

THE UNIVERSITY OF CHICAGO

ASPECTS OF EUKARYOTIC TRANSFER RNA STRUCTUROME, INTERACTOME, AND
EXPRESSION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE BIOLOGICAL SCIENCES AND
THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON CELL AND MOLECULAR BIOLOGY

BY

NOAH ESCAREÑO PEÑA

CHICAGO, ILLINOIS

DECEMBER 2024

Table of Contents

ABSTRACT.....	IV
LIST OF FIGURES	V
ACKNOWLEDGMENT	VI
CHAPTER 1	1
INTRODUCTION	1
1.1 HUMAN CYTOSOLIC AND MITOCHONDRIAL TRNA STRUCTURES	1
1.2 TRNA POST-TRANSCRIPTIONAL CHEMICAL MODIFICATIONS.....	7
1.3. MAJOR CELLULAR TRNA-PROTEIN INTERACTIONS.....	10
1.3 TRNA EXPRESSION.....	15
1.4 SUMMARY OF THESIS RESEARCH.....	19
CHAPTER 2	21
2.1 INTRODUCTION	22
2.2 RESULTS.....	25
2.2.1 DM-DMS-MAPSEQ DEVELOPMENT	25
2.2.2 CHROMOSOMAL-ENCODED TRNA STRUCTOME IN VIVO	32
2.2.3 IN VIVO - IN VITRO DIFFERENCES IN DMS SIGNALS REFLECT CYTOSOLIC TRNA- PROTEIN/RIBOSOME INTERACTIONS.....	36
2.3 METHODS AND MATERIALS	49
2.3.1 CELL CULTURE AND ARSENITE TREATMENT.....	49
2.3.2 IN VIVO DMS TREATMENT OF HEK 293T CELLS	49
2.3.3 IN VITRO DMS TREATMENT OF HEK 293T TOTAL RNA.....	50
2.3.4 WHOLE-CELL TRANSLATION ASSAY BY 35S PULSE LABELING	50
2.3.5 TRNA SEQUENCING WITH MSR-SEQ.....	51
2.3.6 DATA ANALYSIS OF DM-DMS-MAPSEQ DATA	53
2.3.7 DATA AND CODE AVAILABILITY	54
2.4 DISCUSSION	54
CHAPTER 3	59
3.1 INTRODUCTION.....	59
3.2 RESULTS.....	60
3.2.1 STRUCTURAL MAPPING OF MITOCHONDRIAL TRNAs IN VIVO.....	60
3.2.2 ARSENITE STRESS INDUCES MT-TRNA STRUCTURAL CHANGES.....	66

3.3 METHODS	70
3.4 DISCUSSION	70
<u>CHAPTER 4.....</u>	<u>72</u>
4.1 INTRODUCTION.....	72
4.2 RESULTS.....	74
4.2.1 tRNA SEQ OF SARS-CoV-2 VIRAL PARTICLES REVEAL SELECTIVE ENRICHMENT	74
4.2.4 MODIFICATION PROFILE OF ENRICHED tRNA IN SARS-CoV-2 VIRAL PARTICLES.....	81
4.3 METHODS	85
4.3.1 SARS-CoV-2 ISOLATES	85
4.3.2 RNA ISOLATION.....	85
4.3.3 RNA LIBRARY CONSTRUCTION AND SEQUENCING.....	86
4.3.4 SEQUENCING DATA ANALYSIS	86
4.3.5 DATA AVAILABILITY STATEMENT	87
4.4 DISCUSSION	87
<u>CHAPTER 5.....</u>	<u>90</u>
5.1 INTRODUCTION.....	91
5.2 RESULTS.....	95
5.3 METHODS	98
5.3.1 MSR-SEQ	98
5.4 DISCUSSION	99
<u>CHAPTER 6.....</u>	<u>101</u>
6.1 <i>IN VIVO</i> STRUCTURAL MAPPING OF OTHER SMALL RNA	101
6.2 SELECTIVE PACKAGING OF HOST RNA IN VIRAL PARTICES	102
<u>REFERENCES.....</u>	<u>104</u>

ABSTRACT

Transfer ribonucleic acid (tRNA) is an essential molecule for protein translation across kingdoms of life. Canonically, tRNA is responsible for carrying the amino acid to the ribosome during translation. Of all the RNA families, tRNA has the greatest number of nucleotides that are modified with post translational chemical modifications. Recently, it has been shown that these chemical modifications can significantly regulate expression, mediate RNA-protein interactions, and modulate RNA structure. To further study these specific aspects of eukaryotic tRNA, I and fellow colleagues conducted several studies, and the results of these studies will be presented in this thesis. In the Chapter 1 and Chapter 2, I will discuss the results of an analysis using a newly developed DM-DMS-MaPseq to determine the interactome and structome of chromosomally and mitochondrially encoded tRNA *in vivo*. This analysis provides the first *in vivo* evidence that human EF1A, an important protein for elongation and protein synthesis, interacts with tRNA and amino acids with the same interaction paradigm as EF-Tu. Using arsenite stress, we observed distinct structerome and interactome responses by the cytosolic and mitochondrial translation systems. In Chapter 3, we studied the selective packaging of host RNA in SARS-CoV-2 viral particles using RNA sequencing. We found six anticodon families of tRNA was selectively packaged during SARS-CoV-2 viral particle formation in addition to other small RNA such as Y RNA. We found evidence that packaging of specific tRNA isodecoders was modification and sequence dependent as well, suggesting that a specific sequences or modifications may be used as identity elements for selective packaging. In Chapter 4, we validated the expression of tRNA from a synthetic tRNA chromosome in yeast. During this study, we determined that the synthetic tRNA chromosome did not negatively impact endogenous tRNA expression.

List of Figures

FIGURE 1.1.1	2
FIGURE 1.1.2	6
FIGURE 1.1.3	10
FIGURE 1.2.1	14
FIGURE 1.3.1	16
FIGURE 1.3.2	17
FIGURE 2.1.1	29
FIGURE 2.2.S1	31
FIGURE 2.2.2	35
FIGURE 2.2.3	41
FIGURE 2.2.S3	43
FIGURE 2.2.4	47
FIGURE 2.2.S4	48
FIGURE 3.2.1	62
FIGURE 3.2.S1	65
FIGURE 3.2.2	68
FIGURE 3.2.S2	70
FIGURE 4.2.1	76
FIGURE 4.2.2	78
FIGURE 4.2.3	80
FIGURE 4.2.4	82
FIGURE 4.2.S1	83
TABLE 4.2.2	84
FIGURE 5.1.1	94
FIGURE 5.2.1	96
FIGURE 5.2.2	97

ACKNOWLEDGMENT

I am grateful to the help and mentorship offered by my advisor, Dr. Tao Pan, and members of my thesis committee: Dr. Heng-Chi Lee, Dr. Alex Ruthenburg, and Dr. David Pincus. The work within my dissertation could not have been possible without the support of past and present members of the Pan lab, especially the pioneering work by Dr. Wen Zhang, Dr. Chris Katanksi, and Dr. Chris Watkins in the development of MSR-seq. Special thanks to Dr. Sihao Huang, Dr. Yichen Hou, Dr. Marcus Foo and all the other members of the Pan lab who have helped instill in me the scientific spirit to remain curious. Thank you to Dr. Viviana Simon and her lab, whose work in SARS-CoV-2 allowed the project described in Chapter 4 to even be possible. Thank you to Dr. Daniel Schindler and Dr. Roy S.K. Walker whose work in neochromosome is the basis of the work in Chapter 5 and gave me the opportunity to work with an extensive group of scientists. I would like to extend my thanks to the University of Chicago Biological Science Division and the Cell and Molecular Biology program administrators, whose support helped me throughout my time in Chicago. This work could not be possible with the funding and support from the Initiative for Maximizing Student Development, NIH Diversity Supplemental and the Molecular Genetics and Cell Biology Training Grant.

Finally, the completion of my doctoral degree was made easier due to the love and support from my family. Thank you to my mother, father and sibling: Alma Rita Escareño Peña, Ramiro Jimenez Peña, Val Escareño Peña. I also am eternally grateful to the sacrifices made with love that afforded me the opportunity to pursue this degree made by grandparents: Rosaura Jimenez Peña, Rogelio Zamora Peña, Roberto Escareño, and Minerva Torres Escareño. Thank you to all my aunts, uncles, and cousins who helped raise me and have always cheered me on from home: Maria Juana Escareño, Juana Escareño, Sylvia Escareño, Roberto Escareño, Nelda

Escareño, Victor Escareño, Laurie Escareño, Yolanda Peña Loya , Guadalupe "Lupe" Loya , Deborah Lee Herrera, and Maritza Herrera. Thanks also to my extensive family who have always offered their support and are a source of comfort while I am away from home.

This thesis includes text and figures derived from publications co- authored manuscript in *Cell* published Cell Press, relies on data from a co-first authored paper published in *Frontiers of Cell and Developmental Biology* published by Frontiers Media, and a co-first authored manuscript submitted to *Nature Communications* by SpringerNature.

CHAPTER 1

INTRODUCTION

Transfer ribonucleic acid (tRNA) is an essential component in protein translation by delivery amino acids to the ribosome while also decoding messenger RNA (mRNA) codons to ensure translation fidelity. Many nucleotides of tRNA are also modified chemically known as post-translational modifications, these modifications regulate how tRNA decodes codons, interacts with proteins, and mediates how it folds. tRNA structure and chemical modifications are necessary for its function, their dysregulation can lead to human diseases. (Suzuki et al 2018)

Within this chapter, I will review the human structure for chromosomal and mitochondrial encoded tRNA, tRNA interactome within the cell, and tRNA biogenesis.

1.1 HUMAN CYTOSOLIC AND MITOCHONDRIAL TRNA STRUCTURES

Within humans, there are two types of tRNA that are expressed, chromosomal and mitochondrial. Chromosomal encoded tRNA are more numerous compared to mitochondrial encoded tRNA, with approximately 429 highly confident tRNA genes compared to 22 tRNA genes, respectively. Chromosomal tRNA fold into structures that are like the canonical secondary and tertiary tRNA structure, while human mitochondrial encoded tRNAs have canonical secondary structure, but non-canonical tertiary interactions.

Canonical tRNA is a 76 nucleotide long RNA with five structural features from the 5' to 3' end: acceptor stem, D-stem loop, anticodon stem loop, variable arm, and T-stem loop. For

some tRNA, such as those that encode for Serine or Leucine amino acids, the variable arm is approximately 10 - 15 nt in length which is used for recognition for aminoacylation, while other tRNAs have variable loops 4-5 nt in length. (Sharp et al 1985, Maglott et al 1999) The 3' acceptor stem region also contains a CCA (cytosine, cytosine, adenosine) sequence that is necessary for aminoacylation, or "charging", with the cognate amino acid. Interaction between the D-loop (dihydropyridine loop) and T-loop (TΨC loop) occurs to form the "elbow" region to result in "L shape" tertiary model. (Holley et al 1965, Berg et al 2020) (**Figure 1.1.1**). The specific interactions of nucleotides of tRNA necessary for folding into the "L shape" tertiary structure are used in canonical nomenclature. For example, major D-loop and T-loop interactions that allows for the "elbow" region formation is between guanine at position 18 (G18) in the D-loop and uridine at position 55 (A55) in the T-loop. Another interaction, nt 9 located in the D-stem interact with 12-23 base pair in the D-stem, also helps the elbow fold. These canonical base pair interactions are shown in Figure 1.1.1.

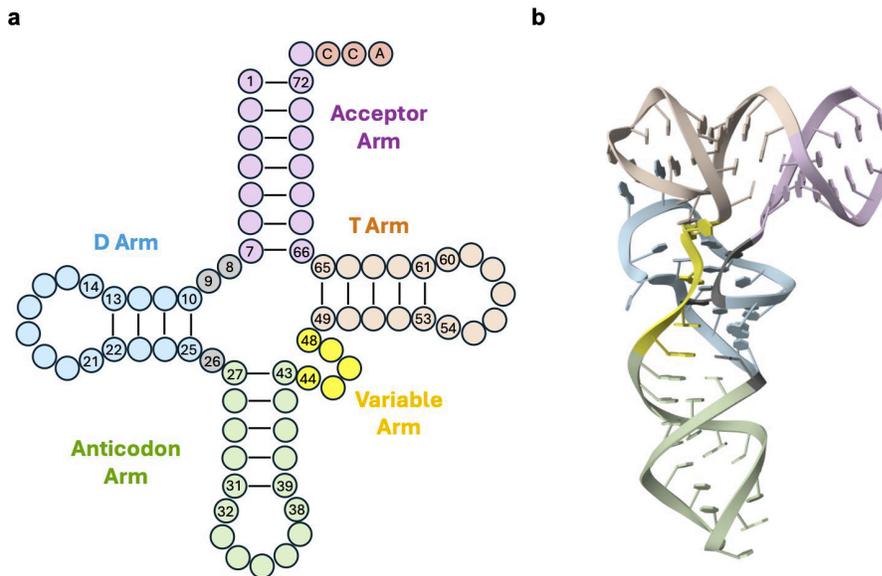


Figure 1.1.1 tRNA structure and tertiary structure and residue nomenclature. (a)

Cloverleaf secondary structure, also showing standard tRNA nomenclature and canonical

position. (b) Tertiary structure modified from a tRNA^{Val} structure (PDB: 8CTI), different stem-loop regions are color coded like in (a).

To decode mRNA codons, tRNA relies upon an anticodon sequence located at the canonical positions 34, 35, and 36 in the anticodon loop. Aminoacylated tRNA carry the amino acid to the ribosome during translation where the anticodon nucleotides can pair with mRNA codons via hydrogen bonding. (Crick et al 1961, Nirenberg et al 1965, Berg et al 2020,) In canonical tRNA, the distance of the anticodon sequence to the 3' CCA sequence is approximately 70 Å, which is the distance between decoding and peptidyltransferase centers within the ribosome indicating how the tertiary structure of the tRNA is important for its function in protein synthesis (Suzuki 2021).

Human mitochondrial tRNA are encoded by the mitochondrial genome, which expresses 22 species of mitochondrial tRNA decoding 60 mitochondrial mRNA sense codons in the mitochondrial genome. The mitochondrial genome is a circular double stranded DNA that encodes 37 genes divided into three major groups: 13 genes encode essential subunits for the respiratory complex and ATP synthase, 22 for each mt tRNA, and two for mitochondrial rRNAs. (Suzuki et al, 2011) To ensure the proper translation of these mitochondrial genes, mitochondrial has its own protein synthesis machinery that is composed of mitochondrial tRNA and mitochondrial ribosomes. Human mitoribosome is composed of three major group of macromolecules: ribosomal proteins import from the cytosol, two mitochondrial rRNA which consist of the 16S and 12S subunits, and mitochondrial tRNA^{Val}. (Suzuki 2018, Laptev et al 2020) The mitochondrial tRNA^{Val} is used as a structural component which also shows how mitochondrial translation system varies from chromosomal encoded translation systems. It

should also be noted that mitoribosomes RNA components are all encoded from the mitochondrial genome. However, the protein components of mitoribosomes are encoded by the nuclear genome, translated in the cytosol, and imported from the cytoplasm to the mitochondria. These protein components of the mitoribosome include the ribosomal proteins, translation factors, aminoacyl tRNA synthetases (aaRS), and other factors that are necessary for mt-RNA processing and maturation. (Steinberg et al 1994, Frazer-Abel et al 2008) Mitochondrial genome decoding deviates from the canonical genetic code by using AUA for Met instead of Ile, UGA for Trp which is normally a stop codon, and using AGR (R = A and G) as stop codons. (Suzuki et al 2021) The result of this is that all amino acid sets of codons can be decoded by one of the 22 mt tRNA isodecoders. 22 tRNAs are the minimum number of tRNAs necessary to encode all amino acids by any organism or organelle.

Human mitochondrial tRNA structural features are more diverse compared to cytosolic tRNA. Mitochondrial tRNA has three types that deviate from the canonical structure. (**Figure 1.1.2**) The first of these mitochondrial tRNA structures was type 3 which was first reported in 1980, when studying the mammalian mt tRNA^{SerAGY}, that can decode the AGU and AGC codons. (Suzuki et al 201). Type III tRNA lacks the entire D stem compared to the canonical structure, but it can still be aminoacylated and function during translation. *In vitro* structural analysis identified that type III tRNA contains a flexible tRNA structural core, which allows for interactions to form a "boomerang" structure that can maintain the necessary distance between the 3' CCA and anticodon to function within the ribosome. Type I was later reported when analyzing tRNA^{SerUCN}, (Shimada et al 2001) it still has all four major regions of the canonical secondary structure but retains some distinct differences: a single base pair between the acceptor stem and D loop stem, an extended anticodon loop by an additional six nucleotides. Despite

these variations, type I still maintains a D-loop and T-loop interaction to fold into the "L" shape. Finally, Type II mitochondrial tRNA are those that do not retain the cytosolic tRNA D-loop and T-loop interacting nucleotides but still retain D-loop and T-loop features. Most of the mitochondrial tRNA have been classified as type II, by either *in silico* modeling or *in vitro* structural analysis. The type II tRNA, although highly variant even amongst each other, have differing D-loop and T-loop sizes by either additional or fewer nucleotides compared to the canonical, cytosolic tRNA structure. The varying sizes of these loop regions disrupt the canonical D-loop and T-loop interaction. To compensate, type II tRNA have interactions between the D-loop stem with the small variable loop which ultimately allows the "L" shape structure to form for type II tRNA and let them maintain their tRNA function.

