specifically and efficiently than free 23S rRNA.* RNA, 2007. **13**(11): p. 1868-76.
125. Ofengand, J., et al., *Pseudouridines and pseudouridine synthases of the ribosome.* Cold Spring Harb Symp Quant Biol, 2001. **66**: p. 147-59.
126. Alian, A., et al., *Crystal structure of an RluF-RNA complex: a base-pair rearrangement is the key to selectivity of RluF for U2604 of the ribosome.* J Mol Biol, 2009. **388**(4): p. 785-800.
127. Czudnochowski, N., et al., *The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs.* Nucleic Acids Res, 2014. **42**(3): p. 2037-48.
128. Wrzesinski, J., et al., *Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from Escherichia coli.* Biochemistry, 1995. **34**(27): p. 8904-13.
129. Bakin, A., et al., *The single pseudouridine residue in Escherichia coli 16S RNA is located at position 516.* Nucleic Acids Res, 1994. **22**(18): p. 3681-4.
130. Jayalath, K., et al., *Pseudouridine Synthase RsuA Captures an Assembly Intermediate that Is Stabilized by Ribosomal Protein S17.* Biomolecules, 2020. **10**(6).
131. Dalluge, J.J., et al., *Conformational flexibility in RNA: the role of dihydrouridine.* Nucleic Acids Res, 1996. **24**(6): p. 1073-9.
132. Bou-Nader, C., et al., *Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases: perspectives on the evolution scenario.* Nucleic Acids Res, 2018. **46**(3): p. 1386-1394.

133. Faivre, B., et al., *Dihydrouridine synthesis in tRNAs is under reductive evolution in Mollicutes.* RNA Biol, 2021: p. 1-12.
134. Mattimore, V. and J.R. Battista, *Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation.* J Bacteriol, 1996. **178**(3): p. 633-7.
135. Kanehisa, M., et al., *Data, information, knowledge and principle: back to metabolism in KEGG.* Nucleic Acids Res, 2014. **42**(Database issue): p. D199-205.
136. Cloutier, P., et al., *A newly uncovered group of distantly related lysine methyltransferases preferentially interact with molecular chaperones to regulate their activity.* PLoS Genet, 2013. **9**(1): p. e1003210.
137. Keffer-Wilkes, L.C., E.F. Soon, and U. Kothe, *The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain.* Nucleic Acids Res, 2020. **48**(14): p. 7981-7990.
138. Miracco, E.J. and E.G. Mueller, *The products of 5-fluorouridine by the action of the pseudouridine synthase TruB disfavor one mechanism and suggest another.* J Am Chem Soc, 2011. **133**(31): p. 11826-9.
139. Veerareddygari, G.R., S.K. Singh, and E.G. Mueller, *The Pseudouridine Synthases Proceed through a Glycal Intermediate.* J Am Chem Soc, 2016. **138**(25): p. 7852-5.
140. Addepalli, B. and P.A. Limbach, *Pseudouridine in the Anticodon of Escherichia coli tRNATyr(QPsiA) Is Catalyzed by the Dual Specificity Enzyme RluF.* J Biol Chem, 2016. **291**(42): p. 22327-22337.
141. Chaudhuri, B.N., et al., *Crystal structure of the apo forms of psi 55 tRNA pseudouridine synthase from Mycobacterium tuberculosis: a hinge at the base of the catalytic cleft.* J Biol Chem, 2004. **279**(23): p. 24585-91.
142. Vieira-Pires, R.S. and J.H. Morais-Cabral, *3(10) helices in channels and other membrane proteins.* J Gen Physiol, 2010. **136**(6): p. 585-92.
143. Kitagawa, M., et al., *Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research.* DNA Res, 2005. **12**(5): p. 291-9.
144. Wright, J.R., et al., *Pre-steady-state kinetic analysis of the three Escherichia coli pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis.* RNA, 2011. **17**(12): p. 2074-84.
145. Scott, K.A., et al., *Conformational entropy of alanine versus glycine in protein denatured states.* Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2661-6.
146. Larentis, A.L., et al., *Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor.* BMC Res Notes, 2014. **7**: p. 671.
147. Shachrai, I., et al., *Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth.* Mol Cell, 2010. **38**(5): p. 758-67.
148. Malakar, P. and K.V. Venkatesh, *Effect of substrate and IPTG concentrations on the burden to growth of Escherichia coli on glycerol due to the expression of Lac proteins.* Appl Microbiol Biotechnol, 2012. **93**(6): p. 2543-9.
149. Andrews, K.J. and G.D. Hegeman, *Selective disadvantage of non-functional protein synthesis in Escherichia coli.* J Mol Evol, 1976. **8**(4): p. 317-28.
150. Farewell, A. and F.C. Neidhardt, *Effect of temperature on in vivo protein synthetic capacity in Escherichia coli.* J Bacteriol, 1998. **180**(17): p. 4704-10.

