MULTIDISCIPLINARY INVESTIGATIONS ON TRNA MODIFICATIONS: A CMNM^5(S^2)U CASE STUDY

CYNTHIA DESIRÉE ETONDI FONDERSON
Bachelor of Science, University of Lethbridge, 2019

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

MULTIDISCIPLINARY INVESTIGATIONS ON TRNA MODIFICATIONS:
A $CMNM^5(S^2)U$ CASE STUDY

CYNTHIA DESIRÉE ETONDI FONDERSON

Date of Defense: May 6, 2022

Dr. U. Kothe	Professor	Ph.D.
Dr. S. Wetmore	Professor	Ph.D.
Thesis Co-Supervisors		
Dr. B. Demeler	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. M. Roussel	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. J. Hamel	Assistant Professor	Ph.D.
Chair, Thesis Examination Committee Member		

ABSTRACT

Although tRNA modifications 5-carboxymethylaminomethyl(-2-thio)uridine ($cmnm^5(s^2)U$; and their human homologs) have been linked to mitochondrial disorders, their effects *in vivo* remain unknown. Furthermore, due to a lack of structural evidence, controversy exists in the literature pertaining to the assembly and catalytic mechanism of the MnmEG complex that introduces these modifications in tRNA. This thesis used a multidisciplinary approach to study the MnmEG modification complex and the effects of $cmnm^5(s^2)U$ on tRNA structure. Biochemical studies reveal the binding propensities of MnmE and MnmG to their tRNA substrate, contributing vital experimental evidence towards the isolation of the MnmEG-tRNA complex. Concurrently, computational studies uncover the structural and cooperative properties of $cmnm^5(s^2)U$ in biologically relevant tRNAs, providing the first atomic-level details of the function of these modifications in translation. Overall, this thesis lays the foundation for further explorations into this tRNA modification family and its relation to disease.

ACKNOWLEDGEMENTS

First, I want to acknowledge my mum who has always been the wind beneath my wings. Her love for science and research set me on the path I follow today. Throughout this journey, she has been a primary source of support and fortitude. Thank you for being my sounding board, shoulder to cry on and everything else in between.

Next, I want to appreciate my amazing supervisors, Drs. Ute Kothe and Stacey Wetmore, who guided and unconditionally supported me over the past years. I am grateful for the assistance and encouragement you have given me throughout my time at the University of Lethbridge. I am also very grateful for past and present colleagues in the Kothe and Wetmore Labs, who provided much needed inspiration, advice, and laughter.

I want to recognize my committee members, Drs. Borries Demeler and Marc Roussel, whose questions, feedback, and ideas have been incredibly valuable in shaping the progress of my thesis.

Lastly, but most importantly, I thank God, who gave me the strength and patience to finish this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3'-AC	3' anticodon base conformation
A-site	Aminoacyl site
aa-tRNA	Aminoacylated tRNA
aaRS	Aminoacyl tRNA synthetases
ADP	Adenosine diphosphate
APBS	Adaptive Poisson-Boltzmann Solver
ASL	Anticodon stem-loop
cMD	Conventional molecular dynamics
$cmnm^5s^2U$	5-carboxymethylaminomethyl-2-thiouridine
$cmnm^5U$	5-carboxymethylaminomethyluridine
CPU	Central processing unit
Cryo-EM	Cryogenic electron microscopy
D-arm	Dihydrouridine arm
DBM	Dinucleotide binding motif
DL	Disordered loop
DTT	Dithiothreitol
E-site	Exit site
EF	Elongation factor
εRMSD	Ensemble averaged root-mean-square deviation
FAD	Flavin adenine dinucleotide
FB	Flanking base conformation
GDP	Guanosine diphosphate
GEF	Guanine-nucleotide exchange factor
GFP	Green fluorescent protein
GPU	Graphic processing unit
GR	Glutathione reductase
GTP	Guanosine triphosphate
HMG-CoA	3-hydroxy3-methylglutaryl coenzyme A
IF	Initiation factor
IPTG	Isopropyl β-D-1-thiogalactopyranoside
mcm^5s^2U	5-methoxycarbonylmethyl-2-thiouridine
MD	Molecular dynamics
MD	Molecular dynamics
MELAS	Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes
MERRF	Myoclonus epilepsy associated with ragged-red fibers
ML-AUC	Multiwavelength analytical ultracentrifugation

mnm^5U	5-methylaminomethyluridine
mt	Mitochondrial
NADH	1,4-dihyronicotinamide adenine dinucleotide
NMR	Nuclear magnetic resistance
nt	Nucleotide
NTP	Nucleoside triphosphate
P-site	Peptidyl site
PAGE	Polyacrylamide gel electrophoresis
PC	Principal component vector
PCA	Principal component analysis
PCR	Polymerase Chain Reaction
PME	Particle mesh Ewald
PMSF	Phenylmethanesulfonylfluoride
PTC	Peptidyl transferase center
rMD	Replica molecular dynamics
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
rRNA	Ribosomal RNA
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate
T-arm	$T\psi C$ arm
THF	Tetrahydrofolate
τm^5s^2U	5-taurinomethyl-2-thiouridine
τm^5U	5-taurinomethyluridine
tRF	Transfer RNA fragment
WB	Wobble base conformation
xm^5s^2U	5-methyl-2-thiouridine derivatives
xm^5U	5-methyluridine derivatives
xo^5U	5-hydroxyuridine derivatives

CHAPTER 1: INTRODUCTION

1.1 Thesis Overview

RNA is a highly flexible nucleic acid that shows vast structural and functional diversity *in vivo.*[1-3] The structural diversity of RNA molecules allows them to perform varied functions, and they have been found to play central roles in many cellular processes, including gene expression and regulation.[4] To facilitate the diverse functions performed by RNA, chemically modified nucleobases are post-transcriptionally inserted into RNA structures.[5-7] These non-canonical nucleobases can directly impact RNA structure and alter intramolecular interactions between RNA and other biomolecules. Although over 150 modifications have been discovered in RNA, 93 of these were exclusively found in transfer RNA (tRNA) molecules.[8-10] This thesis focuses on a family of tRNA modifications inserted at position 34 by the MnmEG modification complex that have been linked to multiple neurological and mitochondrial disorders.[11-14] Specifically, many biochemical and computational studies were conducted to understand the assembly of this modification complex, and the role of its modifications *in vivo.* The current chapter offers an overview of the relevant literature on tRNA, the MnmEG complex and the 5-carboxymethylaminomethyluridine modification family, providing the background required to understand the scope of this thesis.

1.2 Transfer RNA

Transfer RNAs are crucial connectors between the coding messenger RNA and the growing polypeptide chain during protein synthesis.[15-18] These non-coding RNA are approximately 70-100 nucleotides in length, have an average weight of 25,000 g/mol, carry specific amino acids and recognize mRNA codons using their anticodon bases.[19,20] tRNA nucleotides are organized into a series of helical hairpins, which are often illustrated using a cloverleaf secondary structure with five domains – the acceptor stem, the dihydrouridine arm (D-arm), the anticodon stem-loop (ASL), the variable loop and the TψC arm (T-arm; Figure 1.1). The D-arm is named after the dihydrouridine base that contributes to the stabilization of the tRNA tertiary structure, while the T-arm is named due to the presence of universally

conserved thymidine (T), pseudouridine (ψ) and cytidine (C) residues which facilitate interactions with the ribosome.[3,21,22]

The five tRNA domains can generally be classified into two groups, namely the structural and functional domains.[20] Structural tRNA domains are involved in tRNA folding and the achievement of its tertiary structure. The highly conserved L-shape of tRNA arises from the formation of tertiary interactions between the D- and T-loops, while the variable region accommodates the remaining tRNA nucleotides and helps stabilize the tertiary tRNA structure (Figure 1.1).[23-26] The first functional domain in tRNA is the acceptor stem that is charged with an amino acid at its 3' end (CCA) through an ester linkage facilitated by aminoacyl tRNA synthetases (aaRS). The function of these enzymes is critical to translation, as the ribosome does not read the amino acid tRNAs are charged with and will incorporate the wrong amino acid if the tRNA is incorrectly aminoacylated.[20,27,28] Consequently, tRNAs have discriminatory bases 5' of the CCA, which determine aminoacylation specificity.[28] The second functional domain is the ASL that contains the three-base anticodon (34, 35 and 36). Translation specificity and accuracy is ensured by the anticodon as it interacts with mRNA codons via base pairing interactions.[19,29]

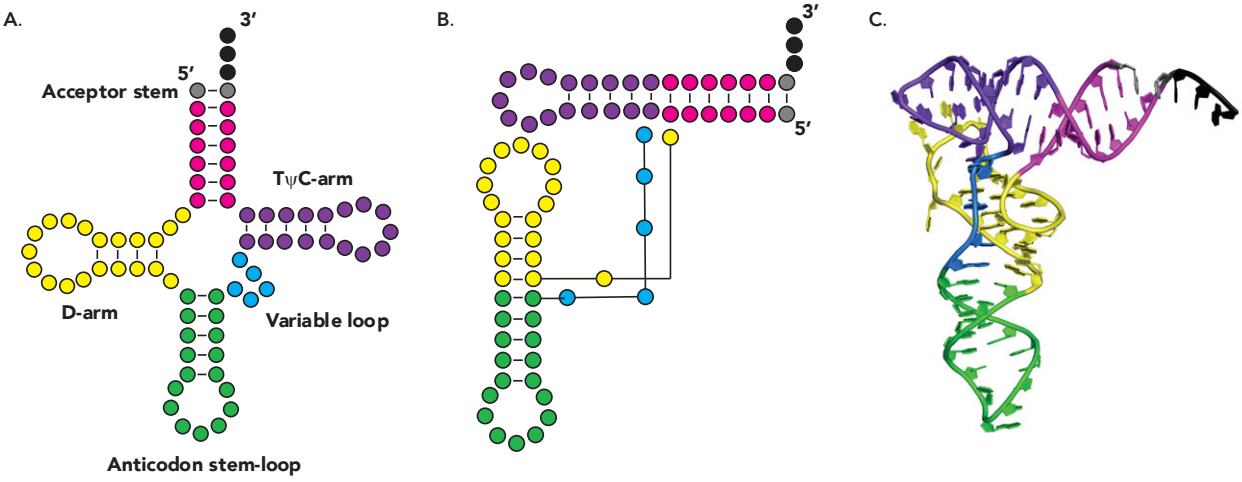


Figure 1.1 – Transfer RNA structure

Graphical representations of the secondary structure of tRNA in its cloverleaf (A) and L-shape (B) forms. Tertiary structure of tRNA in its canonical L-shape (C). The acceptor stem, D-arm, ASL, variable loop and TψC-arm are colored magenta, yellow, green, blue, and purple, respectively.

Transfer RNAs have also evolved functions outside of their direct role in translation.[20] In bacteria, tRNA enzymatically transfer amino acids to the N-terminal of various proteins and this transfer can either add functionality to the proteins or cause them to be targets for turn-over.[30-34] Bacteria also use tRNA to add amino acids to peptides and lipids during the formation of antibiotics and the synthesis of peptidoglycan for their cell walls.[35,36] In eukaryotes, tRNA regulate gene expression during amino acid starvation and can participate in viral gene expression during reverse transcription.[37-41] tRNA fragments, originating from the cleavage of mature tRNA under various stresses including starvation, oxidative stress, and hypoxia, have also been found to play critical roles in cellular function.[42,43] There are four types of tRNA fragments – large tRNA fragments from the cleavage within the ASL generating 5' and 3' tRNA halves, and smaller fragments that arise from cleavage at the D-loop (5' tRF) or the T-loop (3' CCA tRF).[44,45] Some roles of tRNA fragments include the formation of stress granules (5' tRNA half), signalling apoptosis (5' and 3' tRNA halves), halting translation (5' tRF), mRNA degradation and enhanced translation (3' CCA tRF).[46-48]

In humans, tRNAs have been found to play critical roles in pathogenic replication and disease development. For example, the human immunodeficiency virus type 1 (HIV-1) can use range of tRNAs as primers in reverse transcription – a critical step in retroviral replication, during which viral genomic RNA is converted to double-stranded DNA that is subsequently integrated within the host to form a provirus.[40,49] During this process, HIV-1 exploits structural elements of the tRNA elbow and merges its U5-PBS (primer binding site) element with the TψC-acceptor minihelix to build its reverse transcription initiation complex. Furthermore, HIV-1 uses tRNAs to transport its major structural protein (Gag) to the nucleus. To do this, the N-terminal domain of this protein, responsible for cytoplasmic localization, inserts itself between the D and T loops, and the protein uses the tRNA as a carrier to enter the nucleus.[50-52] Interestingly, the involvement of tRNAs in retroviral life cycles has been extensively studied in the literature, and some tRNAs have been considered as targets for antiviral therapeutics.[53-59] All in all, tRNA are important biomolecules that participate and regulate gene expression in various ways.

1.2.1 Transfer RNA modifications

Transfer RNAs have the greatest diversity and largest number of modified nucleosides discovered to date.[60] On average, up to 15% of the tRNA structure can be modified at once, making it the most post-transcriptionally modified RNA in existence.[9,24,25] tRNA modifications arise from structurally diverse chemical changes and can either be small and reversible or bulkier and irreversible.[61,62] Moreover, tRNA are differentially modified, i.e., some nucleobases are more frequently and diversely modified than others. There are two modification hotspots in tRNA – the core region and the anticodon loop.[61,63,64]

The core tRNA region is defined by the folding domains, i.e., D, T, and variable regions, and the nucleobases in these domains are usually hypomodified before tRNA secondary structure is fully attained.[65] Frequent modifications within the tRNA core include pseudouridine (27,55), dihydrouridine (20), methyl-5-uridine (thymine; 54), N^2-N^2-dimethylguanosine (26), N^1-methyladenosine (9) and 4-thiouridine (8,9). Modifications within the core provide global stabilization to the tertiary structure of tRNA and prevent premature tRNA degradation.[66-69] Additionally, tRNA chaperones like TruB and TrmA act on the T-arm, introducing modifications that enhance tRNA folding.[70,71]

ASL is the functional domain of tRNA that reads the mRNA codon through hydrogen bonding interactions.[20,72] The 3 anticodon bases (34, 35 and 36) interact with 3 codon bases (1, 2 and 3). Bases 35 and 36 usually form Watson-Crick base pairs with the second and first codon bases respectively, whereas base 34 interacts with the third codon base, but does not always form canonical interactions.[15,73] For instance, during the translation of phenylalanine codons by $tRNA^{Phe}_{GAA}$, G34 base pairs with C3 and U3, which allows UUU and UUC codons to be translated by a single tRNA. This atypical interaction known as ‘wobble’ pairing is a system through which the set of 61 codons can be translated with a limited number of tRNA species.[74,75] Base 34 is defined as the wobble base, and it is the most frequently modified position within the tRNA.[73,74] This position also has the widest variety of chemical modifications, which play critical roles in modulating codon recognition during translation.[61,63,64] Of the four canonical bases, uracil is often inserted at the wobble position, and is most frequently modified.[76-79] Wobble uridine modifications are always accompanied by modifications at position 37, which is the second most diversely modified nucleotide in tRNA. Located at the 3' end of the anticodon, modifications at this base are proposed to

ensure accurate mRNA decoding and reading frame maintenance during translation.[80,81] Therefore, while investigating the effects of ASL modifications on tRNA structure, this thesis will focus on modifications at the 34th and 37th positions of tRNA.

1.3 Transfer RNA and Protein Synthesis

Ribosomal protein synthesis (translation) is an essential process for the maintenance of cell viability.[82,83] During translation, ribosomes – cellular complexes made of ribosomal RNA (rRNA) and protein, bind aminoacylated tRNAs (aa-tRNAs) in a codon-dependent manner to synthesize mRNA-encoded proteins.[84] Ribosomes have two subunits – a large subunit and a small subunit, and the size of these vary from one organism to another.[84-86] For instance, bacterial ribosomes have large 50S subunits consisting of 23S rRNA, 5S rRNA and 30 proteins, and small 30S subunits consisting of 16S rRNA and 20 proteins, while human ribosomes have large 60S subunits consisting of 28S, 5.8S and 5S rRNA and 47 proteins, and small 40S subunits consisting of 18S rRNA and 33 proteins.[86,87] All ribosomes possess three tRNA binding sites – aminoacyl (A), peptidyl (P), and exit (E) site.[85,86]

There are four steps to ribosomal protein synthesis – initiation, elongation, termination, and recycling.[82] In bacteria, initial interactions between the small ribosomal subunit, the mRNA and the tRNA is mediated by initiation factors 1, 2 and 3 (IF1, IF2, IF3).[88-90] The small subunit binds mRNA via base pairing interactions between mRNA's Shine-Dalgarno sequence and a complementary sequence present in the 3'-end of 16S rRNA. Then, the initiator tRNA binds the 30S subunit at the P-site and its anticodon base pairs with the mRNA start codon. Accurate mRNA-tRNA interactions signal the rapid assembly of the full ribosomal complex. Upon the association of the 50S and 30S subunits, initiation factors are released.[82,89]

Prior to binding the ribosome, an aa-tRNA will form a ternary complex with elongation factor Tu (EF-Tu) and GTP via its acceptor stem to prevent the premature and inaccurate tRNA insertion.[91,92] The aa-tRNA•EF-Tu•GTP complex interacts with the 30S subunit at its A-site and the mRNA codon via the tRNA anticodon stem-loop (ASL). If cognate codon-anticodon pairing is formed, GTP is hydrolyzed by EF-Tu and the aa-tRNA is accommodated into the 50S subunit. Alternatively, tRNA that form non-cognate

pairing are generally rejected by the ribosome in a process known as conformational proofreading.[92-94] During this process, the mRNA-tRNA interaction is detected via base pairing between the 30S subunit and the minor groove of the codon-anticodon interface. This interaction between the ribosome and the mRNA-tRNA complex is sensitive to Watson-Crick geometry, not base sequence. Consequently, cognate and some near-cognate tRNAs are more stable than non-cognate tRNAs, causing the latter to be rejected by the ribosome.[95-98] The full accommodation of the aa-tRNA within the A-site leads to the formation of a peptide bond by the peptidyl transferase center (PTC) located in the 50S subunit.[99-102]

Following this reaction, the P-site tRNA is deacylated while the A-site tRNA bears the nascent peptide chain and the two tRNAs are translocated into the E- and P-sites of the 30S subunit, respectively.[103] The translocation of the mRNA-tRNA complex is catalyzed by elongation factor G (EF-G), which uses GTP hydrolysis to accelerate tRNA movement.[104,105] During this process, the ribosomal complex is reorganized through a ratchet-like movement of the 30S subunit relative to the 50S one.[106-108] After tRNA translocation, the next mRNA codon is in the A-site and a new round of elongation can proceed until a stop codon enters the A-site, instigating the termination phase of protein synthesis. Stop codons are recognized by either release factor 1 or 2 (RF1 and RF2), which bind near the ribosomal A-site and catalyze the release of the completed polypeptide attached to the P-site tRNA.[109-112] Once the peptide chain is released, the ribosomal complex dissociates and is recycled with the help of the ribosome recycling factor (RRF) and EF-G•GTP. The mRNA and deacylated tRNA are removed from the 30S subunit, which is now free to initiate translation afresh.[109,113,114]

Throughout translation, tRNA forms extensive interactions with the ribosomal complex.[82] At the P-site, interactions between the anticodon arm (base pairs 29-41 and 30-40) and 30S subunit are important in discriminating between initiator tRNA and elongator tRNAs.[115-118] Extensive interactions also exist between the ASL and several 16S rRNA bases and 30S proteins S9 and S13. These interactions stabilize the mRNA-tRNA complex at the P-site.[115,117-119] Moreover, ribosomal protein L5 is in proximity to C56 and the 23S rRNA interacts with bases 12 and 13 in the P-site tRNA D-arm.[120,121] At the A-site, the anticodon arm of tRNA interacts with the decoding center of the 30S ribosome via the universally conserved bases in the 16S rRNA – G530, A1492 and A1493.[75,115,122] These bases interact

with the minor groove of the codon-anticodon helix and are proposed to discriminate between cognate and non-cognate base pairing by monitoring the shape of the mRNA-ASL helix. Furthermore, mutations at these bases reduce tRNA translocation by 20-fold, suggesting that these bases are required for rapid translocation.[122-124] Ribosomal proteins S12 and S13 also interact with the anticodon arm within the A-site. These interactions are essential to an efficient translation process as mutations within these proteins promote spontaneous, EF-G-independent translocation.[125-127] Ribosomal protein L16 interacts with the backbone of the tRNA at positions 23 and 54, stabilizing the tRNA within the A-site.[128] Furthermore, helix 38 of the 23S rRNA lies between the elbows of A- and P-site tRNAs and maintains the pre-translocation state of the ribosome.[129,130] Several conserved bases within the PTC interact with the 3'-CCA ends of the A- and P-site tRNA.[99,115,118,119,128,129,131] Some of these bases facilitate translocation (e.g., G2251)[132,133] while others are required for peptide release (e.g., G2252, A2451, U2506, U2585, and A2602).[133-135] The N-terminal end of ribosomal protein L27 also interacts with A- and P-site tRNAs, promoting peptide bond formation by the ribosome.[136,137] At the E-site, the 30S subunit interacts with the anticodon bases exiting tRNA, causing a relaxation in the anticodon loop.[115,118,119,138] The bases in these interactions have also been shown to be necessary for the maintenance of the translational reading frame.[139] On the other hand, extensive interactions exist between the acceptor arm of the E-site tRNA and the 50S subunit.[115,118,128] These interactions control the translocation of P-site tRNA into the E-site and stabilize hybrid P/E-site states.[140-143]

As discussed above, extensive interactions are formed between the ASL of tRNA and 30S subunit within the ribosomal complex.[77,82,91,98,119,123,125,139,144,145] but a lot remains unknown about how modifications in the ASL influence tRNA accommodation and function during protein synthesis. Although structural studies have provided insight into the molecular features of the translation complex, they fail to describe the flexibilities of the macromolecules that could play key roles in the complex's functionality.[115,118,119,122,127,139] Conventionally, the dynamic properties of nucleic acids are investigated using molecular dynamics (MD) simulations.[119,146-150] This computational technique uses the three-dimensional structure of biomolecules, experimentally derived via NMR or X-ray crystallography, to evaluate interactions within (or between) macromolecules as a function of the coordinates of their individual substituents, including but not limited to atoms, residues, or nucleobases. Consequently, MD

studies deliver detailed atomistic motions and interactions over set periods of time and can be used to understand macromolecular structure–function relationships observed in experimental studies.[119,146,147,151,152] A wide variety of MD studies have been carried out on tRNA. The first computational study on tRNA was a 32 ps MD simulation reported by McCammon and Harvey in 1988 on yeast $tRNA^{Phe}$.[153] Since then, MD simulations have also been used to study tRNA folding,[154-156] the aminoacylation process[157,158] and various binding events.[155,159] More interestingly, this technique has been used to investigate the effects of posttranscriptional modifications on tRNA structure using a variety of tRNA models. For instance, McCrate and colleagues studied the structural effects of various ASL modifications at different positions, using an X-ray crystal structure of the $tRNA^{Lys}_{UUU}$ ASL (bases 27-43 only), but more recent MD studies on tRNA use full tRNA structures to study modification effects.[159-162] Nevertheless, despite the increased interest in tRNA modifications and their structural and functional effects, there is no consensus in the literature on the MD protocol to use for these studies and computational investigations on tRNA vary in simulation length (ps to μs) and replication (zero to three replicas).[161-164] This thesis sets out to address this issue by developing a computational protocol for investigations on tRNA structure and dynamics. Subsequently, the methodology will be applied to understand the structural effects the 5-carboxymethylaminomethyluridine modification family has on tRNA.

1.4 The 5-carboxymethylaminomethyl-2-thiouridine Modification Family

Wobble uridines were initially proposed to recognize A and G at the third codon, but their high conformational flexibility allows them to interact with any of the four canonical bases (four-way wobble rule).[165,166] Two types of C5-wobble uridine modifications have been described in the literature, based on their chemical structure and decoding properties: 5-hydroxyuridine derivatives (xo^5U) with an oxygen atom directly bonded to the C5 atom of the uracil base and 5-methyluridine derivatives (xm^5U) with a methylene carbon directly bonded to the C5 atom.[167-169] In general, xo^5U modifications contribute to the efficient reading of A-, G- and U-ending codons and can recognize all four bases in the absence of tRNA isoacceptors.[91] Isoacceptors define chemically different tRNA species that are acylated by the same

amino acid. Consequently, xo^{5}U modifications are often found in tRNA responsible for entire codon boxes such as tRNA$^{Leu}_{CNN}$, tRNA^{Pro}, tRNA^{Ser} and tRNA^{Val}.[9,166-168] In contrast, xm^{5}U modifications are found in tRNA responsible for decoding two codon sets that end in purines (NNR) including but not limited to tRNA$^{Leu}_{UUR}$, tRNA^{Lys}, tRNA^{Gln}, tRNA^{Glu}.[8,170] These modifications prevent the misreading of pyrimidine (Y)-ending near-cognate codons and are generally more rigid than their xo^{5}U counterparts. Interestingly, xm^{5}U modifications include 2-thiouridine (xm^{5}s^{2}U) and 2'-O-methyluridine (xm^{5}Um) derivatives, which are believed to increase rigidity in the modifications and lock their sugar in a C3'-endo conformation.[65,169,171]

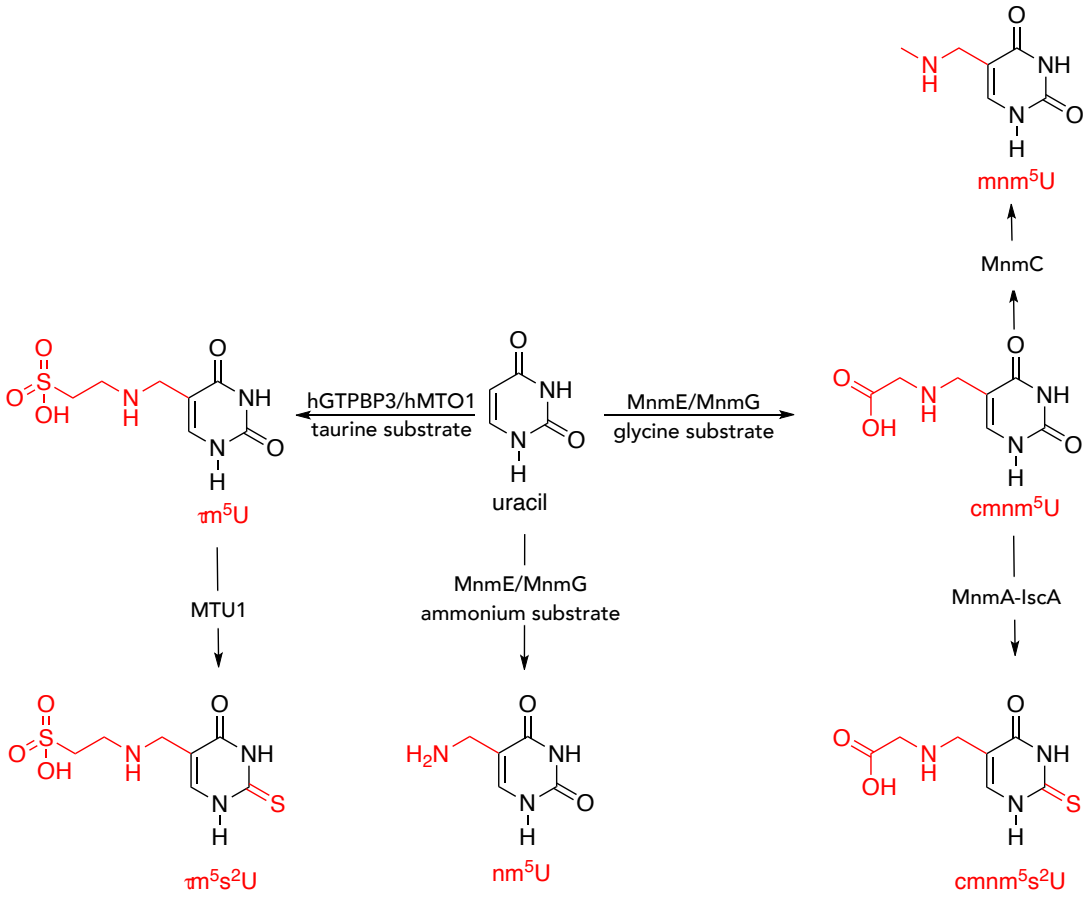


Figure 1.2 – 5-carboxymethylaminomethyluridine modifications

The uridine base is modified by the bacterial MnmE-MnmG complex (glycine substrate) to form 5-carboxymethylaminomethyluridine (cmnm^{5}U) and the hGTPBP3/hMTO1 complex (taurine) to form 5-taurinomethyluridine (τm^{5}U). cmnm^{5}U and τm^{5}U are further modified at C2 by the MnmA-IscA complex and MTU1, forming 5-carboxymethylaminomethyl-2-thiouridine (cmnm^{5}s^{2}U) and 5-taurinomethyl-2-thiouridine (τm^{5}s^{2}U) respectively. In bacteria, the MnmE-MnmG complex glycine can be substituted by ammonium, leading to the formation of 5-aminomethyluridine (nm^{5}U).

5-carboxymethylaminomethyluridine (cmnm^{5}U) and its derivatives can be found in archaeal, bacterial, and eukaryotic tRNA (Figure 1.2).[171-173] The MnmE/MnmG (MnmEG) complex and their eukaryotic homologs (MSS1 and MTO1) are responsible for the insertion of the substituents at C5. The thiolation of this modification at C2 occurs independently of C5 and is facilitated by the MnmA-IscS complex. In addition to being wobble uridines, cmnm^{5}U modifications act as intermediates during the biosynthesis of 5-methylaminomethyluridine (mnm^{5}U) in bacteria. This process is mediated by MnmC(o) and MnmC(m) enzymes that cleave the carboxymethyl substituent and replace it with a methyl group.[171,174] In mammalian mitochondrial (mt)-tRNA, glycine is be replaced by taurine, forming 5-taurinomethyluridine (τm^{5}U) modifications via the GTPBP3/MTOI complex. Like cmnm^{5}U, τm^{5}U is independently thiolated by MTU1, the mitochondrial homolog for MnmA.[170,175,176] Under stress conditions e.g., taurine starvation, cmnm^5(s^2)U34 modifications have been found in mitochondrial tRNA.[62,170]

1.4.1 MnmE

MnmE (formerly TrmE) is a homodimeric protein with each monomer (~ 50 kDa) consisting of an N-terminal domain, a helical domain and a G-domain that is located within its helical domain (Figure 1.3).[177-179] The N-terminal domain of MnmE, made up of five stranded mixed β-sheets and 3 α-helices, is the major contributor at the dimerization interface of the protein. This domain is also responsible for binding tetrahydrofolate (THF) and its derivatives (formyl-THF, methylene-THF) with affinities in the low micromolar and sub-micromolar range.[179,180] The α-helical domain consists of three to six helices and four long helices that form a 4-helix bundle. The C-terminal residues of this domain are not part of any helix and are in proximity of THF bound to the N-terminal domain. These amino acid residues form a highly conserved FC(V/I/L)GK motif, and the cysteine residue was found to be essential during tRNA modification.[181] The G-domain of MnmE closely resembles the canonical G-domain exemplified by the Ras protein and consists of 6 β-strands and 5 α-helices.[179] Furthermore, MnmE G-domains contain at least 4 of the 5 sequence motifs that characterize G-proteins: the GxxxxGK(S/T) motif in the P loop, a conserved threonine in switch I, the DxxG motif in switch II and the NKxD motif responsible for specificity toward a guanine nucleobase.[179,182]

Although the structure of MnmE's G-domain closely follows that of typical G-proteins, its biochemical features are peculiar. Whereas G-proteins like Ras have low intrinsic GTPase activity and high affinities for their substrate (GTP) and product (GDP), the MnmE G-domain demonstrates a relatively high intrinsic rate of GTP hydrolysis and rather low affinities for guanidine nucleotides. Consequently, MnmE's G-domain does not require auxiliary GTPase activating proteins (GAPs) and guanine-nucleotide exchange factors (GEFs) to catalyze GTP hydrolysis and proceed through the GTPase cycle.[183-185]

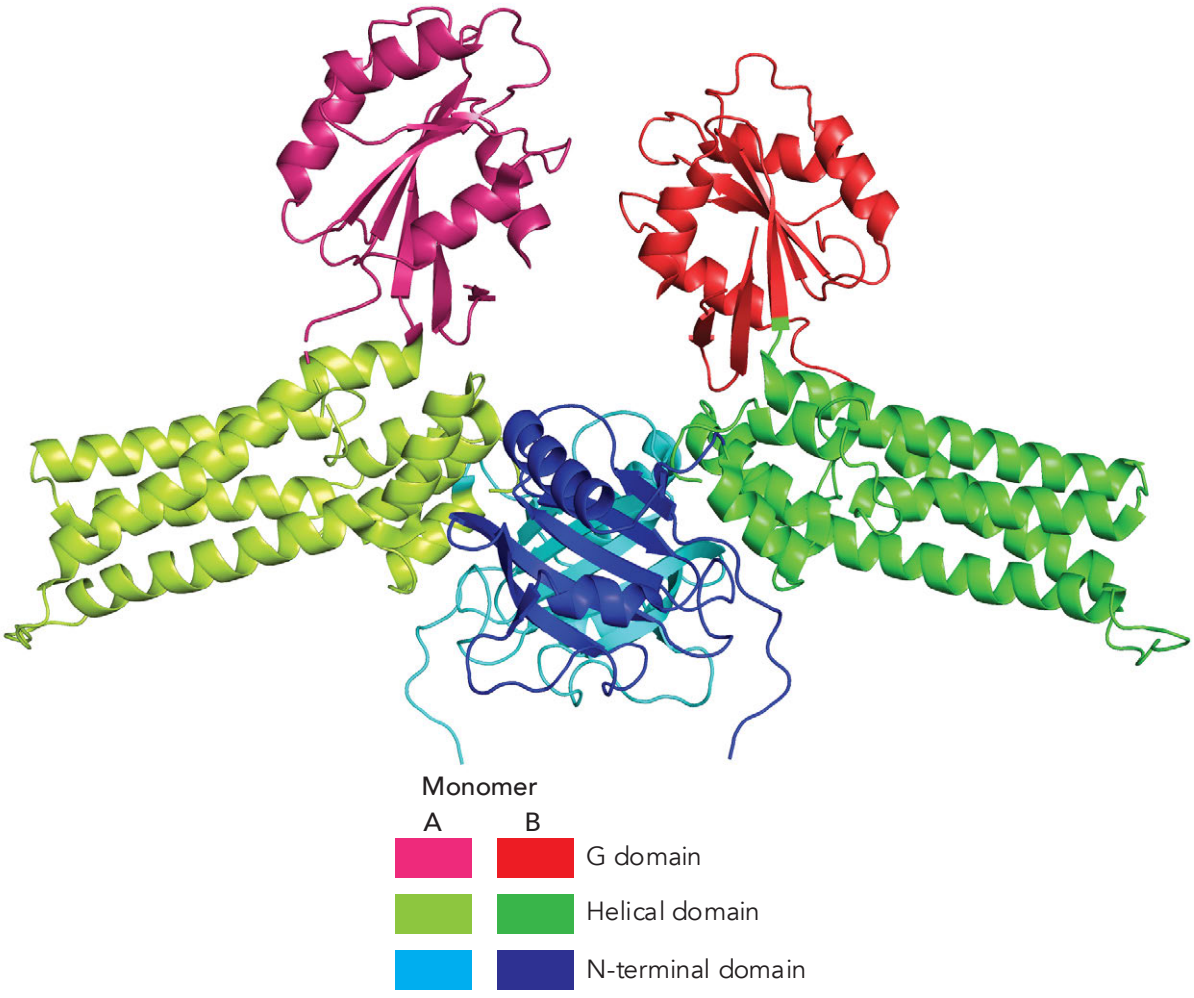


Figure 1.3 – MnmE structure

Ribbon representation of the symmetry model of dimeric MnmE (A) obtained from the X-ray crystal structures of *Chlorobium tepidum* MnmE (PDB: 3GEE). The domains of each protein are color coded according to their monomeric subunits.

In the presence of monovalent ions, the GTPase activity of MnmE increases exponentially, depending on the ionic radius of the ion ($K^+ \geq Rb^+ > Cs^+ > Na^+$).[186-188] MnmE's G-domain can be expressed independently, and the isolated domain maintains the catalytic activity observed in the full protein.[177,187] Size exclusion chromatography revealed that the G-domain acts as a monomeric protein except in the presence of K^+ and the transition-state analog GDP-AlF_x when there is a shift in its elution profile toward a dimeric state. The crystal structure of *E. coli* MnmE G-domain complexed with Mg^{2+}, K^+ and GDP-AlF_4^- revealed the large conformational changes that occur during the dimerization of the domain. This is particularly true at its switch regions, which form most of the inter-subunit interactions at the dimer interface. Here, switch I and II from the first subunit mainly interact with switch II and I from the second subunit, respectively. The AlF_4^- group predictably mimics the γ-phosphate in the transition state, interacting with the main chain amine of the threonine residues from the switch I GTTRD motif and the main chain amine of glycine provided by the switch II DxxG motif.[187] However, unlike classical Ras-like G-proteins, the regions preceding the two threonine residues is involved in binding a K^+ ion that has been identified as being essential for activity and G-domain dimerization. This loop (subsequently dubbed the K-loop) is in MnmE's active site in a similar location as the catalytic arginine finger in the Ras-GAP complex, which suggests that the K^+ may play a role in stabilizing the excess negative charges accumulating in the transition state.[189] Additionally, MnmE has a leucine residue adjacent to its DxxG motif in switch II, divergent from Ras-like GTPases that have a glutamine residue at that position.[190] This substitution is typical of HAS-GTPases (hydrophobic amino acid substituted for catalytic glutamine GTPases), but MnmE shows an alternative mechanism that couples the activation of a nucleophilic water to the dimerization of the G-domain.[191-194] In MnmE, the reorganisation of switch II during K^+-mediated dimerization of the G-domain moves a glutamate residue into the enzyme's active site and the glutamate activates the nucleophilic water that is probably required for catalytic function.[187]

1.4.2 MnmG

MnmG (formerly GidA) is a homodimeric protein with each monomer (~ 70 kDa) consisting of a FAD-binding domain, two insertion domains and a C-terminus domain (Figure 1.4).[178,180,195,196] The FAD-binding domain consists of a large and small β-sheet packed against each other and several

α-helices packed against the large sheet. The large β-sheet is a five-stranded sheet arranged in a Rossman fold, while the small β-sheet consists of three antiparallel strands that cross over the two halves of the Rossman fold, characterizing MnmG as a member of the glutathione reductase (GR) family of FAD-binding proteins.

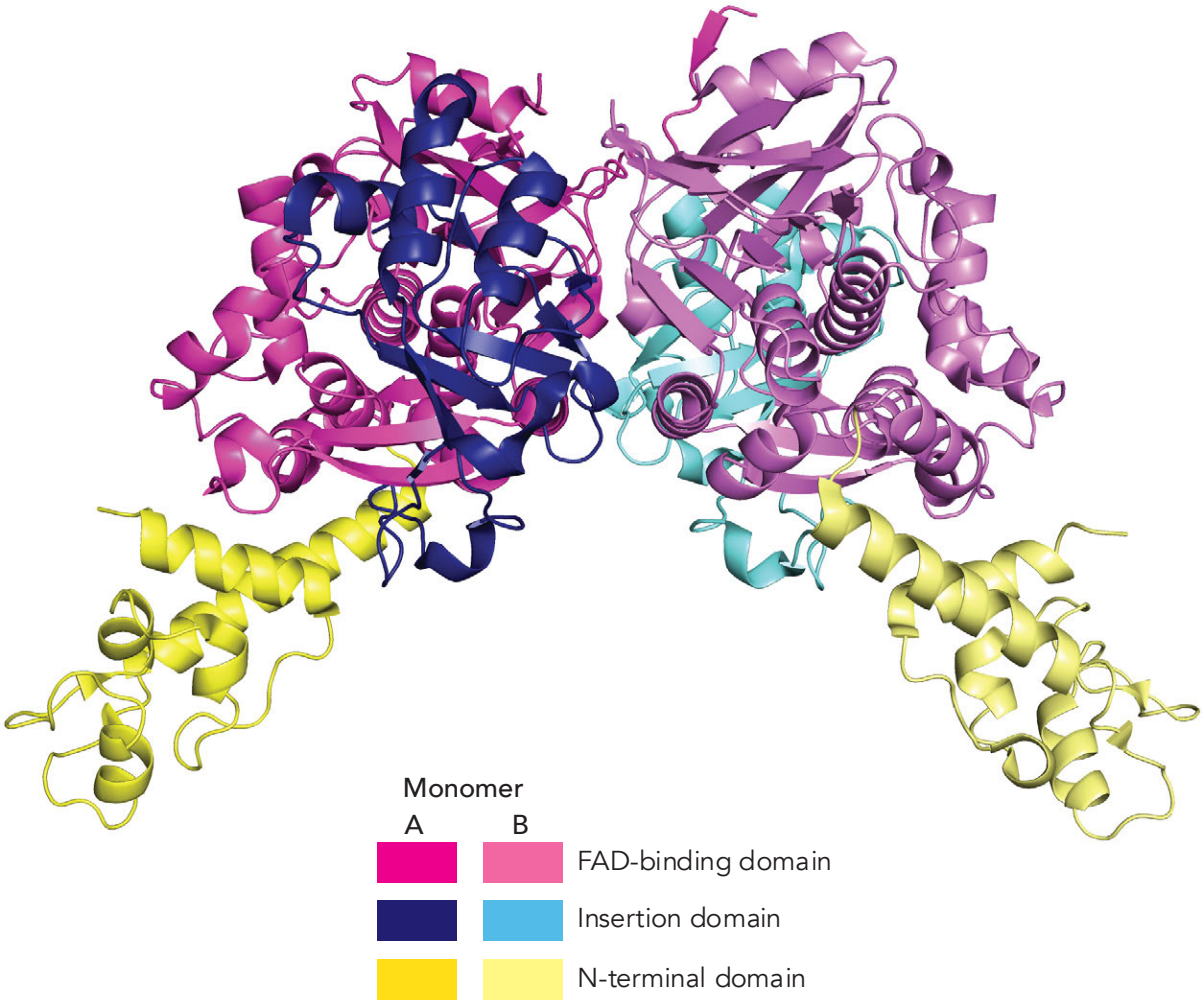


Figure 1.4 – MnmG structure

Ribbon representation of the symmetry model of dimeric MnmG obtained from the X-ray crystal structures of *Escherichia coli* MnmG (PDB: 3CES). The domains of each protein are color coded according to their monomeric subunits.

MnmG has two domains inserted within the FAD-binding domain, which is a defining feature of the GR_2 subfamily of FAD-binding proteins. These insertion domains include a small domain (mixed three-stranded β-sheet and 2 α-helices) inserted between the second and third b-strands of the Rossman fold, and a large domain (two β-sheets, 4- and 6-stranded antiparallel sheets, and 3 helices) inserted between the last two β-strands of the Rossman fold.[178,195,196] As is typical of FAD-binding proteins, MnmG has a dinucleotide-binding motif (DBM) motif (GxGHAGxEA), a loop region that interacts with the FAD moiety within the active site.[197] Also embedded within this domain are two MnmG-specific motifs that supposedly contribute to the positioning of the FAD cofactor within the protein's active site. The first motif is located within the second insertion domain, while the second motif is found at the C-terminus of the FAD-binding domain. The FAD moiety binds at the C-terminal end of the β-strands in the large sheet in an elongated conformation, with its ADP facing the five-stranded β-sheet and its isoalloxazine ring pointing away from the domain.[178,195,196] Interestingly, the FAD-binding domain of MnmG also contains a deep positively charged pocket, which was confirmed to be the tRNA binding site. This pocket is ideal for nucleic acid binding and there are no other domains in MnmG or MnmE that could neutralize this substrate's negative charge.[178] The second insertion domain of MnmG is proposed to be a NADH-binding domain due to high similarities between its topology and that of the small NADH-binding domain of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA reductase), which consists of a 4-stranded antiparallel β-sheet with cross-over helices on one side of the sheet.[198-200] Crystal structures of HMG-CoA reductase complexed with NADH and NAD^+ revealed that the cofactor binds at this domain's N-terminal that is highly analogous to part of the second insertion domain in MnmG.[199,200] Recall that this insertion domain also contains a conserved MnmG-specific motif that contributes to the orientation of the FAD moiety's isoalloxazine ring. Remarkably, this motif corresponds to the NADH-binding site in HMG-CoA reductase, and due to its conservation and high flexibility, this region has been proposed to play a role in substrate binding (glycine) in MnmG.[178] The C-terminal domain of MnmG is an all-helical domain, consisting of two helical bundles connected by two extended α-helices. The first two helices of this domain are tightly packed against the FAD- and NADH-binding domains, while the other helices are more flexible and can adopt different conformations. As far as we know, MnmG readily dimerizes *in vitro* without the aid of any

cofactors, and its dimer interface is made up of a strong network of interactions between the FAD- and NADH-binding domains of both subunits.[178,195,196]

1.4.3 The MnmE/MnmG modification complex

Experimental studies by Yim and colleagues revealed that MnmE and MnmG work interdependently, and the substate of the complex is contingent on cellular growth conditions. Under exponential growth conditions, the MnmEG complex preferentially uses glycine as a substrate and inserts a carboxymethylaminomethyl moiety at C5 of U34 in tRNA. Alternatively, under conditions of high cell density, ammonium is the preferential substrate and an aminomethyl group is added at C5 of the wobble base.[180,181,201,202] Using the preliminary experimental evidence available to them, Scrima and colleagues proposed a catalytic mechanism for the MnmEG complex (Figure 1.5).[179] According to their findings, a highly conserved cysteine residue at the C-terminus of MnmE attacks the double bond at C6 of U34, leading to the formation a carbocation at C5. MnmE-bound 5-formyl-THF is subsequently incorporated at C5 via a nucleophilic attack on the carboxylic group from the formyl entity. MnmE is regenerated through rearrangement in the intermediate that results in the expulsion of THF and the formation of 5-formyl uracil. To incorporate the carboxymethylamino moiety, the amine of glycine – presumably bound within MnmG's second insertion domain – attacks the adduct's carbonyl group, expels a water molecule and forms an intermediary Schiff base that is reduced by $FADH_2$.[179]

Following this proposal, consequent studies confirmed that the methyl donor is co-purified with MnmE, and that the modification reaction could proceed in the absence of an externally provided THF derivative. Furthermore, structural and biochemical studies proposed N^5-N^{10}-methylene tetrahydrofolate (5-methylene-THF) to be a better candidate for this reaction, as was observed in other bacterial uridine methylating enzymes like TrmFO and ThyA and the human homolog of MnmE, the GTPBP3 GTPase.[69,180,203-205] These new discoveries gave rise to a second scheme for the catalytic mechanism of the MnmEG complex by Yim and colleagues (Figure 1.5).[171] They propose that the MnmE-bound 5-methylene-THF is first converted to a reactive iminium ion, possibly by the activated nucleophilic water found in MnmE's active site. This highly reactive ion readily reacts with the amino group of glycine and the resulting adduct is dehydrogenated by MnmG-bound FAD. The iminium group is then transferred from

THF to C5 of U34, while a conserved cysteine residue in the C-terminal domain of MnmG carries out a nucleophilic attack on C6 of U34, thereby forming a covalently bound MnmG-tRNA adduct in an SN_2 mechanism. This adduct is reduced by $FADH_2$, C5 is deprotonated and the C6-MnmG bond is cleaved, completing the modification cycle.[171]

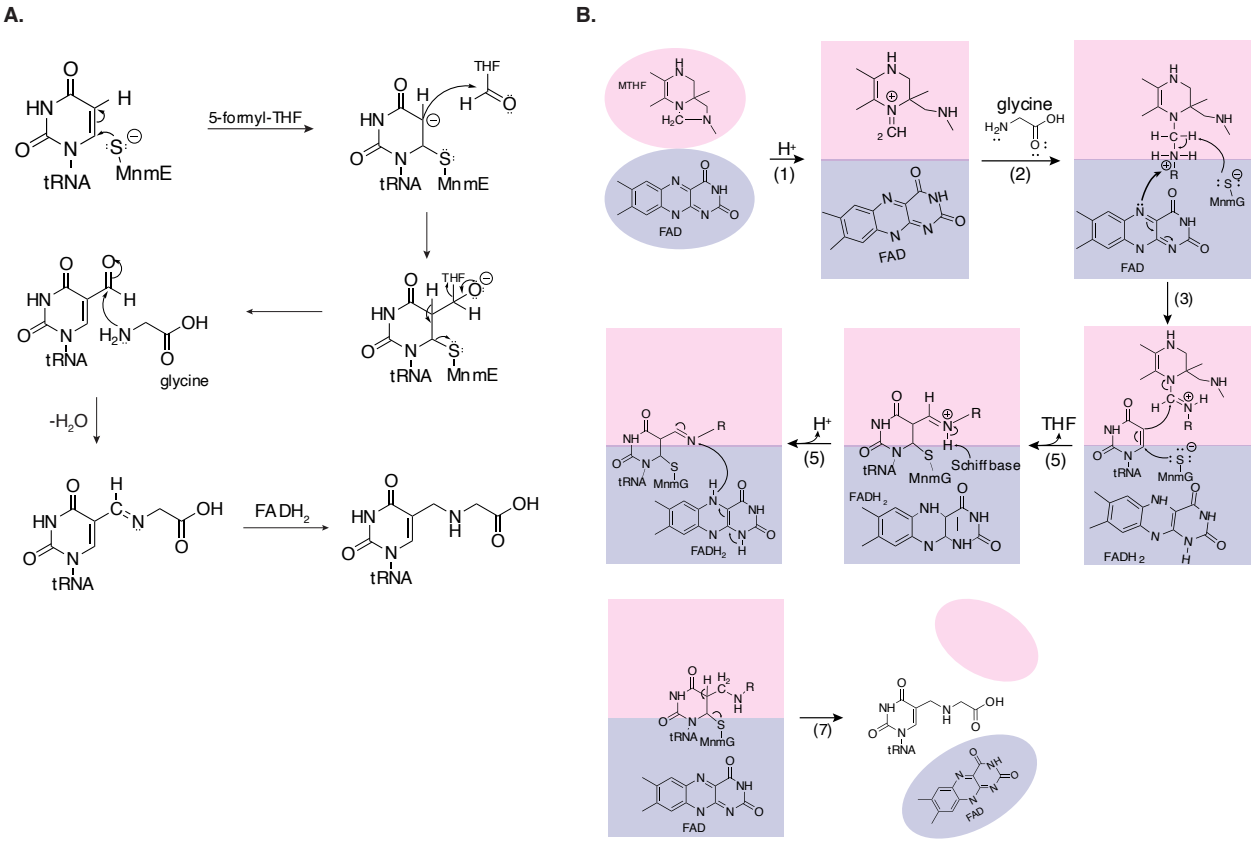


Figure 1.5 – Catalytic mechanism of the MnmE-MnmG complex

Biosynthetic pathways of the MnmEG complex hypothesized by Scrima *et al.*[179] (A) and Moukadiri *et al.*[202] (B) for the formation of cmnm^{5}U34. Scrima and colleagues proposed that the uracil base is activated at C5 by a cysteine residue in MnmE, enabling the transfer of a C1-group from 5-formyl-THF to the uridine base. Glycine is incorporated through the action of a Schiff base, which is then hydrogenated and reduced by MnmG–bound $FADH_2$. The proposal by Moukadiri and collaborators assumes that a general acid converts methylene-THF into a reactive iminium ion (1), which allows for the addition of the carboxylmethylamine group to the methylene group at N5 of THF (2). MnmG-bound FAD performs a dehydrogenation reaction (3), forming $FADH_2$ and facilitating a nucleophilic attack at C6 of the uracil base by a catalytic cysteine residue in MnmG (4). $FADH_2$ acts as a reducing agent of the Schiff base (5), and the latter initiates a series of rearrangement reactions (6,7), resulting in the formation cmnm^{5}U34.

Proposals for the catalytic mechanism of the MnmEG complex principally differ in the role assigned to GTP hydrolysis during the modification cycle (Figure 1.6). In the first scheme, GTP hydrolysis is

proposed to induce large conformational changes in MnmE that are distributed throughout the complex. These rearrangements are necessary for the cofactors from the two enzymes to be in proximity of each other and the reaction to occur efficiently.[177,180,195,201] Based on this catalytic mechanism, it was suggested that the enzymes came together in a symmetric manner, whereby the N- and C-terminal domains of both subunits of MnmE interacted with the FAD-binding domain and the C-terminal helices of MnmG, which concurred with the discovery that the MnmEG complex exists in an $\alpha_2\beta_2$ stoichiometry (Figure 1.7).[178,201]

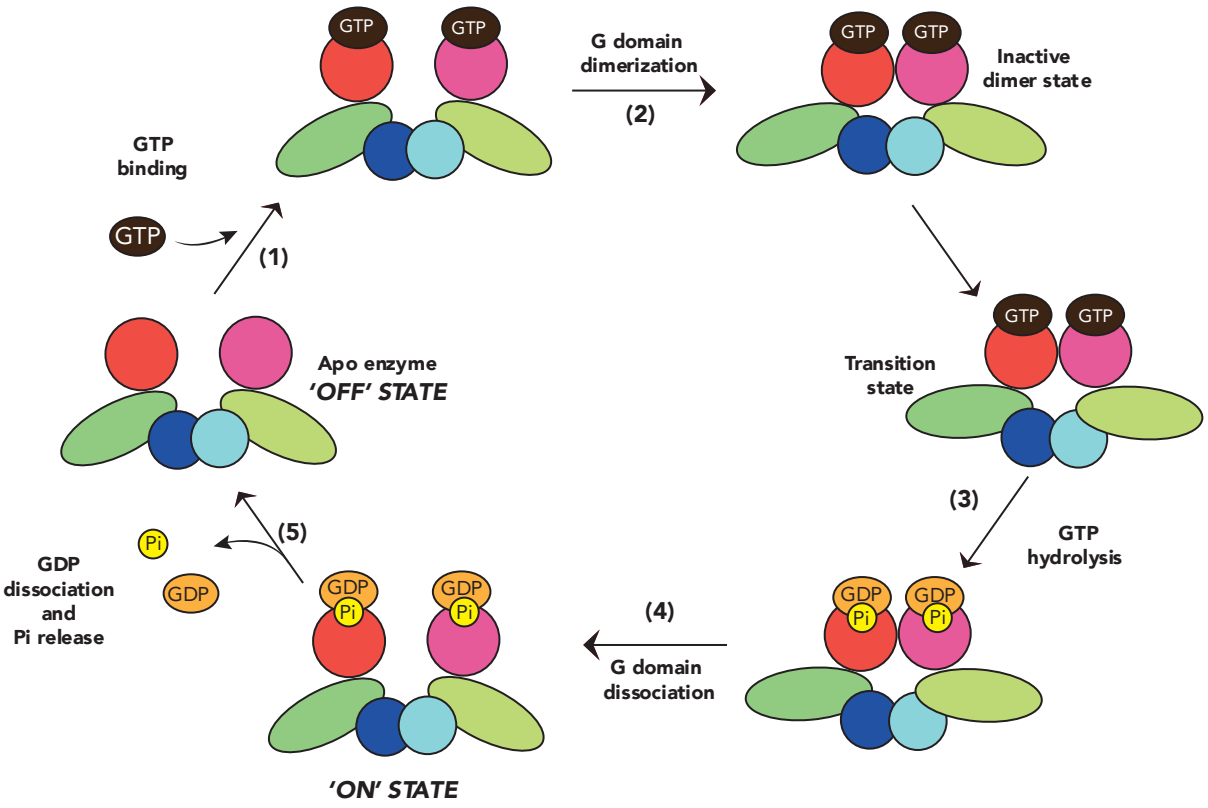


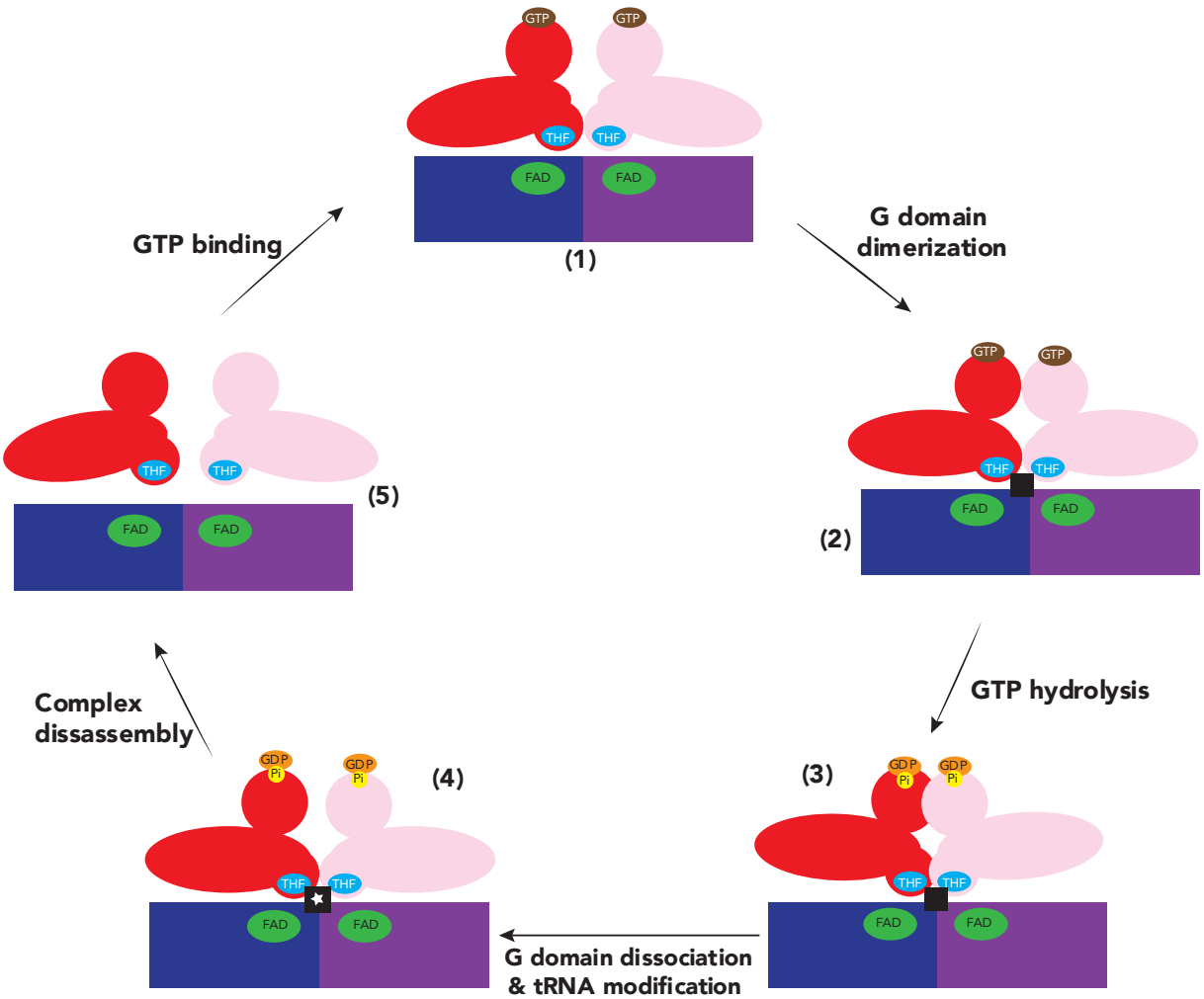
Figure 1.6 – "Kissing" mechanism of MnmE's G domains

Sequential conformational changes in MnmE during its GTPase cycle. MnmE is a dimeric protein, and each monomer is composed of three domains – an N-terminal domain (small circle), a helical domain (oval) and a G domain (large circle). Upon GTP binding (1), MnmE is reorganized (2,3), causing its G domains to move closer to each other. GTP hydrolysis causes the G domains to dissociate (4) and functionally activates MnmE ('ON' state). Finally, the release of Pi and GDP completes the GTPase cycle (5), and high concentrations of these molecules inhibit MnmE activity ('OFF' state). This figure has been adapted from Meyer *et al.*[206]

On the other hand, the second catalytic mechanism suggests that GTP hydrolysis may simply provide energy for the reaction to occur.[171,207] This role is coupled with the discovery that various guanine nucleotide binding events have different effects on the conformation of MnmE's subunits.[183,184,187,201,208]

Considering this, a second proposal was hypothesized for the formation of the MnmEG complex by Fislage and colleagues through a series of SAXS experiments, and they suggest that the two proteins come together asymmetrically, in an L-shaped complex (figure 1.7).[207] According to their study, the N-terminal and helical domains of one subunit from the MnmE dimer bind with the C-terminal domain of one subunit from the MnmG dimer, forming the $\alpha_2\beta_2$ complex. This complex bound one tRNA molecule between the THF-binding and FAD-binding domains of MnmE and MnmG, and conformational changes that occurred following GTP hydrolysis drew the proteins closer to the wobble base. Over the course of their investigations, they were also able to isolate an $\alpha_2\beta_2\alpha_2$ ($\alpha_4\beta_2$) complex, in which the second subunit of the already bound MnmG interacts with a subunit from another MnmE dimer in a similar manner as before. Their results showed that formation of the $\alpha_4\beta_2$ complex is linked to the GTPase cycle of MnmE, particularly GTP binding. The cycle begins in the GDP-bound state of the asymmetric $\alpha_2\beta_2$ complex. When GDP is replaced by GTP, the MnmE undergoes conformational rearrangements that induce allosteric changes on MnmG. These changes promote the binding of a second MnmE dimer, leading to the formation of the $\alpha_4\beta_2$ complex. Each subunit of the $\alpha_4\beta_2$ complex can bind a tRNA molecule, which is modified during (or prior to) GTP hydrolysis. Following the hydrolysis event, the complex dissociates and the $\alpha_2\beta_2$ complex is regenerated.[174,207] Unfortunately, no study has successfully isolated the fully assembled MnmEG complex, causing a standoff between the conflicting proposals in literature.

A.


GTP
GTP
THF
THF
FAD
FAD
(1)
G domain
dimerization
(2)
GTP hydrolysis
(3)
GDP
Pi
G domain dissociation
& tRNA modification
(4)
Complex
dissassembly
(5)
GTP binding

B.

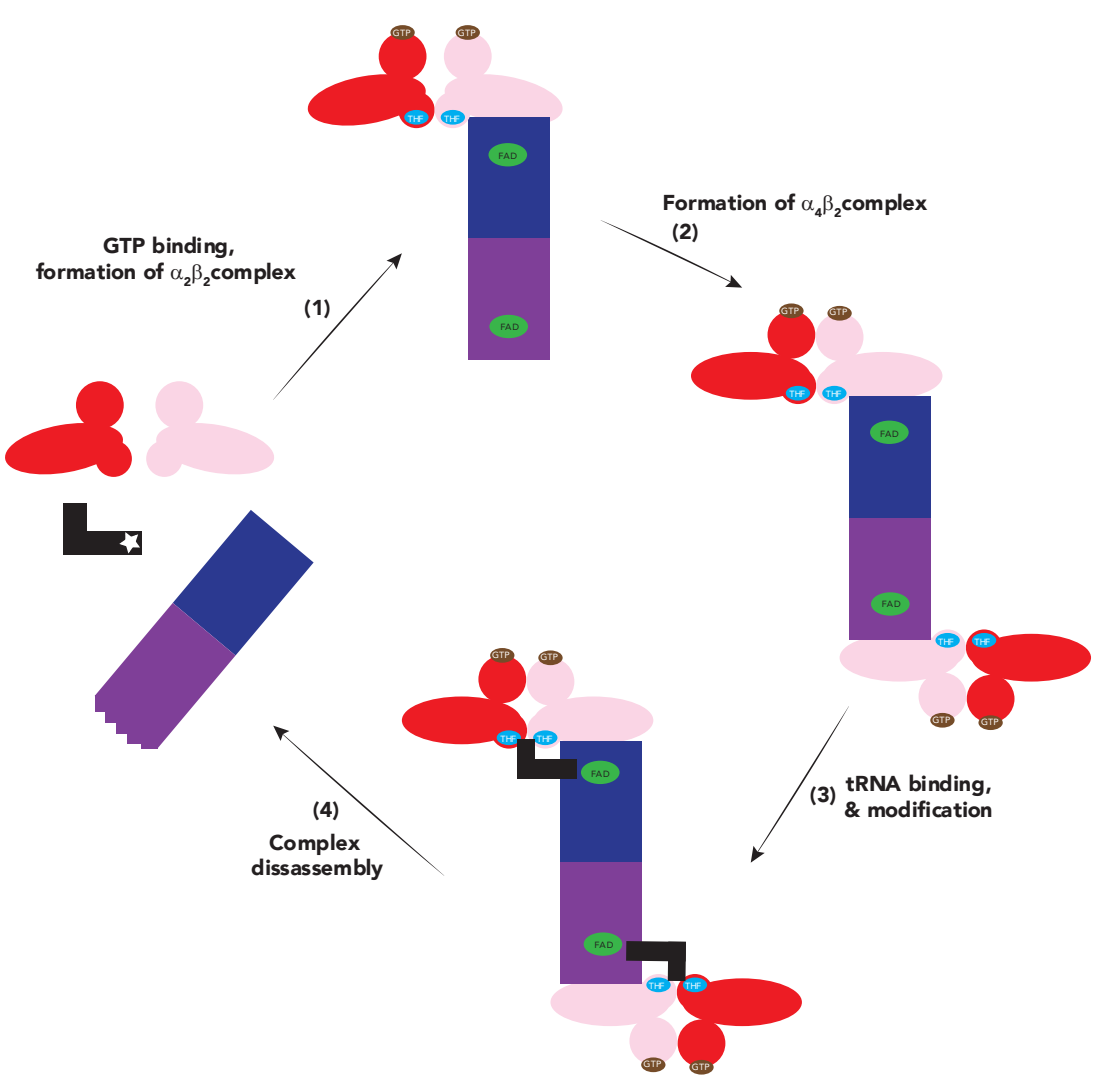


Figure 1.7 – Proposed models for the assembly and activity of MnmE-MnmG complex

(A) MnmE (red and pink) and MnmG (blue and purple) dimers come together in a symmetric $\alpha_2\beta_2$ complex, and the tRNA (black) binds the complex in a positively charged pocket located at the MnmE-MnmG interface (1). Upon GTP binding, the G domains dimerize, and conformational changes are distributed throughout the helical domain of MnmE, inducing GTP hydrolysis (2). In its GDP-bound state, the complex is activated, the G domains dissociate and the tRNA is modified (3,4). Following catalysis, the complex disassembles, and the proteins are ready to begin the next modification cycle (5). (B) When GDP is bound to MnmE, the MnmE and MnmG form an asymmetric $\alpha_2\beta_2$ complex. G domain dimerization is induced by the binding of MnmG to MnmE (1). GTP binding constricts the G domain dimers, causing an upward movement in the helical domains of MnmE (2). This conformational change allows a second MnmE dimer to bind to MnmG, resulting in the formation of an $\alpha_4\beta_2$ complex (3). Next one (possibly two) tRNA molecules bind to the complex, interacting with the FAD binding domain of MnmG and the N-terminal domain of MnmE. Upon GTP hydrolysis, the tRNA is modified and released and the complex returns to its $\alpha_2\beta_2$ state (4). Cofactors are color-coded: GTP – brown, THF – light blue, FAD – green, GDP – orange, Pi – yellow. Figures were adapted from studies conducted by Armengod *et al.*[201] (A) and Fislage *et al.*[207]

1.5 The Modifications and Disease

As previously discussed, the unmodified wobble uridine is able to recognize all four nucleotides during translation and this "four-way wobbling" phenomenon allows a single tRNA molecule to read four codons within a family box.[73,74] On the other hand, modified uridines, particularly members of the xm^5U34 family, have been found to restrict decoding by the wobble base to pyrimidine bases (A and G). The x-methyl substituent at C5 of the uracil base is believed to enable efficient of G-ending codons, while the 2-thio group in xm^5s^2U34 derivatives allegedly promotes the decoding of purine-ending codons.[169,209,210] Therefore, theoretically, $cmnm^5U34$-modified mt-$tRNA^{Leu}_{UUA}$ is able to read UAA and UUG codons, while $cmnm^5s^2U34$-modified mt-$tRNA^{Lys}_{UUU}$ translates AAA and AAG codons.[211] Proposals suggest that in the absence of τm^5U34, $tRNA^{Leu}$ only recognizes the UUA codon, meanwhile $tRNA^{Lys}$ lacking τm^5s^2U34 is unable to read any of its cognate codons. Additionally, a correlation was found between the lack of τm^5s^2U34 modifications and a substantial reduction in overall translation in mitochondria.[212] Of the thirteen respiratory chain complex genes encoded in mt-DNA, ND6 – a subunit of respiratory chain complex I, has the highest usage of UUG codons and its translation is heavily inhibited by the loss of $\tau m^5(s^2)U34$ modifications, leading to defective respiratory activity, a symptom directly linked to mitochondrial dysfunction.[11,14,212]

Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS), and myoclonus epilepsy associated with ragged-red fibers (MERRF) are diseases linked to the anomalous $\tau m^5(s^2)U34$ modification in mitochondrial $tRNA^{Leu}_{UUA}$ and $tRNA^{Lys}_{UUU}$.[12,213,214] MELAS is a childhood disorder that predominantly affects the nervous system. It is characterized by seizures, recurrent headaches, loss of appetite, recurrent vomiting and stroke-like episodes that affect visual, auditory, motor, and intellectual skills. MERRF is characterized by myoclonus epilepsy (muscle twitches), myopathy (muscle weakness), short stature, hearing loss, lactic acidosis, and exercise intolerance.[215] Three reasons have been identified for the presence of these anomalies in mitochondria, and these include mutations within the tRNA genes, pathogenic mutations within the modification enzymes and taurine starvation.[73,170,212,216-218] Interestingly, the mutations within the mt-tRNA genes of leucine and lysine responsible for MELAS and MERRF lie outside the anticodon arm, usually within the D-loop.[170,211,212] Loss of function mutations in the

MTO1 and *GTPBP3* genes are associated with hypertrophic cardiomyopathy and lactic acidosis. Moreover, patients with *GTPBP3* mutations can develop Leigh syndrome, a progressive encephalopathy linked to mitochondrial dysfunction.[13,216,218] Mutations in the mitochondrial tRNA-specific-2-thiouridylase 1 (MTU1) cause a reduction of τm^5s^2U in mt-tRNA^{Lys} and loss of function mutations in this enzyme are linked to reversible infantile liver failure (RILF).[219,220] Taurine is an essential amino acid in carnivores required for efficient development in many animals and the biogenesis of a number of tRNA modifications. When taurine is depleted, glycine is inserted at C5 of U34 in mt-tRNAs, leading to the formation of $cmnm^5(s^2)U34$ modifications that are observed in bacteria.[11,221]

There is no cure for either MELAS or MERRF, and treatment plans mainly focus on symptom management, which varies among patients. Excitingly, reports on the association between taurine starvation and MELAS prompted the use of orally administered taurine as a remedy for the disease.[222] Clinical trials revealed that this therapeutic suppressed stroke recurrence in individuals with MELAS and the drug was recently (2019) approved as a MELAS treatment in Japan.[223] Apart from this, one investigation has studied the human proteins for therapeutic purposes, but fruitful results are yet to be attained.[224]

1.6 Thesis Objectives

The goal of this thesis is to provide more insight into the MnmEG modification complex and work towards the structural isolation of the assembled complex. In this work, particular focus is placed on the binding properties of the complex and the effects of the resulting modifications on tRNA structure and function. In Chapter 2, biochemical studies will be conducted with the aim of gaining a better understanding of the MnmEG modification complex. So far, there have been multiple proposals pertaining to how this complex is formed and independent crystal structures exist for MnmE and MnmG.[178,207] However, there is no crystal structure of the full complex and there are no other structural investigations that confirm or deny the ideas brought forward by the previously reviewed experimental studies. Therefore, more explorations are needed to gain a comprehensive understanding of the modification engine, with the overarching goal of elucidating the structure of the MnmEG complex. To this end,

Chapter 2 uses binding assays to investigate the binding affinity of MnmE, MnmG and the entire modification complex under various conditions, providing insight into protein–tRNA interactions within the complex. Additionally, different RNA substrates were investigated to probe the specificity of each protein component and the complex at large. Having concrete evidence for the binding propensities of this complex will promote further probing into its catalytic mechanism i.e., modification insertion, as this will allow physically appropriate analyses to be made on the two proposed mechanisms in literature. Moreover, MnmE and MnmG are homologous to their human counterparts, GTPBP3 and MTO1, and mutations in these enzymes are heavily linked to defects in mitochondrial function.[11,13,14,170,201,212,217,218] Consequently, the structures of the bacterial enzymes will provide a starting point for investigations on the proteins and their therapeutic potential in MELAS and MERRF treatments.

To further understand the role of the MnmEG complex, this thesis uses computational methods to probe the effects of $cmnm^5(s^2)U34$ on tRNA structure. Although the general role xm^5U34 type tRNA modifications play in translation is known,[167-169] the specific role of $cmnm^5U34$ and its derivatives remains elusive, as there are no studies that investigate the effects of this modification group on tRNA structure and consequently function. Furthermore, unlike modifications like 5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U34),[225,226] no investigation has considered how $cmnm^5(s^2)U34$ modified tRNA is accommodated within the ribosome and their role within the translation machinery is unknown. As previously mentioned, there is a lack of consensus in the literature on how to investigate tRNA structures using computational techniques. Consequently, a computational investigation on tRNA will be carried out in Chapter 3 of this thesis, to determine how to accurately describe the tRNA conformational space using MD simulations. Then, an atomic-level study on the structural and functional roles of $cmnm^5(s^2)U34$ modifications within the context of tRNA structure will be conducted in Chapter 4, applying the computational protocol developed in the preceding chapter. This study will offer insights into the modifications' roles during codon reading and assess the proposals brought forward by previous experimental studies.[5,20,53,227] Finally, Chapter 5 recapitulates the major findings observed over the course of this thesis and provides avenues for future studies. Altogether, this work uses a two-pronged approach to study the insertion and function of $cmnm^5(s^2)U34$ modifications and provides a greater understanding of the modification mechanism and role during protein synthesis.

1.7 References

1 Chastain, M. & Tinoco, I., Jr. Structural elements in RNA. *Prog Nucleic Acid Res Mol Biol* **41**, 131-177, doi:10.1016/s0079-6603(08)60008-2 (1991).

2 Holley, R. W. *et al.* Structure of a ribonucleic acid. *Science* **147**, 1462-1465, doi:10.1126/science.147.3664.1462 (1965).

3 Holley, R. W., Everett, G. A., Madison, J. T. & Zamir, A. Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *J Biol Chem* **240**, 2122-2128 (1965).

4 Caprara, M. G. & Nilsen, T. W. RNA: versatility in form and function. *Nat Struct Biol* **7**, 831-833, doi:10.1038/82816 (2000).

5 Agris, P. F. Decoding the genome: a modified view. *Nucleic Acids Res* **32**, 223-238, doi:10.1093/nar/gkh185 (2004).

6 Agris, P. F. The importance of being modified: an unrealized code to RNA structure and function. *RNA* **21**, 552-554, doi:10.1261/rna.050575.115 (2015).

7 Agris, P. F. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* **9**, 629-635, doi:10.1038/embor.2008.104 (2008).

8 Bjork, G. R. & Hagervall, T. G. Transfer RNA Modification: Presence, Synthesis, and Function. *EcoSal Plus* **6**, doi:10.1128/ecosalplus.ESP-0007-2013 (2014).

9 Bjork, G. R. *et al.* Transfer RNA modification. *Annu Rev Biochem* **56**, 263-287, doi:10.1146/annurev.bi.56.070187.001403 (1987).

10 Grosjean, H. in *Fine-Tuning of RNA Functions by Modification and Editing* (ed Henri Grosjean) 1-22 (Springer Berlin Heidelberg, 2005).

11 Asano, K. *et al.* Metabolic and chemical regulation of tRNA modification associated with taurine deficiency and human disease. *Nucleic Acids Res* **46**, 1565-1583, doi:10.1093/nar/gky068 (2018).

12 Kirino, Y., Goto, Y.-i., Campos, Y., Arenas, J. & Suzuki, T. Specific correlation between the wobble modification deficiency in mutant tRNAs and the clinical features of a human mitochondrial disease. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7127-7132, doi:10.1073/pnas.0500563102 (2005).

13 Peng, G.-X. *et al.* The human tRNA taurine modification enzyme GTPBP3 is an active GTPase linked to mitochondrial diseases. *Nucleic Acids Res* **49**, 2816-2834, doi:10.1093/nar/gkab104 (2021).

14 Torres, A. G., Batlle, E. & Ribas de Pouplana, L. Role of tRNA modifications in human diseases. *Trends Mol Med* **20**, 306-314, doi:10.1016/j.molmed.2014.01.008 (2014).

15 Crick, F. H. On protein synthesis. *Symp Soc Exp Biol* **12**, 138-163 (1958).

16 Crick, F. H., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192**, 1227-1232, doi:10.1038/1921227a0 (1961).

17 Dounce, A. L. [Duplicating mechanism for peptide chain and nucleic acid synthesis]. *Enzymologia* **15**, 251-258 (1952).

18 Nirenberg, M. *et al.* RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc Natl Acad Sci U S A* **53**, 1161-1168, doi:10.1073/pnas.53.5.1161 (1965).

19 Sharp, S. J., Schaack, J., Cooley, L., Burke, D. J. & Soll, D. Structure and transcription of eukaryotic tRNA genes. *CRC Crit Rev Biochem* **19**, 107-144, doi:10.3109/10409238509082541 (1985).

20 Berg, M. D. & Brandl, C. J. Transfer RNAs: diversity in form and function. *RNA Biol* **18**, 316-339, doi:10.1080/15476286.2020.1809197 (2021).

21 Nissen, P., Kjeldgaard, M., Thirup, S., Clark, B. F. C. & Nyborg, J. The ternary complex of aminoacylated tRNA and EF-Tu-GTP. Recognition of a bond and a fold. *Biochimie* **78**, 921-933, doi:Doi 10.1016/S0300-9084(97)86714-4 (1996).

22 Maglott, E. J., Goodwin, J. T. & Glick, G. D. Probing the structure of an RNA tertiary unfolding transition state. *Journal of the American Chemical Society* **121**, 7461-7462, doi:DOI 10.1021/ja9913075 (1999).

23 Kim, S. H. *et al.* Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **185**, 435-440, doi:10.1126/science.185.4149.435 (1974).

24 Shi, H. J. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 angstrom resolution: A classic structure revisited. *RNA* **6**, 1091-1105, doi:Doi 10.1017/S1355838200000364 (2000).

25 Woo, N. H., Roe, B. A. & Rich, A. Three-dimensional structure of Escherichia coli initiator tRNAfMet. *Nature* **286**, 346-351, doi:10.1038/286346a0 (1980).

26 Zhang, J. & Ferre-D'Amare, A. R. The tRNA Elbow in Structure, Recognition and Evolution. *Life (Basel)* **6**, doi:10.3390/life6010003 (2016).

27 Chapeville, F. *et al.* On the role of soluble ribonucleic acid in coding for amino acids. *Proc Natl Acad Sci U S A* **48**, 1086-1092, doi:10.1073/pnas.48.6.1086 (1962).

28 Crothers, D. M., Seno, T. & Söll, D. G. Is There a Discriminator Site in Transfer RNA? *Proceedings of the National Academy of Sciences* **69**, 3063-3067, doi:10.1073/pnas.69.10.3063 (1972).

29 Hall, K. B., Sampson, J. R., Uhlenbeck, O. C. & Redfield, A. G. Structure of an unmodified tRNA molecule. *Biochemistry* **28**, 5794-5801, doi:10.1021/bi00440a014 (1989).

30 Balzi, E., Choder, M., Chen, W. N., Varshavsky, A. & Goffeau, A. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of Saccharomyces cerevisiae. *J Biol Chem* **265**, 7464-7471 (1990).

31 Kaji, H., Novelli, G. D. & Kaji, A. A Soluble Amino Acid-Incorporating System from Rat Liver. *Biochim Biophys Acta* **76**, 474-477 (1963).

32 Rao, K. R. & Tarver, H. A soluble amino acid-incorporating system from the fat-rich supernatant fractions of rat liver. *Arch Biochem Biophys* **141**, 758-761, doi:10.1016/0003-9861(70)90194-3 (1970).

33 Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science* **254**, 1374-1377, doi:10.1126/science.1962196 (1991).

34 Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**, 1298-1345, doi:10.1002/pro.666 (2011).

35 Francklyn, C. S. & Minajigi, A. tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett* **584**, 366-375, doi:10.1016/j.febslet.2009.11.045 (2010).

36 Katz, A., Elgamal, S., Rajkovic, A. & Ibba, M. Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology. *Mol Microbiol* **101**, 545-558, doi:10.1111/mmi.13419 (2016).

37 Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223-1237, doi:10.1016/s0092-8674(05)80070-4 (1992).

38 Hinnebusch, A. G. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59**, 407-450, doi:10.1146/annurev.micro.59.031805.133833 (2005).

39 Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745, doi:10.1016/j.cell.2009.01.042 (2009).

40 Jin, D. & Musier-Forsyth, K. Role of host tRNAs and aminoacyl-tRNA synthetases in retroviral replication. *J Biol Chem* **294**, 5352-5364, doi:10.1074/jbc.REV118.002957 (2019).

41 Sawyer, R. C. & Dahlberg, J. E. Small RNAs of Rous sarcoma virus: characterization by two-dimensional polyacrylamide gel electrophoresis and fingerprint analysis. *J Virol* **12**, 1226-1237, doi:10.1128/JVI.12.6.1226-1237.1973 (1973).

42 Gebetsberger, J. & Polacek, N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol* **10**, 1798-1806, doi:10.4161/rna.27177 (2013).

43 Raina, M. & Ibba, M. tRNAs as regulators of biological processes. *Front Genet* **5**, 171, doi:10.3389/fgene.2014.00171 (2014).

44 Oberbauer, V. & Schaefer, M. R. tRNA-Derived Small RNAs: Biogenesis, Modification, Function and Potential Impact on Human Disease Development. *Genes (Basel)* **9**, doi:10.3390/genes9120607 (2018).

45 Haussecker, D. *et al.* Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**, 673-695, doi:10.1261/rna.2000810 (2010).

46 Kuscu, C. *et al.* tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* **24**, 1093-1105, doi:10.1261/rna.066126.118 (2018).

47 Lee, S. R. & Collins, K. Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. *J Biol Chem* **280**, 42744-42749, doi:10.1074/jbc.M510356200 (2005).