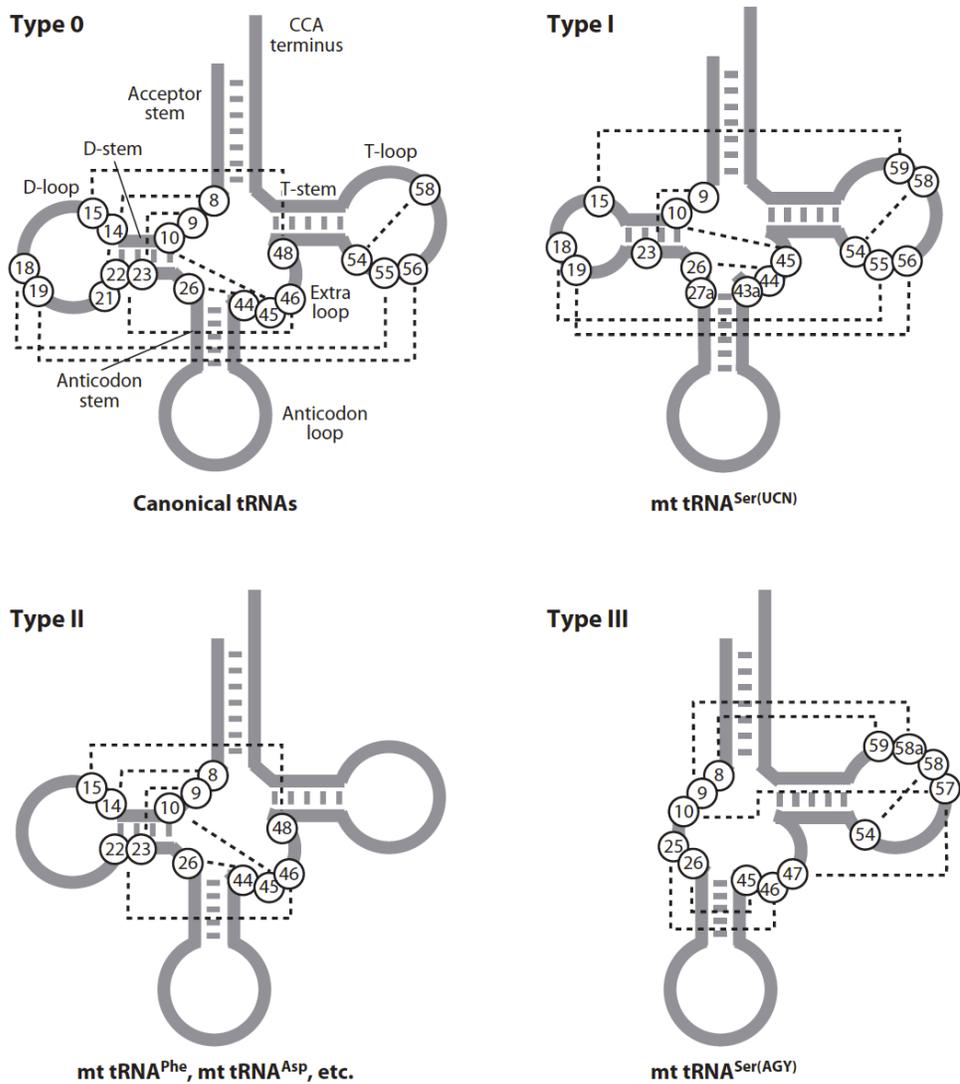


Figure 1.1.2 Structures of human mitochondrial tRNA. (a) Type 0 is canonical tRNA structure, (b) Type I (c) Type II, (d) Type III. Circled numbers represent tRNA positions following tRNA nomenclature. Interactions are indicated with dotted lines. Figure was adapted from Suzuki et al 2011. Copyright by *Annual Reviews of Chemistry*.

1.2 TRNA POST-TRANSCRIPTIONAL CHEMICAL MODIFICATIONS

Of all the RNA types, tRNA is the most highly modified with chemical modifications. These modifications mediate tRNA structure stability, recognition of cognate tRNA by aaRS for aminoacylation, and decoding of mRNA during protein translation. (Goto et al 2008, Guan et al 2006) Based on recently published studies, a total 39 type of modifications have been identified in human tRNA (Wang et al 2023, Suzuki et al 2020, Pan et al 2018, Kimura et al 2020). Cytosolic tRNA contains approximately 13 modifications per molecule on average, whereas mitochondrial tRNA have approximately five modifications per molecule. (Suzuki et al, 2020, Pan 2018, Kimura 2020, Schimmel et al 2018)

Studies attempting to comprehensively map the human tRNA modifications has grown immensely in the past decade. To date, using deep sequencing techniques and LC-MS, 39 types of modifications have been identified in the cytosolic tRNA. (Suzuki 2018, Crécy-Lagard et al 2019). In LC-MS, each tRNA is first purified using affinity chromatography with complementary oligos on solid resin or beads. The purified tRNA is then digested using an RNase that cuts at guanine nucleotide to generate small oligos that can be analyzed using mass spec. Using LC-MS, all the 18 mitochondrial tRNA modification types were identified (Suzuki et al 2014).

Modifications that consist of methylation at 2'-OH or nucleobase are chemically simplest and the most abundant in human tRNA. (Boccaletto et al 2018, Zhang et al 2021) The base methylations on tRNA include N¹-methyladenosine (m¹A), 1-methylguanosine (m¹G), 3-methylcytidine (m³C), and N²-dimethylguanosine (m²₂G). (Zhang et al 2021) Other tRNA modifications include pseudouridine (Ψ), queuosine (Q), and 3-(3-amino-3-carboxypropyl)uridine (acp³U), and wybutosine (W). (**Figure 1.1.3**) These modifications can be

found throughout the body of the tRNA such as the D-arm, T-arm, and variable loop. (Helm et al 2014) There are hotspots of tRNA modifications, specifically the position 34 which is the wobble anticodon nucleotide that can accommodate irregular base pairings as well as position 37 which is the 3' nucleotide adjacent to the anticodon nucleotides.

Modifications on the anticodon loop are involved in codon recognition and ensure accurate reading frame maintenance with mRNA during protein synthesis. (Han et al 2018, Björk et al 1982) As first theorized by Crick in 1966, an unmodified U at position 34 (U34) can base pair with A and G in mRNA. Later, it was described that the wobble U34 can even support four-way base pairing with all nucleotides, such as those occurring in human mitochondrial tRNA-mRNA decoding (Suzuki et al 2011). Therefore, in theory, an amino acid family box of codons with a U34 can be decoded by a single tRNA due to the wobble position pairing. On the other hand, a modified U34 in tRNA restricts decoding to only two codons in an amino acid family box. (Yoshida et al 2015, Suzuki 2018) Examples of wobble U34 modifications for cytoplasmic and mitochondrial tRNA include 5-methoxycarbonylmethyluridine (mcm^5U) and 5-carboxyaminoethyluridine ($cmnm^5U$), respectfully. (Suzuki et al 2020, Asano et al 2018) As explained earlier, mitochondria utilize a single tRNA to decode the four codons per amino acid. This single tRNA has an unmodified U34 to engage in four-way wobbling. However, for amino family codons that contain at least two codons ending with purines (adenosine or guanine) residues, their corresponding tRNA have a modified U34 such as taurinomethyluridine (τm^5U) and 5-taurinomethyluridine 3-thiouridine (τm^5s^2U). These U34 modifications of mitochondrial tRNA prevent misreading of pyrimidine ending codons and stabilize pairing with G-ending codons. (Suzuki et al 2020) Hypomodification of τm^5U in mitochondrial tRNA has been associated to mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes

(MELAS). Another example of wobble position modification is 5-formylcytidine (f⁵C) of mitochondrial tRNA^{Met} and allows the decoding of non-canonical start codon AUA and AUG in mitochondrial translation. Disruption of enzymes that biosynthesize f⁵C in mitochondrial tRNA result in reduced mitochondrial activity, evidence that f⁵C34 is essential for proper mitochondrial function. (Cantara et al 2013, Moriya et al 1994, Kawarada et al 2017).

An interesting tRNA modification is queuosine, Q34 which is a hypermodified guanine that has a 7-deazapurine structure at the wobble anticodon position in human and bacteria. In all organisms, queuosine is found on tRNA^{His}, tRNA^{Asp}, tRNA^{Asn}, and tRNA^{Tyr} in both cytosolic and mitochondrial tRNAs. Q34 enhances translation speed by increasing translation of U-ending codons compared to its unmodified G34 tRNA counterpart; loss of Q34 impedes translation and protein folding which activates stress response pathways. (Muller et al 2019, Tuorto et al 2018) Bacteria can biosynthesize queuine, which is the base moiety of queuosine. (Iwata-Reuyl, et al 2003) Humans cannot synthesize queuine. Thus, humans obtain queuine as a micronutrient from the gut microbiome or dietary sources, which was validated experimentally using germ-free mice that were feed queuine deficient diet, resulting in queuosine modification being absent in cellular tRNA. (Marks et al 1997) To synthesize Q34 modification in humans, a tRNA guanine transglycosylase (TGT) substitutes guanine nucleobase with queuine. Knockouts of TGT subunits resulted in loss of the queuosine modification in tRNA. (Boland et al 2009, Suzuki et al 2020, Zhang et al 2018) Queuosine is an example of nutrient control of tRNA modifications, as well as indicating a link between human translation efficiency with the gut microbiome.

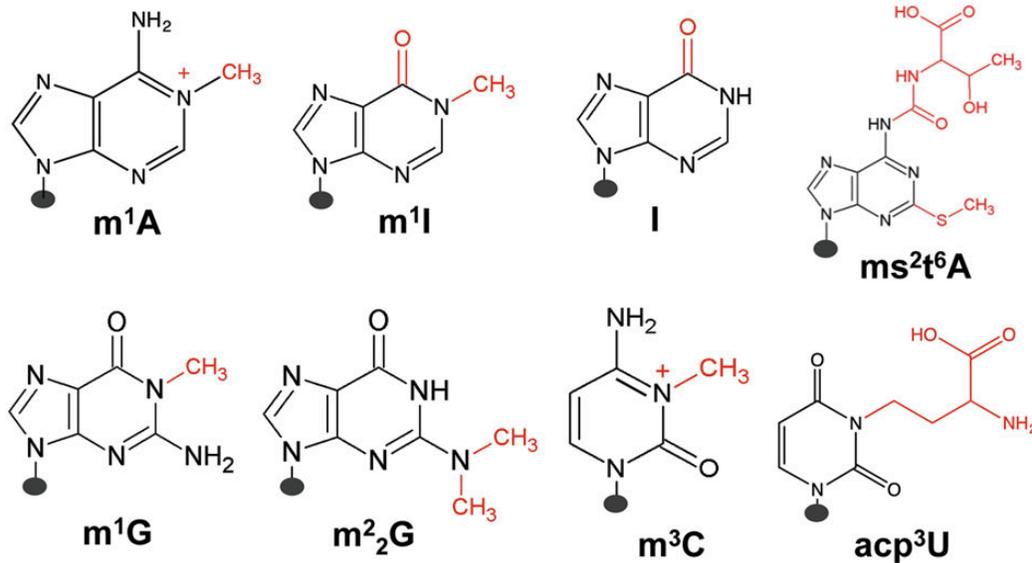


Figure 1.1.3 Common tRNA modifications found on human cytosolic and mitochondrial tRNA. Figure was adapted from Pan 2018. Copyrights by *Cell Research*.

1.3. MAJOR CELLULAR TRNA-PROTEIN INTERACTIONS

Aminoacyl tRNA synthetases are responsible for proper recognition of cognate tRNA and aminoacylate with the appropriate amino acid. Briefly, these synthetases participate in a two-step reaction that binds amino acid to the 3' end of tRNA. Amino acids are activated by ATP at the catalytic site of aminoacyl tRNA synthetases, forming aminoacyl-AMP that stays attached to the synthetases and release P_i. The 3' adenosine at position 76 of the tRNA attacks the carbonyl carbon of AMP, which causes tRNA to be covalently attached to the amino acid at the 3' end. This mechanism of action of aminoacylation in aaRS is universally conserved, but the synthetases do have structural diversity. There are two classes of aminoacyl tRNA synthetases that are organized by their architecture and active site placement. (Shimmel et al 1991, Ribas de Pouplana et al 2001, Cusack 1997) Class I and class II synthetases bind to tRNA differently.

Class I binds to tRNA at their minor groove of the acceptor stem, whereas class II binds to the major groove of the acceptor stem.

To ensure fidelity during protein translation, aaRS must maintain proper recognition of cognate tRNA and amino acids. aaRS binding to tRNA begins as an electrostatic interaction between polar nucleotide residues with the tRNA phosphate backbone. (Beuning & Musier-Forsyth 1999, Perona and Hou 2007) Following this initial electrostatic interaction phase, identity elements of tRNA are utilized to ensure proper recognition of cognate tRNA. These identity elements include certain nucleotide positions, nucleotide modifications, and structural features. (Jordan Ontiveros et al., 2019., Larkin et al 2002) The anticodon nucleotides 34, 35, and 36 of tRNA and position 73 adjacent to the 3' of the acceptor stem are important identity elements for many tRNA-aaRS binding. Using the anticodon loop for recognition is common for most aaRS, however, for amino acids with 6 codons which include Ser and Leu this is not possible because they do not share anticodon nucleotides that allow for recognition. Seryl tRNA synthetase (SerRS) and Leucyl-tRNA synthetase (LeuRS) use an alternative approach, where each enzyme interacts with the long tRNA variable loop to recognize the appropriate cognate tRNA. (Hou and Schimmel 2018) Unique to tRNA^{Ala}, its identity element for its synthetase is a G3-U70 base pair which is absent in all other tRNAs.

In the mitochondria, 19 of the mt aaRS are encoded by the human genome. (Antonellis et al 2008) Humans do not have a mitochondrial aaRS that corresponds to GluRS, instead, tRNA^{Glu} is generated by converting Gln-charged tRNA^{Glu} to Glu-tRNA^{Glu} using another set of enzyme. (Ibba et al 2000) Mitochondrial LysRS and GlyRS share the same gene as the cytosolic LysRS and GlyRS, while the rest of the mitochondrial aaRS has their own genes within the chromosome. Alternative splicing of the LysRS gene results in the production of the cytosolic

and mitochondrial isoforms. (Moukadiri et al 2009) The two isoforms of GlyRS are produced from the same gene but have different initiation sites. (Tolkunova et al 2000) Following translation, mitochondrial aaRS are transported to the mitochondria.

Eukaryotic EF1A and its prokaryotic counterpart, EF-Tu is one of the most abundant cytosolic proteins in any cell. (Jiang et al 2020, Lund et al 1996, Lee et al 1992) Due to this high abundance and rate of translation within the cell, it is the primary protein that interacts with the cytosolic tRNA and delivers the tRNA to the A site of ribosomes. EF1A/EF-Tu binds the acceptor stem and T stem-loop regions of charged aminoacyl-tRNA. EF1A/EF-Tu binds to charged tRNA preferentially, indicating the esterified amino acid is necessary for tRNA binding with EF1A. (Lariviere et al 2001) Measuring binding affinity of the EF1A prokaryotic homolog, EF-Tu, to tRNA shows that binding is determined both by the tRNA sequence and the specific amino acid charged to the tRNA. (Asahara and Uhlenbeck 2002) However, to ensure the fidelity of translation, EF1A must recruit all charged tRNA without preference. To reconcile the evidence of this binding specificity, a uniform EF-Tu binding aa-tRNA paradigm was derived which requires a balancing specificity for specific tRNA sequences and the corresponding amino acids: tRNAs with sequences that confer strong binding are charged with amino acids that confer weak binding to EF-Tu, and vice versa.