151. Pekarsky, A., M. Reininger, and O. Spadiut, *The impact of technical failures on recombinant production of soluble proteins in Escherichia coli: a case study on process and protein robustness.* Bioprocess Biosyst Eng, 2021.
152. Slouka, C., et al., *Monitoring and control strategies for inclusion body production in E. coli based on glycerol consumption.* J Biotechnol, 2019. **296**: p. 75-82.
153. Robichon, C., et al., *Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography.* Appl Environ Microbiol, 2011. **77**(13): p. 4634-46.
154. Sugimoto, S., et al., *Construction of Escherichia coli dnaK-deletion mutant infected by lambdaDE3 for overexpression and purification of recombinant GrpE proteins.* Protein Expr Purif, 2008. **60**(1): p. 31-6.
155. Rohman, M. and K.J. Harrison-Lavoie, *Separation of copurifying GroEL from glutathione-S-transferase fusion proteins.* Protein Expr Purif, 2000. **20**(1): p. 45-7.
156. Sehnal, D., et al., *Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures.* Nucleic Acids Research, 2021. **49**(W1): p. W431-W437.
157. Bugg, T.D., *The development of mechanistic enzymology in the 20th century.* Nat Prod Rep, 2001. **18**(5): p. 465-93.
158. Amyes, T.L., et al., *Enzyme activation through the utilization of intrinsic dianion binding energy.* Protein Eng Des Sel, 2017. **30**(3): p. 157-165.
159. Olson, M.A., *Calculations of free-energy contributions to protein-RNA complex stabilization.* Biophys J, 2001. **81**(4): p. 1841-53.
160. Schultz, S.K. and U. Kothe, *tRNA elbow modifications affect the tRNA pseudouridine synthase TruB and the methyltransferase TrmA.* RNA, 2020. **26**(9): p. 1131-1142.
161. Peng, S., et al., *Evaluation of three reference genes of Escherichia coli for mRNA expression level normalization in view of salt and organic acid stress exposure in food.* FEMS Microbiol Lett, 2014. **355**(1): p. 78-82.
162. Arsene, F., T. Tomoyasu, and B. Bukau, *The heat shock response of Escherichia coli.* Int J Food Microbiol, 2000. **55**(1-3): p. 3-9.
163. Heermann, R., et al., *The universal stress protein UspC scaffolds the KdpD/KdpE signaling cascade of Escherichia coli under salt stress.* J Mol Biol, 2009. **386**(1): p. 134-48.
164. Huang, E.Y., A.M. Mohler, and C.E. Rohlman, *Protein expression in response to folate stress in Escherichia coli.* J Bacteriol, 1997. **179**(17): p. 5648-53.
165. Kobayashi, H., M. Yamamoto, and R. Aono, *Appearance of a stress-response protein, phage-shock protein A, in Escherichia coli exposed to hydrophobic organic solvents.* Microbiology (Reading), 1998. **144 (Pt 2)**: p. 353-359.
166. Kvint, K., et al., *The bacterial universal stress protein: function and regulation.* Curr Opin Microbiol, 2003. **6**(2): p. 140-5.
167. Free, A. and C.J. Dorman, *The Escherichia coli stpA gene is transiently expressed during growth in rich medium and is induced in minimal medium and by stress conditions.* J Bacteriol, 1997. **179**(3): p. 909-18.
168. Hardman, S.J. and K.C. Thompson, *Influence of base stacking and hydrogen bonding on the fluorescence of 2-aminopurine and pyrrolocytosine in nucleic acids.* Biochemistry, 2006. **45**(30): p. 9145-55.
169. Jean, J.M. and K.B. Hall, *2-Aminopurine fluorescence quenching and lifetimes: role of base stacking.* Proc Natl Acad Sci U S A, 2001. **98**(1): p. 37-41.