48 Thompson, D. M., Lu, C., Green, P. J. & Parker, R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14**, 2095-2103, doi:10.1261/rna.1232808 (2008).
49 Marquet, R., Isel, C., Ehresmann, C. & Ehresmann, B. tRNAs as primer of reverse transcriptases. *Biochimie* **77**, 113-124, doi:10.1016/0300-9084(96)88114-4 (1995).
50 Bou-Nader, C. *et al.* HIV-1 matrix-tRNA complex structure reveals basis for host control of Gag localization. *Cell Host Microbe* **29**, 1421-1436.e1427, doi:10.1016/j.chom.2021.07.006 (2021).
51 Larsen, K. P. *et al.* Architecture of an HIV-1 reverse transcriptase initiation complex. *Nature* **557**, 118-122, doi:10.1038/s41586-018-0055-9 (2018).
52 Zhang, J. Interplay between Host tRNAs and HIV-1: A Structural Perspective. *Viruses* **13**, 1819, doi:10.3390/v13091819 (2021).
53 Abbink, T. E. M. & Berkhout, B. HIV-1 reverse transcription initiation: A potential target for novel antivirals? *Virus Research* **134**, 4-18, doi:https://doi.org/10.1016/j.virusres.2007.12.009 (2008).
54 Carine, T. Structural Bases of the Annealing of Primer Lys tRNA to the HIV-1 Viral RNA. *Current HIV Research* **3**, 147-156, doi:http://dx.doi.org/10.2174/1570162053506919 (2005).
55 Gremminger, T. *et al.* Extended Interactions between HIV-1 Viral RNA and tRNA(Lys3) Are Important to Maintain Viral RNA Integrity. *International journal of molecular sciences* **22**, 58, doi:10.3390/ijms22010058 (2020).
56 Kang, S. M., Wakefield, J. K. & Morrow, C. D. Mutations in both the U5 region and the primer-binding site influence the selection of the tRNA used for the initiation of HIV-1 reverse transcription. *Virology* **222**, 401-414, doi:10.1006/viro.1996.0437 (1996).
57 Kelly, N. J. & Morrow, C. D. Yeast tRNA(Phe) expressed in human cells can be selected by HIV-1 for use as a reverse transcription primer. *Virology* **313**, 354-363, doi:10.1016/s0042-6822(03)00243-5 (2003).
58 Yu, Q. & Morrow, C. D. Essential regions of the tRNA primer required for HIV-1 infectivity. *Nucleic Acids Res* **28**, 4783-4789, doi:10.1093/nar/28.23.4783 (2000).
59 Yu, Q. & Morrow, C. D. Identification of critical elements in the tRNA acceptor stem and T(Psi)C loop necessary for human immunodeficiency virus type 1 infectivity. *J Virol* **75**, 4902-4906, doi:10.1128/jvi.75.10.4902-4906.2001 (2001).
60 Ma, J., Zhang, L., Chen, S. & Liu, H. A brief review of RNA modification related database resources. *Methods*, doi:https://doi.org/10.1016/j.ymeth.2021.03.003 (2021).
61 Lorenz, C., Lünse, C. E. & Mörl, M. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* **7**, 35 (2017).
62 Suzuki, T. *et al.* Complete chemical structures of human mitochondrial tRNAs. *Nature Communications* **11**, 4269, doi:10.1038/s41467-020-18068-6 (2020).
63 Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways-2013 update. *Nucleic Acids Res* **41**, D262-D267, doi:10.1093/nar/gks1007 (2013).
64 Machnicka, M. A., Olchowik, A., Grosjean, H. & Bujnicki, J. M. Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol* **11**, 1619-1629, doi:10.4161/15476286.2014.992273 (2014).
65 El Yacoubi, B., Bailly, M. & de Crecy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet* **46**, 69-95, doi:10.1146/annurev-genet-110711-155641 (2012).
66 Helm, M. *et al.* The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res* **26**, 1636-1643, doi:10.1093/nar/26.7.1636 (1998).
67 Herschlag, D. RNA chaperones and the RNA folding problem. *J Biol Chem* **270**, 20871-20874, doi:10.1074/jbc.270.36.20871 (1995).
68 Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip Rev RNA* **4**, 35-48, doi:10.1002/wrna.1144 (2013).
69 Urbonavicius, J., Armengaud, J. & Grosjean, H. Identity elements required for enzymatic formation of N2,N2-dimethylguanosine from N2-monomethylated derivative and its possible role in avoiding alternative conformations in archaeal tRNA. *J Mol Biol* **357**, 387-399, doi:10.1016/j.jmb.2005.12.087 (2006).
70 Keffer-Wilkes, L. C., Soon, E. F. & Kothe, U. The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain. *Nucleic Acids Res* **48**, 7981-7990, doi:10.1093/nar/gkaa548 (2020).

71 Keffer-Wilkes, L. C., Veerareddygari, G. R. & Kothe, U. RNA modification enzyme TruB is a tRNA chaperone. *Proceedings of the National Academy of Sciences* **113**, 14306, doi:10.1073/pnas.1607512113 (2016).
72 Bock, R. M. Structure-function relations in tRNA. *Basic Life Sci* **1**, 189-196 (1973).
73 Suzuki, T. in *Fine-Tuning of RNA Functions by Modification and Editing* (ed Henri Grosjean) 23-69 (Springer Berlin Heidelberg, 2005).
74 Crick, F. H. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* **19**, 548-555, doi:10.1016/s0022-2836(66)80022-0 (1966).
75 Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721-732, doi:10.1016/s0092-8674(02)01086-3 (2002).
76 Barrell, B. G. *et al.* Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc Natl Acad Sci U S A* **77**, 3164-3166, doi:10.1073/pnas.77.6.3164 (1980).
77 Bonitz, S. G. *et al.* Codon recognition rules in yeast mitochondria. *Proc Natl Acad Sci U S A* **77**, 3167-3170, doi:10.1073/pnas.77.6.3167 (1980).
78 Inagaki, Y. *et al.* Translation of Synonymous Codons in Family Boxes byMycoplasma capricolumtRNAs with Unmodified Uridine or Adenosine at the First Anticodon Position. *Journal of Molecular Biology* **251**, 486-492, doi:https://doi.org/10.1006/jmbi.1995.0450 (1995).
79 Andachi, Y., Yamao, F., Iwami, M., Muto, A. & Osawa, S. Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in Mycoplasma capricolum. *Proc Natl Acad Sci U S A* **84**, 7398-7402, doi:10.1073/pnas.84.21.7398 (1987).
80 Grosjean, H. Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour. *Madame Curie Bioscience Database* (2013).
81 Björk, G. R. *et al.* TRANSFER RNA MODIFICATION. *Annual Review of Biochemistry* **56**, 263-285, doi:10.1146/annurev.bi.56.070187.001403 (1987).
82 Khade, P. & Joseph, S. Functional interactions by transfer RNAs in the ribosome. *FEBS Lett* **584**, 420-426, doi:10.1016/j.febslet.2009.11.034 (2010).
83 Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci* **25**, 1390-1406, doi:10.1002/pro.2950 (2016).
84 Green, R. & Noller, H. F. Ribosomes and translation. *Annu Rev Biochem* **66**, 679-716, doi:10.1146/annurev.biochem.66.1.679 (1997).
85 Baßler, J. & Hurt, E. Eukaryotic Ribosome Assembly. *Annual Review of Biochemistry* **88**, 281-306, doi:10.1146/annurev-biochem-013118-110817 (2019).
86 Shajani, Z., Sykes, M. T. & Williamson, J. R. Assembly of Bacterial Ribosomes. *Annual Review of Biochemistry* **80**, 501-526, doi:10.1146/annurev-biochem-062608-160432 (2011).
87 Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520**, 640-645, doi:10.1038/nature14427 (2015).
88 Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* **11**, 113-127, doi:10.1038/nrm2838 (2010).
89 Simonetti, A. *et al.* A structural view of translation initiation in bacteria. *Cell Mol Life Sci* **66**, 423-436, doi:10.1007/s00018-008-8416-4 (2009).
90 Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416-420, doi:10.1038/nature07192 (2008).
91 Kothe, U. & Rodnina, M. V. Codon Reading by tRNAAla with Modified Uridine in the Wobble Position. *Molecular Cell* **25**, 167-174, doi:https://doi.org/10.1016/j.molcel.2006.11.014 (2007).
92 Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142-148, doi:10.1038/342142a0 (1989).
93 Pape, T., Wintermeyer, W. & Rodnina, M. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *Embo J* **18**, 3800-3807, doi:10.1093/emboj/18.13.3800 (1999).
94 Valle, M. *et al.* Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol* **10**, 899-906, doi:10.1038/nsb1003 (2003).
95 Savir, Y. & Tlusty, T. Conformational proofreading: the impact of conformational changes on the specificity of molecular recognition. *PLoS One* **2**, e468, doi:10.1371/journal.pone.0000468 (2007).
96 Savir, Y. & Tlusty, T. The Ribosome as an Optimal Decoder: A Lesson in Molecular Recognition. *Cell* **153**, 471-479, doi:10.1016/j.cell.2013.03.032 (2013).

97 Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S. & Puglisi, J. D. tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol* **11**, 1008-1014, doi:10.1038/nsmb831 (2004).
98 Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256-259, doi:10.1038/nature10913 (2012).
99 Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930, doi:10.1126/science.289.5481.920 (2000).
100 Rodnina, M. V. The ribosome as a versatile catalyst: reactions at the peptidyl transferase center. *Current Opinion in Structural Biology* **23**, 595-602, doi:https://doi.org/10.1016/j.sbi.2013.04.012 (2013).
101 Wohlgemuth, I., Beringer, M. & Rodnina, M. V. Rapid peptide bond formation on isolated 50S ribosomal subunits. *EMBO reports* **7**, 699-703, doi:https://doi.org/10.1038/sj.embor.7400732 (2006).
102 Torres de Farias, S., Gaudencio Rego, T. & Jose, M. V. Peptidyl Transferase Center and the Emergence of the Translation System. *Life (Basel)* **7**, doi:10.3390/life7020021 (2017).
103 Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell* **25**, 505-517, doi:10.1016/j.molcel.2007.01.022 (2007).
104 Rodnina, M. V., Savelsbergh, A., Katunin, V. I. & Wintermeyer, W. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* **385**, 37-41, doi:10.1038/385037a0 (1997).
105 Spiegel, P. C., Ermolenko, D. N. & Noller, H. F. Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome. *RNA* **13**, 1473-1482, doi:10.1261/rna.601507 (2007).
106 Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318-322, doi:10.1038/35018597 (2000).
107 Peske, F., Savelsbergh, A., Katunin, V. I., Rodnina, M. V. & Wintermeyer, W. Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. *J Mol Biol* **343**, 1183-1194, doi:10.1016/j.jmb.2004.08.097 (2004).
108 Rodnina, M. V. & Wintermeyer, W. The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation. *Biochem Soc Trans* **39**, 658-662, doi:10.1042/BST0390658 (2011).
109 Korostelev, A. A. Structural aspects of translation termination on the ribosome. *RNA* **17**, 1409-1421, doi:10.1261/rna.2733411 (2011).
110 Laurberg, M. *et al.* Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852-857, doi:10.1038/nature07115 (2008).
111 Rodnina, M. V. Translation in Prokaryotes. *Cold Spring Harb Perspect Biol* **10**, doi:10.1101/cshperspect.a032664 (2018).
112 Weixlbaumer, A. *et al.* Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953-956, doi:10.1126/science.1164840 (2008).
113 Kiel, M. C., Kaji, H. & Kaji, A. Ribosome recycling: An essential process of protein synthesis. *Biochem Mol Biol Educ* **35**, 40-44, doi:10.1002/bmb.6 (2007).
114 Pavlov, M. Y., Freistroffer, D. V., MacDougall, J., Buckingham, R. H. & Ehrenberg, M. Fast recycling of Escherichia coli ribosomes requires both ribosome recycling factor (RRF) and release factor RF3. *Embo J* **16**, 4134-4141, doi:10.1093/emboj/16.13.4134 (1997).
115 Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**, 1065-1077, doi:10.1016/j.cell.2006.08.032 (2006).
116 Lancaster, L. & Noller, H. F. Involvement of 16S rRNA nucleotides G1338 and A1339 in discrimination of initiator tRNA. *Mol Cell* **20**, 623-632, doi:10.1016/j.molcel.2005.10.006 (2005).
117 Qin, D., Abdi, N. M. & Fredrick, K. Characterization of 16S rRNA mutations that decrease the fidelity of translation initiation. *RNA* **13**, 2348-2355, doi:10.1261/rna.715307 (2007).
118 Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-1942, doi:10.1126/science.1131127 (2006).
119 Abdi, N. M. & Fredrick, K. Contribution of 16S rRNA nucleotides forming the 30S subunit A and P sites to translation in Escherichia coli. *RNA* **11**, 1624-1632, doi:10.1261/rna.2118105 (2005).

120 Ali, I. K., Lancaster, L., Feinberg, J., Joseph, S. & Noller, H. F. Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Mol Cell* **23**, 865-874, doi:10.1016/j.molcel.2006.08.011 (2006).
121 Kipper, K., Hetenyi, C., Sild, S., Remme, J. & Liiv, A. Ribosomal intersubunit bridge B2a is involved in factor-dependent translation initiation and translational processivity. *J Mol Biol* **385**, 405-422, doi:10.1016/j.jmb.2008.10.065 (2009).
122 Cochella, L., Brunelle, J. L. & Green, R. Mutational analysis reveals two independent molecular requirements during transfer RNA selection on the ribosome. *Nat Struct Mol Biol* **14**, 30-36, doi:10.1038/nsmb1183 (2007).
123 Garcia-Ortega, L., Stephen, J. & Joseph, S. Precise alignment of peptidyl tRNA by the decoding center is essential for EF-G-dependent translocation. *Mol Cell* **32**, 292-299, doi:10.1016/j.molcel.2008.09.012 (2008).
124 Phelps, S. S., Jerinic, O. & Joseph, S. Universally conserved interactions between the ribosome and the anticodon stem-loop of A site tRNA important for translocation. *Mol Cell* **10**, 799-807, doi:10.1016/s1097-2765(02)00686-x (2002).
125 Cukras, A. R., Southworth, D. R., Brunelle, J. L., Culver, G. M. & Green, R. Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA:tRNA complex. *Mol Cell* **12**, 321-328, doi:10.1016/s1097-2765(03)00275-2 (2003).
126 Gao, H. *et al.* Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell* **113**, 789-801, doi:10.1016/s0092-8674(03)00427-6 (2003).
127 Ogle, J. M. & Ramakrishnan, V. Structural insights into translational fidelity. *Annu Rev Biochem* **74**, 129-177, doi:10.1146/annurev.biochem.74.061903.155440 (2005).
128 Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C. & Ramakrishnan, V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* **16**, 528-533, doi:10.1038/nsmb.1577 (2009).
129 Hampl, H., Schulze, H. & Nierhaus, K. H. Ribosomal components from Escherichia coli 50 S subunits involved in the reconstitution of peptidyltransferase activity. *J Biol Chem* **256**, 2284-2288 (1981).
130 Moore, V. G., Atchison, R. E., Thomas, G., Moran, M. & Noller, H. F. Identification of a ribosomal protein essential for peptidyl transferase activity. *Proc Natl Acad Sci U S A* **72**, 844-848, doi:10.1073/pnas.72.3.844 (1975).
131 Harms, J. *et al.* High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679-688, doi:10.1016/s0092-8674(01)00546-3 (2001).
132 Dorner, S., Brunelle, J. L., Sharma, D. & Green, R. The hybrid state of tRNA binding is an authentic translation elongation intermediate. *Nat Struct Mol Biol* **13**, 234-241, doi:10.1038/nsmb1060 (2006).
133 Feinberg, J. S. & Joseph, S. A conserved base-pair between tRNA and 23 S rRNA in the peptidyl transferase center is important for peptide release. *J Mol Biol* **364**, 1010-1020, doi:10.1016/j.jmb.2006.09.040 (2006).
134 Polacek, N. *et al.* The critical role of the universally conserved A2602 of 23S ribosomal RNA in the release of the nascent peptide during translation termination. *Mol Cell* **11**, 103-112, doi:10.1016/s1097-2765(02)00825-0 (2003).
135 Youngman, E. M., Brunelle, J. L., Kochaniak, A. B. & Green, R. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell* **117**, 589-599, doi:10.1016/s0092-8674(04)00411-8 (2004).
136 Maguire, B. A., Beniaminov, A. D., Ramu, H., Mankin, A. S. & Zimmermann, R. A. A protein component at the heart of an RNA machine: the importance of protein l27 for the function of the bacterial ribosome. *Mol Cell* **20**, 427-435, doi:10.1016/j.molcel.2005.09.009 (2005).
137 Schmeing, T. M., Huang, K. S., Strobel, S. A. & Steitz, T. A. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520-524, doi:10.1038/nature04152 (2005).
138 Jenner, L., Rees, B., Yusupov, M. & Yusupova, G. Messenger RNA conformations in the ribosomal E site revealed by X-ray crystallography. *EMBO Rep* **8**, 846-850, doi:10.1038/sj.embor.7401044 (2007).

139 Devaraj, A., Shoji, S., Holbrook, E. D. & Fredrick, K. A role for the 30S subunit E site in maintenance of the translational reading frame. *RNA* **15**, 255-265, doi:10.1261/rna.1320109 (2009).
140 Fei, J., Kosuri, P., MacDougall, D. D. & Gonzalez, R. L., Jr. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol Cell* **30**, 348-359, doi:10.1016/j.molcel.2008.03.012 (2008).
141 Lill, R., Robertson, J. M. & Wintermeyer, W. Binding of the 3' terminus of tRNA to 23S rRNA in the ribosomal exit site actively promotes translocation. *Embo J* **8**, 3933-3938 (1989).
142 Sergiev, P. V. *et al.* Function of the ribosomal E-site: a mutagenesis study. *Nucleic Acids Res* **33**, 6048-6056, doi:10.1093/nar/gki910 (2005).
143 Walker, S. E., Shoji, S., Pan, D., Cooperman, B. S. & Fredrick, K. Role of hybrid tRNA-binding states in ribosomal translocation. *Proc Natl Acad Sci U S A* **105**, 9192-9197, doi:10.1073/pnas.0710146105 (2008).
144 Berk, V., Zhang, W., Pai, R. D. & Cate, J. H. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci U S A* **103**, 15830-15834, doi:10.1073/pnas.0607541103 (2006).
145 Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897-902, doi:10.1126/science.1060612 (2001).
146 Cheatham III, T. E., Young, M.A. . Molecular dynamics simulation of nucleic acids: Successes, limitations, and promise*. *Biopolymers* **56**, 24, doi:https://doi.org/10.1002/1097-0282(2000)56:4<232::AID-BIP10037>3.0.CO;2-H (2000).
147 Hospital, A., Goñi, J.R., Orozco, M., Gelpi, J. Molecular dynamics simulations: advances and applications. . *Adv Appl Bioinform Chem.* **8**, 37-47, doi:https://doi.org/10.2147/AABC.S70333 (2015).
148 McDowell, S. E., Špačková, N. a., Šponer, J. & Walter, N. G. Molecular dynamics simulations of RNA: an in silico single molecule approach. *Biopolymers: Original Research on Biomolecules* **85**, 169-184 (2007).
149 Pérez, A., Luque, F. J. & Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Accounts of chemical research* **45**, 196-205 (2012).
150 Sanbonmatsu, K. Y. Computational studies of molecular machines: the ribosome. *Curr Opin Struct Biol* **22**, 168-174, doi:10.1016/j.sbi.2012.01.008 (2012).
151 Šponer, J. *et al.* Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *The Journal of Physical Chemistry Letters* **5**, 1771-1782, doi:10.1021/jz500557y (2014).
152 Šponer, J. *et al.* RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chemical Reviews* **118**, 4177-4338, doi:https://10.1021/acs.chemrev.7b00427 (2018).
153 McCammon, J. A. & Harvey, S. C. *Dynamics of proteins and nucleic acids*. (Cambridge University Press, 1988).
154 Ding, F. *et al.* Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* **14**, 1164-1173, doi:10.1261/rna.894608 (2008).
155 Li, R., Macnamara, L. M., Leuchter, J. D., Alexander, R. W. & Cho, S. S. MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. *International Journal of Molecular Sciences* **16**, 15872-15902 (2015).
156 Sorin, E. J. *et al.* Does native state topology determine the RNA folding mechanism? *J Mol Biol* **337**, 789-797, doi:10.1016/j.jmb.2004.02.024 (2004).
157 Budiman, M. E., Knaggs, M. H., Fetrow, J. S. & Alexander, R. W. Using molecular dynamics to map interaction networks in an aminoacyl-tRNA synthetase. *Proteins* **68**, 670-689, doi:10.1002/prot.21426 (2007).
158 Yamasaki, S., Nakamura, S., Terada, T. & Shimizu, K. Mechanism of the difference in the binding affinity of E. coli tRNAGln to glutaminyl-tRNA synthetase caused by noninterface nucleotides in variable loop. *Biophys J* **92**, 192-200, doi:10.1529/biophysj.106.093351 (2007).
159 Auffinger, P. & Westhof, E. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J Mol Biol* **269**, 326-341, doi:10.1006/jmbi.1997.1022 (1997).

160 McCrate, N. E., Varner, M. E., Kim, K. I. & Nagan, M. C. Molecular dynamics simulations of human tRNA Lys,3 UUU: the role of modified bases in mRNA recognition. *Nucleic Acids Res* **34**, 5361-5368, doi:10.1093/nar/gkl580 (2006).
161 Prabhakar, P. S., Takyi, N. A. & Wetmore, S. D. Posttranscriptional modifications at the 37th position in the anticodon stem-loop of tRNA: structural insights from MD simulations. *RNA* **27**, 202-220, doi:10.1261/rna.078097.120 (2021).
162 Zhang, X., Walker, R. C., Phizicky, E. M. & Mathews, D. H. Influence of Sequence and Covalent Modifications on Yeast tRNA Dynamics. *Journal of Chemical Theory and Computation* **10**, 3473-3483, doi:10.1021/ct500107y (2014).
163 Caulfield, T. R., Devkota, B. & Rollins, G. C. Examinations of tRNA Range of Motion Using Simulations of Cryo-EM Microscopy and X-Ray Data. *Journal of Biophysics* **2011**, 219515, doi:10.1155/2011/219515 (2011).
164 Grzybkowska, A., Jędrzejczyk, D., Rostkowski, M., Chworos, A. & Dybala-Defratyka, A. RNA model evaluation based on MD simulation of four tRNA analogs. *RSC Advances* **6**, 101778-101789, doi:10.1039/C6RA14933B (2016).
165 Björk, G. R. in *tRNA* (ed Uttam L. RajBhandary Dieter Söll) 165-205 (1994).
166 Yokoyama, S. & Nishimura, S. in *tRNA* 207-223 (1994).
167 Nasvall, S. J., Chen, P. & Bjork, G. R. The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA* **13**, 2151-2164, doi:10.1261/rna.731007 (2007).
168 Nasvall, S. J., Chen, P. & Bjork, G. R. The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo. *RNA* **10**, 1662-1673, doi:10.1261/rna.7106404 (2004).
169 Yokoyama, S. *et al.* Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc Natl Acad Sci U S A* **82**, 4905-4909, doi:10.1073/pnas.82.15.4905 (1985).
170 Suzuki, T., Suzuki, T., Wada, T., Saigo, K. & Watanabe, K. Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases. *Embo J* **21**, 6581-6589, doi:10.1093/emboj/cdf656 (2002).
171 Armengod, M. E. *et al.* Enzymology of tRNA modification in the bacterial MnmEG pathway. *Biochimie* **94**, 1510-1520, doi:10.1016/j.biochi.2012.02.019 (2012).
172 Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691, doi:10.1038/nature02026 (2003).
173 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-179, doi:10.1038/nature14447 (2015).
174 Fislage, M., Wauters, L. & Versees, W. MnmE, a GTPase That Drives a Complex tRNA Modification Reaction. *Biopolymers* **105**, 568-579, doi:10.1002/bip.22813 (2016).
175 Li, X. & Guan, M. X. A human mitochondrial GTP binding protein related to tRNA modification may modulate phenotypic expression of the deafness-associated mitochondrial 12S rRNA mutation. *Mol Cell Biol* **22**, 7701-7711, doi:10.1128/mcb.22.21.7701-7711.2002 (2002).
176 Villarroya, M. *et al.* Characterization of human GTPBP3, a GTP-binding protein involved in mitochondrial tRNA modification. *Molecular and cellular biology* **28**, 7514-7531, doi:10.1128/mcb.00946-08 (2008).
177 Cabedo, H. *et al.* The Escherichia coli trmE (mnmE) gene, involved in tRNA modification, codes for an evolutionarily conserved GTPase with unusual biochemical properties. *Embo J* **18**, 7063-7076, doi:10.1093/emboj/18.24.7063 (1999).
178 Meyer, S., Scrima, A., Versees, W. & Wittinghofer, A. Crystal structures of the conserved tRNA-modifying enzyme GidA: implications for its interaction with MnmE and substrate. *J Mol Biol* **380**, 532-547, doi:10.1016/j.jmb.2008.04.072 (2008).
179 Scrima, A., Vetter, I. R., Armengod, M. E. & Wittinghofer, A. The structure of the TrmE GTP-binding protein and its implications for tRNA modification. *Embo J* **24**, 23-33, doi:10.1038/sj.emboj.7600507 (2005).
180 Moukadiri, I. *et al.* Evolutionarily conserved proteins MnmE and GidA catalyze the formation of two methyluridine derivatives at tRNA wobble positions. *Nucleic Acids Res* **37**, 7177-7193, doi:10.1093/nar/gkp762 (2009).

181 Yim, L., Moukadiri, I., Bjork, G. R. & Armengod, M. E. Further insights into the tRNA modification process controlled by proteins MnmE and GidA of Escherichia coli. *Nucleic Acids Res* **34**, 5892-5905, doi:10.1093/nar/gkl752 (2006).
182 Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299-1304, doi:10.1126/science.1062023 (2001).
183 Martinez-Vicente, M. *et al.* Effects of mutagenesis in the switch I region and conserved arginines of Escherichia coli MnmE protein, a GTPase involved in tRNA modification. *J Biol Chem* **280**, 30660-30670, doi:10.1074/jbc.M503223200 (2005).
184 Prado, S., Villarroya, M., Medina, M. & Armengod, M. E. The tRNA-modifying function of MnmE is controlled by post-hydrolysis steps of its GTPase cycle. *Nucleic Acids Res* **41**, 6190-6208, doi:10.1093/nar/gkt320 (2013).
185 Wittinghofer, A. & Vetter, I. R. Structure-function relationships of the G domain, a canonical switch motif. *Annu Rev Biochem* **80**, 943-971, doi:10.1146/annurev-biochem-062708-134043 (2011).
186 Meyer, S., Wittinghofer, A. & Versees, W. G-domain dimerization orchestrates the tRNA wobble modification reaction in the MnmE/GidA complex. *J Mol Biol* **392**, 910-922, doi:10.1016/j.jmb.2009.07.004 (2009).
187 Scrima, A. & Wittinghofer, A. Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. *Embo J* **25**, 2940-2951, doi:10.1038/sj.emboj.7601171 (2006).
188 Yamanaka, K., Hwang, J. & Inouye, M. Characterization of GTPase Activity of TrmE, a Member of a Novel GTPase Superfamily, from Thermotoga maritima. *Journal of Bacteriology* **182**, 7078-7082, doi:doi:10.1128/JB.182.24.7078-7082.2000 (2000).
189 Scheffzek, K. *et al.* Structural analysis of the GAP-related domain from neurofibromin and its implications. *Embo J* **17**, 4313-4327, doi:10.1093/emboj/17.15.4313 (1998).
190 Pai, E. F. *et al.* Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *Embo J* **9**, 2351-2359 (1990).
191 Anand, B., Surana, P. & Prakash, B. Deciphering the catalytic machinery in 30S ribosome assembly GTPase YqeH. *PLoS One* **5**, e9944 (2010).
192 Ash, M.-R., Maher, M. J., Mitchell Guss, J. & Jormakka, M. The cation-dependent G-proteins: In a class of their own. *FEBS Letters* **586**, 2218-2224, doi:https://doi.org/10.1016/j.febslet.2012.06.030 (2012).
193 Robinson, V. L., Hwang, J., Fox, E., Inouye, M. & Stock, A. M. Domain arrangement of Der, a switch protein containing two GTPase domains. *Structure* **10**, 1649-1658, doi:10.1016/s0969-2126(02)00905-x (2002).
194 Verstraeten, N., Fauvart, M., Versees, W. & Michiels, J. The universally conserved prokaryotic GTPases. *Microbiol Mol Biol Rev* **75**, 507-542, second and third pages of table of contents, doi:10.1128/MMBR.00009-11 (2011).
195 Shi, R. *et al.* Structure-function analysis of Escherichia coli MnmG (GidA), a highly conserved tRNA-modifying enzyme. *J Bacteriol* **191**, 7614-7619, doi:10.1128/JB.00650-09 (2009).
196 Osawa, T. *et al.* Conserved cysteine residues of GidA are essential for biogenesis of 5-carboxymethylaminomethyluridine at tRNA anticodon. *Structure* **17**, 713-724, doi:10.1016/j.str.2009.03.013 (2009).
197 Dym, O. & Eisenberg, D. Sequence-structure analysis of FAD-containing proteins. *Protein Sci* **10**, 1712-1728, doi:10.1110/ps.12801 (2001).
198 Holm, L. & Sander, C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**, 316-319, doi:10.1093/nar/26.1.316 (1998).
199 Lawrence, C. M., Rodwell, V. W. & Stauffacher, C. V. Crystal structure of Pseudomonas mevalonii HMG-CoA reductase at 3.0 angstrom resolution. *Science* **268**, 1758-1762, doi:10.1126/science.7792601 (1995).
200 Tabernero, L., Bochar, D. A., Rodwell, V. W. & Stauffacher, C. V. Substrate-induced closure of the flap domain in the ternary complex structures provides insights into the mechanism of catalysis by 3-hydroxy-3-methylglutaryl-CoA reductase. *Proc Natl Acad Sci U S A* **96**, 7167-7171, doi:10.1073/pnas.96.13.7167 (1999).

201 Armengod, M. E. *et al.* Modification of the wobble uridine in bacterial and mitochondrial tRNAs reading NNA/NNG triplets of 2-codon boxes. *RNA Biol* **11**, 1495-1507, doi:10.4161/15476286.2014.992269 (2014).
202 Moukadiri, I., Garzon, M. J., Bjork, G. R. & Armengod, M. E. The output of the tRNA modification pathways controlled by the Escherichia coli MnmEG and MnmC enzymes depends on the growth conditions and the tRNA species. *Nucleic Acids Res* **42**, 2602-2623, doi:10.1093/nar/gkt1228 (2014).
203 Carreras, C. W. & Santi, D. V. The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* **64**, 721-762, doi:10.1146/annurev.bi.64.070195.003445 (1995).
204 Hamdane, D., Grosjean, H. & Fontecave, M. Flavin-Dependent Methylation of RNAs: Complex Chemistry for a Simple Modification. *J Mol Biol* **428**, 4867-4881, doi:10.1016/j.jmb.2016.10.031 (2016).
205 Koehn, E. M. *et al.* An unusual mechanism of thymidylate biosynthesis in organisms containing the thyX gene. *Nature* **458**, 919-923, doi:10.1038/nature07973 (2009).
206 Meyer, S. *et al.* Kissing G domains of MnmE monitored by X-ray crystallography and pulse electron paramagnetic resonance spectroscopy. *PLoS Biol* **7**, e1000212, doi:10.1371/journal.pbio.1000212 (2009).
207 Fislage, M. *et al.* SAXS analysis of the tRNA-modifying enzyme complex MnmE/MnmG reveals a novel interaction mode and GTP-induced oligomerization. *Nucleic Acids Res* **42**, 5978-5992, doi:10.1093/nar/gku213 (2014).
208 Yim, L. *et al.* The GTPase activity and C-terminal cysteine of the Escherichia coli MnmE protein are essential for its tRNA modifying function. *J Biol Chem* **278**, 28378-28387, doi:10.1074/jbc.M301381200 (2003).
209 Smith, H. C., Wedekind, J. E., Xie, K., Sowden, M. P. . *Fine-Tuning of RNA Functions by Modification and Editing*. Vol. 12 (2005).
210 Yoshida, M. *et al.* Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proc Natl Acad Sci U S A* **112**, 2764-2769, doi:10.1073/pnas.1415525112 (2015).
211 Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nature Reviews Molecular Cell Biology* **22**, 375-392, doi:10.1038/s41580-021-00342-0 (2021).
212 Suzuki, T., Nagao, A. & Suzuki, T. Human mitochondrial diseases caused by lack of taurine modification in mitochondrial tRNAs. *Wiley Interdiscip Rev RNA* **2**, 376-386, doi:10.1002/wrna.65 (2011).
213 Yasukawa, T. *et al.* Defect in modification at the anticodon wobble nucleotide of mitochondrial tRNALys with the MERRF encephalomyopathy pathogenic mutation. *FEBS Letters* **467**, 175-178, doi:https://doi.org/10.1016/S0014-5793(00)01145-5 (2000).
214 Yasukawa, T. *et al.* Modification Defect at Anticodon Wobble Nucleotide of Mitochondrial tRNAsLeu(UUR) with Pathogenic Mutations of Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis, and Stroke-like Episodes*. *Journal of Biological Chemistry* **275**, 4251-4257, doi:https://doi.org/10.1074/jbc.275.6.4251 (2000).
215 Nakamura, M. *et al.* MERRF/MELAS overlap syndrome: a double pathogenic mutation in mitochondrial tRNA genes. *Journal of Medical Genetics* **47**, 659-664, doi:10.1136/jmg.2009.072058 (2010).
216 Baruffini, E. *et al.* MTO1 mutations are associated with hypertrophic cardiomyopathy and lactic acidosis and cause respiratory chain deficiency in humans and yeast. *Hum Mutat* **34**, 1501-1509, doi:10.1002/humu.22393 (2013).
217 Fakruddin, M. *et al.* Defective Mitochondrial tRNA Taurine Modification Activates Global Proteostress and Leads to Mitochondrial Disease. *Cell Rep* **22**, 482-496, doi:10.1016/j.celrep.2017.12.051 (2018).
218 Kopajtich, R. *et al.* Mutations in GTPBP3 cause a mitochondrial translation defect associated with hypertrophic cardiomyopathy, lactic acidosis, and encephalopathy. *Am J Hum Genet* **95**, 708-720, doi:10.1016/j.ajhg.2014.10.017 (2014).
219 Wu, Y. *et al.* Mtu1-Mediated Thiouridine Formation of Mitochondrial tRNAs Is Required for Mitochondrial Translation and Is Involved in Reversible Infantile Liver Injury. *PLoS Genet* **12**, e1006355, doi:10.1371/journal.pgen.1006355 (2016).
220 Zeharia, A. *et al.* Acute infantile liver failure due to mutations in the TRMU gene. *Am J Hum Genet* **85**, 401-407, doi:10.1016/j.ajhg.2009.08.004 (2009).

221 Sturman, J. A. Taurine in development. *Physiol Rev* **73**, 119-147, doi:10.1152/physrev.1993.73.1.119 (1993).
222 Rikimaru, M. *et al.* Taurine ameliorates impaired the mitochondrial function and prevents stroke-like episodes in patients with MELAS. *Intern Med* **51**, 3351-3357, doi:10.2169/internalmedicine.51.7529 (2012).
223 Ohsawa, Y. *et al.* Taurine supplementation for prevention of stroke-like episodes in MELAS: a multicentre, open-label, 52-week phase III trial. *J Neurol Neurosurg Psychiatry* **90**, 529-536, doi:10.1136/jnnp-2018-317964 (2019).
224 Kirino, Y. *et al.* Acquisition of the wobble modification in mitochondrial tRNALeu(CUN) bearing the G12300A mutation suppresses the MELAS molecular defect. *Human Molecular Genetics* **15**, 897-904, doi:10.1093/hmg/ddl007 (2006).
225 Joshi, K., Bhatt, M. J. & Farabaugh, P. J. Codon-specific effects of tRNA anticodon loop modifications on translational misreading errors in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res* **46**, 10331-10339, doi:10.1093/nar/gky664 (2018).
226 Manickam, N., Joshi, K., Bhatt, M. J. & Farabaugh, P. J. Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res* **44**, 1871-1881, doi:10.1093/nar/gkv1506 (2016).
227 Ashraf, S. S. *et al.* The uridine in "U-turn": contributions to tRNA-ribosomal binding. *RNA (New York, N.Y.)* **5**, 503-511, doi:10.1017/s1355838299981931 (1999).

CHAPTER 2: BINDING OF THE MNME-MNMG MODIFICATION COMPLEX TO TRNA

2.1 Objectives

As previously discussed, there is a lack of consensus in the literature on the formation and catalytic mechanism of the MnmE-MnmG (MnmEG) complex.[1-4] Furthermore, the absence of a full crystal structure for the complex makes it difficult to computationally study the protein-protein and protein-RNA interactions within the complex. Consequently, more information is required to understand how this modification engine comes together to modify tRNA's wobble uridine. To this end, this thesis seeks to characterize the binding properties of MnmE and MnmG to tRNA and to provide insight into the protein-tRNA interactions present within the complex. Moreover, in a bid to isolate a stable functional complex, the effects known cofactors (besides THF to prevent methylation at C5 of U34) have on the binding propensities of the proteins were also studied.

Additionally, little is understood about the specificity of MnmE and MnmG, and the complex is generally assumed to be a tRNA modification entity. Consequently, this thesis will also investigate binding interactions between MnmE and MnmE and other RNA substrates. For this study, two kinds of RNA substrates will be considered – a short, single-stranded RNA with no secondary structure, and a long RNA with extensive secondary structure. The goal is to provide more insight into protein specificity, but more pertinently the recognition and binding mechanisms of MnmE and MnmG.

2.2 Methodology

2.2.1 Protein expression and purification

2.2.1.1 Protein expression

Wild-type MnmE and MnmG

Recombinant hexahistidine-tagged MnmE and MnmG were expressed from the pCA24N(-) plasmid in AG1(ME5305) *E. coli* cells in the presence of 50 mg/L chloramphenicol. The pCA24N(-)-MnmE and the -MnmG plasmids were a kind gift from the ASKA collection.[5] Cultures were inoculated at an optical density at 600 nm (OD_{600}) of 0.1 and grown in 2 L (4 x 500 mL) of LB medium with 50 μg/mL chloramphenicol at 37°C. Protein expression was induced when cells reached the log phase of growth

(OD_{600} = 0.6) using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cell cultures were grown overnight at 18°C and harvested by centrifugation at 5000×*g* for 15 minutes at 4°C using a JA-14 rotor (Beckman). Cells were flash frozen in liquid nitrogen and stored at -80°C for future use. Protein expression was monitored by removing 1 OD_{600} samples at regular time intervals and resuspending the cell pellets in 0.1 M Tris-HCl pH 8.5 containing 5 M urea. These samples were analyzed by 10% SDS-PAGE. Gels were stained using Coomassie blue, destained and scanned.

Fluorescently tagged protein variants

Green fluorescent protein (GFP) and hexahistidine-tagged variants of MnmE and MnmG were expressed as outlined above using the pCA24N(+) plasmid in AG1(ME5305) *E. coli* cells.[5] Recombinant mCherry and hexahistidine-tagged variants of the proteins were expressed from the pCA24N-MnmE-mCherry and PCA24N-MnmG-mCherry plasmids previously prepared in the Kothe group. In brief, the mCherry gene sequence was inserted using the NcoI and SalI restriction sites of the pCA24N(-) plasmid.

2.2.1.2 Protein purification

Cell opening

All proteins were purified using the same procedure. Frozen cell pellets were resuspended in Buffer A (20 mM Tris-HCl pH 8.1, 400 mM NaCl, 5% (v/v) glycerol, 1 mM β-mercaptoethanol, 0.5 mM phenylmethanesulfonylfluoride (PMSF), 30 mM imidazole) using approximately 5 mL of buffer per gram of cells, and cells were thawed on ice while stirring. Lysozyme (1 mg/mL) was added to the homogenous cell suspension, and the latter was incubated for 30 minutes. Next, sodium deoxycholate (12.5 mg/g of cells) was added to the cell suspension, which was incubated for 30 minutes. Cells were opened on ice via sonication using a ½ inch probe, 1-minute intervals, intensity level 6, and duty cycle 60% for 15 minutes. The cell lysate was centrifuged at 30,000×*g* for 30 minutes at 4°C using a JA-25.5 rotor (Beckman). A 3 μL sample of the lysate was stored at 4°C for future analysis.

Nickel sepharose chromatography

The cleared lysate was added to 2.5 mL of Ni^{2+}-sepharose slurry and incubated with gentle shaking at room temperature for an hour. The slurry was centrifuged at 500×*g* for 5 minutes at 4°C, the supernatant decanted and stored at 4°C for future analysis. Next, the resin was washed and centrifuged (500×*g*, 5 minutes) 6 times with 4 mL Buffer A wash, and the supernatants were pooled and stored at 4°C. Then, the protein was eluted 8 times using 2.25 mL of Buffer B (20 mM Tris-HCl pH 8.1, 400 mM NaCl, 5%(v/v) glycerol, 1 mM β-mercaptoethanol, 500 mM imidazole) per elution (90% resin volume). Each elution step included a 5-minute incubation period before the slurry was centrifuged (500×*g*, 5 minutes) and the elution decanted and stored at 4°C. All washes and elutions were conducted at 4°C. Protein purification was monitored by collecting 50 μL samples after every centrifugation step. These samples were analyzed by 10% SDS-PAGE. Elutions were pooled and concentrated using ultrafiltration (Vivaspin MWCO 10000) and stored at 4°C. Proteins were further purified using size exclusion chromatography the next day.

Size exclusion chromatography

The concentrated protein elutions (approximately 5 mL) were injected onto a Superdex 200 column (10/300 GL column, GE Healthcare) in 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$ (flow rate = 0.5 mL/min, BioLogic DuoFlow chromatography system), and the absorbance was monitored at 280 nm. Peak fractions were analyzed by 10% SDS-PAGE, pooled, and concentrated as before. Then, proteins were aliquoted, flash frozen in liquid nitrogen and stored at -80°C for future use.

Protein quantification

The theoretical extinction coefficients of all proteins calculated using ProtParam (Table 2.1) and their concentrations were determined photometrically at 280 nm. The relative purity for all proteins was determined using a 10% SDS-PAGE, which was analyzed in ImageJ.

Table 2.1 – Theoretical extinction coefficients for MnmE and MnmG proteins

Protein	ε_{280} ($M^{-1} \cdot cm^{-1}$)
Wild-type MnmE	26470
Wild-type MnmG	47330
MnmE-GFP	48610
MnmG-GFP	69595
MnmE-mCherry	60975
MnmG-mCherry	81960

2.2.2 tRNA synthesis, purification and labelling

PCR amplification of tRNA gene

The following gene sequence for the T7 promoter and *E. coli* $tRNA^{Glu}$ was synthesized and inserted into the backbone of the pIDTSMART-KAN plasmid (IDT):

5' – GCTAATACGACTCACTATAGGTCCCCTTCGTCTAGAGGCCCAGGACACCG

CCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGCCCGCCA – 3'

Then, the pIDTSMART-KAN-tRNAGlu_UUC_T7 plasmid was used to generate the template DNA of *E. coli* $tRNA^{Glu}$ through the PCR amplification using the following primers:

T7 promoter sense 5' – TAATACGACTCACTATAGGTCCCCTTCGTCTAGAGGCCCAGGACACCG

CCCTTTCACG – 3'

$tRNA^{Glu}$ antisense 5' – TGGCGGGCCCTAGGGGATTCG – 3'

All PCR reactions were carried out using 1X Pfu buffer with $MgSO_4$, 200 μM of each dNTP, 0.5 μM T7 promoter sense primer, 0.5 μM $tRNA^{Glu}$ antisense reverse primer, 0.02 U/μL Pfu DNA polymerase and 30 ng/μL of miniprepped plasmid DNA. PCR conditions are outlined in Table 2.2. Template amplification was confirmed using a 12% DNA-PAGE.

Table 2.2 – PCR amplification of $tRNA^{Glu}$ gene from plDTSMART-KAN-tRNAGlu_UUC_T7 DNA template

Step	Temperature (°C)	Time	Cycles
Pre-denaturation	95	5 minutes	1
Initial denaturation	95	5 minutes	1
Denaturation	95	30 seconds	35
Annealing	41	30 seconds	
Extension	72	30 seconds	

In vitro transcription and purification of tRNA

In vitro transcriptions were carried out by incubating 10% (v/v) PCR-generated DNA template in transcription buffer (40 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl), 10 mM DTT, 3 mM NTPs (ATP, CTP, GTP, UTP), 5 mM GMP, 0.01 U/μL iPPase, 0.3 mM T7 RNA Polymerase, 0.12 U/μL RNase inhibitor for 4 hours at 37°C. Tritium labelled $tRNA^{Glu}$ was generated by substituting 3 mM UTP with 0.3 mM [5-^{3}H] UTP. Following incubation, template DNA was digested with 1 U/μL DNase over the course of an hour. The non-radioactive tRNA was precipitated in isopropanol and isolated using phenol-chloroform extraction. Then, its concentration was determined photometrically at 260 nm ($\varepsilon = 7 \times 10^5\ M^{-1} \bullet cm^{-1}$; provided by IDT), and the tRNA was stored at -20°C for future use. The radiolabeled tRNA was purified using the Nucleobond Xtra Midi column (Macherey-Nagel). Prior to purification, the column was equilibrated with Buffer R0' (100 mM Tris/acetate, 10 mM $MgCl_2$, 15% ethanol, pH 6.3), and the tRNA was diluted using Buffers R0' and R3' (100 mM Tris/acetate, 10 mM $MgCl_2$, 15% ethanol, 1150 mM NaCl, pH 6.3) to a final salt concentration of 0.2 M NaCl. The *in vitro* transcription mix was loaded onto the column and washed with Buffer R1'a (100 mM Tris/acetate, 10 mM $MgCl_2$, 15% ethanol, 300 mM NaCl, pH 6.3). Then, the tRNA was eluted using Buffer R3' and precipitated using 3 column volumes of cold 100% ethanol overnight at -20°C. Next, the tRNA was centrifuged at 4500xg for 45 minutes at 4°C. The pellet was washed twice using 2 column volumes of 70% (v/v) ethanol. After every wash, the tRNA

was centrifuged at 4500xg for 30 minutes at 4°C and the supernatant decanted. The pellet was air dried, dissolved in MilliQ and stored at -20°C for future use. The tRNA concentration was determined as before.

Preparation of other RNA substrates

To probe the specificity of the MnmEG complex, two other RNA substrates were investigated – an snR34 substate and snR81. The snR34 substrate is a 27-nt single-stranded RNA, with no secondary structure. snR81 is 188-nt is a snoRNA with a bipartite architecture governed by two hairpins, and each hairpin has an internal loop. These RNAs were prepared and radiolabeled by Dr. Erin Kelly, a previous Kothe lab member.[6]

2.2.3 Nitrocellulose filter binding

The RNA substrate was incubated in buffer at 65°C for 3 minutes, then cooled at room temperature for 15 minutes to allow for folding. A low constant concentration of radiolabeled RNA (30 nM) was incubated with increasing concentrations of protein for 10 minutes in the appropriate buffer at room temperature to allow for binding. To measure RNA binding in the absence of cofactors, Buffer 1 was used (50 mM Tris-HCl pH 7.5, 70 mM NH_4Cl, 30 mM KCl, 1 mM EDTA, 4 mM $MgCl_2$). When observing RNA binding in the presence of cofactors, RNA and protein were incubated in Buffer 2 (50 mM Tris-HCl pH 7.5, 70 mM NH_4Cl, 1 mM FAD, 300 mM GTP, 100 mM KCl, 5 mM $MgCl_2$, 100 mM Glycine). The reaction mixture was filtered under vacuum through a nitrocellulose membrane (0.2 μm, Whatman). The membrane was then washed immediately with 1 mL ice cold Buffer 1 and dissolved in 10 mL scintillation cocktail for 30 minutes. The level of tRNA binding was determined through scintillation counting. The dissociation constant (K_D) for tRNA binding was calculated by plotting the fraction of bound RNA against protein concentration and fitting a hyperbolic function to the data (using Prism):

$$\text{Bound} = \text{Bound}_{max} \times [\text{protein}] / (K_D + [\text{protein}])$$

2.2.4 Structural analysis of MnmE and MnmG

The electrostatic potential of high-resolution structures of monomeric MnmE (PDB: 3GEE) and MnmG (PDB: 3CES) were analyzed using PyMOL.[7] First, dimeric models of each protein were generated via symmetry modelling. Then, the electrostatic potential of the dimer was calculated using the APBS Electrostatics module, and the resulting profile was visualized onto the molecule's solvent excluded surface.[8] Finally, the results from this analysis were used to generate hypothetical structures for various assemblies of the MnmE-MnmG complex.

2.3 Results

2.3.1 Protein expression and purification

All recombinant hexahistidine-tagged proteins were overexpressed from previously outlined pCA24N plasmids from the ASKA collection in AG1 (ME5305) *E. coli* cells and the proteins were isolated using a double purification strategy. First, affinity chromatography using nickel-sepharose was used to capture the hexahistidine-tagged proteins and eliminate the majority of cellular contaminants (Figure 2.1A). Elutions from affinity chromatography were pooled and loaded onto the Superdex 200 column and the size exclusion chromatography was employed to further purify the proteins (Figure 2.1B). Peak fractions matching the elution profiles of each protein were analyzed by SDS-PAGE and bands corresponding to the proteins of interest were observed (Figure 2.1C; MnmE is approximately 50 kDa in size). MnmE and MnmG were successfully purified with relative purities greater than 90% (Table 2.3). Some fluorescently tagged proteins showed signs of degradation, as multiple bands were prominent on the gels, and these band sizes corresponded with unlabelled or partially labelled MnmE and MnmG (Figure 2.1D). The purity of each protein was determined using the gel analysis module in ImageJ. In summary, all six proteins were successfully expressed and purified to reasonable concentrations and adequate purity for nitrocellulose filter binding assays.

Table 2.3 – Concentrations and purity of MnmE and MnmG variants based on SDS-PAGE analysis

Protein	Concentration, μM	Number of prominent bands	Protein purity, %
Wild-type MnmE	263	1	91
Wild-type MnmG	40	1	94
MnmE-GFP	37	1	95
MnmG-GFP	82	1	90
MnmE-mCherry	157	2	61
MnmG-mCherry	119	2	69

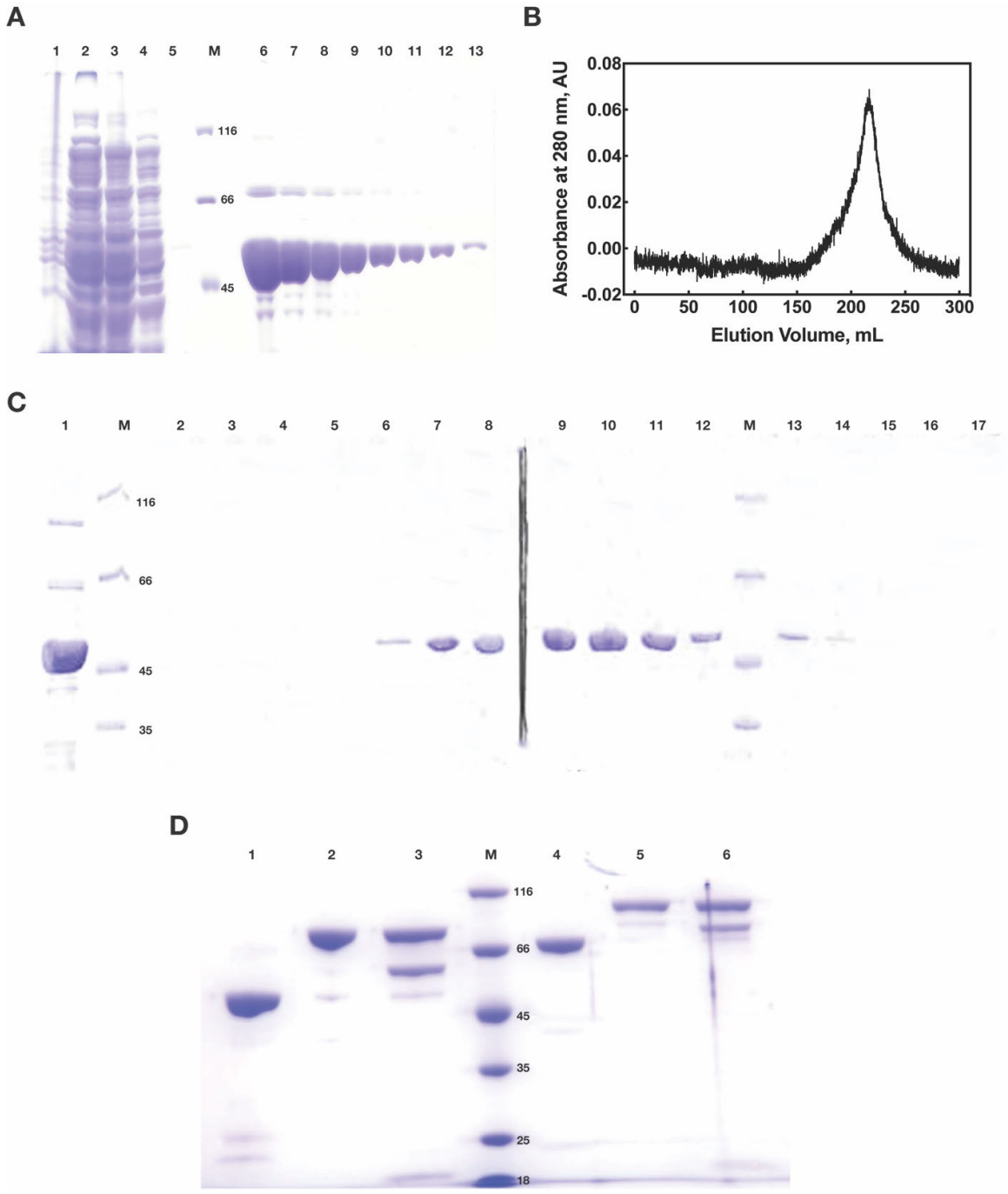


Figure 2.1 – Protein expression and purification summary

(A) Summary of protein expression and nickel affinity purification of wt-MnmE. Lane 1: cell pellet prior to resin binding, lanes 2 & 3: cell lysate prior to and after resin binding, lanes 4 & 5: first and last purification washes, lanes 6-13: elutions. Chromatogram (B) and SDS-PAGE (C) of SEC purified MnmE. Lane1: MnmE purified using nickel sepharose affinity chromatography, lanes 2-17: size exclusion chromatography peak elution samples (145 to 300 mL). (D) SDS-PAGE of purified wt-MnmE (1), MnmE-GFP (2), MnmE-mCherry (3), wt-MnmG (4), MnmG-GFP (5) and MnmG-mCherry (6). In all gels, Lane M represents a 14.4-116 kDa wide range protein molecular weight marker.

2.3.2 tRNA preparation

To probe the binding affinity MnmE and MnmG have for their substrate, $tRNA^{Glu}$ was transcribed, purified, and radiolabeled as it contains the $cmnm^5U34$ modification *in vivo*. The $tRNA^{Glu}$ gene was PCR amplified from the pIDTSMART-KAN-tRNAGlu_UUC_T7 plasmid and analyzed using a DNA-PAGE (Figure 2.2 A). The template is 96 nt long, and one band was observed just below the 100 nt marker at the expected size. Two additional bands were also seen on the gel.

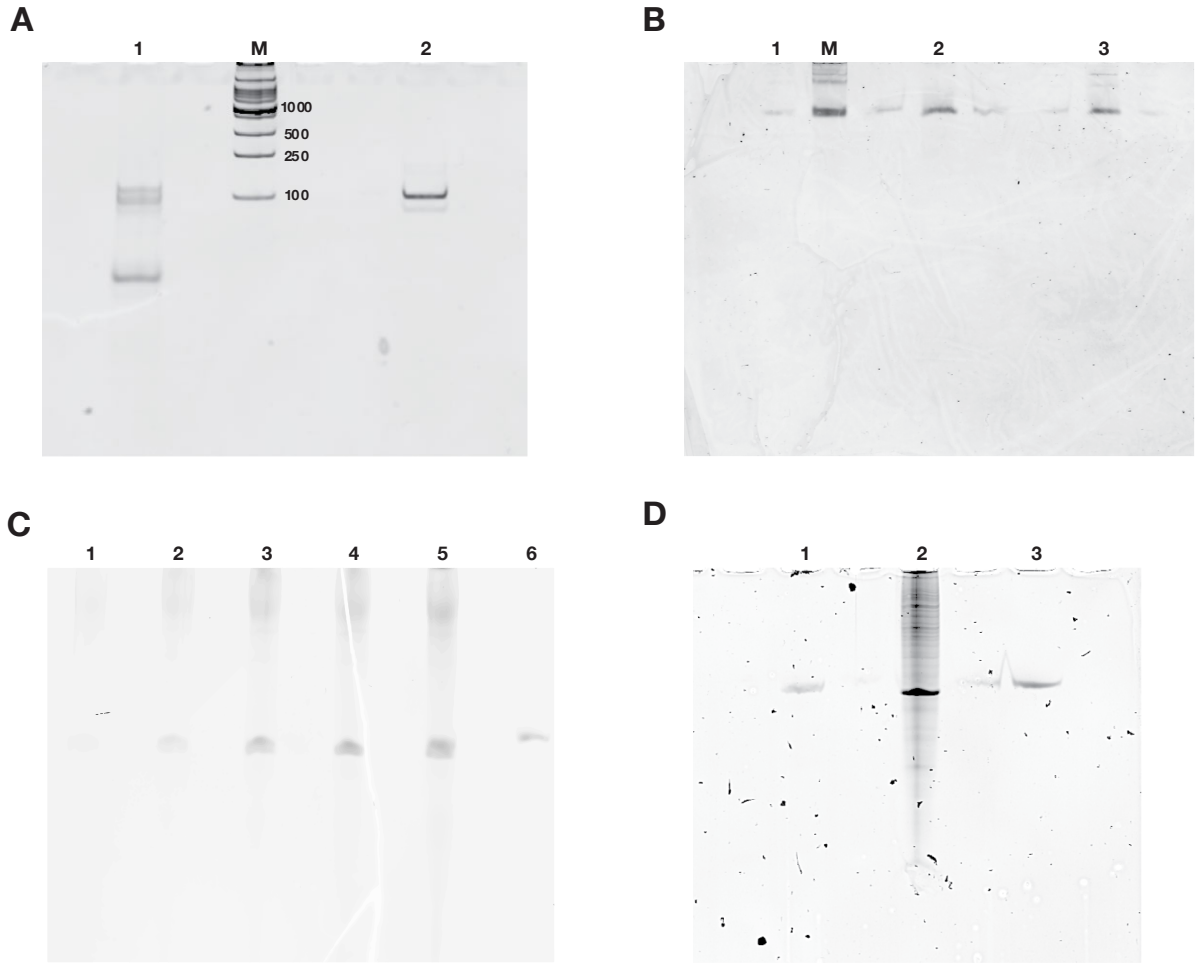


Figure 2.2 – Transcription and purification of $tRNA^{Glu}$

(A) PCR amplification of $tRNA^{Glu}$ gene from the pIDTSMART-KAN DNA template. Lane 1: PCR product, lane 2: positive control ($tRNA^{Phe}$ gene), M: 1kb DNA ladder. (B) Small-scale *in vitro* transcription (IVT) of $tRNA^{Glu}$. Lane 1: $tRNA^{Phe}$ as a positive control, lane 2: $tRNA^{Glu}$ gene PCR product, lane 3: IVT sample after a 1-hour reaction, M: 1kb DNA ladder. (C) Large-scale IVT of $tRNA^{Glu}$. Lanes 1 to 5: 5 mL samples taken at 0, 15, 30, 60 and 180 minutes. Lane 6: $tRNA^{Phe}$ as a positive control. (D) Purification of $tRNA^{Glu}$ extracted from IVT via phenol-chloroform extraction. Lane 1: $tRNA^{Phe}$ as a positive control, Lane 2: IVT $tRNA^{Glu}$, lane 3: purified $tRNA^{Glu}$.

The first band, found a little bit above the 100 nt marker, indicated the formation of primer dimers (the sense primer is 58 nt long), while the second band located towards the bottom of the gel, showed unused primers. Following gene amplification, $tRNA^{Glu}$ was transcribed in a small test *in vitro transcription* (1-hour reaction) to ensure successful tRNA production, which was confirmed using UREA-PAGE analysis (Figure 2.2 B). $tRNA^{Glu}$ migrated in a similar manner as $tRNA^{Phe}$ (positive control) and at the end of the reaction, a single band was observed at the expected size (~ 96 nt). Large-scale reactions were conducted and frequently monitored over the course of the reaction and samples were analyzed using a UREA-PAGE (Figure 2.2 C). As before, $tRNA^{Glu}$ travelled along the gel with the positive control. Furthermore, successful *in vitro* transcription was evident, as the band intensity for $tRNA^{Glu}$ increased with reaction time. Finally, the radiolabeled tRNA was purified using anion exchange chromatography, and this was confirmed using UREA-PAGE analysis. The tRNA purification process was very effective, as all contaminant bands present in the *in vitro* transcription product are absent in the elution (Figure 2.2 D).

2.3.3 Affinity of MnmE and MnmG to tRNA

MnmE and MnmG differentially bind to tRNA

According to the two proposals for the catalytic mechanism of the MnmEG complex, both MnmE and MnmG should interact with the tRNA molecule. MnmG has been posited as the primary binding site for this substrate, as it has a deep positively charged pocket that could accommodate the negative charge from tRNA's phosphate backbone. Nevertheless, theoretically, MnmE should also contribute to substrate binding, as cofactors required for tRNA modification are bound to this protein.

Triplicate nitrocellulose filter binding experiments were used to determine each protein's affinity for their tRNA substrate. The percentage of protein-bound RNA was averaged over the three replicates and plotted against protein concentration (Figure 2.3 A and B). A hyperbolic function was subsequently fitted to this data to determine the K_D values summarized in Table 2.4. Although proteins had similar binding magnitudes (binding amplitude of 42.2 ± 7.2% and 38.5 ± 0.4% for MnmE and MnmG, respectively), their affinities to tRNA were different. Whereas MnmG consistently bound tRNA with a K_D of 0.30 ± 0.06 μM,

MnmE formed weaker interactions the substrate and the strength of these interactions varied between replicates, yielding a K_D of 1.39 ± 0.38 μM.

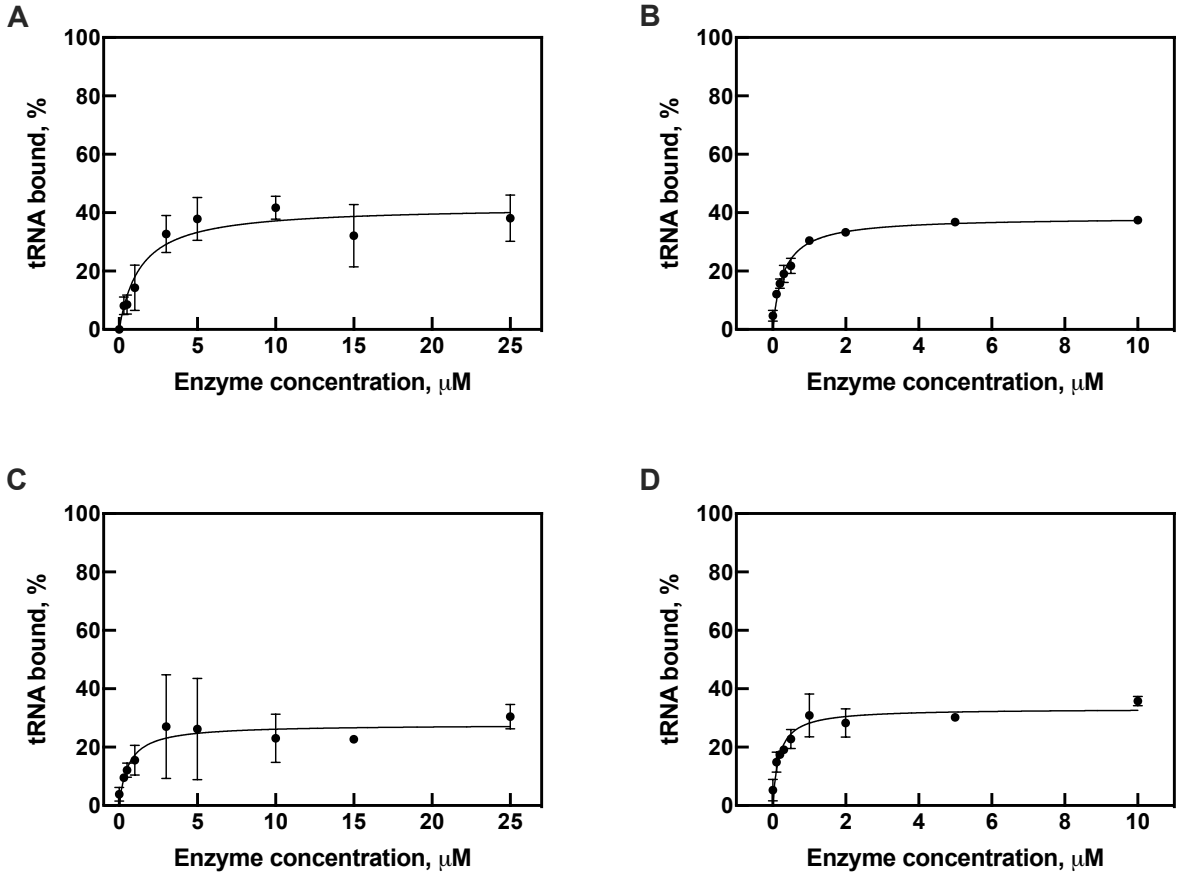


Figure 2.3 – Binding affinities of wild-type MnmE and MnmG for tRNA

Average binding curves for wild-type MnmE and MnmG to the $tRNA^{Glu}$ substrate in the absence (A,B) and presence of cofactors (C,D), N = 3. A hyperbolic function was fitted to each dataset a to determine the dissociation constant (K_D). All K_Ds are listed in Table 2.4.

Effects of cofactors on the binding affinity MnmE and MnmG have for tRNA

As previously discussed, the MnmEG complex is proposed to modify tRNA with the help of many cofactors. MnmE binds guanine nucleotides (GTP and GDP) and the THF-derivative required for uracil methylation, while MnmG binds electron carriers (FADH and NADH) and the amino acid substituent (e.g., glycine) that is inserted into the C5-methylated uracil. To test whether these cofactors affect substrate binding to the complex, triplicate nitrocellulose filter binding experiments were used to determine each protein's affinity for their tRNA substrate in the presence of these molecules. For these

experiments, all cofactors were present except the THF-derivative to prevent catalysis. Guanine nucleotide binding events are hypothesized to act as regulatory switches for MnmEG catalytic activity as they were recently found to induce conformational changes in MnmE.[9] In particular, the GTP binding event was linked to the dimerization of MnmE and the formation of the $\alpha_4\beta_2$ MnmEG complex. Consequently, GTP was the chosen guanine nucleotide for these experiments. As before, the percentage of protein-bound RNA was averaged over the three replicates, plotted against protein concentration and a hyperbolic function was fitted to the data to determine the K_Ds (Figure 2.3 C and D; Table 2.4).

Interestingly, in the presence of cofactors, MnmE and MnmG had similar binding magnitudes, with binding amplitude ranges of 27.5 ± 7.7% and 33.1 ± 2.5%, respectively. Under these conditions, MnmE interacted with tRNA with an average K_D of 0.55 ± 0.23 μM. This affinity is 2.5 times higher than what was observed in the absence of the cofactors (1.39 ± 0.38 μM). A similar phenomenon is observed in MnmG, which had an average K_D of 0.17 ± 0.02 μM, conveying an affinity 2 times higher than the previous experiment (0.30 ± 0.06 μM). Therefore, it can be concluded that cofactors increase MnmE and MnmG's affinity for their tRNA.

Table 2.4 – Affinity of wild-type MnmE and MnmG for $tRNA^{Glu}$

Protein	Dissociation constant (μM)	Binding Amplitude (%)
MnmE	1.39 ± 0.38	42.2 ± 7.2
MnmG	0.30 ± 0.06	38.5 ± 0.4
MnmE + cofactors	0.55 ± 0.23	27.5 ± 7.7
MnmG + cofactors	0.17 ± 0.02	33.1 ± 2.5

Binding specificity of MnmE and MnmG

Results from the previous experiments showed that MnmG (K_D = 0.30 ± 0.06 μM) has a greater binding affinity for tRNA than MnmE (K_D = 1.39 ± 0.38 μM). Cofactors were found to increase substrate binding affinity in both proteins, but MnmG's binding affinity was still greater (K_D = 0.17 ± 0.02 μM) than MnmE's (K_D = 0.55 ± 0.23 μM). Based on these observations, I hypothesized that MnmG may be binding tRNA specifically, whereas MnmE may be a non-specific RNA-binding protein. To test this hypothesis, duplicate nitrocellulose filter binding experiments were used to determine each protein's affinity for other RNA substrates, a short (27 nt), single-stranded RNA with no secondary structure,[6] and a long RNA (188 nt) with extensive secondary structure.

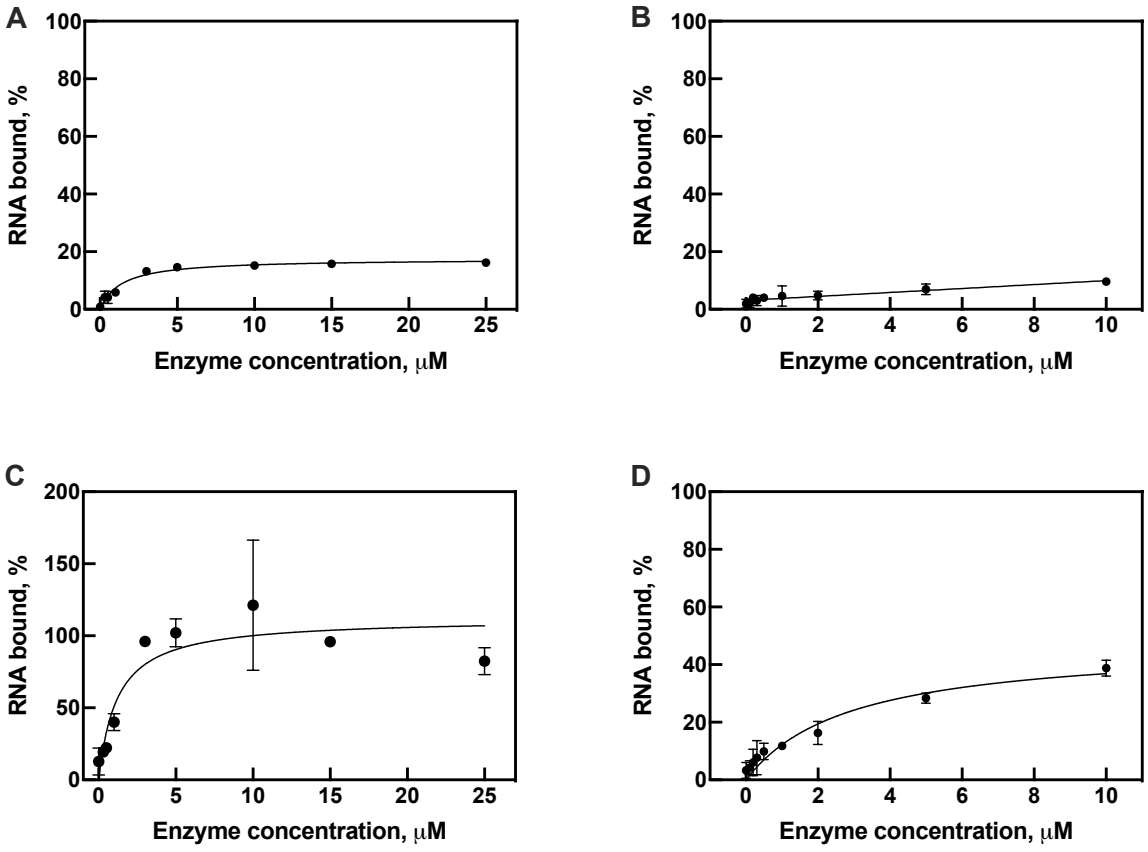


Figure 2.4 – Binding specificity of MnmE and MnmG enzymes

Average binding curves for wild-type MnmE (A: short, unstructured RNA, C: long, structured RNA) and MnmG (B: short, unstructured RNA, D: long, structured RNA) to other RNA substrates, N = 2. Hyperbolic functions were fitted to datasets A, C and D to determine the KDs. Dataset B was analyzed using linear regression at 95% confidence interval. The K_Ds are listed in Table 2.5

MnmE bound unstructured (K_D = 1.40 ± 0.57 μM) and structured RNA (K_D = 1.20 ± 0.34 μM) with comparable affinities to tRNA (K_D = 1.39 ± 0.38 μM; Figure 2.4A and C). However, variation is observed in the binding amplitudes of each RNA. MnmE binds the short, unstructured RNA with a lower amplitude (17.6 ± 1.1%) than tRNA (42.2 ± 7.2%), but binds the long, structured RNA with higher amplitude (112.4 ± 17.6%). This implies that MnmE may be a non-specific RNA-binding protein, that preferentially binds to RNA secondary structure.

MnmG bound the long, structured RNA (K_D = 3.17 ± 1.95 μM) with an affinity ten times lower than tRNA (K_D of 0.30 ± 0.06 μM; Figure 2.4D). Furthermore, although the protein interacted with the short, unstructured RNA, the data was linear and not well-fitted to a hyperbolic function (Figure 2.4B). Consequently, the dissociation constant and affinity could not be determined for this data set. All the binding data for the structured and unstructured RNA substrates are summarized in Table 2.5. This data suggests that MnmG is able to discriminate for secondary structure, and preferentially binds to tRNA.

Table 2.5 – Affinity of wild-type MnmE and MnmG for other RNA substrates

RNA substrate	Protein	Dissociation constant (μM)	Binding Amplitude (%)
Short, unstructured	MnmE	1.40 ± 0.57	17.6 ± 1.1
	MnmG	N/A	N/A
Long, structured	MnmE	1.20 ± 0.34	112.4 ± 17.6
	MnmG	3.17 ± 1.95	48.2 ± 6.3

Binding affinity of the MnmEG complex

Although evidence collected suggests MnmE is a non-specific RNA-binding protein, low concentrations of the protein bound a considerable amount of tRNA (8.6 ± 3.3% tRNA bound to 0.5 μM MnmE). Moreover, as previously stated, MnmE's GTPase cycle has been directly linked to conformational changes that are presumably distributed throughout the modification complex. Consequently, the effect MnmE has on MnmG's affinity for tRNA was investigated using nitrocellulose filter binding assays. In the

presence of 0.1 μM MnmE, MnmG bound tRNA with a K_D of 0.21 ± 0.03 μM, an affinity that is comparable to the protein's interaction with tRNA in the absence of MnmE (K_D = 0.30 ± 0.06 μM; Figure 2.5A). Nevertheless, MnmG bound tRNA with a higher amplitude in the presence of MnmE (52.0 ± 4.8%) than tRNA alone (38.5 ± 0.4%). In the presence of cofactors and 0.1 μM MnmE, the K_D for MnmG's interaction with tRNA was within the range of the protein's affinity for tRNA (0.26 ± 0.03 μM; Figure 2.5B). More interestingly, the presence of cofactors increased the binding amplitude to almost 100% (96.4 ± 3.2%; Table 2.6), which could be an indicator that under these conditions, a highly stable MnmEG complex can be formed. Hence, it can be concluded that MnmE improves tRNA binding to MnmG, and cofactors may play an important role in stabilizing the MnmEG complex once the protein-tRNA interactions are formed.

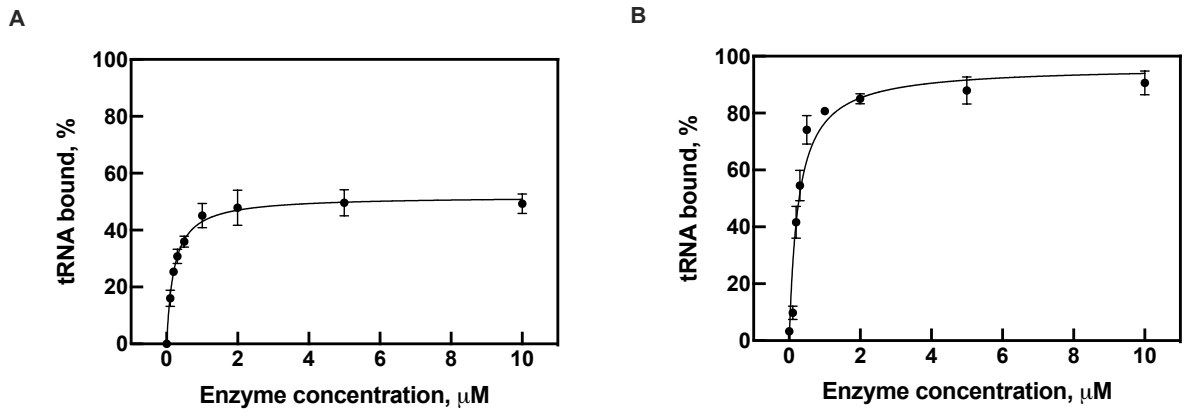


Figure 2.5 – Binding affinities of wild-type MnmG for tRNA in the presence of MnmE

Average binding curves of MnmG and 0.1 μM MnmE in the absence (A) and presence (B) of cofactors, N = 3. A hyperbolic function was fitted to each dataset a to determine the dissociation constant (K_D). All K_Ds are listed in Table 2.6.

Table 2.6 – Affinity of wild-type MnmG for $tRNA^{Glu}$ in the presence of 0.1 μM MnmE

Protein	Dissociation constant (μM)	Binding Amplitude (%)
MnmG	0.21 ± 0.03	52.0 ± 4.8
MnmG + cofactors	0.26 ± 0.03	96.4 ± 3.2

Impact of fluorescent protein fusions on tRNA binding by MnmE and MnmG

Although potential conditions for the formation of the MnmEG and the MnmEG-tRNA complexes have been discovered, ambiguity still abounds when it comes to the complex stoichiometry. Therefore, I planned to investigate the stoichiometric assembly of these complexes using analytical ultracentrifugation (AUC). For these experiments, the goal was to photometrically monitor the formation of the complexes as they sedimented. First, the C-termini of MnmE and MnmG were fused with either green fluorescent protein or mCherry fluorescent protein, generating four fusion proteins – MnmE-GFP, MnmE-mCherry, MnmG-GFP, and MnmG-mCherry. Subsequently, duplicate nitrocellulose filter binding assays were used to determine the tRNA affinity of each fusion protein, the average protein-bound tRNA was plotted against protein concentration and a hyperbolic function was fitted to the data to determine the K_Ds summarized in Table 2.7.

Table 2.7 – Affinity of fusion MnmE and MnmG proteins for $tRNA^{Glu}$

Protein	Dissociation constant (μM)	Binding Amplitude (%)
MnmE-GFP	0.76 ± 0.05	29.5 ± 0.6
MnmG-GFP	1.10 ± 0.41	66.7 ± 30.7
MnmE-mCherry	1.95 ± 0.23	61.4 ± 3.1
MnmG-mCherry	1.25 ± 0.10	48.0 ± 14.4

MnmE-GFP bound tRNA with a K_D of 0.76 ± 0.05 μM, an affinity two-fold higher than its wild-type variant (K_D = 1.39 ± 0.38 μM; Figure 2.6A). However, the fusion protein bound tRNA with a lower amplitude (29.5 ± 0.6%) than MnmE (42.2 ± 7.2%). On the other hand, while MnmE-mCherry's affinity for tRNA (K_D = 1.95 ± 0.23 μM; Figure 2.6C) is comparable to its wild-type variant (K_D = 1.39 ± 0.38 μM), it bound tRNA with a slightly higher amplitude (61.4 ± 3.1%) than MnmE (42.2 ± 7.2%). Hence, it was concluded that MnmE-mCherry best models the wild-type protein.

MnmG-GFP bound tRNA with a K_D of 1.10 ± 0.41 μM, while MnmG-mCherry bound the substrate with a K_D of 1.25 ± 0.10 μM (Figure 2.6B and D). Although these proteins had comparable affinities, they bound tRNA to different extents. MnmG-GFP bound tRNA with an amplitude of 66.7 ± 30.7%, while MnmG-mCherry bound the substrate with an amplitude of 48.0 ± 14.4%, which is closer to what was observed in the wild-type protein (42.2 ± 7.2%). MnmG-GFP and MnmG-mCherry showed reduced affinity for tRNA (3.7 and 4.2 times lower, respectively) relative to wild-type MnmG (K_D =0.30 ± 0.06 μM). Therefore, the addition of fluorescent proteins at the C-terminus of MnmG was found to be detrimental for the formation of tRNA-protein interactions.

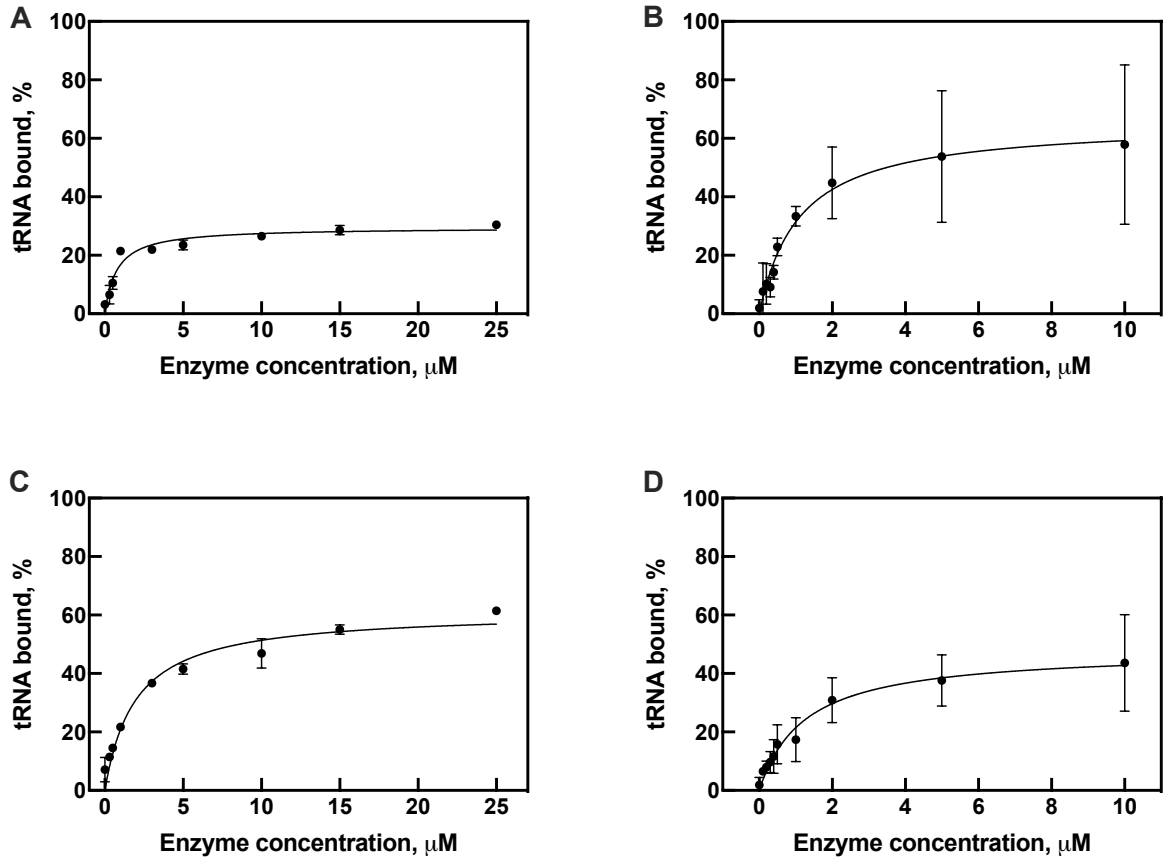


Figure 2.6 – Binding affinity of fusion MnmE and MnmG proteins for $tRNA^{Glu}$

Average binding curves of GFP and mCherry fused to MnmE (A, C) and MnmG (B, D) to the $tRNA^{Glu}$ substrate, N = 2. Hyperbolic functions were fitted to all datasets to determine the K_Ds and the K_Ds are listed in Table 2.7

2.3.4 Structural analysis of MnmE and MnmG

Structural analyses of proteins have shown that charged and polar residues are critical for the formation of protein-protein and protein-nucleic acid complexes. For proteins to complex with nucleic acids, they need positively charged surfaces or pockets to interact with the negatively charged nucleic acid phosphate backbone. Consequently, positively charged regions on protein surfaces have conventionally been identified as potential binding regions for nucleic acid molecules.[10-13]

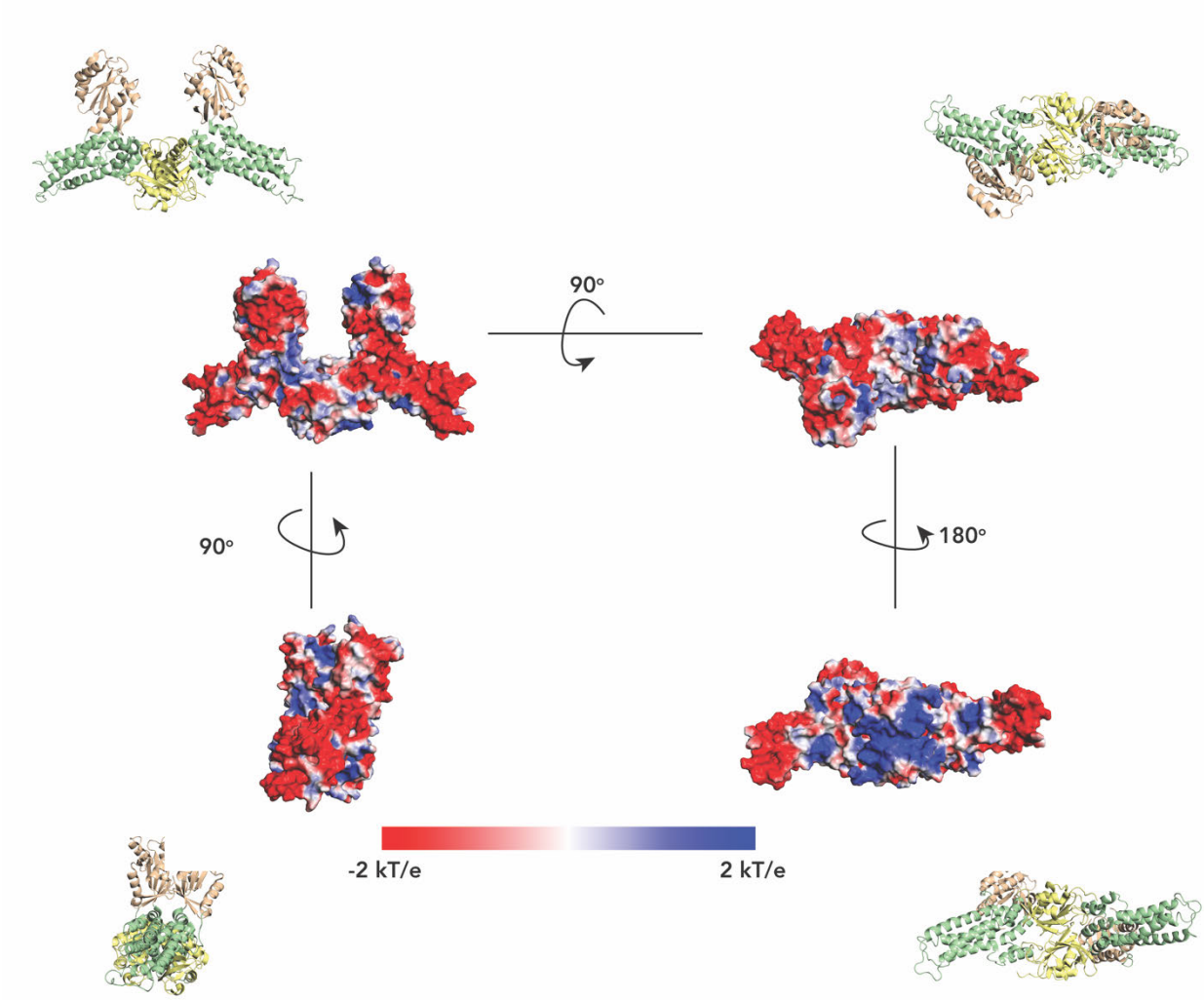


Figure 2.7 – Structural analysis of MnmE

Electrostatic potential mapping of symmetry modelled MnmE (PDB: 3GEE) derived from the Adaptive Poisson-Boltzmann Solver (APBS) module in PyMOL. Ribbon representations of MnmE are peripheral to their corresponding electrostatic surface maps and domains are color coded – G-domains in orange, helical domains in green and N-terminal domains in yellow.