Ribosome interacts with three tRNAs during translation. Charged tRNA is recruited to the A site of the ribosome and undergoes intermediate tRNA conformation changes known as A/T state of binding. (Moazed et al 1989) Within this conformation, aa-tRNA is still bound to EF-Tu at the 3' acceptor region while the anticodon stem loop region interacts with the mRNA. Proper pairwise binding between the tRNA anticodon and the mRNA codon occurs through the interaction of conserved nucleotides of A1493, A1492, and G530 of the small ribosomal subunit

(SSU). These 16S rRNA residues interact with the minor groove of the tRNA-mRNA duplex to confirm proper tRNA-mRNA base pairing which induces a 30S ribosome conformational closure. (Schmeing and Ramakrishna, 2009) Near cognate tRNA however fails to induce this conformation change, which results in lower stability and rejection of the aa-tRNA, thus acting as a proofreading mechanism. 30S domain closure brings the sarcin-ricin loop (SRL) of the large ribosomal subunit (LSU) to interact with EF-Tu which catalyzes the GTP hydrolysis of EF-Tu. After GTP hydrolysis, EF-Tu is still attached to aa-tRNA for a short time due to the delayed release of inorganic phosphate (Pi), a byproduct of the hydrolysis reaction. During this period, the ribosome can still dynamically change from open to close domains. This acts as an additional proofreading mechanism to ensure proper cognate aa-tRNA interaction, by allowing EF-Tu and near-cognate aa-tRNA to be released before further accommodation in the ribosome occurs. Following release of the inorganic phosphate, EF-Tu is also released from the ribosome which leaves tRNA in the A site of the ribosome. (Thompson et al 1977) During this period, the anticodon and elbow region of tRNA interacts with the ribosome and ribosomal proteins which determines whether tRNA is also accommodated into the peptidyl transferase center. If a near cognate tRNA is present, the ribosome converts into an open conformation and releases the tRNA. If elbow and anticodon interactions do not trigger release from the ribosome, the SSU of the ribosome conducts a forward swivel which moves the 3' acceptor stem of the tRNA into the peptidyl transferase center. Once this occurs, the amino acid from the tRNA is quickly transferred to the elongating peptide chain. (Pape et al 1998) After peptide bond formation, the elongation factor G (EF-G) stabilizes tRNA interactions within the ribosome and increases the speed the translocation of tRNA using GTP hydrolysis. Hydrolysis EF-G-GTP aids the change of the SSU orientation that results in the necessary forward swivel for translocation to occur.

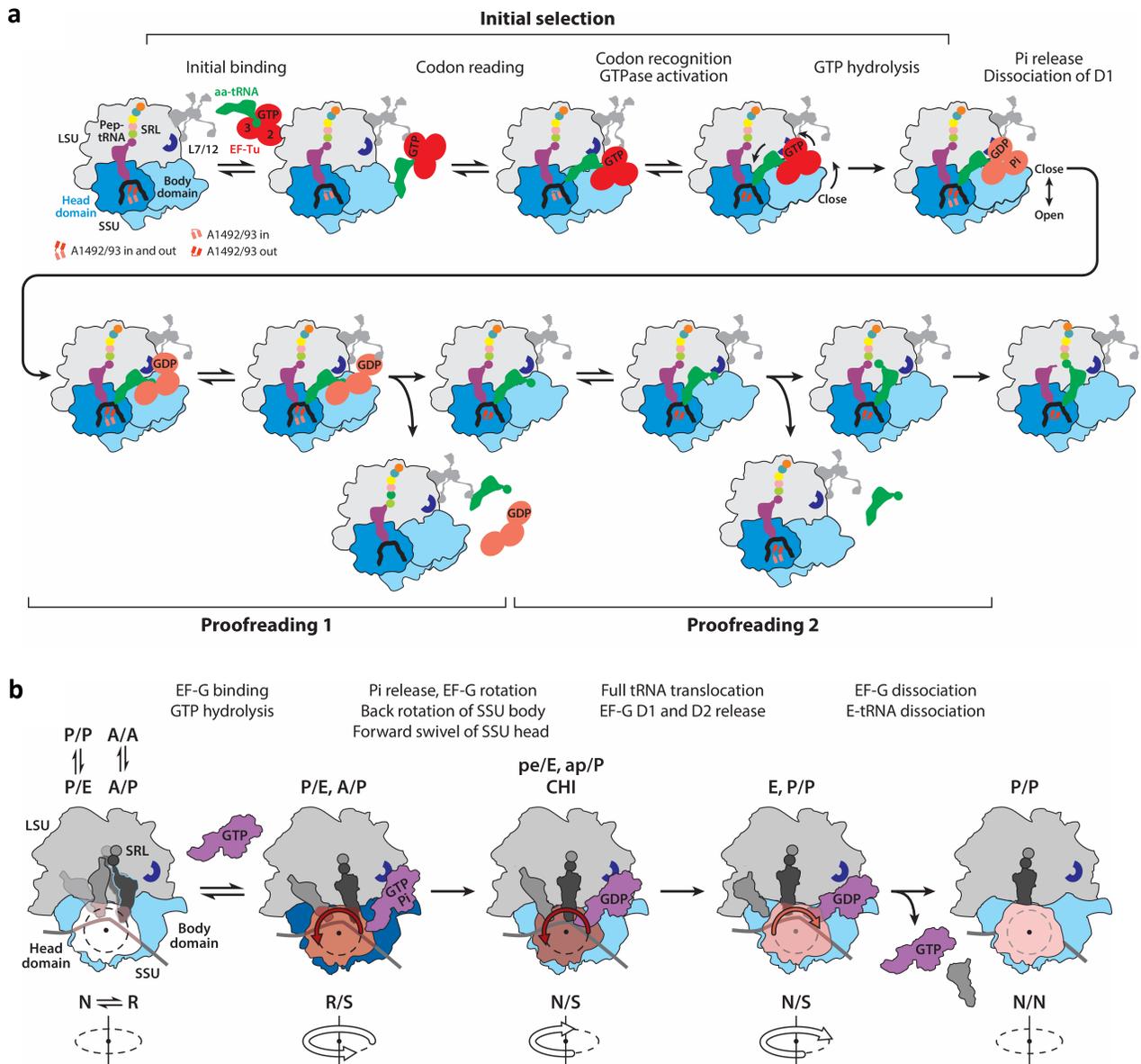


Figure 1.2.1 tRNA incorporation, accommodation, and translocation in ribosome. (a) This is modeling of charged tRNA bound by EF1A homolog, EF-Tu, being recruited to the ribosome A site, where it binds with mRNA. Upon proper formation of a tRNA-mRNA duplex, the ribosome SSU closes its domain and the SSR hydrolysis GTP which products Pi. Near cognate tRNA can be removed from the ribosome when EF-Tu-GDP (Pi) is attached, which is proofreading 1, or when EF-Tu has fully dissociated, proofreading 2. (b) N denotes the orientational of ribosomal proteins and the tRNA. that are nonrotated, R denotes the rotated orientation of the ribosomal

proteins and tRNA in hybrid states. EF-Tu binds to ribosome and accelerates the R orientation. GTP hydrolysis leads to a release of Pi and causes the SSU to swivel to an N orientation. This causes the charge tRNA to translocate to the P site. SSU swivels back to S orientation and EF-Tu is released. This figure was adapted from Rodina, M.V., 2023. Copyright by *Annual Review of Biophysics*.

1.3 tRNA EXPRESSION

In humans, tRNA genes are transcribed by RNA polymerase III (RNAPIII) (Gabrielsen et al, 1991). RNAPIII is an 18 subunit enzyme that also transcribes 5S rRNA, a component of the cytosolic ribosome, U6 snRNA, a component of the spliceosome, and several other abundant non-coding RNAs. For RNAPIII to transcribe tRNA genes, two sequences must be present: the A box and B box. These are located in the internal region of the tRNA gene. In 1980 Segal et al (ref.) determined that TFIIC and TFIIIB from column fractions that could transcribe tRNA genes. During tRNA gene transcription, TFIIC recognizes and binds to A and B box. (**Figure 1.3.1**) The distance between A and B box is determined by length of the variable region of the tRNA gene. Upon binding to the tRNA gene, TFIIC recruits TFIIIB to the upstream region of the tRNA gene, resulting in a TFIIC and TFIIIB complex. This complex recruits RNAPIII to the tRNA gene leading to promoter melting and transcription until reaching the termination sequence, which for tRNA genes is generally consecutive 4-5 T residues at the non-template strand. (Arimbasseri et al 2015, Braglia et al 2005)

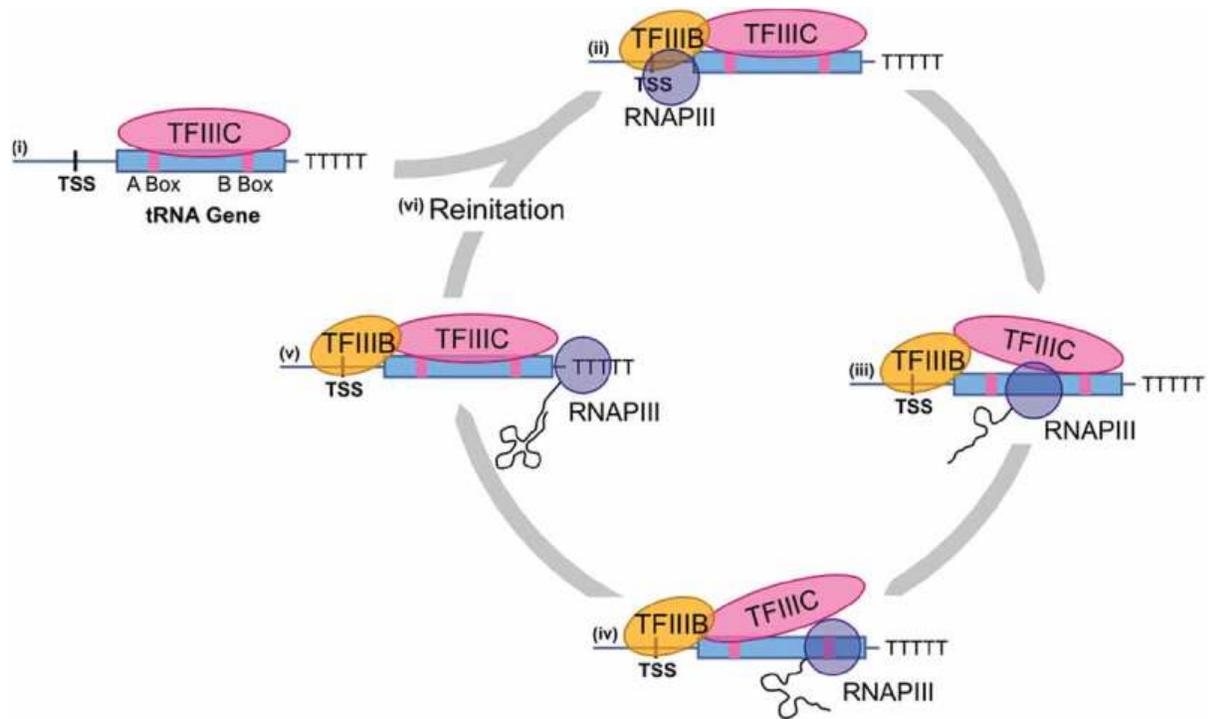


Figure 1.3.1 tRNA transcription by RNAPIII. (i) TFIIC interacts with the A box and B box sequence, recognizing it as a tRNA gene. (ii) TFIIB is recruited to TFIIC which in turn recruits RNAPIII to the tRNA gene. (iii) The promoter is melted by TFIIB and (iv) RNAPIII which leads to the transcription of the tRNA gene. (v) Transcription is terminated by a series of thymine residues on the non-template strand. (vi) RNAPIII can be recruited back tRNA gene by reinitiation which increases tRNA gene transcription. This figure was adapted from Berg and Brandl 2021. Copyrighted by *RNA Biology*.

Human mitochondrial genome is a double stranded circular genome, that encodes 37 mitochondrial genes including 22 tRNAs, 13 mRNAs and 2 rRNAs. **(Figure 1.3.1)** The strands of the mitochondrial genome are named heavy and light strand, each having their own respective

composed of mitochondrial RNA polymerase (POLMRT), Tfam, and mtTFB (Tiranti et al 1997, Flierl et al 1997). For the light strand, a single LSP starts transcription for the whole strand. The heavy strand has two HSP. HSP1 only transcribes tRNA^{Phe}, 12S rRNA, tRNA^{Val}, and 16S rRNA, which make up the short heavy strand, whereas HSP2 transcribes the mitochondrial genome starting from 12S rRNA. Termination of HSP1 transcription is mediated by transcription factor mTERF (mitochondrial termination factor) that recognizes HSP1 and the mitochondrial termination sequences, which is located at the tRNA^{Leu(UUR)} gene. (Martin et al 2005)

After transcription, the tRNA transcripts still need to be processed into mature tRNA. These initial tRNA transcripts are known as "pre-tRNA". For cytosolic tRNA, RNase P cleaves the upstream nucleotides in the pre-tRNA sequence, producing the 5' end of mature tRNA. This RNase P cleavage requires the first and second base pairs of the acceptor stem. The cleavage also generates the canonical seven base pairs of the acceptor stem. (Carrara et al 1989, Paisley et al 1994, Maraia et al 2011) The 3' of pre-tRNA contains a tract of uridine nucleotides, which is a binding site for the La protein. The La protein prevents premature cleavage of the 3' U tract and aids in tRNA folding. After folding, RNase Z cleaves the 3' U tract. At this point, the tRNA nucleotidyltransferase catalyzes the addition of 3' CCA to form the mature tRNA 3'CCA terminus. This addition of CCA is done without a template. Mitochondria RNA transcripts also need to be processed to generate the mature mitochondrial tRNAs. Due to the tRNA genes located between rRNA and mRNA genes, precursor mitochondrial tRNAs do not have intervening 5' or 3' sequences between them and other mitochondrial RNAs. Instead, both the 5' and 3' ends of tRNA transcripts are processed to separate them from mitochondrial rRNA and/or mRNA. For ribonucleases, the mitochondrial RNase P generates the mature 5' end, whereas

RNAse Z (ELAC2) generates the mature 3' end. For the 3' CCA addition, mitochondrial CCA adding enzyme is used (TRNT1). Following this tRNA processing to mature tRNA, both tRNA are modified with tRNA modifications by various RNA modification enzymes.

1.4 SUMMARY OF THESIS RESEARCH

In this thesis, I will be presenting three major projects. These projects focus on studying *in vivo* structurome and interactome of eukaryotic tRNA, selective RNA packing by viral particles, and validating tRNA expression from a synthetic yeast chromosome. I will summarize the major results and conclusion from each of these projects in their respective chapters. It should be noted that the work presented in this thesis is either currently under review or has been published.

In Chapter 2 and 3, I will present the results of using DM-DMS-MaPseq to study eukaryotic tRNA structure and interactions *in vivo*. This technique utilized dimethyl sulfate (DMS) chemical probing coupled with demethylase (DM) treatment to conduct mutational profiling (MaP) of eukaryotic tRNA *in vivo*. DMS is known to methylate exposed cytosine or adenosine nucleotides residues on RNA and has been previously used to study RNA structures and their response to stress (Wang et al 2019, Tomezko et al 2020, Yamagami et al 2022) Methylated residues disrupt reverse transcription, these disruptions can be detected by RNA sequencing by mutational signatures, thereby allowing for mutational profiling of RNA residues that have been modified by DMS. Due to DMS methylating single stranded residues, the mutational signatures of DMS-induced methylation can greatly inform RNA structures as well as regions that are protected from protein interaction. Use of DMS mutational profiling to study RNA structure has been previously limited to mRNA, but small RNA such as tRNA, has been

limited with only plant or bacteria tRNA being well investigated. To my knowledge, *in vivo* structome and interactome analysis have not been conducted for human tRNA. The results for cytoplasmic and mitochondrial eukaryotic tRNA are presented in Chapter 2 and Chapter 3, respectively. In Chapter 4, I will focus on the results of studying the host RNA that was selectively packaged by SARS-CoV-2 viral particles during viral particle formation. To determine selective packaging of SARS-CoV-2, we utilize MSR-seq, an RNA sequencing developed by the Tao Pan lab. It has been known that during viral particle formation, host RNA is selectively packaged with the most well studied instance of this phenomenon being the selective packaging of tRNA^{LysTTT} during HIV-1 viral particle formation. However, studies of RNA selective packaging have been limited to retroviruses and none have been conducted for SARS-CoV-2 viral particles, which make our study one of the first to rigorously investigate the host RNA selectively packaged by SARS-CoV-2. In Chapter 4, I with the aid of colleagues of the Pan lab, utilized precursor tRNA to validate the expression from a synthetic yeast chromosome to determine whether the synthetic chromosome impacted endogenous tRNA expression. The results of this study validated the expression of a synthetic yeast chromosome was possible in yeast and further moved the field of synthetic biology closure to accomplishing the goals of developing a synthetic and modular *S. cerevisiae*.

CHAPTER 2

IN VIVO STRUCTURE PROFILING OF HUMAN CYTOSOLIC tRNA STRUCTURE AND INTERACTOME RESPONSE TO STRESS

Acknowledgment: This chapter is derived from an article that has been submitted for publication (Peña, Hou, Watkins, et al). The authors of this article were Noah Peña, Yichen Hou, Christopher P. Watkins, Sihao Huang, Wen Zhang, Christopher D. Katanski, and Tao Pan. N.P., C.P.W., T.P. conceived the project. N.P., Y.H., T.P. analyzed and interpreted data, with help from S.H., C.D.K. C.P.W. performed the DMS experiments and built sequencing libraries, with help from W.Z. Y.H., N.P., C.P.W., T.P. wrote the paper.