APPENDIX

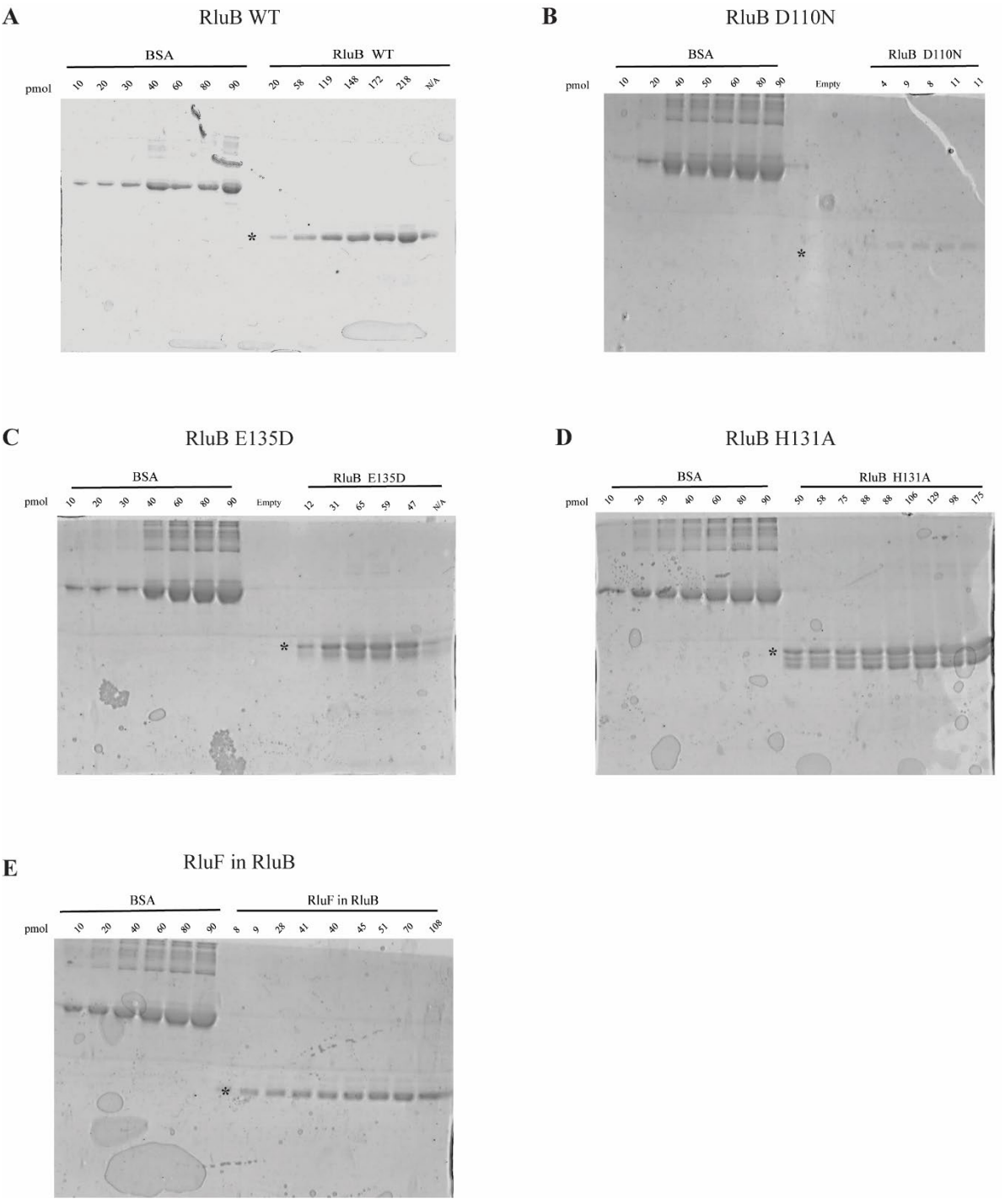


Figure A.1 SDS-PAGEs of RluB variants to determine the concentrations of protein preparations (A-E). SDS-PAGE of enzyme variant to measure the concentration compared to a BSA standard. Protein concentrations of each band were determined using ImageJ software, and the percent purity calculated (Table 3.1). Bands indicated with an asterisk were determined to be the full-length protein.