Electrostatic analysis of MnmE reveals two positively charged regions on the protein's surface. The first is located within the predominantly negative G domains and is composed of small, interspersed patches located at dimerization interface, while the second is a larger, concentrated region at the bottom of the N-terminal domains of the protein (Figure 2.7). The helical domains of MnmE are predominantly composed of negatively charged residues, but positive patches are found at interdomain boundaries, especially the boundary between the helical and N-terminal domains.

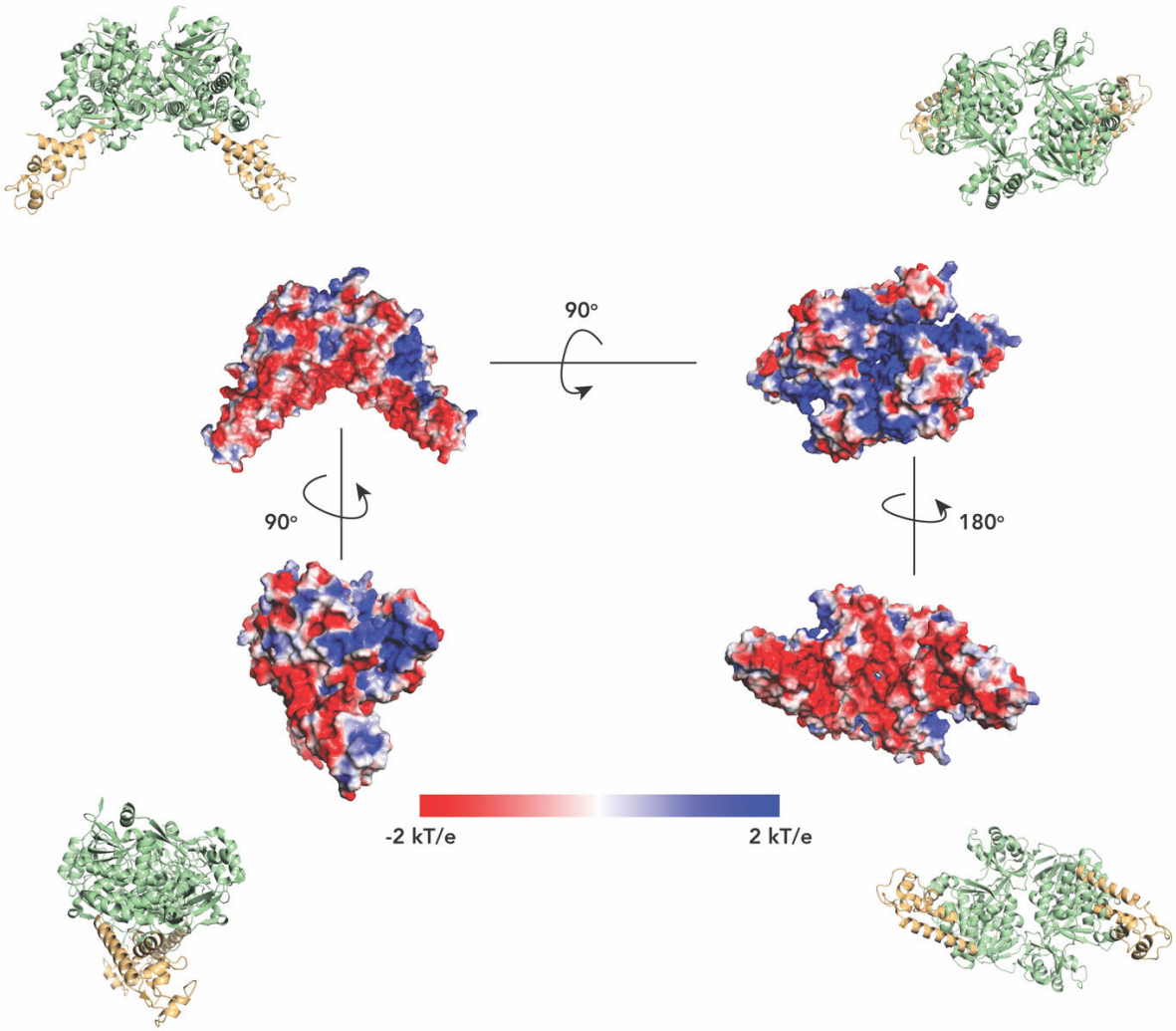


Figure 2.8 – Structural analysis of MnmG

Electrostatic potential mapping of symmetry modelled MnmG (PDB: 3CES) derived from the APBS module in PyMOL. Ribbon representations of the protein are peripheral to their corresponding electrostatic surface maps and domains are color coded – FAD-binding domains and insertion domains in green and helical domains in yellow.

Electrostatic analysis of MnmG revealed the protein has three deep, positively charged pockets at the top of the protein, located within its FAD-binding and insertion domains and all these pockets can partially accommodate the ASL and stabilize interactions with the tRNA substrate (Figure 2.8). On the other hand, there is a large negative pocket at the bottom of MnmG that encompasses the helical domains of the protein and the bottom of the FAD-binding domains.

2.4 Discussion

2.4.1 Contributions towards project objectives

Two proposals have been brought forward for the assembly of the MnmEG complex and its resulting action on the wobble tRNA uridine, and these reports diverge on the protein association, complex activation, and complex stoichiometry.[1,4] Therefore, the overarching goal of this project was to develop a deeper understanding of the MnmE-MnmG tRNA modification engine and work towards the isolation of a fully assembled complex. This thesis has characterized the binding properties of MnmE and MnmG to tRNA, in the absence and presence of cofactors, providing insight into the protein-tRNA interactions present within the complex. Furthermore, binding propensities MnmE and MnmG have for other RNA substrates have also been examined, providing insight into the complex's specificity. Lastly, in a bid to understand the complex's binding interactions, I have potentially isolated conditions under which a full, stable MnmEG-tRNA complex is formed.

2.4.2 Insights on the assembly of the MnmE-MnmG tRNA modification complex

Structural analyses of MnmE and MnmG revealed that both proteins have positively charged regions that could potentially interact with the tRNA substrate. However, it can be inferred that MnmE is less likely to contribute to the stabilization of the tRNA substrate within the complex's active site. Unlike MnmG that has three potential binding pockets for the substrate, MnmE has two positive regions on its surface, and neither of them can fully accommodate the tRNA's ASL. Moreover, the positively charged region at the protein's G domains participates in conformational changes that occur upon GTP binding, leaving only one possible binding site for the tRNA molecule.[9] The binding studies in this thesis have

shown that MnmG binds tRNA with greater affinity than MnmE, which concurs with the electrostatic potential profiles of each protein and proposals in literature that attribute tRNA binding to MnmG within this complex.

There are two proposals for the assembly of the MnmEG-tRNA complex – a symmetric $\alpha_2\beta_2\gamma_1$ complex proposed by Scrima *et al.*,[4,9] and an asymmetric complex that alternates between an $\alpha_2\beta_2\gamma_1$ and an $\alpha_4\beta_2\gamma_2$ stoichiometric assembly discovered by Fislage *et al*.[1,3,14,15] For the symmetric MnmEG complex to form, the tRNA substrate binds to the positively charged pocket in the center of the FAD-binding domains, and MnmE sits on top of the MnmG-tRNA complex, interacting through its N-terminal domain (Figure 2.9). In this complex, the ASL backbone is stabilized via interactions with MnmE and MnmG but the remainder of the tRNA is theoretically exposed to solvent. Moreover, in this configuration, MnmE and MnmG only interact at the tRNA binding site.

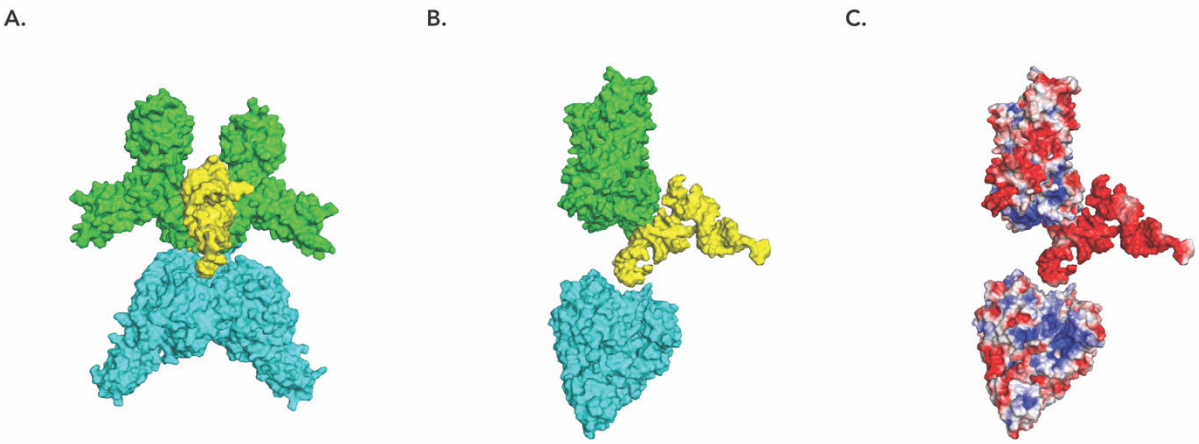


Figure 2.9 – Symmetric assembly of the MnmE-MnmG tRNA modification complex

Front (A) and side-views (B) for the symmetric assembly of the MnmEG-tRNA complex based on observations made by Scrima and collaborators.[4,9] Complex components (MnmE-green, MnmG-cyan, tRNA-yellow) were manually docked in PyMOL, using their electrostatic potential maps as reference. Electrostatic potential mapping of the MnmEG complex derived from the APBS module (C). The protein surface is colored according to electrostatic charge on a red (-2 kT/e) to blue (2 kT/e) scale.

On the other hand, the asymmetric complex is arranged in an $\alpha_2\beta_2\gamma_1$ manner when MnmE and MnmG come together in a transversal manner, with a single N-terminal and helical domain from MnmE interacting with one monomer of MnmG at its FAD-binding domain. Binding of the first MnmE dimer hypothetically induces conformational changes in MnmG, allowing a second MnmE dimer to bind the

second MnmG monomer in a transversal manner as well. The ASL is situated within a peripheral positively charged pocket on the MnmE-MnmG interface, and potentially two tRNA molecules can bind to the complex at once (Figure 2.10). Based on the electrostatic analyses of the proteins, I speculate that the asymmetric MnmEG-tRNA complex likely exists in an angular, “twisted” conformation rather than a perpendicular, “head-on” conformation. If MnmE and MnmG came together in a perpendicular manner to form the $\alpha_2\beta_2\gamma_1$ variant of the asymmetric complex, the tRNA’s ASL would lay with a peripheral positively charged pocket of MnmG and its acceptor stem would interact with the positively charged cavity at the center of MnmG. However, in this conformation, the binding of a tRNA substrate within the second peripheral pocket of MnmG would result in steric and electrostatic charges between the two tRNA molecules, destabilizing the complex. Consequently, if the asymmetric complex truly exists, the formation of an angular complex will likely be favored over that of a perpendicular complex.

Although both proposals for the assembly of the MnmEG complex are feasible, the asymmetric complex is more structurally favorable than the symmetric complex. Firstly, the tRNA substrate forms more stabilizing electrostatic interactions with MnmE and MnmG in the asymmetric complex that it does in the symmetric assembly. Secondly, the asymmetric assembly of the MnmEG complex allows extensive interactions to be formed between the MnmE and MnmG, stabilizing the overall structure of the complex. These interactions are lacking in the symmetric complex, making it less stable than the alternative.

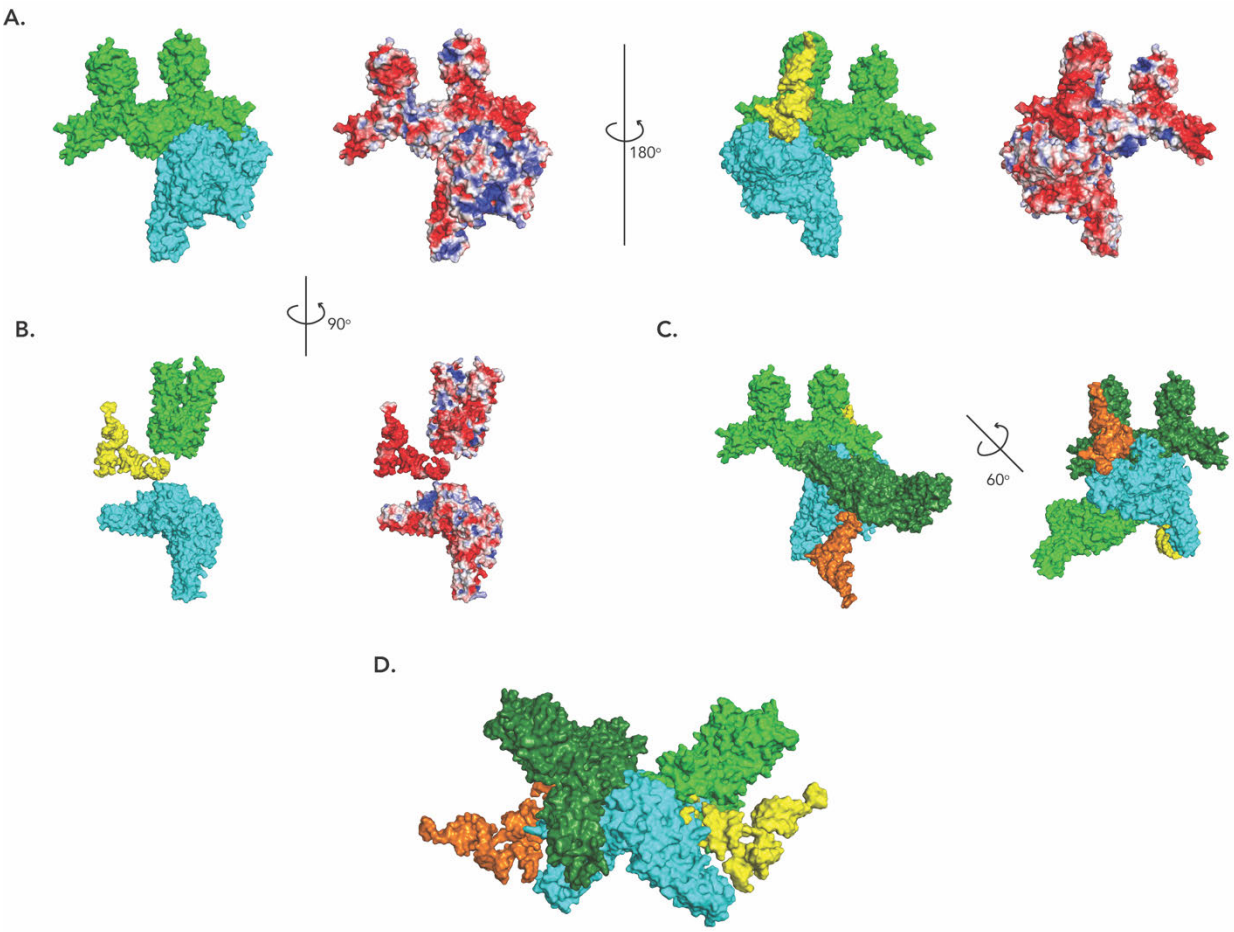


Figure 2.10 – Asymmetric assembly of the MnmE-MnmG tRNA modification complex

Front, back (A) and side views (B) for the asymmetric assembly of the $\alpha_2\beta_2\gamma_1$ MnmEG-tRNA complex based on observations made by Fislage and collaborators.[1,14] Complex components (MnmE-green, MnmG-cyan, tRNA-yellow) and were manually docked in PyMOL, using their electrostatic potential maps as reference. Electrostatic potential maps were derived from the APBS module, and the protein surface is colored according to electrostatic charge on a red (-2 kT/e) to blue (2 kT/e) scale. (C) Front and side views of the $\alpha_4\beta_2\gamma_2$ MnmEG-tRNA complex. The second MnmE dimer is colored forest green, while the second tRNA molecule is colored orange. (D) MnmG-centered view of the $\alpha_4\beta_2\gamma_2$ MnmEG complex.

2.4.3 Rationale for the formation of the MnmE-MnmG complex

Transfer RNAs are highly modified molecules across all domains of life. Although the most prevalent modifications are methylations and pseudouridylations, extensive modifications have been found at the ASL of tRNAs.[16-19] Interestingly, tRNAs are usually methylated and pseudouridylated by single enzymes like TrmA and RluA respectively, and tRNA modification complexes have rarely been discovered.[20] In bacteria two evolutionarily conserved tRNA modification complexes have been identified – the MnmEG complex, and the TsaBDE complex that inserts *N*^{6}-threonylcarbamoyladenosine (t^{6}A) at

position 37 of ANN decoding-tRNAs.[21,22] In eukaryotes, three evolutionarily conserved tRNA modification complexes have been discovered, and both complexes are responsible for the modification of wobble uridine bases. The MSS1/MTO1 complex (homologous to MnmEG), the KEOPS/EKC complex (homologous to TsaBDE), and the elongator complex – a conserved protein complex required for the formation of 5-carbamoylmethyluridine (ncm^5U34) and 5-methoxycarbonylmethyl(-2-thio)uridine ($mcm^5(s^{2)}U34$) in $tRNA^{Lys}_{UUU}$, $tRNA^{Glu}_{UUC}$ and $tRNA^{Gln}_{UUG}$.[23-26] The presence of the MnmEG complex across evolutionary domains and its uniqueness relative to other tRNA modification proteins begs the question, "Why are two proteins required for the formation of $cmnm^5(s^2)U34$?". This question could be somewhat addressed by considering the complex's components. MnmG is a FAD and NADH-binding protein that has been shown in this work to preferentially bind the tRNA substrate. On the other hand, MnmE is – according to my binding studies, a non-specific RNA binding protein, whose GTPase cycle has been proposed to control the initiation of $cmnm^5U34$ formation. Additionally, my binding data revealed that MnmE improves tRNA binding to MnmG to a manner akin to cofactors and existing proposals for the catalytic mechanism of this complex have attributed the formation of the modification's side chain to MnmE. Taken together, these observations suggest that MnmE may be an accessory protein to MnmG, and its purpose is to stabilize the substrate within the active site and provide the cofactors needed for tRNA modification. As a non-specific RNA binding protein, MnmE could hypothetically interact with other RNA molecules and participate in their modification as well. Apart from MnmG, MnmE has been found to bind other RNA modification enzymes MnmA (a tRNA-specific thiouridylase) and RluD (a pseudouridine synthetase for the ribosomal large subunit), and stress and replication proteins (DnaA, YhbZ, GyrB). The *mnmE* gene has also been linked to the pathogenicity of various bacteria including but not limited to *Salmonella*, *Francisella tularensis* and *Streptococcus mutans*.[27-29] Nevertheless, further investigations on MnmE are required to determine whether this protein has other cellular functions in addition to tRNA modification.

2.4.4 Future directions

The overarching goal of this project was to isolate the MnmEG complex and confirm recent reports on the formation of the $\alpha_4\beta_2$ complex. To this end, I planned to use multiwavelength analytical

ultracentrifugation (MWL-AUC)[30,31] to characterize the protein-protein and protein-tRNA interactions within the complex. Analytical ultracentrifugation (AUC) is a technique that employs centrifugal and optical systems to analyze the sedimentation and diffusion behavior of molecules under biologically relevant conditions. Additionally, employing a multi-wavelength detection system during an AUC experiment is ideal for studying protein-nucleic acid interactions *in vitro*, as it allows for the distinction of protein and nucleic acids signals based on their spectral differences, providing clear insight on the molar stoichiometry of such complexes.[32-35] To clearly distinguish between the spectral profiles of MnmE and MnmG, GFP and mCherry fluorescent proteins were fused to the C-termini of the complex proteins and the binding propensity of each fusion protein was evaluated to ensure they retained their catalytic abilities. The MnmE-mCherry protein best modelled wild-type MnmE's interaction with tRNA, but none of the MnmG fusion proteins were able to replicate the wild-type protein's propensity for the tRNA substrate. To create alternative fusion proteins of MnmG with mCherry or GFP in future, the fluorescent proteins could be fused to the N-terminus of MnmG. Therefore, I suggest that only MnmE-mCherry be used in the upcoming AUC experiments. These experiments will be used to characterize each protein's interaction with tRNA, as well as each other. Titration experiments will then be conducted to test for the formation of the $\alpha_4\beta_2$ complex, titrating the MnmE(-mCherry) against MnmG. All AUC experiments will take into consideration the effects cofactors have on complex assembly, especially GTP and its derivatives as their binding have been proposed to induce complex-wide conformational changes.

Although the structures of the monomeric MnmE and MnmG have been resolved using X-ray crystallography, the dimeric structure of both proteins are yet to be fully described, though models have been computationally constructed via symmetry modelling and SAXS experiments.[1,4,36,37] Therefore, in addition to AUC experiments, future structural studies on the MnmEG complex could use cryogenic electron microscopy (cryo-EM) to elucidate the structure of the MnmE-MnmG complex. With this technique, an electron beam is fired at a frozen biomolecule and uses the scattered electrons to resolve microscopic images of individual molecules.[38] Cryo-EM surpasses X-ray crystallography in that it does not require the generation of protein crystals to reconstruct the three-dimensional shape of a molecule. Protein crystallization is a limiting factor in X-ray crystallography, as some molecules can take years to crystallize, while others do not crystallize at all.[39] More importantly, whereas X-ray crystallography

provides a single, well-defined conformation of a crystallized protein, cryo-EM freezes proteins from solution, allowing the sampling of multiple conformations. Recent advances in this area have resulted in the imaging of high-resolution structures of large complexes like the human Drosha-DGCR8-pri-miRNA complex, making this technique ideal for structural investigations on the MnmE-MnmG complex.[40]

Another point of contention in the literature on the MnmE-MnmG complex is its catalytic mechanism, and two hypotheses have been proposed – the first by Scrima *et al*.[4,9] and the other by Moukadiri and collaborators (Chapter 1, Figure 1.5).[2,3] These proposals overlap in that they attribute the creation of the modification moiety to MnmE and the rearrangement of the C5-moeity to MnmG. Nevertheless, they diverge in the cofactors and mechanism used in each protein, especially MnmE.[2,4,10,21] To gain more insight into this complex's catalytic mechanism, future studies could probe the active site of the complex using mutational studies or use transition-state analogues to capture the complex in transient conformational states. The caveat here is the lack of complete structural information on the active site of this modification engine. Consequently, the revelation of its structure is critical to a deeper understanding of the complex's catalytic activity. Gaining full structural information on this complex will also allow for computational probing of the complex's mechanism. Advances in computational chemistry have resulted in the deciphering of the catalytic mechanisms of many proteins, including the human DNA repair enzyme APE1 and various DNA glycosylases, and these studies often complement their experimental counterparts.[41,42] Similarly, this method could be used to investigate the MnmEG complex and gain insight into its catalytic mechanism.

2.5 References

1 Fislage, M. *et al.* SAXS analysis of the tRNA-modifying enzyme complex MnmE/MnmG reveals a novel interaction mode and GTP-induced oligomerization. *Nucleic Acids Res* **42**, 5978-5992, doi:10.1093/nar/gku213 (2014).

2 Moukadiri, I., Garzon, M. J., Bjork, G. R. & Armengod, M. E. The output of the tRNA modification pathways controlled by the Escherichia coli MnmEG and MnmC enzymes depends on the growth conditions and the tRNA species. *Nucleic Acids Res* **42**, 2602-2623, doi:10.1093/nar/gkt1228 (2014).

3 Moukadiri, I. *et al.* Evolutionarily conserved proteins MnmE and GidA catalyze the formation of two methyluridine derivatives at tRNA wobble positions. *Nucleic Acids Res* **37**, 7177-7193, doi:10.1093/nar/gkp762 (2009).

4 Scrima, A., Vetter, I. R., Armengod, M. E. & Wittinghofer, A. The structure of the TrmE GTP-binding protein and its implications for tRNA modification. *Embo J* **24**, 23-33, doi:10.1038/sj.emboj.7600507 (2005).

5 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* **12**, 291-299, doi:10.1093/dnares/dsi012 (2005).

6 Kelly, E. K., Czekay, D. P. & Kothe, U. 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* **25**, 1393-1404, doi:10.1261/rna.071043.119 (2019).

7 Schrödinger, L. The PyMOL Molecular Graphics System, Version 2.0

8 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **98**, 10037-10041, doi:10.1073/pnas.181342398 (2001).

9 Scrima, A. & Wittinghofer, A. Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. *Embo J* **25**, 2940-2951, doi:10.1038/sj.emboj.7601171 (2006).

10 Abdi, N. M. & Fredrick, K. Contribution of 16S rRNA nucleotides forming the 30S subunit A and P sites to translation in Escherichia coli. *RNA* **11**, 1624-1632, doi:10.1261/rna.2118105 (2005).

11 Sheinerman, F. B., Norel, R. & Honig, B. Electrostatic aspects of protein–protein interactions. *Current Opinion in Structural Biology* **10**, 153-159, doi:https://doi.org/10.1016/S0959-440X(00)00065-8 (2000).

12 Vascon, F. *et al.* Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design. *Computational and Structural Biotechnology Journal* **18**, 1774-1789, doi:https://doi.org/10.1016/j.csbj.2020.06.029 (2020).

13 Zhou, H.-X. & Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical reviews* **118**, 1691-1741, doi:10.1021/acs.chemrev.7b00305 (2018).

14 Fislage, M., Wauters, L. & Versees, W. MnmE, a GTPase That Drives a Complex tRNA Modification Reaction. *Biopolymers* **105**, 568-579, doi:10.1002/bip.22813 (2016).

15 Yim, L., Moukadiri, I., Bjork, G. R. & Armengod, M. E. Further insights into the tRNA modification process controlled by proteins MnmE and GidA of Escherichia coli. *Nucleic Acids Res* **34**, 5892-5905, doi:10.1093/nar/gkl752 (2006).

16 Agris, P. F. Decoding the genome: a modified view. *Nucleic Acids Res* **32**, 223-238, doi:10.1093/nar/gkh185 (2004).

17 Agris, P. F. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* **9**, 629-635, doi:10.1038/embor.2008.104 (2008).

18 Gustilo, E. M., Vendeix, F. A. & Agris, P. F. tRNA's modifications bring order to gene expression. *Curr Opin Microbiol* **11**, 134-140, doi:10.1016/j.mib.2008.02.003 (2008).

19 Zhang, J. & Ferre-D'Amare, A. R. The tRNA Elbow in Structure, Recognition and Evolution. *Life (Basel)* **6**, doi:10.3390/life6010003 (2016).

20 Krutyhołowa, R., Zakrzewski, K. & Glatt, S. Charging the code — tRNA modification complexes. *Current Opinion in Structural Biology* **55**, 138-146, doi:https://doi.org/10.1016/j.sbi.2019.03.014 (2019).

21 Armengod, M. E. *et al.* Enzymology of tRNA modification in the bacterial MnmEG pathway. *Biochimie* **94**, 1510-1520, doi:10.1016/j.biochi.2012.02.019 (2012).
22 Missoury, S. *et al.* The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t6A tRNA-modification. *Nucleic Acids Res* **46**, 5850-5860, doi:10.1093/nar/gky323 (2018).
23 Chen, C., Huang, B., Eliasson, M., Rydén, P. & Byström, A. S. Elongator Complex Influences Telomeric Gene Silencing and DNA Damage Response by Its Role in Wobble Uridine tRNA Modification. *PLOS Genetics* **7**, e1002258, doi:10.1371/journal.pgen.1002258 (2011).
24 Srinivasan, M. *et al.* The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *The EMBO Journal* **30**, 873-881, doi:https://doi.org/10.1038/emboj.2010.343 (2011).
25 Colby, G., Wu, M. & Tzagoloff, A. MTO1 codes for a mitochondrial protein required for respiration in paromomycin-resistant mutants of Saccharomyces cerevisiae. *J Biol Chem* **273**, 27945-27952, doi:10.1074/jbc.273.43.27945 (1998).
26 Decoster, E., Vassal, A. & Faye, G. MSS1, a nuclear-encoded mitochondrial GTPase involved in the expression of COX1 subunit of cytochrome c oxidase. *J Mol Biol* **232**, 79-88, doi:10.1006/jmbi.1993.1371 (1993).
27 Faron, M. *et al.* The Francisella tularensis migR, trmE, and cphA genes contribute to F. tularensis pathogenicity island gene regulation and intracellular growth by modulation of the stress alarmone ppGpp. *Infect Immun* **81**, 2800-2811, doi:10.1128/iai.00073-13 (2013).
28 Shippy, D. C. & Fadl, A. A. tRNA modification enzymes GidA and MnmE: potential role in virulence of bacterial pathogens. *Int J Mol Sci* **15**, 18267-18280, doi:10.3390/ijms151018267 (2014).
29 Shippy, D. C., Eakley, N. M., Lauhon, C. T., Bochsler, P. N. & Fadl, A. A. Virulence characteristics of Salmonella following deletion of genes encoding the tRNA modification enzymes GidA and MnmE. *Microb Pathog* **57**, 1-9, doi:10.1016/j.micpath.2013.01.004 (2013).
30 Howlett, G. J., Minton, A. P. & Rivas, G. Analytical ultracentrifugation for the study of protein association and assembly. *Current Opinion in Chemical Biology* **10**, 430-436, doi:https://doi.org/10.1016/j.cbpa.2006.08.017 (2006).
31 Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science* **11**, 2067-2079 (2002).
32 Gorbet, G. E., Pearson, J. Z., Demeler, A. K., Cölfen, H. & Demeler, B. in *Methods in Enzymology* Vol. 562 (ed James L. Cole) 27-47 (Academic Press, 2015).
33 Horne, C. R., Henrickson, A., Demeler, B. & Dobson, R. C. J. Multi-wavelength analytical ultracentrifugation as a tool to characterise protein–DNA interactions in solution. *European Biophysics Journal* **49**, 819-827, doi:10.1007/s00249-020-01481-6 (2020).
34 Mitra, S. & Demeler, B. in *RNA Spectroscopy* 281-317 (Springer, 2020).
35 Cole, J. L., Lary, J. W., P. Moody, T. & Laue, T. M. in *Methods in Cell Biology* Vol. 84 143-179 (Academic Press, 2008).
36 Meyer, S. *et al.* Kissing G domains of MnmE monitored by X-ray crystallography and pulse electron paramagnetic resonance spectroscopy. *PLoS Biol* **7**, e1000212, doi:10.1371/journal.pbio.1000212 (2009).
37 Meyer, S., Scrima, A., Versees, W. & Wittinghofer, A. Crystal structures of the conserved tRNA-modifying enzyme GidA: implications for its interaction with MnmE and substrate. *J Mol Biol* **380**, 532-547, doi:10.1016/j.jmb.2008.04.072 (2008).
38 Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1862**, 324-334, doi:https://doi.org/10.1016/j.bbagen.2017.07.020 (2018).
39 Bai, X.-c., McMullan, G. & Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences* **40**, 49-57, doi:https://doi.org/10.1016/j.tibs.2014.10.005 (2015).
40 Partin, A. C. *et al.* Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. *Molecular Cell* **78**, 411-422.e414, doi:https://doi.org/10.1016/j.molcel.2020.02.016 (2020).
41 Aboelnga, M. M. & Wetmore, S. D. Unveiling a Single-Metal-Mediated Phosphodiester Bond Cleavage Mechanism for Nucleic Acids: A Multiscale Computational Investigation of a Human

DNA Repair Enzyme. *Journal of the American Chemical Society* **141**, 8646-8656, doi:10.1021/jacs.9b03986 (2019).
42 Kaur, R., Nikkel, D. J. & Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Computational Molecular Science* **10**, e1471, doi:https://doi.org/10.1002/wcms.1471 (2020).

CHAPTER 3: DEVELOPING AN ACCURATE MD SIMULATION PROTOCOL FOR SAMPLING tRNA STRUCTURE

3.1 Objectives

Prior to computationally investigating the effects of cmnm^5(s^2)U34 modifications on tRNA structure, an efficient MD protocol needed to be developed for representative sampling of the molecule's conformational space. Historically, single classical all-atom MD (cMD) simulations have been used to describe dynamic properties of tRNA, but structural models and timescales for these simulations vary from one study to another.[1-11] Two structural models of tRNA have been probed in the literature using a cMD approach – a full tRNA model[12,13] and an ASL model, which only includes residues within the anticodon stem-loop (residues 27 to 43).[13,14] To the best of my knowledge, the first cMD tRNA simulation ever run was a 32 ps all-atom trajectory on tRNA^{Asp}.[4] Since then, sampling times of tRNA cMD simulations have been extended to the μs timescale, with the longest tRNA cMD being 1 μs long.[10,12]

A handful of studies have used a replica MD (rDM) approach, which involves running a number of identically prepared short cMD simulations (replicas) in parallel, to study tRNA dynamics. However, the protocols for these works also vary from one study to another. The first tRNA rMD protocol involved a six 500 ps replica ensemble and trajectories were performed on the ASL of tRNA^{Asp}.[5] As of today, the timescale (500 ps to 1 μs) and replica number (3 to 6) for tRNA rMD simulations varies greatly.[2,3,15] This use of inconsistent approaches to study different tRNA sequences across a plethora of organisms makes it difficult to draw accurate cross-study comparisons on the structural dynamics of tRNA. Furthermore, while a few studies have sampled multiple experimentally isolated tRNA conformations across various sequences,[16-22] no computational study has successfully sampled and described all states observed in literature. Consequently, the development of a computational protocol that ensures representative sampling of tRNA conformations is required to achieve conformational convergence in tRNA MD investigations. More importantly, and directly related to the present thesis, such a protocol can be used in future work to obtain a deeper understanding of tRNA structural dynamics in the presence and absence of modifications, and therefore enhance our knowledge on the role of tRNAs and tRNA modifications *in vivo.*

The goal of this study is to compare the performance of a single, long MD simulation to that of various replica ensembles in order to establish an accurate MD protocol for sampling the phase space of tRNA. MD simulations were performed on unmodified *E. coli* $tRNA^{Phe}$ to capture the structural dynamics of the molecule in the absence of modifications. This representative system was chosen because experimental evidence reveals that tRNAs adopt the same tertiary structure regardless of sequence origin, modification or environment (unbound, protein-bound, ribosome-bound). Modified nucleotides have been proven to affect local and global arrangements within tRNA molecules.[11,14,21,23-34] Clustering analysis, principal component analysis and residue-based free energy decomposition were carried out on a range of trajectories to describe the conformational space of tRNA surveyed by cMD and rMD methods, and a variety of statistical analyses were also conducted to determine convergence in the conformational sampling across the simulation sets. In particular, this study explores the structural arrangements within the ASL across thirty-one MD trajectories spanning 45 μs of simulation time, with the hope to sample the conformations observed in experimental and computational literature. The results obtained: (i) reveal stark differences in the structural sampling efficiencies of cMD and rMD, (ii) aid in the development of a structure-based classification of ASL conformations of tRNA, and (iii) provide recommendations on target simulations for the accurate description of tRNA in future studies. In summary, this inquiry reveals insights into the convergence of tRNA MD simulations – especially structural convergence within the ASL – and establishes a computational protocol for representative and efficient sampling of tRNA structural dynamics using MD simulations.

3.2 Computational Background

MD simulations investigate the atomic structure and associated dynamics of a molecular system by solving Newton's equations of motion for that assembly.[35] As of today, cMD simulations have been used to infer physical and chemical properties of many biomolecules including proteins, lipids, carbohydrates, DNA and RNA.[35-39] Furthermore, the technique has been employed to explain experimental data and inform the design of new experiments.[40-42] Nevertheless, despite the great successes cMD simulations have achieved in providing insight into the structures and functions of many biomolecules, representative sampling in these simulations is an issue for concern as biomolecules have complex conformational

spaces.[43] The potential energy surface of a biomolecule is often rugged, with many accessible local minima that represent key conformational states. Transitions between these states take time as low energy conformations of some biomolecules are separated by large energy barriers that need to be overcome to achieve representative sampling. Some biomolecules can also adopt a large number of stable states connected by a limited number of transition pathways, which results in longer computing time for conformational changes and reduced transition frequencies among stable states (sampling bottleneck).[44-46] These impediments imply that MD simulations can be trapped at local minima and the resultant sampling may not be representative of reality. Consequently, the concern associated with cMD is determining whether the simulation attains conformational convergence, which is defined by the accurate and representative sampling of all possible states (phase space) of the biomolecular system of interest.[44,47]

RNA is a complex molecule that takes on a wide variety of roles in the cellular processing of genetic information, and its structure and dynamics have been proven to be critical to its function.[48-54] Many studies have investigated the effects of simulation conditions, such as forcefields, water models and ionic strength, on the stability of RNA molecules during MD studies,[55-66] but investigations pertaining to the conformational convergence of specific RNAs in MD simulations have yet to be conducted. Nevertheless, computational enquiries that aimed to probe a variety of RNA structures, including short oligonucleotides, hairpins and aptamers, have employed the replica protocol to extend the simulation time-scale and increase the probability of sampling infrequent dynamic events on the molecule's potential energy surface.[67,68] Relative to single, long simulations, multiple-replica protocols have shown high agreement between the computational predictions and experimentally observed states.[69,70] For instance, a recent study by Bottaro and colleagues used 24-replica ensembles of 1 μs-simulations to computationally describe the conformational states of four RNA tetranucleotides, and the predicted structural states complemented NMR data and helped identify artifacts associated with solution-state NMR experiments.[71] Another study by Steuer *et al.*[72] employed a 15-replica ensemble of 2 μs-trajectories to describe conformational changes in the guanidine-II riboswitch aptamer upon ligand binding. Their simulations successfully characterized aptamer states isolated in X-ray diffraction studies and identified potential transition pathways for conformational switching upon interactions with various ligands. Finally,

Sorin and colleagues showed that 200-replica ensembles of ~ 10 ns trajectories can accurately describe folding and unfolding dynamics of small (12 nt) RNA hairpins.[73] Although multiple copies of short MD trajectories have been used to probe conformational changes in RNA, no study has compared the differential performance of a single, long simulation to that of a replica ensemble in the conformational sampling of RNA molecules. Furthermore, because these molecules vary in size, motifs, tertiary and quaternary arrangements, [74,75] multiple studies may be required to accurately describe the conformational space of various RNA molecules.

In contrast to RNA, sampling convergence studies have been conducted on other biomolecules, but the conclusions in terms of timescale and number of replications varies from one biosystem to another. For example, in a study devoted to the behaviour of a B-DNA duplex, Gallindo-Murillo *et al.* compared the performance of 200 trajectories in the ns range with single trajectories in the μs range. The authors concluded that ensembles of multiple short simulations sample the phase space of a DNA duplex to the same extent as one long simulation.[76] In contrast, Caves and colleagues compared the ability of ten 120 ps trajectories to that of a 1 ns trajectory to sample the dynamic properties of the peptide crambin.[77] They concluded that multiple short trajectories started at different points in phase space achieve more efficient sampling than a single long trajectory. Another study by Genheden and Ryde compared the ability of twenty 400 ps trajectories to that of a 10 ns trajectory to sample the dynamic properties of avidin, a biotin-binding protein found in bacteria, avians and amphibians.[78] The authors used the molecular mechanics/generalized Born surface area (MM/GBSA) method to evaluate statistical convergence of four binding events to avidin, and concluded that several independent short simulations are needed to obtain converged results with a statistical precision of 1 kJ/mol. A more recent study by Perez *et al.* examined the dynamics of the 16-mer peptide $Bak_{16}BH3$ using an 8 μs MD simulation, eight 1 μs simulations and eighty 0.1 μs (100 ns) simulations, and also concluded the ensemble of eight simulations can provide better conformational sampling than a unique trajectory.[79] However, they caution that the minimum MD trajectory length should be chosen wisely, as the short trajectories need to be long enough to overcome energy barriers along the potential energy surface of the biomolecule being investigated. These select examples emphasize the importance of testing the performance of cMD and

rMD protocols for each type of RNA molecules like tRNA, as conformational convergence varies from one structural system to another.

3.3 Methodology

3.3.1 Model preparation

The initial coordinates of standalone, full-length, unmodified *Escherichia coli* $tRNA^{Phe}$ were taken from a 3.0 Å resolution X-ray crystal structure (PDB ID: 3L0U).[80] The tRNA model was neutralized using Na^+ ions and excess Na^+ and Cl^- ions were added to attain a physiological concentration of 150 mM.[81] The system was solvated in an explicit TIP3P octahedral water box such that the solute was at least 10.0 Å away from the box edge in all directions. The model was prepared using the LEaP module in the AMBER 18 package and the tRNA was described using the ff99bsc0χ_{OL3} forcefield.[82]

3.3.2 MD simulation protocol

The positions of all solvent molecules and ions were initially minimized using 2500 steps of steepest descent, followed by 2500 steps of conjugate gradient minimization, using a force constant of 100 kcal mol^{-1} $Å^{-2}$ to constrain the tRNA. Subsequently, heavy atoms of the solvent molecules and ions were constrained using a force constant of 100 kcal mol^{-1} $Å^{-2}$ and the positions of all hydrogen atoms within the model were minimized using 2500 steps of steepest descent, followed by 2500 steps of conjugate gradient minimization. The solute was then minimized using 2500 steps of steepest descent, followed by 2500 steps of conjugate gradient minimization, while a force constant of 100 kcal mol^{-1} $Å^{-2}$ was applied to all solvent and ion molecules. Finally, the entire system was minimized using 2500 steps of unrestrained steepest descent, followed by 2500 steps of unrestrained conjugate gradient minimization.

Following minimization, the solute was restrained using a 25 kcal mol^{-1} $Å^{-2}$ force constant and the system was heated from 0 to 310 K in 50 K increments using the Langevin thermostat (γ = 1) and a 1 fs time step. Subsequently, the force restraints on the solute were reduced in a stepwise manner, from 25 to 1.5 kcal mol^{-1} $Å^{-2}$, using a time step of 2 fs and the SHAKE algorithm under NPT conditions (1 atm, 310 K). The PMEMD cuda module of AMBER 18 was used to perform MD production simulations of

various lengths using a time step of 2 fs.[82] The production trajectories were started from the same minimized conformation and all simulations were carried out with the periodic boundary condition, using a 10 Å non-bonded cut-off and accounting for electrostatic interactions using the particle mesh Ewald (PME) method. In total, thirty-one production simulations were performed, including thirty 500-ns simulations and a 5 μs simulation.

3.3.3 Analyses

Trajectories were sampled for analysis every 200 ps over the course of the production phase. The AmberTools 20 version of CPPTRAJ was used to analyze all trajectories.[83,84] To assess system stability, heavy-atom root-mean-square deviations (RMSDs) for each simulation were evaluated with respect to the crystal structure coordinates. Hydrogen-bond interactions were evaluated using a distance cut-off of 3.4 Å and an angle cut-off of 120°, while stacking occupancies were determined using a distance cut-off of ≤ 6 Å between the center of masses and an angle cut-off ≤ 40° or ≥ 140° between the normal vectors of the planes of the two bases.

The internal structural dynamics at the ASL were analyzed using the Barnaba library in Python.[85] The solvent, ions and all other tRNA domains were stripped from the trajectories and conformational analyses were performed on residues 27 to 43. Heavy atom RMSD and ensemble averaged RMSD (εRMSD) were obtained to visualize the dynamics of the ASL using the minimized crystal structure as reference. εRMSD is a contact map-based distance that describes the relative arrangements of the tRNA nucleotides, making εRMSD ideal for detecting conformational changes.[86] To calculate the εRMSD for a three-dimensional nucleic acid structure *a*, position vectors R_{ij}^{a} are calculated for every residue in the molecule and rescaled to introduce an ellipsoidal anisotropy particular to base–base interactions:

$$\tilde{r}_{ij}^{a} = \left(\frac{x_{ij}^{a}}{a}, \frac{y_{ij}^{a}}{a}, \frac{z_{ij}^{a}}{b}\right), \text{ with } a = 5 \text{ Å and } b = 3 \text{ Å}.$$

Given two structures *a* and *b* consisting of *N* residues, εRMSD is calculated as

$$\varepsilon\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i,j} |\mathbf{G}(\tilde{\mathbf{r}}_{ij}^{a}) - \mathbf{G}(\tilde{\mathbf{r}}_{ij}^{b})|^2}.$$

G is a nonlinear, Heaviside function of $\tilde{\mathbf{r}}$ defined as

$$\mathbf{G}(\tilde{\mathbf{r}}) = \begin{pmatrix} \sin(\gamma\tilde{r})\frac{\tilde{r}_x}{\tilde{r}} \\ \sin(\gamma\tilde{r})\frac{\tilde{r}_y}{\tilde{r}} \\ \sin(\gamma\tilde{r})\frac{\tilde{r}_z}{\tilde{r}} \\ 1+\cos(\gamma\tilde{r}) \end{pmatrix} \times \frac{\Theta(\tilde{r}_{cutoff} - \tilde{r})}{\gamma}, \text{ where } \tilde{r}_{cutoff} = 2.4 \text{ Å and } \gamma = \frac{\pi}{\tilde{r}_{cutoff}}.$$

Following the εRMSD calculation, structures from each trajectory were clustered using the DBSCAN algorithm.[87] The clustering analysis was visualized by projecting the trajectory onto the first two components of a principal component analysis performed on the collection of G-vectors. For every cluster, structures with the lowest average distance from all other cluster members were identified as the centroid. Upon visual inspection of the three-dimensional representation of each centroid in PyMOL, similar clusters were classified into conformational groups.[88] To determine the variability within each conformational group, the clustered frames were saved to independent trajectories and the base–base interactions were quantified using CPPTRAJ. A dynamic secondary structure representation of each conformational group was built using BARNABA to visualize variations in base-pairing and base-stacking interactions for each conformational group. BARNABA defaults to a distance cut-off of 3.4 Å and an angle cut-off of 120° for hydrogen-bond interactions and a distance cut-off of ≤ 4 Å between the center of masses and an angle cut-off < 40° or > 140° between the normal vectors of the planes of the bases for stacking interactions.

The conformational occupancies across the thirty 500-ns simulations were tallied and confidence intervals about the population mean were evaluated for the dominant conformations. To evaluate the ASL phase space sampled by the replica and the 5 μs simulations, dihedral principal component analyses (dPCA) were conducted on the simulations using the backbone torsions (α, β, γ, δ, ε, ζ, χ) of all ASL residues.[89] Free energy landscapes of the trajectory sets were calculated and the results were projected onto the first two principal components from the subspace of the replica ensemble.

3.4 Results and Discussion

3.4.1 The global structural features of tRNA are maintained across thirty 500-ns replicas

To investigate the conformational space of tRNA sampled using the replica protocol, thirty 500-ns MD simulations were conducted. Each trajectory was independently analyzed, and average replica properties were assessed over all replicas. For each simulation, RMSDs of all tRNA heavy atoms were estimated relative to the crystal structure to measure the trajectory equilibrium. Overall, RMSDs for each replica showed stable fluctuations throughout the simulations, and the ensemble had a median average RMSD of 4.0 ± 0.6 Å (Figure 3A.1). To evaluate the structural flexibility of each tRNA region, root-mean-square fluctuations (RMSFs) of all heavy atoms in each nucleotide were calculated. Predictably, RMSF analysis revealed the loop regions to be the most flexible, especially the D arm, ASL and the variable loop, with the average RMSF for residues within these regions ranging from 2 to 12 Å (Figure 3A.2). In contrast, the stem regions of all hairpin domains had lower amplitude fluctuations, resulting in RMSF values as low as 1.5 Å for these regions. To visualize tRNA dynamics within the replica ensemble, representative structures for all simulations were overlayed onto the crystal structure reference. Concurrent with the RMSF analysis, the structural overlays revealed that the backbone atoms at the tRNA elbow and ASL have the greatest deviation relative to the crystal structure. More importantly, RMSF analysis and backbone overlays of the replica ensemble revealed that variable conformations were adopted by the ASL, highlighting the flexible nature of this domain that has been observed in the experimental literature.[1,20,90-93]

Despite the great flexibility at the elbow region, the non-covalent interactions that govern tRNA folding were maintained throughout all trajectories. Specifically, hydrogen bonding interactions involved in tertiary base-pairs were maintained throughout all trajectories (occupancies > 85%; Figure 3A.3), while stacking interactions between the nucleobases within D and TψC loop were also preserved, with occupancies generally > 85%. The only exception is the G19/G57 stack that had an average occupancy of 60 ± 30% due to the flexible nature of G19, which arises from it being one of the outermost and solvent exposed bases in the elbow region. Like G19, C56 is also an outermost base in this region but C56 is less dynamic as it partakes in a strong stacking network with residues G57, C18 and A58. Nevertheless,

the dynamics at the elbow region did not disrupt the L-shape of tRNA and the molecule remained folded throughout all replica simulations.

Relative to the elbow region, larger motions were observed in the ASL domain over the course of all replica trajectories. Non-covalent interactions were well maintained within the stem regions across the ensemble (occupancies > 90%; Figure 3A.4), but they varied greatly within the loop regions. Hydrogen bonding in the U32–A38 and U33–A37 base pairs was reduced relative to the helical bases, with average occupancies of 50 ± 27% and 30 ± 17%, respectively. The U-turn motif was preserved in the replica ensemble as no stacking was observed between U33 and G34. Apart from the non-existent U33/U34 stack, the occupancy for other interactions within the loop varied from one replica to another. This variation is reflected in the high standard deviations for the average stacking occupancies within the loop, which ranged from 16% to 38%. In spite of these flexibilities, the hairpin-like motif of the ASL was conserved throughout all replica simulations.

In summary, although variations in the non-covalent interactions that govern tRNA folding were observed between replicas, the global structure of $tRNA^{Phe}$ was maintained across the entire replica ensemble. The ASL was the most dynamic domain of the tRNA, and the high deviations observed in non-covalent interactions across the replica ensemble indicate that the anticodon loop may have potentially adopted a wide variety of conformations over the course of the trajectories.

3.4.2 Seven unique ASL conformations were identified across the thirty 500-ns replica ensemble

To develop an MD protocol for representative sampling in tRNA studies, it is critical to describe the conformational space of interest and define the structural arrangements the molecule can adopt. To this end, thirty replica simulations were conducted and conformational analyses of the ASL were conducted on each replica trajectory using the relative positions of each nucleobase within the domain (residues 27 to 43; Figure 1). To provide a representative description of the ASL conformational space, the sampling occupancy for each isolated conformation was averaged over all thirty replicas. The nomenclature used throughout this study is based on the position and non-covalent interactions of the most dynamic nucleobase in the ASL.

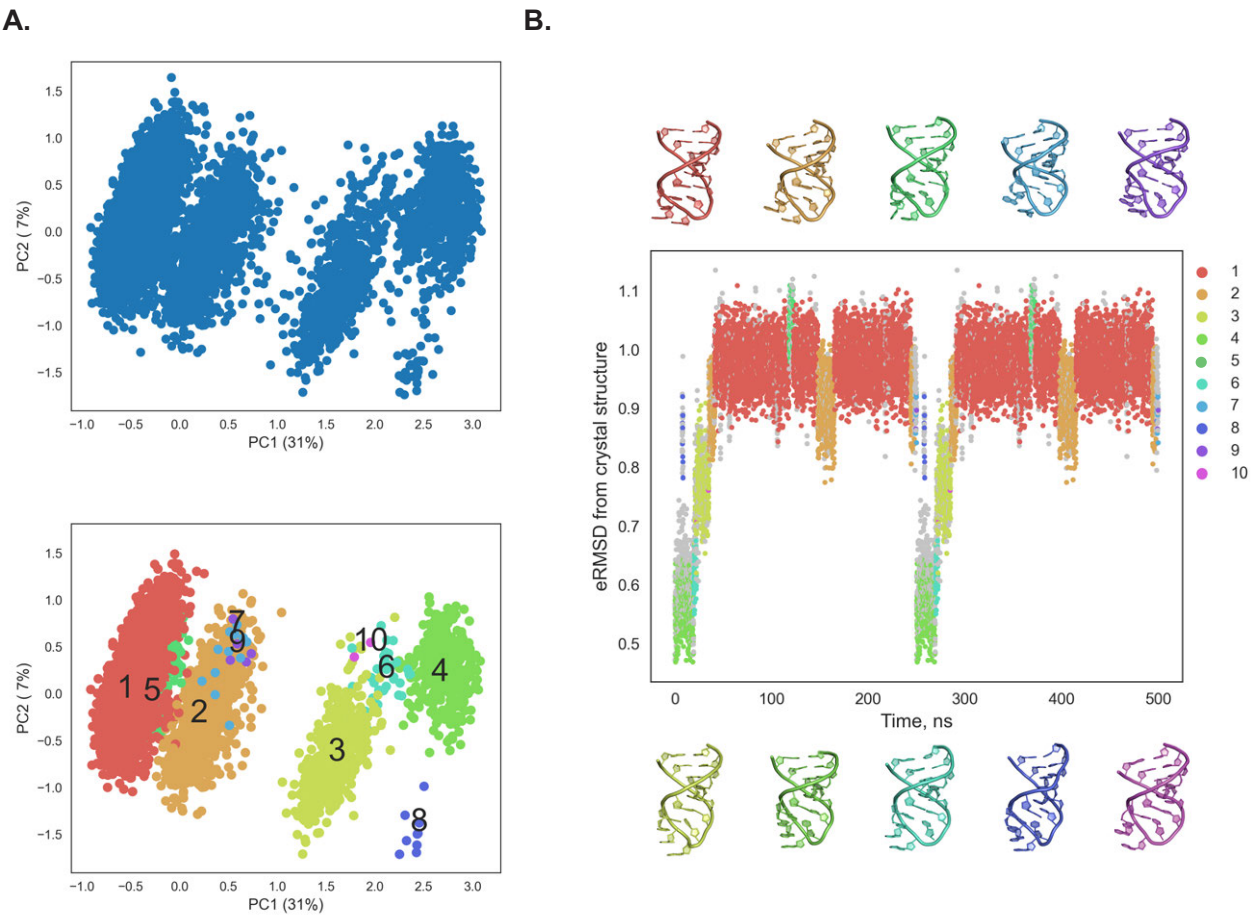


Figure 3.1 – Schematic for εRMSD clustering

(A) Projections of replica 1 from the rMD ensemble onto the first two eigenvector components calculated from εRMSD values (top) and identified clusters within the trajectory, determined using the DBSCAN algorithm (bottom). (B) εRMSD-time plot of the trajectory, color-coded by cluster. Grey points represent unassigned frames.

In the crystal structure of $tRNA^{Phe}$ provided by Byrne and colleagues,[80] the ASL adopted an organized state, in which continuous stacking was observed within the stem and loop regions of the hairpin, with the exception of U33/G34, indicating the presence of the U-turn. For the purpose of this study, the crystal structure conformation was labelled the 34-stacked conformation. In addition to this state, six other conformers were sampled across the replica trajectories, which were denoted the 34-unstacked, 35-unstacked, 36-unstacked, 33-out, 37-out and disordered conformations (Figure 3.2). Generally, these structural states can be grouped into four categories based on the position of the most displaced base (relative to the crystal structure reference) within the anticodon loop. These are: (i) the wobble base conformations (WB; 34-stacked and 34-unstacked), which describe deviations at position

34, (ii) the 3' anticodon base conformations (3'-AC; 35- and 36-unstacked) that encompass the dynamics at positions 35 and 36, (iii) the anticodon flanking base conformations (FB; 33- and 37-out), which describe fluctuations in residues 33 and 37, and (iv) the disordered loop (DL) conformation that is defined by the random, disorganized state of the loop region. Unsurprisingly, the sampled WB conformations most resemble the crystal structure reference, with the corresponding representative structures having RMSDs between 2.1 and 2.4 Å when overlayed onto the crystal structure (Table 1). On the other hand, the 3' AC and disordered conformations are less comparable to the starting structure and overlays of their representative structures had an RMSD range of 2.2 to 3.1 Å.

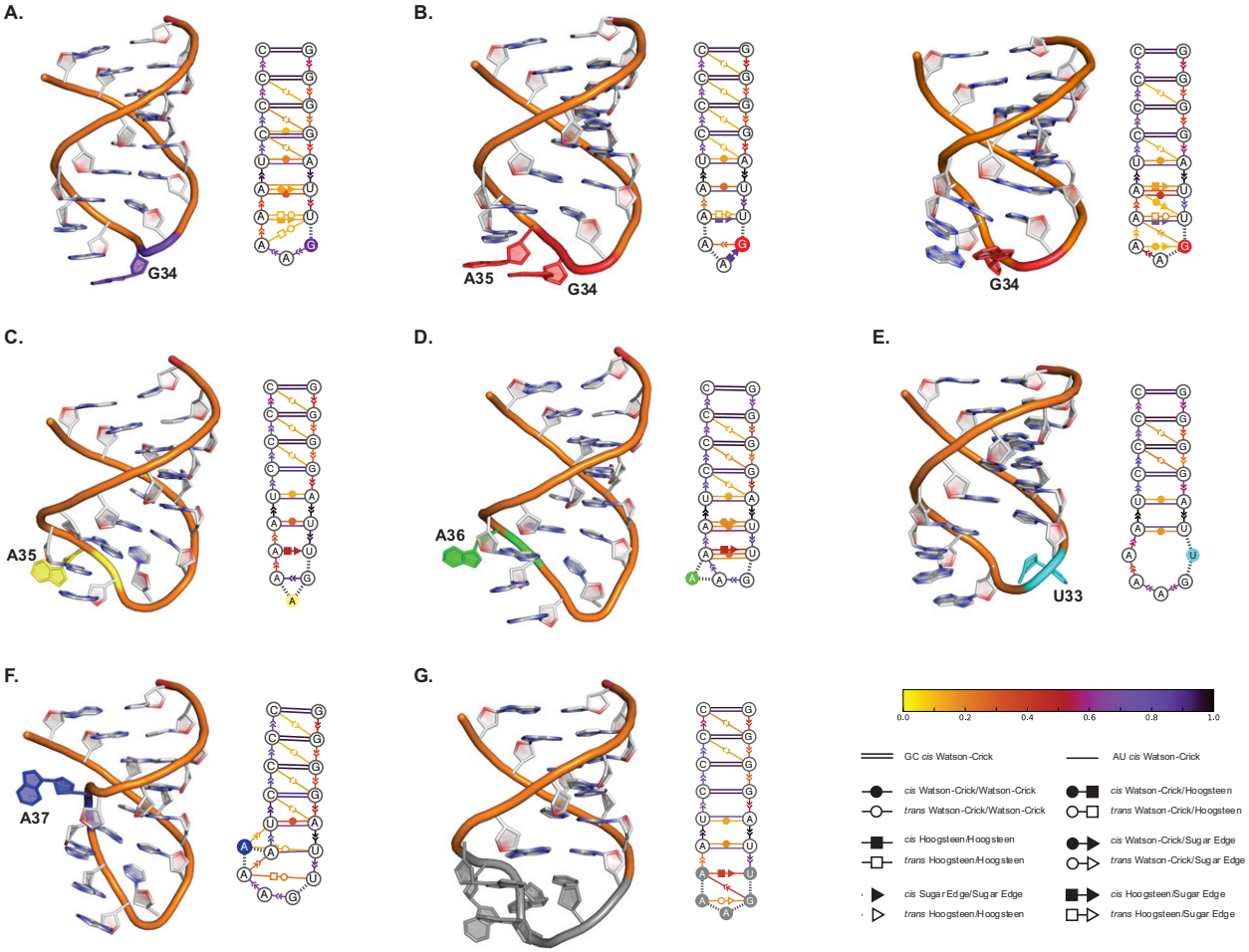


Figure 3.2 – Conformational states of the tRNA^{Phe} ASL identified using rMD

Seven structural arrangements for ASLPhe identified over the course of this study. Cartoon (left) and secondary structure (right) representations of each conformation are provided, and interactions are denoted using the Leontis–Westhof notation for non-covalent RNA interactions. From left to right, top to bottom: 34-stacked (A), 34-unstacked (B), 35-unstacked (C), 36-unstacked (D), 33-out (E), 37-out (F), disordered (G).

With the exception of the 35-unstacked conformation, all configurational states sampled by the replica simulations have been observed in experimental studies and the computed and experimental structures are highly similar (representative structure overlay RMSDs range from 1.5 to 2.8 Å; Figure 3). The 34-unstacked conformation could not be overlayed onto its experimental reference due to artifacts in the solution-state NMR experiment.[19] Nonetheless, similarities were found in the positions of residues 34 and 35, and the representative structure from the present study matched the description of the deposited NMR structure by Cabello-Villegas *et al.*, validating its existence in literature.

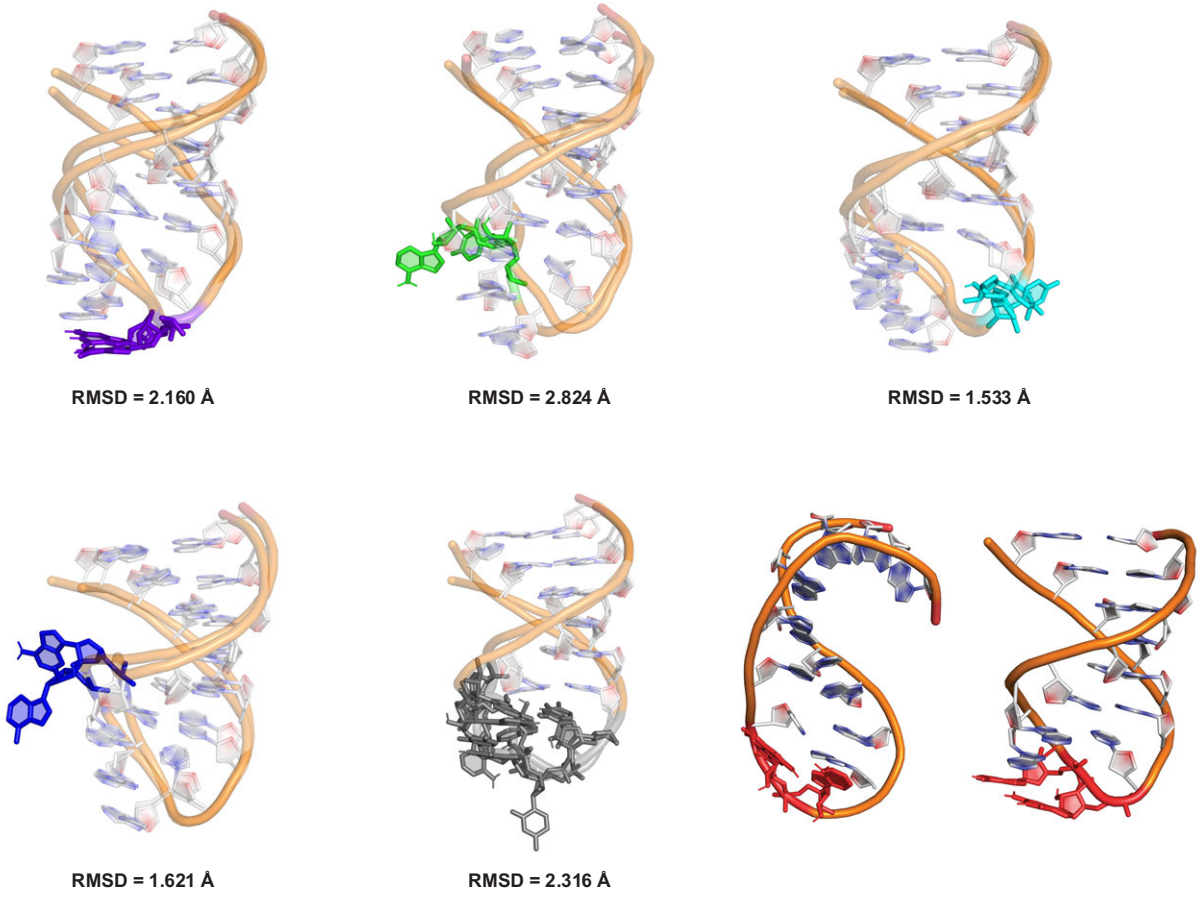


Figure 3.3 – Conformational sampling in the 30-replica ensemble

Overlays of representative structures for the seven ASL states isolated in the rMD replicas with experimentally observed ASL states. Conformation–PDB ID, from left to right, top to bottom: 34-stacked–3L0U, 36-unstancked–2KYR, 33-out–6UGG, 37-out–2FMT, disordered–1GTR, and 34-unstacked–1J4Y.

Dynamics observed within wobble base conformations of tRNA^{Phe}

Among the loop residues, G34 was observed to be the most flexible nucleobase. Per residue root-mean-square fluctuation (RMSF) analysis over all trajectories revealed high ranges of motion at this nucleotide (average RMSF ranging from 3.4 to 10.4 Å, Figure 3A.2). The pseudo-rotational profile of the anticodon loop agreed with this observation, as the backbone of G34 adopted a greater variety of η and θ pseudotorsions compared to other nucleotides in this region (Figure 3A.5). Moreover, G34 adopted both *syn* and *anti* conformations around its glycosidic bond (Figure 3A.6). Therefore, it is unsurprising that two of the dominantly sampled conformations in this study differ in the positions of nucleotide 34.

The first conformation describing dynamics at position 34 is the 34-stacked state, which is defined by the presence of stacking interactions between bases 34 and 35. Structures classified as 34-stacked had high stacking interaction occupancies (total stacking per base > 85%; Figure 3A.7) throughout the ASL but maintained the distinct U-turn motif between bases 33 and 34. Watson-Crick hydrogen bonding was also maintained at high occupancies for all base pairs within the stem region. This high maintenance of stacking and hydrogen bonding interactions in the stem region indicates that the helical alignment of the ASL was preserved. In the loop, continuous stacking interactions were observed between the anticodon bases (34, 35 and 36), with average interaction occupancies over 85%. The average loop opening for this conformation, which corresponds to the C1'–C1' distance between nucleotides 33 and 37, was 10.2 ± 1.3 Å. This distance matches that of an open loop conformation adopted by functional tRNAs in the ribosomal complex during translation.[94-97] Structures within the 34-stacked conformational group mainly differed in the position of U33 within the loop, as this base formed Watson-Crick/Hoogsteen interactions with A36 (~ 20%), Hoogsteen/Sugar Edge (~ 25%) and Watson-Crick/Sugar Edge (~ 25%) interactions with A37. This variability in the pairing partner of U33 resulted in reduced stacking interaction with U32 (~ 55%). The backbone-atom fluctuations within this conformation were slight (Figure 3A.7), except for U33 as this base moved to accommodate the base-pairing interactions with A36 and A37 discussed above. This conformational state most closely resembles the crystal structure reference (PDB ID: 3L0U) and is the most frequently identified ASL conformation in free and ribosome-bound tRNAs.[16,80,98-103]

The second conformer that describes base displacement at position 34 is the “34-unstacked” conformation, whose distinctive feature is the absence of the G34/A35 stacking interaction observed in the experimental starting structure (unmodified $tRNA^{Phe}$).[80] Once unstacked, G34 either hydrogen bonded with the Hoogsteen edge of A35 (type-I, 91%) or was exposed to solvent with the lack of non-covalent interactions with the 35-flanking base (type-II, 9%). Nevertheless, the type-I/II states occupy similar basins on the ASL phase space and were interchanged over the course of the trajectories.

In the type-I 34-unstacked conformation, the Watson-Crick/Hoogsteen edge hydrogen bonding interactions between G34 and A35 stabilized the position of the first codon base, as evidenced by a high level of planarity (21.2 ± 12.6º) and the glycosidic distance (6.4 ± 0.4 Å) between the two nucleotides. The base-pairing geometry between G34 and A35 concurred with that of other G–A pairs observed in RNA structures, which have a planarity range of 0 to 50° and C1'–C1' distance of ~ 7.1 Å.[104,105] In this state, no major fluctuations were observed in the backbone atoms of G34 and A35 and the distinctive U-motif was maintained at U33 and G34 (Figure 3A.8). However, the G34-A35 base-pairing interaction mildly disrupted the stacking interactions between A35 and A36 (~ 65%) relative to the 34-stacked conformation. Members of the type-I 34-unstacked group also had high stacking (total stacking per base > 85%) and hydrogen bonding (> 90%) occupancies within the helix, and the average loop opening was 9.5 ± 0.6 Å, similar to the crystal structure reference. Unlike the 34-stacked conformation, U33 formed long-lasting sugar edge/Hoogsteen interactions with A37 (> 90%), stabilizing the backbone atoms at position 33. The general structural features of the type-I 34-unstacked state were observed in solution NMR structures of the ASL domains from unmodified $tRNA^{Phe}$ (PDB ID: 1KKA),[19] and $tRNA^{Tyr}$ in the presence of the N^6-isopentanyladenosine (i^6A) modification at position 37 and/or pseudouridine at position 39 (PDB ID: 2LA9, 2LBQ, 2LBR).[21]

In contrast to the type-I 34-unstacked state, when G34 is exposed to solvent (type-II), the nucleotide formed short-lived sugar edge/Watson-Crick hydrogen bonds with A36 (20%) and transient hydrogen bonds with the phosphate backbone (15 interactions, lowest occupancy = 0.7%, highest occupancy = 2.5%). Members of this class still maintained the U-motif, but the backbone dihedral angles of G34 fluctuated greatly, with the η dihedral angle ranging from 130 to 350º, compared to an η range of 150 to 300º in the 34-stacked conformation (Figure 3A.9). Whereas the backbone of G34 was highly

dynamic within this group, other nucleotides remained stable, and high stacking and hydrogen-bond occupancies were observed across the ASL. U33 always paired with A37, but base stacking with A37 (A36/A37 and A37/38) was reduced to ~ 80% due to steric clashes between A37 and the flexible G34. All the interactions that define this conformation were recently observed experimentally in a crystallographic study on unmodified $tRNA^{Val}$ from *E. coli* (PDB ID: 7EQJ, unreleased).[22] This conformation was also sampled in $tRNA^{Gly}$ isoacceptor molecules during a previous MD study that investigated the structural recognition elements found in unmodified $tRNA^{Gly}$, $tRNA^{Arg}$, $tRNA^{Ala}$ and $tRNA^{Val}$.[106]

Nucleotide fluctuations at the 3' anticodon bases

Besides G34, the second (A35) and third (A36) anticodon nucleotides were also displaced over the course of the trajectories, resulting in the 35-unstacked and 36-unstacked conformations respectively. In both cases, the bases (A35 or A36) moved from their original position relative to the crystal structure reference, disrupting stacking interactions with their flanking bases. Fluctuations in A35 or A36 did not result in complete disarray at the anticodon loop as stable stacking interactions were formed between G34 and A36 (97%) or A35 and A37 (99%), respectively. Moreover, the integrity of the ASL helix was preserved as the stacking and base-pairing interactions were maintained within the stem region of both the 35-unstacked and 36-unstacked conformations (interaction occupancies over 80%). Although the loop openings for both conformational states were comparable (9.6 ± 1.2 Å and 10.5 ± 0.4 Å respectively), the movement of A35 and A36 had different effects on the ASL. Specifically, movement of A35 increased the backbone flexibility at G34 ($120° < \theta < 360°$, $60° < \eta < 180°$ and $240° < \eta < 360°$) and A36 ($60° < \eta < 240°$), leading to a wider anticodon loop (9.6 ± 1.2 Å) relative to the 34-stacked conformation (Figure 3A.10). On the other hand, movement of A36 restricted backbone movement at G34 ($160° < \theta < 250°$, $0° < \eta < 120°$), resulting in a more angled U-turn relative to the 34-stacked conformation (Figure 3A.11). Although the 35-unstacked conformation has yet to be reported experimentally, the 36-unstacked conformation was detected in a solution NMR structure of human mitochondrial $tRNA^{fMet}$ (PDB ID: 2KRY, 2KRZ)[17] and was sampled in an MD investigation on the $tRNA^{Gly}_{ACC}$ isoacceptor.[106]

Dynamics of anticodon flanking bases in tRNA^{Phe}

Fluctuations at the flanking bases of the anticodon (U33 and A37) were observed and classified as two conformational states – 33-out and 37-out, respectively. The distinct feature of these groups is the flipping of the nominal base out of the anticodon loop. When U33 flips out of the loop, the ASL lost its distinguishing U-motif and adopted a broad range of pseudorotational angles ($0° < \theta < 360°$, $0° < \eta < 60°$ and $120° < \eta < 360°$), resulting in the formation of an internal loop-like structure (Figure 3A.12). On the other hand, movement of A37 out of the anticodon loop more extensively restricted the backbone flexibility at G34 than the 36-unstacked conformation, leading to a more pronounced U-turn between U33 and G34 (Figure 3A.13). With either U33 or A37 out of the loop, the loop opening was wider relative to WB and 3'-AC conformations (14.7 ± 2.4 Å and 15.3 ± 1.7 Å, respectively), but this did not cause the stem to unravel, with stacking and hydrogen bonding within the stem region being maintained (occupancies > 80%). Interestingly, A37 was more flexible than U33 once flipped out of the loop, as evident by the wide range of glycosidic torsions ($0° < \chi < 360°$) adopted by A37 sampled within the 37-out conformer. Furthermore, whereas U33 formed no long-lasting interactions once flipped out, A37 formed hydrogen-bond interactions with U32 (~ 25%) and stacks with U40 (~ 15%; Fig A14). In spite of sequence differences, the structural features of the 37-out conformation were observed in tRNA^{fMet} and were suggested to be a discriminatory element for the initiator tRNA during translation (PDB ID: 2FMT).[90] Similarly, the 33-out conformation was isolated in a study of unmodified *E. coli* tRNA^{Asp}, which reiterated the intrinsic flexibility of the ASL domain (PDB ID: 6UGG).[103]

Disorder within the anticodon loop of tRNA^{Phe}

The last conformation sampled is the disordered conformation, defined by a complete lack of interactions observed within the crystal structure reference at the anticodon loop (U32 to A38). Although the centroids for this group describe a variety of interactions, all members have lost local nucleotide arrangements within the loop present in the 34-stacked conformation. Specifically, none of the continuous stacking interactions present at the anticodon loop in the crystal structure of unmodified tRNA$^{Phe\ 80}$ were well-preserved (occupancies ranged from 0 to 15%; Figure 3A.15), and the loop lost its distinct U-turn motif. Furthermore, the anticodon loop adopted a wider opening (16.5 ± 1.7 Å) relative to all other

configurational states. The backbone atoms of the disordered conformation were also highly dynamic, and fluctuations were especially unrestrained at A35 (60° < θ < 240°), A36 (60° < η < 330°), A37 (0° < θ < 60° and 240° < θ < 360°, 0° < η < 360°) and A38 (120° < η < 300°). Although the continuous stacking of anticodon loop bases observed in the experimental starting structure is absent in this conformational group, other non-covalent interactions were observed. For instance, G34 stacked with A37 (< 10%), while a stacking interaction was observed between U33 and A35 (~ 80%). Watson-Crick hydrogen-bond interactions were also observed between U33 and A36, as A37 sometimes interacted with the phosphate backbone of G27 (~ 35%). Many NMR studies have identified similar disordered ASL states for different tRNA molecules.[18,19,21,26,28] Additionally, the disordered ASL conformation has been observed in tRNA molecules complexed with aminoacyl synthetases, where interactions with anticodon bases are critical for the accurate amino acid charging at the 3' of the tRNA molecule.[107-109]

Torsional deviations of ASL conformations identified in this study

To further validate the conformational structures identified with BARNABA, a dPCA was performed on the entire trajectory ensemble, using all backbone dihedral angles of the ASL. In the literature, six dihedral angles have been used to describe the backbone conformations of RNA molecules (α, β, γ, δ, ε, and ζ) and a wide range of tertiary structures are attained through the fine tuning of these torsional angles.[110-112] Moreover, the glycosidic torsion angle (χ) of nucleotides was considered, as it accounts for the location of nucleobases with respect to the backbone. Variations in the backbone and glycosidic torsions of the nucleotides within the anticodon loop are summarized in Figures A5 and A6, respectively, while Figure 4 displays the two-dimensional representation of the dihedral angles projected onto the subspace of the first and second principal component vectors (PC1 and PC2). In this analysis, PC1 and PC2 describe 1.2% and 0.7%, respectively, of the total variance of the motions in the ensemble. Correlations were observed between the dPCA and εRMSD-based conformational sampling, as in general, structurally different states occupy different wells on the dihedral-based free energy landscape of the domain. This suggests that structural deviations observed in the conformational sampling were mostly dictated by changes in the backbone of the anticodon loop, relative to the crystal structure reference.

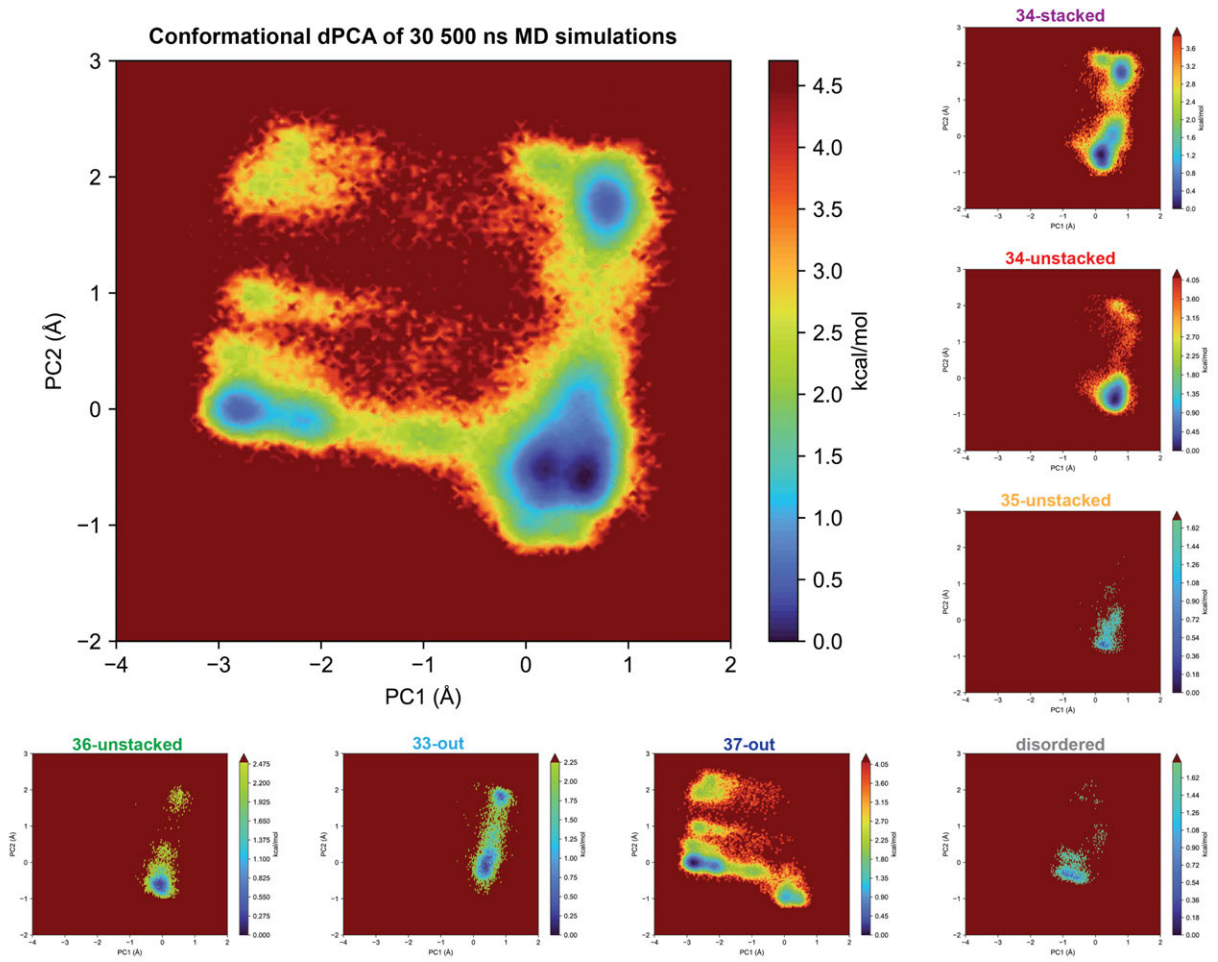


Figure 3.4 – Conformational space sampled across replica ensemble

Dihedral principal component analysis on the backbone dihedrals of the ASL sampled across the 30-replica ensemble.

Three densely populated local minima were identified on the free energy landscape for the replica trajectories, which correspond to the 34-stacked, 34-unstacked and 37-out conformations (Figure 3.4). The local minimum for the 34-stacked state overlapped that of the 34-unstacked conformation, indicating that the backbone dihedral angles of these conformations are very similar. Indeed, overlays of the representative structures for these conformations results in an RMSD of 0.98 Å and, except for the nitrogenous bases at positions 34 and 35, the coordinates of all bases within the domain are the same for both conformations. In contrast, the local minimum for the 37-out conformation was furthest from all other conformational states, which suggests that this conformation has a highly altered backbone relative to other ASL conformations. This hypothesis was confirmed via overlays of the representative structures for

37-out and 34-stacked (RMSD = 1.30 Å), which revealed a kink in the backbone at position 37 that results in a narrower loop region for the 37-out conformation. Contrary to the previous conformations, the 35-unstacked, 36-unstacked, 33-out, and disorder conformations occupied higher energy regions on the ASL potential energy surface, indicating that the backbone configurations for these states are less stable compared to the 34-stacked, 34-unstacked and 37-out states. Thus, the dPCA explains how reversible conformational changes are at the anticodon loop by describing the torsional dynamics at this region and outlining the similarities and differences between backbone torsions of each conformational group.

In conclusion, seven unique ASL conformations were identified across the thirty 500-ns replica ensemble. Six of these conformations have observed in experimental structural data for a variety of tRNAs and tRNA environments. Nevertheless, the question still remains as to whether a reduced replica set can reproducibly sample the same range of conformations.

3.4.3 Single 500-ns simulations do not adequately sample the tRNA phase space

Although seven ASL conformational states were isolated using the 30 replica trajectories, varied conformational sampling was observed within each simulation (Figure 3A.16). Specifically, a maximum of four ASL states were sampled in replicas 6, 7, 8, 10, 11, 23, and 25, with each trajectory isolating diverse conformational combinations. For instance, while replica 6 sampled the 34-stacked, 34-unstacked, 35-unstacked, and disordered conformations, replica 7 identified the 34-stacked, 34-unstacked, 35-unstacked, and 33-out conformations. In contrast, six replicas sampled only one conformation (34-unstacked in replica 28; 37-out in replicas 2, 5, 14, 17 and 30), indicating that each trajectory may have been trapped in a local minimum in phase space. Furthermore, replicas that isolated the same conformational states sampled the conformations with different frequencies. For instance, replicas 1 and 19 both sampled the 34-stacked and 34-unstacked conformations, but replica 1 had a sampling ratio of 4:1, while replica 19 exhibited an occupancy ratio of 2:3. Therefore, no two replicas had identical conformational profiles, which highlights the necessity for replication in tRNA MD studies to achieve a comprehensive and representative description of the ASL conformational space.