Note: There are three terms that will be repeated throughout this thesis that require definition, isodecoders, isoacceptors, and anticodon families

- Isodecoders are tRNAs that have the same anticodon sequences, but different sequences in the body, and are encoded by different genes in the human genome.

- Isoacceptors are tRNAs that are charged with the same amino acid, but have different anticodon sequences.

- An anticodon family represent all tRNA isodecoders that share the same anticodon sequence.

2.1 INTRODUCTION

Within cells, secondary and tertiary structures of RNA can allow the detection of metabolites, recruit proteins, regulate phase separation, mediate RNA-RNA interactions, and affect RNA stabilities. (Kertesz et al 2010, Talkish et al 2014, Langdon et al 2018) However, methods to study tRNA structure have been primarily conducted *in vitro*, such as using X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (Cryo-EM). These methods, although powerful, do not allow a study of how RNA structure in a cellular context. Therefore, utilization of next generation sequencing (NGS) coupled with chemical probing of cells to map RNA structures *in vivo* have been developed.

Chemical probing of RNA looks at the reactivity of RNA nucleotides to specific chemical treatments, which in turn inform accessibility and provide information on RNA structures. Common applied chemicals utilized for chemical probing reagents include dimethyl sulfate (DMS), kethoxal, selective hydroxyl acylation analyzed by primer extension (SHAPE), and RNAases. These reagents induce structure-dependent chemical adducts or strand scissions on RNAs, which can be detected using NGS (Ding et al 2014, Rouskin et al 2014,). Of all these reagents, DMS was the first used for *in vivo* RNA structure mapping, due to its endowed properties that allow it to have high cell permeability, high oxidative reactivity for RNA within cells, and a reaction time in orders of minutes which would greatly reduce the oxidative stress of treated cells. (Tijerina et al 2007, Tomesko et al 2020). DMS alkylates N¹ of adenosine and N³ of cytosine nucleotides on the Watson-Crick face of single-stranded RNA, which are detected as either stops or mutations upon reverse transcription to generate cDNA. In 2014, three parallel studies were conducted using DMS chemical probing with deep sequencing to provide insight on the cellular structure of *A. thaliana*, *S. cerevisiae*, and human cells. (Ding et al 2014, Rouskin et al

2014, Talkish et al 2014) From these reported structures, a striking finding was that *in vivo* mRNA structures were drastically different compared to their *in vitro* forms. Notably, stable mRNA structural motifs *in vivo* were less structured *in vitro*. The explanation for this was that thermodynamics drive structural stability for mRNA *in vivo*, while *in vitro* RNA structures are heavily influenced and determined by interactions with energy dependent unwinding, such as with helicases, as well as other intricate interactions with metabolites, proteins, and other nucleic acids. These findings show that a comprehensive understanding of RNA function and regulatory roles requires a realistic view of RNA native structures within cells.

Chemical probing *in vivo* can also allow for "foot printing" to study RNA-protein interactions, as regions that where proteins interact with RNA are shield from chemical modifications. (Smola et al 2015) As referred to earlier, large differences between *in vitro* and *in vivo* chemical probing signals were seen. It has since been proposed that a large part of these differences between *in vitro* and *in vivo* chemical probing signals could be indicative of the protein-binding motifs on mRNAs. Therefore, a comparison between the *in vivo* and *in vitro* chemical probing signals at RNA regions that interact with proteins could allow *in vivo* of RNA-protein interactions at a whole cell level to be studied. However, the complexity of whole-cell RNA-protein interactions, the transient nature of many binding events, and the low coverage of individual mRNA structure elements limited these studies to the analyses of a few well-characterized mRNA-protein interactions and those between rRNAs and ribosomal proteins.

RNA structures are also fundamental to the function of other types of RNAs, including long non-coding RNAs (lncRNAs), microRNAs (miRNAs), small-nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and tRNAs. Luo et al. applied a SHAPE-based method to study the pre-miRNA substrates for Dicer and uncovered genetic elements crucial to Dicer

binding and cleavage, providing an example of the small RNA structurome contributing to exciting biology. Recently, Yamagami et al. applied tRNA structure-seq to examine *in vivo* structures of *E. coli* tRNAs and revealed dynamics of tRNA folding under heat stress.

(Yamagami et al 2022)

The reference human genome encodes a diverse set of over 600 putative tRNA genes, among which 429 are considered tRNA genes with high confidence, i.e., they likely fold into the canonical tRNA structures based on their predicated base pairing and tertiary interaction properties. Nevertheless, whether any given chromosomal-encoded tRNA transcript folds into a canonical tRNA structure *in vivo* have not been tested systematically. As tRNA function in translation requires the canonical structure, *in vivo* structural information would serve as a validation of tRNA functionality. Moreover, tRNAs are also an excellent model system for studying RNA-protein interactions by *in vivo* chemical probing since the structures of tRNAs and many of their interacting protein complexes are well characterized.

Here, we develop an *in vivo* structural mapping method for tRNA and other structured noncoding RNAs using DMS mapping with mutation profiling coupled with m¹A/m³C demethylase (DM) treatment to enhance the efficiency and accuracy of structural mapping for the highly structured tRNA. DMS reaction in cells with highly structured RNA generates lower reactivities than mRNA. As reverse transcriptases may also generate mutations at modified RNA nucleotides and structural motifs, an AlkB demethylase treatment was applied to selectively remove the m¹A/m³C methyl groups before cDNA synthesis and thereby revealing DMS-specific mutations in sequencing reads. We term the method DM-DMS-MaPseq. We unveiled the *in vivo* tRNA structurome for cytosolic tRNAs at the isodecoder level and for mitochondrial tRNAs. We found that small sequence variations between isodecoders can have a significant

impact on tRNA structural features. Furthermore, we revealed the dynamic tRNA folding and protein interaction during arsenite stress-induced translational reprogramming.

2.2 RESULTS

2.2.1 DM-DMS-MaPseq development

DM-DMS-MaPseq is a combination of *in vivo* DMS chemical probing and Multiplexed Small RNA sequencing (MSR-seq), which was recently developed to achieve efficient sequencing of highly structured and highly modified tRNA species. (Watkins et al 2022, Yamagami et al 2022, Zubradt et al 2016) Although tRNAs had been difficult to sequence due to their stable structure and numerous modifications, recent advances in tRNA-seq have overcome most of these challenges. (Watkins et al 2022) The major steps of DM-DMS-MaPseq are depicted in **Figure 2.2.1a**. For *in vivo* structure profiling, the protocol started with DMS treatment on live human HEK293T cells. Total RNAs were extracted and used directly as the input for library preparation. For *in vitro* structure profiling, DMS treatment was performed on extracted, deproteinated total RNA. Following the MSR-seq protocol, total RNAs, including deacylated tRNAs, were ligated to barcoded, biotinylated multiplexing adapters. Multiple barcoded samples were combined and immobilized on streptavidin beads. The combined samples were split into two, one mock and the other treated with the *E. coli* wild-type AlkB demethylase, which selectively demethylated m¹A and m³C. (Watkins et al 2022, Fedeles et al 2015) Library construction proceeded with reverse transcription using the thermostable Superscript IV reverse transcriptase and an optimized reaction condition, followed by on-bead second adapter ligation and PCR to generate DNA libraries for Illumina sequencing. The sequencing data was processed to obtain mutation signatures at each nucleotide position. We obtained 4.4-16.8 million reads

mapped to human chromosomal and mitochondrial encoded tRNA genes (to be discussed further in Chapter 3) from these libraries, with mapping rates ranging from 55-71% (**Table 2.2.1**). The read counts between biological replicates are highly correlated (**Figure 2.2.1.S1a**). As expected, most mapped reads are tRNAs, with the remainder mostly mapped to 5S/5.8S rRNAs and other abundant non-coding RNAs, such as snRNAs and snoRNAs, that are of similar length to tRNAs (**Figure 2.2.1.S1b**). Our data have very high positional coverage for many tRNAs, which enables in-depth analysis of the tRNA structure and interactions *in vivo*.

DMS predominantly methylates the Watson-Crick face of the unpaired A and C residues. The resulting m¹A and m³C can be read as mutation signatures in our sequencing data, as thermostable RTs can read through them. Given the presence of abundant and diverse tRNA modifications and the interference of tRNA structure to RT reaction, it is critical to validate bona fide DMS signatures in our experiment. The *E. coli* wild-type AlkB demethylase (DM) selectively reverses the DMS products of m¹A or m³C back to the unmodified A or C. Therefore, comparing sequencing libraries with and without DM treatment could more confidently assign the mutation signatures derived from m¹A or m³C. Globally, when comparing the mutation rate without and with DM treatment ($DM_{diff} = \text{Mutation rate (-DM)} - \text{Mutation rate (+DM)}$), **Figure 2.2.1a**) at different nucleotides, only A and C residues show high and positive mutation rate differences values in DMS-treated samples, validating the selectivity of demethylation (Supplemental Figure 1c).

We present two abundant non-coding RNAs to illustrate the feasibility of DM-DMS MaPseq. The 5S rRNA is part of the 60S ribosomal subunit *in vivo* and has a well-defined secondary structure and high-resolution structure in the ribosome. To showcase the effect of DM

treatment, DM_{diff} under various treatment conditions *in vitro* and *in vivo* are shown for the 5S rRNA (**Figure 2.2.1b**). Our demethylase treatment was highly effective in removing the methylation induced mutation signatures for both *in vitro* and *in vivo* samples. To quantify the DMS reactivity at each nucleotide, we introduce the “DMS signal,” which is equal to the difference between DM_{diff} with and without DMS reaction (DMS signal = $DM_{diff}(+DMS) - DM_{diff}(-DMS)$), **Figure 2.2.1a**). The *in vivo* DMS signal is drastically lower than the *in vitro* DMS signal, consistent with 5S rRNA being protected as part of the ribosome *in vivo* (**Figure 2.2.1c**). Zooming into the *in vivo* high reactivity region around A42, the three *in vivo* DMS conditions show the expected results of zero DMS signal for untreated, high DMS signal for high DMS concentration, and intermediate DMS signal for low DMS concentration (**Figure 2.2.1d**). The *in vitro* DMS signal pattern largely recapitulates the known 5S rRNA secondary structure (**Figure 2.2.1e**). The residues with the highest DMS signals *in vivo* include A42, A90, A101, and A103, among which A42, A101, and A103 are located in the loop regions and on the surface of the ribosome, thus are expected to be more prone to DMS reaction (**Figure 2.2.1e**). These results validate our experimental approach for the assessment of highly structured RNA *in vitro* and *in vivo*.

The DMS signals for the chromosomally encoded cytosolic initiator tRNA^{Met} (tRNA^{iMet}) are also much higher *in vitro* than *in vivo* (**Figure 2.2.1f**). The known natural modifications m¹A58 and m¹G9 in this tRNA³⁰ produce strong mutation rates, which are largely reversed upon DM treatment. DMS reaction generates additional mutation signatures primarily in the D, anticodon, variable, and anticodon loop regions *in vitro*, consistent with tRNA^{iMet} folding into the canonical secondary structure (**Figure 2.2.1g**). DMS signals in these regions are drastically reduced *in vivo*, consistent with tRNA^{iMet} being protected *in vivo*, presumably through its

interactions with the initiation factor eIF2B and the 40S ribosome. Zooming into the anticodon loop region of the *in vivo* samples again shows the expected no DMS signal for untreated, high DMS signal for high DMS concentration, and intermediate DMS signal for low DMS concentration (**Figure 2.2.1h**). Hence, we focus our downstream tRNA analysis on the *in vivo* untreated and high DMS concentration samples.

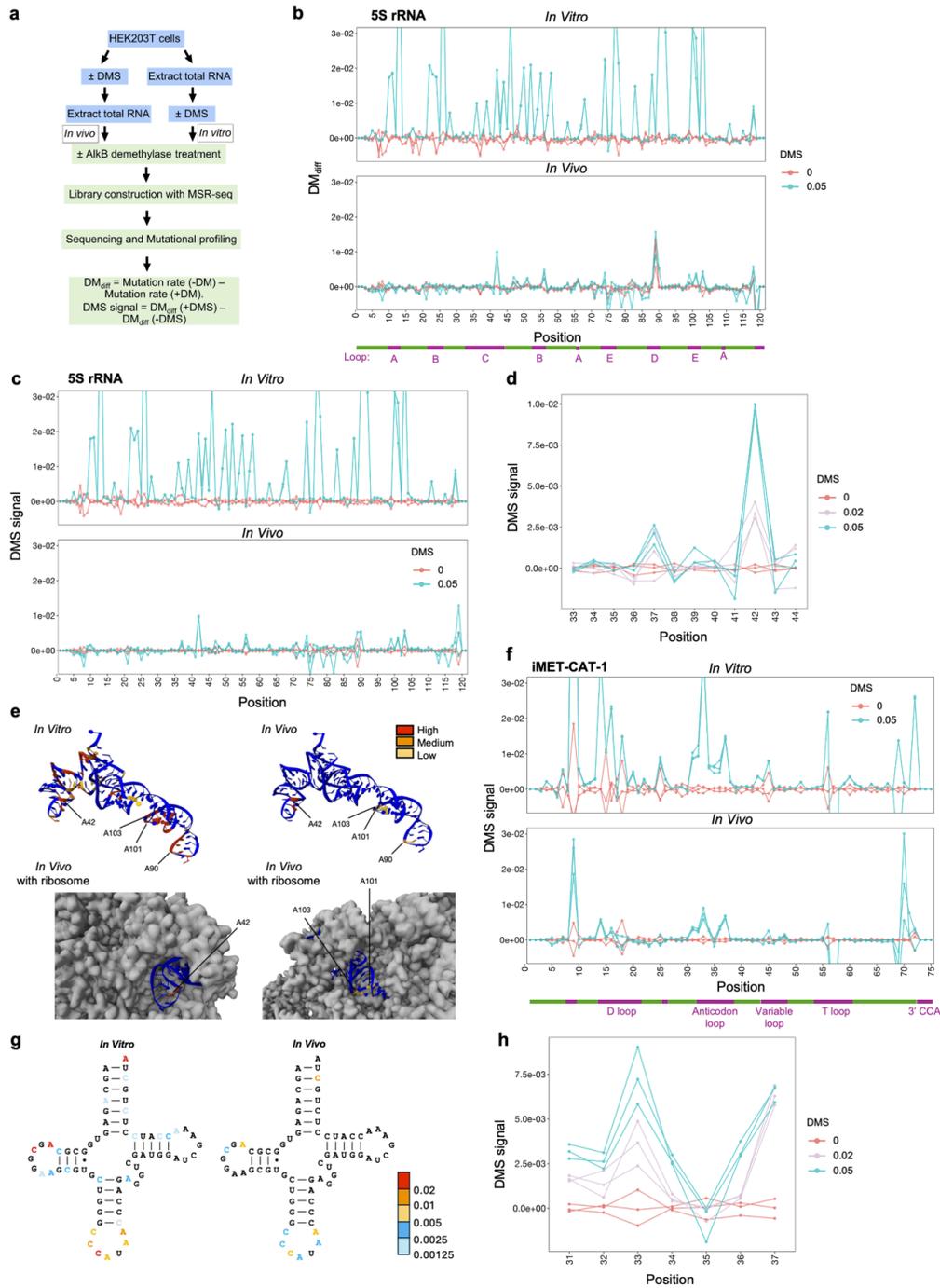


Figure 2.1.1 DM-DMS-MaPseq method recapitulates rRNA and tRNA conformations *in vitro* and *in vivo*.