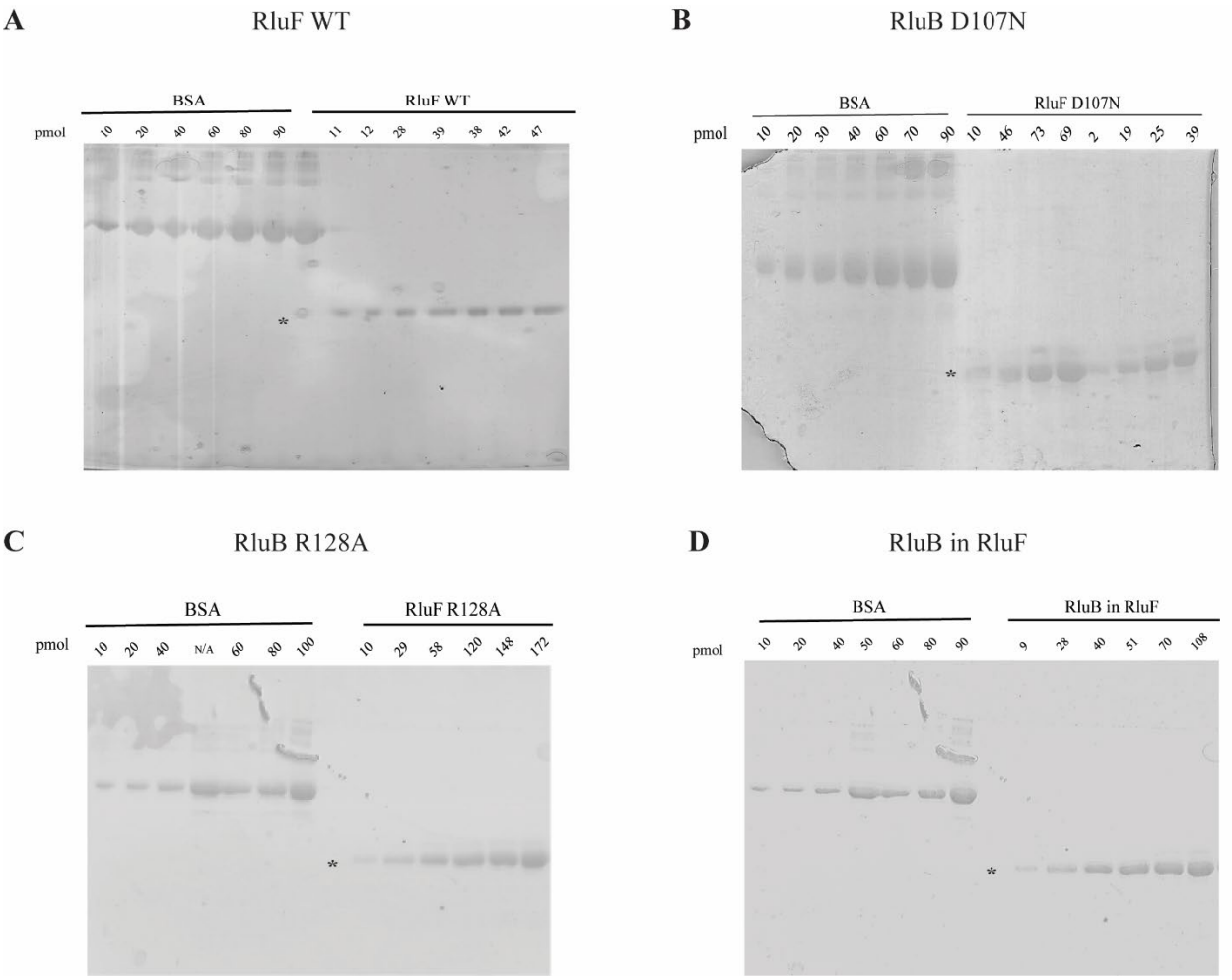


Figure A.2 SDS-PAGEs of RluF variants to determine the concentrations of protein preparations (A-D). SDS-PAGE of enzyme variant to measure the concentration compared to a BSA standard. Protein concentrations of each band were determined using ImageJ software, and the percent purity calculated (Table 3.1). Bands indicated with an asterisk were determined to be the full-length protein.