Of the seven conformations identified in this study, three conformations were predominantly sampled across the replica ensemble (Figure 3.5A) – the 34-stacked (38%), 34-unstacked (34%) and the

37-out (20%). Interestingly, positions 34 and 37 are the most posttranscriptionally modified regions of the ASL across all phylogenetic domains, highlighting their functional importance *in vivo*.[23,113-116] All other conformations were sampled > 5% of the entire trajectory time, with least prominent conformations being the 35-unstacked and disorder groups that were sampled ~ 1% of the time.

In summary, the rMD protocol surveyed the expanse of the conformational space of the ASL and conformational analyses across the replica trajectories revealed that different structural arrangements with varying occupancies were sampled within each simulation in the ensemble. These analyses also showed that fluctuations at positions 34 and 37 are responsible for most of the dynamics observed within the tRNA ASL, as together their ASL states occupy ~ 90% of the conformational profile uncovered by replica simulations. More importantly, analysis of this large set of replica simulations provides a basis for the evaluation of smaller rMD protocols in representative sampling of the conformational space of tRNA.

3.4.4 At least 10 500-ns replicas are required to accurately sample dominant conformations of the ASL

Although the 30-simulation replica ensemble successfully sampled known and undiscovered ASL conformations, performing numerous replicas is computationally expensive. Therefore, it is important to determine the minimum number of replicas required for representative sampling of the tRNA phase space. For this investigation, the sampling performance of randomly grouped replica ensembles composed of 3, 5, 10 and 15 replicas were assessed. The overall sampling of each ensemble was evaluated with respect to the sampling across all thirty replicas, as well as the reproducibility of the conformational sampling among different ensembles of equal sizes. Recall that the dominantly sampled states were the 34-stacked, 34-unstacked and 37-out conformations (Figure 3.5A). All other isolated conformations (35-unstacked, 36-unstacked, and 33-out) were classified as minor conformations.

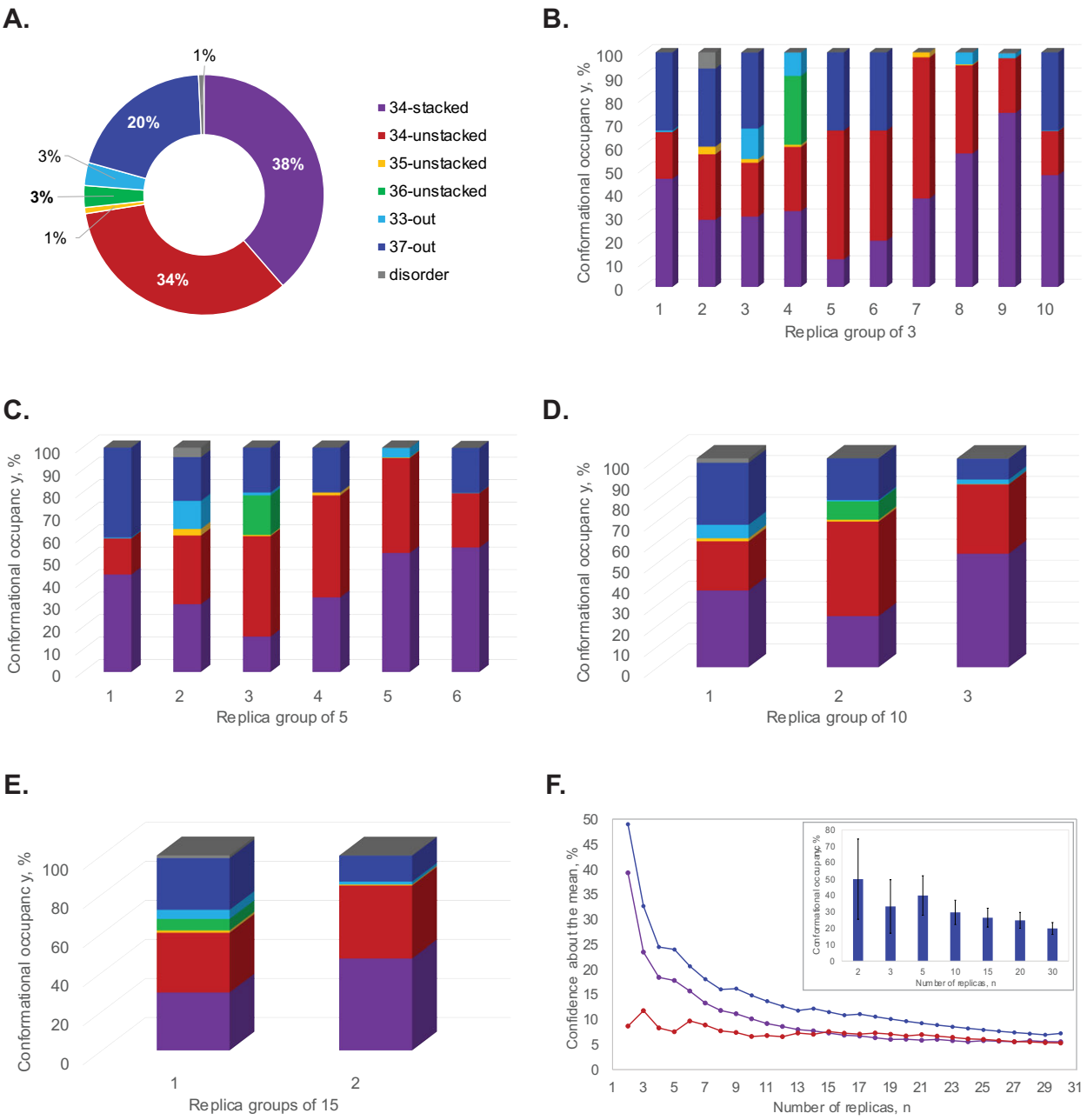


Figure 3.5 – Statistical analyses of replica simulations

(A) Average conformational occupancies across the 30-replica ensemble. (B–E) Variance in conformational sampling across replica groups of 3, 5, 10 and 15 simulations. Conformations are color-coded using the legend defined in plot A. (F) Two-sided confidence intervals ($P < 0.05$) about the population mean occupancies of the three dominant conformations sampled in the rMD simulations. The inner panel describes the fluctuations in uncertainty of the conformational occupancy for the 37-out conformation with respect to the number of samples.

Datasets composed of 3 and 5 simulations failed to sample dominant conformations in the ensemble of thirty simulations. Among the ten datasets comprised of 3 trajectories (Figure 3.5B), high

variance was observed in the sampling occupancies of the dominant conformations (standard deviations of 18%, 15% and 17% for 34-stacked, 34-unstacked and 37-out, respectively). Furthermore, four out of the ten datasets (sets 4, 7, 8 and 9) failed to sample the dominant 37-out conformation, and minor conformations like 33-out and 36-unstacked are highly represented in certain datasets (> 10% occupancy; sets 3 and 4), distorting the reality of the conformational space. Replica groups of 5 simulations better described the phase space and only one out of the five groups failed to sample the 37-out conformation (Figure 3.5C). However, minor conformations were still overrepresented within groups of 5 simulations (> 10% occupancy; 33-out in set 2 and 36-unstacked in set 3), indicating the need for more replicas. Datasets composed of 10 or 15 simulations accurately represented the ASL conformational space (Figure 3.5D and E), as variations between members of these replica groups were reduced relative to sampling datasets composed of 3 and 5 replicas (standard deviation ranges of 15% and 12% for the 34-stacked, 11% and 9% for 34-unstacked, 10% and 9% for the 37-out across datasets of 10 and 15 simulations, respectively). Moreover, dominant conformations were accurately identified across each ten- and fifteen-replica ensemble evaluated, and the minor conformations occupied > 10% of the conformational profiles of each group. These observations concur with the conformational profile of the thirty-replica ensemble, suggesting that a minimum of 10 replicas are required to representatively sample the configurational space of the ASL.

To further confirm that a minimum of 10 simulations are required for representative sampling of the ASL phase space, confidence intervals about the population mean were evaluated on the conformational occupancies for the three dominant conformations (34-stacked, 34-unstacked and 37-out) over all ensemble trajectories (Figure 5F). In general, a sharp decrease in the confidence interval was observed as we progressed from 1 to 10 replicas, with confidence interval dropping from 40 to 10%, 11 to 7% and 50 to 16% for 34-stacked, 34-unstacked and 37-out conformations, respectively. Moreover, the subsequent addition of replicas only provided small improvements in the confidence interval, with the intervals across all 30 replicas being 6%, 5% and 7% for 34-stacked, 34-unstacked and 37-out conformations, respectively. Therefore, this analysis confirms that a replica ensemble of at least ten 500-ns simulations is required to representatively sample the configuration space of the ASL and attain conformational convergence.

3.4.5 A single, long MD simulation does not adequately sample tRNA phase space

Thus far, this study has confirmed that ten 500-ns simulations can accurately identify dominant and minor ASL states. Nevertheless, cMD simulations have predominantly been used in the literature to study tRNA dynamics.[1-11] Therefore, it is important to evaluate the ability of an equivalent cMD trajectory to sample the configurational space of the ASL, to determine the best MD protocol for future tRNA investigations. Therefore, a 5 μs cMD simulation was run and analyzed in the same way as the independent replica trajectories. The RMSDs was relatively stable throughout this trajectory, being 4.6 ± 0.6 Å (Figure 3A.2). Like the replica trajectories, residues in the loop regions of the tRNA D arm, ASL and variable region were the most flexible (RMSFs ranging from 3 to 10 Å). To visualize tRNA dynamics within the cMD trajectory, representative structures were extracted at 500-ns intervals, and overlayed with respect to the crystal structure of $tRNA^{Phe}$ (PDB: 3L0U;[80] Figure 3A.2). In agreement with the RMSF analysis, the structural overlays revealed the elbow region and ASL domain were the most dynamic areas of the tRNA. As with the trajectory ensemble, the dynamics at the elbow region did not disrupt the non-covalent interactions that govern tRNA folding. Specifically, tertiary hydrogen bonding and stacking interactions between the nucleobases within the D and TψC loop were highly preserved over the course of the trajectory (hydrogen bonding occupancies > 95% and stacking occupancies > 80%; Figure 3A.3).

Relative to the elbow region, more varied motions were observed in the ASL in the cMD simulation. Non-covalent interactions were preserved within the stem regions (interaction occupancies > 90%), but more variations were observed within the loop (Figure 3A.4). The U32–A38 and U33–A37 base pairs had reduced occupancies (approximately 80% and 20%, respectively) relative to their helical counterparts. Furthermore, base stacking in the stem extended to the anticodon loop but varied among interacting nucleobases. No stacking was observed between U33 and G34, denoting the presence of the U-turn motif at this location. Low stacking was also observed between G34 and A35 (< 10% occupancy), due to the dynamic nature of the first codon base (G34). Recall that residue 34 is the most solvent exposed base within the anticodon loop, and its position is only stabilized through interactions with one flanking base (A35). Consequently, the nucleobase at position 34 has more degrees of freedom relative to other bases in the region, which remain stacked with their flanking bases over 70% of the time. In summary, the

dynamics of tRNA observed in the cMD trajectory revealed that the global fold and overall structure of the tRNA were well maintained throughout the trajectory, which concurs with the observations made in the replica ensemble.

Conformational analysis of the ASL was run on the cMD simulation using the relative positions of every nucleotide within the domain (residues 27 to 43). To evaluate sampling of the ASL phase space at different simulation lengths, the cMD trajectory was analyzed in segments using its first 500 ns, 1.5, 2.5, 3.5 and 5 μs, and conformational sampling within each segment was assessed using the εRMSD-based clustering analysis (Figure 3.6). In the first 500 ns, the 34-stacked and 34-unstacked conformations were sampled interchangeably with occupancies of 20% and 80%, respectively.

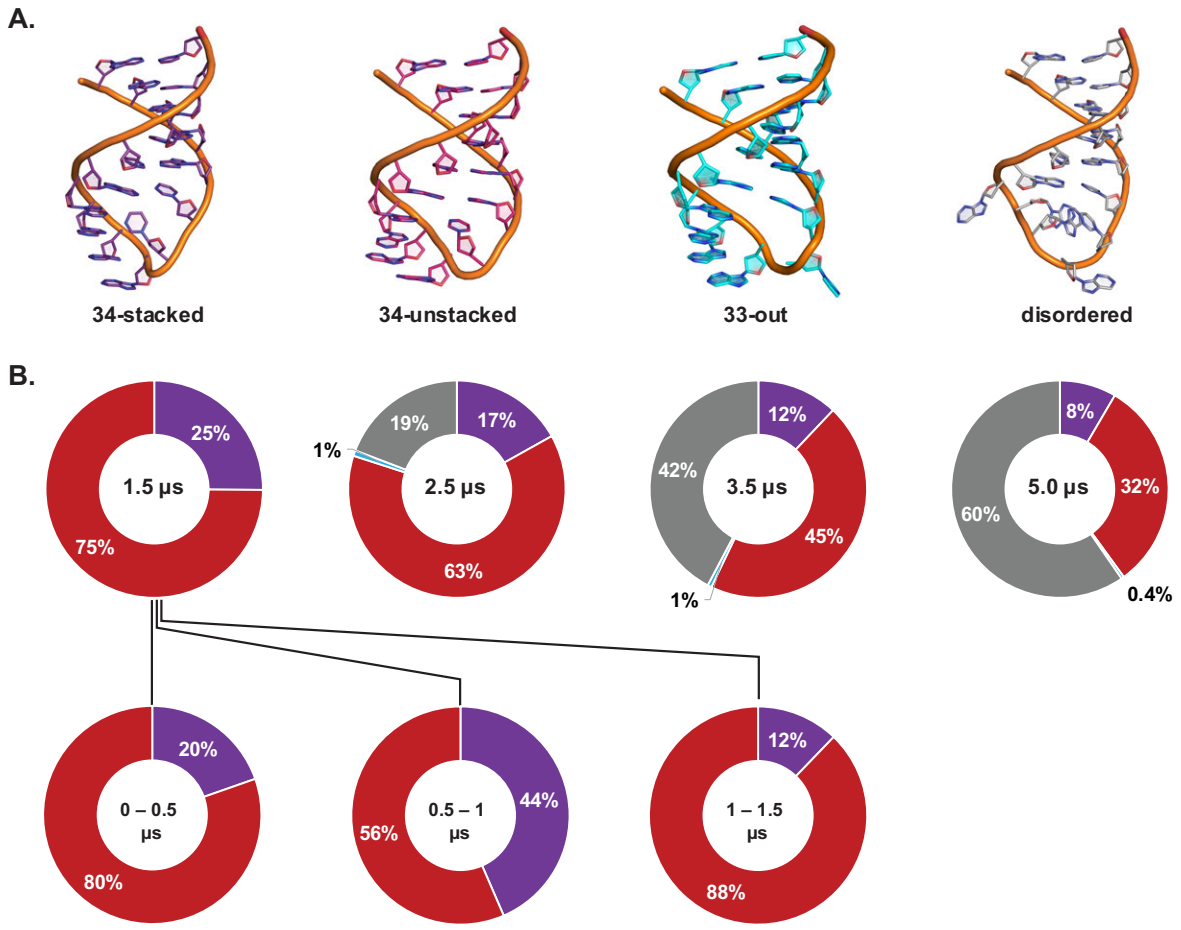


Figure 3.6 – Detailed analyses of the cMD trajectory

(A) Representative structures for all conformations sampled during the cMD trajectory in order of first appearance. (B) Conformational profiles at various time intervals of the 5 μs trajectory. The conformations are color-coded using the code defined in plot A.

Subsequently, varied sampling of the 34-stacked and 34-unstacked conformations was observed until the 2 μs mark (occupancies of 17% and 63%, respectively). At ~ 2 μs, the 33-out conformation was briefly sampled (1%), and the disordered conformation was subsequently sampled for 19% of the trajectory. After 2.5 μs, no other conformation was sampled, with only the disordered conformation being adopted for the remainder of the trajectory, indicating that the trajectory may have been trapped in a local minimum. At the end of the 5 μs trajectory, four ASL conformations were isolated – the 34-stacked (8%), 34-unstacked (32%), 33-out (0.4%) and disordered (60%) conformations.

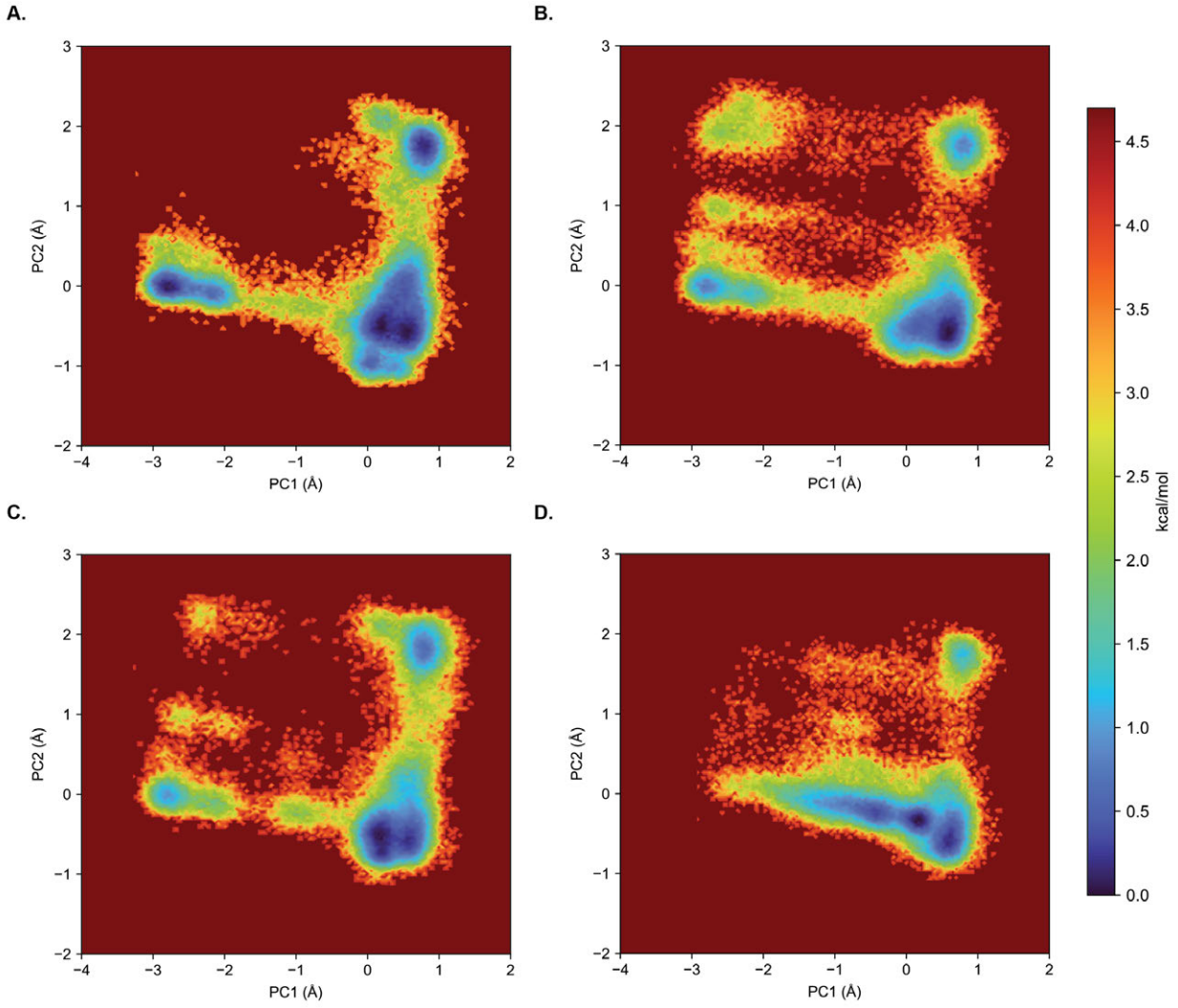


Figure 3.7 – Sampling performance of the cMD simulation relative to 10-replica ensembles

Principal component analysis of the backbone dihedrals of the ASL from the three 10-replica ensembles – replicas 1 to 10 (A), replicas 11 to 20 (B), replicas 21 to 30 (C) – and the cMD simulation. All trajectories were projected onto the PC space of the 30-replica ensemble.

The conformational profile for this simulation revealed an underrepresentation of the highly prevalent 34-stacked conformation and an overrepresentation of the disordered conformation which is less frequently observed in the literature. Nevertheless, the computed and experimental structures of identified conformations were well-matched, with low RMSDs ranging from 1.5 to 2.4 Å (Figure 3A.17). The structural representations for each conformation isolated in the cMD simulation also matched those from the replica ensemble, with RMSDs ranging from 0.8 to ~ 1.0 Å.

In addition to the conformational analysis, the configurational space sampled by the cMD simulation relative to the replica ensemble was assessed using dPCA. To do this, the 5 μs simulation was projected onto the subspace defined by the thirty-replica ensemble (Figure 3.7). For equivalent comparisons, the three sampling datasets composed of ten replicas were also projected on the subspace defined by the thirty-replica ensemble to assess their performance in conformational sampling of the ASL. The dPCA revealed that the 5 μs simulation failed to explore parts of the configurational space covered by all replica ensembles considered.

In conclusion, the conformational and torsional analyses of the 5 μs trajectory show that a single, long-time scale simulation does not provide a comprehensive overview of the ASL phase space. Consequently, the present study concludes that the singular cMD simulation inadequately samples tRNA dynamics, and at least ten 500-ns simulations are required to identify the major modes of motion for this molecule.

3.5 Conclusions

Molecular dynamics simulations have been widely employed to study tRNA structure and dynamics, with two protocols (rMD and cMD) having been historically applied in this area.[1,94,117,118] Nevertheless, when it comes to an accurate and reproducible method, no consensus exists in the literature, and computational methodologies vary from one study to another. To evaluate the relative sampling performance of short, replicate simulations and long, single trajectories, the present study compares the conformations sampled using multiple ns-timescale replica simulations to those attained using a single μs-timescale simulation. When taken together, thirty 500-ns replicas were found to identify all ASL states previously isolated in experimental studies. Nevertheless, because running numerous

replicas is a computationally demanding endeavor, the minimum number of replicas required to representatively sample the configurational space of the ASL was also evaluated. Ten 500-ns simulations (total simulation time of 5 μs) were found to be proficient at accurately describing the major and minor conformations identified in the larger replica ensemble. To assess whether the cMD protocol provides similar sampling of the phase space, a single 5 μs simulation was conducted on the tRNA molecule. Analyses of this trajectory revealed a long standalone simulation is unable to describe the configurational space of the anticodon loop region. To conclude, this study recommends the use of at least ten 500-ns simulations to investigate tRNA systems in the future. The rMD protocol and analyses recommended in this study can also be applied to RNA structures with similar global folds – for instance, tRNA-like structures,[119,120] or to investigate small hairpin domains similar to the ASL. For larger molecules (> ~ 100 nucleotides) and molecules with less stable tertiary arrangements, the length and number of MD simulations may need to be reevaluated as sampling convergence has been shown to differ from one biomolecule to another.[56,76,121-123]

3.5.1 Future directions

The protocol recommended in this study models full tRNA to investigate the structural dynamics, particularly in the ASL. However, due to the abundance of posttranscriptional modifications within the anticodon stem-loop and the domain's functional role in protein synthesis,[23,24,114,124-126] some studies have employed a truncated tRNA model (ASL domain only) to study the structural dynamics of the region. Yet, to this day, no study has investigated conformational sampling in the truncated tRNA model. Furthermore, as with the full-length tRNA model, the simulation length and replicate number employed in these studies vary.[4,5,21,26,127-129] Consequently, future work could expand on the protocol proposed in the present study by investigating the conformational space surveyed using an ASL model. Similar analyses to those performed in this thesis will deliver insight on the conformational states the anticodon loop adopts in the absence of other tRNA domains, and potentially provide a more efficient way of studying tRNA modifications at the ASL, as well as a basis for future MD investigations on the RNA hairpin motif in general.

In addition to hairpins, RNA molecules can adopt a wide range of secondary structure arrangements (motifs).[130-132] For instance, packaging RNA of the phi29 bacteriophage was observed to adopt the open and stacked conformations of three-way junctions,[133] while kissing loops, 90°-kink and four-way junctions were identified in viral ribosomal RNAs.[134,135] Furthermore, in addition to their cellular significance, synthetic biologists have recently employed various RNA motifs to develop RNA nanostructures that could be used to control molecular functions by acting as carriers for functional molecules and scaffolds for molecular processes like apoptosis.[136-139] Despite their relevance, the convergence of simulation protocols on most RNA motifs is unclear and the structural dynamics of most RNA motifs remain unknown. Consequently, future work could develop protocols for the accurate and efficient conformational sampling of other RNA motifs. Understanding the structural states adopted by other RNA motifs will provide insight into their cellular functions and expand their application in biotechnology endeavors. Beyond RNA motifs, the convergence of simulation protocols for other small RNAs (e.g., aptamers, riboswitches) could also be evaluated to determine if, generally, short, replicated simulations are sufficient to describe their conformational space.

3.6 References

1 Alexander, R. W., Eargle, J. & Luthey-Schulten, Z. Experimental and computational determination of tRNA dynamics. *FEBS Letters* **584**, 376-386, doi:https://doi.org/10.1016/j.febslet.2009.11.061 (2010).

2 Auffinger, P., Louise-May, S. & Westhof, E. Molecular Dynamics Simulations of Solvated Yeast tRNAAsp. *Biophysical Journal* **76**, 50-64, doi:https://doi.org/10.1016/S0006-3495(99)77177-8 (1999).

3 Auffinger, P. & Westhof, E. in *Modification and Editing of RNA* 103-112 (1998).

4 Auffinger, P. & Westhof, E. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J Mol Biol* **269**, 326-341, doi:10.1006/jmbi.1997.1022 (1997).

5 Auffinger, P. & Westhof, E. H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophysical Journal* **71**, 940-954, doi:https://doi.org/10.1016/S0006-3495(96)79298-6 (1996).

6 Black Pyrkosz, A., Eargle, J., Sethi, A. & Luthey-Schulten, Z. Exit Strategies for Charged tRNA from GluRS. *Journal of Molecular Biology* **397**, 1350-1371, doi:https://doi.org/10.1016/j.jmb.2010.02.003 (2010).

7 Eargle, J., Black, A. A., Sethi, A., Trabuco, L. G. & Luthey-Schulten, Z. Dynamics of Recognition between tRNA and Elongation Factor Tu. *Journal of Molecular Biology* **377**, 1382-1405, doi:https://doi.org/10.1016/j.jmb.2008.01.073 (2008).

8 Roh, J. H. *et al.* Dynamics of tRNA at Different Levels of Hydration. *Biophysical Journal* **96**, 2755-2762, doi:https://doi.org/10.1016/j.bpj.2008.12.3895 (2009).

9 Seelam Prabhakar, P., Takyi, N. A. & Wetmore, S. D. Posttranscriptional modifications at the 37th position in the anticodon stem–loop of tRNA: structural insights from MD simulations. *RNA* **27**, 202-220, doi:10.1261/rna.078097.120 (2021).

10 Srivastava, A., Yesudhas, D., Ramakrishnan, C., Ahmad, S. & Gromiha, M. M. Role of disordered regions in transferring tyrosine to its cognate tRNA. *International Journal of Biological Macromolecules* **150**, 705-713, doi:https://doi.org/10.1016/j.ijbiomac.2020.02.070 (2020).

11 Zhang, X., Walker, R. C., Phizicky, E. M. & Mathews, D. H. Influence of Sequence and Covalent Modifications on Yeast tRNA Dynamics. *Journal of Chemical Theory and Computation* **10**, 3473-3483, doi:10.1021/ct500107y (2014).

12 Prabhakar, P. S., Takyi, N. A. & Wetmore, S. D. Posttranscriptional modifications at the 37th position in the anticodon stem-loop of tRNA: structural insights from MD simulations. *RNA* **27**, 202-220, doi:10.1261/rna.078097.120 (2021).

13 Šponer, J. *et al.* RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chemical Reviews* **118**, 4177-4338, doi:https://10.1021/acs.chemrev.7b00427 (2018).

14 Murphy, F. V. t., Ramakrishnan, V., Malkiewicz, A. & Agris, P. F. The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat Struct Mol Biol* **11**, 1186-1191, doi:10.1038/nsmb861 (2004).

15 Auffinger, P. & Westhof, E. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNAAsp anticodon hairpin11Edited by J. Karn. *Journal of Molecular Biology* **269**, 326-341, doi:https://doi.org/10.1006/jmbi.1997.1022 (1997).

16 Bénas, P. *et al.* The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA* **6**, 1347-1355, doi:10.1017/s1355838200000911 (2000).

17 Bilbille, Y. *et al.* The human mitochondrial tRNAMet: structure/function relationship of a unique modification in the decoding of unconventional codons. *Journal of molecular biology* **406**, 257-274, doi:10.1016/j.jmb.2010.11.042 (2011).

18 Cabello-Villegas, J. & Nikonowicz, E. P. Solution structure of ψ32-modified anticodon stem–loop of Escherichia coli tRNAPhe. *Nucleic Acids Res* **33**, 6961-6971, doi:10.1093/nar/gki1004 (2005).

19 Cabello-Villegas, J., Winkler, M. E. & Nikonowicz, E. P. Solution Conformations of Unmodified and A37N6-dimethylallyl Modified Anticodon Stem-loops of Escherichia coli tRNAPhe. *Journal of Molecular Biology* **319**, 1015-1034, doi:https://doi.org/10.1016/S0022-2836(02)00382-0 (2002).

20 Caulfield, T. R., Devkota, B. & Rollins, G. C. Examinations of tRNA Range of Motion Using Simulations of Cryo-EM Microscopy and X-Ray Data. *Journal of Biophysics* **2011**, 219515, doi:10.1155/2011/219515 (2011).
21 Denmon, A. P., Wang, J. & Nikonowicz, E. P. Conformation effects of base modification on the anticodon stem-loop of Bacillus subtilis tRNA(Tyr). *J Mol Biol* **412**, 285-303, doi:10.1016/j.jmb.2011.07.010 (2011).
22 Jeong, H. & Kim, J. Unique anticodon loop conformation with the flipped-out wobble nucleotide in the crystal structure of unbound tRNAVal. *RNA* **27**, 1330-1338, doi:10.1261/rna.078863.121 (2021).
23 Agris, P. F. Decoding the genome: a modified view. *Nucleic Acids Res* **32**, 223-238, doi:10.1093/nar/gkh185 (2004).
24 Allnér, O. & Nilsson, L. Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome. *RNA (New York, N.Y.)* **17**, 2177-2188, doi:10.1261/rna.029231.111 (2011).
25 Duechler, M., Leszczyńska, G., Sochacka, E. & Nawrot, B. Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cell Mol Life Sci* **73**, 3075-3095, doi:10.1007/s00018-016-2217-y (2016).
26 Durant, P. C. & Davis, D. R. Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine11Edited by I. Tinoco. *Journal of Molecular Biology* **285**, 115-131, doi:https://doi.org/10.1006/jmbi.1998.2297 (1999).
27 Kothe, U. & Rodnina, M. V. Codon Reading by tRNAAla with Modified Uridine in the Wobble Position. *Molecular Cell* **25**, 167-174, doi:https://doi.org/10.1016/j.molcel.2006.11.014 (2007).
28 Lescrinier, E. *et al.* The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res* **34**, 2878-2886, doi:10.1093/nar/gkl081 (2006).
29 Morris, R. C., Brown, K. G. & Elliott, M. S. The effect of queuosine on tRNA structure and function. *J Biomol Struct Dyn* **16**, 757-774, doi:10.1080/07391102.1999.10508291 (1999).
30 Thiaville, P. C. *et al.* Global translational impacts of the loss of the tRNA modification t(6)A in yeast. *Microbial cell (Graz, Austria)* **3**, 29-45, doi:10.15698/mic2016.01.473 (2016).
31 Torres, A. G., Batlle, E. & Ribas de Pouplana, L. Role of tRNA modifications in human diseases. *Trends Mol Med* **20**, 306-314, doi:10.1016/j.molmed.2014.01.008 (2014).
32 Vare, V. Y., Eruysal, E. R., Narendran, A., Sarachan, K. L. & Agris, P. F. Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules* **7**, doi:10.3390/biom7010029 (2017).
33 Vendeix, F. A. *et al.* Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J Mol Biol* **416**, 467-485, doi:10.1016/j.jmb.2011.12.048 (2012).
34 Zhou, J.-B. *et al.* Molecular basis for t6A modification in human mitochondria. *Nucleic Acids Res* **48**, 3181-3194, doi:10.1093/nar/gkaa093 (2020).
35 Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646-652, doi:10.1038/nsb0902-646 (2002).
36 Cheatham, T. E. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Current Opinion in Structural Biology* **14**, 360-367, doi:https://doi.org/10.1016/j.sbi.2004.05.001 (2004).
37 Fadda, E. & Woods, R. J. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today* **15**, 596-609, doi:10.1016/j.drudis.2010.06.001 (2010).
38 Giudice, E. & Lavery, R. Simulations of Nucleic Acids and Their Complexes. *Accounts of Chemical Research* **35**, 350-357, doi:10.1021/ar010023y (2002).
39 Norberg, J. & Nilsson, L. Molecular Dynamics Applied to Nucleic Acids. *Accounts of Chemical Research* **35**, 465-472, doi:10.1021/ar010026a (2002).
40 Krawczyk, K., Sim, A. Y. L., Knapp, B., Deane, C. M. & Minary, P. Tertiary Element Interaction in HIV-1 TAR. *Journal of Chemical Information and Modeling* **56**, 1746-1754, doi:10.1021/acs.jcim.6b00152 (2016).
41 Kung, R. W., Deak, T. K., Griffith-Salik, C. A., Takyi, N. A. & Wetmore, S. D. Impact of DNA Adduct Size, Number, and Relative Position on the Toxicity of Aromatic Amines: A Molecular

Dynamics Case Study of ANdG- and APdG-Containing DNA Duplexes. *Journal of Chemical Information and Modeling* **61**, 2313-2327, doi:10.1021/acs.jcim.1c00202 (2021).
42 Sharma, K. D., Kathuria, P., Wetmore, S. D. & Sharma, P. Can modified DNA base pairs with chalcogen bonding expand the genetic alphabet? A combined quantum chemical and molecular dynamics simulation study. *Physical Chemistry Chemical Physics* **22**, 23754-23765, doi:10.1039/D0CP04921B (2020).
43 Cheatham III, T. E. & Young, M. A. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* **56**, 232-256 (2001).
44 Luitz, M., Bomblies, R., Ostermeir, K. & Zacharias, M. Exploring biomolecular dynamics and interactions using advanced sampling methods. *Journal of Physics: Condensed Matter* **27**, 323101, doi:10.1088/0953-8984/27/32/323101 (2015).
45 van Gunsteren, W. F. & Mark, A. E. Validation of molecular dynamics simulation. *The Journal of Chemical Physics* **108**, 6109-6116, doi:10.1063/1.476021 (1998).
46 Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annual Review of Biophysics* **40**, 41-62, doi:10.1146/annurev-biophys-042910-155255 (2011).
47 Higo, J. & Nakamura, H. Virtual states introduced for overcoming entropic barriers in conformational space. *Biophysics* **8**, 139-144 (2012).
48 Sharp, P. A. The Centrality of RNA. *Cell* **136**, 577-580, doi:10.1016/j.cell.2009.02.007 (2009).
49 Cruz, J. A. & Westhof, E. The Dynamic Landscapes of RNA Architecture. *Cell* **136**, 604-609, doi:10.1016/j.cell.2009.02.003 (2009).
50 Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nature Reviews Genetics* **15**, 469-479, doi:10.1038/nrg3681 (2014).
51 Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13-37 (2005).
52 Montange, R. K. & Batey, R. T. Riboswitches: Emerging Themes in RNA Structure and Function. *Annual Review of Biophysics* **37**, 117-133, doi:10.1146/annurev.biophys.37.032807.130000 (2008).
53 Warf, M. B. & Berglund, J. A. Role of RNA structure in regulating pre-mRNA splicing. *Trends in biochemical sciences* **35**, 169-178 (2010).
54 Woese, C. R. & Pace, N. R. Probing RNA structure, function, and history by comparative analysis. *Cold Spring Harbor Monograph Series* **24**, 91-91 (1993).
55 Banáš, P. *et al.* Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of Chemical Theory and Computation* **6**, 3836-3849, doi:10.1021/ct100481h (2010).
56 Beššeová, I. *et al.* Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration. *The journal of physical chemistry B* **116**, 9899-9916 (2012).
57 Ditzler, M. A., Otyepka, M., Šponer, J. & Walter, N. G. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Accounts of Chemical Research* **43**, 40-47, doi:10.1021/ar900093g (2010).
58 Fadrná, E. *et al.* Molecular dynamics simulations of guanine quadruplex loops: advances and force field limitations. *Biophysical journal* **87**, 227-242 (2004).
59 Krepl, M., Réblová, K., Koča, J. & Šponer, J. Bioinformatics and Molecular Dynamics Simulation Study of L1 Stalk Non-Canonical rRNA Elements: Kink-Turns, Loops, and Tetraloops. *The Journal of Physical Chemistry B* **117**, 5540-5555, doi:10.1021/jp401482m (2013).
60 Kuhrova, P. *et al.* Computer folding of RNA tetraloops: identification of key force field deficiencies. *Journal of chemical theory and computation* **12**, 4534-4548 (2016).
61 Mlýnský, V. & Bussi, G. Exploring RNA structure and dynamics through enhanced sampling simulations. *Current Opinion in Structural Biology* **49**, 63-71, doi:https://doi.org/10.1016/j.sbi.2018.01.004 (2018).
62 Mráziková, K. *et al.* UUCG RNA Tetraloop as a Formidable Force-Field Challenge for MD Simulations. *Journal of Chemical Theory and Computation* **16**, 7601-7617, doi:10.1021/acs.jctc.0c00801 (2020).
63 Pal, S. & Paul, S. Effect of Hydrated and Nonhydrated Choline Chloride–Urea Deep Eutectic Solvent (Reline) on Thrombin-Binding G-quadruplex Aptamer (TBA): A Classical Molecular Dynamics Simulation Study. *The Journal of Physical Chemistry C* **123**, 11686-11698, doi:10.1021/acs.jpcc.9b01111 (2019).

64 Rebič, M., Laaksonen, A., Šponer, J., Uličný, J. & Mocci, F. Molecular Dynamics Simulation Study of Parallel Telomeric DNA Quadruplexes at Different Ionic Strengths: Evaluation of Water and Ion Models. *The Journal of Physical Chemistry B* **120**, 7380-7391, doi:10.1021/acs.jpcb.6b06485 (2016).
65 Šponer, J., Islam, B., Stadlbauer, P. & Haider, S. in *Annual Reports in Medicinal Chemistry* Vol. 54 197-241 (Elsevier, 2020).
66 Zgarbová, M., Otyepka, M., Šponer, J., Lankaš, F. & Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *Journal of Chemical Theory and Computation* **10**, 3177-3189, doi:10.1021/ct500120v (2014).
67 Voter, A. F. Parallel replica method for dynamics of infrequent events. *Physical Review B* **57**, R13985-R13988, doi:10.1103/PhysRevB.57.R13985 (1998).
68 Shirts, M. R. & Pande, V. S. Mathematical Analysis of Coupled Parallel Simulations. *Physical Review Letters* **86**, 4983-4987, doi:10.1103/PhysRevLett.86.4983 (2001).
69 Christen, M. & van Gunsteren, W. F. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *Journal of Computational Chemistry* **29**, 157-166, doi:https://doi.org/10.1002/jcc.20725 (2008).
70 Fersht, A. R. On the simulation of protein folding by short time scale molecular dynamics and distributed computing. *Proceedings of the National Academy of Sciences* **99**, 14122, doi:10.1073/pnas.182542699 (2002).
71 Bottaro, S., Bussi, G., Kennedy, S. D., Turner, D. H. & Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Science advances* **4**, eaar8521-eaar8521, doi:10.1126/sciadv.aar8521 (2018).
72 Steuer, J., Kukharenko, O., Riedmiller, K., Hartig, J. S. & Peter, C. Guanidine-II aptamer conformations and ligand binding modes through the lens of molecular simulation. *Nucleic Acids Res* **49**, 7954-7965, doi:10.1093/nar/gkab592 (2021).
73 Sorin, E. J., Rhee, Y. M., Nakatani, B. J. & Pande, V. S. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophysical journal* **85**, 790-803, doi:10.1016/S0006-3495(03)74520-2 (2003).
74 Batey, R. T., Rambo, R. P. & Doudna, J. A. Tertiary Motifs in RNA Structure and Folding. *Angewandte Chemie International Edition* **38**, 2326-2343, doi:https://doi.org/10.1002/(SICI)1521-3773(19990816)38:16<2326::AID-ANIE2326>3.0.CO;2-3 (1999).
75 Weeks, K. M. Review toward all RNA structures, concisely. *Biopolymers* **103**, 438-448, doi:https://doi.org/10.1002/bip.22601 (2015).
76 Galindo-Murillo, R., Roe, D. R. & Cheatham, T. E. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 1041-1058, doi:https://doi.org/10.1016/j.bbagen.2014.09.007 (2015).
77 Caves, L. S., Evanseck, J. D. & Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Science* **7**, 649-666 (1998).
78 Genheden, S. & Ryde, U. How to obtain statistically converged MM/GBSA results. *Journal of Computational Chemistry* **31**, 837-846, doi:https://doi.org/10.1002/jcc.21366 (2010).
79 Perez, J. J., Tomas, M. S. & Rubio-Martinez, J. Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory. *Journal of Chemical Information and Modeling* **56**, 1950-1962, doi:10.1021/acs.jcim.6b00347 (2016).
80 Byrne, R. T., Konevega, A. L., Rodnina, M. V. & Antson, A. A. The crystal structure of unmodified tRNAPhe from Escherichia coli. *Nucleic Acids Res* **38**, 4154-4162, doi:10.1093/nar/gkq133 (2010).
81 Schmit, J. D., Kariyawasam, N. L., Needham, V. & Smith, P. E. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *Journal of Chemical Theory and Computation* **14**, 1823-1827, doi:10.1021/acs.jctc.7b01254 (2018).
82 AMBER 2018 (University of California, San Francisco, 2018).
83 Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095, doi:10.1021/ct400341p (2013).
84 AMBER 2020 (University of California, San Francisco, 2020).

85 Bottaro, S. *et al.* Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* **25**, 219-231, doi:10.1261/rna.067678.118 (2019).
86 Bottaro, S., Di Palma, F. & Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res* **42**, 13306-13314, doi:10.1093/nar/gku972 (2014).
87 Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD* **96**, 226-231 (1996).
88 Schrödinger, L. The PyMOL Molecular Graphics System, Version 2.0
89 Altis, A., Nguyen, P. H., Hegger, R. & Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *The Journal of Chemical Physics* **126**, 244111, doi:10.1063/1.2746330 (2007).
90 Barraud, P., Schmitt, E., Mechulam, Y., Dardel, F. & Tisné, C. A unique conformation of the anticodon stem-loop is associated with the capacity of tRNAfMet to initiate protein synthesis. *Nucleic Acids Res* **36**, 4894-4901, doi:10.1093/nar/gkn462 (2008).
91 Barrell, B. G. *et al.* Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc Natl Acad Sci U S A* **77**, 3164-3166, doi:10.1073/pnas.77.6.3164 (1980).
92 Carter, C. W. & Wolfenden, R. tRNA acceptor stem and anticodon bases form independent codes related to protein folding. *Proceedings of the National Academy of Sciences* **112**, 7489, doi:10.1073/pnas.1507569112 (2015).
93 Carter, C. W. & Wolfenden, R. tRNA acceptor-stem and anticodon bases embed separate features of amino acid chemistry. *RNA Biology* **13**, 145-151, doi:10.1080/15476286.2015.1112488 (2016).
94 Agirrezabala, X. & Valle, M. Structural Insights into tRNA Dynamics on the Ribosome. *International Journal of Molecular Sciences* **16**, 9866-9895 (2015).
95 Berk, V., Zhang, W., Pai, R. D. & Cate, J. H. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci U S A* **103**, 15830-15834, doi:10.1073/pnas.0607541103 (2006).
96 Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S. & Puglisi, J. D. tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol* **11**, 1008-1014, doi:10.1038/nsmb831 (2004).
97 Khade, P. & Joseph, S. Functional interactions by transfer RNAs in the ribosome. *FEBS Lett* **584**, 420-426, doi:10.1016/j.febslet.2009.11.034 (2010).
98 Hall, K. B., Sampson, J. R., Uhlenbeck, O. C. & Redfield, A. G. Structure of an unmodified tRNA molecule. *Biochemistry* **28**, 5794-5801, doi:10.1021/bi00440a014 (1989).
99 Holley, R. W. *et al.* Structure of a ribonucleic acid. *Science* **147**, 1462-1465, doi:10.1126/science.147.3664.1462 (1965).
100 Bou-Nader, C. *et al.* HIV-1 matrix-tRNA complex structure reveals basis for host control of Gag localization. *Cell Host & Microbe* **29**, 1421-1436.e1427, doi:10.1016/j.chom.2021.07.006 (2021).
101 Jovine, L., Djordjevic, S. & Rhodes, D. The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg2+ in 15-year old crystals11Edited by J. Karn. *Journal of Molecular Biology* **301**, 401-414, doi:https://doi.org/10.1006/jmbi.2000.3950 (2000).
102 Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S.-H. Crystal structure of yeast phenylalanine transfer RNA: I. Crystallographic refinement. *Journal of Molecular Biology* **123**, 607-630, doi:https://doi.org/10.1016/0022-2836(78)90209-7 (1978).
103 Chan, C. W., Badong, D., Rajan, R. & Mondragón, A. Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA* **26**, 278-289, doi:10.1261/rna.073478.119 (2020).
104 Halder, S. & Bhattacharyya, D. RNA structure and dynamics: A base pairing perspective. *Progress in Biophysics and Molecular Biology* **113**, 264-283, doi:https://doi.org/10.1016/j.pbiomolbio.2013.07.003 (2013).
105 Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30**, 3497-3531, doi:10.1093/nar/gkf481 (2002).
106 Saint-Léger, A. *et al.* Saturation of recognition elements blocks evolution of new tRNA identities. *Science Advances* **2**, e1501860, doi:doi:10.1126/sciadv.1501860 (2016).
107 Eiler, S., Dock-Bregeon, A.-C., Moulinier, L., Thierry, J.-C. & Moras, D. Synthesis of aspartyl-tRNAAsp in Escherichia coli—a snapshot of the second step. *The EMBO Journal* **18**, 6532-6541, doi:https://doi.org/10.1093/emboj/18.22.6532 (1999).

108 Hur, S. & Stroud, R. M. How U38, 39, and 40 of Many tRNAs Become the Targets for Pseudouridylation by TruA. *Molecular Cell* **26**, 189-203, doi:10.1016/j.molcel.2007.02.027 (2007).

109 Sherlin, L. D. *et al.* Influence of transfer RNA tertiary structure on aminoacylation efficiency by glutaminyl and cysteinyl-tRNA synthetases11Edited by J. Doudna. *Journal of Molecular Biology* **299**, 431-446, doi:https://doi.org/10.1006/jmbi.2000.3749 (2000).

110 Murray, L. J. W., Arendall, W. B., Richardson, D. C. & Richardson, J. S. RNA backbone is rotameric. *Proceedings of the National Academy of Sciences* **100**, 13904, doi:10.1073/pnas.1835769100 (2003).

111 Reijmers, T., Wehrens, R. & Buydens, L. Circular effects in representations of an RNA nucleotides data set in relation with principal components analysis. *Chemometrics and Intelligent Laboratory Systems* **56**, 61-71 (2001).

112 Richardson, J. S. *et al.* RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA (New York, N.Y.)* **14**, 465-481, doi:10.1261/rna.657708 (2008).

113 Agris, P. F. The importance of being modified: an unrealized code to RNA structure and function. *RNA* **21**, 552-554, doi:10.1261/rna.050575.115 (2015).

114 Agris, P. F. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* **9**, 629-635, doi:10.1038/embor.2008.104 (2008).

115 Bjork, G. R. & Hagervall, T. G. Transfer RNA Modification: Presence, Synthesis, and Function. *EcoSal Plus* **6**, doi:10.1128/ecosalplus.ESP-0007-2013 (2014).

116 Cantara, W. A., Murphy, F. V. t., Demirci, H. & Agris, P. F. Expanded use of sense codons is regulated by modified cytidines in tRNA. *Proc Natl Acad Sci U S A* **110**, 10964-10969, doi:10.1073/pnas.1222641110 (2013).

117 Šponer, J. *et al.* Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *The Journal of Physical Chemistry Letters* **5**, 1771-1782, doi:10.1021/jz500557y (2014).

118 Hospital, A., Goñi, J.R., Orozco, M., Gelpi, J. Molecular dynamics simulations: advances and applications. . *Adv Appl Bioinform Chem.* **8**, 37-47, doi:https://doi.org/10.2147/AABC.S70333 (2015).

119 Hammond, J. A., Rambo, R. P., Filbin, M. E. & Kieft, J. S. Comparison and functional implications of the 3D architectures of viral tRNA-like structures. *RNA* **15**, 294-307, doi:10.1261/rna.1360709 (2009).

120 Wu, S., Li, X. & Wang, G. tRNA-like structures and their functions. *The FEBS Journal* (2021).

121 Amadei, A., Ceruso, M. A. & Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **36**, 419-424, doi:https://doi.org/10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO;2-U (1999).

122 Grossfield, A. & Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu Rep Comput Chem* **5**, 23-48, doi:10.1016/S1574-1400(09)00502-7 (2009).

123 Smith, L. J., Daura, X. & van Gunsteren, W. F. Assessing equilibration and convergence in biomolecular simulations. *Proteins: Structure, Function, and Bioinformatics* **48**, 487-496, doi:https://doi.org/10.1002/prot.10144 (2002).

124 Ashraf, S. S. *et al.* The uridine in "U-turn": contributions to tRNA-ribosomal binding. *RNA (New York, N.Y.)* **5**, 503-511, doi:10.1017/s1355838299981931 (1999).

125 Ashraf, S. S., Guenther, R. & Agris, P. F. Orientation of the tRNA anticodon in the ribosomal P-site: quantitative footprinting with U33-modified, anticodon stem and loop domains. *RNA* **5**, 1191-1199, doi:10.1017/s1355838299990933 (1999).

126 Han, L. & Phizicky, E. M. A rationale for tRNA modification circuits in the anticodon loop. *RNA* **24**, 1277-1284, doi:10.1261/rna.067736.118 (2018).

127 Sambhare, S. B. *et al.* Structural significance of modified nucleosides k2C and t6A present in the anticodon loop of tRNAIle. *RSC Advances* **4**, 14176-14188, doi:10.1039/C3RA47335J (2014).

128 Xiao, X., Agris, P. F. & Hall, C. K. Molecular recognition mechanism of peptide chain bound to the tRNALys3 anticodon loop in silico. *Journal of Biomolecular Structure and Dynamics* **33**, 14-27, doi:10.1080/07391102.2013.869660 (2015).

129 Durant, P. C., Bajji, A. C., Sundaram, M., Kumar, R. K. & Davis, D. R. Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. *Biochemistry* **44**, 8078-8089, doi:10.1021/bi050343f (2005).
130 Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Current opinion in structural biology* **16**, 279-287, doi:10.1016/j.sbi.2006.05.009 (2006).
131 Leontis, N. B. & Westhof, E. Analysis of RNA motifs. *Current Opinion in Structural Biology* **13**, 300-308, doi:https://doi.org/10.1016/S0959-440X(03)00076-9 (2003).
132 Moore, P. B. Structural Motifs in RNA. *Annual Review of Biochemistry* **68**, 287-300, doi:10.1146/annurev.biochem.68.1.287 (1999).
133 Lescoute, A. & Westhof, E. Topology of three-way junctions in folded RNAs. *RNA* **12**, 83-93, doi:10.1261/rna.2208106 (2006).
134 Dibrov, S. M., Johnston-Cox, H., Weng, Y. H. & Hermann, T. Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew Chem Int Ed Engl* **46**, 226-229, doi:10.1002/anie.200603807 (2007).
135 Laughrea, M. & Jetté, L. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation. *Biochemistry* **35**, 1589-1598, doi:10.1021/bi951838f (1996).
136 Shu, D., Shu, Y., Haque, F., Abdelmawla, S. & Guo, P. Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nat Nanotechnol* **6**, 658-667, doi:10.1038/nnano.2011.105 (2011).
137 Li, M. *et al.* In vivo production of RNA nanostructures via programmed folding of single-stranded RNAs. *Nature Communications* **9**, 2196, doi:10.1038/s41467-018-04652-4 (2018).
138 Ohno, H., Akamine, S. & Saito, H. RNA nanostructures and scaffolds for biotechnology applications. *Current Opinion in Biotechnology* **58**, 53-61, doi:https://doi.org/10.1016/j.copbio.2018.11.006 (2019).
139 Saito, H. & Inoue, T. Synthetic biology with RNA motifs. *The International Journal of Biochemistry & Cell Biology* **41**, 398-404, doi:https://doi.org/10.1016/j.biocel.2008.08.017 (2009).

CHAPTER 4: EFFECTS OF CMNM^5(S^2)U34 AND ASSOCIATED A37 MODIFICATIONS IN TRNA

4.1 Objectives

As previously discussed, a broad range of modified nucleotides have been identified at positions 34 (wobble base) and 37 of tRNA, and these modifications have been proposed to facilitate tRNA accommodation within the ribosomal binding site.[1-3] 5-carboxymethylaminomethyl(-2-thio)uridine (cmnm^5(s^2)U) and related modifications are bacterial wobble base modifications present in tRNA$^{Leu}_{UUR}$, tRNA^{Lys}, tRNA^{Trp}, tRNA^{Glu} and tRNA^{Gln} that selectively recognize purine bases at the third codon position during translation.[4-6] *In vivo*, these modifications are always accompanied by modified adenosines, 2-methylthio-N^6-isopentenyladenosine (ms^{2}i^{6}A) or N^6-threonylcarbamoyladenosine (t^{6}A), at position 37. cmnm^5(s^2)U and their taurine analogs, 5-taurinomethyl(-2-thio)uridine (τm^5(s^2)U), are also prevalent in human mitochondrial tRNA^{Trp}, tRNA^{Leu} and tRNA^{Lys}. Anomalies in these modifications have been linked to substantial reductions in mitochondrial translation and the development of mitochondrial syndromes, MELAS and MERRF.[7-12] Unfortunately, cures have yet to be found for these diseases and limited knowledge exists on the dynamic nature of these modifications *in vivo*. Consequently, the present study uses replica MD simulations to investigate the structural impact of cmnm^5(s^2)U$_{34}$ modifications on tRNA and their synergistic effects with modifications at position 37, as evidence suggests that modifications within the ASL can influence the effects imparted by other modified nucleobases.[3,13]

Toward this end, cmnm^{5}U and cmnm^{5}s^{2}U were computationally inserted at the wobble position of models for completely unmodified tRNA^{Trp} and tRNA^{Lys}, respectively, as these carry the modifications of interest *in vivo* (Table 4.1). The unmodified tRNA variants were also considered for use as reference to reveal the impact of the individual modifications. Subsequently, the effects of double modifications at positions 34 and 37 on each tRNA system were also considered. For these investigations, ms^{2}i^{6}A and t^{6}A were inserted into the 37th position of cmnm^{5}U34-tRNA^{Trp} and cmnm^{5}s^{2}U34-tRNA^{Lys}, respectively. Note that multiple computational and experimental studies have investigated the effects of ms^{2}i^{6}A37 and t^{6}A37 on tRNA structure in the absence of other modifications.[14-25] Consequently, the primary focus of this study is to determine the effects of the cmnm^{5}U modification family on tRNA structure before considering their additive effects with associated modifications at position 37. The detailed computational analyses within

this study uncover: (i) the spatial configurations adopted by each modification; (ii) the conformational impacts of the modifications on the ASL, including the effects of single atom substitutions in modified nucleobases; (iii) the structural impacts of the modifications on tRNA, and (iv) the synergistic effects of cmnm^5(s^2)U34 and modified adenosines on tRNA structure. Moreover, the tRNA systems modeled in this study will be compared to experimentally isolated protein-bound and ribosome-bound tRNAs to discover the effects these modifications may have *in vivo*. The results show marked differences in the dynamics of tRNAs containing cmnm^{5}U and cmnm^{5}s^{2}U and reveal the role double modifications play in structurally ordering the ASL for translation. In summary, the molecular details provided in this study are critical for understanding the biological functions of cmnm^5(s^2)U and associated A37 modifications and take an important step toward revealing the structure–function properties of different tRNAs. This information is valuable as it could inform the future design of tRNA-based or nucleotide-based therapeutics in the battle against MELAS and MERRF.

4.2 Methodology

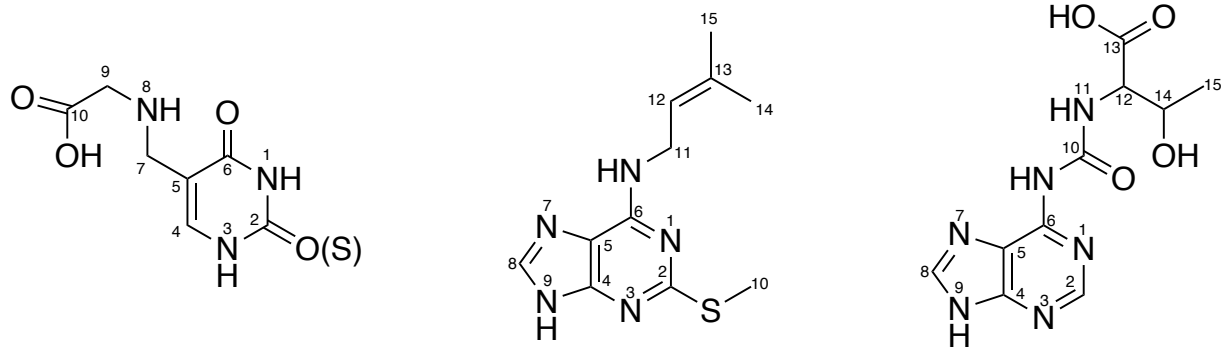
4.2.1 Model choice

An unmodified *E. coli* tRNA^{Phe} model was used to develop a computational protocol to accurately sample the tRNA phase space (Chapter 3). To ensure that this protocol can be used to reveal the structural changes that arise in the presence of posttranscriptional modifications, the configurational effects of cmnm^5(s^2)U34 modifications were investigated in the same tRNA structure (*E. coli* tRNA^{Phe}, PDB ID: 3L0U) using 10-replica ensembles. Results from this study (presented in Appendix III) revealed that 10-replica ensembles can be used to uncover the structural differences that arise due to the presence of modified nucleobases. Consequently, the protocol was applied to tRNA^{Trp} and tRNA^{Lys} models that carry cmnm^5(s^2)U modifications *in vivo*.

4.2.2 Model preparation

Initial coordinates of standalone, full-length *Escherichia coli* tRNA^{Trp} and *Bos Taurus* tRNA^{Lys} were taken from X-ray crystal structures with PDB IDs 4V5R (3.1 Å) and 1FIR (3.3 Å), respectively.[26,27] To build

comparable tRNA models (73 nt), the 3' terminal residues of $tRNA^{Trp}$ (CCAW) and $tRNA^{Lys}$ (CCA) were deleted, ensuring all tRNA structures investigated were composed of 73 nucleotides. To uncover the structural effects of modifications of interest, all chemical modifications within $tRNA^{Trp}$ and $tRNA^{Lys}$ were manually removed and replaced with the corresponding canonical bases using the PyMOL suite.[28] $cmnm^5(s^2)U34$ and associated modified adenosines were then built by editing the uridine and adenine nucleobase templates from AMBER using Avogadro (Scheme 4.1) and inserted into unmodified $tRNA^{Trp}$ and $tRNA^{Lys}$.[29]



Scheme 4.1 – Atom numbering of modified nucleobases investigated in the present work

These nucleobases were parameterized (atom type assignments and charges) using the PyRED program, and the parameters were supplemented by the Generalized Amber Force Field (GAFF).[30-35] tRNA models were neutralized using Na^+ ions and excess Na^+ and Cl^- ions were added to attain a physiological salt concentration of 150 mM.[36] All systems were solvated in an explicit TIP3P octahedral water box such that the solute was at least 10.0 Å from the box edge in all directions. The models were prepared using the LEaP module in the AMBER 18 package and the tRNA was described using the ff99bsc0χ_{OL3} forcefield.[37]

4.2.3 MD simulation protocol

The position of all solvent molecules and ions were initially minimized using 2500 steps of steepest descent, followed by 2500 steps of conjugate gradient minimization, with a force constant of

100 kcal mol^{-1} $Å^{-2}$ applied to constrain the tRNA. Subsequently, heavy atoms of the solvent molecules and ions were constrained using a force constant of 100 kcal mol^{-1} $Å^{-2}$ and the positions of all hydrogen atoms within the model were minimized using 2500 steps of steepest descent, followed by 2500 steps of conjugate gradient minimization. The solute was then minimized using 2500 steps of steepest descent, followed by 2500 steps of conjugate gradient minimization, while a force constant of 100 kcal mol^{-1} $Å^{-2}$ was applied to all solvent and ion molecules. Finally, the entire system was minimized using 2500 steps of unrestrained steepest descent, followed by 2500 steps of unrestrained conjugate gradient minimization.

Following minimization, the solute was restrained using a 25 kcal mol^{-1} $Å^{-2}$ force constant and the system was heated from 0 to 310 K in 50 K increments using the Langevin thermostat (γ = 1) and a 1 fs time step. Subsequently, the force restraints on the solute were reduced in a stepwise manner, from 25 to 1.5 kcal mol^{-1} $Å^{-2}$, using a time step of 2 fs and the SHAKE algorithm under NPT conditions (1 atm, 310 K). The PMEMD CUDA module of AMBER 18 was used to perform 500-ns MD production simulations using a time step of 2 fs.[37] All simulations were carried out with the periodic boundary condition, using a 10 Å nonbonded cut-off and accounting for electrostatic interactions using the particle mesh Ewald (PME) method. In accordance with the findings presented in Chapter 3, each system was simulated using a replica ensemble of ten 500-ns simulations, and six systems were investigated (Table 4.1), yielding a simulation time of 5 μs per system and a total simulation time of 30 μs for this study.

Table 4.1 – Transfer RNA systems investigated in this study

tRNA system	Modification(s) present
$tRNA^{Trp}$	none
	$cmnm^5U34$
	$cmnm^5U34 + ms^2i^6A37$
$tRNA^{Lys}$	none
	$cmnm^5s^2U34$
	$cmnm^5s^2U34 + t^6A37$

4.2.4 Analyses

Trajectories were sampled for analysis every 200 ps over the course of the production phase. The AmberTools 20 version of CPPTRAJ was used to analyze all trajectories.[38,39] To assess system stability, heavy-atom RMSDs for each simulation were evaluated with respect to the corresponding unmodified crystal structure coordinates. Hydrogen-bond occupancies were evaluated using a distance cut-off of 3.4 Å and an angle cut-off of 120°, while stacking occupancies were determined using a distance cut-off of ≤ 6 Å between the center of masses and an angle cut-off ≤ 40° or ≥ 140° between the normal vectors of the planes of the two bases.

The internal structural dynamics at the ASL were analyzed using the Barnaba library in Python.[40] The solvent, ions and all other tRNA domains were stripped from the trajectories and conformational analyses were performed on residues 27 to 43. Heavy atom RMSD and εRMSD were obtained to visualize the dynamics of the ASL using the minimized crystal structure as a reference. Following the εRMSD calculation, structures from each trajectory were clustered using the DBSCAN algorithm, and the clustering analysis was visualized using the first two principal components of the trajectory.[41] Representative structures from each cluster were visually inspected using PyMOL, and used to classify the clusters into conformational groups.

To determine the variability within each conformational group, the clustered frames were saved to independent trajectories and the base–base interactions were quantified using CPPTRAJ. A dynamic secondary structure representation of each conformational group was built using BARNABA. BARNABA defaults to a distance cut-off of 3.4 Å and an angle cut-off of 120° for hydrogen-bond interactions and a distance cut-off of ≤ 4 Å between the center of masses and an angle cut-off < 40° or > 140° between the normal vectors of the planes of the bases for stacking interactions.

The structural dynamics within residues 34 and 37 were also considered, and structural changes within the modified nucleobases were captured every 5 ns. For each system, backbone dihedral angles (α, β, γ, δ, ε, ζ, χ), glycosidic torsions (χ), and dihedral angles within the side chain of all modifications were calculated and averaged across the replica ensemble using CPPTRAJ.

4.3 Results and Discussion

4.3.1 cmnm^{5}U34 and ms^{2}i^{6}A37 work together to reduce inherent dynamics in the anticodon loop of tRNA^{Trp} by stabilizing backbone torsions and enhancing nucleotide–nucleotide interactions

The cmnm^{5}U34 modification is derived from the addition of an amino-methylated glycine substituent at C5 of a wobble uridine base. Present in bacterial and mitochondrial tRNA^{Trp}, tRNA$^{Arg\text{-}1}$, tRNA^{Leu} and tRNA^{Lys},[42,43] the side chain of this modified base adopts a wide range of torsions,[44] leading to increased flexibility at the first anticodon position. *In vivo,* cmnm^{5}U34 is accompanied by ms^{2}i^{6}A37, a modified nucleotide that arises from the addition of an isopentenyl substituent at N6 and a thiomethyl substituent at C2 of canonical adenine. Previous experimental and computational studies have investigated the impact of ms^{2}i^{6}A37 on the structure of the anticodon loop,[45,46] but none considered the additive effect of cmnm^{5}U34 and ms^{2}i^{6}A37 on this region. Consequently, this study considers the local and global structural changes induced by the simultaneous presence of these modifications in *E. coli* tRNA^{Trp}, which carries these modifications in nature. Specifically, two modified tRNA^{Trp} systems were considered – a singly modified tRNA^{Trp} with cmnm^{5}U at position 34 (cmnm^{5}U34-tRNA^{Trp}) and a doubly modified system with cmnm^{5}U at position 34 and ms^{2}i^{6}A at position 37 (cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}).

To establish a point of reference for investigating the impact of cmnm^{5}U34 and ms^{2}i^{6}A37 in tRNA^{Trp}, ten 500-ns MD simulations were performed on the unmodified tRNA structure. On average, tRNA^{Trp} was stable throughout the trajectories, as indicated by the low average root-mean-square deviations (RMSD) for all heavy atoms with respect to the experimental starting structure (3.9 to 6.4 Å; Figure 4A.1). The root-mean-square fluctuations (RMSF; Figure 4.1) for each residue averaged over each replica ensemble revealed higher stability in the stem regions (RMSF < 5 Å) relative to the loop regions (RMSF > 6 Å). Specifically, the nucleobases at the elbow region of tRNA^{Trp} were highly mobile, resulting in varied hydrogen bonding and stacking interactions throughout independent replicas. Nevertheless, the D and TψC loops of tRNA^{Trp} did not unfold, and the kissing interaction between the loops persisted throughout the replica ensemble (Figure 4.1, 4A.2). Moreover, tRNA^{Trp} maintained its canonical L-shape over the course of all simulations, as the coaxial stacks of the D-ASL domains and the TψC-acceptor stem were preserved.

The most dynamic region of unmodified tRNA^{Trp} was the anticodon loop (C32 to A38), with RMSFs ranging between 7 and 14 Å. Concurring evidence was obtained from the torsional analysis of the backbone atoms for the loop region, which uncovered high amplitude motions and the formation of varied non-covalent interactions between the anticodon bases (positions 34, 35 and 36) of tRNA^{Trp} (Figure 4A.2). Nevertheless, the high degree of structural variations at the anticodon loop did not disrupt interactions within the stem of the ASL domain. Watson-Crick hydrogen bonding was maintained in the helical base pairs, with average occupancies > 99%, and all helical pairs stacked with adjacent nucleobases over 80% of the time (Figure 4A.2).

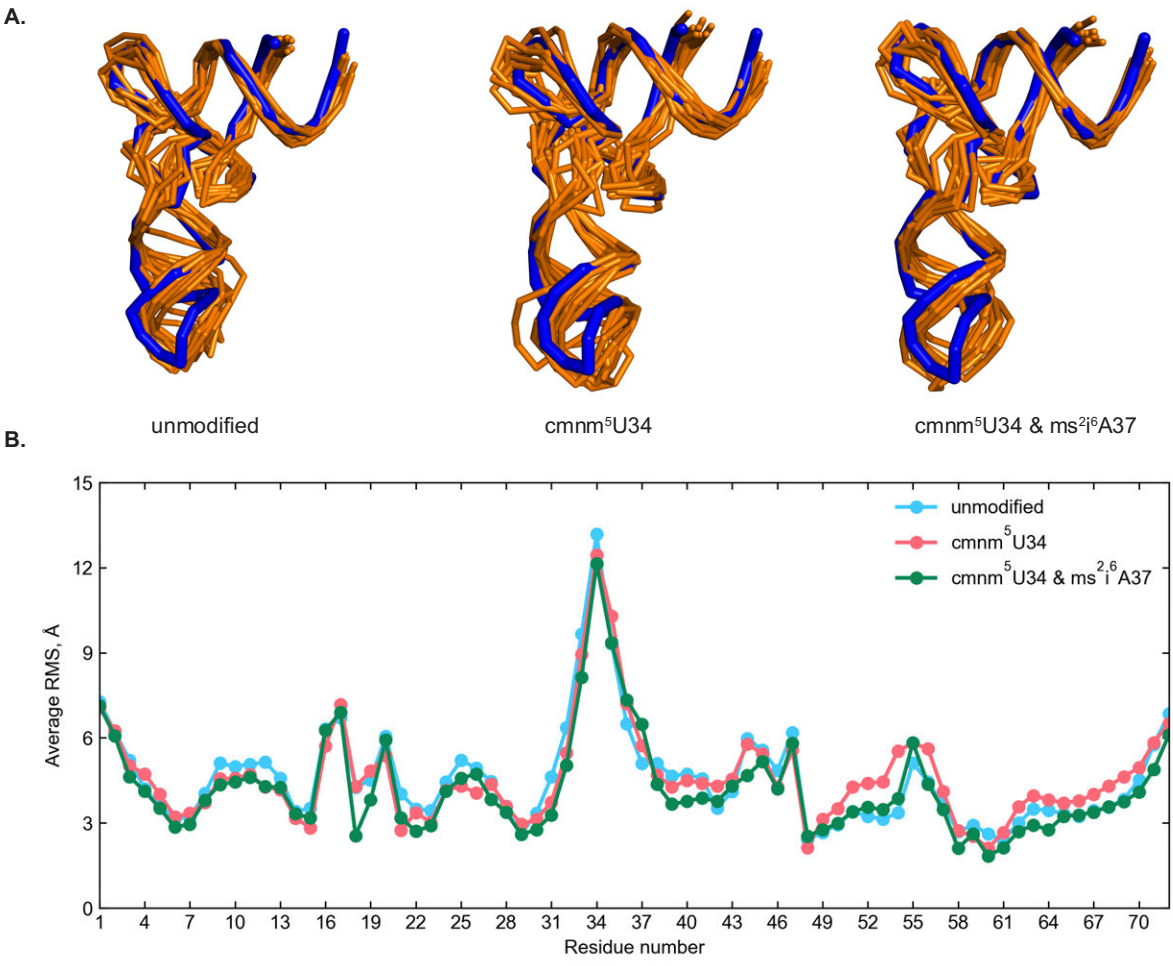


Figure 4.1 – Full tRNA dynamics across the tRNA^{Trp} systems

(A) Representative structure overlays (orange ribbons) from replica trajectories of unmodified tRNA^{Trp}, cmnm^{5}U34-tRNA^{Trp} and cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}. The reference structure is the experimental crystal structure (PDB ID: 4V5R, blue). (B) Average per residue fluctuations in unmodified tRNA^{Trp} (cyan), cmnm^{5}U34-tRNA^{Trp} (pink) and cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp} (green). The ASL domain spans from residue 27 to 43.

Stacking interactions were also observed within the anticodon loop, but interaction occupancies differed from one stacking pair to another. No stacking was observed between bases 33 and 34, implying that the U-turn motif was preserved. This concurs with the backbone profiles for these two nucleotides (Figure 4A.3), which revealed different η and θ torsions relative to their helical counterparts. Apart from the C34/C35 pair (stacking occupancy ~ 20%), other nucleobases within the ASL of $tRNA^{Trp}$ persistently stacked with their flanking bases, and the average interaction occupancies ranged from 60 to 80%.

Detailed analyses of the dynamics at positions C34 and A37 were conducted to provide a point of reference for understanding the impact of modified nucleobases at these positions. Similarly to the experimental starting structure, C34 adopted an *anti* conformation over the course of all replica simulations (average χ torsional angle of 209 ± 19°; Figure 4A.4). In contrast, the backbone torsion angles at position 34 varied greatly. Particularly, the α, ε, and ζ torsional angles occupied a wide range of values (31° < α < 341°, 72° < ε < 324°, 11° < ζ < 360°). Taken together, the wide range of torsional angles adopted by C34 and the low stacking occupancy in the C34/C35 (20%) pair indicate that the wobble base is highly flexible in the absence of modifications. This observation concurs with previous experimental and computational studies that have shown that, regardless of its identity, base 34 can form diverse interactions with nearby bases and adopt a variety of positions within the anticodon loop.[3,13,45,47-53] In contrast to C34, A37 was relatively stable within the anticodon loop (RMSF ~ 5 Å) and no significant deviations were observed in the backbone torsions at this position (Figure 4A.3). In unmodified $tRNA^{Trp}$, the nucleotide adopted an *anti* conformation over the course of all replica simulations (average χ torsional angle of 190 ± 9°; Figure 4A.5), and persistent stacking was observed between A37 and its flanking bases A36 (75%) and A38 (95%).

The varied interactions observed at positions 34 of $tRNA^{Trp}$ indicate that diverse structural arrangements may be present within the anticodon loop. Therefore, a conformational analysis was conducted to uncover the conformational states adopted by unmodified $tRNA^{Trp}$. The present study uses the conformational classification developed in Chapter 3. Four ASL conformations were identified in unmodified $tRNA^{Trp}$ (Figure 4.2) – 34-stacked (3%), 34-unstacked (58%), 36-unstacked (20%) and disordered (19%). Note that fluctuations at positions 34 and 36 did not affect the global structure of the

tRNA, but led to new non-covalent interactions within the ASL. The anticodon loop of $tRNA^{Trp}$ is also prone to disorder, with varied structural arrangements being observed 19% of the total simulation time relative to the crystal structure reference. The prevalence of the disordered conformation in unmodified $tRNA^{Trp}$, combined with high conformational variability, underscores the instability of the U-turn structure in the absence of modifications.

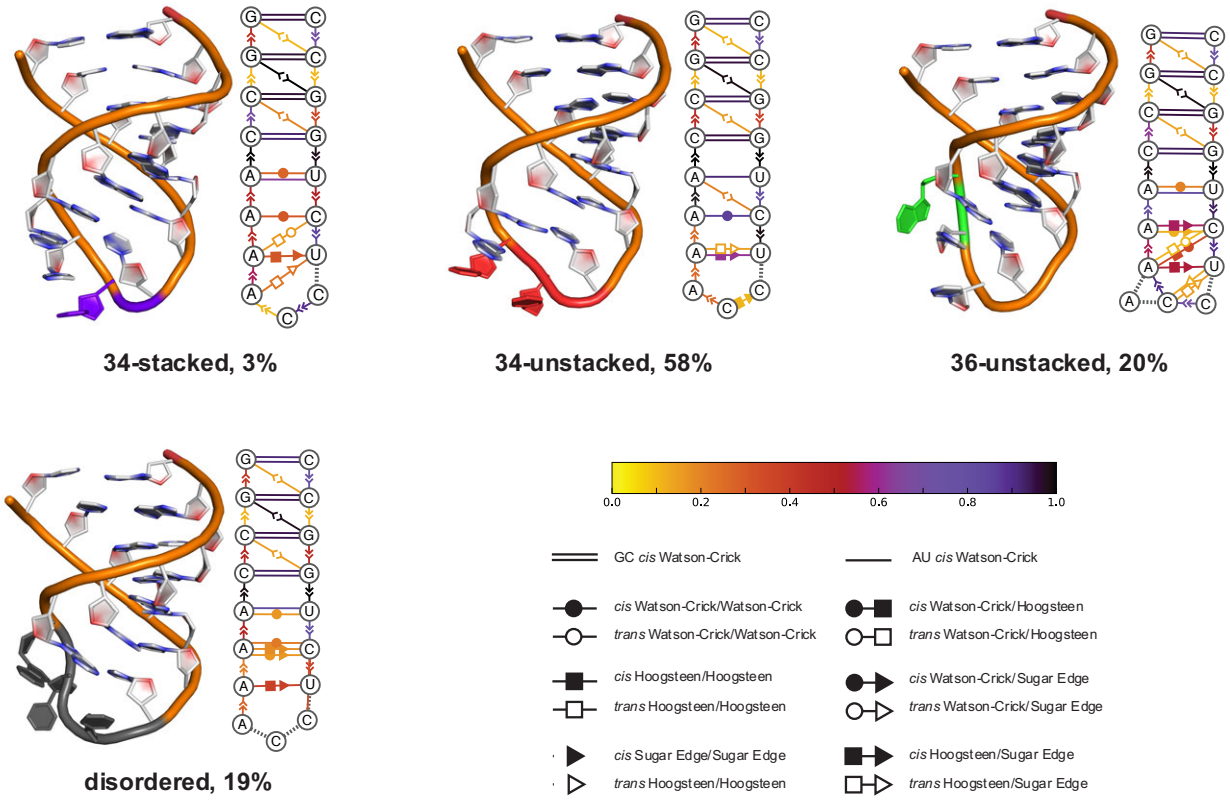


Figure 4.2 – Conformational profile for the ASL of unmodified $tRNA^{Trp}$

Cartoon and secondary structure representations of conformational states adopted by the unmodified $tRNA^{Trp}$ ASL. Conformations are defined and color-coded as in Chapter 3. Non-covalent interactions are denoted using the Leontis-Westhof notation for RNA molecules.

Despite the high flexibilities in the modification side chains of $cmnm^5U34$ and ms^2i^6A37 (Figure 4.3), the modified tRNA systems are stable throughout their replica simulations and exhibit comparable RMSDs to unmodified $tRNA^{Trp}$ (mean replica RMSDs of 5.1 ± 0.8 and 4.8 ± 0.7 Å for $cmnm^5U34$-$tRNA^{Trp}$ and $cmnm^5U34/ms^2i^6A37$-$tRNA^{Trp}$, respectively; Figure 4A.1). Moreover, similar trends were observed in the time averaged RMSFs for each nucleotide across all three systems (Figure 4.1).

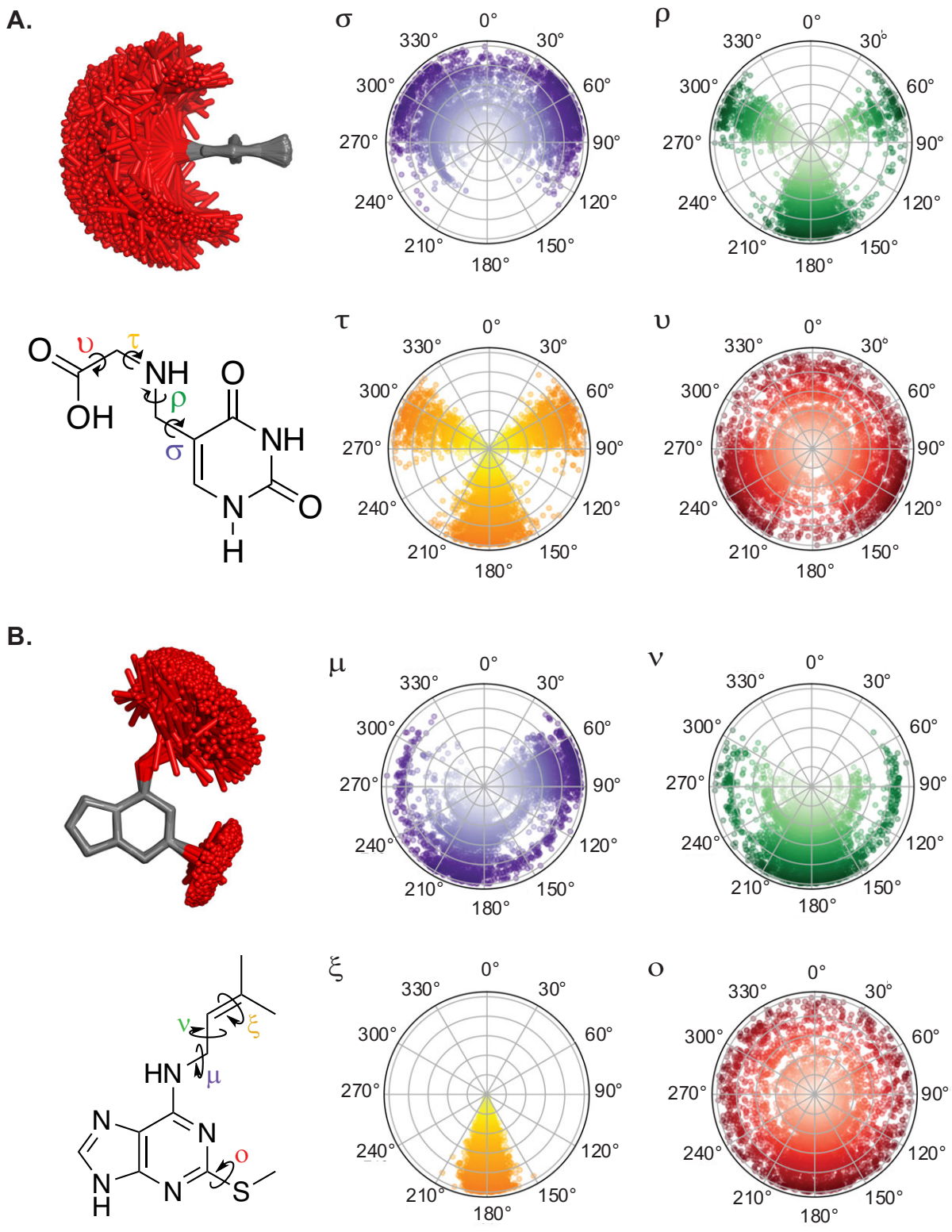


Figure 4.3 – Sidechain flexibilities of cmnm^{5}U34 and ms^{2}i^{6}A37

Structural representation and density maps of torsion angles adopted by cmnm^{5}U34 (A) and ms^{2}i^{6}A37 (B) across ten 500-ns trajectories. Simulation time (5 μs) is represented on the r-axes of each polar plot.

The insertion of modified nucleotides at the ASL neither enhanced nor weakened the overall interaction network at the elbow region of tRNA, as the non-covalent interactions that govern the tertiary fold of the tRNA were maintained at comparable occupancies across the three tRNA systems (Figure 4A.2). All in all, the modified $tRNA^{Trp}$ systems retained the expected L-shaped structure, through three-dimensional folding interactions present in the unmodified system. Nevertheless, structural analyses of the replica ensembles for each modified system revealed that bulky modifications at positions 34 and 37 differentially alter the structure of the ASL, especially in the anticodon loop (Figure 4.1).