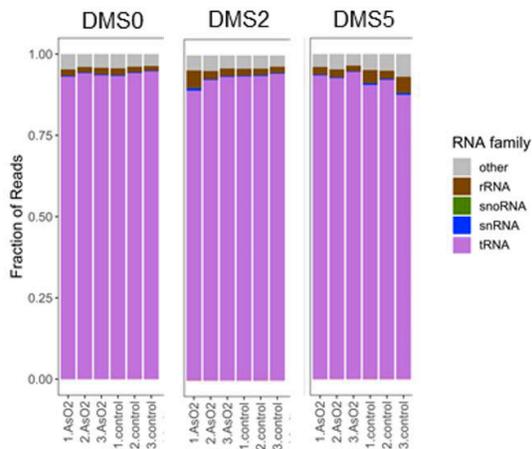
(a) Schematic diagram showing the DM-DMS-MaPseq experimental 725 setup and data analysis

workflow. (b) Line graphs of in vitro and in vivo positional DM_{diff} values of 5S rRNA with (blue) or without (red) 5% DMS treatment (n = 3 biological replicates for each). (c) Line graphs of in vitro and in vivo positional DMS signals of 5S rRNA with (blue) or without (red) 5% DMS treatment (n = 3 biological replicates for each). (d) Line graphs of in vivo positional DMS signals in the loop C region of 5S rRNA with 0% (red), 2% (light purple), or 5% DMS (blue) treatment (n = 3 biological replicates for each). (e) Cryo-EM structures of 5S rRNAs in the human 80S ribosome (shown as 5S rRNA alone or in the ribosome, PDB: 4UG0) overlaid with in vitro and in vivo DMS signals of high (red), medium (orange), or low (yellow). (f) Line graphs of in vitro and in vivo positional DMS signals of tRNA^{iMet} with (blue) or without (red) 5% DMS treatment (n = 3 biological replicates for each). (g) Secondary structure of tRNA^{iMet} overlaid with in vitro and in vivo DMS signals. (h) Line graphs of in vivo positional DMS signals in the anticodon region of tRNA^{iMet} with 0% (red), 2% (light purple), or 5% DMS (blue) treatment (n = 3 biological replicates).

a r^2 -coefficient

Replicates	DMS0	DMS2	DMS5	DM	AsO ₂ stress
<i>In vivo</i>					
1 vs 2	0.97	0.97	0.97	-	-
1 vs 3	0.99	0.98	0.99	-	-
2 vs 3	0.99	0.99	0.99	-	-
1 vs 2	0.97	0.98	0.98	+	-
1 vs 3	0.97	0.98	0.98	+	-
2 vs 3	0.95	0.94	0.96	+	-
1 vs 2	0.95	0.95	0.95	-	+
1 vs 3	0.98	0.99	0.96	-	+
2 vs 3	0.99	0.99	0.99	-	+
1 vs 2	0.96	0.96	0.96	+	+
1 vs 3	0.93	0.96	0.90	+	+
2 vs 3	0.98	0.98	0.98	+	+
<i>In vitro</i>					
1 vs 2	0.99		0.99	-	
1 vs 3	0.99		0.99	-	
2 vs 3	0.99		0.98	-	
1 vs 2	0.99		0.98	+	
1 vs 3	1		0.98	+	
2 vs 3	0.99		0.98	+	

b



c

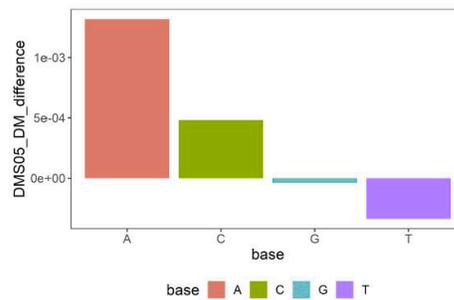


Figure 2.2.S1 Additional results for Figure 1 showing quality control of DM-DMS-MaPseq reads.

(a) Statistics of replicates. (b) Mapping of fractions of tRNA and other RNAs. DMS0: no DMS, DMS2: 2% DMS, DMS5: 5% DMS. (c) Nucleotide identity of DM_{diff} samples with 5% DMS treatment

2.2.2 Chromosomal-encoded tRNA structome *in vivo*

More than 600 tRNA genes are annotated in the human chromosome. Each tRNA gene is assigned a tRNAScan score in the genomic tRNA database, which correlates with their predicted secondary structure formation strength and conservation of residues that form tertiary interactions.³³ By convention, a tRNAScan score of >50 is considered a high confidence tRNA that likely forms a canonical tRNA structure, a cloverleaf secondary structure comprised of four stems and four loops with the nucleotides 14-21 (conventional tRNA nomenclature) in D, 32-39 in anticodon, 44-48 in variable, and 54-60 in T loops (**Figure 1.1a**). The tertiary structure is formed by a network of interactions among nucleotides in the D, T, and variable loops (**Figure 1.1b**). A total of 429 high-confidence tRNA genes are present in the hg38 genome, composed of 267 unique tRNA sequences (isodecoders). (Chan et al 2016) Given the diversity of tRNA sequences and the variation among their predicted structural stabilities, we systematically tested to what extent cytosolic tRNAs conform with canonical structures *in vivo*.

The *in vitro* and *in vivo* DMS signal profiles for the most abundant cytosolic tRNA isodecoders among each of the 47 tRNA families with different anticodons are shown in **Figure 2.2.2a**. Both *in vitro* and *in vivo* DMS signals show clearly discernable patterns based on the tRNA secondary structure, where loop regions and the 3'CCA have significantly higher DMS signals than stem regions. The anticodon loop region shows the highest DMS signals which is consistent with this region forming the least extent of tertiary interactions among all loops in tRNA. For these most abundant tRNA in each anticodon family, the DMS signals are all consistent with the formation of the canonical secondary structure.

Putting aside the mutations derived from incomplete demethylation of the naturally occurring m¹A, m³C, m¹G, and m²²G modifications (marked in **Figure 2.2.2a**), the *in vitro* DMS

signals are always higher than *in vivo* DMS signals. This is also confirmed by examining the *in vitro* and *in vivo* averaged DMS signal distribution among all tRNAs in each stem-loop region (**Figure 2.2.2b**). Given that tRNAs form highly stable structures *in vitro*, lower DMS signal *in vivo* may be attributed to lower DMS reactivity *in vivo*, as DMS diffusion into cells may result in lower *in vivo* DMS concentration. On the other hand, tRNAs interact extensively with proteins and ribosomes in cells, which can protect them from reacting with DMS. As an exemplar, the interaction model is more consistent with the much higher DMS signals seen in the 3'NCCA region *in vitro* than *in vivo*: for each tRNA, this region binds to the EF-1A protein or is buried in the ribosome *in vivo*, which can explain the dramatic difference.

To further explore how sequence variations between high confidence isodecoders impact tRNA folding, we analyzed the *in vivo* DMS signal profiles of two pairs of tRNA isodecoders with different tRNAScan scores. tRNA^{Ser}(AGA)-1 and -4 have three nucleotide differences: C25 and U25 (standard tRNA nomenclature) that form a G10-C25 or G10-U25 pair at the end of D-stem, U39 and C39 that forms an A31-U39 pair in SerAGA-1, but an A31-C39 in SerAGA-4, and U50e and C50e in the loop of the variable stem-loop that is not involved in any base pairing in the tRNA. Despite just three sequence changes, tRNA^{Ser}(AGA)-4 has significantly higher DMS signals, suggesting that it has a less stable structure *in vivo* compared to tRNA^{Ser}(AGA)-1 (**Figure 2.2.2c**). The conversion of A31-U39 in SerAGA-1 to A31-C39 in SerAGA-4 markedly increases the DMS signals in the anticodon loop. The conversion of G10-C25 to G10-U25 increases the DMS signals on the opposite end of the variable loop region at ~15 Å away (**Figure 2.2.2d**), suggesting that the tertiary interaction network of tRNA can transmit long-range effects in structural change. tRNA^{Val}(UAC)-1 and tRNA^{Val}(UAC)-4 differ at 15 out of 76 positions in acceptor, anticodon, and T stems and multiple loop regions. All stem sequence changes are

matched with compensatory changes that maintained the Watson-Crick base pairs, whereas the loop changes can weaken noncanonical base pairs or tertiary interaction such as U54 in ValUAC-1 to G54 in ValUAC-4 which can disrupt G18-U54 tertiary interaction. These sequence changes result in higher DMS signals in ValUAC-4 than in ValUAC-1 (**Figure 2.2.2e**). Again, the structural change permeates throughout the tRNA structure (**Figure 2.2.2f**). In both tRNA^{Ser}(AGA) and tRNA^{Val}(UAC) cases, the respective high scoring isodecoder indeed has lower DMS signals than the low-scoring isodecoder.

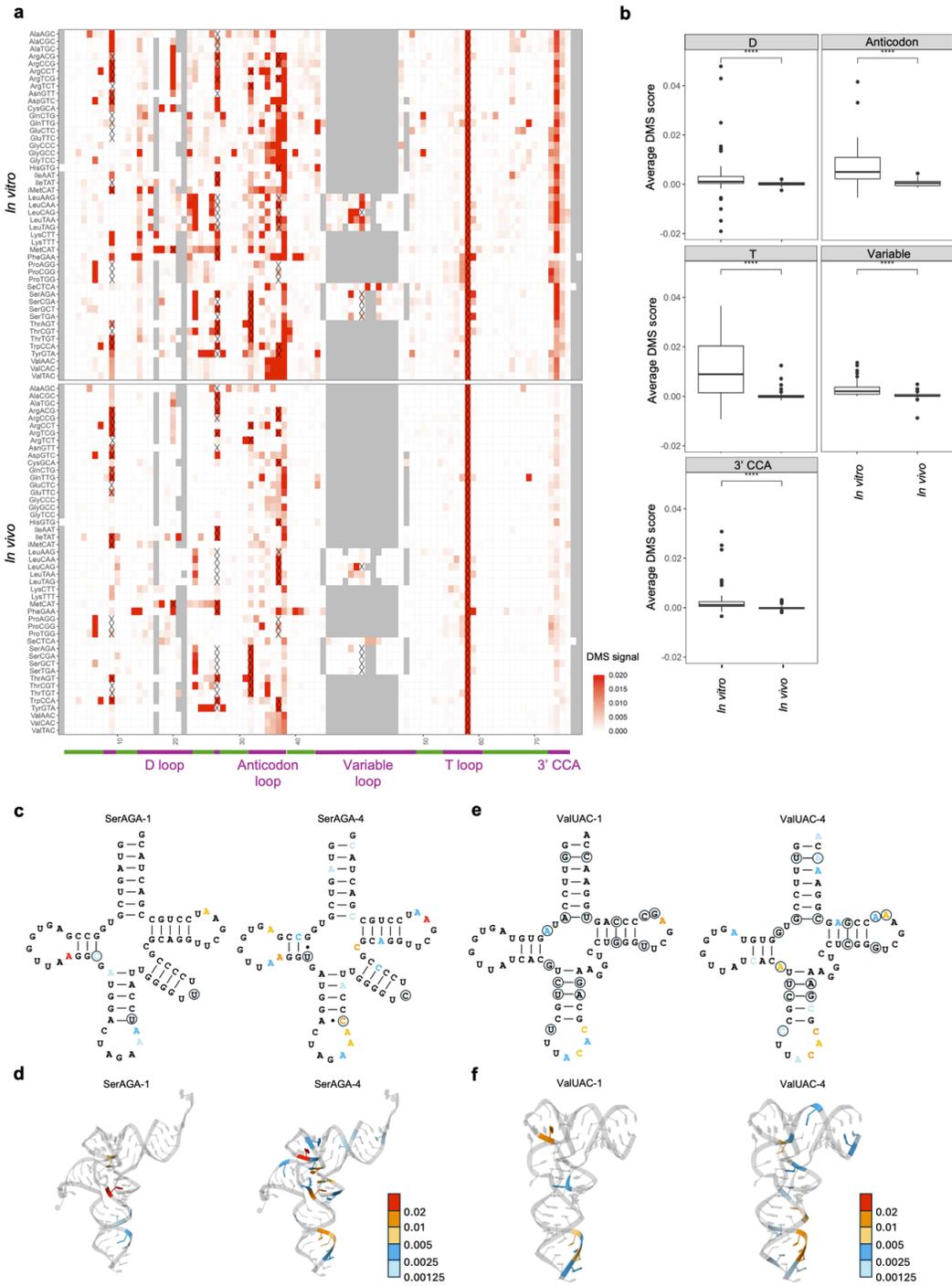


Figure 2.2.2 DMS Mapping of chromosomal-encoded tRNA reveals *in vitro* versus *in vivo*, and *in vivo* differences.

(a) Heatmap showing the *in vitro* and *in vivo* positional DMS signals for the most abundant

tRNA isodecoder in each anticodon family. Native modifications³⁰ that are incompletely removed by demethylases are marked with “X”. (b) *In vitro* and *in vivo* average DMS signals in D, T, anticodon, variable loops, and the 3’ unpaired CCA regions. P-values are calculated with standard Wilcoxon t-test; ****: $p < 10^{-4}$. (c) Secondary structures of two tRNA^{Ser}(AGA) isodecoders, Ser-AGA-1 and Ser-AGA-4, overlaid with *in vivo* DMS signals. Sequence differences are indicated as circled nucleotides. tRNAscan score: Ser-AGA-1, 90.4, Ser-AGA-4, 78.8 (d) Tertiary structures of two tRNA^{Ser}(AGA) isodecoders, Ser-AGA-1 and Ser-AGA-4, overlaid with *in vivo* DMS signals (modified from the E site tRNA in the human 80S ribosome, PDB: 6Z6M) (e) Secondary structures of two tRNA^{Val}(UAC) isodecoders, Val-UAC-1 and Val-UAC-4, overlaid with *in vivo* DMS signals. Sequence differences are indicated as circled nucleotides. tRNAscan score: Val-TAC-1, 82.9, Val-TAC-4, 68.7. (f) Tertiary structures of two tRNA^{Val}(UAC) isodecoders, Val-UAC-1 and Val-UAC-4, overlaid with *in vivo* DMS signals (modified from the tRNA^{Val} in the human METTL1-WDR4-tRNA(Val) complex, PDB: 8CTI).

2.2.3 *In vivo - in vitro* differences in DMS signals reflect cytosolic tRNA-protein/ribosome interactions.

Previous structure mapping studies on mRNAs have concluded that mRNAs are prone to denature *in vivo* due to energy-driven cellular interactions, thereby generally showing higher DMS reactivities *in vivo*. (Zubradt et al 2016) In our case, however, the *in vitro* DMS signal for each chromosomal encoded tRNA is substantially higher than its *in vivo* counterpart (**Figure 2.2.2a**). Cytosolic tRNAs interact with two major groups of proteins and the ribosome required for their participation in translation. These interactions can protect tRNAs from reacting with DMS *in vivo*, thereby reducing *in vivo* DMS signals of the interacting tRNA motifs, resulting in

a region-specific decrease in DMS signals. Dominant cytosolic tRNA binding partners include the highly abundant aminoacyl-tRNA synthetases (aaRSs), the elongation factor EF1A, and the ribosome. tRNAs are aminoacylated (a.k.a., charged) by 20 aaRSs, one for each amino acid. Cellular tRNA charging levels are very high, but charged tRNAs turn over quickly during translation so that tRNA-aaRS interaction occurs constantly. Charged tRNAs are bound by EF1A, which delivers tRNAs to the ribosome for protein synthesis. EF1A is among the most abundant cytosolic proteins in a human cell³⁸ and the primary interacting protein for all cytosolic tRNAs. During translation, tRNAs occupy the A, P, and E-sites on the ribosome, and approximately one-third of all tRNAs can be bound by ribosomes in cells. As all three types of tRNA-protein interactions involve distinct combinations of tRNA structural elements, we set out to deduce these interaction modes by comparing the *in vivo* and *in vitro* DMS signals. To assess the *in vivo* and *in vitro* differential DMS signals in individual tRNA regions and infer information on tRNA-protein interaction, we introduce the τ score, which is the normalized difference between *in vitro* and *in vivo* DMS signals at each nucleotide for each tRNA isodecoder (**Figure 2.2.3.S3a**). The τ value is between 0-1; a higher τ score implicates stronger protection of DMS reaction by tRNA interacting partners *in vivo*. The positional τ scores of the most abundant isodecoder from all tRNA anticodon families show that the loop regions tend to have higher τ scores, which are derived from generally higher DMS reactivities both *in vivo* and *in vitro* (**Figure 2.2.3a**). Disregarding the incomplete demethylation of the common, natural tRNA modifications, we readily identify the D-loop, anticodon loop, T-loop, and the 3'CCA as hotspots of high τ scores. We subsequently used the τ scores from these regions to infer tRNA interacting with the aaRS, EF-1A, and the ribosome.

AaRSs recognize their cognate tRNA substrates through identity elements that usually involve many regions in tRNA in an idiosyncratic manner.^{39,40} To ensure the faithful coupling of anticodons and amino acids, 17 of the 20 aaRSs require anticodon recognition for charging, with the exceptions being seryl-, leucyl- and alanyl-tRNA synthetases (SerRS, LeuRS, and AlaRS) (**Figure 2.2.3.S3b**). SerRS and LeuRS interact with the unique extra-long variable arm of their cognate tRNAs, while AlaRS recognizes a G3:U70 base pair in the acceptor stem. (Tamaki et al 2018, Lenhard et al 1999, Shimada et al 2001) We predict that these two distinct recognition modes by aaRSs would lead to differential protein binding events in the anticodon loop. Indeed, the median anticodon loop τ scores of tRNA^{Ser}, tRNA^{Leu}, and tRNA^{Ala} as a group are significantly lower than τ scores of other tRNAs as a group, consistent with a lack of protein protection at the anticodon region for the tRNAs for these three amino acids (**Figure 2.2.3b**). Unexpectedly, tRNA^{Arg}(ACG) shows a much higher τ score for the anticodon loop than all other tRNA^{Arg}s (**Figure 2.2.3.S3c**). Human tRNA^{Arg}(ACG) has an unusual C13-A22 mismatch at the end of the D stem, whereas all other tRNA^{Arg}s have a conventional U13-A22 or C13-G22 base pair (Supplemental Figure S3d). The D loop sequences are identical in all tRNA^{Arg}s,²¹ which accommodate the U20 identity element for ArgRS recognition (Supplemental Figure S3e). The loosening of the D loop structure through the C13- A22 mismatch may increase the accessibility of U20 for more frequent interaction with ArgRS in vivo, which can, in part, explain the abnormally high τ score of tRNA^{Arg}(ACG).