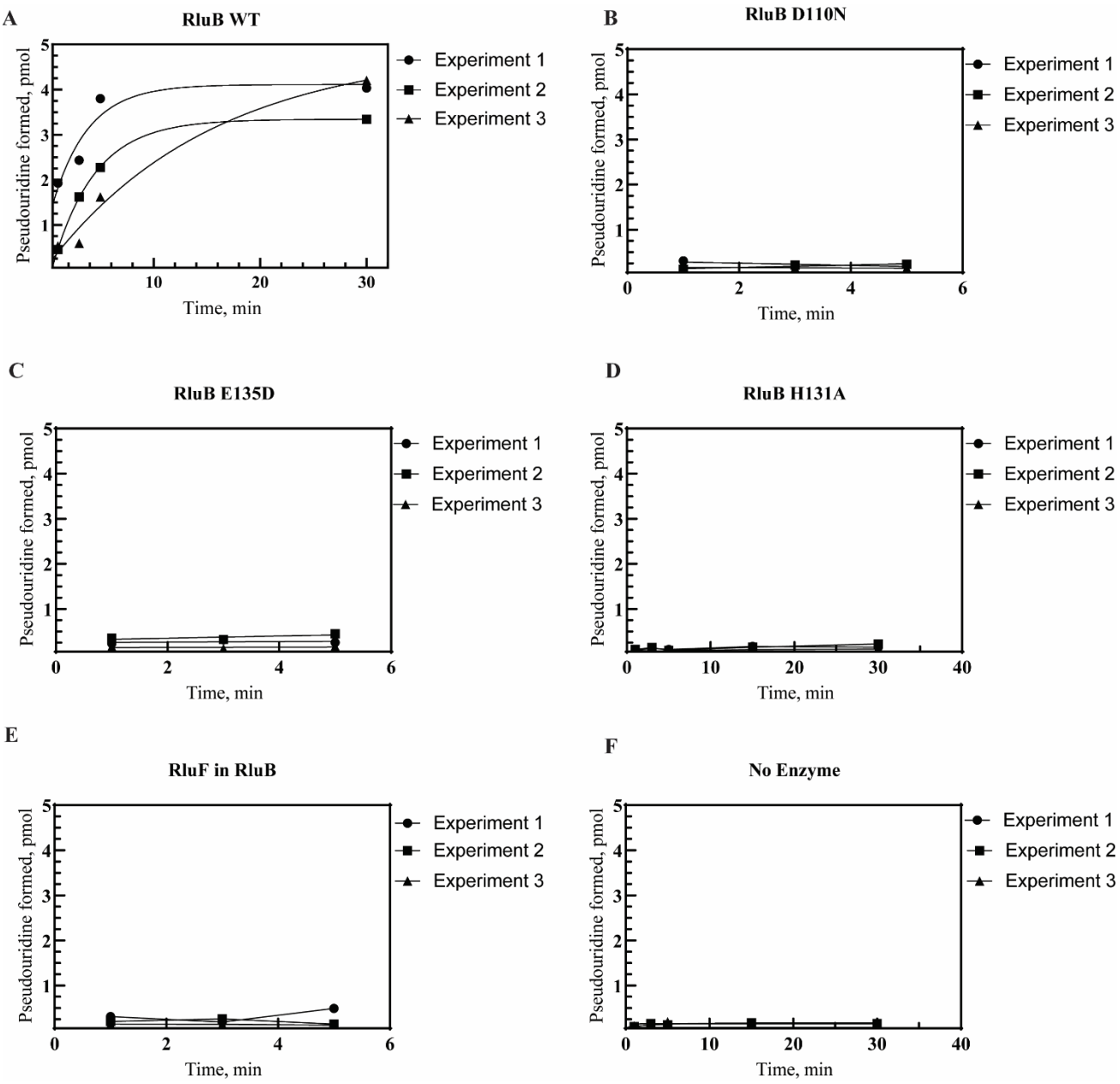


Figure A.3 Pseudouridylation activity of all RluB enzyme variants with PTC 195 (A-F) Individual replicates of tritium release assays of respective enzyme variants from RluB examining the pseudouridylation activity of *in vitro* transcribed PTC 195 substrate. Reactions were performed under single turnover conditions (1000 nM enzyme, 300 nM substrate RNA) and shown are 3 experiments. Each reaction sample has 3 pmol of radiolabeled substrate.