Base pairs within the ASL helix were well-maintained in modified $tRNA^{Trp}$, with average occupancies over 80%. However, slightly reduced hydrogen bonding was observed in the U31–A39 base pair for $cmnm^5U34$-$tRNA^{Trp}$ (Figure 4A.2). The U31–A39 base pair had an interaction occupancy of 99% in the unmodified tRNA and 89% in its singly modified counterpart, which is further reduced to 83% in the doubly modified system. On the other hand, $cmnm^5U34$ and ms^2i^6A37 enhance hydrogen bonding between opposite bases in the anticodon loop. The C32–A38 interaction increased from 50% in the unmodified tRNA to 58% and 55% in the $cmnm^5U34$-$tRNA^{Trp}$ and $cmnm^5U34/ms^2i^6A37$-$tRNA^{Trp}$ systems respectively. More notably, hydrogen bonding interactions between A37 and U33, which were non-existent in unmodified $tRNA^{Trp}$, increased in the doubly modified system, to an average interaction occupancy of 27%. Stacking propensities within the ASL domain of $tRNA^{Trp}$ were also impacted by the presence of $cmnm^5U34$ and ms^2i^6A37. Although stacking within the stem was well preserved, both modified systems exhibited reduced stacking between C32 and U33 relative to the unmodified system (mean stacking occupancy of 68%, 56%, and 47% for unmodified, $cmnm^5U34$-$tRNA^{Trp}$ and $cmnm^5U34/ms^2i^6A37$-$tRNA^{Trp}$, respectively; Figure 4A.2). In the $cmnm^5U34$-$tRNA^{Trp}$ system, this mild reduction in stacking interactions extended to the 3' side of the loop, and average interaction occupancies for the 35/36, 36/37 and 37/38 stacks ranged from 53 to 81%, compared an average interaction range of 59 to 95% in unmodified $tRNA^{Trp}$. In contrast, the presence of ms^2i^6A37 stabilized stacking interactions within the anticodon loop, as evidenced by higher interaction occupancies for the 35/36 (65%), 36/37 (85%) and 37/38 (83%) stacks in the $cmnm^5U34/ms^2i^6A37$-$tRNA^{Trp}$ relative to the unmodified system. Stacking interactions involving the modified base at position 34 also varied among the modified systems. In the singly modified tRNA, the occupancy for the 34/35 stack is comparable to unmodified $tRNA^{Trp}$ (23%

and 20%, respectively), but this interaction is reduced to 13% in cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}. Regardless of the divergent hydrogen bonding and stacking interactions within the ASL of the modified systems, little to no stacking was observed between residues 33 and 34 in both structures (3% and 0% for cmn-tRNA^{Trp} and cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}, respectively) indicating the maintenance of the canonical U-turn motif in all modified structures. In summary, the formation and stability of non-covalent interactions were affected by the presence of cmnm^{5}U34 in the anticodon loop, and these effects were maintained (hydrogen bonding within the 31–39 and 32–38 base pairs and stacking interactions involving U32 and U33) or overturned (hydrogen bonding at position 37 and stacking interactions involving bases U34 to A37) following the addition of ms^{2}i^{6}A at position 37.

A previous MD study on the isolated cmnm^{5}U nucleotide (pcmnm^{5}U) revealed that the nucleotide preferentially adopted the *anti* conformation around its glycosidic bond, while its C5-substituent mostly maintained an extended form, which allowed for hydrogen bonding interactions between the side chain at N11 (Scheme 4.1) and the backbone.[44] In the present study, cmnm^{5}U34 predominantly adopted an *anti* conformation around its glycosidic bond, but unlike its unmodified counterpart, the base also adopted the *syn* conformation ~ 20% of the time (Figure 4A.4, 4A.6). The presence of the *syn* conformation in the singly modified tRNA^{Trp} system implies that cmnm^{5}U can alter the tRNA–mRNA interaction face as non-Watson-Crick interactions can now be formed between the modification and the third codon base. This concurs with experimental reports that cmnm^{5}U and its taurine homolog stabilize U·G wobble pairing during translation.[54,55] Nevertheless, cmnm^{5}U did not affect backbone torsions at position 34 (cmnm^{5}U34; Figure 4A.7) and the distributions for all backbone torsions at position 34 did not change in the presence of ms^{2}i^{6}A37, indicating that the latter modification does not restrain the movement of cmnm^{5}U34. Distributions for the side chain torsions of cmnm^{5}U34 concur with this observation, as similar torsions were sampled in the cmnm^{5}U34-tRNA^{Trp} and the cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp} structures (Figure 4A.7). In general, two side chain conformations were adopted with respect to the C5–C7 linkage, defined by the torsional angle $\sigma = \angle$(C4C5C7N10). Conformer-I ($180° \leq \sigma \leq 360°$) is populated for 45% of the time, while conformer-II ($0° \leq \sigma \leq 180°$) is sampled 55% of the simulation time. In each conformer, other dihedral angles adopt broad torsional ranges, and the side chain rotates between extended and bent forms (Figure 4A.8). In both states, the C5-substituent of cmnm^{5}U34 was able to interact with the phosphate

sugar backbone, stabilizing the position of the nucleobase within the loop. In the bent state, the modified residue also formed hydrogen bonding interactions with nucleobases at the 3' end of the anticodon loop, further enhancing the stability at this position (Figure 4A.9). Thus, my simulations confirm previous computational reports on the flexibility of the isolated cmnm^{5}U34[44] and uncover the dynamics of this nucleotide within the context of the ASL domain and the tRNA.

In contrast to cmnm^{5}U34, modifying A37 altered the backbone torsions at this position (Figure 4A.5, 4A.10, and 4A.11). In particular, the α and γ angles of ms^{2}i^{6}A37 differed from those observed in unmodified tRNA^{Trp}. In the unmodified system, the α angles ranged between 240 and 360°, with a mean value of 300°. In contrast, the α torsion of ms^{2}i^{6}A37 had two conformers, the first conformer (~ 30%) ranged between 60 and 180°, with a mean value of 120° and the second conformer (~ 70%) ranged between 240 and 360°, with a mean value of 300°. A similar trend was observed with the γ torsion of the modified nucleotide, suggesting that ms^{2}i^{6}A37 promotes local rearrangements at this position of the anticodon loop. The N6-substituent of ms^{2}i^{6}A adopted two main conformations with respect to the N6–C10 linkage, defined by the torsional angle $\mu = \angle$(C6N6C10C11). The first conformer ($30° \leq \mu \leq 150°$) was sampled ~ 45% of the time, while conformer-II was populated ~ 52% of the simulation time (Figure 4A.12). Consequently, this investigation provides evidence that modifications within the anticodon loop can alter dynamics at the modified position and at other locations within the anticodon loop.

Conformational sampling of the ASLs in the modified systems revealed that cmnm^{5}U34 and ms^{2}i^{6}A37 promote nucleotide rearrangements within the anticodon loop. Recall that in the absence of modifications, the ASL of tRNA^{Trp} is very flexible and highly susceptible to disarray, with the disordered and 36-unstacked conformations being sampled 19% and 20% of the time, respectively. When cmnm^{5}U is inserted at position 34, inherent dynamics of the anticodon loop increased, but disorder within the loop was reduced (5%). Specifically, in addition to the 36-unstacked conformation sampled in unmodified tRNA^{Trp} (tRNA^{Trp}: 20%, cmnm^{5}U34-tRNA^{Trp}: 29%), U33 (33-out, 1%) and U35 (35-unstacked, 2%) also move away from their crystal structure starting positions (Figure 4.4). The increased number of accessible conformations and increased range of motion observed at positions 33, 35 and 36 suggest that cmnm^{5}U34 may increase the flexibility of its neighboring nucleobases. Analysis of the pseudorotational

torsions at the anticodon loop confirms this hypothesis, as increased variations were noted in the η–θ torsions of U33, U35 and U36 (Figure 4A.13, 4A.14).

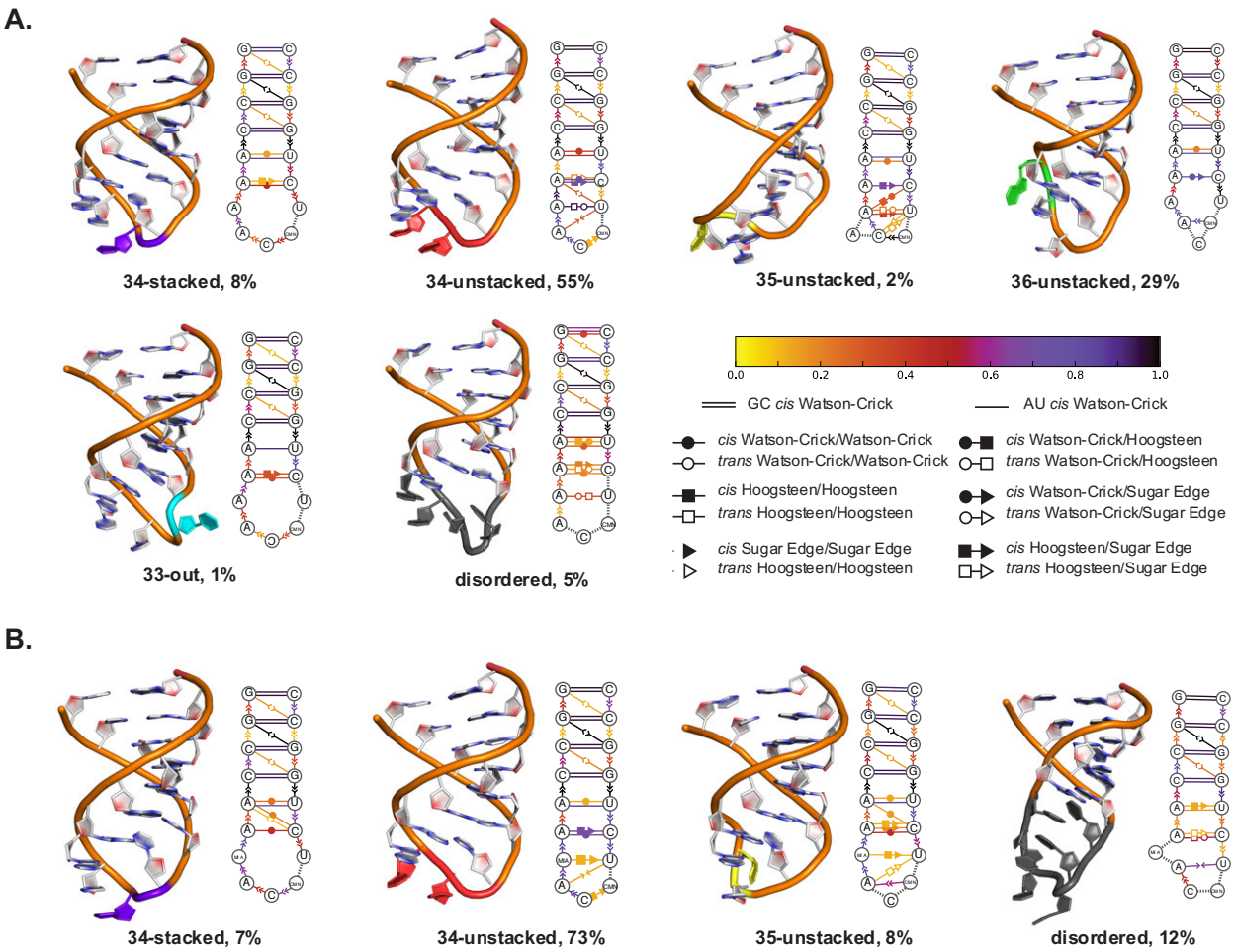


Figure 4.4 – Conformational profiles for the ASL of modified tRNA^{Trp}

(A) Structural representation and density maps of torsion angles adopted by cmnm^{5}U34 across ten 500-ns trajectories. (B) Structural representation and density maps of torsion angles adopted by ms^{2}i^{6}A37 across ten 500-ns trajectories.

The addition of ms^{2}i^{6}A37 adjacent to the anticodon bases reduced the flexibility and amplitude of motion introduced by cmnm^{5}U34 and stabilized the backbone conformation of the entire loop. Specifically, the accommodation of the N6-substituent of ms^{2}i^{6}A37 within the loop restricted the mobility of U33 and promoted the formation of Watson-Crick hydrogen bonds between U33 and A37 (Figure 4A.15). Non-covalent interactions between ms^{2}i^{6}A37 and U33 also facilitated better accommodation of flexible bases, especially 36, which was unstacked for 20% and 29% of the time in unmodified tRNA^{Trp} and cmnm^{5}U34-

$tRNA^{Trp}$, respectively, but remained stacked with A37 ~ 85% of the time in the doubly modified tRNA system (Figure 4.4).

In summary, despite the inherent flexibility of the modified nucleotides, $cmnm^5U34$ and ms^2i^6A37 do not alter the global fold of $tRNA^{Trp}$. $cmnm^5U34$ reduces backbone dynamics at the wobble base and decreases the predisposition for a disordered loop conformation. Nevertheless, the modification increases flexibility at its flanking bases. ms^2i^6A37 reduces these dynamics by instigating structural changes within the backbone at position 37 and enhancing nucleotide–nucleotide stacking and hydrogen bonding interactions within the anticodon loop. Although previous computational studies on $cmnm^5U$ and ms^2i^6A highlighted the dynamic nature of these nucleotides,[25,44] the present work uncovers new atomic-level details on the additive impacts of these modifications within the context of $tRNA^{Trp}$, which carries these modifications *in vivo*.

4.3.2 $cmnm^5s^2U34$ increases the flexibility of the anticodon loop of $tRNA^{Lys}$, while t^6A37 counters the effects of $cmnm^5s^2U34$ and reduces dynamics within the ASL

The 2-thio-modification of $cmnm^5U$ ($cmnm^5s^2U$; Figure 4.7) was first isolated in *Bacillus subtilis* $tRNA^{Lys}$ and was later detected in human mt-$tRNA^{Lys}$ and mt-$tRNA^{Gln}$.[6,17,56] Like its parent modification, this wobble uridine is always accompanied by an adjacent modification at position 37, and sequence analyses revealed t^6A37 to be the most prevalent modification in tRNAs that have $cmnm^5s^2U34$ *in vivo*.[4,57] Arising from the addition of a ureido-threonyl group at N6 of A37, t^6A37 is larger and more flexible than ms^2i^6A37 (Figure 4.7). Multiple studies have also reported that t^6A37 improves translation efficiency.[15,19,58,59] In particular, cellular reduction of t^6A was linked to reduced reading of AGR and CGN codons, protein folding defects and increased sensitivity to cellular stresses.[19,60-62] A recent study by Prabhakar *et al.* revealed that t^6A37 enhances non-covalent interactions between A37 and neighboring bases, thereby altering the structural arrangement within the ASL.[45] Nonetheless, the additive impact of t^6A37 and wobble uridine modifications within the context of tRNA remains elusive. To understand the effect of $cmnm^5s^2U34$ on the tRNA structure, the tRNA was singly modified at position 34

(cmnm^{5}s^{2}U34-tRNA^{Lys}). Subsequently, t^{6}A was added at position 37 to gain insight into the combinatory effect these two modifications have on the molecular structure of tRNA^{Lys} (cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}).

In the absence of modifications, tRNA^{Lys} was stable throughout the 10-replica ensemble, as evidenced by a low average RMSD range of 3.0 to 3.5 Å (Figure 4A.16) relative to the experimental starting structure. Furthermore, the unmodified tRNA maintained its canonical L-shape as the non-covalent interactions that govern the global fold of tRNA^{Lys} were persistent throughout the trajectories (average stacking occupancies > 85%, average hydrogen bond occupancies > 85%). The coaxial stacks of the D-ASL domains and TψC-acceptor stem were also sustained (Figure 4.5, 4A.17), maintaining the tertiary structure of tRNA^{Lys}. RMSF analysis for each residue averaged over the replica ensemble revealed higher stability in the stem regions (RMSFs < 3 Å) relative to the loop regions (RMSFs between 3 and 7 Å). In particular, the anticodon loop of unmodified tRNA^{Lys} was found to be highly flexible, with the anticodon bases being the most dynamic regions in the loop. Torsional analyses of the backbone atoms within the ASL concurred with these observations as they uncovered high fluctuations in the η and θ angles of U34, U35 and U36 relative to other ASL nucleotides (Figure 4A.18).

Large structural fluctuations were observed within the ASL over the course of the replica simulations (Figure 4A.17). At the stem, Watson-Crick hydrogen bonding interactions were prevalent over 80% of the simulation time and stacking interactions were persistent between the stem base pairs over 80% of the time, preserving the ASL helix. Stacking interactions were also observed within the anticodon loop, but interaction occupancies varied between stacking pairs. Little stacking (~ 3%) was detected between U33 and U34, implying that the U-turn is maintained. Other stacking pairs in the loop showed varied propensities, and average interaction occupancy ranged from 15 to 80%, with the least persistent stacking interactions being observed between the anticodon bases (U34 to U36; average occupancy between 15% to 36%), highlighting the dynamic nature of the ASL at these positions.

To provide points of reference for understanding the impact of modified nucleobases at positions 34 and 37, the dynamics of U34 and A37 were further considered. In unmodified tRNA^{Lys}, U34 adopted an *anti* conformation around its glycosidic bond (average χ torsional angle of 208 ± 32°; Figure 4A.19). The backbone of U34 was very dynamic, particularly at its 3' and 5' ends, as shown by the wide ranges of

the α, ε, and ζ angles the nucleotide adopted (25° < α < 326°, 31° < ε < 319°, 5° < ζ < 339°). Compared to U34, A37 was relatively stable within the anticodon loop and no significant fluctuations were detected in the backbone torsions at this position (Figure 4A.18, 4A.20). A37 adopted an *anti* conformation around its glycosidic bond (average χ torsional angle of 191 ± 15°; Figure 4A.20) and formed stacking interactions with its neighboring bases A36 (30%) and A38 (75%) averaged over the course of the replica ensemble.

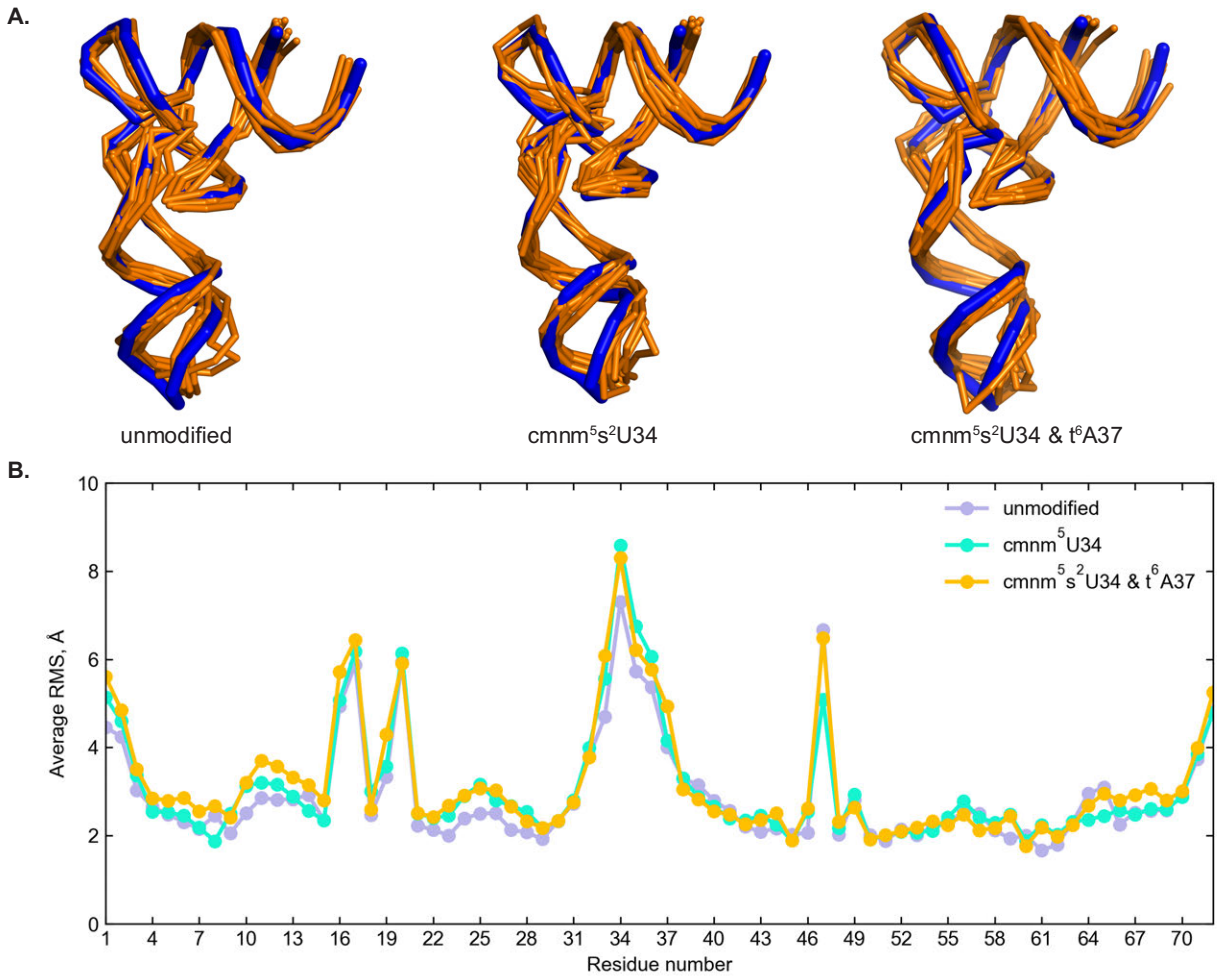


Figure 4.5 – Full tRNA dynamics across the tRNA^{Lys} systems

(A) Representative structure overlays (orange ribbons) from replica trajectories of unmodified tRNA^{Lys}, cmnm^{5}s^{2}U34-tRNA^{Lys} and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}. The reference structure is the experimental crystal structure (PDB ID: 1FIR, blue). (B) Average per residue fluctuations in unmodified tRNA^{Lys} (lavender), cmnm^{5}s^{2}U34-tRNA^{Lys} (teal) and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} (yellow). The ASL domain spans from residue 27 to 43.

Conformational analysis of unmodified $tRNA^{Lys}$ (Figure 4.6) revealed that the unmodified ASL adopted four conformations – 34-stacked (25%), 34-unstacked (10%), 36-unstacked (42%) and disordered (23%). According to this analysis, the most mobile base in the $tRNA^{Lys}$ ASL was U36, as the base did not form stacking interaction with its neighboring bases for the majority of the simulation time.

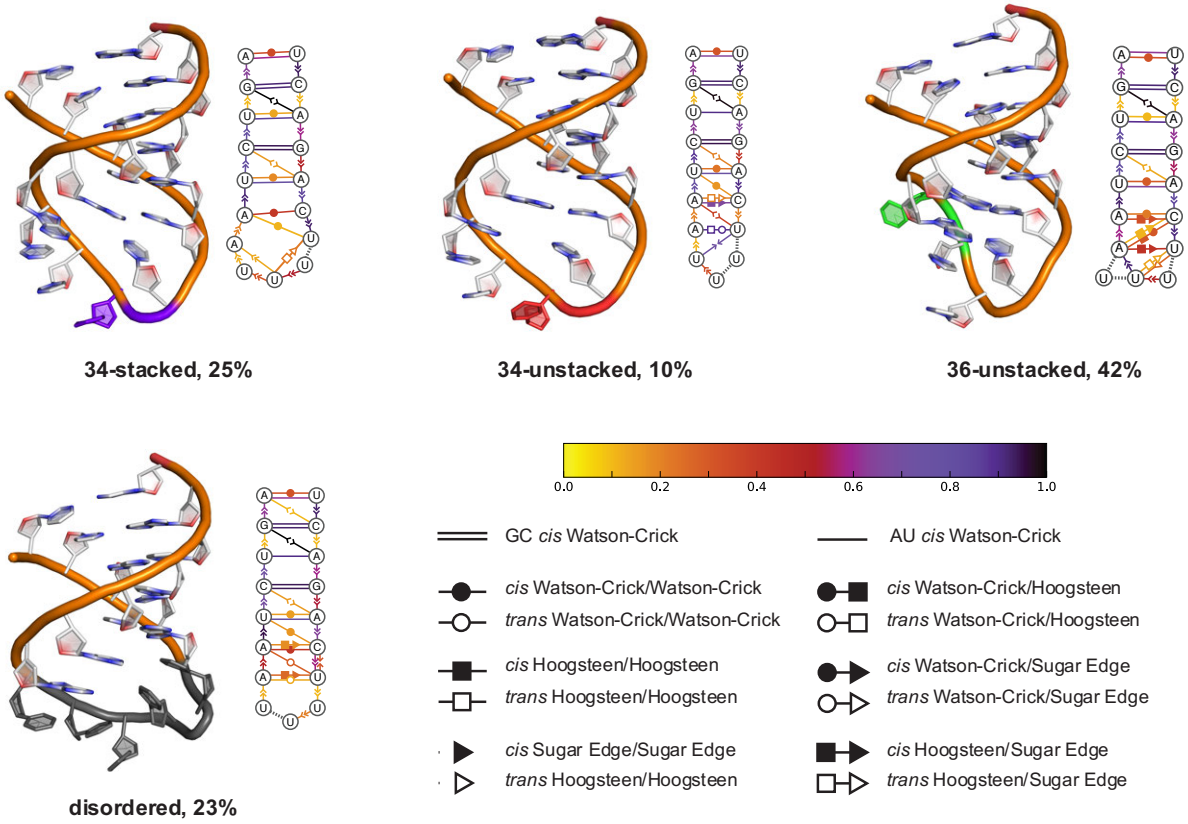


Figure 4.6 – Conformational profile of the ASL of unmodified $tRNA^{Lys}$

Cartoon and secondary structure representations of conformational states adopted by the unmodified tRNATrp ASL. Conformations are defined and color-coded as in Chapter 3. Non-covalent interactions are denoted using the Leontis-Westhof notation for RNA molecules.[63,64]

Nevertheless, the displacement of U36 did not completely disrupt the global structure of the ASL as new non-covalent interactions were observed within the loop (the 35/36 and 36/37 stacking interactions were replaced by 35/37 interaction). U34 also adopted multiple positions throughout the trajectories and remained unstacked 10% of the simulation time. Interestingly, like $tRNA^{Trp}$, the anticodon loop of $tRNA^{Lys}$ is prone to disorder, with a lack of structural arrangement being observed 23% of the time. The presence

of the disordered conformation in unmodified tRNAs underscores the dynamic nature of the anticodon loop and highlights the importance of modifications *in vivo*.

The presence of cmnm^{5}s^{2}U34 and t^{6}A37 did not significantly alter the overall dynamics of tRNA^{Lys}, and on average, replica ensembles exhibited similar RMSDs relative to unmodified tRNA^{Lys} (average RMSDs of 3.3 ± 0.5, 3.5 ± 0.6 and 3.7 ± 0.6 Å for the unmodified, cmnm^{5}s^{2}U34-tRNA^{Lys} and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}, respectively; Figure 4A.16). The time averaged RMSF analyses of the three tRNA^{Lys} systems revealed similar dynamics within the four main domains of the molecule, confirming that the modifications at the anticodon loop do not alter the global structure of tRNA (Figure 4.5). Additionally, the network of tertiary interactions at the tRNA elbow remain prevalent over the course of all replica simulations, and neither modified system showed large-scale deviations from the unmodified tRNA (Figure 4A.17). Overall, the modified tRNA^{Lys} systems maintained the global tRNA^{Lys} structure, and no long-range modification effects were observed in this study.

Nevertheless, structural analyses of each replica ensemble revealed substantial and divergent dynamics at the ASL for of tRNA^{Lys} system, indicating that cmnm^{5}s^{2}U34 and t^{6}A37 differentially alter the structure of this domain. Although hydrogen bonding within the ASL helix was well preserved across the three tRNA^{Lys} models, interactions in the anticodon loop were enhanced in the presence of cmnm^{5}s^{2}U34 and t^{6}A37 (Figure 4A.17). In particular, the U33–A37 base pair, which was non-existent in the unmodified system was moderately present within the cmnm^{5}s^{2}U34-tRNA^{Lys} and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} systems (45% occupancy in both systems). Hydrogen bonding between U32 and A38 was mildly enhanced in the presence of both modifications, and the occupancy for these interactions increased from 15% in the unmodified model to ~ 20% in the modified systems. The strengthening of hydrogen bonding interactions extended to the first base pair in the stem region, as the U31–A39 base pair was also more prevalent in the presence of cmnm^{5}s^{2}U34 and t^{6}A37 (89% occupancy in unmodified tRNA^{Lys} compared to 98% and 94% in cmnm^{5}s^{2}U34-tRNA^{Lys} and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}, respectively). In contrast to hydrogen bonding, stacking propensities within the ASL differed from one modified system to another.

Although stacking interactions within the ASL helix of modified tRNA^{Lys} were well preserved (occupancies > 85% in cmnm^{5}s^{2}U34-tRNA^{Lys} and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}), the 32/33, 36/37 and

37/38 stacking interactions reduced conspicuously relative to unmodified $tRNA^{Lys}$ (32/33 stack: 68% to 48%, 36/37 stack: 32% to 17%, 37/38 stack: 77% to 65 in the unmodified and $cmnm^5s^2U34$-$tRNA^{Lys}$ systems, respectively), indicating that $cmnm^5s^2U34$ increases structural mobility within the loop region (Figure 4A.17). In general, the addition of t^6A at position 37 restored and enhanced the stacking interactions impeded by $cmnm^5s^2U34$, stabilizing the anticodon loop in an organized state. In singly modified tRNA, except for the 34/35 stack, the occupancies for anticodon stacks were lower than unmodified $tRNA^{Lys}$ (34/35: 42% and 36%, 35/36: 10% and 15%, 36/37 stack: 17% and 32%, respectively), but these interactions were increased or reinstated in the $cmnm^5s^2U34/t^6A37$-$tRNA^{Lys}$ system (34/35: 50%, 35/36: 21%, 36/37 stack: 34%). However, stacking interactions with t^6A37 (36/37 and 37/38) and U33 (32/33) are impeded by the modified nucleobase to accommodate for hydrogen bonding interactions at this site. Diverse stacking and hydrogen bonding within the anticodon loop did not disrupt its distinctive U-turn motif, as no stacking was observed between residues 33 and 34 of both modified models. Analyses of non-covalent interactions within the ASL of $tRNA^{Lys}$ reinforce the previous observations regarding the differential effects of $cmnm^5s^2U34$ and t^6A37 on the region. On one hand, the wobble uridine modification increases hydrogen bonding within the loop, but impedes stacking interactions with its neighboring bases. On the other hand, the subsequent addition t6A37 increases dynamics at U33, but strengthens stacking interactions between the anticodon bases (34 to 36), stabilizing their position within the loop.

Similarly to U34 in unmodified $tRNA^{Lys}$, $cmnm^5s^2U34$ preferably adopted an *anti* conformation (~ 95%) around its glycosidic bond ($180° < \chi < 270°$; Figure 4A.19, 4A.21). Interestingly, although the presence of t^6A in the anticodon loop did not change the glycosidic preference of $cmnm^5s^2U34$, an increased *syn*-identity was observed in the $cmnm^5s^2U34/t^6A37$-$tRNA^{Lys}$ system (~ 4:1 *anti–syn* ratio), indicating that the modification at position 37 may influence the positioning of the wobble uridine base within the loop. The thiolation of $cmnm^5U$ altered the backbone configuration at position 34 in both modified systems, especially at the α and δ angles that exhibited different dihedral distributions relative to the unmodified uridine (Figure 4A.22). Recall that $cmnm^5U34$ did not affect the backbone dihedral angles in $tRNA^{Trp}$. Therefore, this different effect is likely due to thiolation at C2 of the uridine base. Be that as it may, the thiol group did not influence the side chain torsions of the C5 substituent, and similar torsions

were observed in $tRNA^{Trp}$ and $tRNA^{Lys}$ systems. As before, two conformers were adopted with respect to the C5–C7 linkage (conformer-I: 46%, conformer-II: 54%; Figure 4A.23). Other side chain torsions adopted similar angles as the unthiolated uridine and no structural deviations were observed between the C5-substituent of $cmnm^5U$ and $cmnm^5s^2U$ at position 34 of $tRNA^{Trp}$ and $tRNA^{Lys}$, respectively. More importantly, the C5-substituent of $cmnm^5s^2U34$ formed similar hydrogen bonding interactions with the phosphate sugar backbone as its unthiolated counterpart, accentuating the similarities between the two nucleotides (Figure 4A.23).

The addition of the ureido threonyl group at A37 altered the backbone torsions at this location. In particular, α and γ angles of t^6A37 differed from those observed in unmodified $tRNA^{Lys}$. In the unmodified model, α angles ranged between 240 and 360° with a mean value of 300° (Figure 4A.20). In contrast, t^6A37 exhibited a broad range of dihedral angles (0 to 360°), highlighting the flexible nature of this base. A similar phenomenon was observed with the γ torsion for t^6A37 (Figure 4A.25, 4A.26). In the unmodified system, γ angles generally ranged between 0 and 120° with a mean value of 60°. In contrast, t^6A37 adopted two γ conformers, the first conformer ($0° < \gamma < 120°$) occupied ~ 57% of the simulation time, while the second ($120° < \gamma < 240°$) was observed ~ 43% of the time. It is important to note that despite its flexible nature, t^6A37 never moves out of the loop, implying that motion in this nucleobase primarily affects other members of the anticodon loop. A previous study on the effects various modifications have at position 37 of $tRNA^{Phe}$ revealed that in the absence of other modifications, the N6-substituent of t^6A and its derivatives adopts a single conformation with respect to the N6–C10 bond, defined by the torsional angle $\mu = \angle(C6N6C10C11)$.[45] In contrast, the present study reveals that in the presence of $cmnm^5s^2U34$, the bulky moiety of t^6A37 can adopt two main conformations with respect to the linking bond. The first conformer ($150° \leq \mu \leq 180°$) was sampled ~ 54% of the time, while conformer II ($180° \leq \mu \leq 210°$) populated ~ 46% of the simulation time (Figure 4A.27). As such, this investigation provides evidence that bulky modifications can influence the positioning of other modified nucleotides within the anticodon loop.

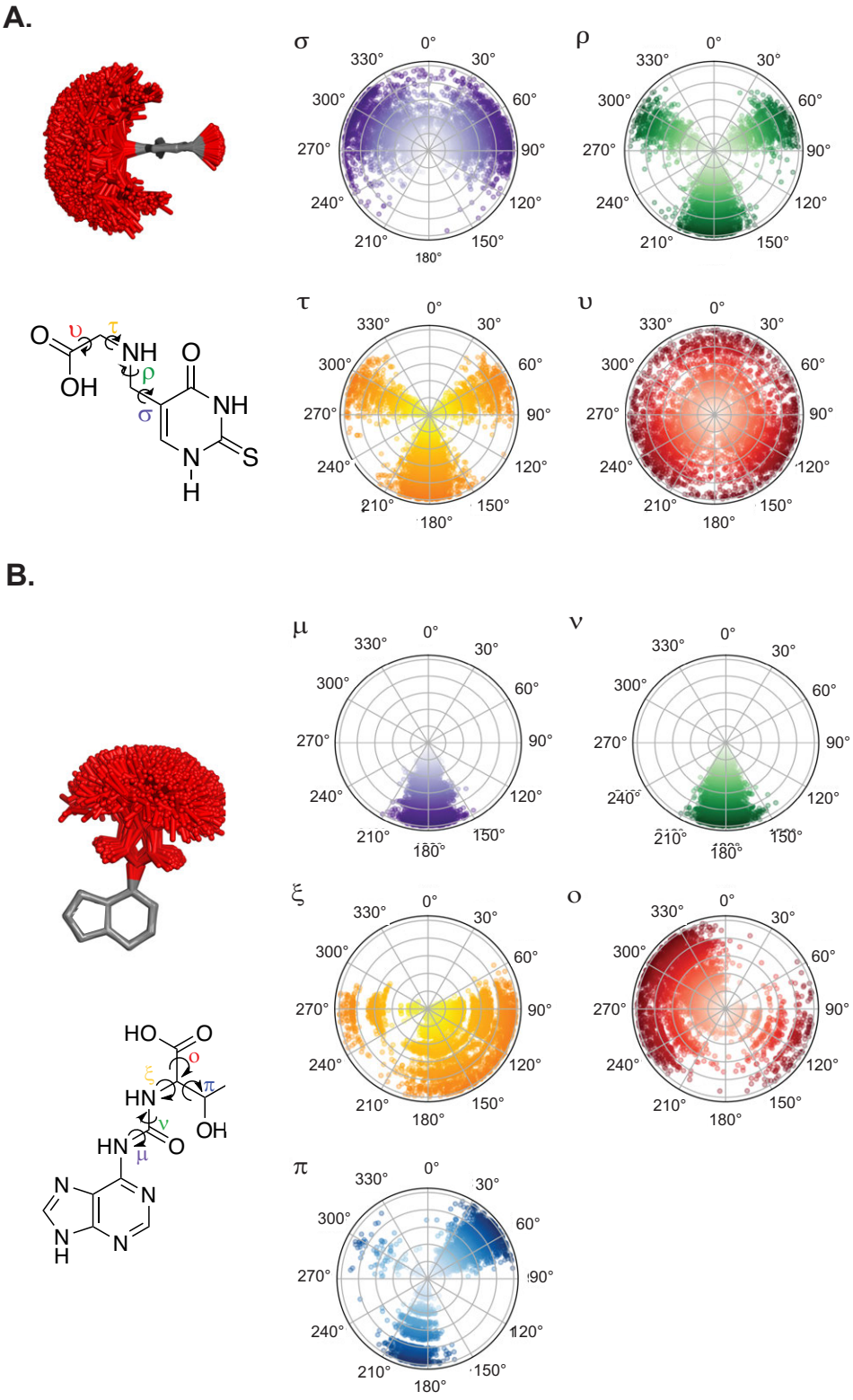


Figure 4.7 – Sidechain flexibilities of cmnm^{5}s^{2}U34 and t^{6}A37

Structural representation and density maps of torsion angles adopted by cmnm^{5}s^{2}U34 (A) and t^{6}A37 (B) across ten 500-ns trajectories. Simulation time (5 μs) is represented on the r-axes of each polar plot.

Conformational analysis of the ASL from the modified $tRNA^{Lys}$ systems revealed that, like $cmnm^5U34$, $cmnm^5s^2U34$ greatly increased flexibility within the anticodon loop. Recall that in the absence of modifications, the ASL of $tRNA^{Lys}$ is highly dynamic and susceptible to disorder, with the disordered state being observed ~ 23% of the time. Moreover, U36 is particularly flexible and moves away from its starting position in the $tRNA^{Lys}$ crystal structure ~ 40% of the simulation time. In the presence of $cmnm^5s^2U34$, U36 remained highly dynamic (36-unstacked: 40%) and increased flexibility was observed at position 35 (35-unstacked: 13% in $cmnm^5s^2U$-$tRNA^{Lys}$ compared to 0% in unmodified $tRNA^{Lys}$; Figure 4.8). Moreover, the ASL was more prone to disorder in the $cmnm^5s^2U$-$tRNA^{Lys}$ system (32% compared to 23% in unmodified $tRNA^{Lys}$), suggesting that $cmnm^5s^2U34$ increases the flexibility of loop bases and hinders overall organization within the anticodon loop. Interestingly, despite the wide-ranging disruptions caused by this modification, the nucleobase at position 34 remained relatively stable over the course of the replica ensemble, and $\eta-\theta$ analyses of the loop region confirmed that motion at the backbone of this base was restrained, in comparison to its unmodified counterpart (Figure 4A.28). Therefore, like $cmnm^5U34$, the function of $cmnm^5s^2U34$ is to reduce the mobility of the wobble base.

The addition of t^6A37 adjacent to the anticodon bases reduced the dynamics induced by the wobble modification. In comparison to the $cmnm^5s^2U34$-$tRNA^{Lys}$ model, the doubly modified $tRNA^{Lys}$ system exhibited a decrease in dynamics and variation in conformational sampling (34-stacked: 47%, 36-unstacked: 28%, disordered: 25%; Figure 4.8). Like ms^2i^6A37 in $tRNA^{Trp}$, the presence of the N6-substituent of t^6A37 restricted the movement of U33 within the loop, priming it for hydrogen bonding interactions with the sugar-phosphate backbone and the Watson-Crick face of A37 (Figure 4A.29). Despite the stabilizing effects of t^6A37, the anticodon loop of $tRNA^{Lys}$ remained susceptible to disorder, and the pseudorotational backbone profile for this system showed high ranges of motion within the loop region, especially in bases 32, 35 and 36 (Figure 4A.30). What is more, the N6-substituent of t^6A is bulky and very flexible and its motion may also impede organization within the anticodon loop and increase flexibility in this region.

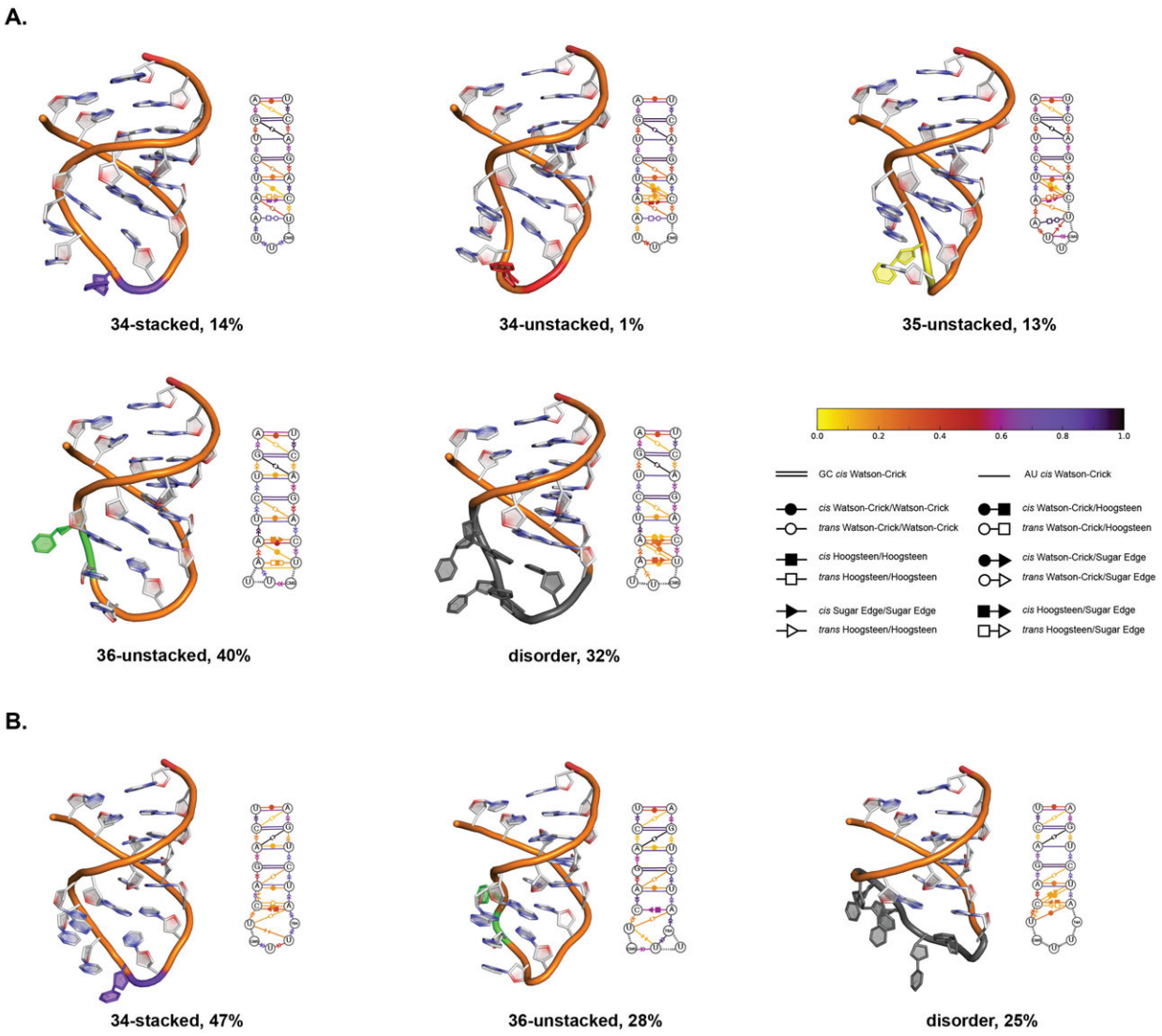


Figure 4.8 – Conformational profiles for the ASL of modified tRNA^{Lys}

Cartoon and secondary structure representations of conformational states adopted by the cmnm^{5}s^{2}U34-tRNA^{Lys} (A) and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} (B) ASL. Conformations are defined and color-coded as in Chapter 3 and non-covalent interactions are denoted using the Leontis-Westhof notation for RNA molecules.

In summary, while long range structural effects were not observed in cmnm^{5}s^{2}U-modified tRNA^{Lys} systems, the substituents of cmnm^{5}s^{2}U34 were found to increase the amplitude of motion within the anticodon loop, particularly at other anticodon positions. Be that as it may, like its unthiolated equivalent, cmnm^{5}s^{2}U34 restrains backbone flexibility at the wobble base and stabilizes its position within the anticodon loop.

The molecular-level analyses presented in this investigation reveal that t^{6}A37 has an antagonistic effect to

cmnm^{5}s^{2}U34, as it restores some stability to the anticodon loop, enhancing stacking and hydrogen bonding interactions within the region. *In vivo,* tRNA^{Lys} usually has a pseudouridine modification at position 38 or 39 and this versatile modification has been found to stabilize single- and double-stranded RNA molecules, including the anticodon tRNA^{Lys}, tRNA^{His} and tRNA^{Tyr}.[65-67] Consequently, in the future, it will be worthwhile to investigate how these three modifications work in tandem to further stabilize the ASL of tRNA^{Lys}.

4.3.3 Posttranscriptional modifications in the anticodon loop predispose the loop towards the adoption of wobble base conformations

The present investigation uncovers a wide range of conformations adopted by the ASL domain of tRNA in the absence and presence of posttranscriptional modifications at positions 34 and 37. In general, the structural states isolated in this study can be grouped into four categories based on the location of the dynamic base within the anticodon loop. These are: (i) the wobble base conformation (WB), which describe motions at position 34, (ii) the 3' anticodon base conformation (3'-AC) that encompass the flexibilities at positions 35 and 36, (iii) the anticodon flanking base conformation (FB), which describes large fluctuations in residues 33 and 37, and (iv) the disordered loop (DL) conformation that is defined by all disordered states adopted by the loop region (Figure 4.9). The DL conformation was sampled ~ 20% of the time across all six systems studied, highlighting the inherent flexibility of the anticodon loop region (Figure 4.10). Nonetheless, the extent of nucleotide dynamics varied from one position to another. Across the six systems examined in this study, FB conformations were the least prevalent conformations sampled (identified in one of the six systems with an overall occupancy of 0.2%), indicating that bases at these positions are relatively stable within the ASL. In contrast, regardless of sequence, large fluctuations were observed in all three anticodon bases (34, 35 and 36). 3'-AC conformations were sampled across all six tRNA systems with an occupancy of 31%, while wobble base conformations were identified in all tRNA systems at an occupancy of 50%. The frequent presence and sampling of 3'-AC and WB conformations highlight the variability of this region and suggest that structure of the anticodon loop allows for dynamics within the anticodon bases.

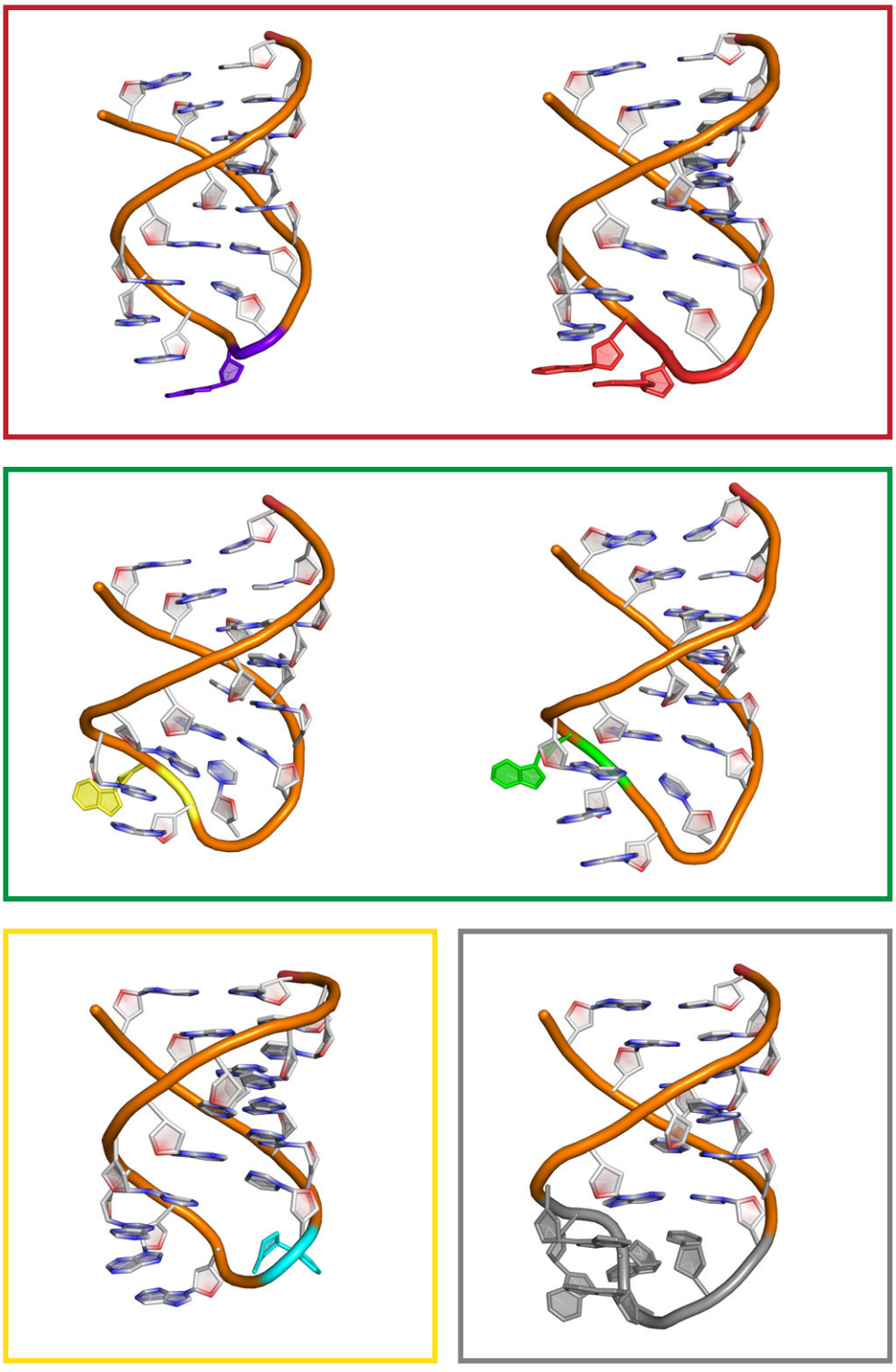


Figure 4.9 – Dynamics at the anticodon loops of tRNAs

The ASL conformational states identified in the present work have been grouped into four categories based on the location of the mobile base relative to crystal structure references. These groups are the wobble base (WB; red), 3' anticodon bases (3'-AC; green), anticodon flanking bases (FB; yellow), and disordered loop (DL; grey).

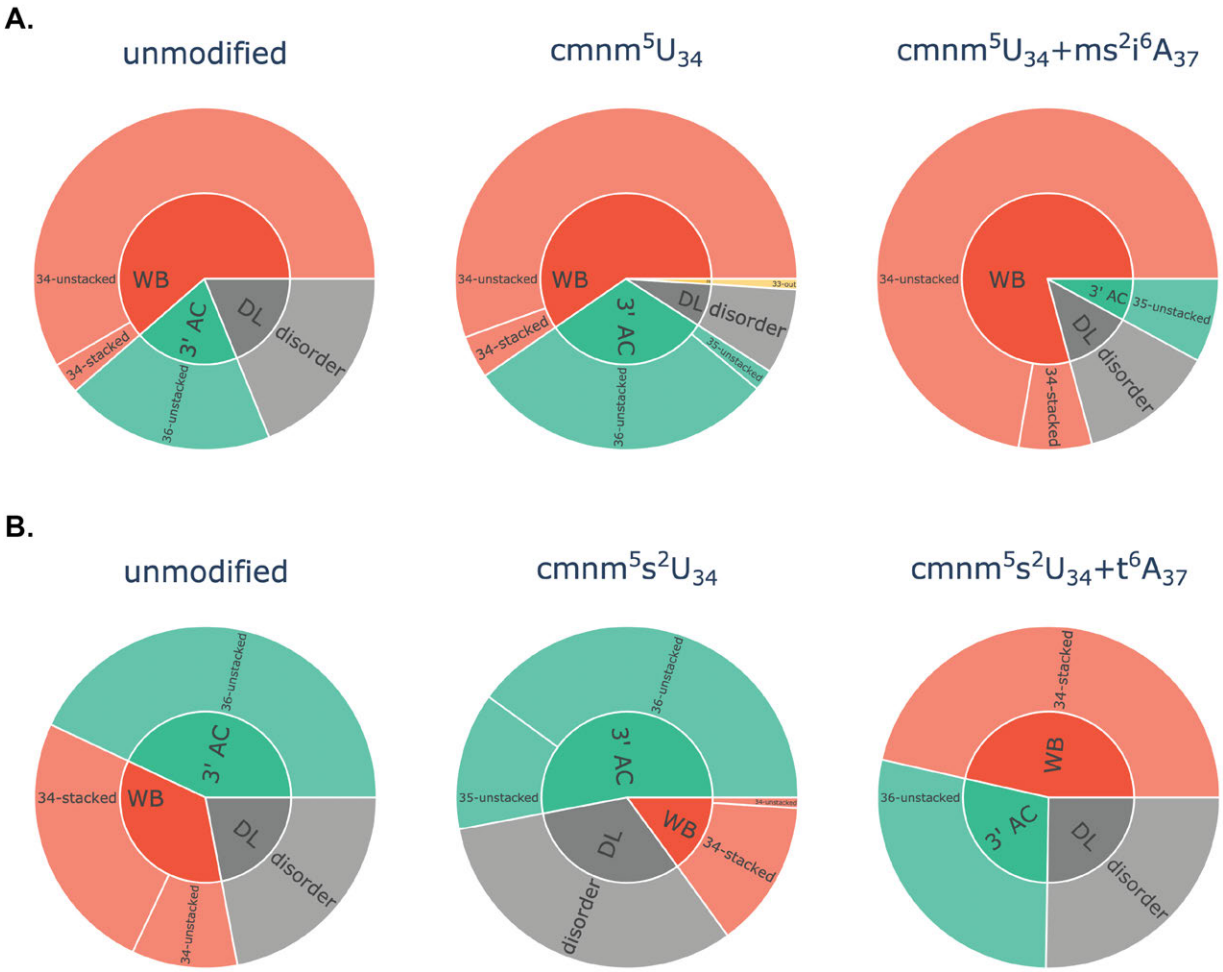


Figure 4.10 – Conformational analysis summary

Stabilization of WB conformation in modified $tRNA^{Trp}$ (B) and $tRNA^{Lys}$ (C) systems . Conformational groups are color-coded according to the notation in Figure 4.9.

Interestingly, posttranscriptional modification of positions 34 and 37 generally reduce disorder and flexibility within the ASL, and lead to the increased adoption of WB conformations. For instance, in $tRNA^{Trp}$, the WB conformations were sampled 80% of the time in the doubly modified ASL relative to 62% in the unmodified structure, while disorder was reduced to 13% in the cmnm^{5}U34/ms^{2}i^{6}A37-$tRNA^{Trp}$ system compared to ~ 20% in the unmodified system (Figure 4.10). A similar trend was observed for $tRNA^{Lys}$, wherein WB conformations were sampled for 35% of the time in the unmodified model and 45% in the cmnm^{5}s^{2}U34/t^{6}A37-$tRNA^{Lys}$ model. cmnm^{5}s^{2}U34 and t^{6}A37 did not reduce the presence of the disordered loop within $tRNA^{Lys}$ (22% and 25% in the unmodified and doubly modified systems, respectively). However, reduced flexibility was observed in other parts of the anticodon loop, particularly

U36, in the cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} model relative to the unmodified system (43% and 28% in the unmodified and doubly modified systems, respectively). This bias towards the adoption of WB conformations in modified ASLs of different tRNAs suggests that these conformations may be functionally advantageous *in vivo*.

The structure of the ASL directly affects tRNA function in translation.[68-70] Specifically, the backbone conformation of the ASL determines how well the tRNA is accommodated within the smaller ribosomal subunit, which affects the fidelity of mRNA–tRNA interactions during translation.[71-73] To determine whether WB conformations are functionally advantageous over other ASL states, the representative structures of all ASL states isolated in this study were compared to experimentally observed ASL structures of enzyme-bound and ribosome-bound tRNA molecules. Elongator factor Tu (EF-Tu) was chosen as the enzyme for this comparison, as it is involved in the conformational proofreading of aminoacylated tRNAs during translation.[74-76] EF-Tu was bound to modified tRNA^{Phe} (PDB ID: 1TTT), while unmodified tRNA^{fMet} was bound to the A-site of *E. coli* ribosome (PDB ID: 6WD0). The ribosome-bound tRNA^{fMet} also base paired with an mRNA molecule. The reference ASL structures compared in this study have an open-loop state, which is defined by the potential to form two Watson-Crick hydrogen bond pairs (N32–N38 and N33–N37) within the anticodon loop. This open-loop structure has been proposed to be important for ribosomal binding during translation.[77-80]

Relative to other conformational groups isolated in this study, the backbone orientations of WB conformations best emulate those of functional tRNA molecules. This is shown by their small RMSDs (as low as 1.5 Å and 1.2 Å across the tRNA^{Trp} and tRNA^{Lys} systems, respectively) relative to the experimentally observed ASL structures (Figure 4.11, 4A.31-32, Table 2-5). Interestingly, the open-loop state observed in the reference ASL structures were only observed in the WB conformations of tRNA^{Trp} and tRNA^{Lys}. Moreover, doubly modified tRNA systems adopted this open-loop state more frequently than their unmodified and singly modified counterparts. In the cmnm^{5}U/ms^{2}i^{6}A37-tRNA^{Trp} system, the open-loop structure was observed 80% of the time compared to 60% and 63% in unmodified and cmnm^{5}U34-tRNA^{Trp}, respectively.

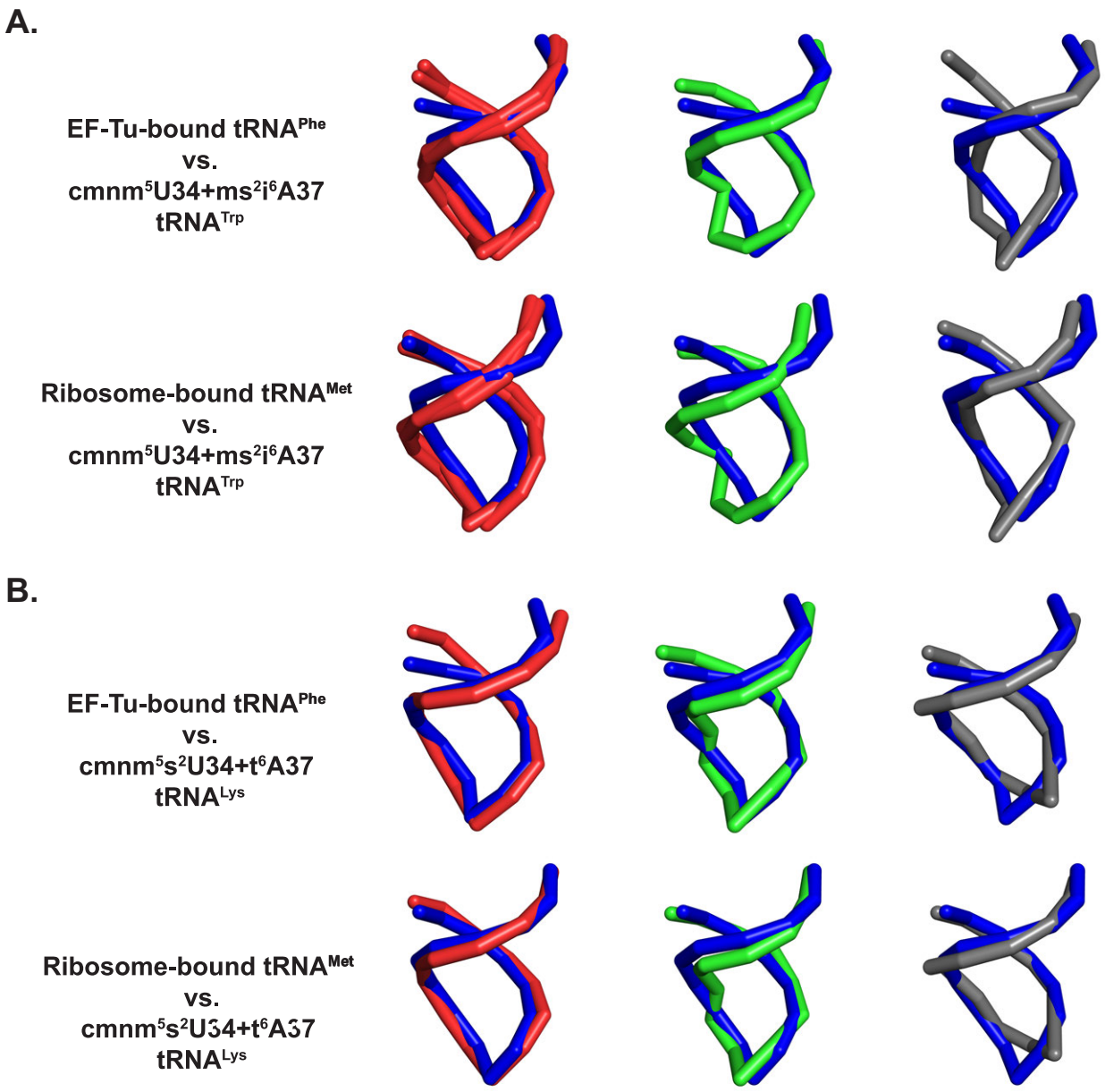


Figure 4.11 – Comparison of simulated ASL states to experimentally-derived functional states

Backbone comparisons of doubly modified $tRNA^{Trp}$ (A) and $tRNA^{Lys}$ (B) against. EF-Tu-bound (top, PDB ID: 1TTT) and ribosome-bound (bottom; PDB ID: 6WD0) tRNA. The experimental reference structures are colored blue, while the WB, FB and DL conformational groups are represented in red, green, and grey, respectively.

Similarly, in $cmnm^5s^2U34/t^6A37$-$tRNA^{Lys}$, the open-loop state was maintained ~ 50% of the time, compared to 35% and 15% in unmodified and $cmnm^5s^2U34$-$tRNA^{Lys}$, respectively. In contrast, the presence of dynamics at all other nucleobases of the anticodon loop significantly distorts the backbone of the ASL relative to the reference structures (overlay RMSDs ranged from 2.0 to 4.3 Å across $tRNA^{Trp}$

systems and 1.9 to 3.9 Å in $tRNA^{Lys}$ models). Consequently, it can be concluded that $cmnm^5(s^2)U34$ modifications and their accompanying modifications at position 37 stabilize functional ASL states. More importantly, similarities between the WB conformations and the ASLs from protein- and ribosome-bound tRNAs indicate that these modifications preorder the anticodon bases prior to translation, and may improve translation accuracy and efficiency by reducing conformational variance at the ASL reading interface.

4.4 Conclusions

The present work uses replica molecular dynamics simulations to study the structural effects of $cmnm^5(s^2)U$ wobble modifications on the ASL and the tRNA at large. Two tRNA sequences were considered to investigate the impact of this modification family and the partner modifications at position 37. Studies on completely unmodified $tRNA^{Trp}$ and $tRNA^{Lys}$ revealed subtle variations in the dynamics with sequence. Nevertheless, irrespective of sequence context, the presence of single and double modifications in the anticodon loop did not affect the three-dimensional fold of the tRNA nor alter inter-domain interactions, indicating that the modifications do not have long-range structural effects on tRNA. While the general structure of the ASL was maintained over the course of all simulations, the anticodon loops of $tRNA^{Trp}$ and $tRNA^{Lys}$ adopted different configurational states to accommodate the diversity in sequence and modification. In general, $cmnm^5(s^2)U$ modifications stabilized backbone and nucleobase dynamics at position 34, but increased the dynamics and flexibility of other nucleotides within the anticodon loop. The presence of ms^2i^6A37 and t^6A37 did not restrict the movement of $cmnm^5(s^2)U$ modifications at position 34. However, these modified nucleotides afforded stability to the remainder of the anticodon loop by altering non-covalent interactions within the loop and promoting the formation of an open-loop ASL structure – a proposed requirement for tRNA binding to the ribosome. Additionally, the combined presence of $cmnm^5(s^2)U34$ and A37 modifications predisposed the ASL toward the adoption of functionally favored states, highlighting their importance *in vivo*. Specifically, the decrease of disordered states and reduced flexibility at positions 35 and 36 drive the ASL to adopt wobble base conformations, which highly mimic functional ASL states isolated in experimental tRNA structures. In summary, results from this investigation provide the previously lacking molecular-level details for how $cmnm^5U34/ms^2i^6A37$

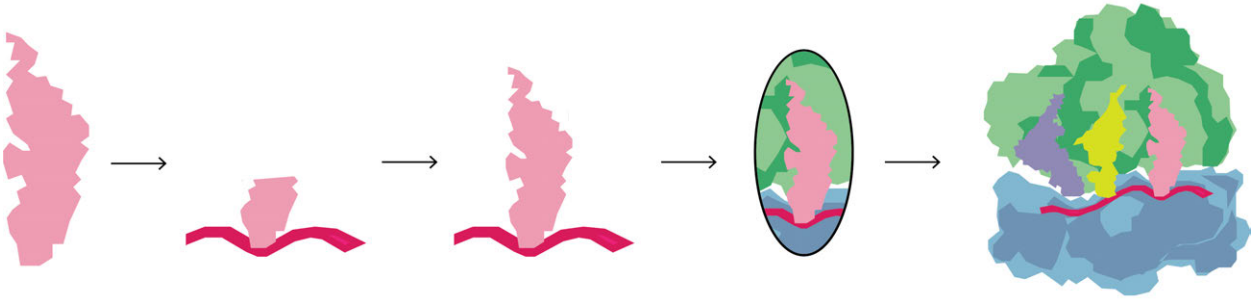
and cmnm^{5}s^{2}U34/t^{6}A37 modifications work in tandem to preorder the anticodon loop for protein synthesis, ensuring proper tRNA accommodation within the ribosome and promoting accurate mRNA–tRNA interactions during translation.

4.4.1 Future directions

Although the present work contributes to our understanding of the structural effects of cmnm^5(s^2)U34 modifications on tRNA, questions pertaining to other modifications in this family remain. Recall that in bacteria, cmnm^{5}U34 acts as an intermediate during the biosynthesis of 5-methylaminomethyluridine (mnm^{5}U34) and 5-aminomethyluridine (nm^{5}U34),[81,82] while the cmnm^5(s^2)U34 homologs 5-taurinomethyl(-2-thio)uridine (τm^5(s^2)U34) have been found in mitochondrial tRNAs (Figure 1.2).[5,8,83,84] Unfortunately, although τm^5(s^2)U34 have been linked to mitochondrial diseases[5,9,11,48,85,86] and (m)nm^{5}U34 have been associated with oxidative phosphorylation diseases,[87,88] little is known about the structural and functional effects of these modifications on tRNA. Using MD simulations to probe these modifications in tRNA will provide insight into their structural dynamics within biologically relevant environments and strengthen our understanding of this modification family in general. Notably, the results from the present study reveal that small changes on the atomic level can lead to large local structural changes. Future work will expand on this idea and reveal the effects of sequentially removing (i.e., cmnm^{5}U → mnm^{5}U → nm^{5}U) or substantially altering (i.e., cmnm^{5}U → τm^{5}U) the chemical groups from the C5 substituent in this modification family.

This study reveals that cmnm^5(s^2)U34 and their accompanying A37 modifications predispose the ASL towards conformations found in ribosome-bound tRNA. Nevertheless, other conformational states were also observed in the modified tRNA models. Therefore, it is important to determine whether these conformations have positive or detrimental effects on translation. Unfortunately, this question is not easily addressed as the ribosome is a large ribonucleoprotein machine with many simultaneous movements within different subunits.[68,89-91] To address this complexity, one could run MD simulations on a ribosomal subsystem, such as an mRNA–ASL or an mRNA–tRNA complex, to gain insight into the decoding properties of the modified ASL. In this case, one could characterize interactions between the different

conformational states of the ASL isolated in the present study and the mRNA codon and compare them to functional ASL states observed in literature.



Scheme 4.2 – Models to probe the role of tRNA modifications in translation using MD simulations
Potential models for MD investigations on translation, from left to right: standalone tRNA, mRNA–ASL complex, mRNA–tRNA complex, partial ribosome complex including tRNA-binding sites, mRNA, and an A- or P-site tRNA and a full translation complex with tRNAs and mRNA bound to the ribosome.

A more sophisticated study would involve investigating tRNA-binding sites within the ribosome–tRNA–mRNA complex and characterize the intermolecular interactions that govern the accommodation of a modified ASL within the ribosome. Alternatively, one could use a multi-basin all-atom structure-based model, which uses multiple experimentally derived structures to define a global potential energy minimum and analyzes a desired motion from that minimum using MD simulations. Nevertheless, while this model has been used to study tRNA translocation and ribosomal subunit rotation,[92-97] it has not been applied to study the effects of tRNA modifications in the ribosome. Overall, these studies will provide the first atomic-level insight on the accommodation and structural dynamics of modified tRNA within the ribosome and further our understanding of the effects this modification family have on translation.

4.5 References

1 Grosjean, H. & Westhof, E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res* **44**, 8020-8040, doi:10.1093/nar/gkw608 (2016).

2 Rozov, A. *et al.* Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nature Communications* **7**, 10457, doi:10.1038/ncomms10457 (2016).

3 Denmon, A. P., Wang, J. & Nikonowicz, E. P. Conformation effects of base modification on the anticodon stem-loop of Bacillus subtilis tRNA(Tyr). *J Mol Biol* **412**, 285-303, doi:10.1016/j.jmb.2011.07.010 (2011).

4 Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways-2013 update. *Nucleic Acids Res* **41**, D262-D267, doi:10.1093/nar/gks1007 (2013).

5 Suzuki, T., Suzuki, T., Wada, T., Saigo, K. & Watanabe, K. Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases. *Embo J* **21**, 6581-6589, doi:10.1093/emboj/cdf656 (2002).

6 Suzuki, T. *et al.* Complete chemical structures of human mitochondrial tRNAs. *Nature Communications* **11**, 4269, doi:10.1038/s41467-020-18068-6 (2020).

7 Leszczynska, G., Leonczak, P., Wozniak, K. & Malkiewicz, A. Chemical synthesis of the 5-taurinomethyl(-2-thio)uridine modified anticodon arm of the human mitochondrial tRNA(Leu(UUR)) and tRNA(Lys). *RNA (New York, N.Y.)* **20**, 938-947, doi:10.1261/rna.044412.114 (2014).

8 Li, X., Li, R., Lin, X. & Guan, M. X. Isolation and characterization of the putative nuclear modifier gene MTO1 involved in the pathogenesis of deafness-associated mitochondrial 12 S rRNA A1555G mutation. *J Biol Chem* **277**, 27256-27264, doi:10.1074/jbc.M203267200 (2002).

9 Suzuki, T., Nagao, A. & Suzuki, T. Human mitochondrial diseases caused by lack of taurine modification in mitochondrial tRNAs. *Wiley Interdiscip Rev RNA* **2**, 376-386, doi:10.1002/wrna.65 (2011).

10 Yasukawa, T., Suzuki, T., Ishii, N., Ohta, S. & Watanabe, K. Wobble modification defect in tRNA disturbs codon-anticodon interaction in a mitochondrial disease. *Embo J* **20**, 4794-4802, doi:10.1093/emboj/20.17.4794 (2001).

11 Yasukawa, T. *et al.* Defect in modification at the anticodon wobble nucleotide of mitochondrial tRNALys with the MERRF encephalomyopathy pathogenic mutation. *FEBS Letters* **467**, 175-178, doi:https://doi.org/10.1016/S0014-5793(00)01145-5 (2000).

12 Yasukawa, T. *et al.* Modification Defect at Anticodon Wobble Nucleotide of Mitochondrial tRNAsLeu(UUR) with Pathogenic Mutations of Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis, and Stroke-like Episodes*. *Journal of Biological Chemistry* **275**, 4251-4257, doi:https://doi.org/10.1074/jbc.275.6.4251 (2000).

13 Han, L. & Phizicky, E. M. A rationale for tRNA modification circuits in the anticodon loop. *RNA* **24**, 1277-1284, doi:10.1261/rna.067736.118 (2018).

14 Durant, P. C., Bajji, A. C., Sundaram, M., Kumar, R. K. & Davis, D. R. Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. *Biochemistry* **44**, 8078-8089, doi:10.1021/bi050343f (2005).

15 Joshi, K., Bhatt, M. J. & Farabaugh, P. J. Codon-specific effects of tRNA anticodon loop modifications on translational misreading errors in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res* **46**, 10331-10339, doi:10.1093/nar/gky664 (2018).

16 Lescrinier, E. *et al.* The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res* **34**, 2878-2886, doi:10.1093/nar/gkl081 (2006).

17 McCrate, N. E., Varner, M. E., Kim, K. I. & Nagan, M. C. Molecular dynamics simulations of human tRNA Lys,3 UUU: the role of modified bases in mRNA recognition. *Nucleic Acids Res* **34**, 5361-5368, doi:10.1093/nar/gkl580 (2006).

18 Sambhare, S. B. *et al.* Structural significance of modified nucleosides k2C and t6A present in the anticodon loop of tRNAIle. *RSC Advances* **4**, 14176-14188, doi:10.1039/C3RA47335J (2014).

19 Thiaville, P. C. *et al.* Global translational impacts of the loss of the tRNA modification t(6)A in yeast. *Microb Cell* **3**, 29-45, doi:10.15698/mic2016.01.473 (2016).

20 Fakruddin, M. *et al.* Cdk5rap1-mediated 2-methylthio-N6-isopentenyladenosine modification is absent from nuclear-derived RNA species. *Nucleic Acids Res* **45**, 11954-11961, doi:10.1093/nar/gkx819 (2017).
21 McLennan, B. D., Buck, M., Humphreys, J. & Griffiths, E. Iron-related modification of bacterial transfer RNA. *Nucleic Acids Res* **9**, 2629-2640, doi:10.1093/nar/9.11.2629 (1981).
22 Manickam, N., Joshi, K., Bhatt, M. J. & Farabaugh, P. J. Effects of tRNA modification on translational accuracy depend on intrinsic codon–anticodon strength. *Nucleic Acids Res* **44**, 1871-1881, doi:10.1093/nar/gkv1506 (2015).
23 Wilson, R. K. & Roe, B. A. Presence of the hypermodified nucleotide N6-(delta 2-isopentenyl)-2-methylthioadenosine prevents codon misreading by Escherichia coli phenylalanyl-transfer RNA. *Proceedings of the National Academy of Sciences* **86**, 409-413, doi:10.1073/pnas.86.2.409 (1989).
24 Lamichhane, T. N. *et al.* Lack of tRNA Modification Isopentenyl-A37 Alters mRNA Decoding and Causes Metabolic Deficiencies in Fission Yeast. *Molecular and Cellular Biology* **33**, 2918-2929, doi:doi:10.1128/MCB.00278-13 (2013).
25 Seelam Prabhakar, P., Takyi, N. A. & Wetmore, S. D. Posttranscriptional modifications at the 37th position in the anticodon stem–loop of tRNA: structural insights from MD simulations. *RNA* **27**, 202-220, doi:10.1261/rna.078097.120 (2021).
26 Bénas, P. *et al.* The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA* **6**, 1347-1355, doi:10.1017/s1355838200000911 (2000).
27 Schmeing, T. M., Voorhees, R. M., Kelley, A. C. & Ramakrishnan, V. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol* **18**, 432-436, doi:10.1038/nsmb.2003 (2011).
28 Schrödinger, L. The PyMOL Molecular Graphics System, Version 2.0
29 Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **4**, 17, doi:10.1186/1758-2946-4-17 (2012).
30 Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **97**, 10269-10280, doi:10.1021/j100142a004 (1993).
31 Dupradeau, F.-Y. *et al.* The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* **12**, 7821-7839, doi:10.1039/c0cp00111b (2010).
32 Gaussian 09, Revision D.01 (Gaussian, Inc., Wallingford CT, 2009).
33 R.E.D. Python: Object oriented programming for Amber force fields (Sanford Burnham Prebys Medical Discovery Institute, Université de Picardie - Jules Vern, 2013).
34 Vanquelef, E. *et al.* R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* **39**, W511-W517, doi:10.1093/nar/gkr288 (2011).
35 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157-1174, doi:https://doi.org/10.1002/jcc.20035 (2004).
36 Schmit, J. D., Kariyawasam, N. L., Needham, V. & Smith, P. E. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *Journal of Chemical Theory and Computation* **14**, 1823-1827, doi:10.1021/acs.jctc.7b01254 (2018).
37 AMBER 2018 (University of California, San Francisco, 2018).
38 Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095, doi:10.1021/ct400341p (2013).
39 AMBER 2020 (University of California, San Francisco, 2020).
40 Bottaro, S. *et al.* Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* **25**, 219-231, doi:10.1261/rna.067678.118 (2019).
41 Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD* **96**, 226-231 (1996).
42 Sibler, A. P., Dirheimer, G. & Martin, R. P. Codon reading patterns in Saccharomyces cerevisiae mitochondria based on sequences of mitochondrial tRNAs. *FEBS Lett* **194**, 131-138, doi:10.1016/0014-5793(86)80064-3 (1986).

43 Sibler, A. P., Dirheimer, G. & Martin, R. P. The primary structure of yeast mitochondrial tyrosine tRNA. *FEBS Letters* **152**, 153-156, doi:https://doi.org/10.1016/0014-5793(83)80368-8 (1983).
44 Kumbhar, N. M. & Gopal, J. S. Structural significance of hypermodified nucleoside 5-carboxymethylaminomethyluridine (cmnm5U) from 'wobble' (34th) position of mitochondrial tRNAs: Molecular modeling and Markov state model studies. *Journal of Molecular Graphics and Modelling* **86**, 66-83, doi:https://doi.org/10.1016/j.jmgm.2018.10.004 (2019).
45 Prabhakar, P. S., Takyi, N. A. & Wetmore, S. D. Posttranscriptional modifications at the 37th position in the anticodon stem-loop of tRNA: structural insights from MD simulations. *RNA* **27**, 202-220, doi:10.1261/rna.078097.120 (2021).
46 Schweizer, U., Bohleber, S. & Fradejas-Villar, N. The modified base isopentenyladenosine and its derivatives in tRNA. *RNA biology* **14**, 1197-1208, doi:10.1080/15476286.2017.1294309 (2017).
47 Alexander, R. W., Eargle, J. & Luthey-Schulten, Z. Experimental and computational determination of tRNA dynamics. *FEBS Letters* **584**, 376-386, doi:https://doi.org/10.1016/j.febslet.2009.11.061 (2010).
48 Armengod, M. E. *et al.* Modification of the wobble uridine in bacterial and mitochondrial tRNAs reading NNA/NNG triplets of 2-codon boxes. *RNA Biol* **11**, 1495-1507, doi:10.4161/15476286.2014.992269 (2014).
49 Barraud, P., Schmitt, E., Mechulam, Y., Dardel, F. & Tisné, C. A unique conformation of the anticodon stem-loop is associated with the capacity of tRNAfMet to initiate protein synthesis. *Nucleic Acids Res* **36**, 4894-4901, doi:10.1093/nar/gkn462 (2008).
50 Carter, C. W. & Wolfenden, R. tRNA acceptor stem and anticodon bases form independent codes related to protein folding. *Proceedings of the National Academy of Sciences* **112**, 7489, doi:10.1073/pnas.1507569112 (2015).
51 Carter, C. W. & Wolfenden, R. tRNA acceptor-stem and anticodon bases embed separate features of amino acid chemistry. *RNA Biology* **13**, 145-151, doi:10.1080/15476286.2015.1112488 (2016).
52 Jeong, H. & Kim, J. Unique anticodon loop conformation with the flipped-out wobble nucleotide in the crystal structure of unbound tRNAVal. *RNA* **27**, 1330-1338, doi:10.1261/rna.078863.121 (2021).
53 Saint-Léger, A. *et al.* Saturation of recognition elements blocks evolution of new tRNA identities. *Science Advances* **2**, e1501860, doi:doi:10.1126/sciadv.1501860 (2016).
54 Kurata, S. *et al.* Decoding property of C5 uridine modification at the wobble position of tRNA anticodon. *Nucleic Acids Symposium Series* **3**, 245-246, doi:10.1093/nass/3.1.245 (2003).
55 Kurata, S. *et al.* Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J Biol Chem* **283**, 18801-18811, doi:10.1074/jbc.M800233200 (2008).
56 Yamada, Y., Murao, K. & Ishikura, H. 5-(Carboxymethylaminomethyl)-2-thiouridine, a new modified nucleoside found at the first letter position of the anticodon. *Nucleic Acids Res* **9**, 1933-1939, doi:10.1093/nar/9.8.1933 (1981).
57 Juhling, F. *et al.* tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**, D159-162, doi:10.1093/nar/gkn772 (2009).
58 Narendran, A. *et al.* Silencing of the tRNA Modification Enzyme Cdkal1 Effects Functional Insulin Synthesis in NIT-1 Cells: tRNALys3 Lacking ms2- (ms2t6A37) is Unable to Establish Sufficient Anticodon:Codon Interactions to Decode the Wobble Codon AAG. *Frontiers in Molecular Biosciences* **7**, doi:10.3389/fmolb.2020.584228 (2021).
59 Zhou, J.-B. *et al.* Molecular basis for t6A modification in human mitochondria. *Nucleic Acids Res* **48**, 3181-3194, doi:10.1093/nar/gkaa093 (2020).
60 Abbink, T. E. M. & Berkhout, B. HIV-1 reverse transcription initiation: A potential target for novel antivirals? *Virus Research* **134**, 4-18, doi:https://doi.org/10.1016/j.virusres.2007.12.009 (2008).
61 Rojas-Benítez, D., Ibar, C. & Glavic, Á. The Drosophila EKC/KEOPS complex: roles in protein synthesis homeostasis and animal growth. *Fly (Austin)* **7**, 168-172, doi:10.4161/fly.25227 (2013).
62 Rojas-Benitez, D., Thiaville, P. C., de Crécy-Lagard, V. & Glavic, A. The Levels of a Universally Conserved tRNA Modification Regulate Cell Growth. *The Journal of biological chemistry* **290**, 18699-18707, doi:10.1074/jbc.M115.665406 (2015).
63 Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Current opinion in structural biology* **16**, 279-287, doi:10.1016/j.sbi.2006.05.009 (2006).

64 Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499-512, doi:10.1017/s1355838201002515 (2001).
65 Entelis, N. S., Kieffer, S., Kolesnikova, O. A., Martin, R. P. & Tarassov, I. A. Structural requirements of tRNA^{Lys} for its import into yeast mitochondria. *Proceedings of the National Academy of Sciences* **95**, 2838-2843, doi:10.1073/pnas.95.6.2838 (1998).
66 Söll, D., Becker, H. D., Plateau, P., Blanquet, S. & Ibba, M. Context-dependent anticodon recognition by class I lysyl-tRNA synthetases. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 14224-14228, doi:10.1073/pnas.97.26.14224 (2000).
67 Davis, D. R., Veltri, C. A. & Nielsen, L. An RNA Model System for Investigation of Pseudouridine Stabilization of the Codon-Anticodon Interaction in tRNALys, tRNAHis and tRNATyr. *Journal of Biomolecular Structure and Dynamics* **15**, 1121-1132, doi:10.1080/07391102.1998.10509006 (1998).
68 Agirrezabala, X. & Valle, M. Structural Insights into tRNA Dynamics on the Ribosome. *International Journal of Molecular Sciences* **16**, 9866-9895 (2015).
69 Agris, P. F. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* **9**, 629-635, doi:10.1038/embor.2008.104 (2008).
70 Allnér, O. & Nilsson, L. Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome. *RNA (New York, N.Y.)* **17**, 2177-2188, doi:10.1261/rna.029231.111 (2011).
71 Khade, P. & Joseph, S. Functional interactions by transfer RNAs in the ribosome. *FEBS Lett* **584**, 420-426, doi:10.1016/j.febslet.2009.11.034 (2010).
72 Simonetti, A. *et al.* A structural view of translation initiation in bacteria. *Cell Mol Life Sci* **66**, 423-436, doi:10.1007/s00018-008-8416-4 (2009).
73 Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416-420, doi:10.1038/nature07192 (2008).
74 Valle, M. *et al.* Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol* **10**, 899-906, doi:10.1038/nsb1003 (2003).
75 Savir, Y. & Tlusty, T. Conformational proofreading: the impact of conformational changes on the specificity of molecular recognition. *PLoS One* **2**, e468, doi:10.1371/journal.pone.0000468 (2007).
76 Savir, Y. & Tlusty, T. The Ribosome as an Optimal Decoder: A Lesson in Molecular Recognition. *Cell* **153**, 471-479, doi:10.1016/j.cell.2013.03.032 (2013).
77 Dao, V. *et al.* Ribosome binding of DNA analogs of tRNA requires base modifications and supports the "extended anticodon". *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2125-2129, doi:10.1073/pnas.91.6.2125 (1994).
78 Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721-732, doi:10.1016/s0092-8674(02)01086-3 (2002).
79 Maehigashi, T., Dunkle, J. A., Miles, S. J. & Dunham, C. M. Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 12740-12745, doi:10.1073/pnas.1409436111 (2014).
80 Yarian, C. *et al.* Modified nucleoside dependent Watson-Crick and wobble codon binding by tRNALysUUU species. *Biochemistry* **39**, 13390-13395, doi:10.1021/bi001302g (2000).
81 Fislage, M., Wauters, L. & Versees, W. MnmE, a GTPase That Drives a Complex tRNA Modification Reaction. *Biopolymers* **105**, 568-579, doi:10.1002/bip.22813 (2016).
82 Armengod, M. E. *et al.* Enzymology of tRNA modification in the bacterial MnmEG pathway. *Biochimie* **94**, 1510-1520, doi:10.1016/j.biochi.2012.02.019 (2012).
83 Li, X. & Guan, M. X. A human mitochondrial GTP binding protein related to tRNA modification may modulate phenotypic expression of the deafness-associated mitochondrial 12S rRNA mutation. *Mol Cell Biol* **22**, 7701-7711, doi:10.1128/mcb.22.21.7701-7711.2002 (2002).
84 Villarroya, M. *et al.* Characterization of human GTPBP3, a GTP-binding protein involved in mitochondrial tRNA modification. *Molecular and cellular biology* **28**, 7514-7531, doi:10.1128/mcb.00946-08 (2008).
85 Ohsawa, Y. *et al.* Taurine supplementation for prevention of stroke-like episodes in MELAS: a multicentre, open-label, 52-week phase III trial. *J Neurol Neurosurg Psychiatry* **90**, 529-536, doi:10.1136/jnnp-2018-317964 (2019).