EF1A delivers all aminoacylated tRNAs (aa-tRNAs) to the A-site of the elongating ribosome and forms extensive interactions with the T stem, the acceptor stem, and the 3'CCA of tRNAs.⁴⁸ Previous studies with acylated and misacylated prokaryotic tRNAs have established a

paradigm of EF1A-aa-tRNA interaction: EF1A has near-uniform binding affinities to all correctly aminoacylated aa-tRNAs which minimizes decoding bias and the loading of misacylated tRNAs on the ribosome. (Asahara et al 2002, LaRiviere et al 2001) This is accomplished by balancing the binding strengths to the esterified amino acids and the T stem sequences independently: tRNAs that bind to EF-Tu, the prokaryotic counterpart of EF1A, at high affinity alone have cognate amino acids with low binding affinity to EF1A and vice versa. The energetic contribution of each tRNA body sequence to EF-Tu binding was determined mostly using *in vitro* transcribed *E. coli* tRNAs and recombinant *T. thermophilus* EF-Tu (Ashara et al 2002); it remains to be determined whether this model also works for human EF1A-aa-tRNA interactions *in vivo*. As both the acceptor and T stem residues show very low DMS signals *in vivo*, we used the τ scores of the 3' C74C75A as a proxy to study EF1A binding affinity. A closer examination of the high-resolution structure of EF-1A-tRNA reveals that C75 and A76 interact directly with EF1A but not C74 (**Figure 2.2.3.S3f**). However, A76 is completely masked by EF1A, so only the C75 τ score serves as a good proxy for EF1A-tRNA binding *in vivo*. We plotted the average C75 τ score across isoacceptors for each amino acid with the estimated energy contribution by the tRNA body alone or by the amino acid alone. (Asahara et al 2002) The energetic contribution by the tRNA body and by the amino acid are negatively correlated for the established *E. coli* tRNA EF-Tu interaction, and lower energetic contribution indicates higher binding affinity. Intriguingly, between the amino acid average C75 τ score and energetic contributions (**Figure 2.2.3c**), we find a moderate positive correlation of tRNA (Pearson's $r = 0.64$) and a moderate negative correlation of the amino acid (Pearson's $r = -0.58$). This result suggests that the interaction strength at the tRNA 3' end also inversely correlates with the amino acid binding affinity and the tRNA body sequence binding affinity to EF1A. Since tRNAs are

generally highly charged *in vivo* and C75 is located closely to the charged amino acid, this result provides evidence for a similar EF1A aa-tRNA interaction paradigm in human cells as previously described for *E. coli* EF-Tu-aa-tRNA.

Next, we examined tRNA engagement with the ribosome using the τ scores of the tRNA “elbow” region where the D- and the T-loops form tertiary interactions in the L-shaped three-dimensional structure. Among the three major tRNA interacting partners in the cytoplasm, the tRNA elbow region interacts mostly with the ribosome (**Figure 2.2.3.S3g**).⁵⁴ The ribosome interacts with the elbow in all three of its tRNA binding sites to stabilize tRNA positioning, propel translation elongation, and enable proofreading. (Zhang et al 2016) At the A and E sites, the 50S subunit positions tRNA through its helix 38 (A-site Finger) and L1 stalk, respectively. At the P site, ribosomal protein L5 contacts the tertiary G19•C56 base pair of the tRNA elbow. Therefore, the tRNA D- and T-loop τ scores can be proxies to examine ribosome binding to aa-tRNAs. We first find a strong positive correlation between the D- and T-loop median τ scores of all most abundant tRNA isodecoders from each anticodon family (Pearson's $r = 0.95$ excluding tRNA^{Arg}(ACG), **Figure 2.2.3d**), which also validates stable D- and T-loop tertiary interactions *in vivo*. Recent findings suggest that the human ribosome decodes G/C-ending codons more efficiently than A/U-ending codons. (Watkins et al 2022, Zhang et al 2016) We reason that this decoding property may also be reflected in tRNA ribosome interactions. We therefore group tRNAs based on their wobble anticodon nucleotide identities: A/U-ending codons are more readily decoded by I34/U34 tRNAs (A34 in tRNA gene is always modified to Inosine in tRNA transcript)⁵⁶ and C/G-ending codons by G34/C34 tRNAs. Comparing the D loop average τ scores in tRNA groups with I/U or G/C at the wobble position, we find tRNAs with G34/C34 with higher τ scores, consistent with stronger interaction with the ribosome (**Figure 2.2.3e**).

tRNA^{Arg}(ACG) is once again an outlier for both D-/T-loop τ score correlation and ribosome interaction. It is possible that the C13-A22 mismatch in the D-stem of this tRNA also strengthens its interaction with the ribosome.

Even though C74 shows a wide range of τ scores (**Figure 2.2.3a**), we fail to identify a good correlation of it to anything specifically. One possibility is that because of its proximity to the interaction centers of tRNA, C74 is involved in interacting with all three partners in similar proportions.

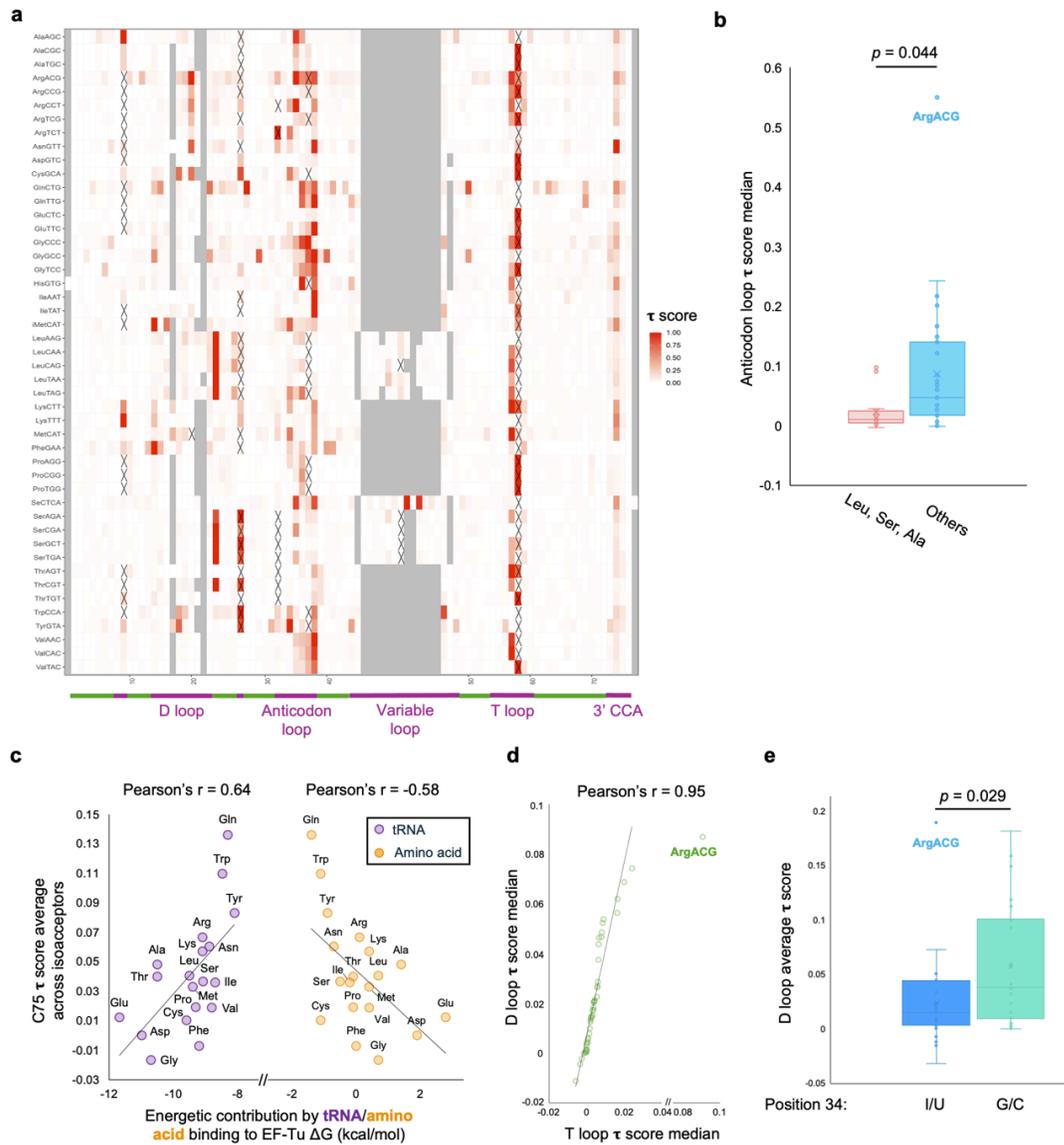


Figure 2.2.3 *In vivo* DMS signals are associated with tRNA interactions with cellular proteins and the ribosome.

(a) Heatmap showing the positional τ scores on the most abundant tRNA isodecoder in each anticodon family. Native modifications³⁰ that are incompletely removed by demethylases are marked with “X”. (b) Box-whisker plot showing anticodon loop τ score medians of tRNAs with two distinct modes of aaRS interactions. Charging of tRNA^{Leu}, tRNA^{Ser}, and tRNA^{Ala} does not require anticodon recognition, while others do. The p-value is calculated with a standard t-test. tRNA^{Arg}(ACG) is indicated as an outlier. (c) Scatter plot showing energetic contribution by tRNA binding or by amino acid binding to EFTu ΔG (kcal/mol) and τ scores at C75 averaged across all isoacceptors for each tRNA anticodon family. The x-axis scale was from Asahara et al. 2002 and 2005. (d) Scatter plot showing D and T loop τ score medians for the most abundant isodecoders in each anticodon family. Pearson's r is calculated excluding the outlier tRNA^{Arg}(ACG). (e) Box-whisker plot showing D loop τ score averages in the most abundant isodecoders grouped by the wobble nucleotide identity. The p-value is calculated with a standard t-test. tRNA^{Arg}(ACG) is indicated as an outlier.

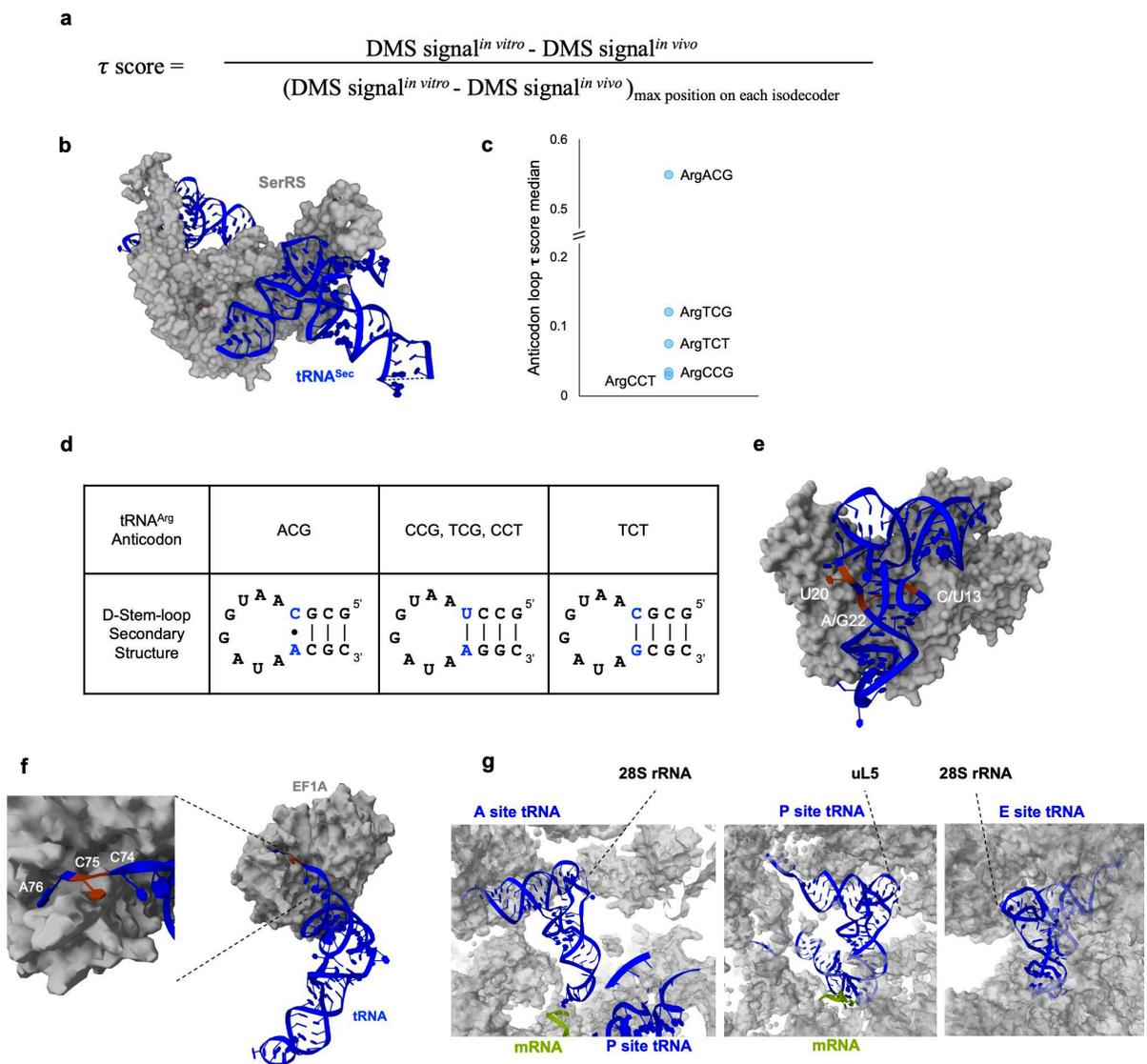


Figure 2.2.S3 Additional results for Fig 3.

(a) Definition of the τ score. (b) Human SerRS-tRNA^{Sec} structure showing the anticodon loop of the substrate tRNA is not bound by the SerRS (PDB: 4RQE). (c) Box-whisker plot showing anticodon loop τ score medians of 5 tRNA^{Arg} isodecoders ($n = 3$ biological replicates). (d) D-stem loop sequence and secondary structures of 5 tRNA^{Arg} isoacceptors. Base pair differences are shown in blue. (e) Structure of yeast ArgRS interaction with tRNA^{Arg} (PDB: 1F7V). (f) EF-1A-tRNA structure showing protection of C75 residue (PDB: 8G60). (g) Structures of A, P, and

E sites tRNAs in the human ribosome showing protection of the tRNA elbow region (PDB: 8G60 (A and P site), PDB: 6QZP (E site)).

2.2.4 Oxidative stress induces cytosolic tRNA structure and interactome changes.

In cells, global translation regulation is tightly linked to cellular metabolism and stress. As the critical components of the translation system, tRNAs undergo significant reprogramming in their expression, modification, charging, and fragmentation under stress. (Watkins et al 2022, Endres et al 2015, McEwen et al 2005) To understand the dynamics of cellular tRNA structure and interactions in response to stress, we treated cells with sodium arsenite to induce oxidative stress. Arsenite exposure induces strong phosphorylation of eIF2 α , which leads to significant downregulation of translation initiation.⁵⁹ Indeed, under our experimental settings, global translation activities were reduced by ~12-fold (**Figure 2.2.4.S4a**). Since arsenite stress inhibits translation initiation, we first examine the DMS signal changes of the initiator tRNA^{Met} (tRNA^{iMet}, **Figure 2.2.4a**). We use the term “ Δ DMS signal” to describe the DMS signal changes with and without stress (**Figure 2.2.4.S4b**). Consistent with a high reduction of translation initiation, which corresponds to decreased utilization of tRNA^{iMet} in the scanning ribosome, the DMS signal strongly increases for the 5' half of the anticodon loop of residues C32C33C34 in the presence of arsenite. DMS signal in the D-loop also increases, which is consistent with decreased ribosome interaction with tRNA^{iMet}. Intriguingly, tRNA^{iMe} exhibits a high *in vivo* signal of C71 (tRNA nomenclature; C70 in the tRNA^{iMet} position, **Figure 2.2.1f**) in the acceptor stem, which base pairs with G2 in its structure. This DMS signal is absent *in vitro*, suggesting that the C71-G2 base pair is weakened *in vivo*, potentially through its interaction with the eIF2B complex (**Figure 2.2.4.S4c**) or the 48S scanning ribosome. (Wang et al 2022) Arsenite exposure

leads to a strong reduction of the DMS signal at C71, consistent with a loss or altered interaction between tRNA^{iMet} with eIF2B or the scanning ribosome.