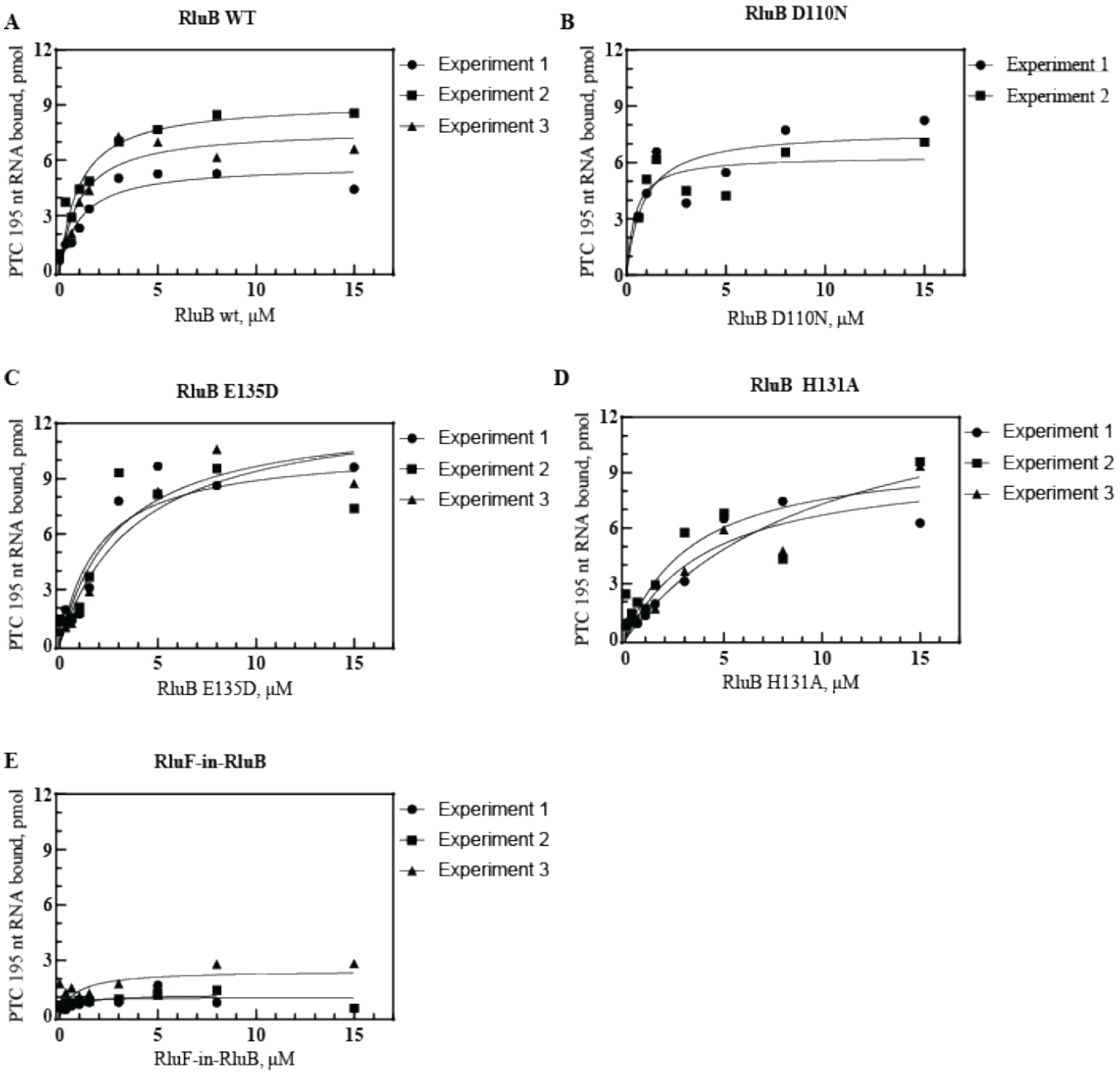


Figure A.4 Nitrocellulose filter binding of all RluB enzyme variants with PTC 195. Nitrocellulose filter binding assays were performed using 3 pmol of ^{3}H labelled PTC 195 with enzyme concentrations spanning from 0-15 μM. Different enzyme variants of RluB were tested in triplicate **(A-E).** The dissociation constants for each enzyme are listed in Table A.1.

Table A.1 – Individual binding affinities of RluB enzyme variants binding the PTC 195 radiolabeled RNA substrate.

RLUB VARIANT	TRIAL 1 K_D (μM)	TRIAL 2 K_D (μM)	TRIAL 3 K_D (μM)	AVERAGE DISSOCIATION CONSTANT (μM)
RluB WT	1.07	0.99	1.02	0.96 ± 0.29
D110N	0.85	0.42		0.64
E135D	2.97	1.99	3.87	2.76 ± 0.73
H131A	4.30	3.39	10.19	5.0 ± 1.6
RluF in RluB	0.84	0.15	0.82	1.21 ± 0.96