86 Rikimaru, M. *et al.* Taurine ameliorates impaired the mitochondrial function and prevents stroke-like episodes in patients with MELAS. *Intern Med* **51**, 3351-3357, doi:10.2169/internalmedicine.51.7529 (2012).
87 Suzuki, T., Nagao, A. & Suzuki, T. Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases. *Annual Review of Genetics* **45**, 299-329, doi:10.1146/annurev-genet-110410-132531 (2011).
88 Umeda, N. *et al.* Mitochondria-specific RNA-modifying Enzymes Responsible for the Biosynthesis of the Wobble Base in Mitochondrial tRNAs: IMPLICATIONS FOR THE MOLECULAR PATHOGENESIS OF HUMAN MITOCHONDRIAL DISEASES *. *Journal of Biological Chemistry* **280**, 1613-1624, doi:10.1074/jbc.M409306200 (2005).
89 Baßler, J. & Hurt, E. Eukaryotic Ribosome Assembly. *Annual Review of Biochemistry* **88**, 281-306, doi:10.1146/annurev-biochem-013118-110817 (2019).
90 Dahlberg, A. E. Ribosome structure. The ribosome in action. *Science* **292**, 868-869, doi:10.1126/science.1061513 (2001).
91 Gao, H. *et al.* Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell* **113**, 789-801, doi:10.1016/s0092-8674(03)00427-6 (2003).
92 Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins11Edited by F. E. Cohen. *Journal of Molecular Biology* **298**, 937-953, doi:https://doi.org/10.1006/jmbi.2000.3693 (2000).
93 Freitas, F. C., Fuchs, G., de Oliveira, R. J. & Whitford, P. C. The Dynamics of Subunit Rotation in a Eukaryotic Ribosome. *Biophysica* **1**, 204-221 (2021).
94 Nguyen, K. & Whitford, P. C. Steric interactions lead to collective tilting motion in the ribosome during mRNA–tRNA translocation. *Nature Communications* **7**, 10586, doi:10.1038/ncomms10586 (2016).
95 Noel, Jeffrey K., Chahine, J., Leite, Vitor B. P. & Whitford, Paul C. Capturing Transition Paths and Transition States for Conformational Rearrangements in the Ribosome. *Biophysical Journal* **107**, 2881-2890, doi:https://doi.org/10.1016/j.bpj.2014.10.022 (2014).
96 Noller, H. F., Yusupov, M. M., Yusupova, G. Z., Baucom, A. & Cate, J. H. D. Translocation of tRNA during protein synthesis. *FEBS Letters* **514**, 11-16, doi:https://doi.org/10.1016/S0014-5793(02)02327-X (2002).
97 Whitford, P. C. *et al.* An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430-441, doi:10.1002/prot.22253 (2009).

CHAPTER 5: THESIS SUMMARY

5.1 Thesis Review

The overall focus of this thesis was to provide insight into the role of 5-carboxymethylaminomethyl(-2-thio)uridine ($cmnm^5(s^2)U$) modifications on tRNA structure and function. Previous investigations revealed these modifications are inserted into tRNA by the evolutionarily conserved MnmE-MnmG (MnmEG) complex,[1-5] and their absence is linked to the development of mitochondrial diseases MELAS and MERRF.[6-10] Nevertheless, uncertainty surrounds the MnmEG complex due to the lack of concurring experimental evidence for its formation and catalytic mechanism,[1,3,4,11-17] as well as the absence of crystal structures for the fully assembled complex. Moreover, the local and global structural changes imparted onto the tRNA structure by this family of modifications remained unknown and no investigation considered the accommodation of $cmnm^5(s^2)U34$ modified tRNA within the ribosome. Consequently, this thesis used a dual-pronged approach to address these gaps in the literature. Specifically, 1) biochemical studies were used to investigate the formation of the MnmEG complex and its interaction with tRNA, and 2) concurrent computational studies were conducted to uncover the effects of $cmnm^5(s^2)U34$ modifications on tRNA structure.

To provide further insight into the structure and assembly of the MnmEG complex, Chapter 2 characterized the binding properties of MnmE and MnmG to tRNA in the absence and presence of known cofactors GTP, FADH and NADH and glycine. In the absence of cofactors, MnmG was found to bind tRNA with a higher affinity than MnmE. Furthermore, although cofactors were found to increase the binding affinities MnmE and MnmG have for tRNA, MnmG always had a higher binding affinity for the tRNA substrate than MnmE. To probe the specificity of the MnmEG complex, the binding affinities of MnmE and MnmG for other RNA substrates were investigated. MnmE was found to indiscriminately bind to unstructured and structured RNA substrates with similar affinities as tRNA, while MnmE discriminated secondary structure and preferentially bound to tRNA. To reveal potential binding sites within the modification complex, electrostatic mapping was also conducted on symmetry modelled MnmE and MnmG. MnmE was found to have a predominantly negative surface, which is detrimental in binding tRNA's negatively charged phosphate backbone. In contrast, MnmG was found to have three positive grooves that could act as tRNA-binding pockets. Taken together, the results from both studies suggest

that the MnmEG complex likely assembles in an asymmetric manner. Nevertheless, further structural experiments like AUC, are required to confirm this hypothesis.

In order to accurately study the effects of cmnm^5(s^2)U34 on tRNA structure, a computational protocol for studying tRNA structure using molecular dynamics simulations was established. Prior to this study, multiple conformational studies had only been conducted for DNA and protein molecules,[18-26] not RNA. Consequently, Chapter 3 of this thesis compared the conformational sampling by a single, long MD simulation to that of various replica ensembles to establish an accurate MD protocol for sampling tRNA phase space. Because the global shape of tRNA was maintained in both rMD and cMD simulations and high amplitudes of motion were observed in the anticodon stem-loop (ASL) domain, a detailed conformational analysis was conducted on the ASL. As an ensemble, thirty 500-ns MD simulations sampled all ASL states previously reported across different sequences in isolated, protein-bound and ribosome-bound tRNAs. On the other hand, replica ensembles of 10 or more independent simulations were also found to representatively sample the ASL phase space, providing a more efficient protocol for computational investigations on tRNA structure. In contrast, a single 5 μs simulation was not able to describe the conformational space of the anticodon loop region. Consequently, replica ensembles of ten 500-ns simulations were recommended to accurately study the structural dynamics of tRNA structures.

Chapter 4 used the computational protocol developed in Chapter 3 to investigate the structural impact of cmnm^5(s)^{2}U34 modifications on tRNA and their synergistic effects with ms^{2}i^{6}A37 or t^{6}A37 in tRNA^{Trp} or tRNA^{Lys}, respectively. The presence of the cmnm^5(s^2)U34 modifications and their accompanying modifications at position 37 did not affect the global shape of tRNA^{Trp} and tRNA^{Lys} and no long-range modification effects were detected in either tRNAs. However, cmnm^{5}U34 and cmnm^{5}s^{2}U34 had different effects on the flexibility of the ASL. Specifically, while cmnm^{5}U34 reduced overall flexibility and disorder in the anticodon loop, cmnm^{5}s^{2}U34 increased the motion of the loop residues and made the region more prone to disorder. Nevertheless, both cmnm^5(s^2)U modifications stabilized the backbone of the nucleotide at position 34. On the other hand, ms^{2}i^{6}A37 and t^{6}A37 had similar effects on tRNA^{Trp} and tRNA^{Lys}. Specifically, these modified bases afforded structural stability to the remainder of the anticodon loop and promoted the adoption of open loop conformations that are functionally favored in protein- and ribosome-bound tRNAs.[27-37] This investigation provided atomic-level details of the role of tandem

cmnm^{5}U34/ms^{2}i^{6}A37 and cmnm^{5}s^{2}U34/t^{6}A37 modifications in preparing the tRNA for protein synthesis. This insight into the structure and dynamics of the modified tRNAs strengthens our understanding of their importance *in vivo*. More notably, due to the links between cmnm^5(s^2)U34 and mitochondrial disorders,[38-42] unveiling their function provides avenues for further research as RNA modifications have recently emerged as therapeutic targets in the battle against multiple diseases.[43,44]

5.2 Final Remarks

This thesis used an interdisciplinary approach to provide key insight into the insertion and effects of the cmnm^5(s^2)U34 modifications in tRNA structure. Specifically, the biochemical studies and structural analyses conducted on the MnmE and MnmG proteins uncovered the binding propensities of each protein and provided the groundwork for future studies that aim to isolate the full MnmEG complex. On the other hand, the computational studies presented in this thesis provided (i) an accurate and efficient protocol for investigating tRNA structures with MD simulations and (ii) molecular-level details of the structural and synergistic effects cmnm^5(s^2)U34 modifications on tRNA. These computational studies provide the basis for future investigations on tRNA structure and a starting point for understanding the role of cmnm^5(s^2)U34 modifications in biologically relevant environments. Overall, this thesis employed a bi-directional approach to investigate cmnm^5(s^2)U modifications, addressing some of the gaps in the experimental literature, and providing missing insight into the structural dynamics of the modification family.

5.3 References

1 Ruiz-Partida, R. *et al.* An Alternative Homodimerization Interface of MnmG Reveals a Conformational Dynamics that Is Essential for Its tRNA Modification Function. *J Mol Biol* **430**, 2822-2842, doi:10.1016/j.jmb.2018.05.035 (2018).

2 Armengod, M. E. *et al.* Modification of the wobble uridine in bacterial and mitochondrial tRNAs reading NNA/NNG triplets of 2-codon boxes. *RNA Biol* **11**, 1495-1507, doi:10.4161/15476286.2014.992269 (2014).

3 Moukadiri, I., Garzon, M. J., Bjork, G. R. & Armengod, M. E. The output of the tRNA modification pathways controlled by the Escherichia coli MnmEG and MnmC enzymes depends on the growth conditions and the tRNA species. *Nucleic Acids Res* **42**, 2602-2623, doi:10.1093/nar/gkt1228 (2014).

4 Armengod, M. E. *et al.* Enzymology of tRNA modification in the bacterial MnmEG pathway. *Biochimie* **94**, 1510-1520, doi:10.1016/j.biochi.2012.02.019 (2012).

5 Moukadiri, I. *et al.* Evolutionarily conserved proteins MnmE and GidA catalyze the formation of two methyluridine derivatives at tRNA wobble positions. *Nucleic Acids Res* **37**, 7177-7193, doi:10.1093/nar/gkp762 (2009).

6 Shippy, D. C. & Fadl, A. A. tRNA modification enzymes GidA and MnmE: potential role in virulence of bacterial pathogens. *Int J Mol Sci* **15**, 18267-18280, doi:10.3390/ijms151018267 (2014).

7 Ohsawa, Y. *et al.* Taurine supplementation for prevention of stroke-like episodes in MELAS: a multicentre, open-label, 52-week phase III trial. *J Neurol Neurosurg Psychiatry* **90**, 529-536, doi:10.1136/jnnp-2018-317964 (2019).

8 Rikimaru, M. *et al.* Taurine ameliorates impaired the mitochondrial function and prevents stroke-like episodes in patients with MELAS. *Intern Med* **51**, 3351-3357, doi:10.2169/internalmedicine.51.7529 (2012).

9 Suzuki, T., Nagao, A. & Suzuki, T. Human mitochondrial diseases caused by lack of taurine modification in mitochondrial tRNAs. *Wiley Interdiscip Rev RNA* **2**, 376-386, doi:10.1002/wrna.65 (2011).

10 Nakamura, M. *et al.* MERRF/MELAS overlap syndrome: a double pathogenic mutation in mitochondrial tRNA genes. *Journal of Medical Genetics* **47**, 659-664, doi:10.1136/jmg.2009.072058 (2010).

11 Fislage, M. *et al.* SAXS analysis of the tRNA-modifying enzyme complex MnmE/MnmG reveals a novel interaction mode and GTP-induced oligomerization. *Nucleic Acids Res* **42**, 5978-5992, doi:10.1093/nar/gku213 (2014).

12 Shi, R. *et al.* Structure-function analysis of Escherichia coli MnmG (GidA), a highly conserved tRNA-modifying enzyme. *J Bacteriol* **191**, 7614-7619, doi:10.1128/JB.00650-09 (2009).

13 Meyer, S., Wittinghofer, A. & Versees, W. G-domain dimerization orchestrates the tRNA wobble modification reaction in the MnmE/GidA complex. *J Mol Biol* **392**, 910-922, doi:10.1016/j.jmb.2009.07.004 (2009).

14 Meyer, S., Scrima, A., Versees, W. & Wittinghofer, A. Crystal structures of the conserved tRNA-modifying enzyme GidA: implications for its interaction with MnmE and substrate. *J Mol Biol* **380**, 532-547, doi:10.1016/j.jmb.2008.04.072 (2008).

15 Yim, L., Moukadiri, I., Bjork, G. R. & Armengod, M. E. Further insights into the tRNA modification process controlled by proteins MnmE and GidA of Escherichia coli. *Nucleic Acids Res* **34**, 5892-5905, doi:10.1093/nar/gkl752 (2006).

16 Scrima, A. & Wittinghofer, A. Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. *Embo J* **25**, 2940-2951, doi:10.1038/sj.emboj.7601171 (2006).

17 Yim, L. *et al.* The GTPase activity and C-terminal cysteine of the Escherichia coli MnmE protein are essential for its tRNA modifying function. *J Biol Chem* **278**, 28378-28387, doi:10.1074/jbc.M301381200 (2003).

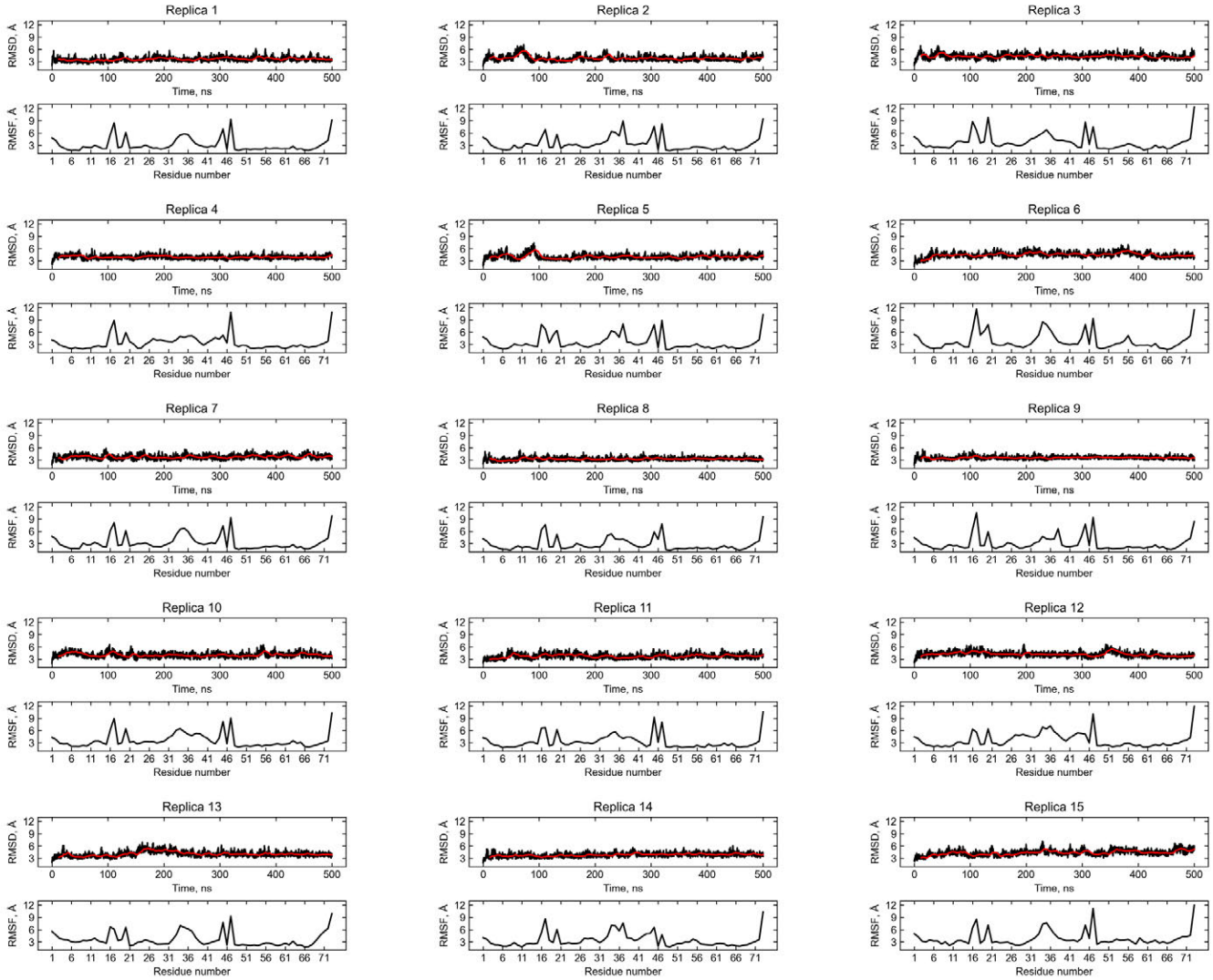
18 Galindo-Murillo, R., Roe, D. R. & Cheatham, T. E. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 1041-1058, doi:https://doi.org/10.1016/j.bbagen.2014.09.007 (2015).

19 Ng, H. W., Laughton, C. A. & Doughty, S. W. Molecular Dynamics Simulations of the Adenosine A2a Receptor: Structural Stability, Sampling, and Convergence. *Journal of Chemical Information and Modeling* **53**, 1168-1178, doi:10.1021/ci300610w (2013).
20 Knapp, B., Frantal, S., Cibena, M., Schreiner, W. , and Bauer, P. . Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? *Journal of Computational Biology* **18**, 997-1005, doi:10.1089/cmb.2010.0237 (2011).
21 Jang, H. *et al.* Structural Convergence Among Diverse, Toxic β-Sheet Ion Channels. *The Journal of Physical Chemistry B* **114**, 9445-9451, doi:10.1021/jp104073k (2010).
22 Grossfield, A. & Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu Rep Comput Chem* **5**, 23-48, doi:10.1016/S1574-1400(09)00502-7 (2009).
23 Soares, T. A. & Straatsma, T. P. Assessment of the convergence of molecular dynamics simulations of lipopolysaccharide membranes. *Molecular Simulation* **34**, 295-307, doi:10.1080/08927020701829880 (2008).
24 Maisuradze, G. G. & Leitner, D. M. Free energy landscape of a biomolecule in dihedral principal component space: Sampling convergence and correspondence between structures and minima. *Proteins: Structure, Function, and Bioinformatics* **67**, 569-578, doi:https://doi.org/10.1002/prot.21344 (2007).
25 Smith, L. J., Daura, X. & van Gunsteren, W. F. Assessing equilibration and convergence in biomolecular simulations. *Proteins: Structure, Function, and Bioinformatics* **48**, 487-496, doi:https://doi.org/10.1002/prot.10144 (2002).
26 Amadei, A., Ceruso, M. A. & Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **36**, 419-424, doi:https://doi.org/10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO;2-U (1999).
27 Giegé, R. & Eriani, G. in *eLS*.
28 Moor, N., Kotik-Kogan, O., Tworowski, D., Sukhanova, M. & Safro, M. The Crystal Structure of the Ternary Complex of Phenylalanyl-tRNA Synthetase with tRNAPhe and a Phenylalanyl-Adenylate Analogue Reveals a Conformational Switch of the CCA End. *Biochemistry* **45**, 10572-10583, doi:10.1021/bi060491l (2006).
29 Petry, S. *et al.* Crystal Structures of the Ribosome in Complex with Release Factors RF1 and RF2 Bound to a Cognate Stop Codon. *Cell* **123**, 1255-1266, doi:10.1016/j.cell.2005.09.039 (2005).
30 Sekine, S.-i. *et al.* Structural Bases of Transfer RNA-Dependent Amino Acid Recognition and Activation by Glutamyl-tRNA Synthetase. *Structure* **14**, 1791-1799, doi:10.1016/j.str.2006.10.005 (2006).
31 Agirrezabala, X. & Valle, M. Structural Insights into tRNA Dynamics on the Ribosome. *International Journal of Molecular Sciences* **16**, 9866-9895 (2015).
32 Allnér, O. & Nilsson, L. Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome. *RNA (New York, N.Y.)* **17**, 2177-2188, doi:10.1261/rna.029231.111 (2011).
33 Berk, V., Zhang, W., Pai, R. D. & Cate, J. H. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci U S A* **103**, 15830-15834, doi:10.1073/pnas.0607541103 (2006).
34 Dahlberg, A. E. Ribosome structure. The ribosome in action. *Science* **292**, 868-869, doi:10.1126/science.1061513 (2001).
35 Khade, P. & Joseph, S. Functional interactions by transfer RNAs in the ribosome. *FEBS Lett* **584**, 420-426, doi:10.1016/j.febslet.2009.11.034 (2010).
36 Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**, 1065-1077, doi:10.1016/j.cell.2006.08.032 (2006).
37 Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416-420, doi:10.1038/nature07192 (2008).
38 Asano, K. *et al.* Metabolic and chemical regulation of tRNA modification associated with taurine deficiency and human disease. *Nucleic Acids Res* **46**, 1565-1583, doi:10.1093/nar/gky068 (2018).

39 Berg, M. D. & Brandl, C. J. Transfer RNAs: diversity in form and function. *RNA Biol* **18**, 316-339, doi:10.1080/15476286.2020.1809197 (2021).
40 Peng, G.-X. *et al.* The human tRNA taurine modification enzyme GTPBP3 is an active GTPase linked to mitochondrial diseases. *Nucleic Acids Res* **49**, 2816-2834, doi:10.1093/nar/gkab104 (2021).
41 Yasukawa, T., Suzuki, T., Ishii, N., Ohta, S. & Watanabe, K. Wobble modification defect in tRNA disturbs codon-anticodon interaction in a mitochondrial disease. *Embo J* **20**, 4794-4802, doi:10.1093/emboj/20.17.4794 (2001).
42 Yasukawa, T. *et al.* Defect in modification at the anticodon wobble nucleotide of mitochondrial tRNALys with the MERRF encephalomyopathy pathogenic mutation. *FEBS Letters* **467**, 175-178, doi:https://doi.org/10.1016/S0014-5793(00)01145-5 (2000).
43 Cayir, A. RNA modifications as emerging therapeutic targets. *Wiley Interdiscip Rev RNA*, e1702, doi:10.1002/wrna.1702 (2021).
44 Uddin, M. B., Wang, Z. & Yang, C. Dysregulations of Functional RNA Modifications in Cancer, Cancer Stemness and Cancer Therapeutics. *Theranostics* **10**, 3164-3189, doi:10.7150/thno.41687 (2020).

APPENIDIX I: SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Figures 3A.1 – 3A.17 and Table 3A.1


Replica 1
Replica 2
Replica 3
Replica 4
Replica 5
Replica 6
Replica 7
Replica 8
Replica 9
Replica 10
Replica 11
Replica 12
Replica 13
Replica 14
Replica 15
RMSD, Å
RMSF, Å
Time, ns
Residue number

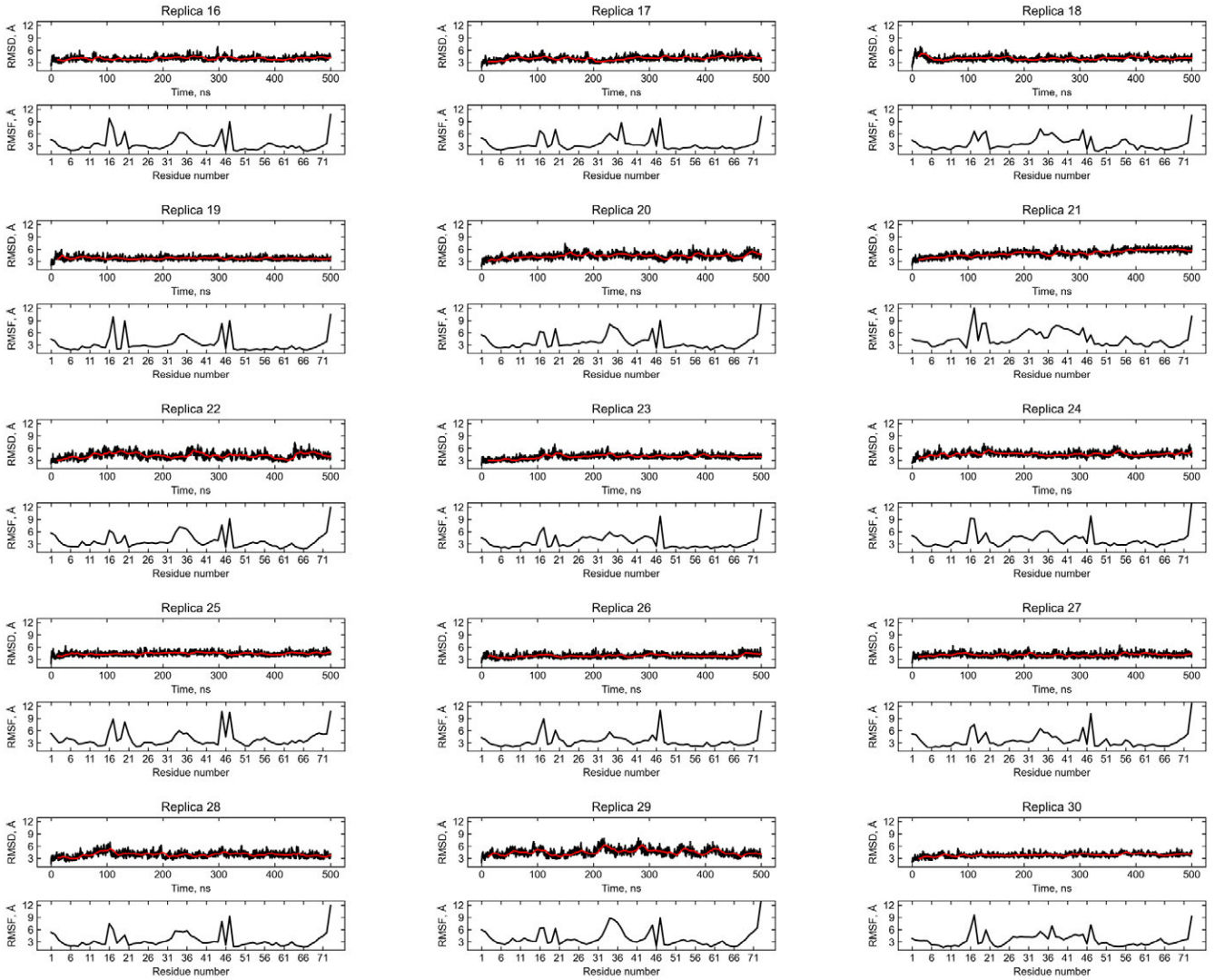


Figure 3A.1 – Fluctuations of the rMD replica ensemble

The heavy atom RMSDs (top) and per residue RMSF (bottom) are presented for each 500-ns replica and the plots read from left to right, top to bottom.

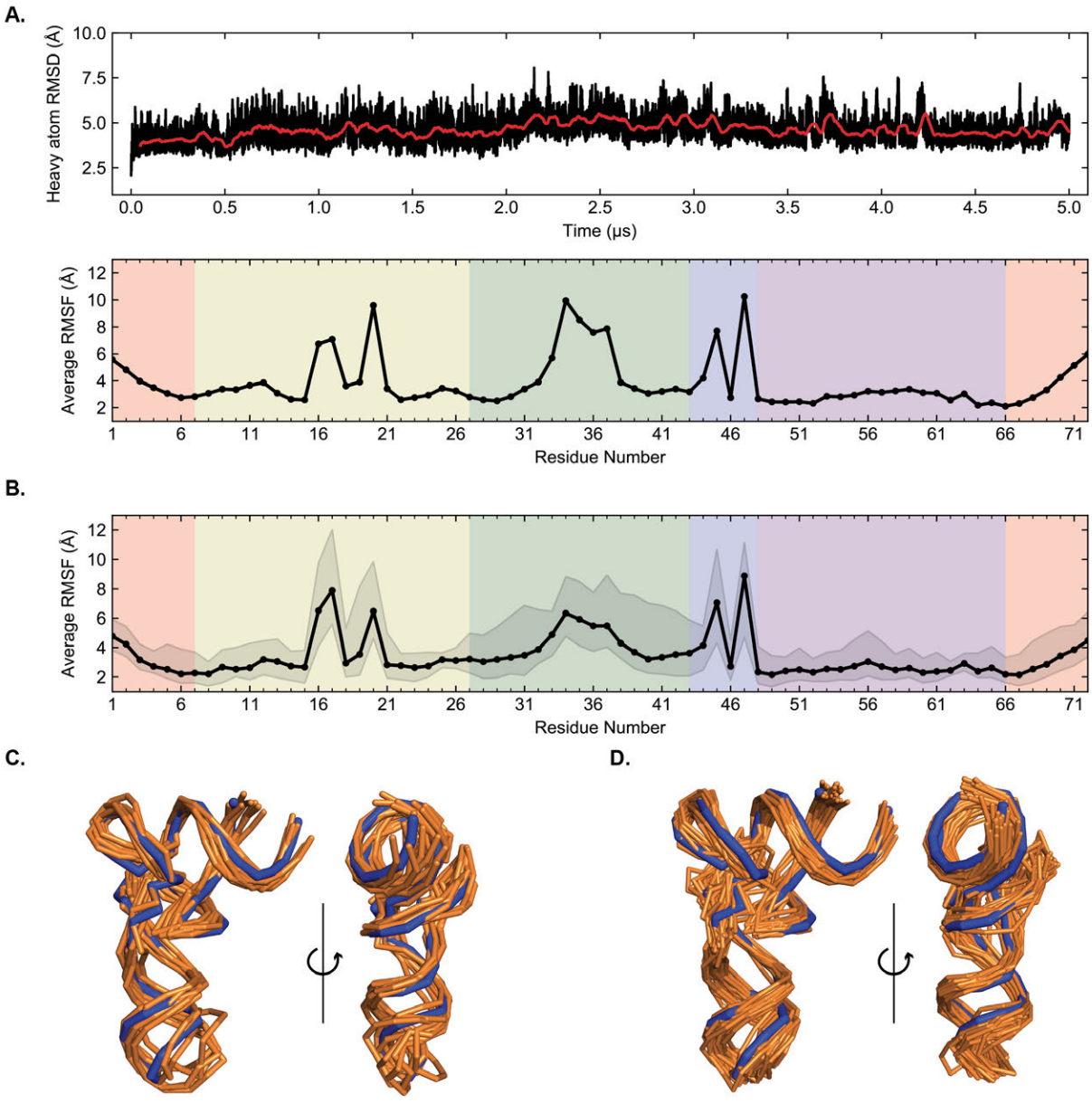


Figure 3A.2 – tRNA dynamics in the cMD and rMD simulations

(A) Heavy atom RMSDs (top) and per residue RMSF (bottom) of the 5 ms cMD trajectory. (B) Per residue RMSF for the rMD simulations. The minimum and maximum fluctuations for each residue are presented in the highlighted area. (C) Representative structure overlays at 500-ns intervals of the cMD trajectory. (D) Representative structure overlays of all replica simulations in the rMD ensemble.

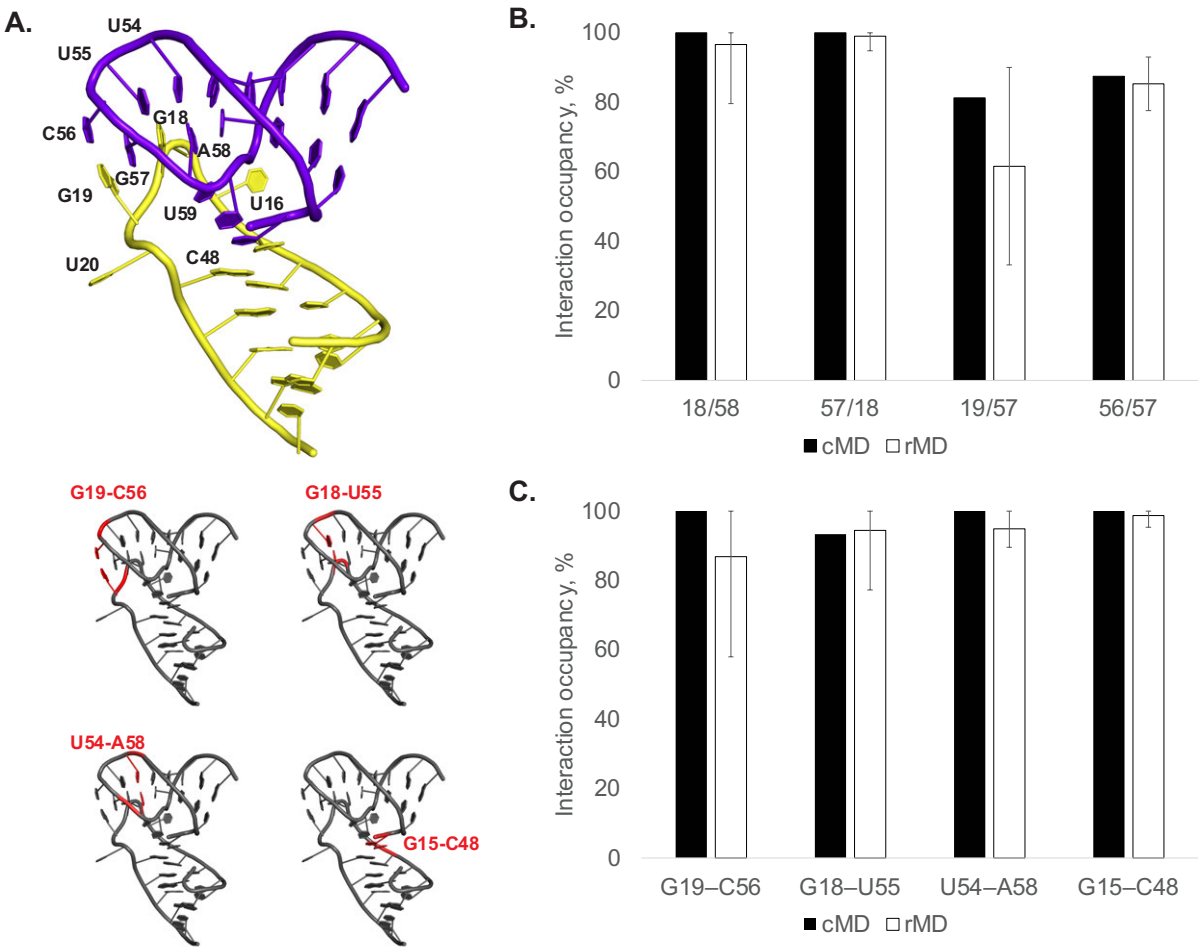


Figure 3A.3 – Maintenance of the global fold of tRNA^{Phe}

(A) Tertiary stacking (top) and hydrogen-bonding interactions at the elbow region of tRNA^{Phe}. (B) Stacking interaction occupancies over cMD and rMD trajectories at the TyC-D loops. Averages and standard deviations are presented for the rMD ensemble. (C) H-bonding interaction occupancies over cMD and rMD trajectories at the TyC-D loops. Averages and standard deviations are presented for the rMD ensemble.

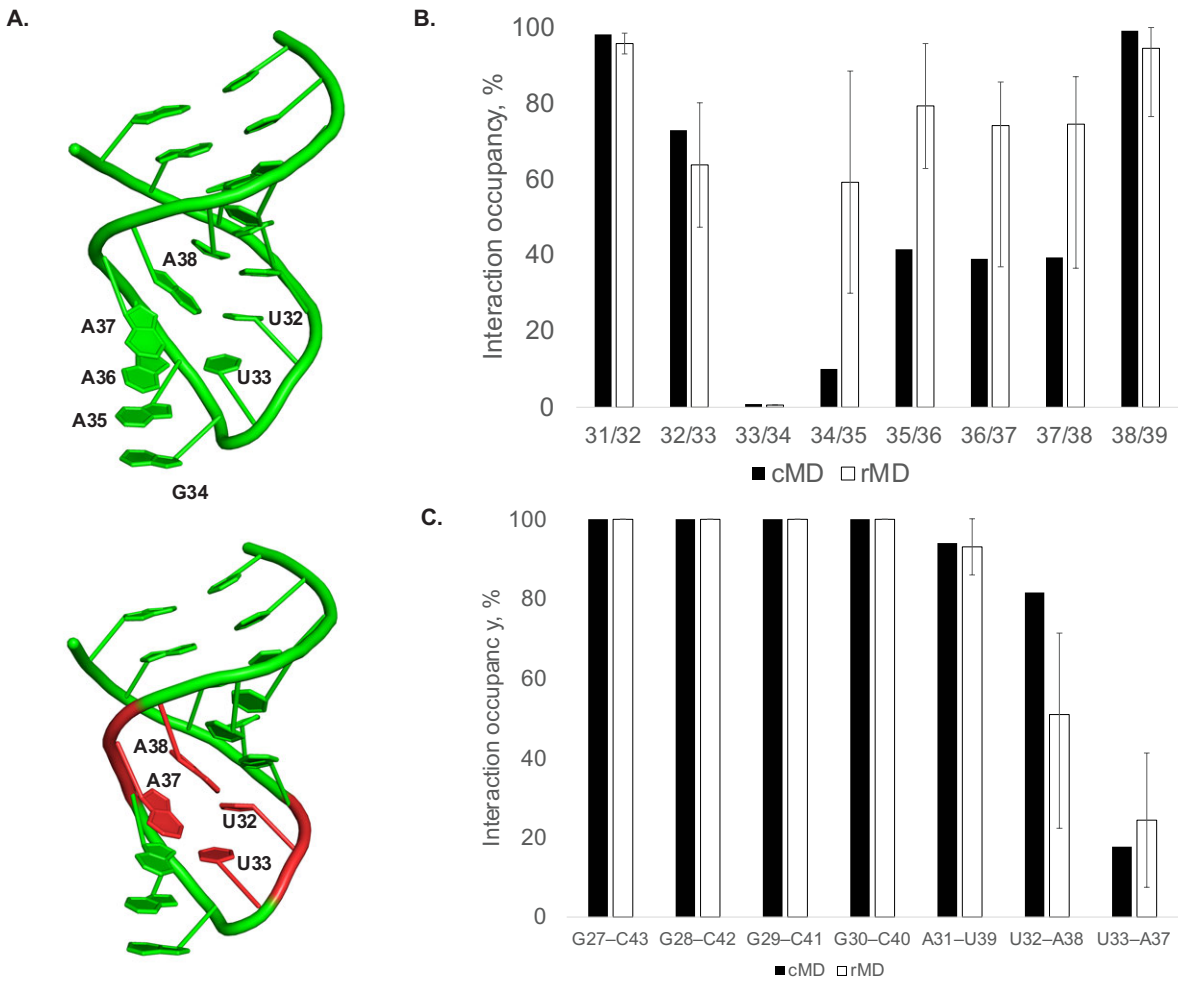


Figure 3A.4 – Non-covalent interactions at the ASL of tRNA^{Phe}

(A) Structural arrangement of the tRNA^{Phe} ASL in the starting crystal structure (PDB ID: 3L0U). The top plot highlights stacking interactions at the loop region, while the bottom plot highlights the two base pairs within the anticodon loop. (B) Stacking interaction occupancies over the cMD and rMD trajectories at the anticodon loop. Averages and standard deviations are presented for the rMD ensemble. (C) Hydrogen bonding interactions over the cMD and rMD trajectories at the anticodon loop. Averages and standard deviations are presented for the rMD ensemble.

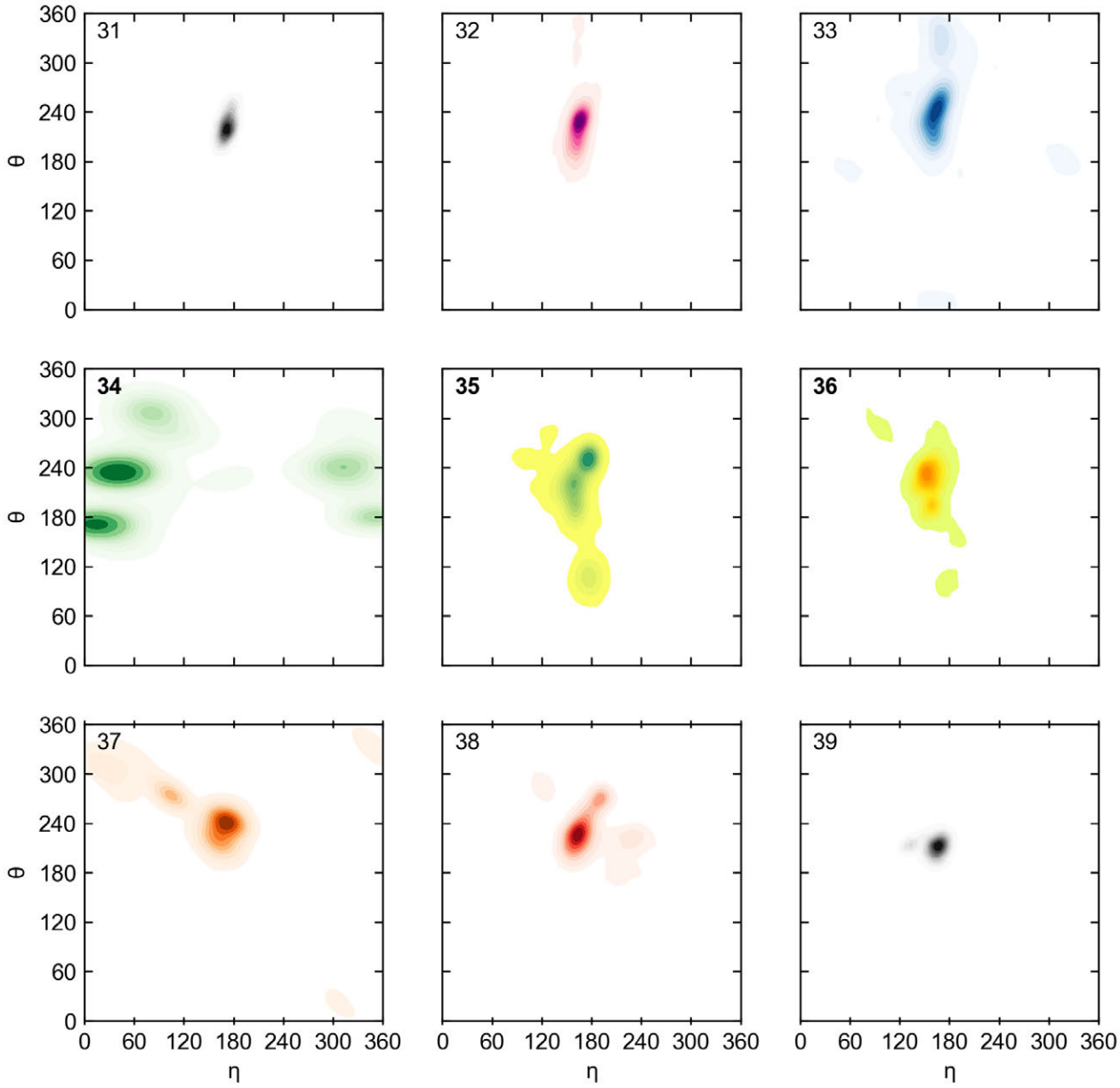


Figure 3A.5 – Backbone dynamics of tRNA^{Phe} throughout rMD simulations

Backbone pseudotorsions (η = ∠C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = ∠C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) of residues 31 to 39 over the course of the rMD trajectories in this study. Helical bases are colored grey, and each loop base has been assigned a color (32: purple, 33: blue, 34: green, 35: yellow-green, 36: yellow, 37: orange, 38: red).

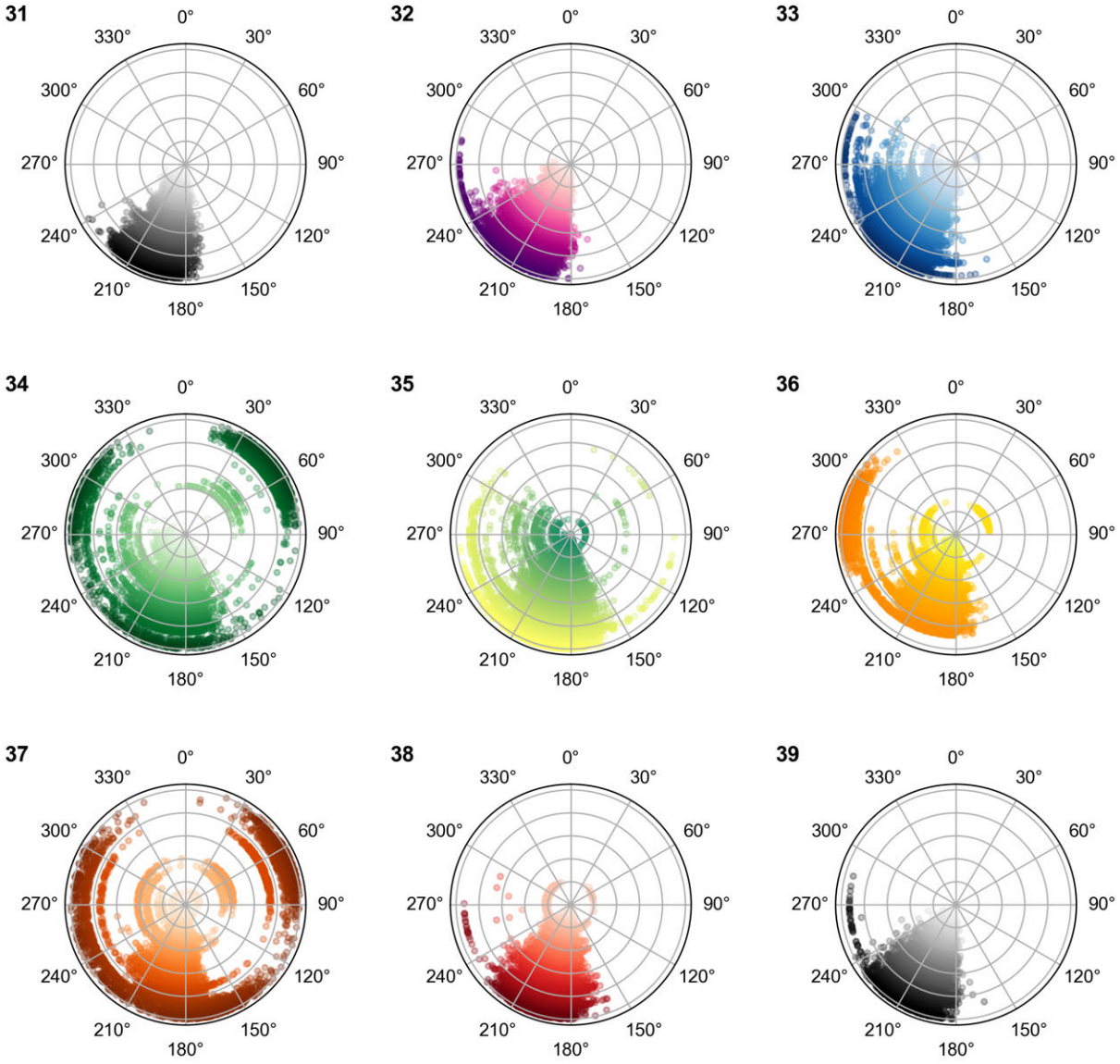


Figure 3A.6 – Dynamics of nitrogenous bases at the anticodon loop of tRNA^{Phe} throughout rMD simulations

Glycosidic torsions (χ = ∠O4'-C1'-N1-C2 or ∠O4'-C1'-N9-C4) of residues 31 to 39 over the course of the rMD trajectories in this study. Helical bases are colored grey, and each loop base has been assigned a color (32: purple, 33: blue, 34: green, 35: yellow-green, 36: yellow, 37: orange, 38: red).

Table 3A.1 – RMSDs from the representative structure overlays of the rMD-predicted ASL conformations and the crystal structure reference (PDB ID: 3L0U)

Conformation	RMSD (Å)
34-stacked	2.1
34-unstacked-I	2.2
34-unstacked-II	2.4
35-unstacked	2.7
36-unstacked	3.1
33-out	2.2
37-out	3.6
disordered	3.1

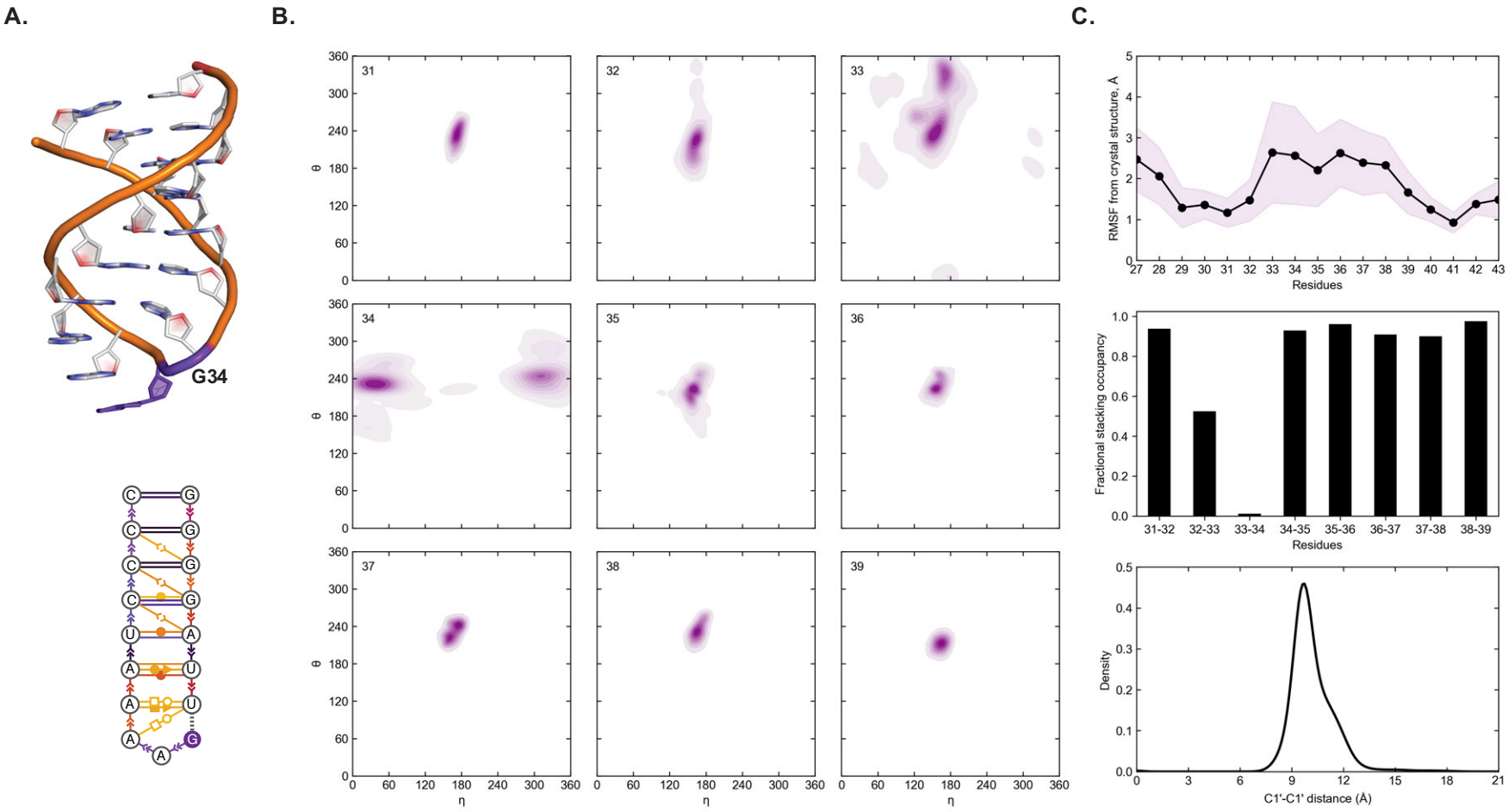


Figure 3A.7 – Description of the 34-stacked conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 34-stacked conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

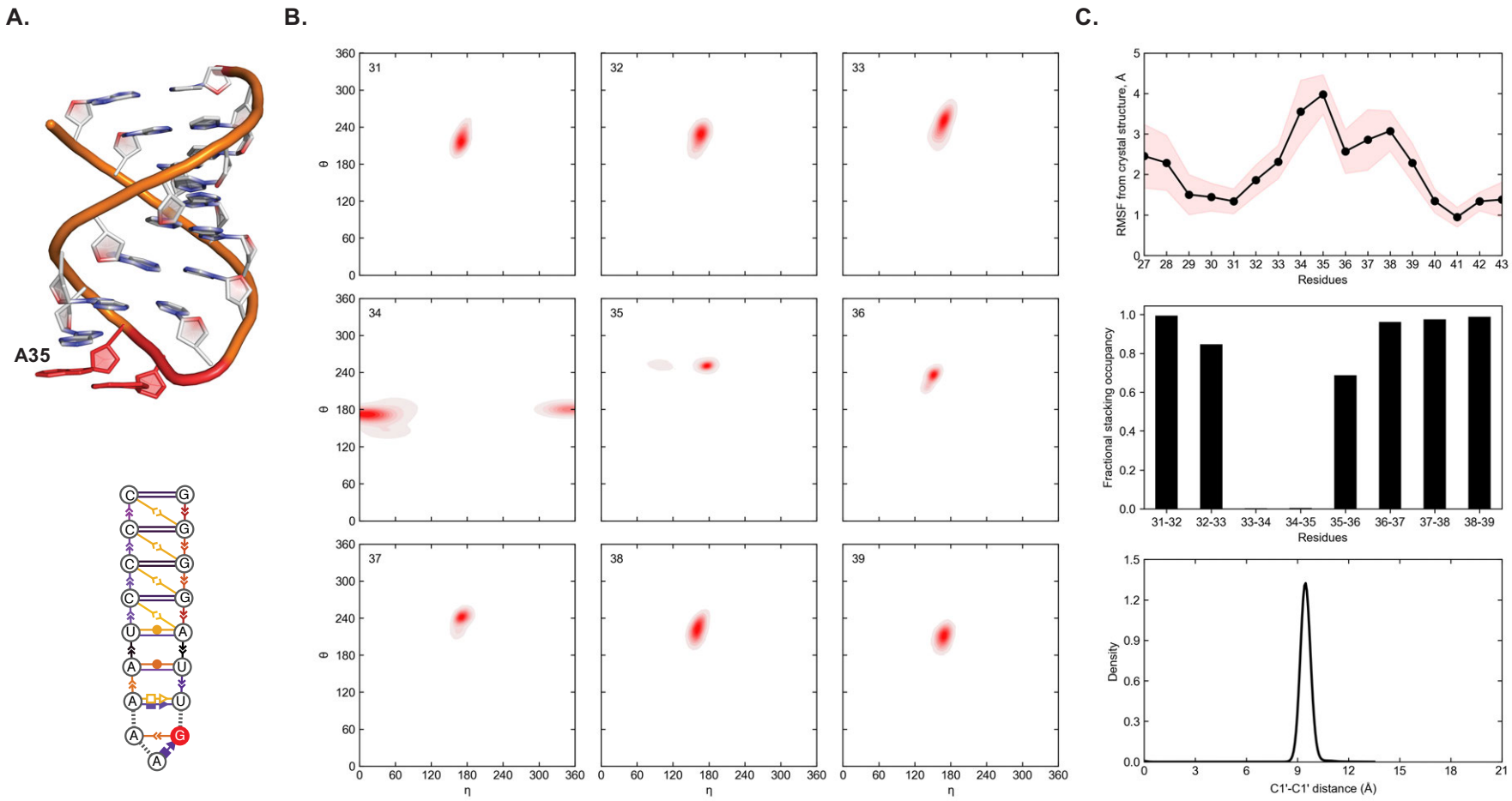


Figure 3A.8 – Description of the 34-unstacked (type-I) conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 34-unstacked-I conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

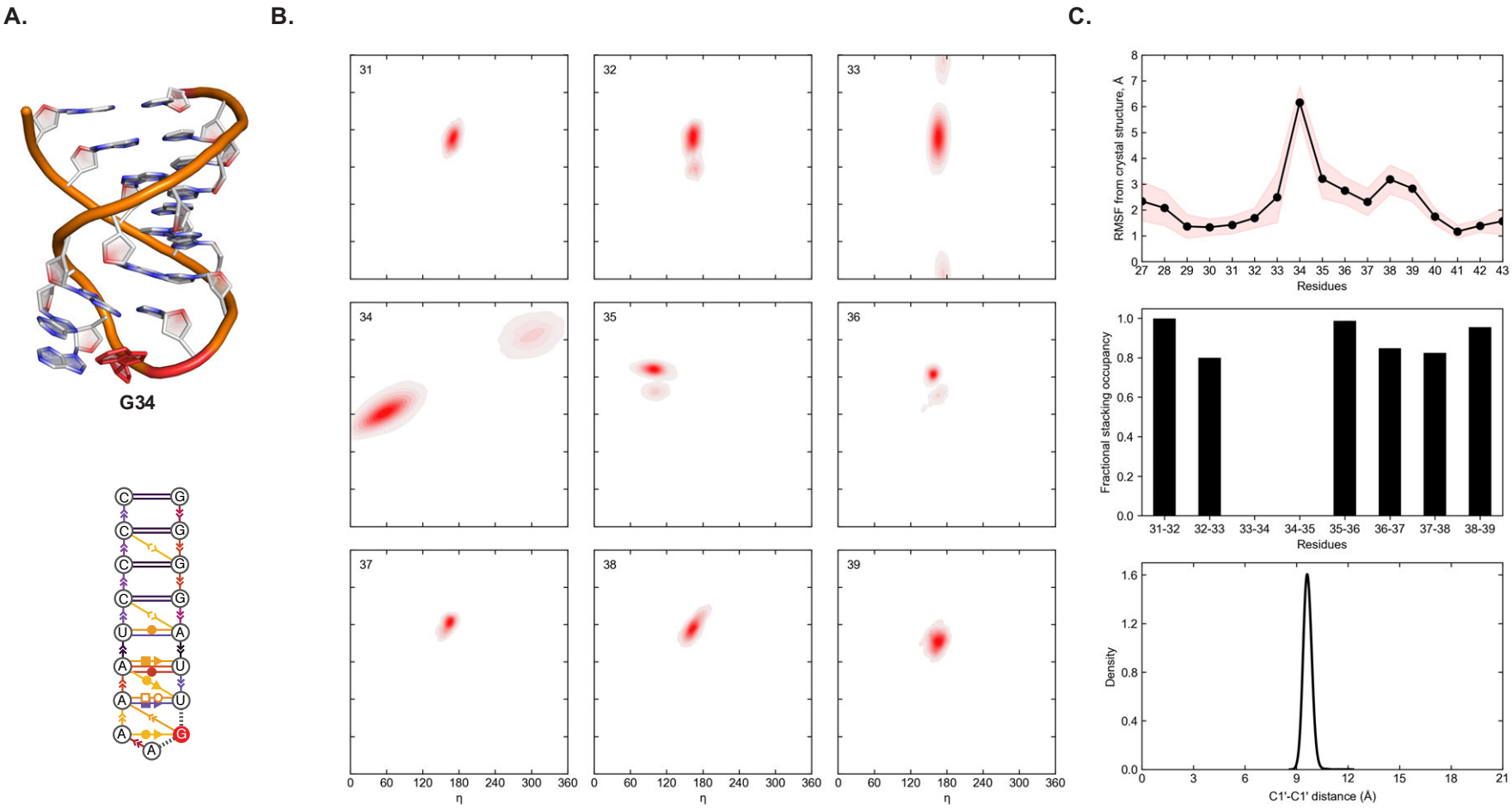


Figure 3A.9 – Description of the 34-unstacked (type-II) conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 34-unstacked-II conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

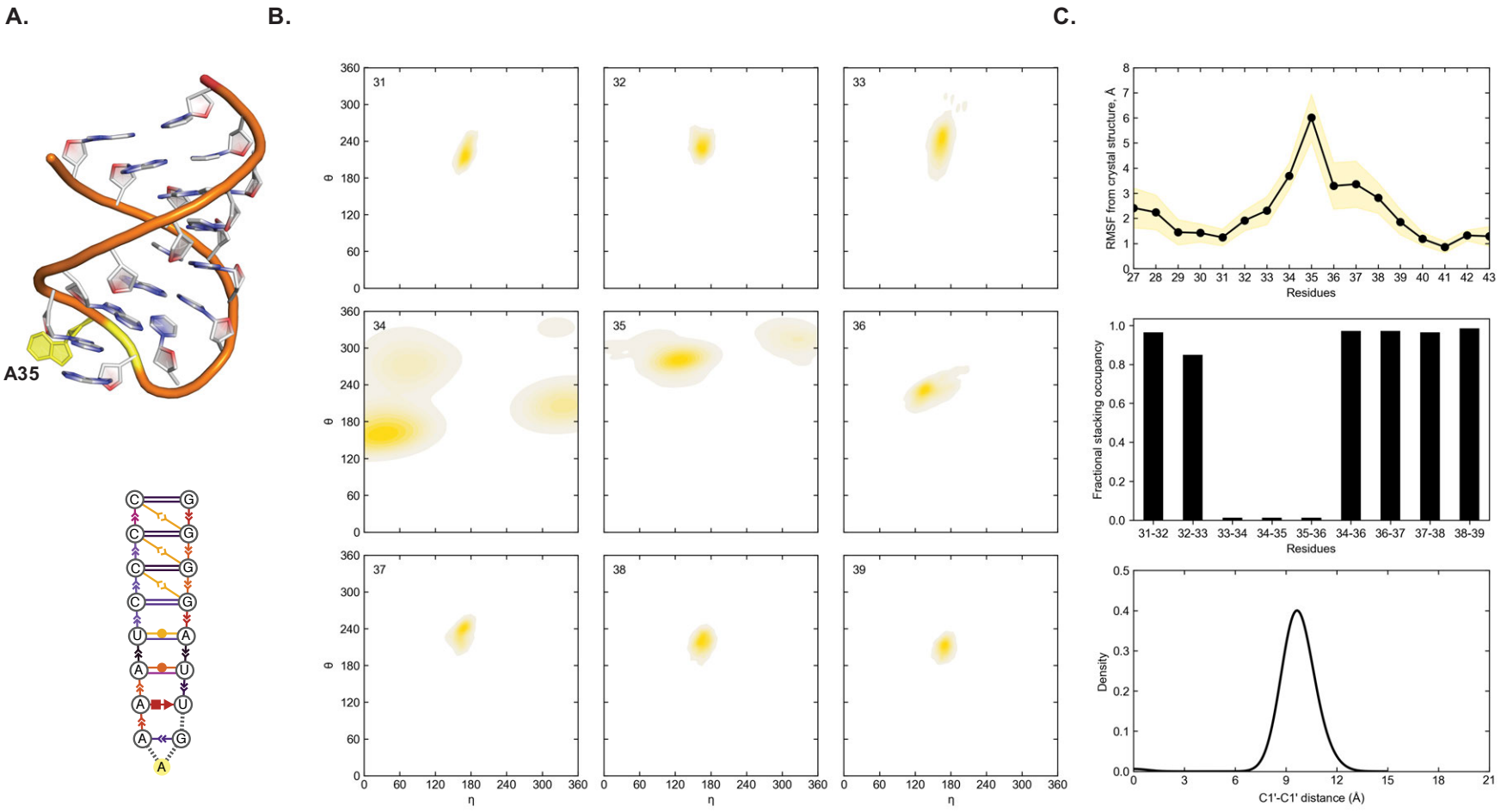


Figure 3A.10 – Description of the 35-unstacked conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 35-unstacked conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

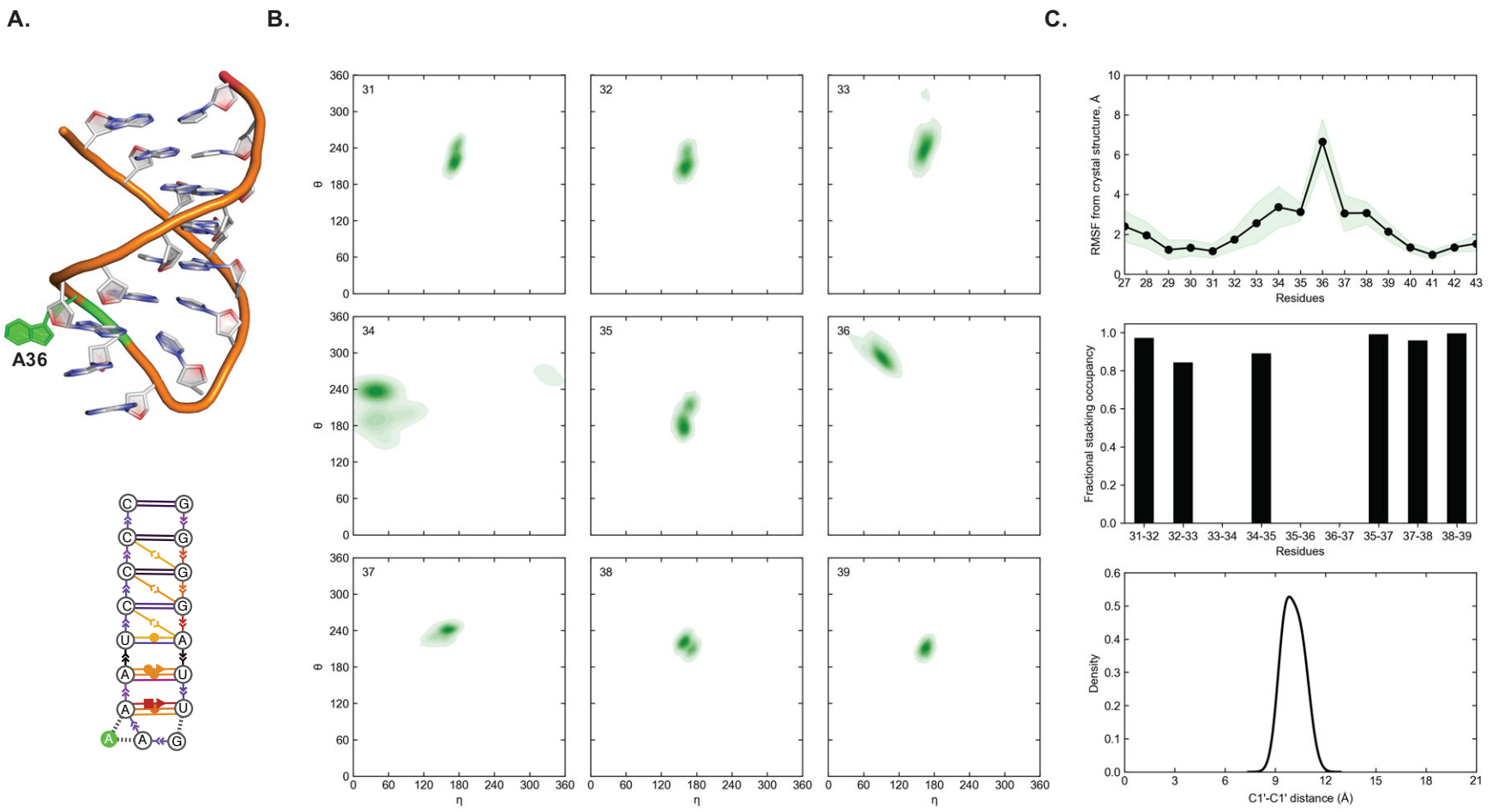


Figure 3A.11 – Description of the 36-unstacked conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 36-unstacked conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

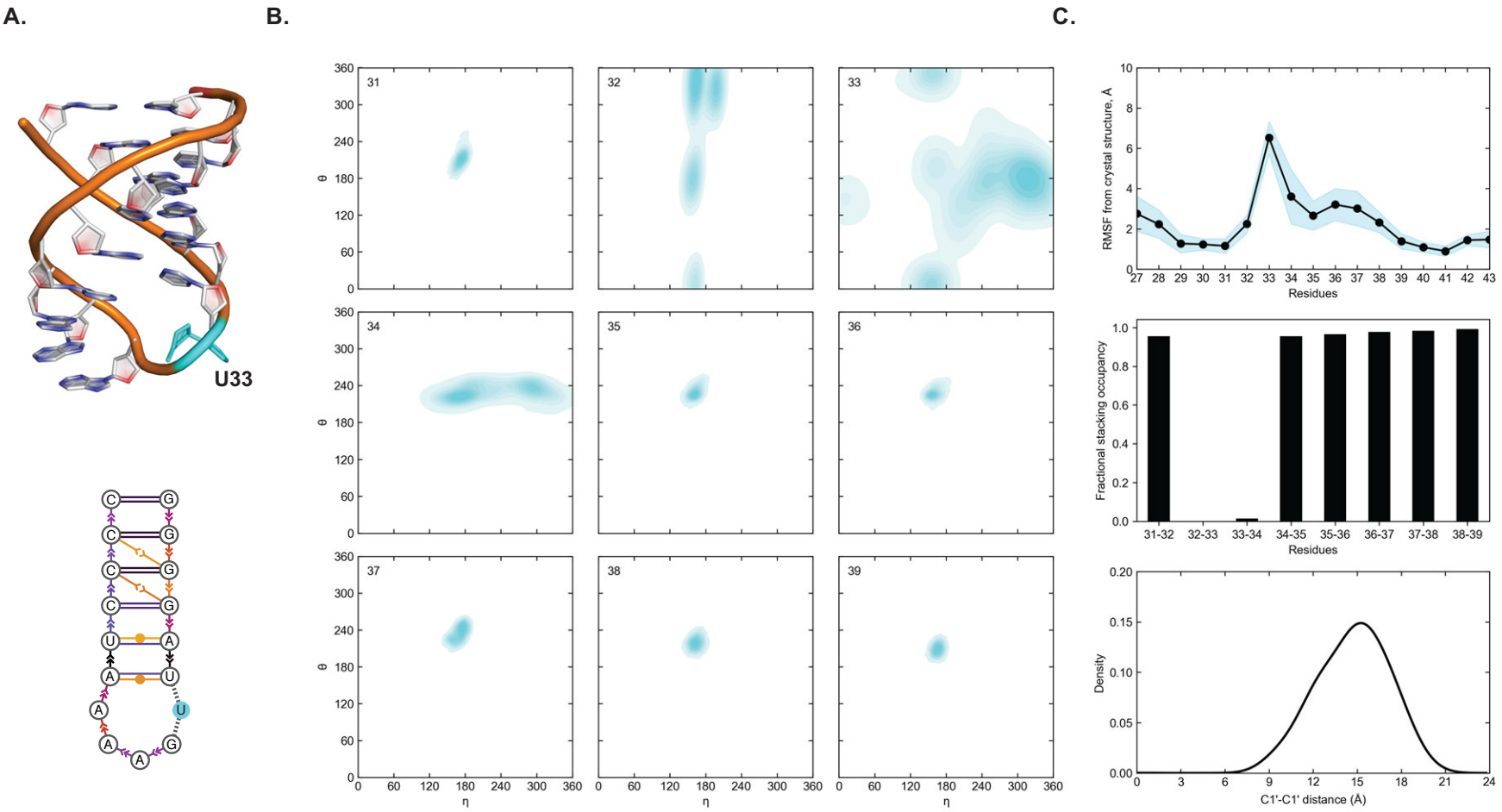


Figure 3A.12 – Description of the 33-out conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 33-out conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

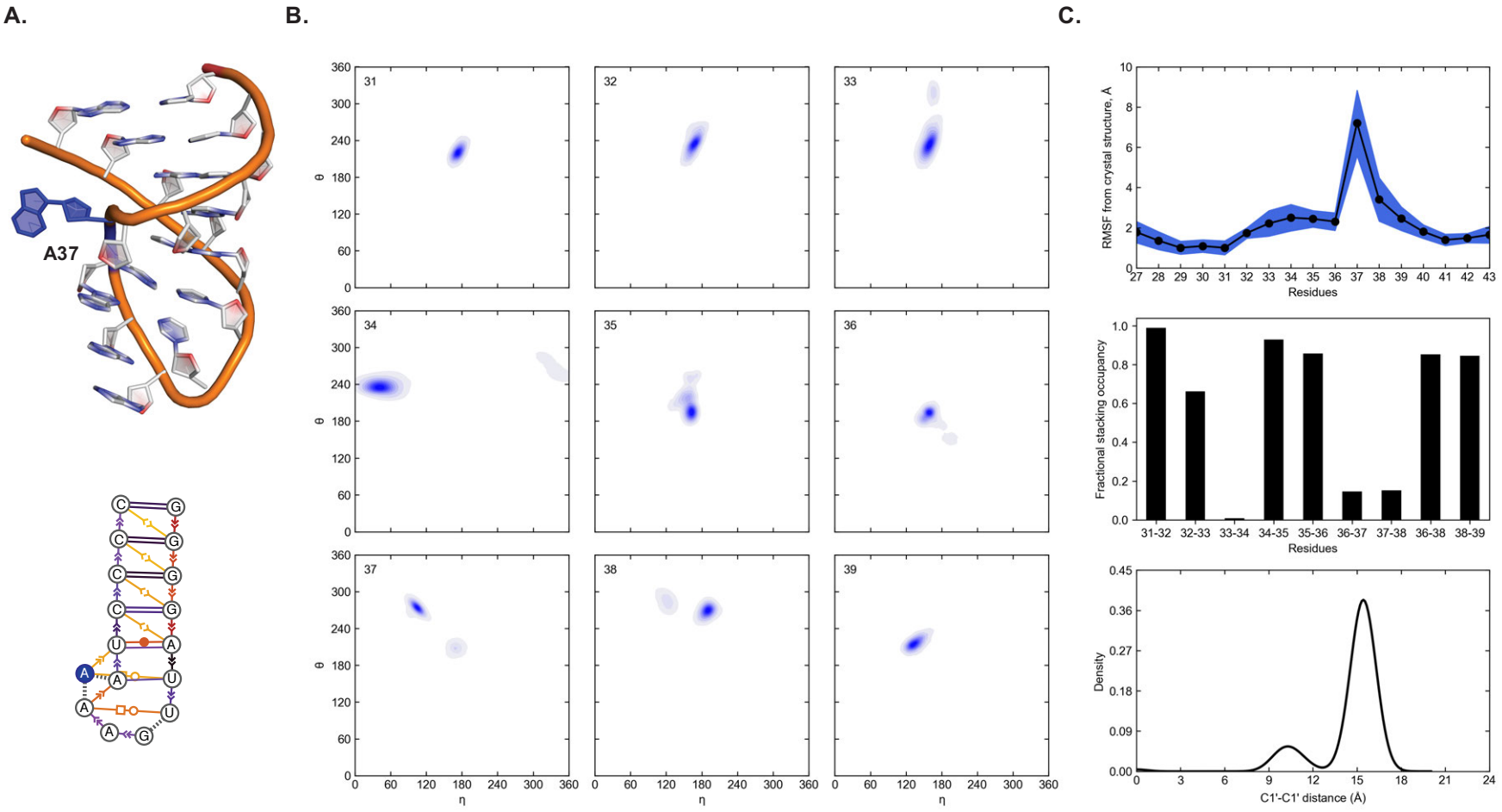


Figure 3A.13 – Description of the 37-out conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the 37-out conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

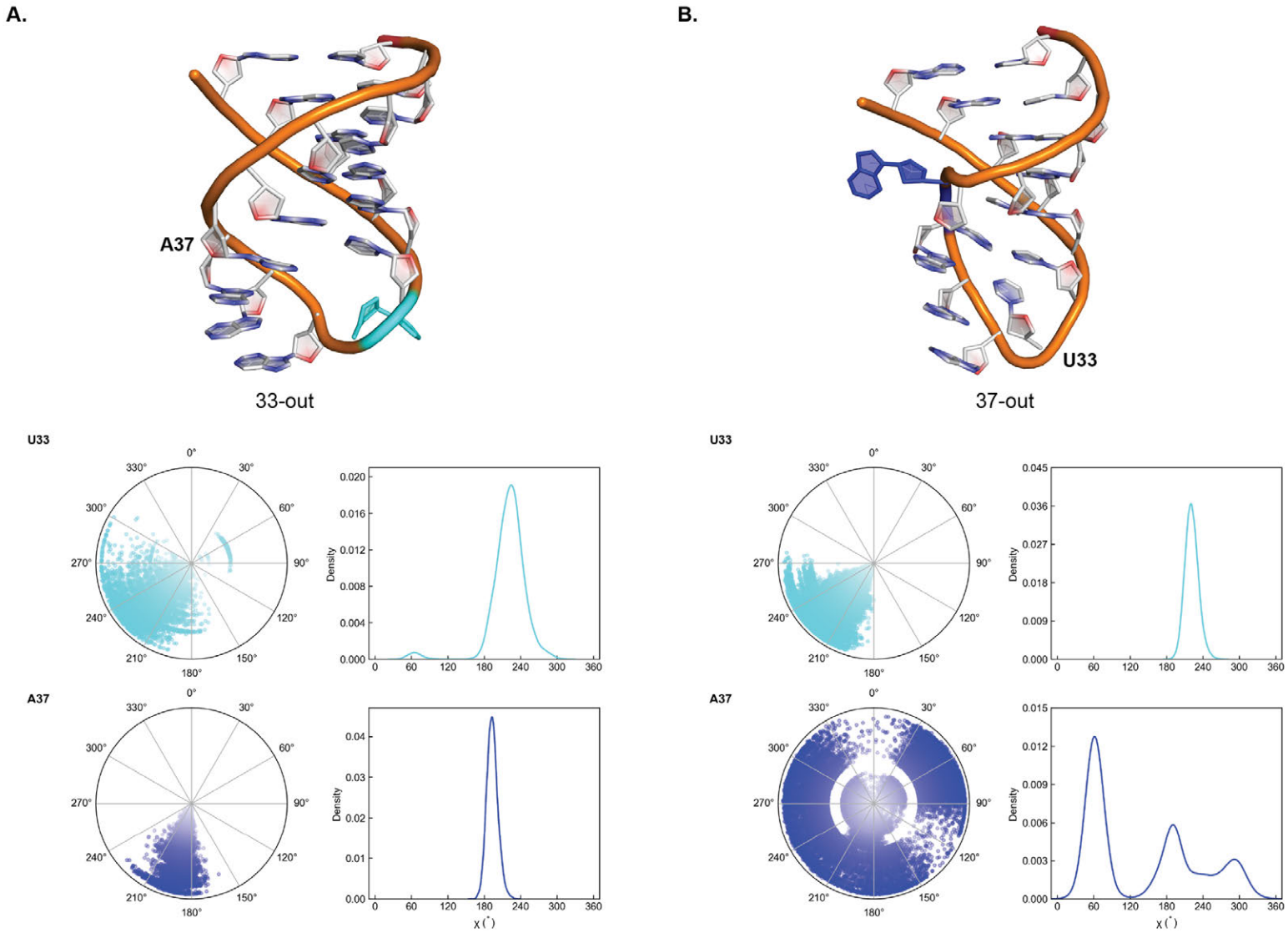


Figure 3A.14 – Flexibility of U33 and A37 when the adjacent base is flipped out of the loop

(A) Time (left) and density (right) maps for the glycosidic torsions (c) of U33 (top) and A37 (bottom) in the 33-out conformation. (B) Time (left) and density (right) maps for the glycosidic torsions (c) of U33 (top) and A37 (bottom) in the 37-out conformation.

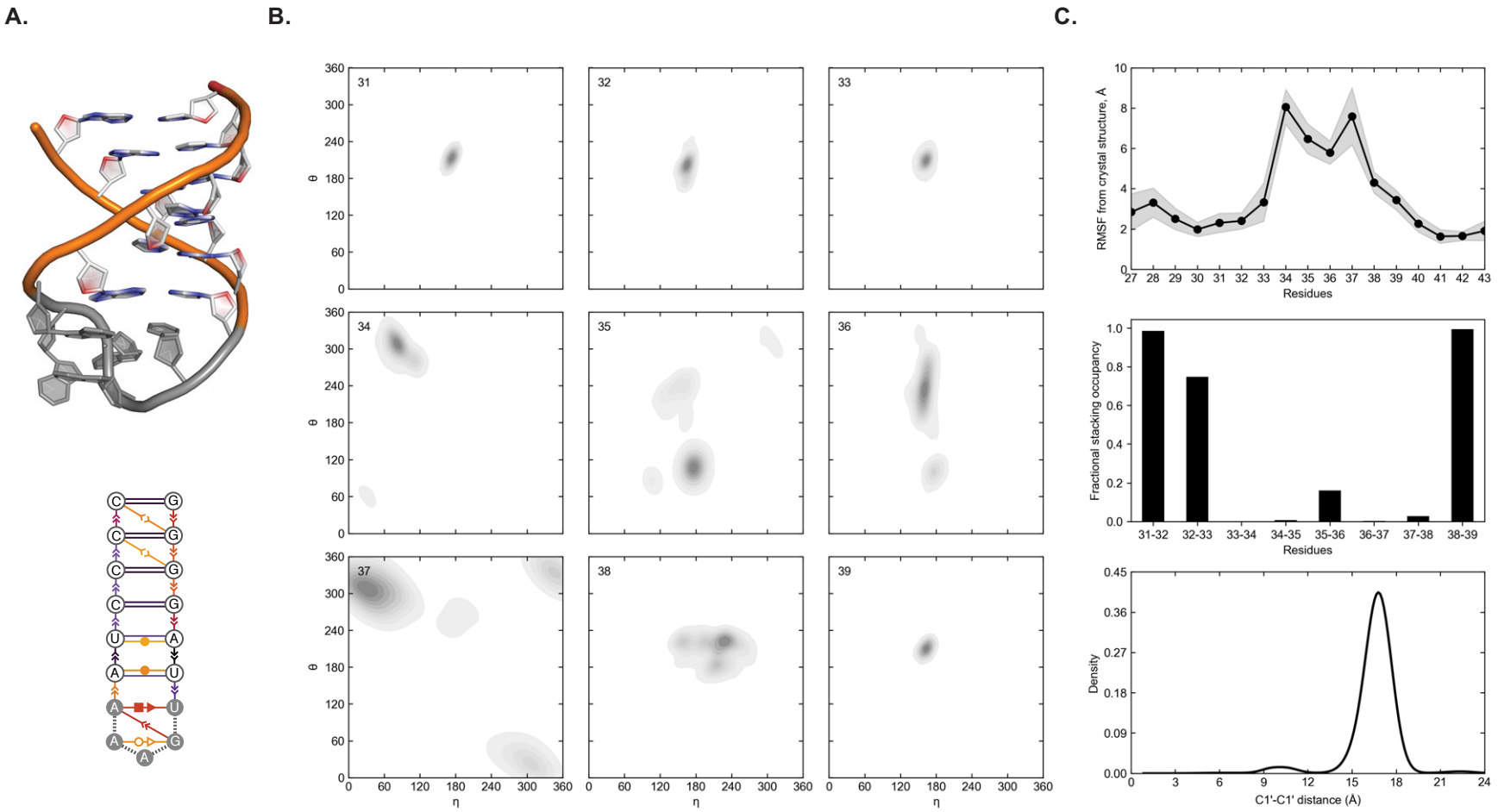


Figure 3A.15 – Description of the disordered conformation

(A) Representative structure (top) and non-covalent interactions present (bottom) in the disordered conformation. (B) η–θ plots for residues 31 to 39 of the ASL. (C) Per residue RMSFs at the ASL (top), prevalence of stacking interactions within the anticodon loop (middle) and the loop opening described using the C1'–C1' distance of U33 and A37.