We examined the DMS signal changes for the most abundant isodecoders from all anticodon families (**Figure 2.2.4b**). Arsenite stress induces widespread changes in cytosolic tRNA structures and/or protein/ribosome interactions. In general, the loop regions have the most DMS signal changes under stress, but changes in the acceptor stem are also detectable, and the direction of the changes is not consistent across tRNAs. To provide a better focus, we zoom into the anticodon loop region by grouping the tRNAs according to their modification status at position 37 (**Figure 2.2.4c**), the immediate 3' nucleotide to the anticodon nucleotides: group 1 are A37 and unmodified, group 2 are A37 and modified to N6-threonylcarbamoyl adenosine (t6A) or N6-isopentenyl adenosine (i6A), and group 3 are G37 that are modified to m1G, m1I or wybutosine (W). (Pan 2018, Suzuki et al 2011) Intriguingly, group 1 tRNAs without A37 modifications generally have positive Δ DMS signals, i.e., higher DMS signals in the presence of arsenite, whereas group 2 tRNAs with A37 modifications have zero Δ DMS signals on average, and group 3 tRNAs with G37 modifications have negative Δ DMS signals on average. This result is consistent with tRNA modifications at position 37 positively contribute to ribosome usage of tRNA, and thus, tRNAs that lack modification at position 37 are the first to be affected by arsenite stress.

Cytosolic tRNA translational usage was profiled under the same arsenite stress in HEK293T cells with polysome-enriched tRNA sequencing in our previous study. (Watkins et al 2022) We reanalyzed the data and identified three significantly enriched and five significantly depleted tRNAs in polysome with or without arsenite treatment (**Figure 2.2.4.S4d**). The three polysome enriched tRNAs, tRNA^{Ile}(AAU), tRNA^{Ile}(UAU), and tRNA^{Lys}(UUU), have been

described in our previous study with Northern blot validation. (Watkins et al 2022) We analyzed the Δ DMS signal by comparing the group of significantly polysome-enriched tRNAs and the group of significantly polysome-depleted tRNAs by >1.5-fold (**Figure 2.2.4d**). For the anticodon, D loop, and variable loop, the average Δ DMS signal is negative for the polysome-enriched tRNAs but positive for the polysome-depleted tRNAs, consistent with the persistent polysome association of the enriched tRNAs under arsenite stress. The Δ DMS signal of the T loop is spread widely without a significant difference between the polysome-enriched and -depleted tRNAs. Intriguingly, the average Δ DMS signal in the acceptor stem is positive for polysome-enriched tRNAs but neutral for the depleted tRNAs. This unexpected result may be explained by the polysome-enriched tRNAs under arsenite stress not fully engaged in protein synthesis, such as not forming the peptide bond, even though they are still bound by the ribosome. This altered accessibility of the acceptor stem in the ribosome may contribute to translational repression and reprogramming during arsenite stress.

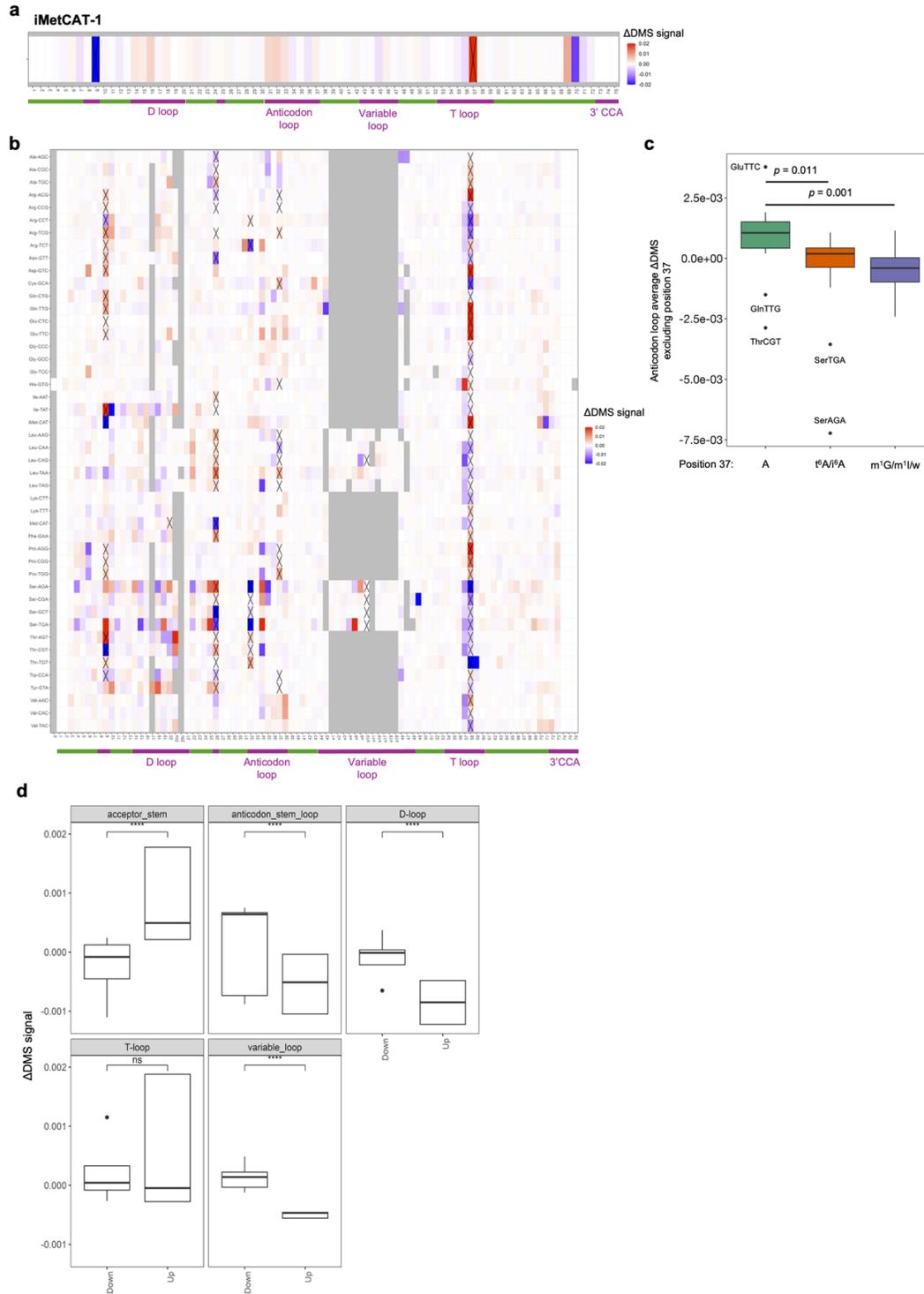


Figure 2.2.4 Cytosolic tRNA structure and interactions in response to arsenite stress

(a) Heatmap of the positional Δ DMS signal on tRNA^{iMet}. Native 784 modifications that are incompletely removed by demethylases are marked with “X”. (b) Heatmap of the positional

Δ DMS signal of the most abundant isodecoder in each anticodon family. Native modifications that are incompletely removed by demethylases are marked with “X”. (c) Anticodon loop (excluding position 37) average Δ DMS signals in tRNA groups with different modification status at position 37. P-values are calculated with Wilcoxon signed-rank tests. (d) The Δ DMS signals in D, T, anticodon, variable loops, and the acceptor stem in tRNAs that are enriched (up) or depleted (down) in polysomes. P-values are calculated with standard Wilcox t-test; n.s.: not significant, ****: $p < 10^{-4}$.

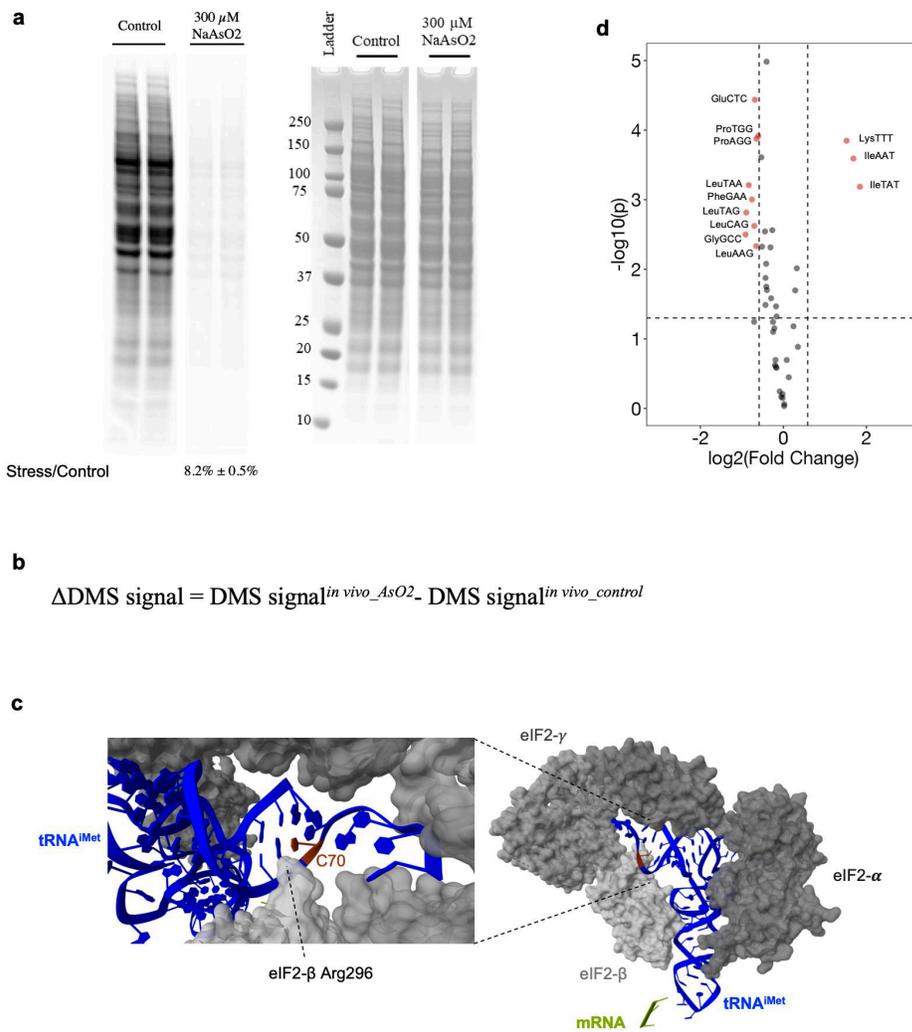


Figure 2.2.S4 Additional results for Figure 4

(a) ^{35}S -Met pulse chase showing a strong reduction of translation upon arsenite stress. Loading controls are on the right. (b) Formula for ΔDMS signal. (c) eLF2b-tRNA^{iMet}-tRNA ternary complex structure showing eLF2- β Arg296 residue in proximity to C70 (C71 in tRNA nomenclature) of tRNA^{iMet} (PDB: 6YBV). (d) Polysome cytosolic tRNA changes under stress (AsO₂ polysome/control polysome). Vertical lines indicate a 1.5-fold change; the horizontal line indicates a $p = 0.05$. Data from reference (Watkins et al 2022) and NCBI GEO GSE198441.

2.3 METHODS AND MATERIALS

2.3.1 Cell culture and Arsenite treatment

HEK 293T cells (ATCC, CRL-11268) were cultured with complete DMEM medium under standard conditions according to ATCC. Briefly, HEK 293T cells were grown in Hyclone DMEM medium (GE Healthcare Life Sciences, SH30022.01) with 10% FBS and 1% Pen–Strep (Penicillin–Streptomycin) to 80% confluency and passaged. Cells were collected, and total RNA was extracted using TRIzol (Thermo Fisher, 15596026) by following the manufacturer's protocol when cells reached 80–90% confluency. Arsenite stress was induced by incubating cells in media containing 300 μM NaAsO₂ at 37 °C for 2 h.

2.3.2 *In vivo* DMS treatment of HEK 293T cells

To a 10-cm dish of HEK 293T cells cultured in 15 mL media, 300 μL or 750 μL of DMS were added to reach a final concentration of 2% or 5%, respectively. After the addition of DMS, cells were incubated at 37 °C for 5 min. The media was then aspirated off, and the cells were washed with PBS twice and harvested for total RNA extraction with TRIzol (Thermo Fisher, 15596026) following the manufacturer's protocol.

2.3.3 In vitro DMS treatment of HEK 293T total RNA

Total RNA was extracted from 80-90% confluent HEK 293T cells with TRIzol (Thermo Fisher, 15596026) following the manufacturer's protocol. 4 µg of HEK 293T total RNA in 10 µL H₂O was combined with 10 µL of the 2x refolding buffer (200 mM NaCl, 12 mM MgCl₂, 20 mM Tris, pH 8.0) and incubated at 30 °C for 30 minutes. DMS was added to each sample to reach a final concentration of 5% at room temperature for 5 minutes. DMS treatment was stopped by adding an equal volume of 2x DMS quench buffer (60% β-mercaptoethanol, 605 0.6 M NaOAc, pH 5.5). DMS-treated total RNA samples were precipitated with ethanol.

2.3.4 Whole-cell translation assay by 35S pulse labeling

HEK293T cells were cultured in a CO₂ incubator at 37 °C with complete DMEM (Cytiva, SH30022.01) supplemented with 10% fetal bovine serum (Thermo, 16140071) and 1% Pen/Strep (Thermo, 15070063). For whole-cell translation assay, 0.3×10⁶ HEK293T cells were seeded in each well of a 6-well plate. Cells were incubated in a CO₂ incubator at 37 °C for 2 days. Cells were treated with either sterile H₂O or 300 µM NaAsO₂ (Sigma, 1062771000) for 2 hours in a CO₂ incubator at 37 °C. Labeling medium without methionine were prepared by supplementing DMEM medium without methionine and cysteine (Thermo, 21013024) with 10% fetal bovine serum, 4 mM glutamine (25030081, Thermo), 0.063 g/L (0.2 mM) L-Cystine·2HCl (Sigma, C2526-100G). The labeling medium and normal complete medium were warmed in a cell incubator for >30 min. After NaAsO₂ treatment, the culture medium was removed, and the cells were washed with warmed labeling media twice. 3 ml labeling medium was then added to each well, and cells were incubated in a CO₂ incubator at 37 °C for 30 min. 62.5 µCi (1 mCi/16)

³⁵S-labeled methionine (PerkinElmer, NEG709A001MC) was added into each well and incubated for 60 min. The labeling mixture was then removed from the plate and disposed of into radioactive waste. 3 ml of warmed normal medium was added into the plates and incubated for 10 min. The medium was then removed from the plate, and cells were washed with ice-cold PBS three times. Cells were then harvested with cell lifter, and cell lysate was then extracted with CelLytic M lysis buffer (Sigma, C2978) supplemented with 1× protease inhibitor (Roche, 11873580001) and 25-250 U/ ml Benzonase (Sigma, E1014) and rotate for 2h or overnight at 4 °C. Equal amount of cell lysate from each sample was then loaded onto SDS-PAGE gel for protein separation. Gel was first stained with Coomassie blue to visualize the total protein loading amount. The gel was then dried using a gel dryer and exposed to a phosphorimaging plate for scanning and analysis.