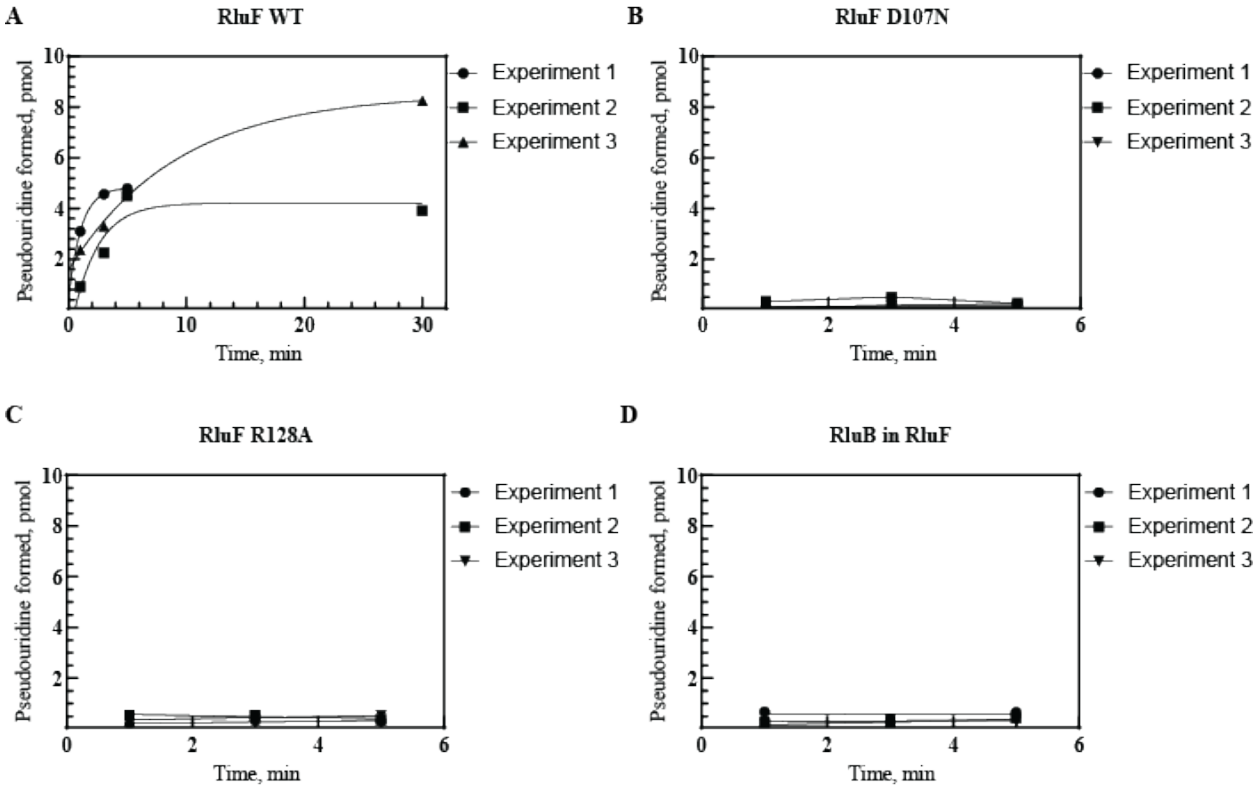


Figure A.5 Pseudouridylation activity of all RluF enzyme variants with PTC 195. (A-D). Individual replicates of tritium release assays of respective enzyme variants from RluB examining the pseudouridylation activity of *in vitro* transcribed PTC 195 substrate. Reactions were performed under single turnover conditions (1000 nM enzyme, 300 nM substrate RNA) and shown are 3 experiments each. Each reaction sample has 3 pmol of radiolabeled substrate.