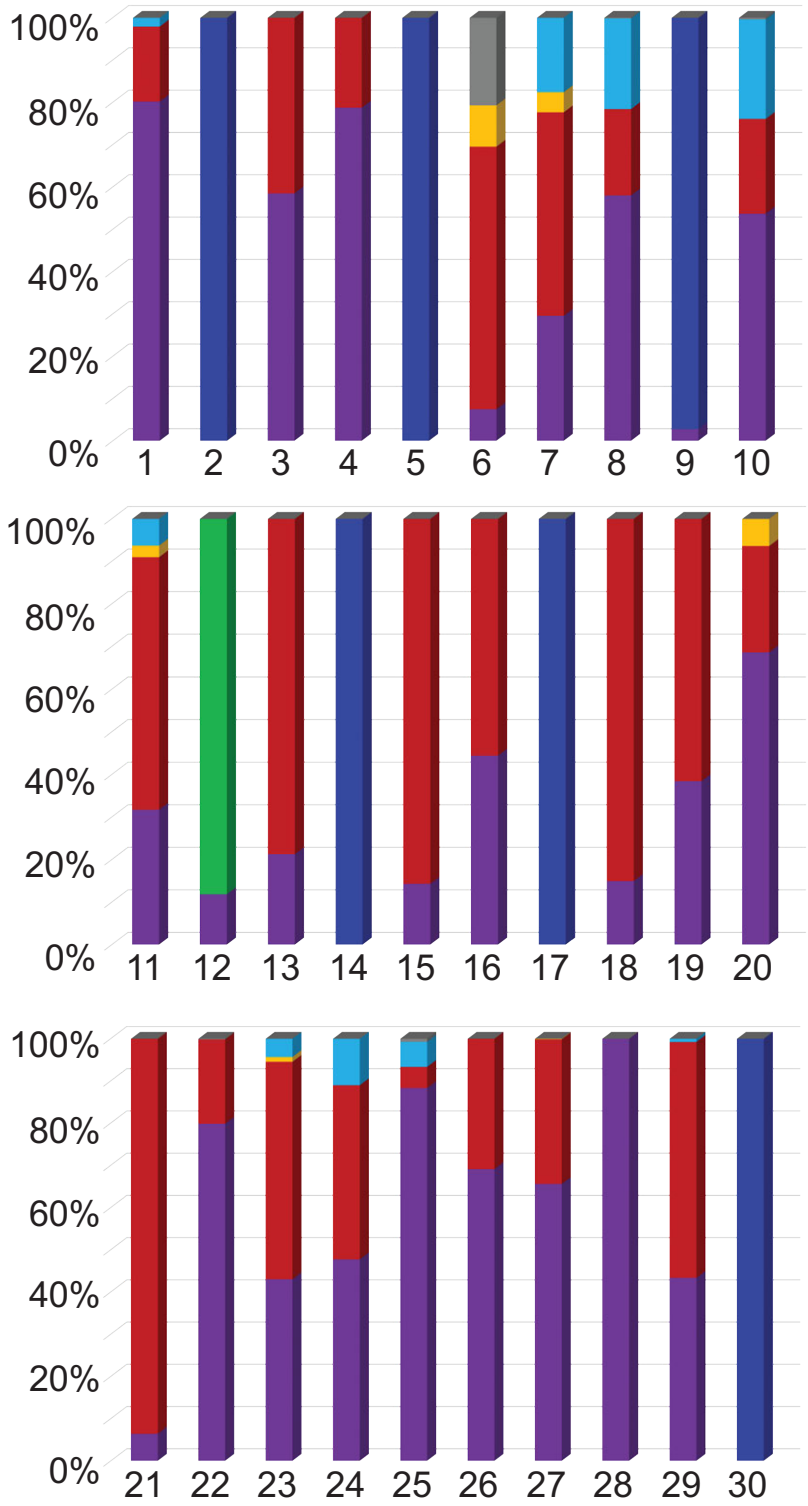


Figure 3A.16 – Conformational sampling across each replica simulation

Conformational sampling for each 500-ns replica trajectory, color-coded according to the scheme from Figure 2; 34-stacked–purple, 34-unstacked–red, 35-unstacked–yellow, 36-unstacked–green, 33-out–light blue, 37-out–dark blue, disordered–grey.

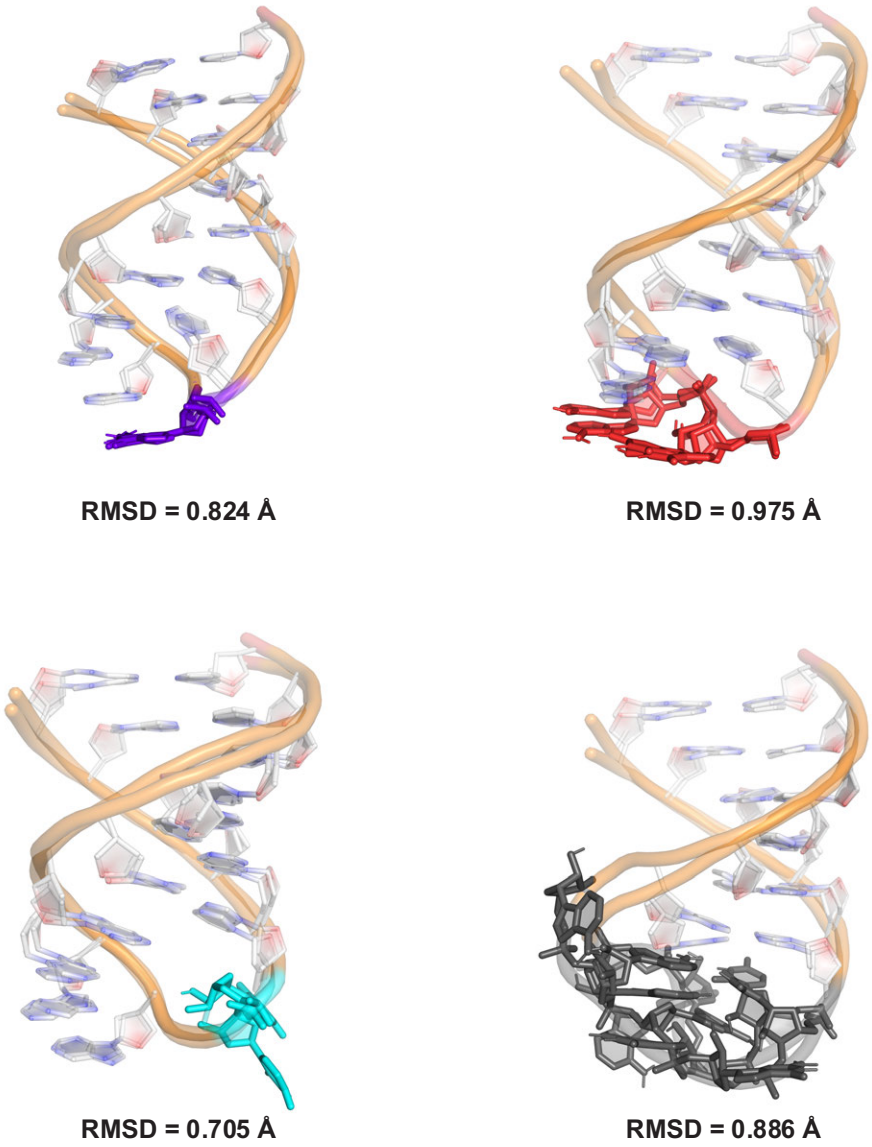


Figure 3A.17 – Relative conformational sampling in the cMD and rMD ensemble

Comparisons of representative structures for the four ASL states isolated in the cMD trajectory relative to their rMD counterparts. From left to right, top to bottom: 34-stacked, 34-unstacked, 33-out, and disorder.

APPENIDIX II: SUPPLEMENTARY INFORMATION FOR CHAPTER 4

Figures 4A.1 – 4A.32 and Tables 4A.1 – 4A.4

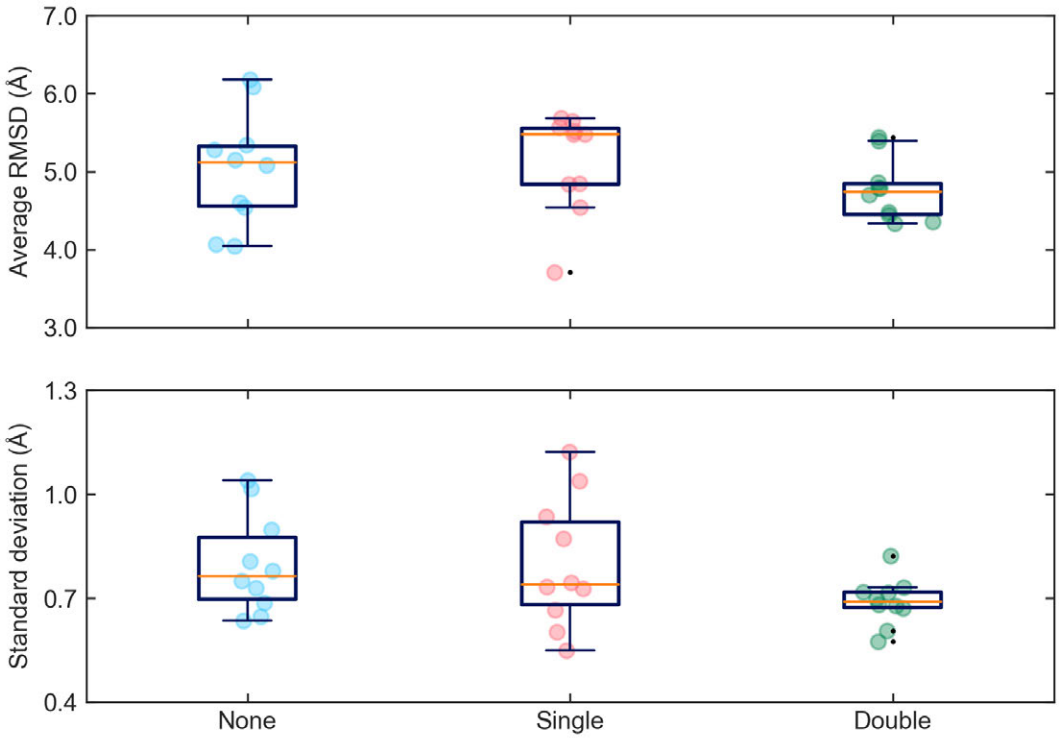


Figure 4A.1– Stability of MD simulations on tRNA^{Trp}

Statistical analysis of RMSDs for unmodified tRNA^{Trp}, cmnm^{5}U34-tRNA^{Trp} and cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}

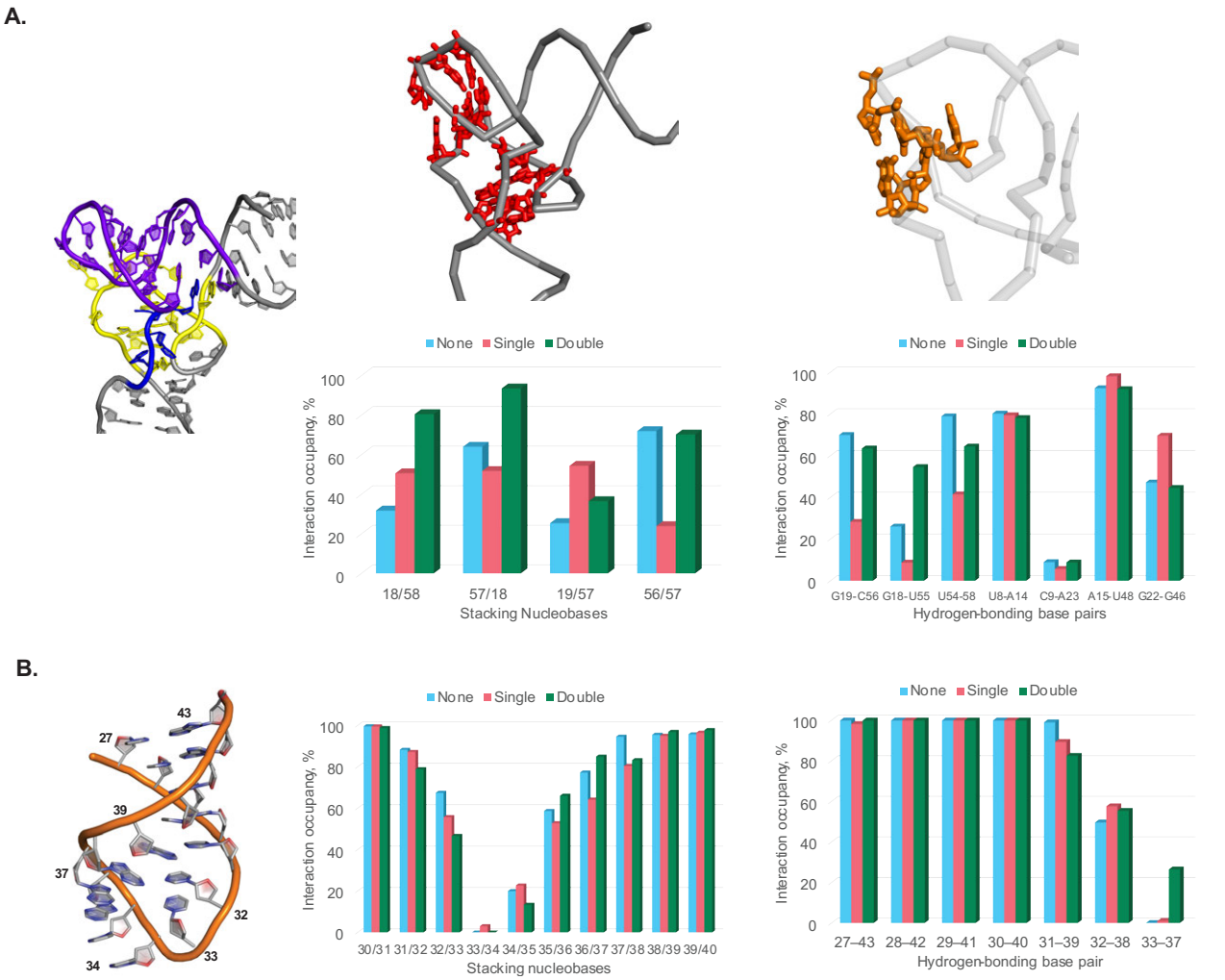


Figure 4A.2 – Non-covalent interactions within tRNA^{Trp}

(A) Persistence of tertiary hydrogen-bonding (red) and stacking (orange) interactions at the tRNA^{Trp} elbow. The tRNA elbow is color-coded by the domains involved – D arm (yellow), variable loop (blue), TyC arm (purple). Interactions are color-coded by system, namely unmodified tRNA^{Trp} (cyan), cmnm^{5}U34-tRNA^{Trp} (pink) and cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp} (green). (B) Non-covalent interactions within the ASL domain of tRNA^{Trp}. Occupancies for the hydrogen-bonding (middle) and stacking (right) interactions within the ASL across ten 500-ns replica ensembles. Interactions are color-coded by system, namely unmodified tRNA^{Trp} (cyan), cmnm^{5}U34-tRNA^{Trp} (pink) and cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp} (green).

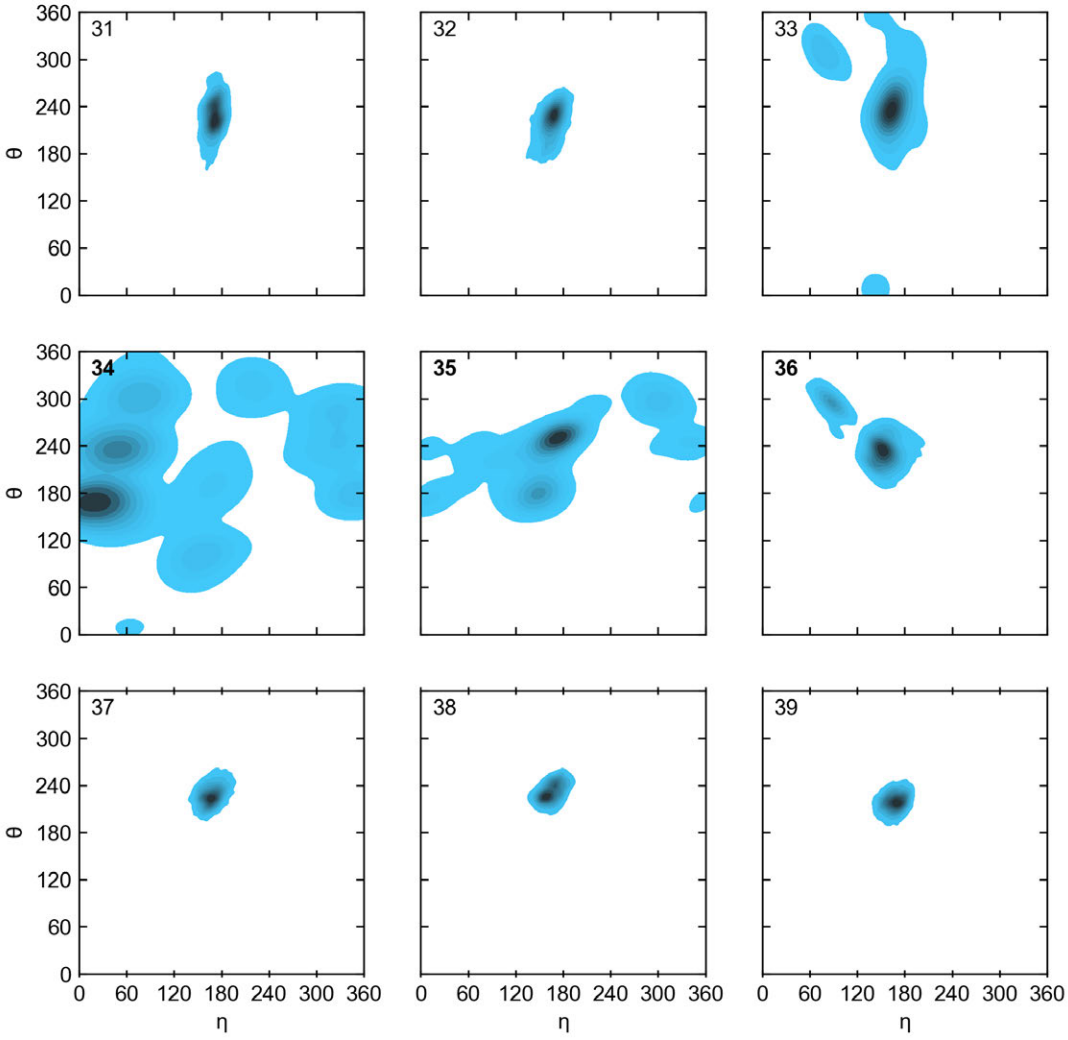


Figure 4A.3 – Pseudorotational analysis of the anticodon loop of unmodified tRNA^{Trp}

Backbone pseudotorsions (η = ∠C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = ∠C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in unmodified tRNA^{Trp}.

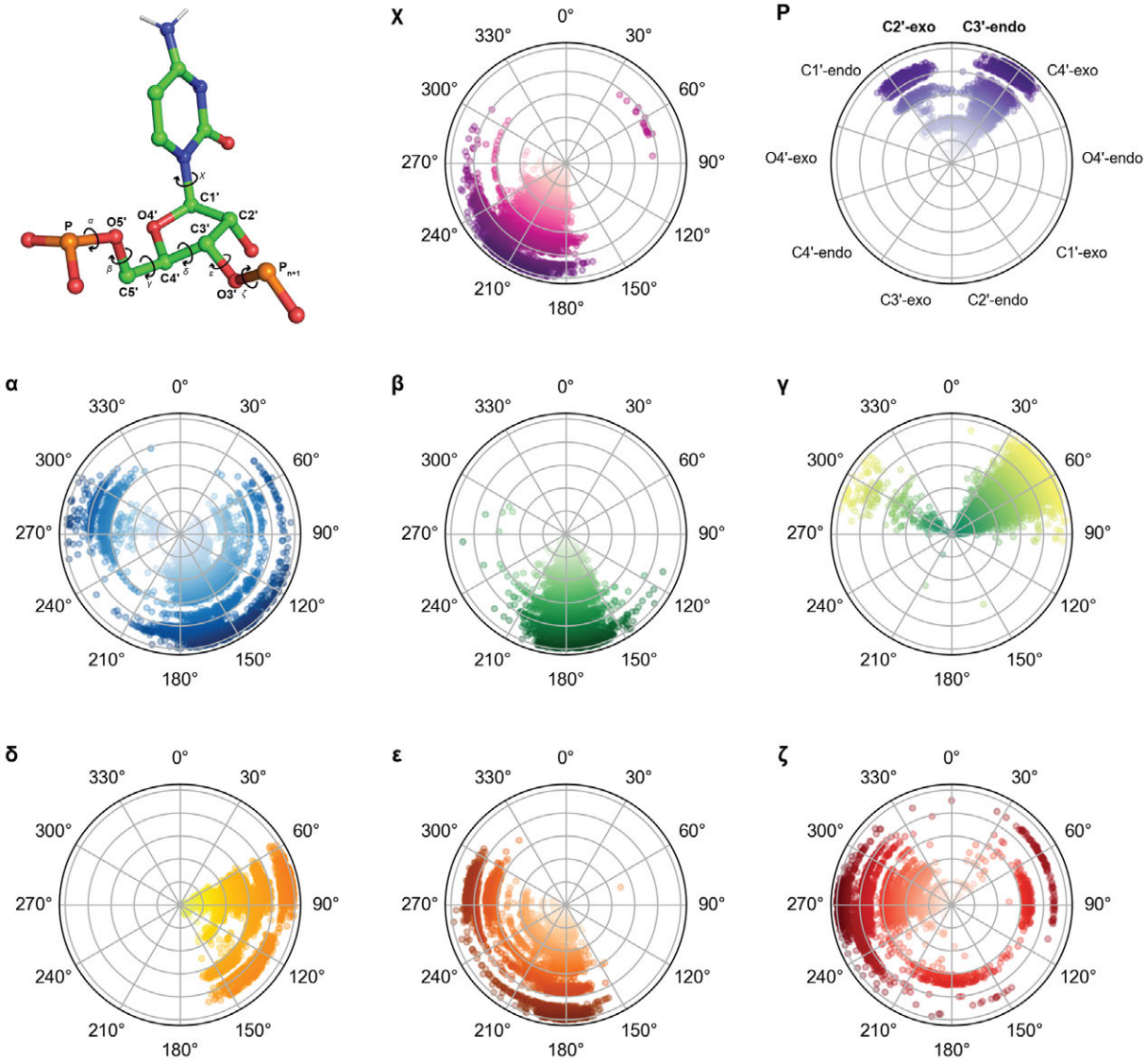


Figure 4A.4 – Backbone dynamics of C34 in unmodified tRNA^{Trp}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of C34 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red). Simulation time (5 μs) is represented on the r-axes of each polar plot.

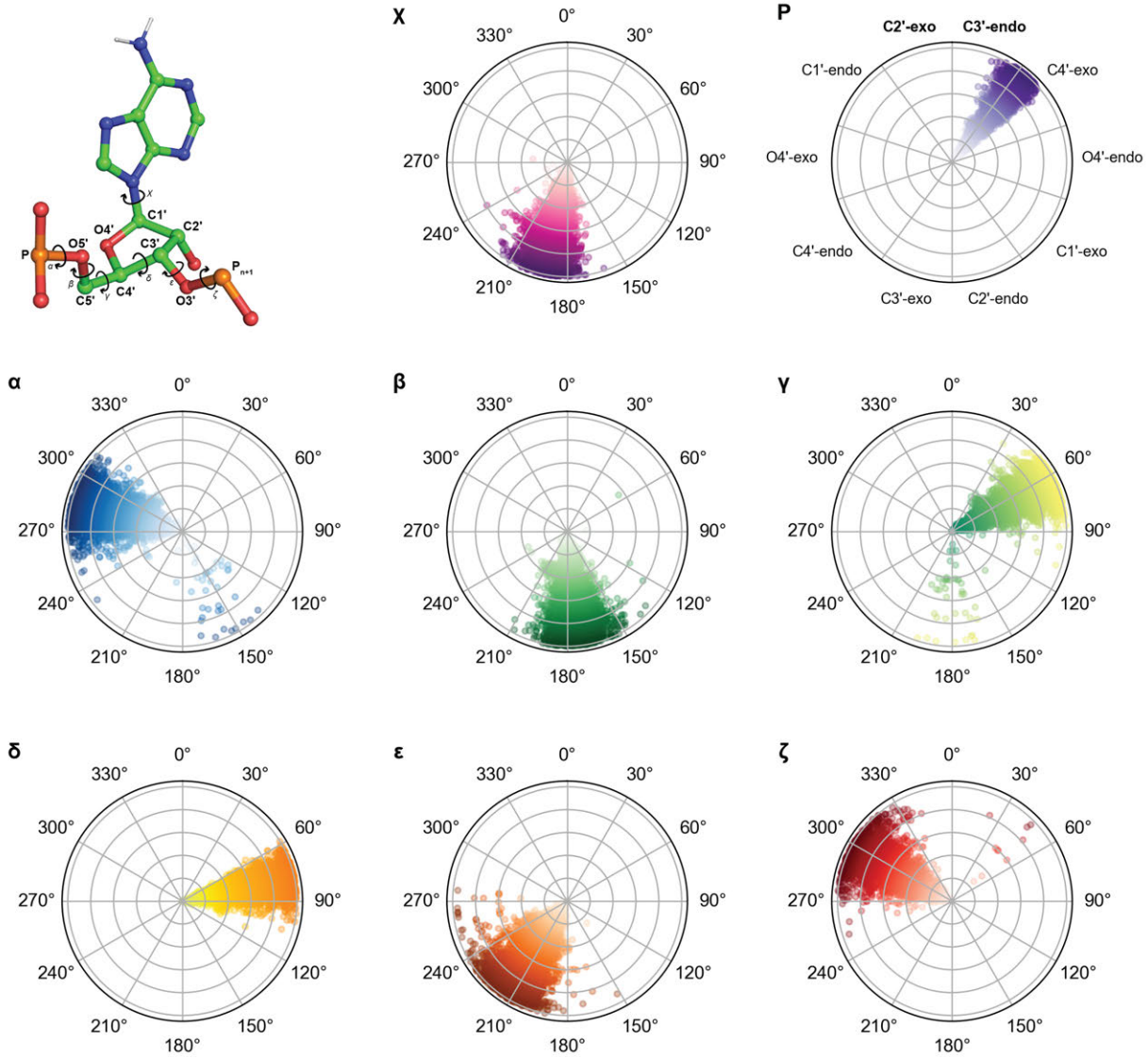


Figure 4A.5 – Backbone dynamics of A37 in unmodified tRNA^{Trp}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of A37 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red). Simulation time (5 μs) is represented on the r-axes of each polar plot.

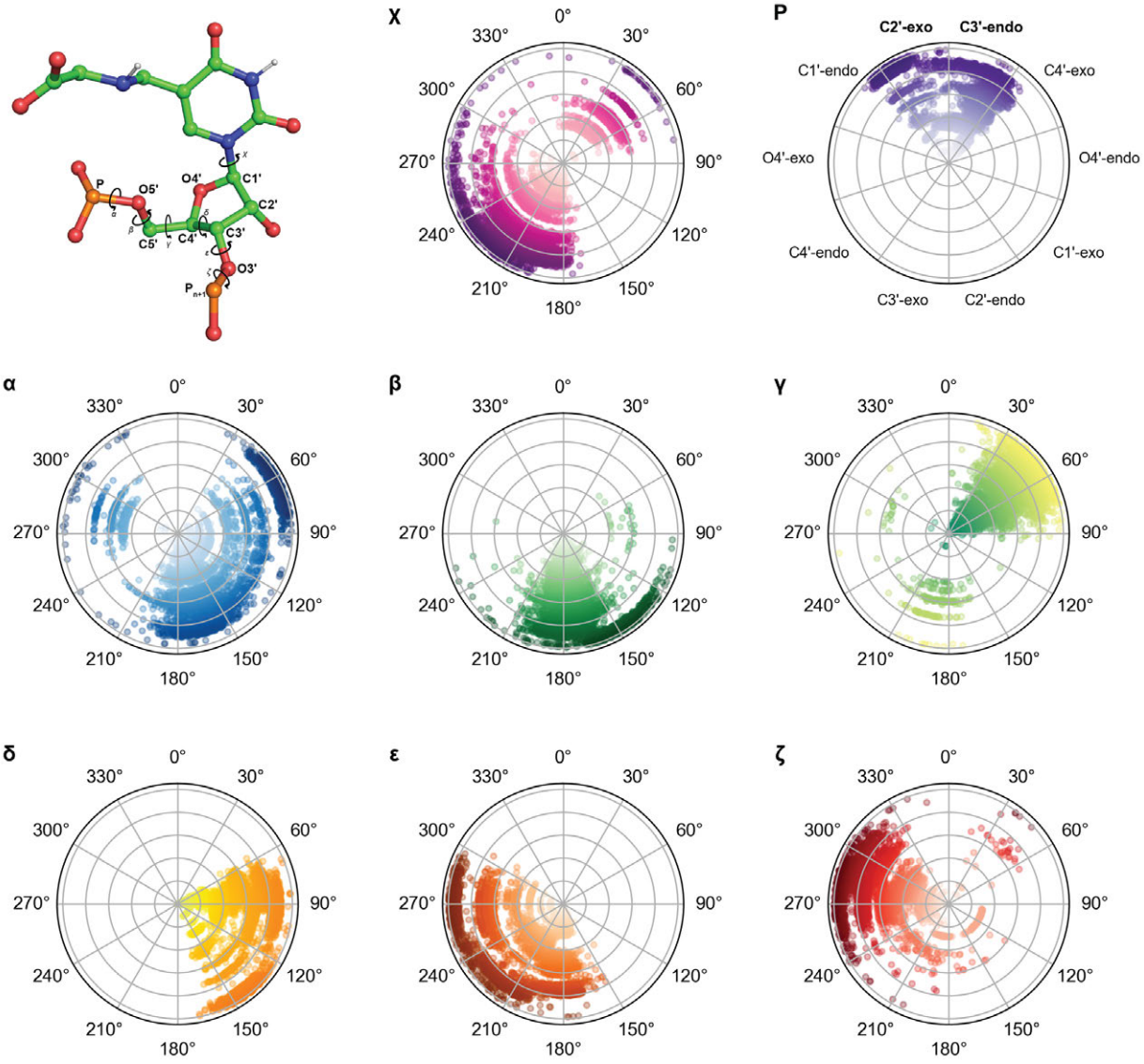


Figure 4A.6 – Backbone dynamics of cmnm^{5}U34 in singly modified tRNA^{Trp}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of cmnm^{5}U34 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red). Simulation time (5 μs) is represented on the r-axes of each polar plot.

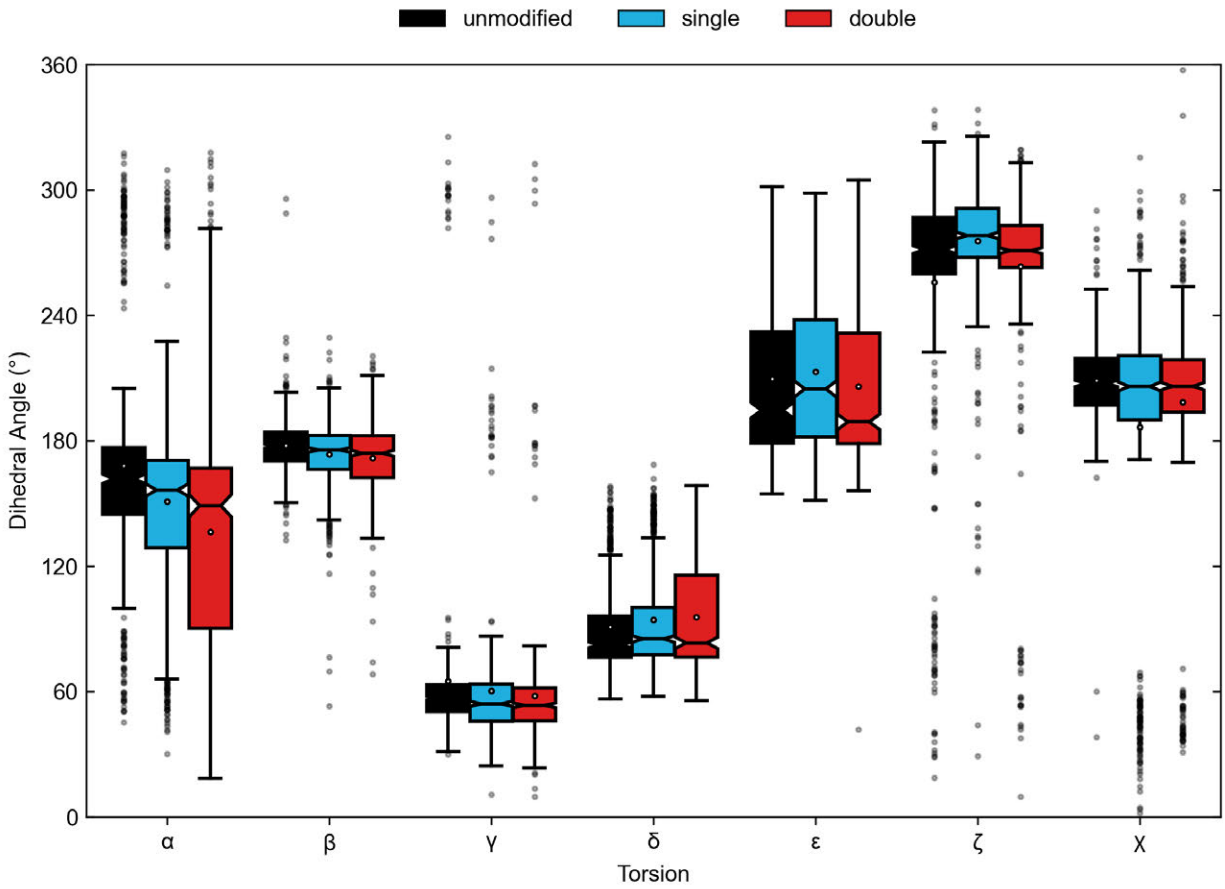


Figure 4A.7 – Dynamics of the backbone conformations at position 34 of tRNA^{Trp} in the presence of cmnm^{5}U34 and ms^{2}i^{6}A37 modifications

Statistical analysis of backbone torsions adopted at position 34 in unmodified, singly (cmnm^{5}U34-tRNA^{Trp}) and doubly (cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}) modified tRNA^{Trp}. The white dots represent the mean, rectangles represent interquartile ranges.

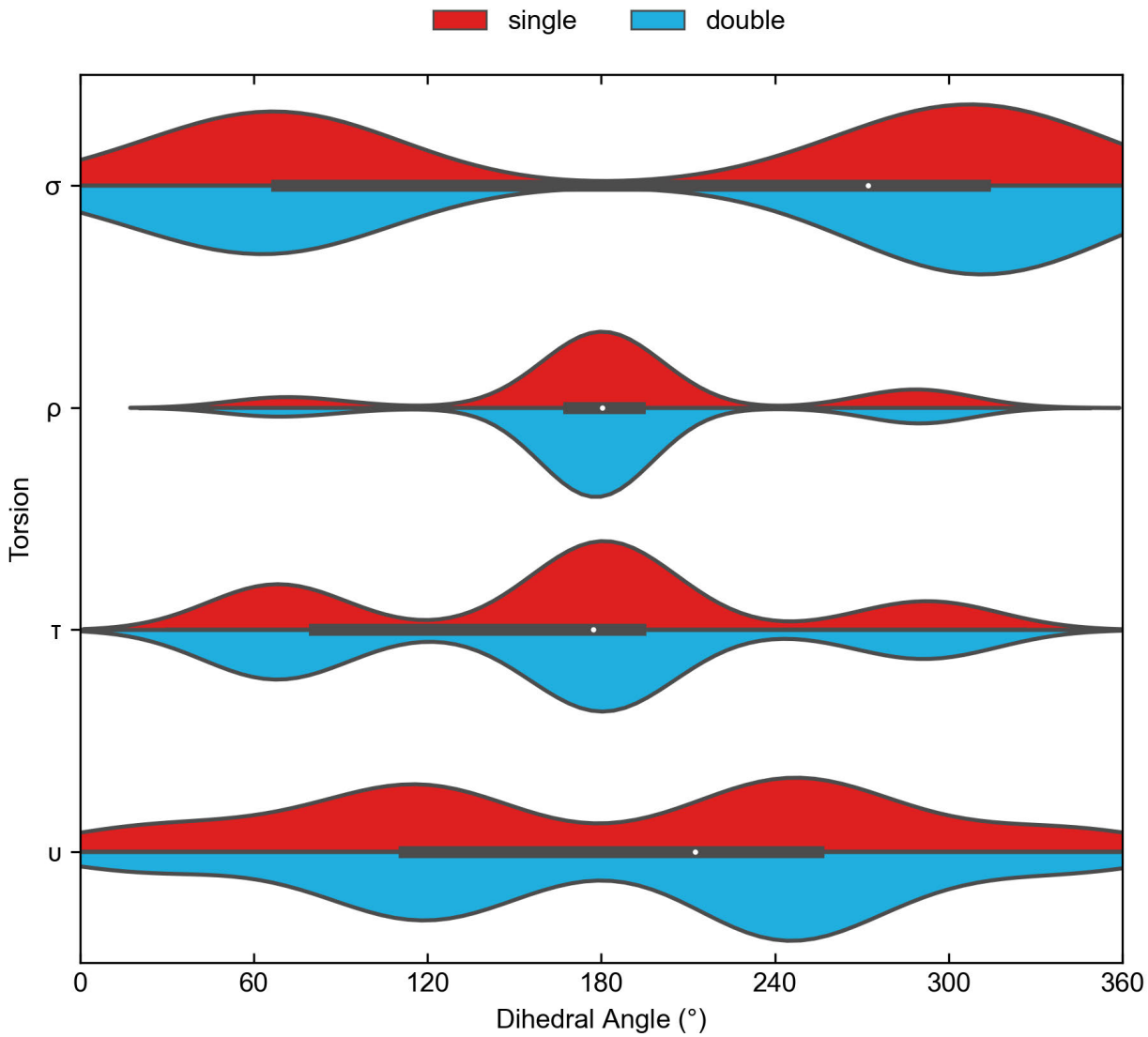


Figure 4A.8 – Dynamics of the cmnm^{5}U34 sidechain within the tRNA^{Trp} ASL

Violinplots comparing the dihedral angle distributions for the C5 moiety of cmnm^{5}U34 in single and double modification models of tRNA^{Trp}. The white dots represent the mean, rectangles represent interquartile ranges. Torsions are defined in Figure 4.3.

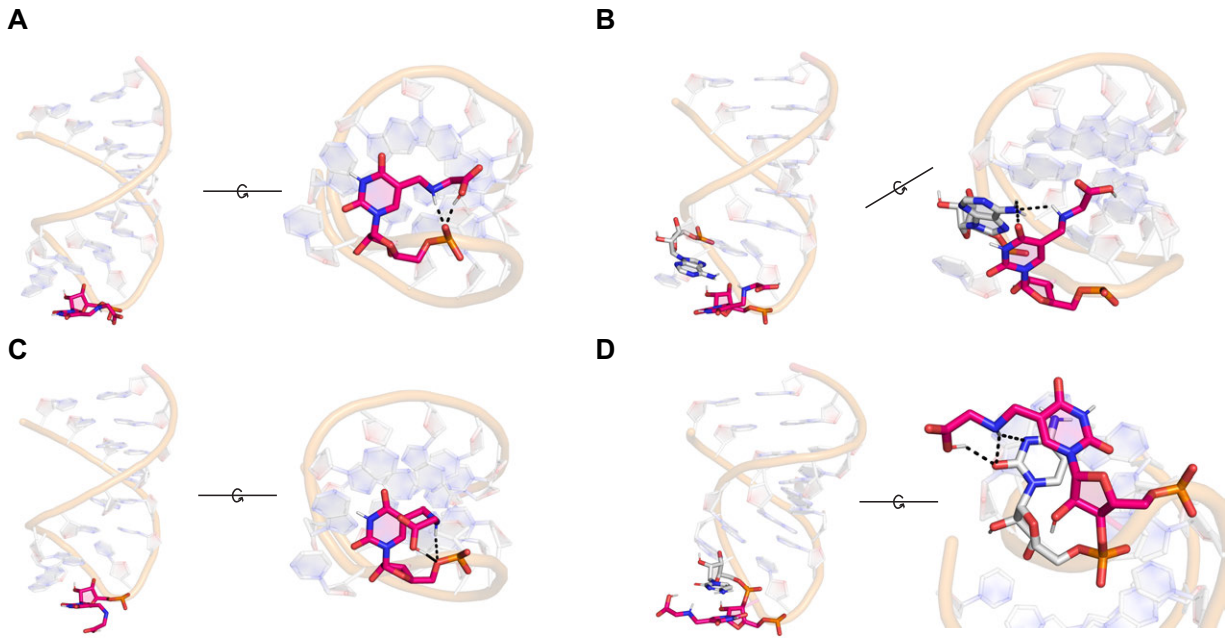


Figure 4A.9 – Persistent (< 50%) hydrogen-bonding interactions formed by cmnm^{5}U34 in singly modified tRNA^{Trp}

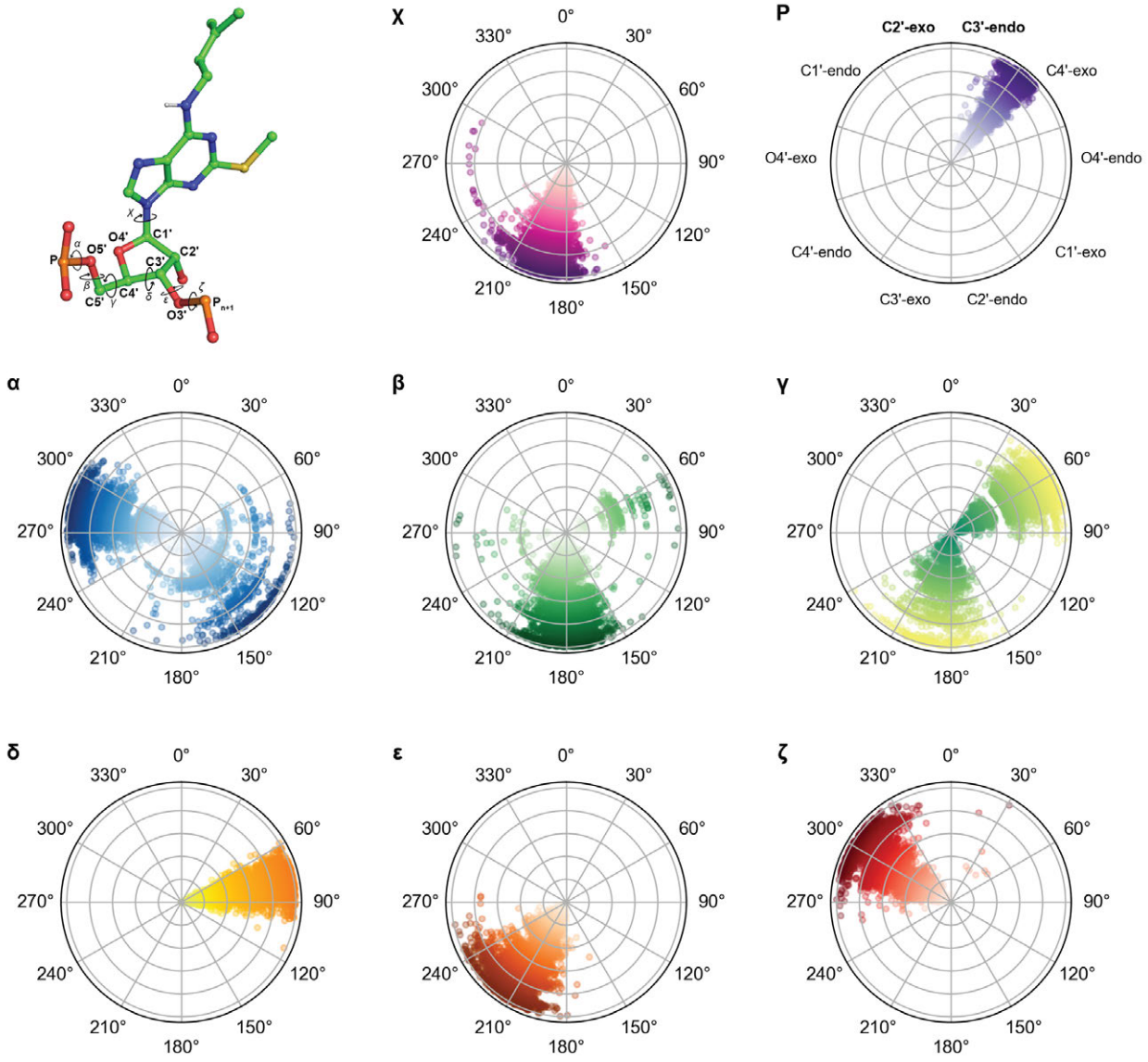


Figure 4A.10 – Backbone dynamics of ms^{2}i^{6}A37 in doubly modified tRNA^{Trp}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of ms^{2}i^{6}A37 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red). Simulation time (5 μs) is represented on the r-axes of each polar plot.

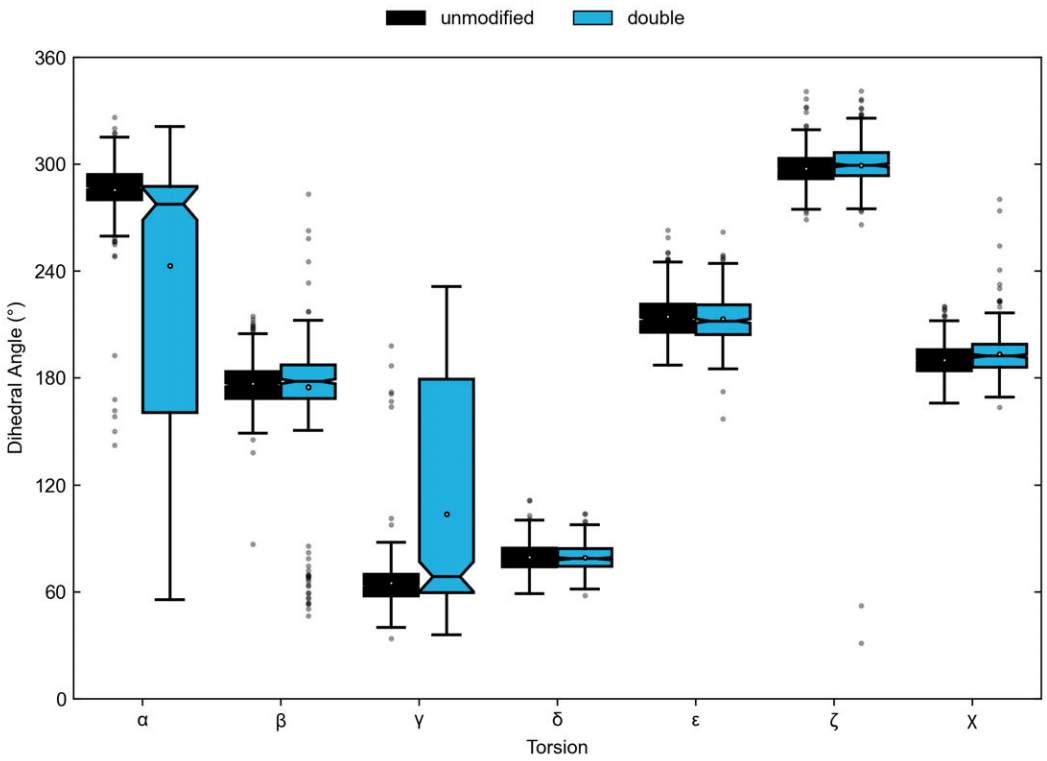


Figure 4A.11 – Backbone dynamics of ms^{2}i^{6}A37 within the tRNA^{Trp} ASL

Violin plots comparing the dihedral angle distributions for the backbone torsions at position 37 in unmodified and double modified models of tRNA^{Trp}. The white dots represent the mean, rectangles represent interquartile ranges.

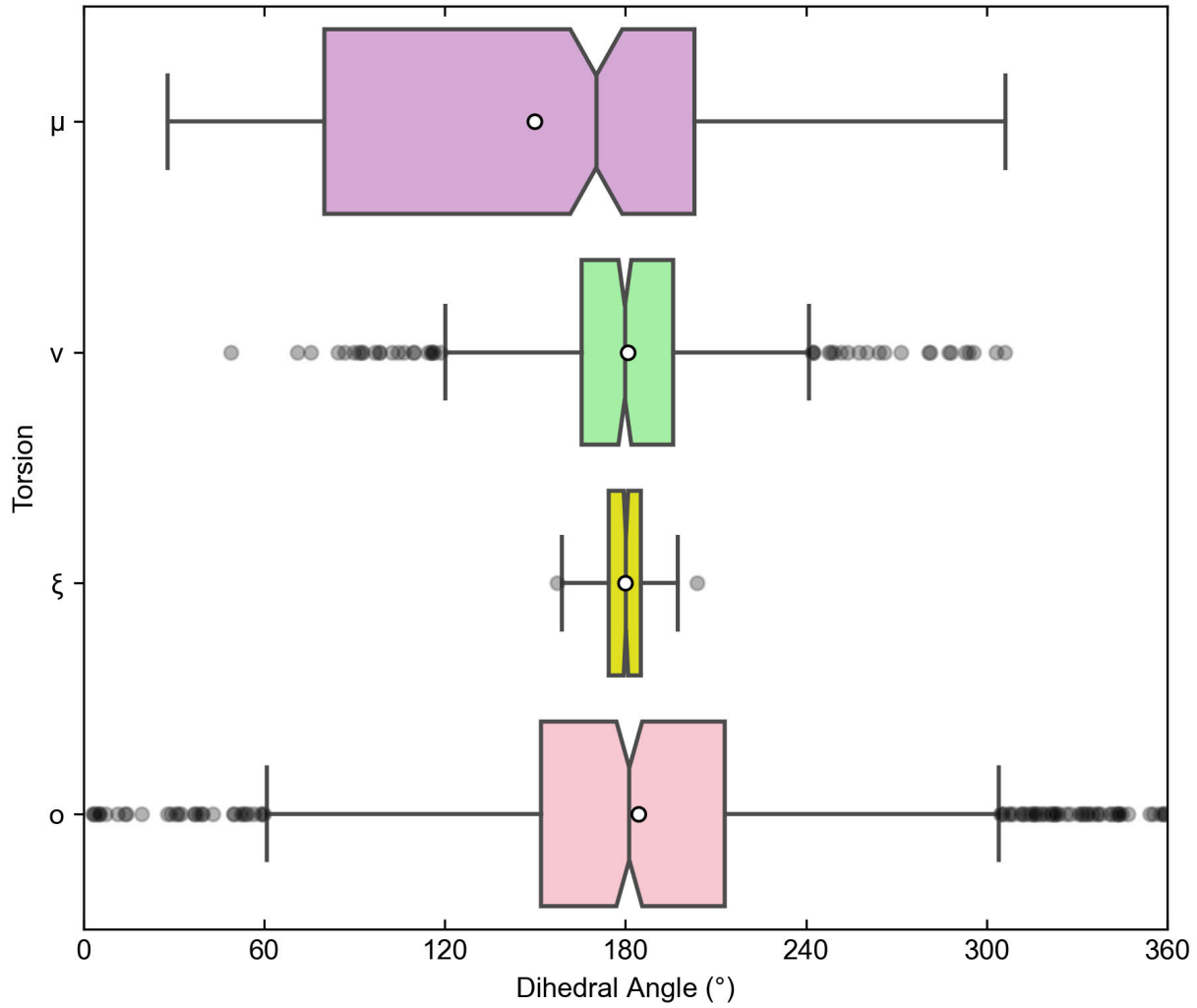


Figure 4A.12 – Dynamics of the ms^{2}i^{6}A37 sidechain within the tRNA^{Trp} ASL

Statistical analysis of side chain torsions adopted in ms^{2}i^{6}A37 doubly modified tRNA^{Trp}. The white dots represent the mean, rectangles represent interquartile ranges. Dihedral torsions are defined in Figure 4.3.

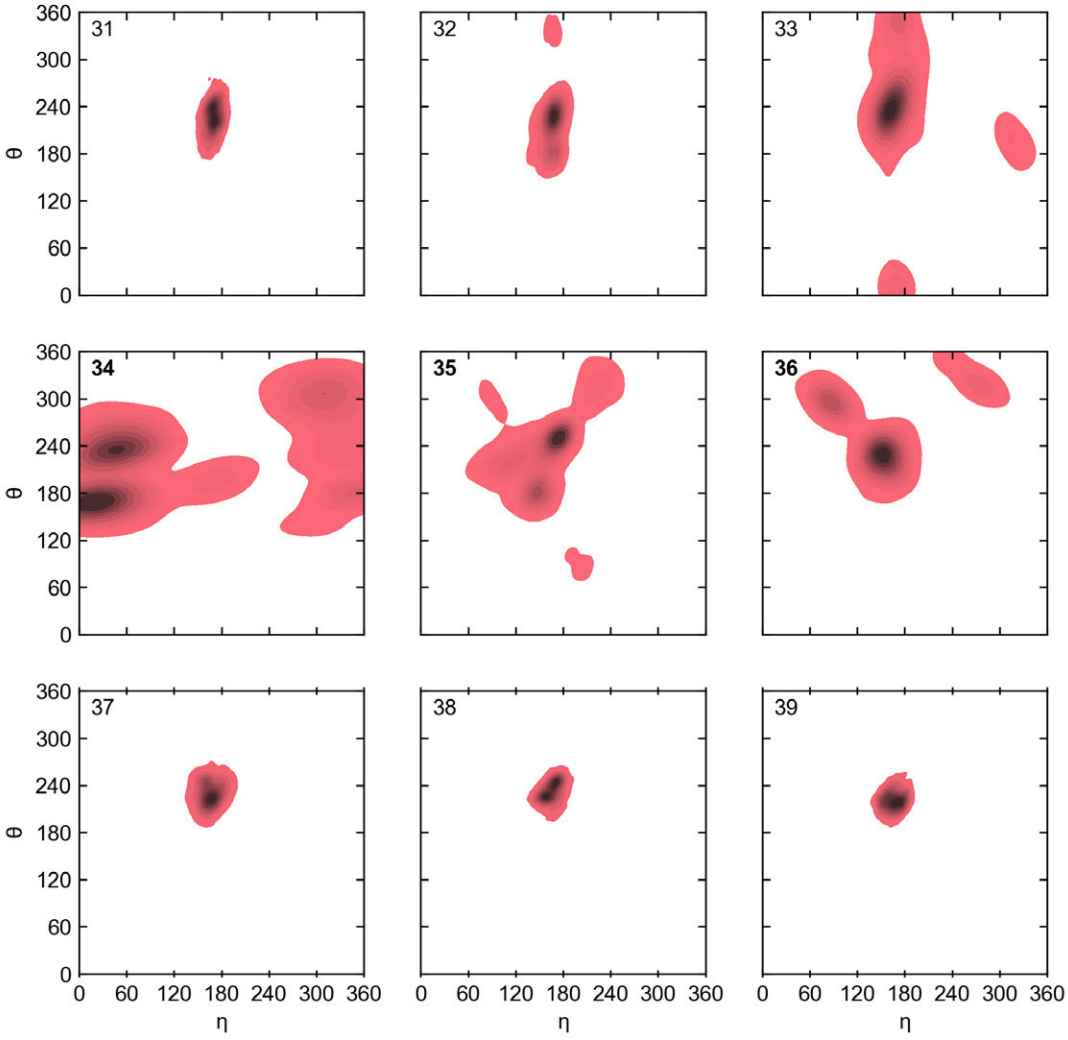


Figure 4A.13 – Pseudorotational analysis of the anticodon loop of cmnm^{5}U34-modified tRNA^{Trp}

Backbone pseudotorsions (η = ∠C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = ∠C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in cmnm^{5}U34-tRNA^{Trp}.

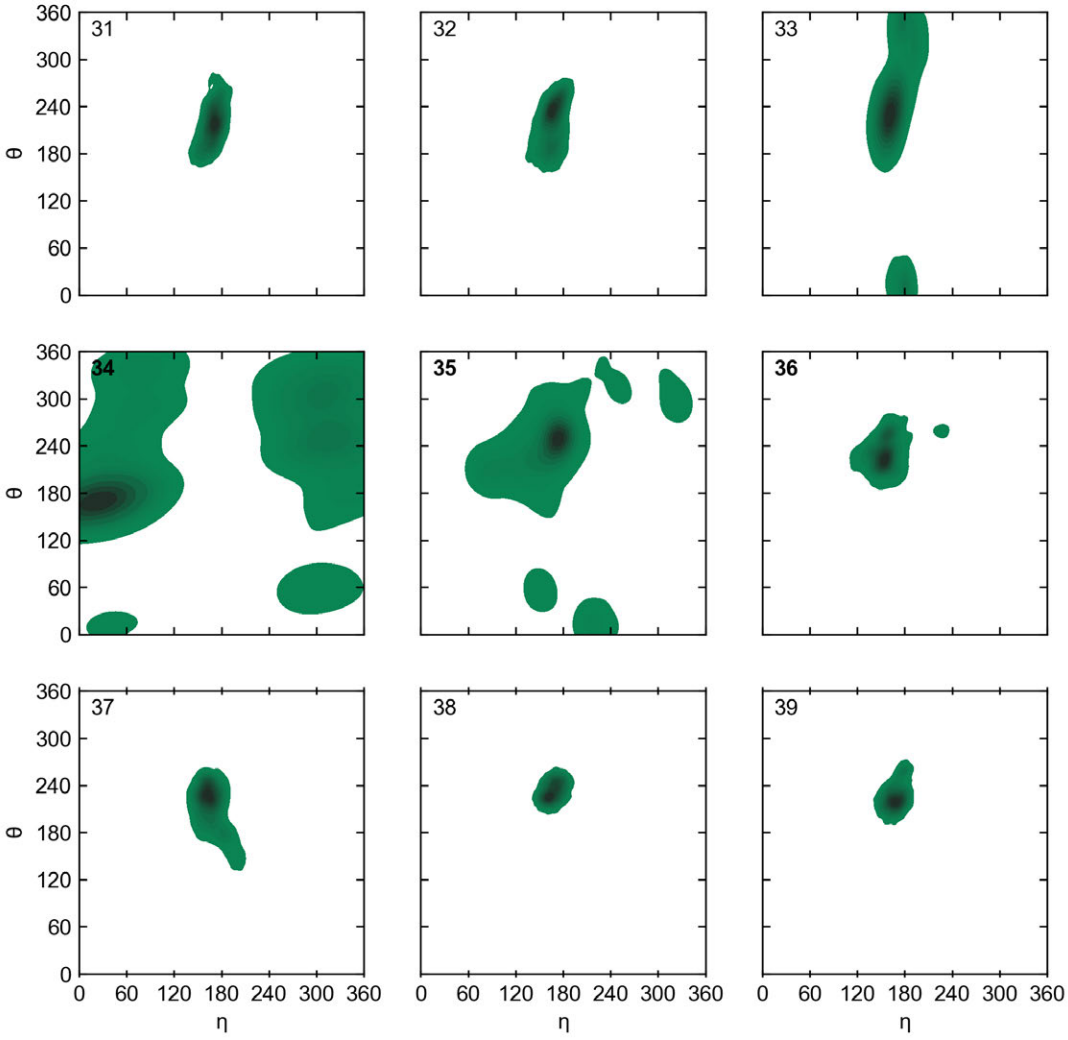


Figure 4A.14 – Pseudorotational analysis of the anticodon loop of cmnm^{5}U34-modified tRNA^{Trp}

Backbone pseudotorsions (η = $\angle$C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = $\angle$C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp}.

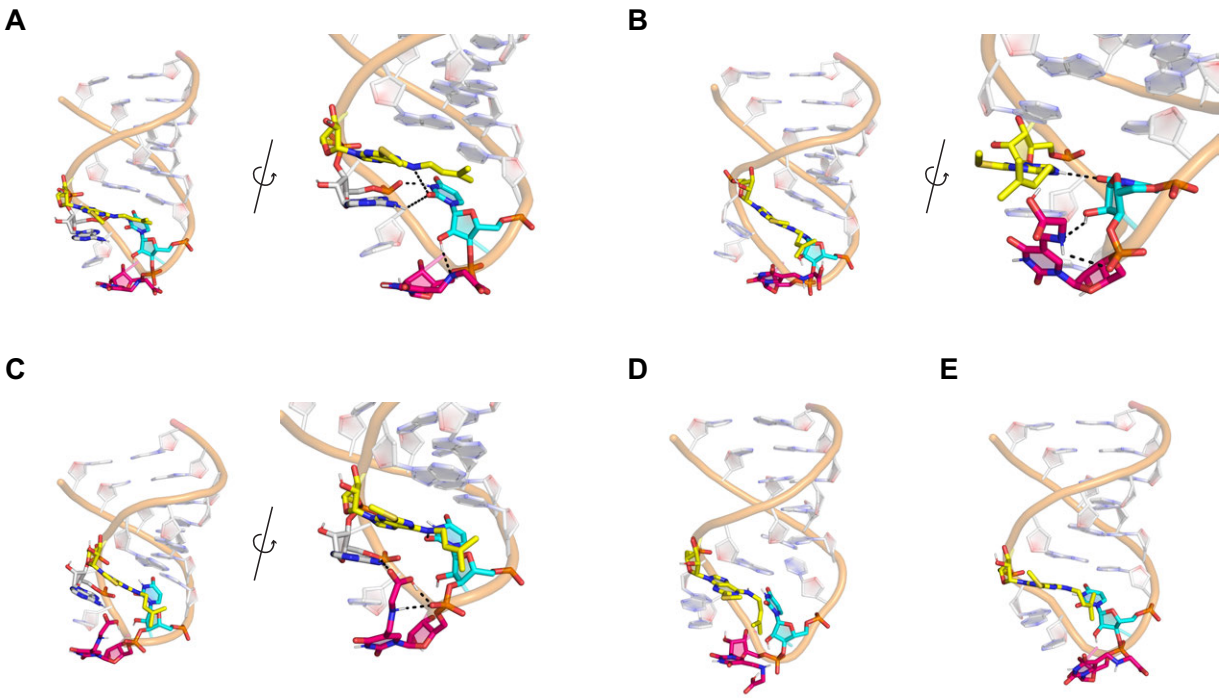


Figure 4A.15 – Widening of the ASL loop by cmnm^{5}U34 and ms^{2}i^{6}A37 in doubly modified tRNA^{Trp}.

Persistent hydrogen-bonding interactions between U33 (cyan), cmnm^{5}U34 (magenta) and ms^{2}i^{6}A37 (yellow) that stabilize an open loop conformation.

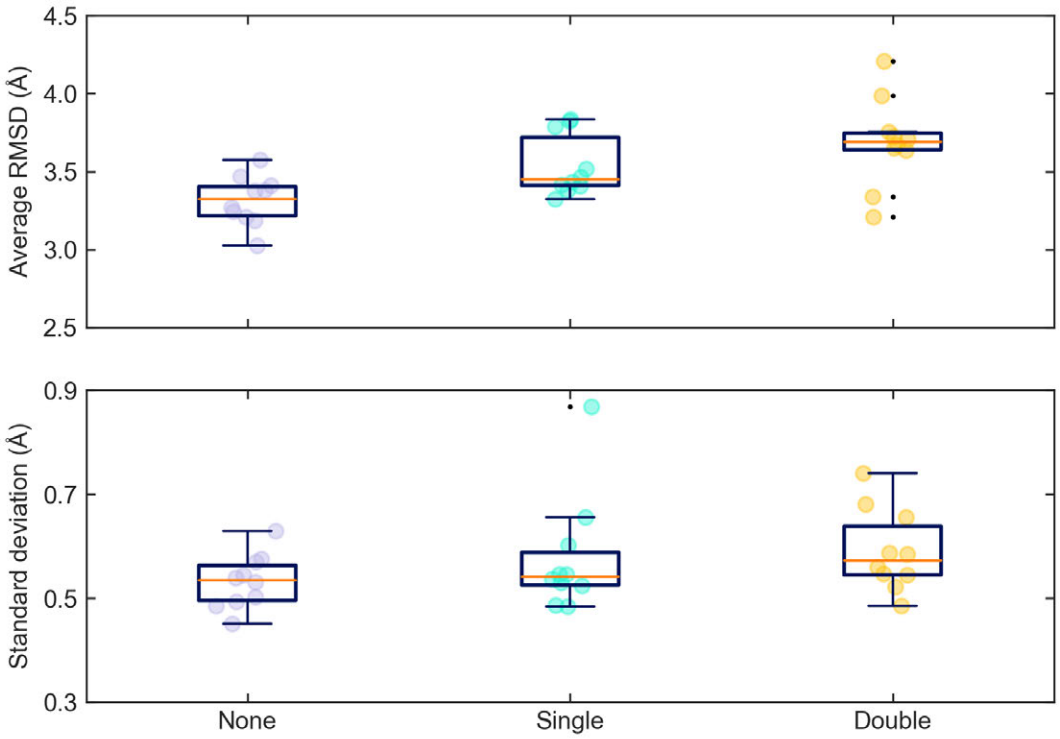


Figure 4A.16– Stability of MD simulations on tRNA^{Lys}

Statistical analysis of RMSDs for unmodified tRNA^{Lys}, cmnm^{5}s^{2}U34-tRNA^{Lys} and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}.

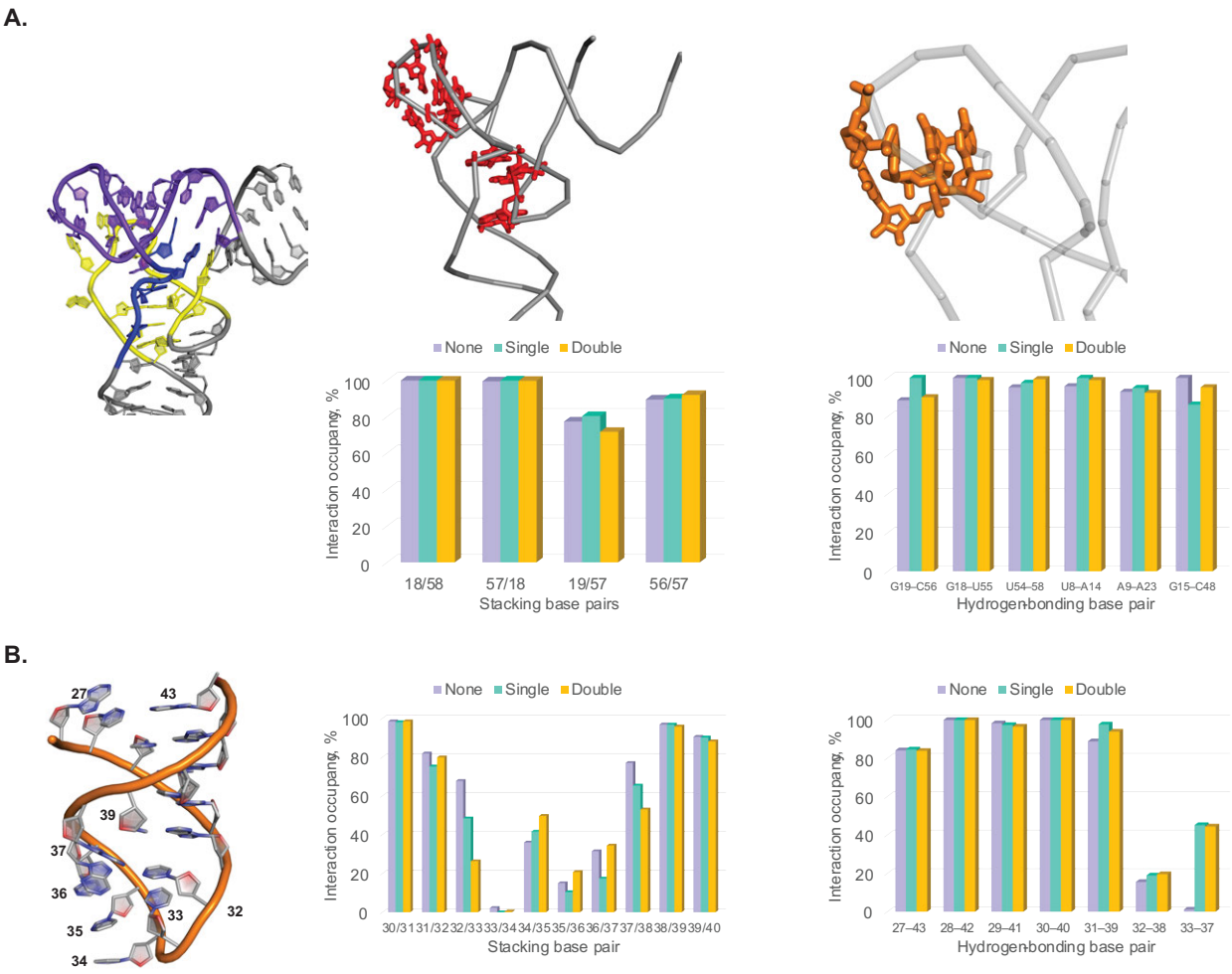


Figure 4A.17 – Non-covalent interactions within tRNA^{Lys}

(A) Persistence of tertiary hydrogen-bonding (red) and stacking (orange) interactions at the tRNA^{Lys} elbow. The tRNA elbow is color-coded by the domains involved – D arm (yellow), variable loop (blue), TyC arm (purple). Interactions are color-coded by system, namely unmodified tRNA^{Lys} (lavender), cmnm^{5}s^{2}U34-tRNA^{Lys} (teal) and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} (yellow). (B) Non-covalent interactions within the ASL domain of tRNA^{Lys}. Occupancies for the hydrogen-bonding (middle) and stacking (right) interactions within the ASL across ten 500-ns replica ensembles. Interactions are color-coded by system, namely unmodified tRNA^{Lys} (lavender), cmnm^{5}s^{2}U34-tRNA^{Lys} (teal) and cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} (yellow).

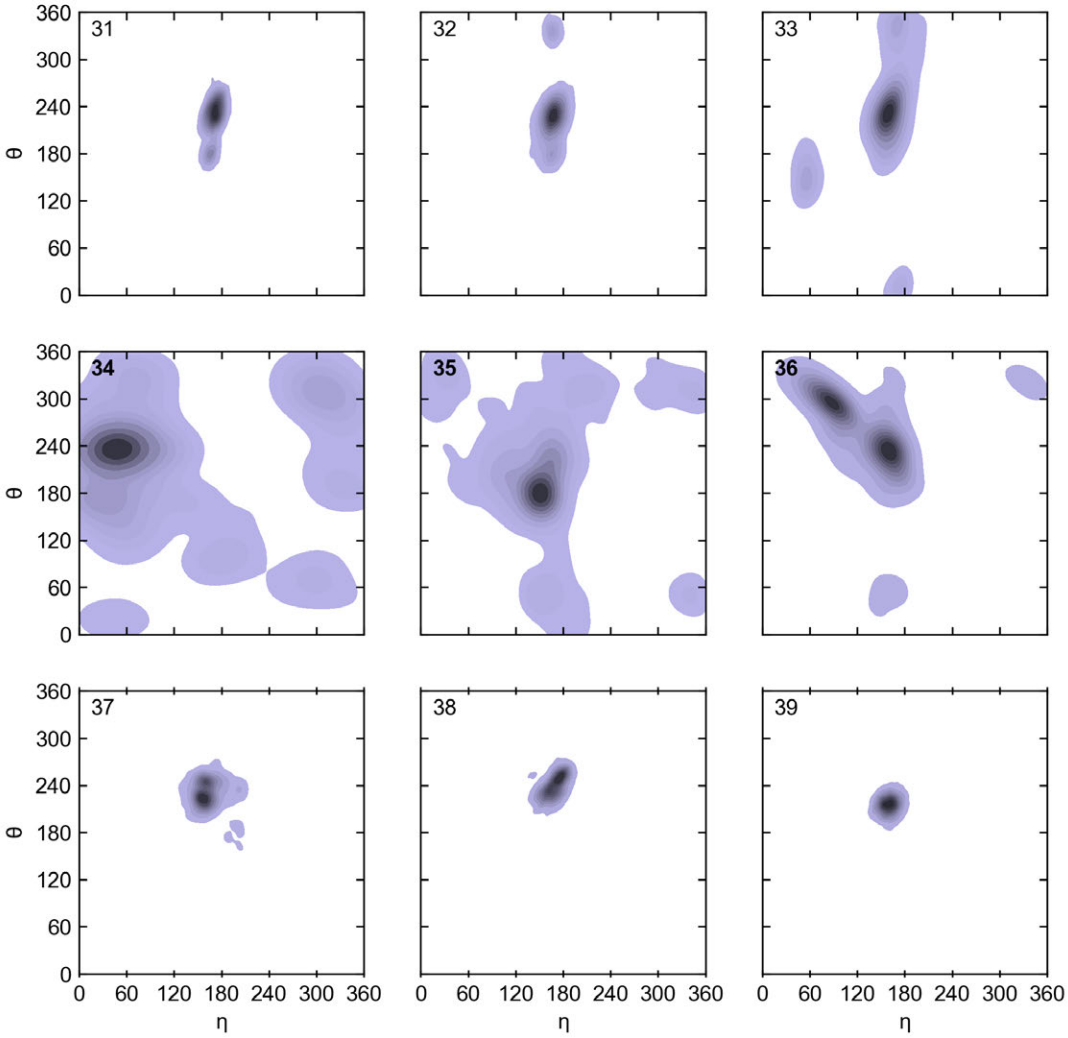


Figure 4A.18 – Pseudorotational analysis of the anticodon loop of unmodified tRNA^{Lys}

Backbone pseudotorsions (η = ∠C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = ∠C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in unmodified tRNA^{Lys}.

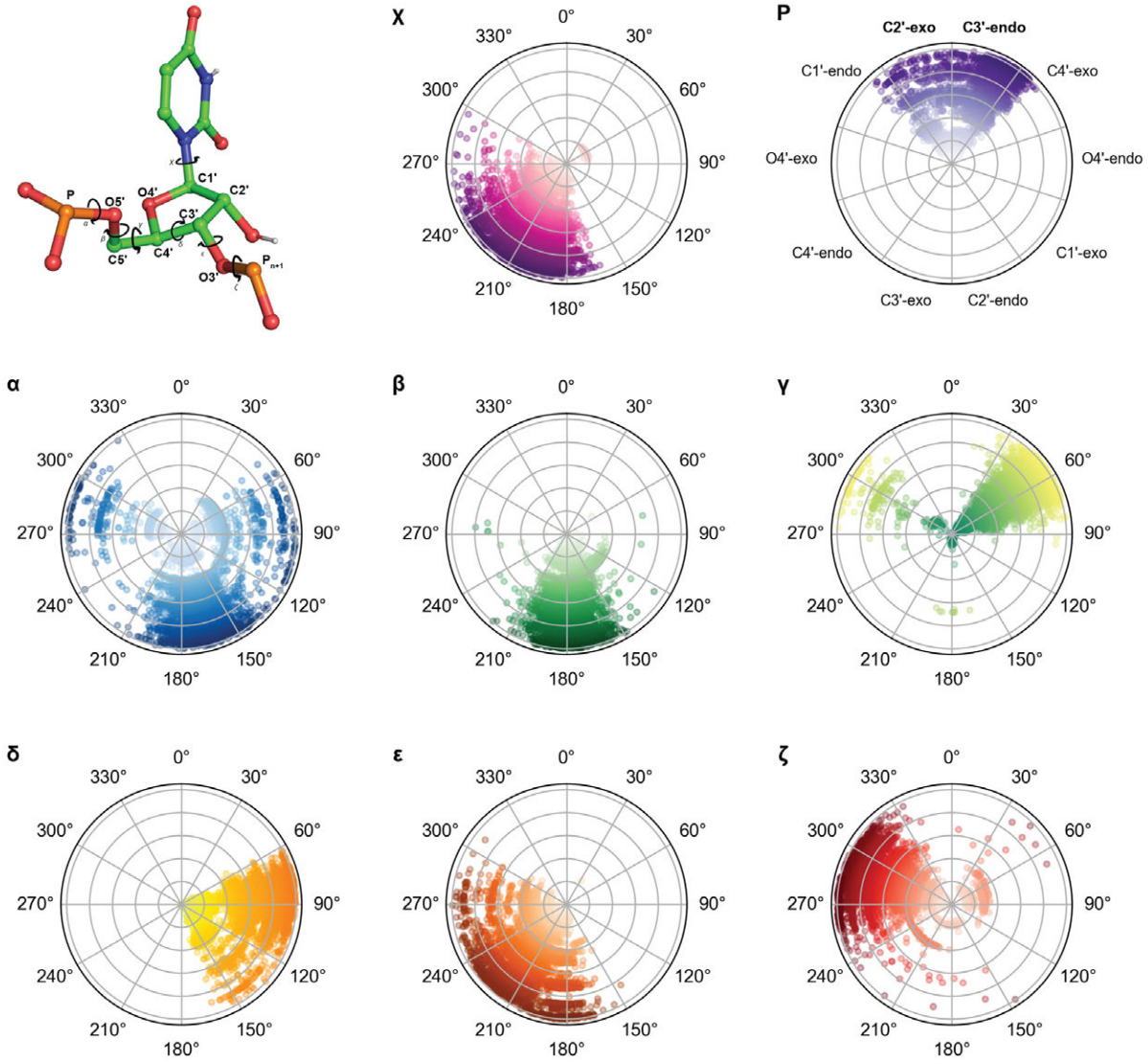


Figure 4A.19 – Backbone dynamics of C34 in unmodified tRNA^{Lys}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of C34 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red).

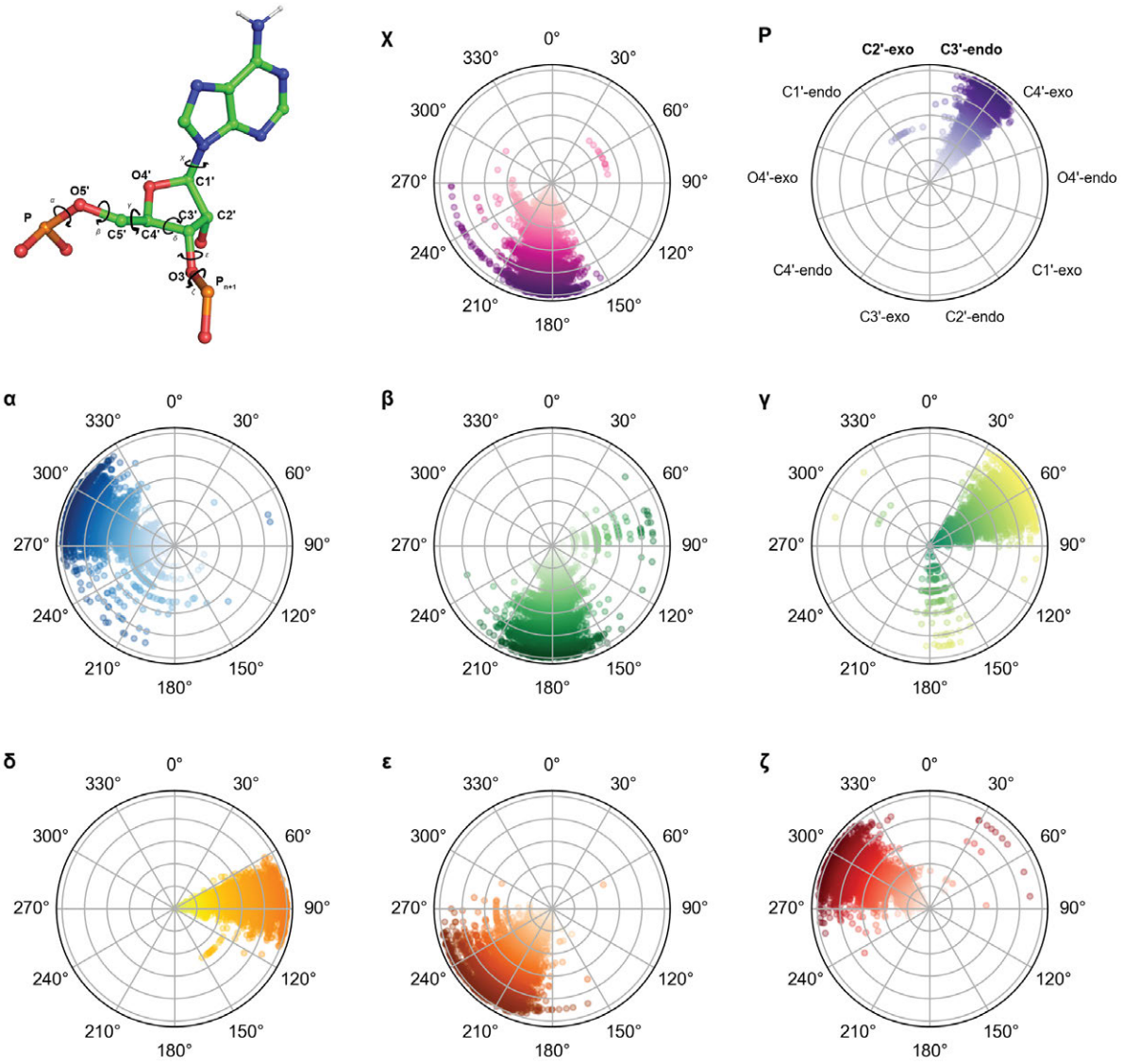


Figure 4A.20 – Backbone dynamics of A37 in unmodified tRNA^{Lys}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of A37 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red). Simulation time (5 μs) is represented on the r-axes of each polar plot.

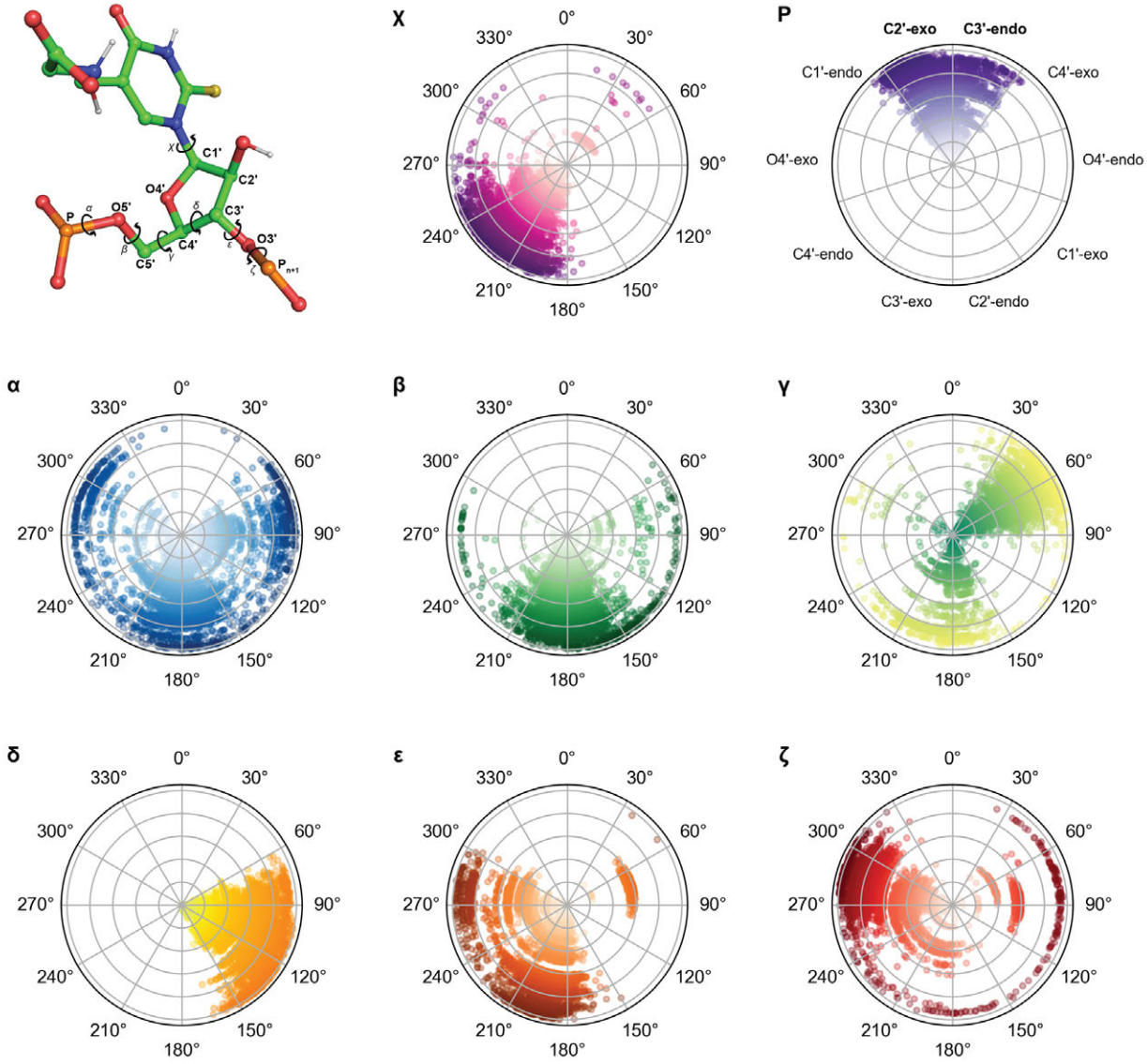


Figure 4A.21 – Backbone dynamics of cmnm^{5}s^{2}U34 in singly modified tRNA^{Lys}

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of cmnm^{5}s^{2}U34 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red).

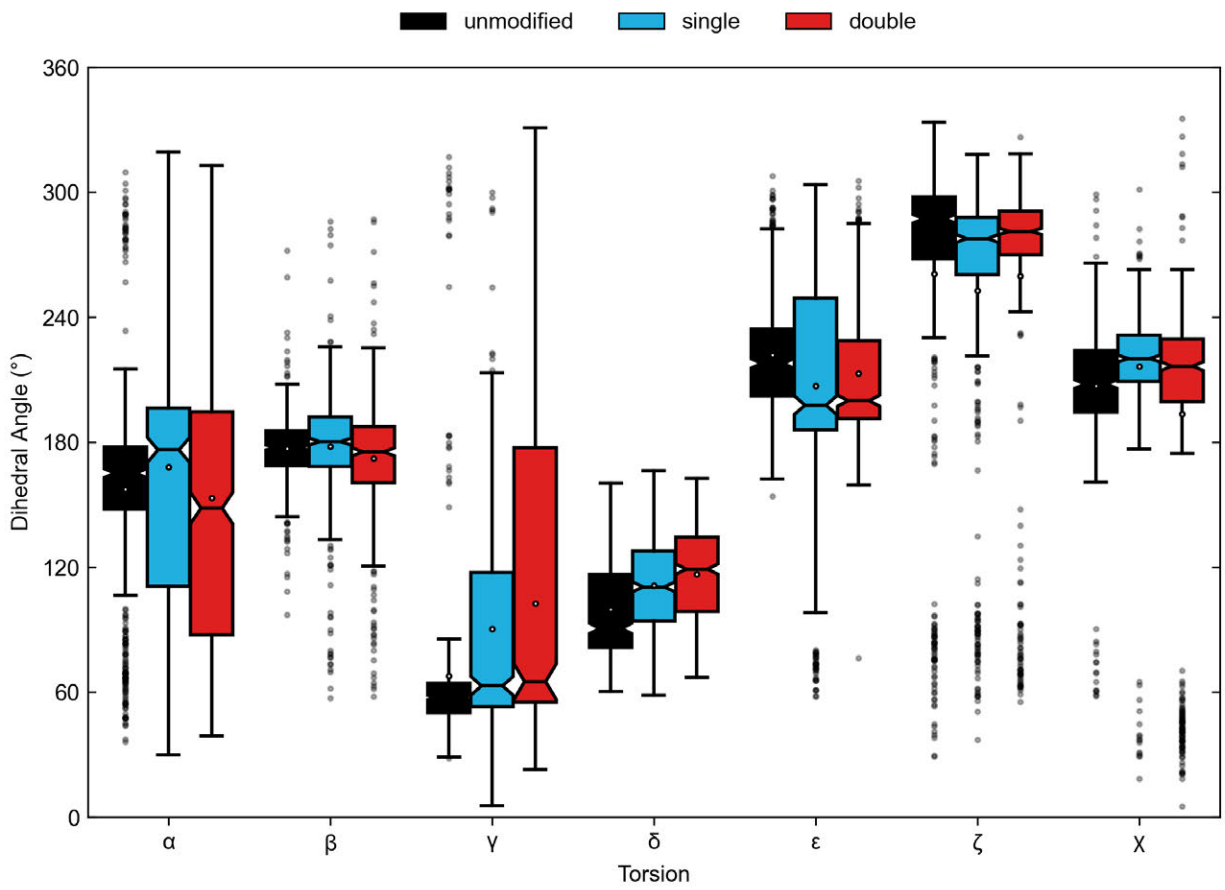


Figure 4A.22 – Analysis of backbone dynamics at position 34 of tRNA^{Lys} in the presence of cmnm^{5}s^{2}U34 and t^{6}A37 modifications

Statistical analysis of backbone torsions adopted at position 34 in unmodified, singly (cmnm^{5}s^{2}U34-tRNA^{Lys}) and doubly (cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys}) modified tRNA^{Lys}. The white dots represent the mean, rectangles represent interquartile ranges.

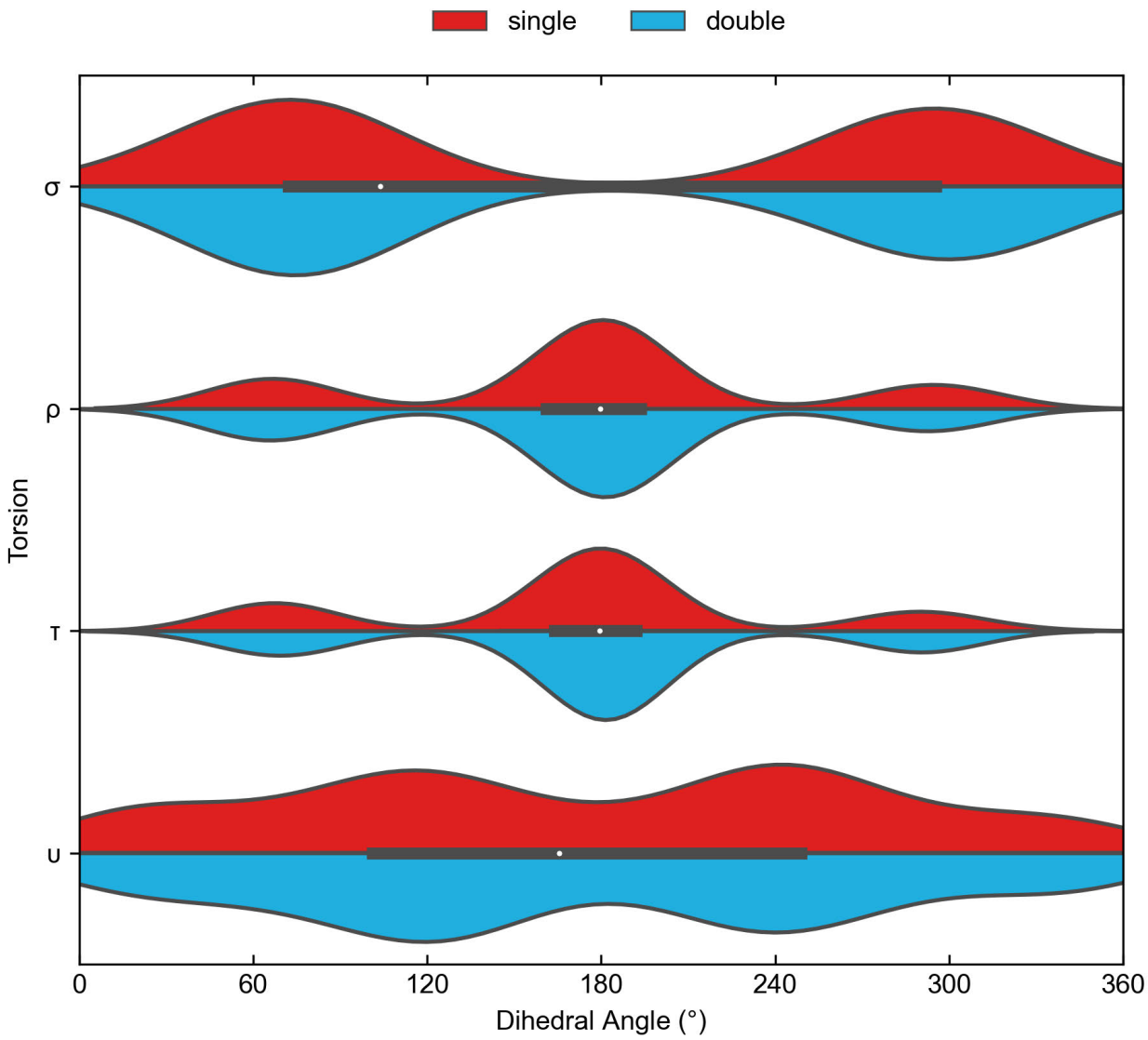


Figure 4A.23 – Dynamics of the cmnm^{5}s^{2}U34 sidechain within the tRNA^{Lys} ASL

Violin plots comparing the dihedral angle distributions for the C5 moiety of cmnm^{5}s^{2}U34 in single and double modification models of tRNA^{Trp}. The white dots represent the mean, rectangles represent interquartile ranges. Torsions are defined in Figure 4.7.

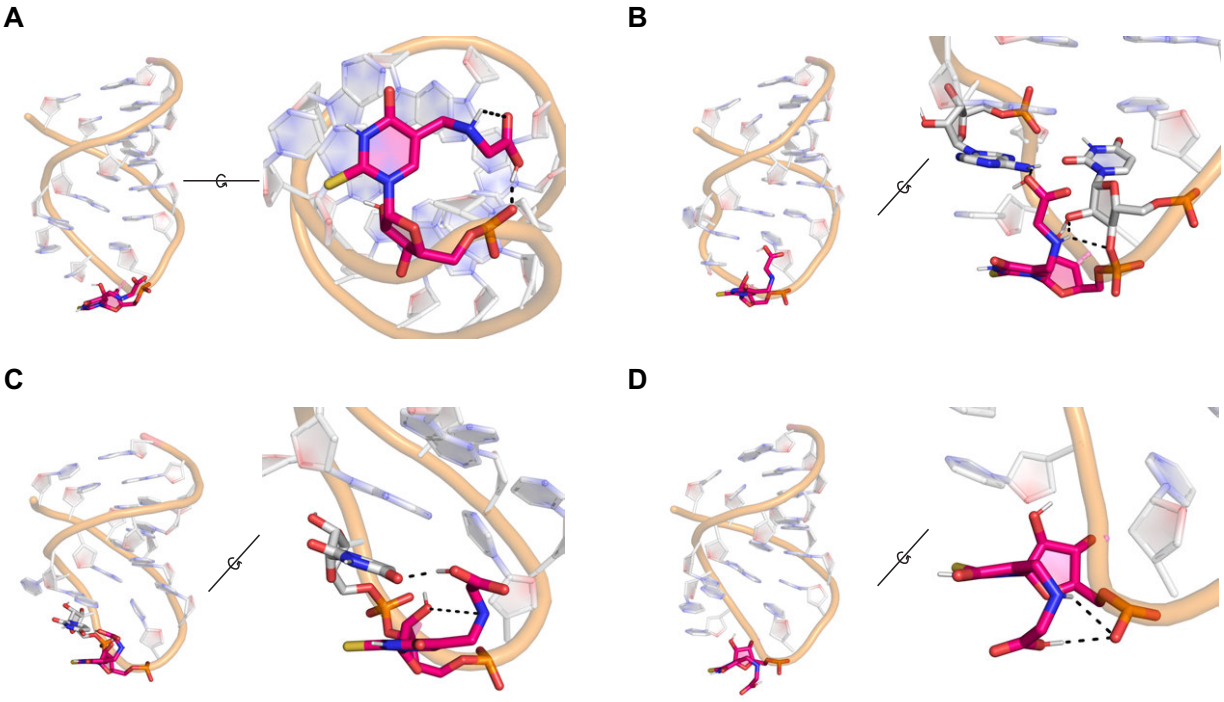


Figure 4A.24 – Persistent (< 50%) hydrogen-bonding interactions formed by cmnm^{5}s^{2}U34 in singly modified tRNA^{Lys}

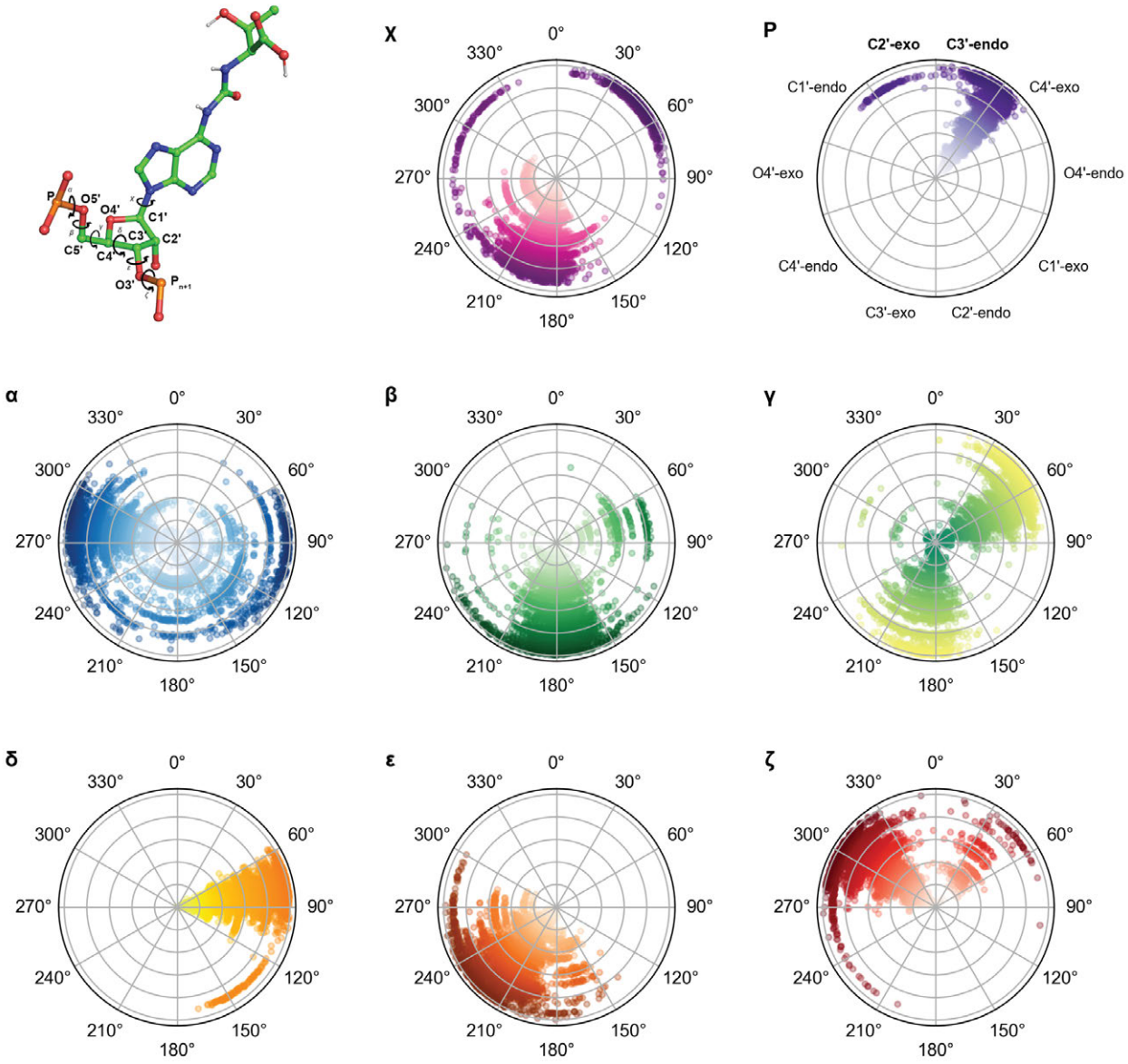


Figure 4A.25 – Backbone dynamics of t^6A37 in doubly modified $tRNA^{Lys}$

Density maps of the glycosidic torsion angle (χ), pseudorotation phase angle (P) and every backbone angle (α–ζ) of t^6A37 color-coded by torsion angle (χ: magenta, P: purple, α: blue, β: green, γ: yellow-green, δ: yellow, ε: orange, ζ: red). Simulation time (5 μs) is represented on the r-axes of each polar plot.

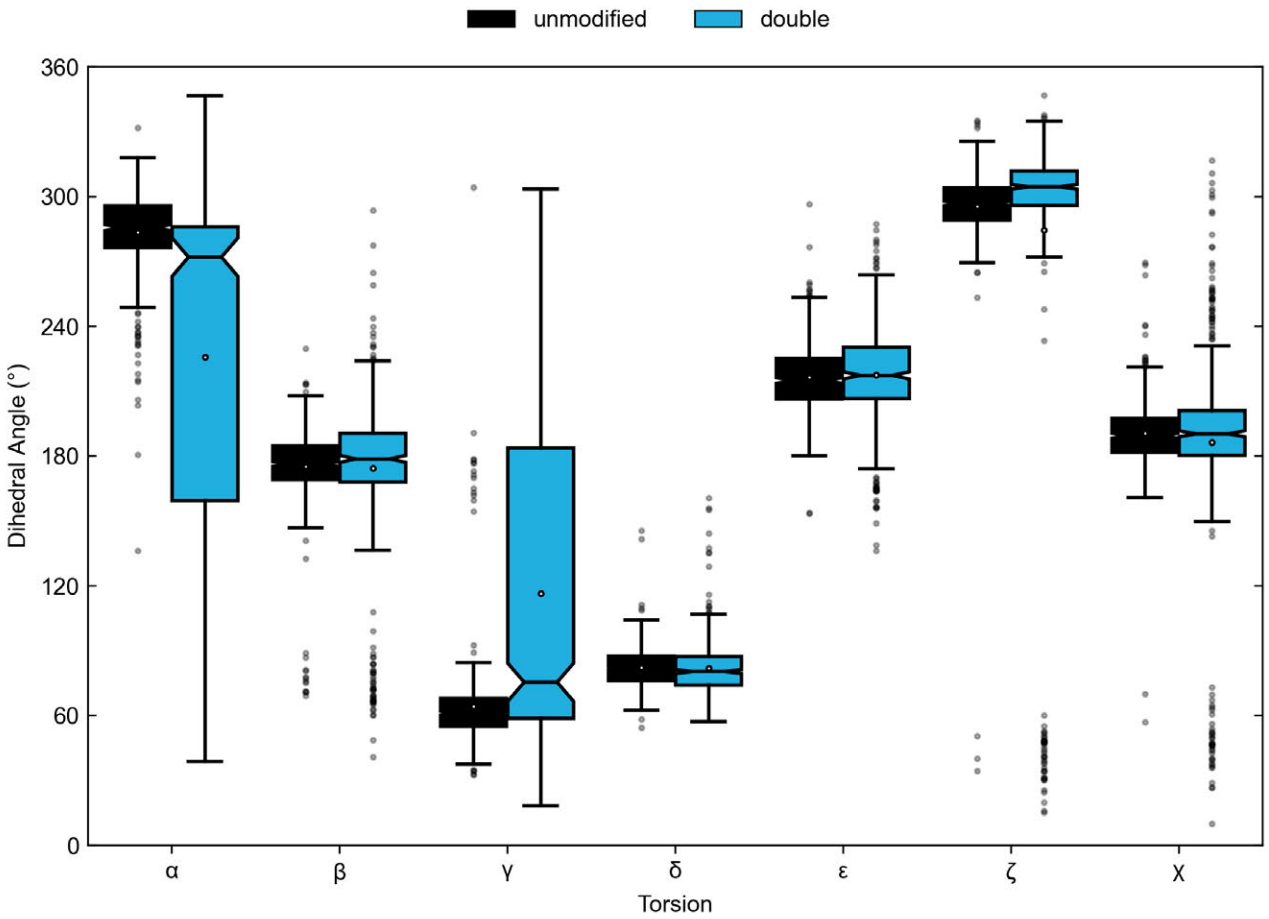


Figure 4A.26 – Analysis of backbone dynamics at position 37 of tRNA^{Lys}

Statistical analysis of backbone torsions adopted at position 37 in unmodified and doubly modified tRNA^{Lys}. The white dots represent the mean, rectangles represent interquartile ranges.

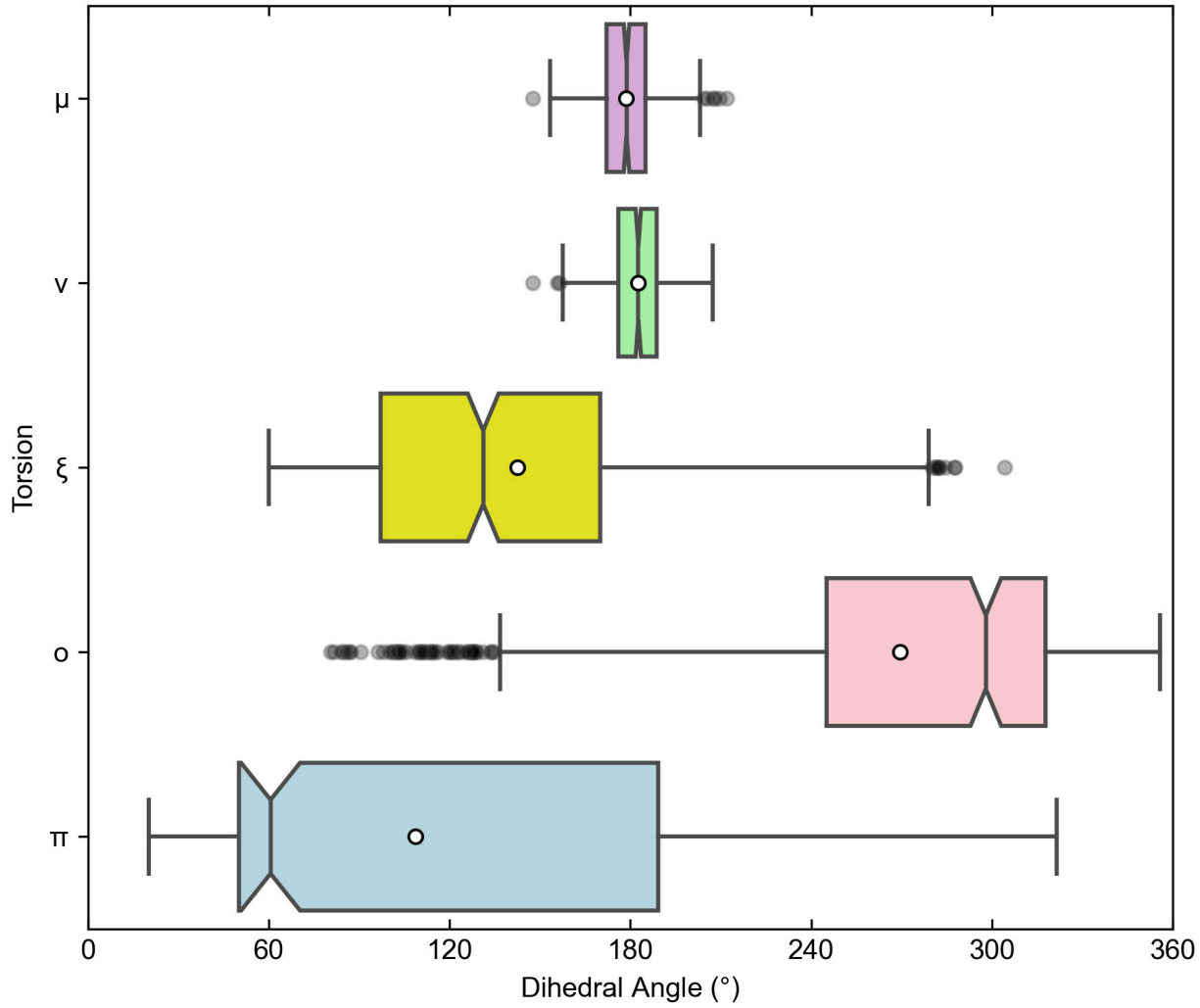


Figure 4A.27 – Dynamics of the t^6A37 sidechain within the $tRNA^{Lys}$ ASL

Statistical analysis of side chain torsions adopted in t^6A37 $cmnm^5s^2U34/t^6A37$-$tRNA^{Lys}$. The white dots represent the mean, rectangles represent interquartile ranges. Torsions are defined in Figure 4.7.

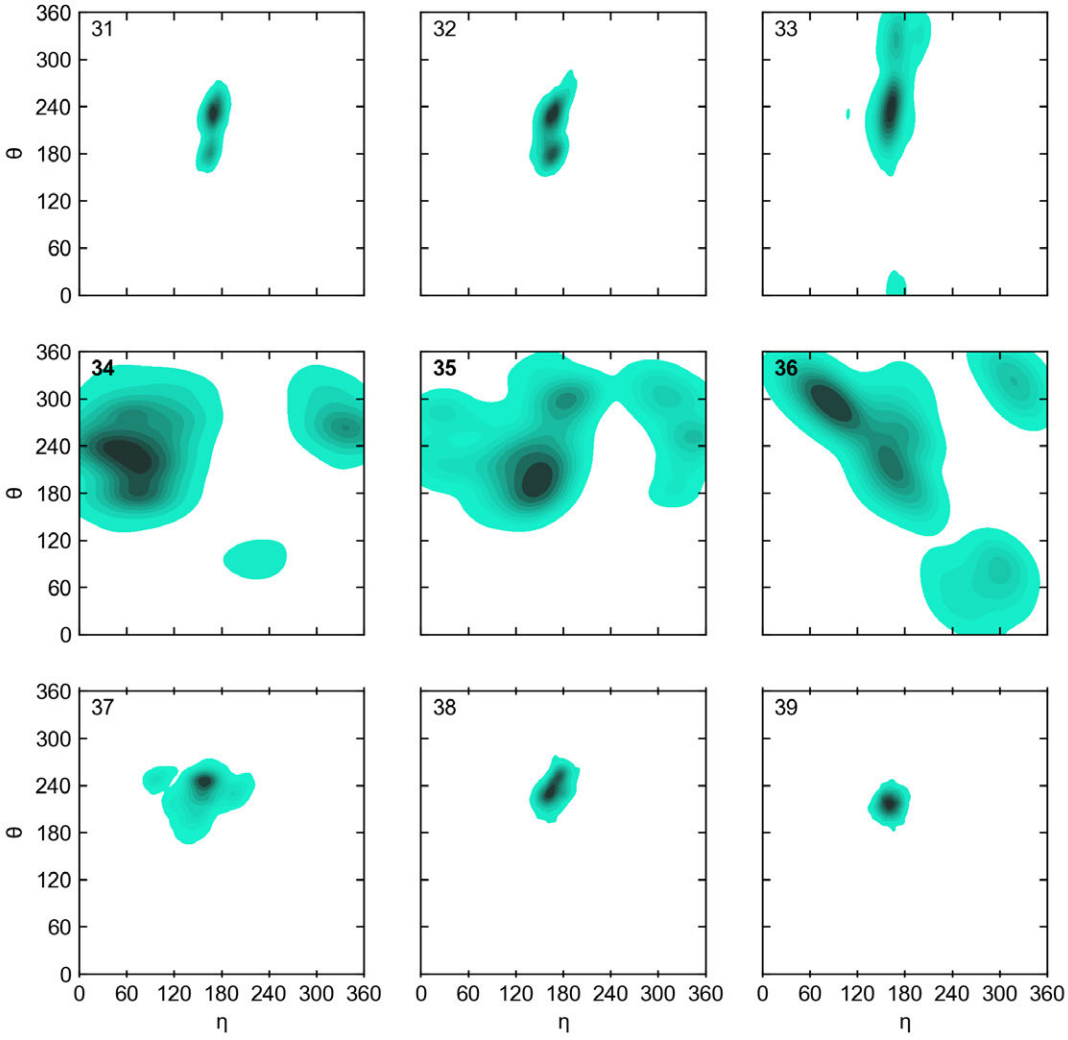


Figure 4A.28 – Pseudorotational analysis of the anticodon loop of cmnm^{5}s^{2}U34-modified tRNA^{Lys}

Backbone pseudotorsions (η = ∠C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = ∠C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in singly modified tRNA^{Lys}.

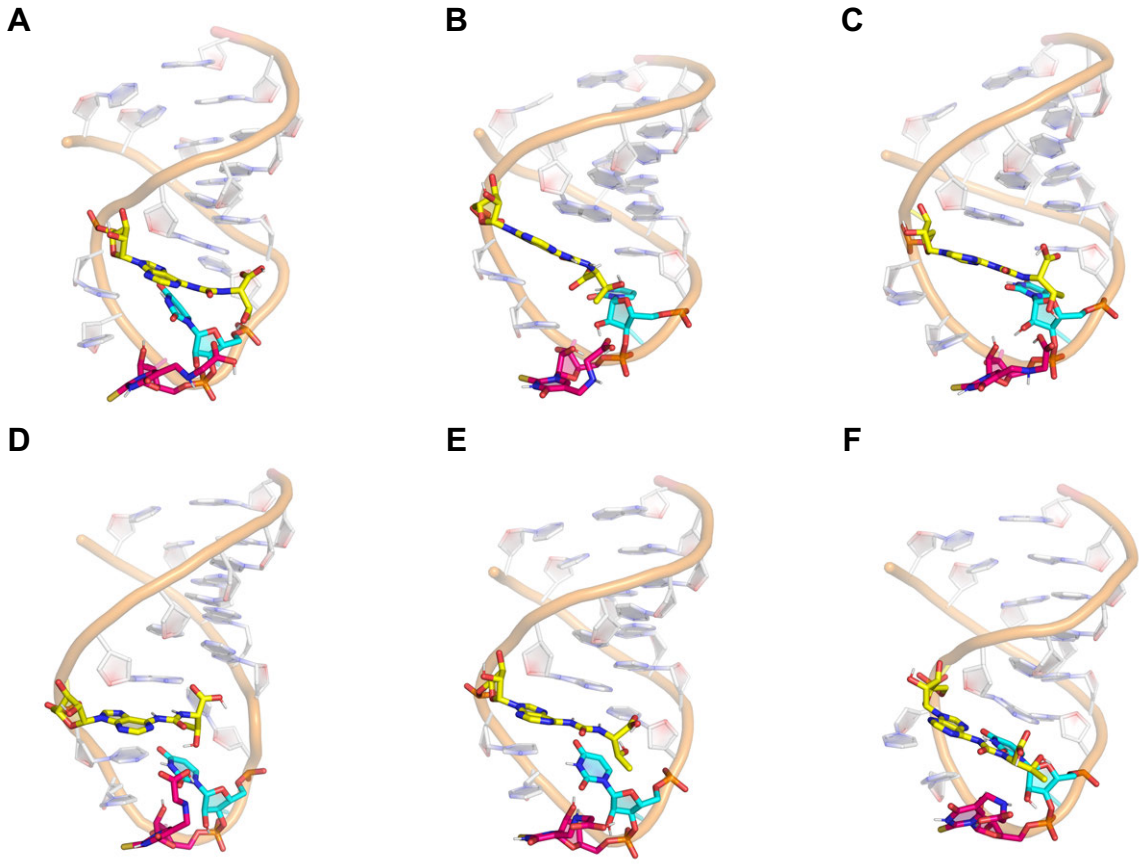


Figure 4A.29 – Widening of the ASL loop by cmnm^{5}s^{2}U34 and t^{6}A37 in doubly modified tRNA^{Lys}.

Persistent hydrogen-bonding interactions between U33 (cyan), cmnm^{5}s^{2}U34 (magenta) and t^{6}A37 (yellow) that stabilize an open loop conformation.

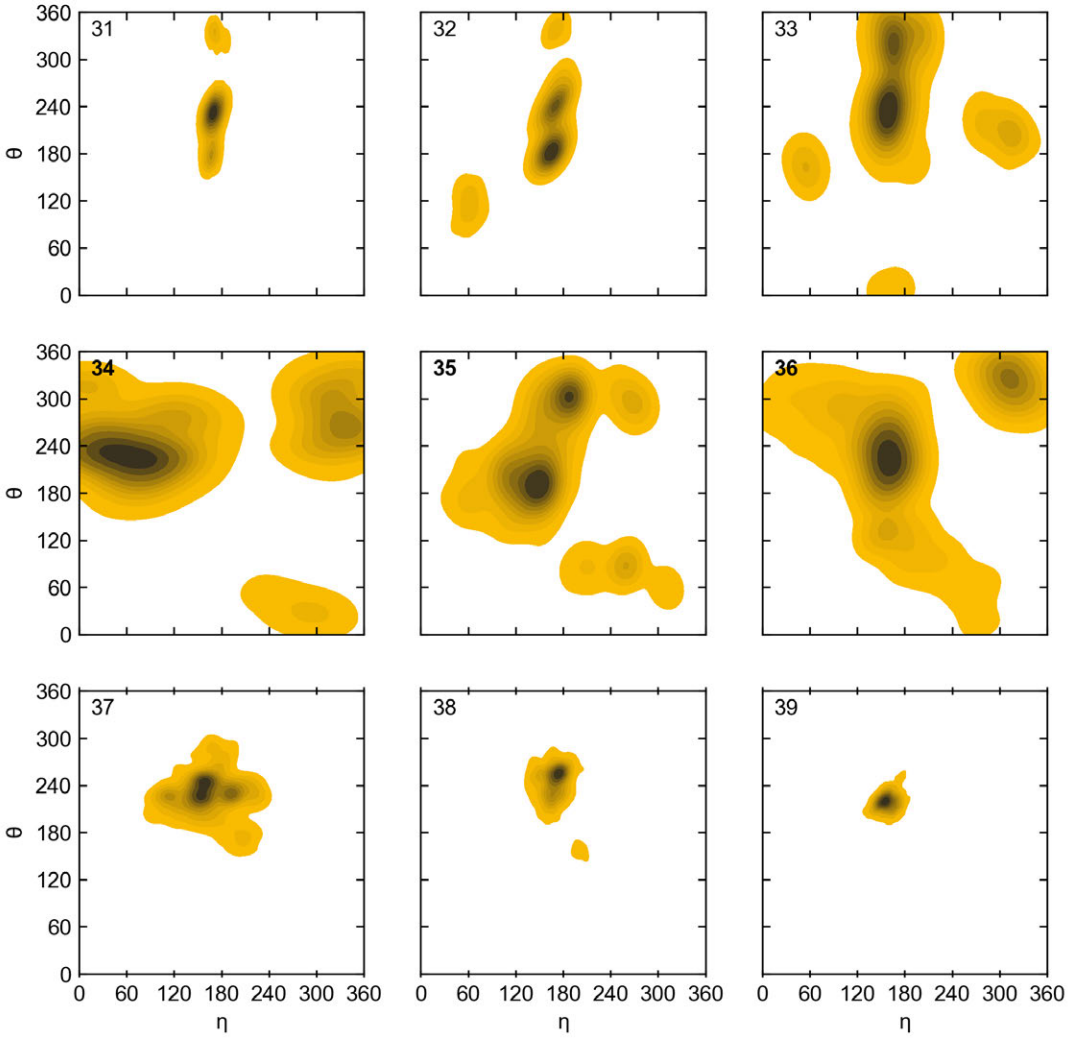


Figure 4A.30 – Pseudorotational analysis of the anticodon loop of cmnm^{5}s^{2}U34/t^{6}A37- tRNA^{Lys}

Backbone pseudotorsions (η = $\angle$C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = $\angle$C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in doubly modified tRNA^{Lys}.

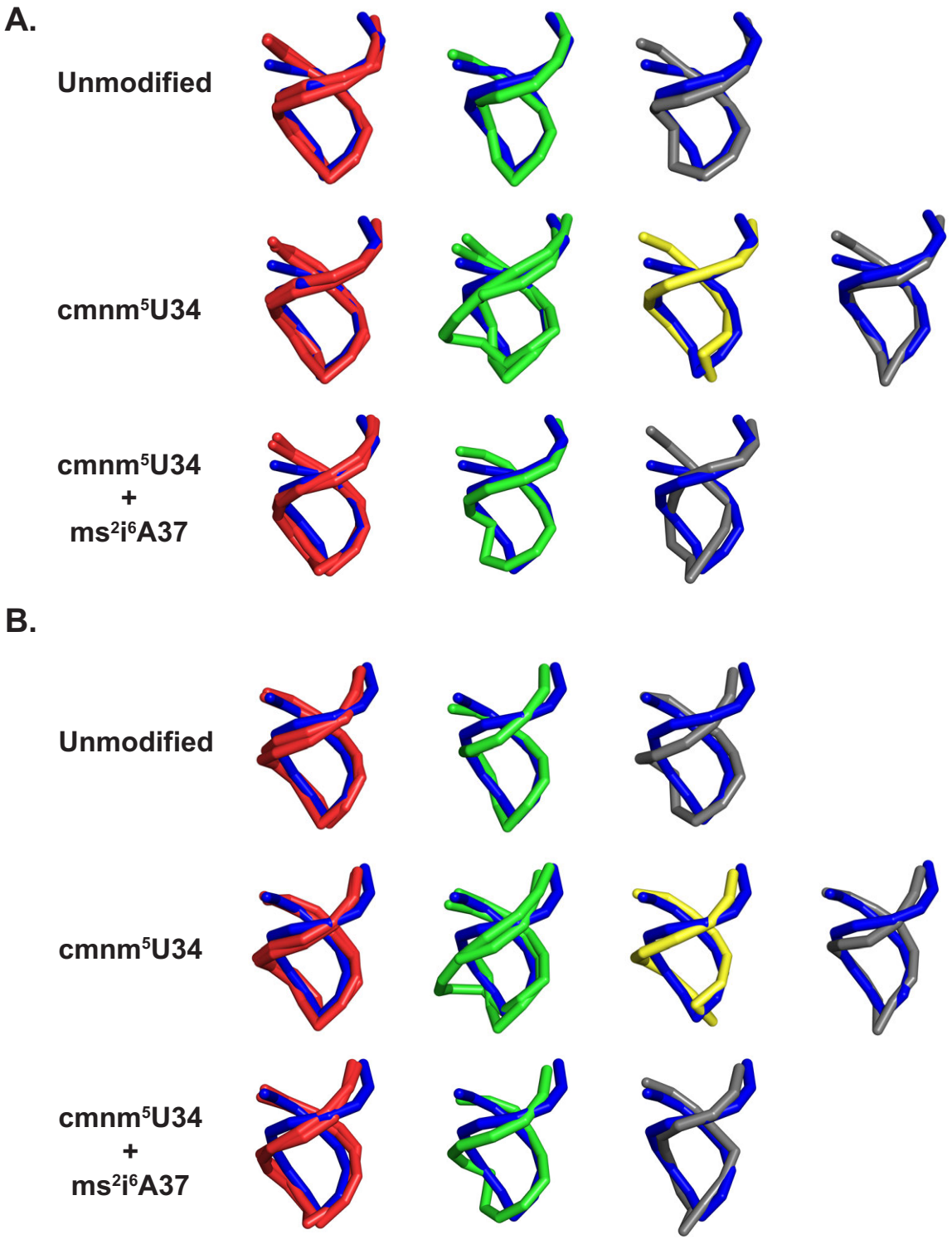


Figure 4A.31– Comparison of simulated tRNA^{Trp} ASL states to experimentally-derived functional states

Backbone comparisons of unmodified tRNA^{Trp}, cmnm^{5}U34-tRNATrp cmnm^{5}U34/ms^{2}i^{6}A37-tRNA^{Trp} and against. EF-Tu-bound (left, PDB ID: 1TTT) and ribosome-bound (right; PDB ID: 6WD0) tRNA. The experimental reference structures are colored blue, while the WB, FB and DL conformational groups are represented in red, green, and grey, respectively.

Table 4A.1 – ASL states from $tRNA^{Trp}$ vs. EF-Tu-bound $tRNA^{Phe}$

Model	Conformation	RMSD (Å)
Unmodified	34-stacked	1.574
Unmodified	34-unstacked	2.196
Unmodified	36-unstacked	2.350
Unmodified	disorder	2.622
cmnm^{5}U-modified	34-stacked	1.633
cmnm^{5}U-modified	34-unstacked	2.325
cmnm^{5}U-modified	35-unstacked	3.226
cmnm^{5}U-modified	36-unstacked	2.544
cmnm^{5}U-modified	33-out	2.716
cmnm^{5}U-modified	disorder	2.826
cmnm^{5}U+ms^{2}i^{6}A-modified	34-stacked	2.403
cmnm^{5}U+ms^{2}i^{6}A-modified	34-unstacked	2.036
cmnm^{5}U+ms^{2}i^{6}A-modified	35-unstacked	2.613
cmnm^{5}U+ms^{2}i^{6}A-modified	disorder	3.426

Table 4A.2 – ASL states from $tRNA^{Trp}$ vs. ribosome-bound $tRNA^{fMet}$

Model	Conformation	RMSD (Å)
Unmodified	34-stacked	2.937
Unmodified	34-unstacked	2.887
Unmodified	36-unstacked	3.563
Unmodified	disorder	3.203
cmnm^{5}U-modified	34-stacked	3.004
cmnm^{5}U-modified	34-unstacked	3.190
cmnm^{5}U-modified	35-unstacked	4.583
cmnm^{5}U-modified	36-unstacked	3.769
cmnm^{5}U-modified	33-out	3.731
cmnm^{5}U-modified	disorder	3.258
cmnm^{5}U+ms^{2}i^{6}A-modified	34-stacked	3.111
cmnm^{5}U+ms^{2}i^{6}A-modified	34-unstacked	3.339
cmnm^{5}U+ms^{2}i^{6}A-modified	35-unstacked	3.505
cmnm^{5}U+ms^{2}i^{6}A-modified	disorder	3.090

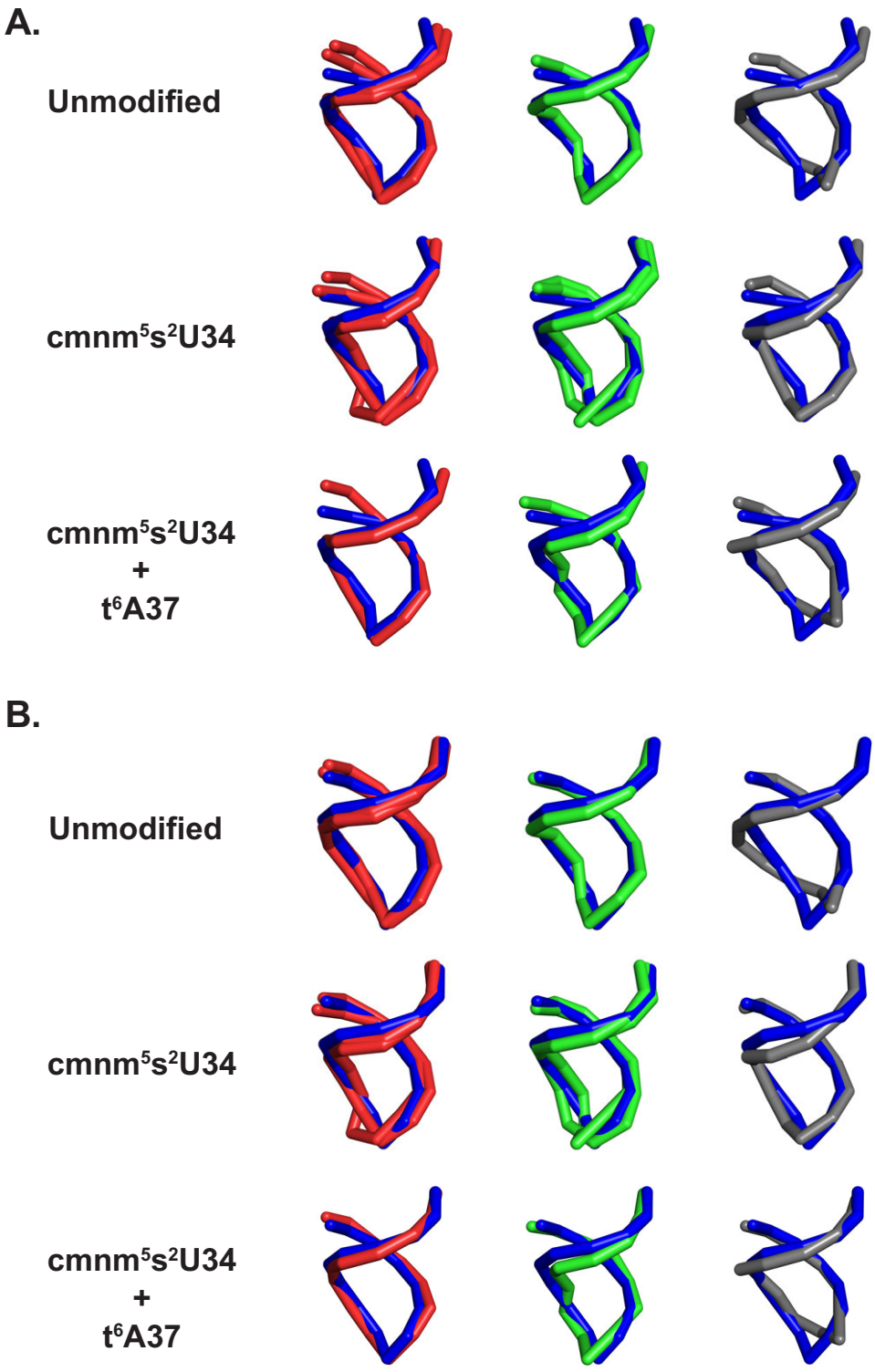


Figure 4A.32– Comparison of simulated tRNA^{Lys} ASL states to experimentally-derived functional states

Backbone comparisons of unmodified tRNA^{Lys}, cmnm^{5}s^{2}U34-tRNA^{Lys} cmnm^{5}s^{2}U34/t^{6}A37-tRNA^{Lys} and against. EF-Tu-bound (left, PDB ID: 1TTT) and ribosome-bound (right; PDB ID: 6WD0) tRNA. The experimental reference structures are colored blue, while the WB, FB and DL conformational groups are represented in red, green, and grey, respectively.

Table 4A.3 – ASL states from tRNA^{Lys} vs. EF-Tu-bound tRNA^{Phe}

Model	Conformation	RMSD (Å)
Unmodified	34-stacked	1.819
Unmodified	34-unstacked	2.203
Unmodified	36-unstacked	2.499
Unmodified	disorder	2.464
cmnm^{5}s^{2}U-modified	34-stacked	1.845
cmnm^{5}s^{2}U-modified	34-unstacked	2.537
cmnm^{5}s^{2}U-modified	35-unstacked	2.521
cmnm^{5}s^{2}U-modified	36-unstacked	2.494
cmnm^{5}s^{2}U-modified	disorder	2.635
cmnm^{5}s^{2}U+t^{6}A-modified	34-stacked	1.848
cmnm^{5}s^{2}U+t^{6}A-modified	36-unstacked	3.110
cmnm^{5}s^{2}U+t^{6}A-modified	disorder	3.051

Table 4A.4 – ASL states from tRNA^{Lys} vs. ribosome-bound tRNA^{fMet}

Model	Conformation	RMSD (Å)
Unmodified	34-stacked	1.302
Unmodified	34-unstacked	1.523
Unmodified	36-unstacked	2.185
Unmodified	disorder	1.949
cmnm^{5}s^{2}U-modified	34-stacked	1.668
cmnm^{5}s^{2}U-modified	34-unstacked	2.101
cmnm^{5}s^{2}U-modified	35-unstacked	2.014
cmnm^{5}s^{2}U-modified	36-unstacked	2.163
cmnm^{5}s^{2}U-modified	disorder	2.435
cmnm^{5}s^{2}U+t^{6}A-modified	34-stacked	1.223
cmnm^{5}s^{2}U+t^{6}A-modified	36-unstacked	2.792
cmnm^{5}s^{2}U+t^{6}A-modified	disorder	2.754

APPENIDIX III: VALIDATION OF tRNA MD PROTOCOL FOR INVESTIGATIONS ON POSTTRANSCRIPTIONAL MODIFICATIONS AT THE ASL

In Chapter 3, an unmodified *E. coli* $tRNA^{Phe}$ (PDB ID: 3L0U) model was used to develop a computational protocol for representative and efficient sampling of the tRNA conformational space. In that study, ten 500-ns replica simulations were found to representatively sample dominant conformations of the ASL and provide efficient use of computational resources. To ensure this protocol can be used to reveal the structural changes that arise in the presence of posttranscriptional modifications, the configurational effects of $cmnm^5(s^2)U34$ modifications were independently investigated in the same tRNA structure (*E. coli* $tRNA^{Phe}$, PDB ID: 3L0U) using 10-replica ensembles. For comparison, a ten 500-ns replica simulations were also performed on the unmodified $tRNA^{Phe}$. All model preparation, MD simulation and analyses protocols were carried out as described in section 4.2.2 of this thesis.

In the absence of modifications, $tRNA^{Phe}$ maintained the tertiary structure isolated in its experimental crystal structure, and the average RMSD across the replica ensemble relative to the reference structure was 3.9 ± 0.4 Å (Figure 4B.1). Per nucleotide RMSF analyses of the unmodified $tRNA^{Phe}$ model revealed the loop regions within D, ASL and variable domains to be more dynamic than helical counterparts (RMSF ranged between 3 and 9 Å in the loop regions, while RMSF is ~ 2.5 Å for their respective stem regions; Figure 4B.2).

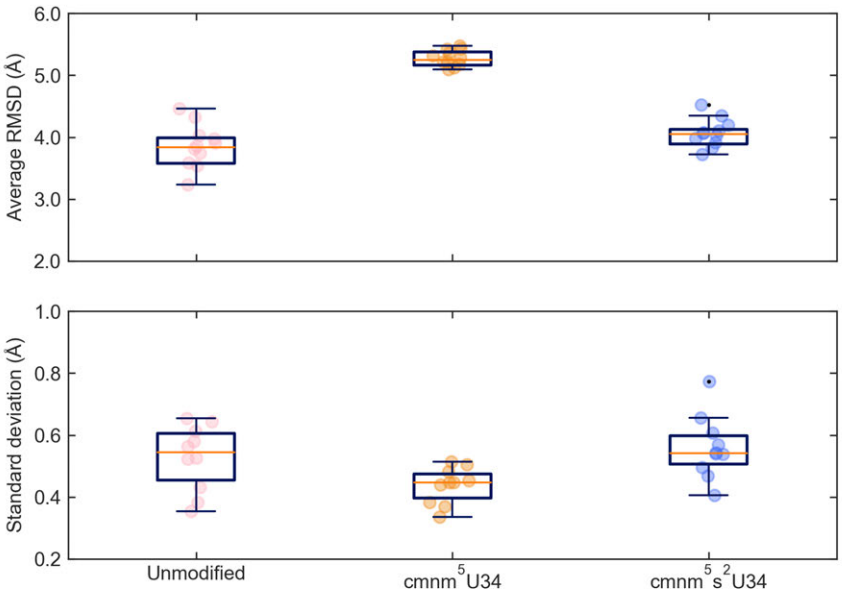


Figure 4B.1 – Stability of MD simulations on $tRNA^{Phe}$

Statistical analysis of RMSDs for unmodified $tRNA^{Phe}$, $cmnm^5U34$-$tRNA^{Phe}$ and $cmnm^5s^2U34$-$tRNA^{Phe}$.

Large ranges of motion in the loop regions of D and variable loops did not interfere with the network of kissing interactions at the tRNA elbow, as stacking and hydrogen-bonding interactions between the TψC, D and variable domains remained prevalent throughout the replica ensemble, with average occupancies > 60% (Figure 4B.3). At the ASL, non-covalent interactions were well-maintained within the domain's stem (occupancies > 85%), but less prevalent within the loop region. Specifically, the U32–A38 and U33–A37 base pairs were only present for ~ 50% and 20% respectively and stacking interactions between consecutive nucleobases range from 60% to 80%, except for the U33/G34 stack that had an occupancy of 2% (Figure 4B.3). The absence of stacking between U33 and G34 at this position indicates that the canonical tRNA U-turn is maintained within this system.

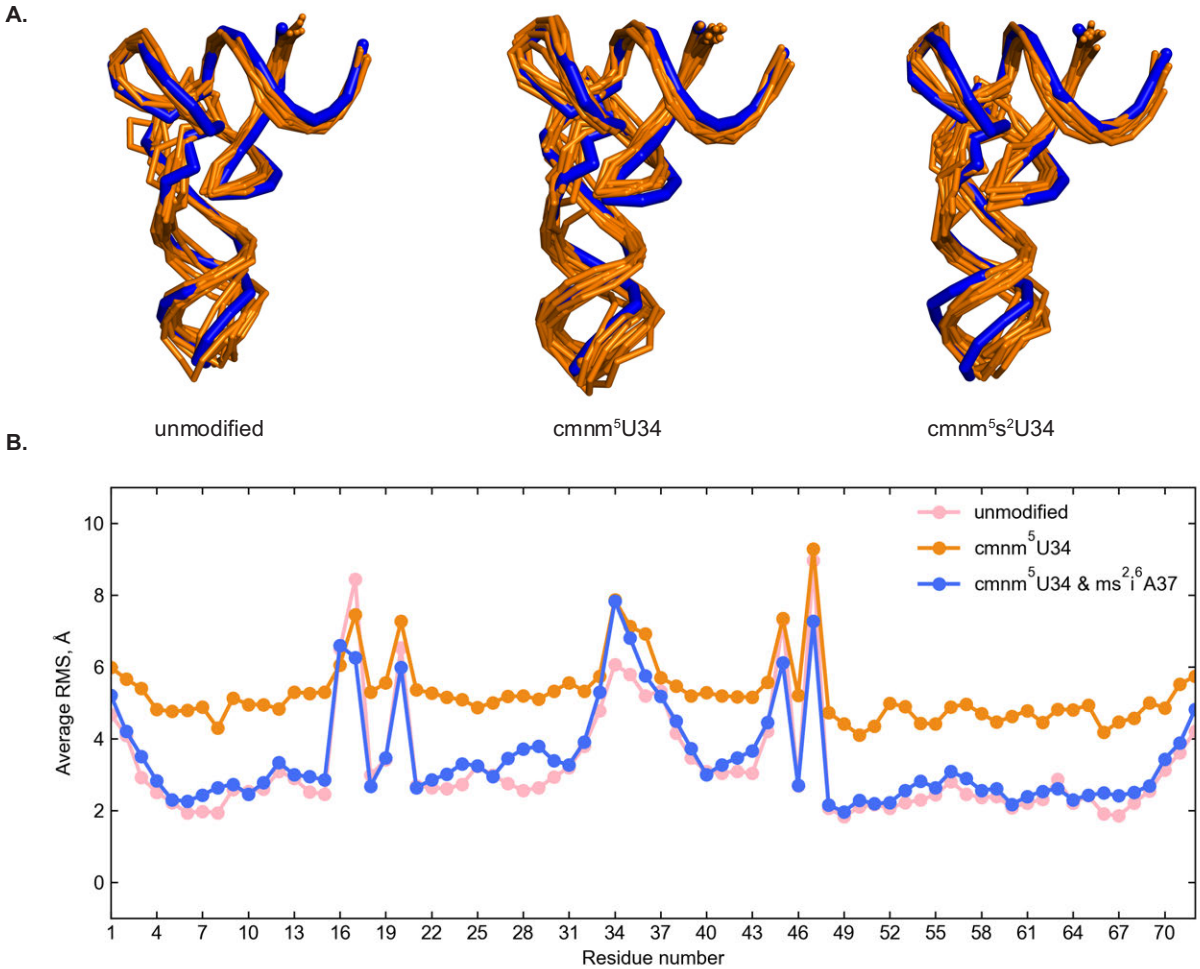


Figure 4B.2 – Full tRNA dynamics across the tRNA^{Phe} systems

(A) Representative structure overlays (orange ribbons) from replica trajectories of unmodified tRNA^{Phe}, cmnm^{5}U34-tRNA^{Phe} and cmnm^{5}s^{2}U34-tRNA$^{Phe.}$ The reference structure is the experimental crystal structure (PDB ID: 3L0U, blue). (B) Average per residue fluctuations in unmodified tRNA^{Phe} (pink), cmnm^{5}U34-tRNA^{Phe} (orange) and cmnm^{5}s^{2}U34-tRNA^{Phe} (blue).

The presence of cmnm^{5}U and cmnm^{5}s^{2}U at position 34 of unmodified tRNA^{Phe} did not alter the global fold of the molecule and all tertiary hydrogen-bonding and stacking interactions were similarly preserved in the modified systems as they were in the unmodified reference. Although trajectories within the replica ensemble of cmnm^{5}U34-modified tRNA^{Phe} (cmnm^{5}U34-tRNA^{Phe}) had high RMSDs (average ensemble RMSD = 5.3 ± 0.1 Å; Figure 4B.1) relative to the unmodified and cmnm^{5}s^{2}U34-modified (cmnm^{5}s^{2}U34-tRNA^{Phe}) models (average ensemble RMSD = 4.1 ± 0.2 Å for the cmnm^{5}s^{2}U34-tRNA^{Phe}), no structural deviations were observed in the global tRNA structure across all three models. Moreover, time averaged per residue RMSF analyses of the modified systems showed large fluctuations in the loop regions (RMSF range of 5 to 10 Å in cmnm^{5}U34-tRNA^{Phe} and 3 to 10 Å in cmnm^{5}s^{2}U34-tRNA^{Phe}) and reduced flexibilities at the helical regions (RMSF of ~ 5 Å in cmnm^{5}U-tRNA^{Phe} and ~ 3 Å in cmnm^{5}s^{2}U34-tRNA^{Phe}), which matches the observations from unmodified tRNA^{Phe} (Figure 4B.2). Therefore, it can be concluded that cmnm^5(s^2)U34 modifications do not have long ranging effects on tRNA structure.

In contrast to the global structure, non-covalent interactions within the anticodon loops of singly modified models of tRNA^{Phe} differed from those observed in the unmodified system. In general, stacking interactions in this region were reduced in the presence of cmnm^{5}U34 (Figure 4B.3). Specifically, the 34/35, 35/36 and 36/37 stacks reduced from 63%, 80% and 74% in unmodified tRNA^{Phe} to 31%, 55% and 55% in the cmnm^{5}U34-tRNA^{Phe} system. Nevertheless, non-covalent interactions adjacent to anticodon bases were enhanced in the cmnm^{5}U34-tRNA^{Phe} system. Hydrogen-bonding interactions between U32 and A38 increased from ~ 50% in the unmodified system to ~ 90% in the modified model. Moreover, stacking interactions between U32 and U33 and A37 and A38 increased to 65% and 88%, respectively, compared to 59% and 66% in unmodified tRNA^{Phe}. Similar trends were observed in the cmnm^{5}s^{2}U34-tRNA^{Phe} system, as reduced stacking was observed between anticodon bases but non-covalent interactions between anticodon flanking bases (U32, U33, A37 and A38) were enhanced. Despite the varied dynamics within the modified ASLs, on average, the U-turn motif was maintained in these systems as evidenced by the lack of stacking between U33 and G34 over the course of all replica trajectories. All in all, these observations suggest that cmnm^5(s^2)U34 modifications increase the flexibility of other anticodon bases. Be that as it may, different occupancies were observed for all non-covalent

interactions in the modified tRNA^{Phe} systems, indicating that cmnm^{5}U34 and cmnm^{5}s^{2}U34 affect structural arrangement within the anticodon loop to different extents.

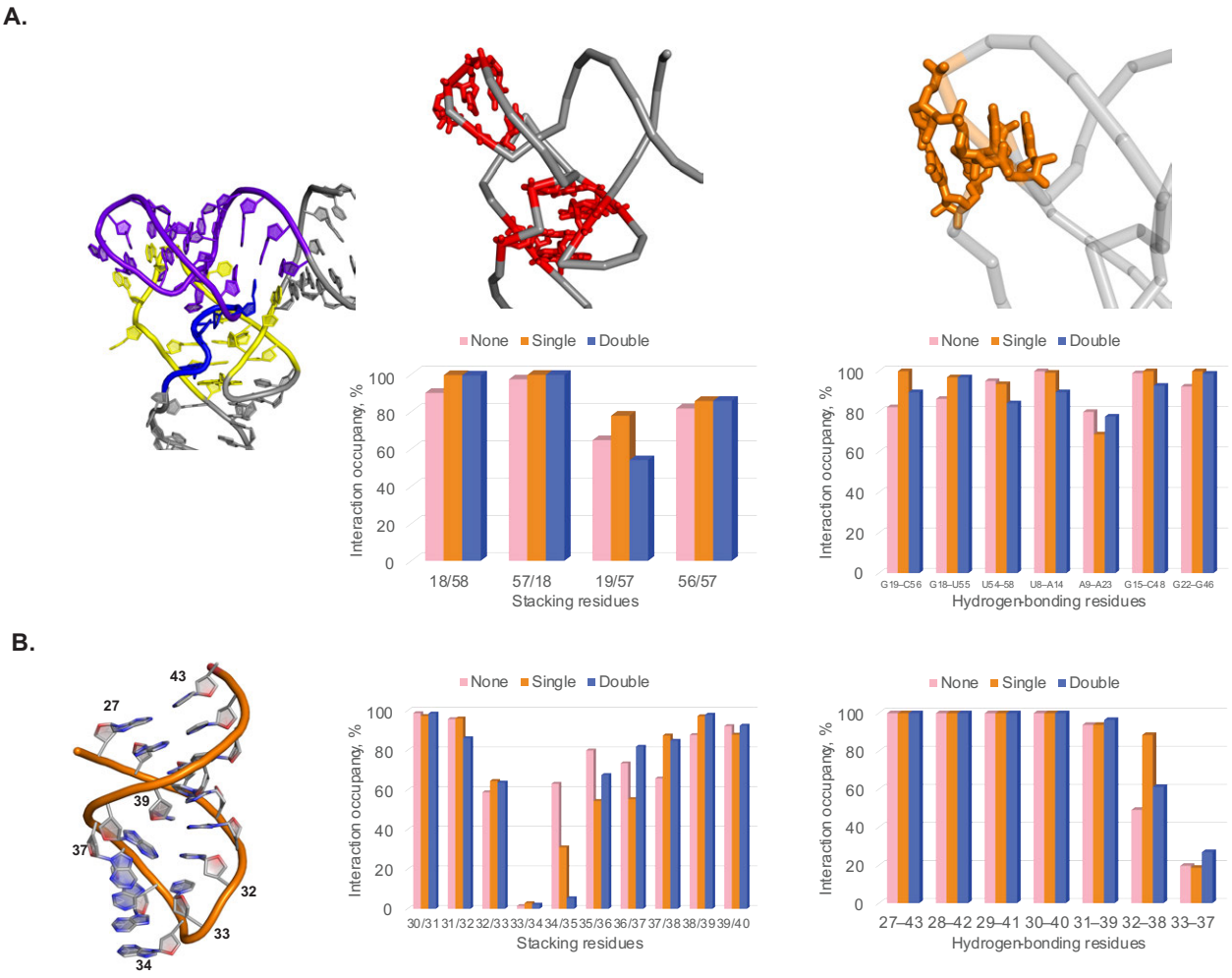


Figure 4B.3 – Non-covalent interactions within tRNA^{Phe}

(A) Persistence of tertiary hydrogen-bonding (red) and stacking (orange) interactions at the tRNA^{Phe} elbow. The tRNA elbow is color-coded by the domains involved – D arm (yellow), variable loop (blue), TyC arm (purple). Interactions are color-coded by system, namely unmodified tRNA^{Phe} (pink), cmnm^{5}U34-tRNA^{Phe} (orange) and cmnm^{5}s^{2}U34-tRNA^{Phe} (blue). (B) Non-covalent interactions within the ASL domain of tRNA^{Phe}. Occupancies for the hydrogen-bonding (middle) and stacking (right) interactions within the ASL of tRNA^{Phe} across ten 500-ns replica ensembles. Interactions are color-coded by system, namely unmodified tRNA^{Phe} (pink), cmnm^{5}U34-tRNA^{Phe} (orange) and cmnm^{5}s^{2}U34-tRNA^{Phe} (blue).

When inserted into tRNA^{Phe}, cmnm^5(s^2)U34 adopted similar side chain conformations around the C5–C7 linkage, defined by the torsional angle $\sigma = \angle$(C4C5C7N10). Conformer I ($180° \leq \sigma \leq 360°$) was occupied 49% of the time in cmnm^{5}U34-tRNA^{Phe} and 46% in cmnm^{5}s^{2}U34-tRNA^{Phe} (Figure 4B.4). On the other hand, conformer II ($0° \leq \sigma \leq 180°$) was sampled 51% and 54% of the time in cmnm^{5}U34-tRNA^{Phe}

and cmnm^{5}s^{2}U34-tRNA^{Phe}, respectively. Furthermore, both cmnm^{5}U and cmnm^{5}s^{2}U altered the backbone conformation at position 34. Specifically, distributions of the α, γ, δ and ε torsions adopted by the modified bases varied significantly from those observed in G34. Although this divergence may partly be due to the difference in base identity (G is a purine while cmnm^5(s^2)U are pyrimidines), differences were also observed in the backbone torsions of the modified systems at position 34. In particular, the δ torsion which had a distinct bimodal distribution with medians of 90° and 150° in cmnm^{5}U34-tRNA^{Phe}, adopted a broader range of dihedrals in cmnm^{5}s^{2}U34-tRNA^{Phe}, leading to trimodal distributions around the mean (Figure 4B.5). These differences observed in the backbone dihedrals of cmnm^5(s^2)U34 in tRNA^{Phe} suggest the modified nucleotides may behave differently, and therefore, have distinctive structural effects within the anticodon loop.

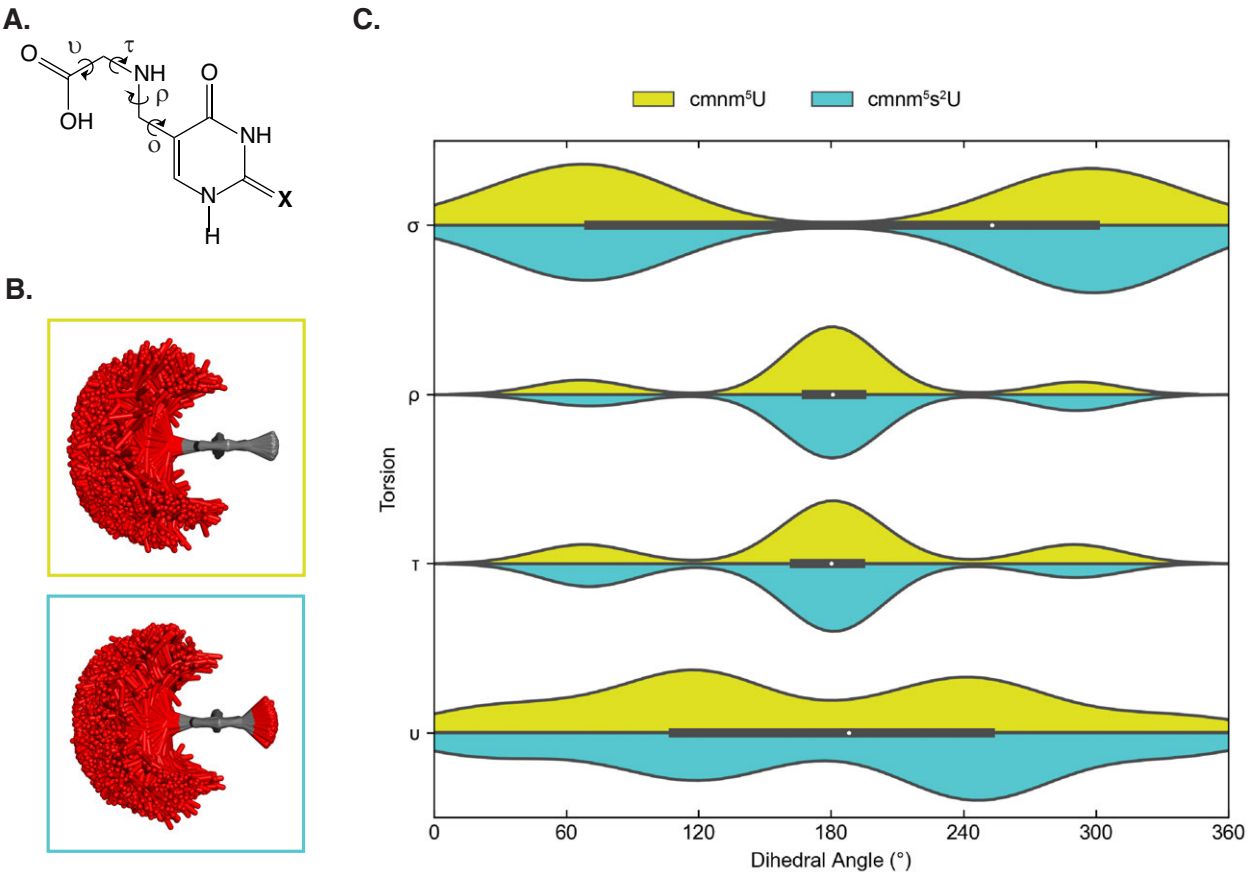


Figure 4B.4 – Sidechain flexibilities of cmnm^{5}U34 and cmnm^{5}s^{2}U34 in tRNA^{Phe}

(A) Torsional angles within the C5 moiety of cmnm^5(s^2)U34. (B) Structural representation of torsion angles adopted by cmnm^{5}U34 (top) and cmnm^{5}s^{2}U34 (bottom) across ten 500-ns trajectories. (C) Voilinplots comparing the distribution of torsion angles adopted by the C5 moiety of cmnm^5(s^2)U modifications inserted at position 34 of tRNA^{Phe}. The white dots represent the mean, rectangles represent interquartile ranges.

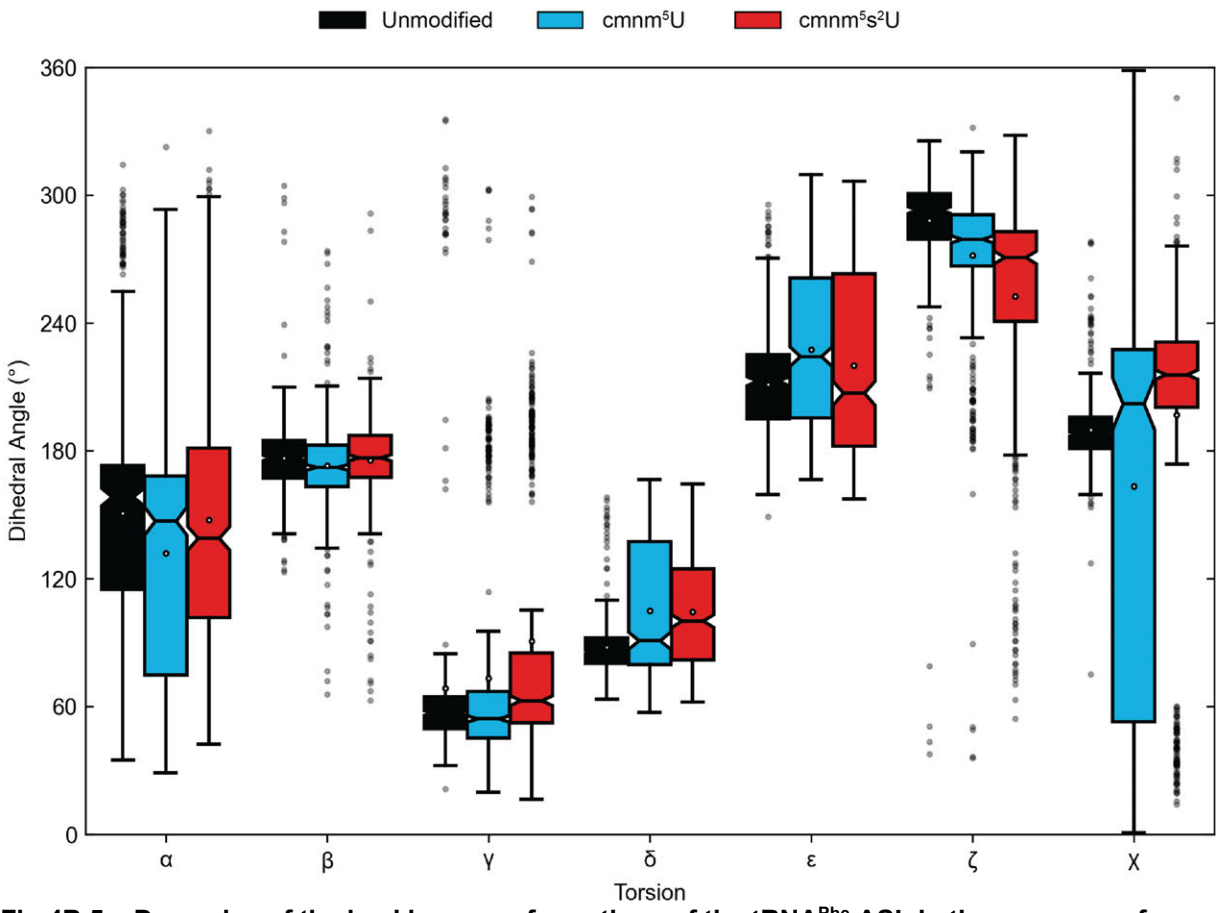


Fig 4B.5 – Dynamics of the backbone conformations of the tRNA^{Phe} ASL in the presence of cmnm^5(s^2)U34 modifications

Statistical analysis of backbone torsions adopted at position 34 in unmodified, cmnm^{5}U-tRNA^{Phe} and cmnm^{5}s^{2}U-tRNA^{Phe} tRNA^{Phe}. The white dots represent the mean, rectangles represent interquartile ranges.

Conformational analyses of the ASL domains of cmnm^{5}U-modified and cmnm^{5}s^{2}U-modified tRNA^{Phe} reveal that both modifications increase dynamics at the anticodon bases, but broader effects are observed in the thiolated system relative to the parent modification. The ASL of unmodified tRNA^{Phe} adopted a wide range conformational states, denoting the flexible nature of this domain (Figure 4B.6, 4B.7A). In particular, bases 34 and 37 were found to be highly dynamic, and these bases were displaced (in comparison to the crystal structure) 64% and 30% of the time respectively. Despite the dynamic nature of the tRNA^{Phe} ASL, its anticodon loop was not susceptible to disorder, as this state was only sampled 2% of the time over the entire replica ensemble. In the presence of cmnm^{5}U34, fewer conformations were adopted by the tRNA^{Phe} ASL, and the majority of the dynamics resided at the wobble base (64%).

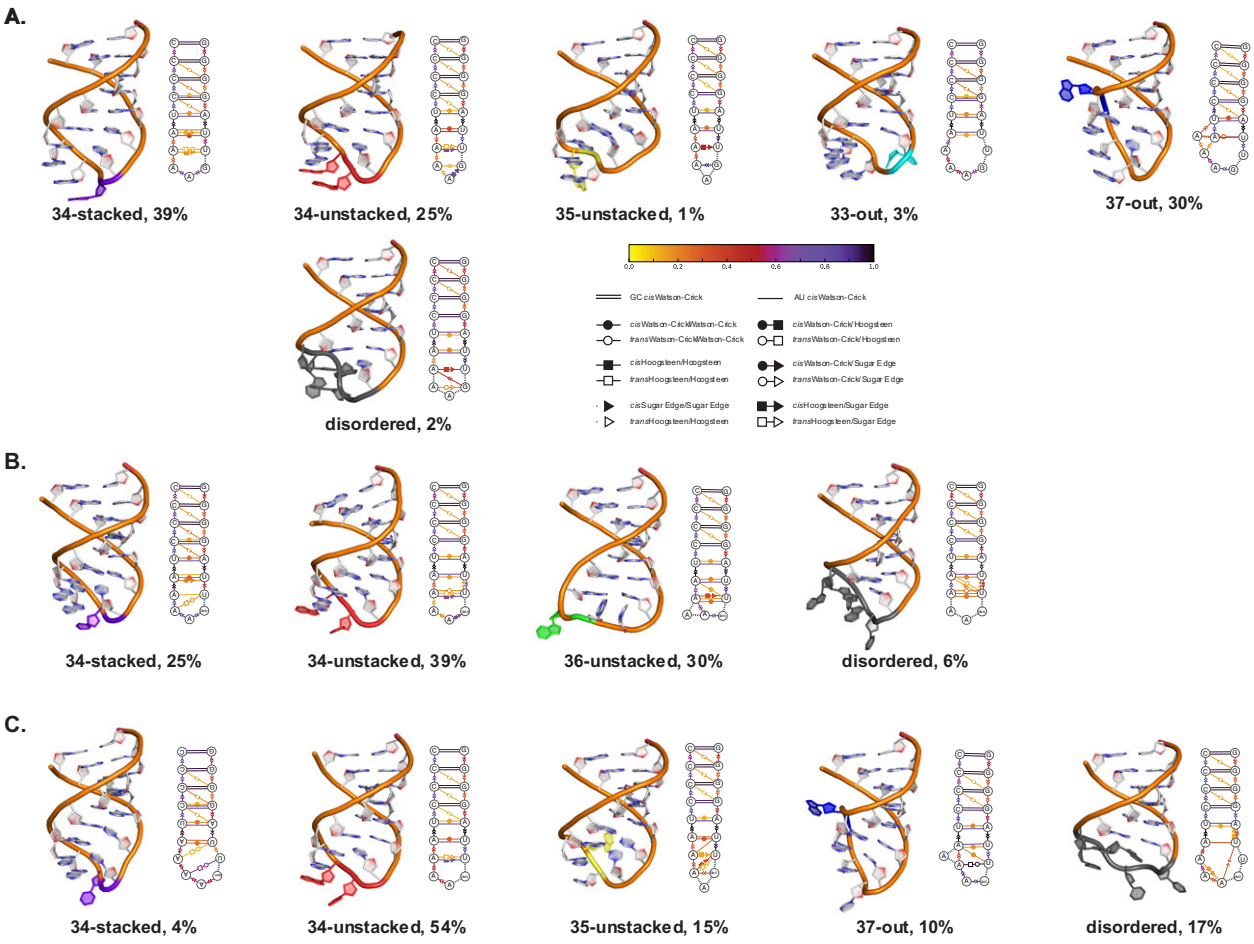


Figure 4B.6 – Conformational profile of the ASL of unmodified and cmnm^5(s^2)U34-modified tRNA^{Phe}

Cartoon and secondary structure representations of conformational states adopted by unmodified tRNA^{Phe} (A), cmnm^{5}U34-tRNA^{Phe} (B) and cmnm^{5}s^{2}U34-tRNA^{Phe} (C) ASLs. Conformations are defined and color-coded as in Chapter 3 and non-covalent interactions are denoted using the Leontis-Westhof notation for RNA molecules.

Notably, conformations that characterize large amplitudes of motion in U33 and A37 were not sampled in the cmnm^{5}U34-tRNA^{Phe} system, implying that the presence of cmnm^{5}U at position 34 may stabilize anticodon flanking bases within the loop. Pseudorotational analyses of the anticodon loop agree with these observations, as backbone torsions were constrained at U34, and fewer conformations were occupied by the backbone of U33 and A37 in the cmnm^{5}U34-tRNA^{Phe} system compared to the unmodified counterpart (Figure 4B.6, 4B.7B). Nevertheless, the cmnm^{5}U34-tRNA^{Phe} model reveals that the presence of the modified nucleotide increases dynamics at other anticodon bases, especially A35, which was found

to be more flexible in the modified system of tRNA^{Phe}. This is substantiated by η–θ analyses that revealed an increase in the backbone flexibility of A35 and A36 relative to the unmodified tRNA^{Phe} model.

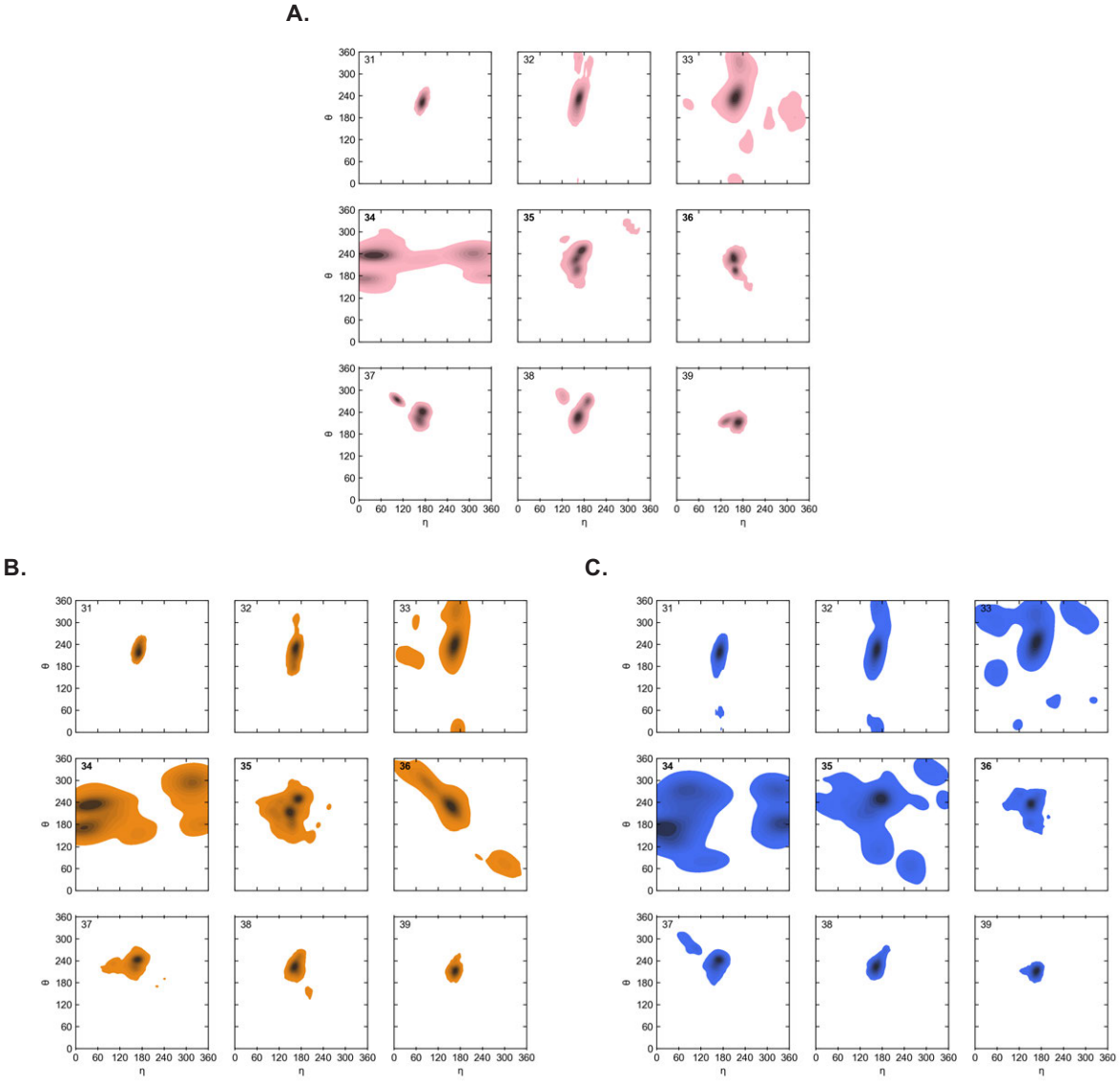


Figure 4B.7 – Pseudorotational analysis of the anticodon loop of tRNA^{Phe}

Backbone pseudotorsions (η = ∠C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$, θ = ∠C4'$_n$-P$_n$-C4'$_{n+1}$-P$_{n+1}$) occupied by residues 31 to 39 in unmodified (A), cmnm^{5}U34-tRNA^{Phe} (B) and cmnm^{5}s^{2}U34-tRNA^{Phe} (C).

Pseudorotational analyses revealed that cmnm^{5}s^{2}U34 restrained the backbone torsions at position 34, but this stability was accompanied by large ranges of motion within the loop's backbone, especially at positions 32, 33 and 35 (Figure 4B.7C). Similarly to its parent modification, the presence of cmnm^{5}s^{2}U at position 34 increased dynamics in A35 ~ 15% of the time (Figure 4B.7A). However, unlike cmnm^{5}U34, the

thiolated modification did not stabilize A37 within the loop, and nucleotide displacements were also sampled at this position 15% of the time. Moreover, cmnm^{5}s^{2}U34 significantly increased the anticodon loop's affinity for a disordered state, which was sampled 17% of time, relative to 2% and 6% in the unmodified and cmnm^{5}U34-modified models of tRNA^{Phe}. The increase in range of motion observed in the cmnm^{5}s^{2}U34-tRNA^{Phe} system relative to the cmnm^{5}U34-tRNA^{Phe} suggests that a single atomic substitution can have far reaching effects on the structural arrangement within the anticodon loop. In this case, the substitution of oxygen with sulfur at C2 of the uridine increased the flexibility of the anticodon loop and reduced non-covalent interactions and structural arrangements at this region.

Overall, the analyses conducted on the three tRNA^{Phe} models (unmodified, cmnm^{5}U34-modified and cmnm^{5}s^{2}U34-modified) reveal that 10-replica ensembles can uncover the structural differences that arise due to the presence of modified nucleobases. Indeed, even conformational changes arising from a single atom substitution were detected using this protocol, validating its utility in future ASL modification studies. More importantly, this preliminary investigation suggests that the primary role of cmnm^5(s^2)U34 modifications in tRNA is to stabilize backbone torsions at the wobble position. Nevertheless, because these modifications add varying amounts of flexibilities within the region, it is important to investigate the synergistic effects between cmnm^5(s^2)U34 and modifications at position 37, especially in tRNAs that have these modifications *in vivo.*