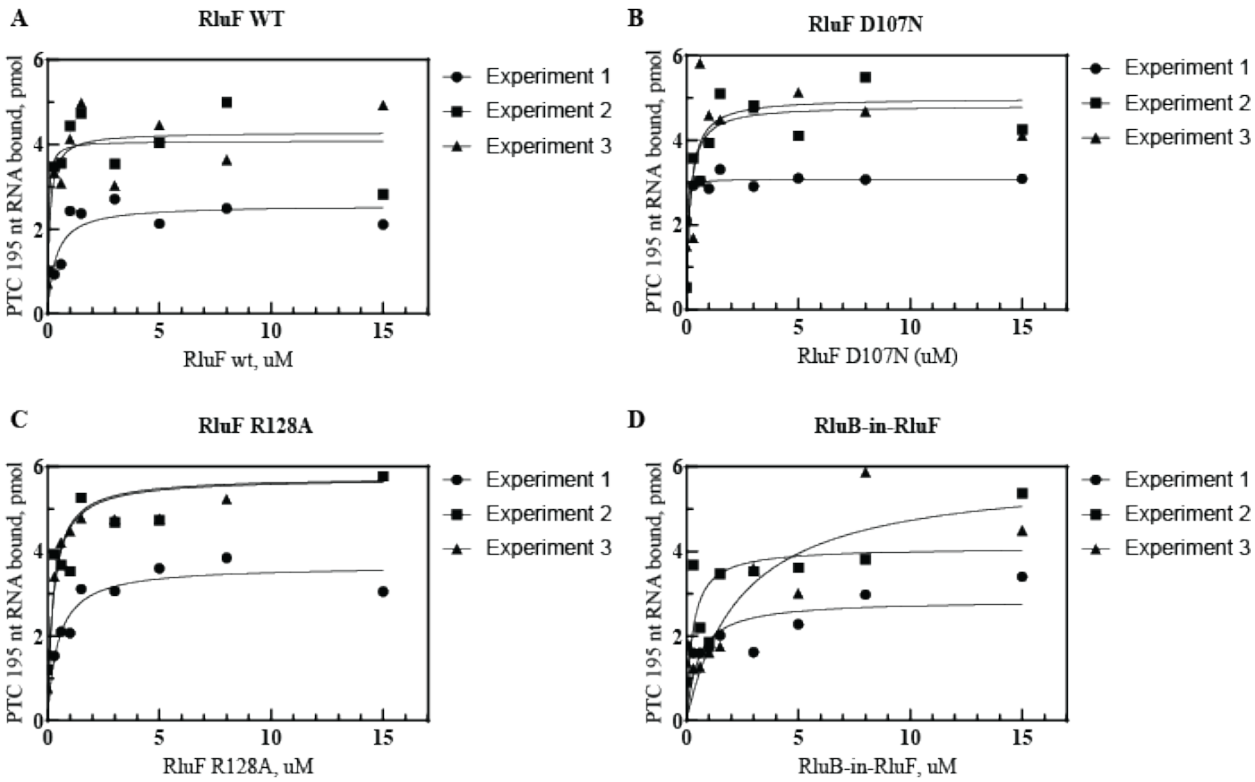


Figure A.6 – Nitrocellulose filter binding of all RluF enzyme variants with PTC 195. Nitrocellulose filter binding assays were performed using 3 pmol of ^{3}H labelled PTC 195 with enzyme concentrations spanning from 0-15 μM. Different enzyme variants of RluF were tested in triplicate **(A-D).** The dissociation constants for each enzyme are listed in Table A.2.

Table A.2 – Individual binding affinities of RluF enzyme variants binding the PTC 195 radiolabeled RNA substrate

RLUF VARIANT	TRIAL 1 K_D (μM)	TRIAL 2 K_D (μM)	TRIAL 3 K_D (μM)	AVERAGE DISSOCIATION CONSTANT (μM)
RluF WT	0.36	0.04	0.1	0.12 ± 0.10
D107N	0.01	0.16	0.16	0.12 ± 0.08
R128A	0.46	0.26	0.25	0.30 ± 0.14
RluB in RluF	0.51	0.27	2.5	0.98 ± 0.44

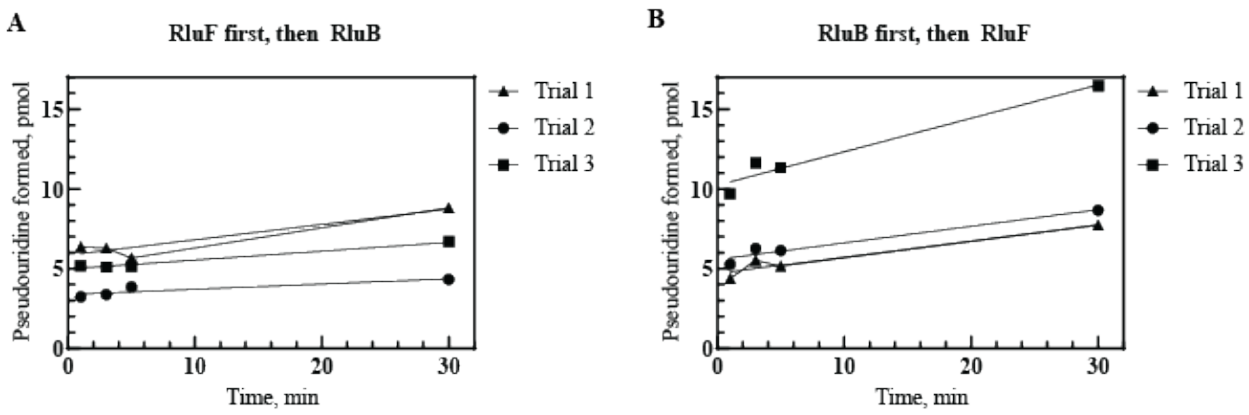


Figure A.7 – Individual trials of pseudouridylation activity of RluB and RluF with pre-modified RNA substrate. Tritium release assay of wildtype enzymes RluB and RluF examining the pseudouridylation activity of pre-modified *in vitro* transcribed PTC 195 substrate. Reactions were performed under single turnover conditions (1000 nM enzyme, 300 nM substrate RNA) and shown are 3 individual experiments. Each reaction sample has 3 pmol of radiolabeled substrate. **(A)** RluF incubated with the RNA substrate for thirty minutes prior to the addition of RluB **(B)** RluB incubated with the RNA substrate for thirty minutes prior to the addition of RluF**.**