2.3.5 tRNA sequencing with MSR-seq

Approximately 1 µg of in vivo or in vitro DMS treated-total RNA samples were used to build tRNA sequencing libraries, following a previously published MSR-seq protocol.²⁴ Briefly, RNAs were first deacylated by incubating at 37 °C in a 33 mM sodium tetraborate buffer for 30 minutes. Next, 5 µL of a PNK reaction stock (4 U/µL NEB T4 PNK, 40 mM MgCl₂, 200 mM Tris-HCl, pH 6.8) was added to repair the 3' end, and the sample was incubated at 37 °C for 20 minutes and then incubated at 65 °C for 10 minutes to inactivate the PNK enzyme. Following 3' end repair, 30 µL of an RNA ligation reaction mix was added to the sample while still in the 3' end repair mix (final concentration: 15% PEG 8000, 1x T4 RNA ligase I buffer, 50 µM ATP, 5% DMSO, 1 mM hexaamide cobalt (III) chloride, and 1 U/µL T4 RNA ligase I) and incubated overnight at 16 °C. This mix also included the barcoded RNA ligation linker/RT primer oligo at a

1.2:1 molar ratio to the input RNA. After ligating the RNA overnight, the sample was bound to streptavidin-coated MyOne C1 dynabeads (ThermoFisher) at room temperature on rotation for 15 minutes to facilitate the library construction process by minimizing sample loss and enabling rapid washes and buffer exchanges between reactions. “+DM” samples were treated according to the established demethylase protocol below. “+DM” and “-DM” samples were then dephosphorylated using a CIP reaction mix (0.2 U/ μ L CIP, 10 mM MgCl₂, 0.5 mM ZnCl₂, 20 mM HEPES, pH 7.5) for 30 minutes at 37 °C to liberate the 3' end of the RT primer. Next, samples were resuspended in 25 μ L of 1x SueprScript IV VILO mix and incubated at 55 °C for 10 minutes and then at 37 °C overnight. Next, samples were treated with an RNase H reaction mix (0.4 U/ μ L 651 NEB RNaseH, 1x RNase H buffer) for 15 minutes at 37 °C. Subsequently, samples were treated with a solution of 50 mM sodium periodate in 15 mM sodium acetate, pH 5.0, for 30 minutes at room temperature to oxidize any non-extended RT primer. Following this, an overnight, room temperature cDNA ligation step was performed to enable PCR amplification of the cDNA (50 μ L reaction; final concentrations: 2 U/ μ L T4 RNA ligase I, 25% PEG 8000, 7.5% DMSO, 50 μ M ATP, 1 mM hexamine cobalt (III) chloride, 2 mM barcoded cDNA ligation oligo, 1x T4 RNA ligase I buffer). Finally, the libraries were amplified by PCR with Illumina primers. Sequencing was conducted using Illumina NovaX 6000, 100-bp paired-end. AlkB demethylase treatment Demethylase buffer conditions were modified from published protocols.^{24,75} Three stock solutions were made fresh immediately before the reaction: l-ascorbic acid 200 mM, 2- ketoglutarate 3 mM, and ammonium iron sulfate 5 mM. The final reaction mixture contained 2 mM l-ascorbic acid, 1 mM 2-ketoglutarate, 0.3 mM ammonium iron sulfate, 100 mM KCl, 50 mM MES pH 6, 50 ng/ μ L BSA, 4 μ M wild-type AlkB, and 4 μ M AlkB-D135S (purified as described previously (Watkins et al 2022)). About 50 μ L of the reaction

mixture was added to 5–20 μ L of decanted streptavidin bead slurry after ligation, immobilization, and washing. The reaction continued for 30 min at 37 °C. Following the reaction, beads were washed once with high salt wash buffer (1 M NaCl, 20 mM Tris-HCl pH 7.4) and once with low salt wash buffer (100 mM NaCl, 20 mM Tris-HCl pH 7.4). Overexpression plasmids for recombinant E. coli AlkB and AlkB-D135S are available from Addgene (https://www.addgene.org/Tao_Pan/).

2.3.6 Data analysis of DM-DMS-MapSeq Data

The data analysis followed the MSR-seq data processing pipeline. Libraries were sequenced from Illumina NovaSeq platforms as previously described. The resulting paired-end reads were demultiplexed by the identification of barcode sequences using Je demultiplex with the following parameters: BPOS=BOTH BM=READ_2 LEN=6:4 FORCE=true C=false.⁷⁶ These options were optimized for samples where the barcode sequence is present on read 2. Barcode sequences were previously described.²⁴ Following demultiplexing, data was aligned using bowtie2 (version 2.3.3.1) with the following parameters: -q -p 10 --local --no-unal.⁷⁷ These reads were aligned to a curated human tRNA reference. This reference contained sequences of HG38 versions of 5S and 5.8S rRNA, snoRNA, Y tRNA, and tRNA genes that were curated for non-redundancy, with a tRNA-scan SE score > 47, removing intron sequences, and 3' "CCA" appended. The Bowtie2 output sam files were converted to bam files, which were then sorted by samtools' sort function.⁷⁸ IGVtools count was used to collapse reads into 1nt windows using the following parameters: -z 5 -w 1 -e 250 —bases. The resulting IGV output wig files were reformatted using a custom Python script to obtain mutation rate and read coverage

compatible with R for data visualization and analysis. To obtain canonical position information for all tRNA genes, the paired reads were processed by mim-tRNAseq⁷⁹ with the following parameters: --species Hsap --cluster-id 0.95 --threads 4 --min-cov 0.0005 --max-mismatches 0.1 --control-condition --no-cca analysis -n hek_mods_correl. The resulting mismatch table file contained deconvoluted tRNA gene sequences with canonical tRNA position information. For tRNA genes that could not be deconvoluted, canonical positioning was done by hand following tRNA nomenclature guidelines as described in Sprinzel et al. 1987. For data visualization, a ≥ 600 read coverage filter was used to ensure only positions with sufficient coverage. Following this, the mutation rates were used for analysis as previously described above for DMS_{diff}, DMS signal, and τ score.

2.3.7 Data and Code availability

The sequencing data has been deposited to NCBI GEO accession number GSE262888. The MSR-seq analysis code, including mutation rate analysis with and without demethylase treatment, is available on GitHub (<https://github.com/ckatanski/CHRIS712seq>).

2.4 DISCUSSION

The rapid development of high-throughput sequencing methods for highly structured RNAs made it possible to examine transcriptome-wide tRNA structures and interactions in living cells. With DM-DMS-MaPseq, we present comprehensive *in vivo* structure and interaction studies for human chromosomal and mitochondrial encoded tRNAs (will be discussed in Chapter 3). Compared to DMS mapping of mRNAs, *in vivo* DMS signals for tRNA are much lower;

therefore, additional considerations need to be implemented to ensure that the observed mutations are indeed from DMS reactions, not spurious reverse transcriptase events or other processes. This was accomplished here by treating the DMS-reacted RNA samples with a m¹A/m³C demethylase before RT and considering only the mutation differences at nucleotides with and without DM treatment. This strategy ensures a more accurate interpretation of the DMS mapping results.

Our result is consistent with human chromosomal-encoded tRNAs folding into canonical tRNA structures *in vivo*, similar to the results in an *E. coli* tRNA study *in vivo*. (Yamagami et al 2022) In contrast to *E. coli*, the human genome contains many tRNA genes that have the same anticodon, but different body sequences termed isodecoders. This exceptional sequence diversity of human tRNAs offers an opportunity to compare the folding patterns of tRNA isodecoders that all have sufficiently high tRNAScan scores to predict to fold into canonical tRNA structures but possess distinct base pairing or tertiary interaction patterns. A brief inspection of all tRNAScan predicted “high confidence” tRNA isodecoders present at sufficiently high abundance in our sample did not identify a tRNA that may fold into a non-canonical structure *in vivo*. We, therefore, compared more subtle but still obvious differences between pairs of tRNA isodecoders to evaluate how sequence diversity between isodecoders can influence tRNA structure. Our results show that a merely 3 nucleotide difference between tRNA^{Ser(AGA)}-1 and -4 can generate local structural variations in the proximity of the sequence changes. In both tRNA isodecoder pairs examined here, the lower tRNAScan scoring tRNA shows more DMS reactivity than the higher scoring one. This result provides experimental evidence of the tRNAScan score as a valuable tool for predicting local structural variations.

The most striking difference we observe is between the *in vitro* and *in vivo* DMS signatures, where tRNAs *in vitro* generally always have significantly higher DMS reactivities than *in vivo*. In contrast to the *E. coli* study, where the *in vitro* results were from unmodified tRNA transcripts, our *in vitro* samples were total RNA isolates from cells that contained the same set of tRNA modifications as *in vivo* samples. The most plausible explanation is that cellular protein/ribosome interactions with tRNA contribute greatly to the *in vivo* DMS footprints. The idea of using the DMS reactivity differences between *in vitro* and *in vivo* to study RNA-protein interactions has been proposed and implemented for well-characterized interactions between ribosomal proteins and rRNAs, as well as RNA binding protein RBFOX2 with its mRNA substrate. (Spitale et al 2015) Transcriptome-wide mRNA-protein interaction studies can be very challenging due to the low coverage of individual structural motifs in mRNAs and the transient nature of these interactions. Therefore, cytosolic tRNAs can be a promising model system to study RNA-protein interactions using the DMS mapping data. Indeed, given the abundant prior knowledge of tRNA-protein and tRNA-ribosome interactions, we were able to approximate each of the three most predominant tRNA-protein/ribosome interactions to aaRS, EF1A, and the ribosome semi-quantitatively by extracting DMS probing information from distinct sets of structural elements in tRNA. Our analysis of tRNA anticodon τ scores conforms with the known two modes of aaRS substrate recognition and provides a plausible explanation for the abnormally high τ score of tRNA^{Arg}(ACG). Our analysis of tRNA C75 τ scores supports the same EF1A-tRNA uniform binding paradigm originally identified from *in vitro* prokaryotic EF-Tu-tRNA studies. Our analysis of tRNA D- and T-loop τ scores supports the model of stronger tRNA-mRNA interactions on the ribosome for C/G ending codons than those for A/U-ending codons.

To examine the tRNA structure and protein/ribosome-interaction dynamics in response to

environmental changes, we applied arsenite treatment to induce oxidative stress in cells and compared the *in vivo* DM-DMS-MaPseq results with and without stress. Arsenite stress induces large-scale changes in *in vivo* DMS footprints that can be interpreted in the context of severely reduced global translation activity and detailed translational reprogramming. The tRNA^{iMet} DMS signal changes are consistent with its decreased engagement in translation initiation. However, we also identify an unexpectedly strong DMS signal at C71 without stress, and this signal is markedly reduced with stress. C71 forms the second base pair in the acceptor stem of tRNA^{iMet}, so this result may implicate the melting of two base pairs in the eIF2B-tRNA^{iMet}-40S or other higher-order complexes in its active form. tRNA^{iMet} has an A1-U72 base pair instead of the most common G1- C72 base pair in elongator tRNAs, which may facilitate this melting at the end of the acceptor stem. We also find that the modification status at position 37, the nucleotide immediately 3' to the anticodon, determines the DMS sensitivity of tRNA towards stress. tRNAs with unmodified 37 bases have higher DMS reactivity for the anticodon loop residues in stress, whereas tRNAs with modified 37 bases have equal or lower DMS reactivity in stress. This result implicates a distinct modification-dependent translational response to arsenite stress for subgroups of tRNA.

We revisited our previously published data on polysome-enriched and -depleted tRNAs with and without arsenite stress to gain additional insights into the translational response of different tRNA groups. Perhaps not surprisingly, the polysome-enriched tRNA group has lower DMS reactivity in the interior tRNA regions compared to the polysome-depleted tRNA group under arsenite stress. To our surprise, this result is reversed for the acceptor stem, where the polysome-associated tRNAs show higher DMS reactivities under stress. As described previously, the polysome-enriched tRNAs are likely located in stalled ribosomes under stress. Ribosome

stalling may be derived from multiple factors, including EF-1A not releasing tRNA, slowed peptide formation, slowed translocation, or tRNA not leaving the E site. It remains to be seen which one of these processes plays a more dominant role in the explanation of the acceptor stem result.

In summary, we present the DM-DMS-MaPseq method to investigate in vivo structure and interactions of tRNA. In addition to tRNAs, DM-DMS-MaPseq also captures reads for other abundant and highly structured non-coding RNAs, such as Y RNA and snRNAs in their cellular context. Similar analyses can be conducted to understand the structural dynamics of these RNAs and their associated RNPs. In the future, a simple add-on procedure would be to first fractionate different RNPs from DMS-treated cells before sequencing library construction. This would highly enrich the in vivo DMS footprints in specific cellular components such as the monosome, polysome, snRNPs, and larger snRNP complexes, which will facilitate more in-depth and accurate interpretation and reveal new biological insights about the cellular dynamics of these RNPs.

CHAPTER 3

IN VIVO STRUCTURE PROFILING OF HUMAN MITOCHONDRIAL tRNA STRUCTROME AND INTERACTOME RESPONSE TO STRESS

Acknowledgment: This chapter is derived from an article that has been submitted for publication (Peña, Hou, Watkins, et al). The authors of this article were Noah Peña, Yichen Hou, Christopher P. Watkins, Sihao Huang, Wen Zhang, Christopher D. Katanski, and Tao Pan. N.P., C.P.W., T.P. conceived the project. N.P., Y.H., T.P. analyzed and interpreted data, with help from S.H., C.D.K. C.P.W. performed the DMS experiments and built sequencing libraries, with help from W.Z. Y.H., N.P., C.P.W., T.P. wrote the paper.

3.1 INTRODUCTION

In human cells, mitochondria have their own translation system that utilizes an exclusive set of 22 tRNA genes encoded by the mitochondrial DNA. Mitochondrial tRNAs (mt-tRNAs) have non-canonical tertiary interactions and lower melting temperatures *in vitro* due to higher numbers of A-U base pairs in the stems. Mitochondrial tRNA interacts with mitochondrial ribosome and elongation factors, making it a prime candidate to study the mitochondrial translation system under stress conditions. However, a comprehensive *in vivo* structural analysis of human mitochondrial tRNA is lacking. Therefore, we propose to utilize DM-DMS-MaPseq to characterize the *in vivo* mitochondrial tRNA structurome and interactome. Using DM-DMS-MaPseq, we found that for mt-tRNAs with non-canonical base-pairing and tertiary interactions, the *in vivo* environment seemed to be crucial for stabilizing their structures. We derived tRNA-protein interaction insights by comparing *in vivo* and *in vitro* DMS signals.

3.2 RESULTS

3.2.1 Structural mapping of mitochondrial tRNAs *in vivo*

To examine the folding and interaction properties of mt-tRNAs, we first plot the positional DMS signals for all 22 mt-tRNAs both *in vitro* and *in vivo* (**Figure 3.2.15a**). Consistent with their weaker structures than cytosolic tRNAs, mt-tRNAs showed high DMS signals throughout the tRNA bodies *in vitro*, including the stem regions. However, the *in vivo* DMS signals for mt-tRNAs are still much lower than those *in vitro*, reminiscent of the *in vitro* and *in vivo* differences seen for cytosolic tRNAs, consistent with mitochondrial protein and ribosome interactions making a major contribution to the stability of mt-tRNA structures in the cellular environment.

To examine the structure data in more detail, we overlaid the *in vivo* and *in vitro* DMS signals to the predicted secondary structures of all mt-tRNAs. The folding of mt-tRNAs is categorized into three types. (Suzuki et al 2011) Type I secondary structure is represented by mt-tRNA^{Ser}(UGA), which has only one nucleotide between the acceptor stem and the D stem, a small D loop, and an extended anticodon stem. Most other mt-tRNAs fold into type II structures. They lack the conserved interaction between G18G19 in the D-loop and U55C56 in the T loop, which maintains the stable “L-shape” tertiary structure of cytosolic tRNAs. Type III structure is unique for mt tRNA^{Ser}(GCU), which misses the entire D stem. 63,66 Mt-tRNA^{Ser}(UGA) and most type II mt tRNAs roughly resemble the reference structure but display various structural nuances in at least one of the stems (**Figure 3.2.S1a-b**). Mt-tRNA^{Leu}(UAA) is one of the most stable mt-tRNAs, and the *in vivo* DMS signals confirm its mostly stable stems. Despite that, we see DMS signals for the unpaired A12:C23 at the D stem and A31:C39 at the anticodon stem *in vitro*

(**Figure 3.2.1b**). As expected, type III mt-tRNA^{Ser}(GCU) is the most unstable among all mt tRNAs *in vitro* (**Figure 3.2.1c**). However, its structure is largely stabilized *in vivo*, as shown by the reduced DMS reactive nucleotides in all three stems.

The mt-tRNA^{Lys}(UUU) has been shown to misfold in the absence of the m¹A9 modification *in vitro*. (Helm et al 2004) In the presence of m¹A9, mt-tRNA^{Lys} folding is consistent with its predicted cloverleaf secondary structure *in vitro* and *in vivo* (**Figure 3.2.1d**). Mt-tRNA^{Val} is unique among all mt-tRNAs because it is also a component of the mitochondrial ribosome in place of the 5S rRNA in the cytosolic ribosome. (Brown et al 2014, Chrzanowska-Lightowlers et al 2007) The DMS signals for mt-tRNA^{Val} show strong signatures throughout this tRNA *in vitro* but were largely reduced and more restricted in the D and anticodon loops *in vivo* (**Figure 3.2.1e**). Even the D and anticodon loop regions interact extensively with the components of the mitochondrial ribosome (Figure 5f), suggesting that its ribosome interaction, at least in part, explains its reduced DMS signals *in vivo*.

We further examined mt-tRNA interactions by examining the positional τ scores, i.e., the normalized *in vitro-in vivo* differences for each mt-tRNA (**Figure 3.2.S1c**). The mt-tRNA τ score interpretation is more complex than cytosolic tRNAs through its inclusion of both folding and interaction components. Focusing on the C75 τ scores, which are strongly associated with mt-tRNA interaction with mitochondrial EF-Tu protein (**Figure 3.2.S1d**), we still find an inverse correlation between tRNA-EF-Tu and amino acid-EF-Tu interactions (Pearson's r-values of -0.64 and 0.62, Figure 5g), suggesting that the principle of balancing tRNA and charged amino acid binding strength to EF-Tu still holds for human mitochondrial translation to maintain translation fidelity.

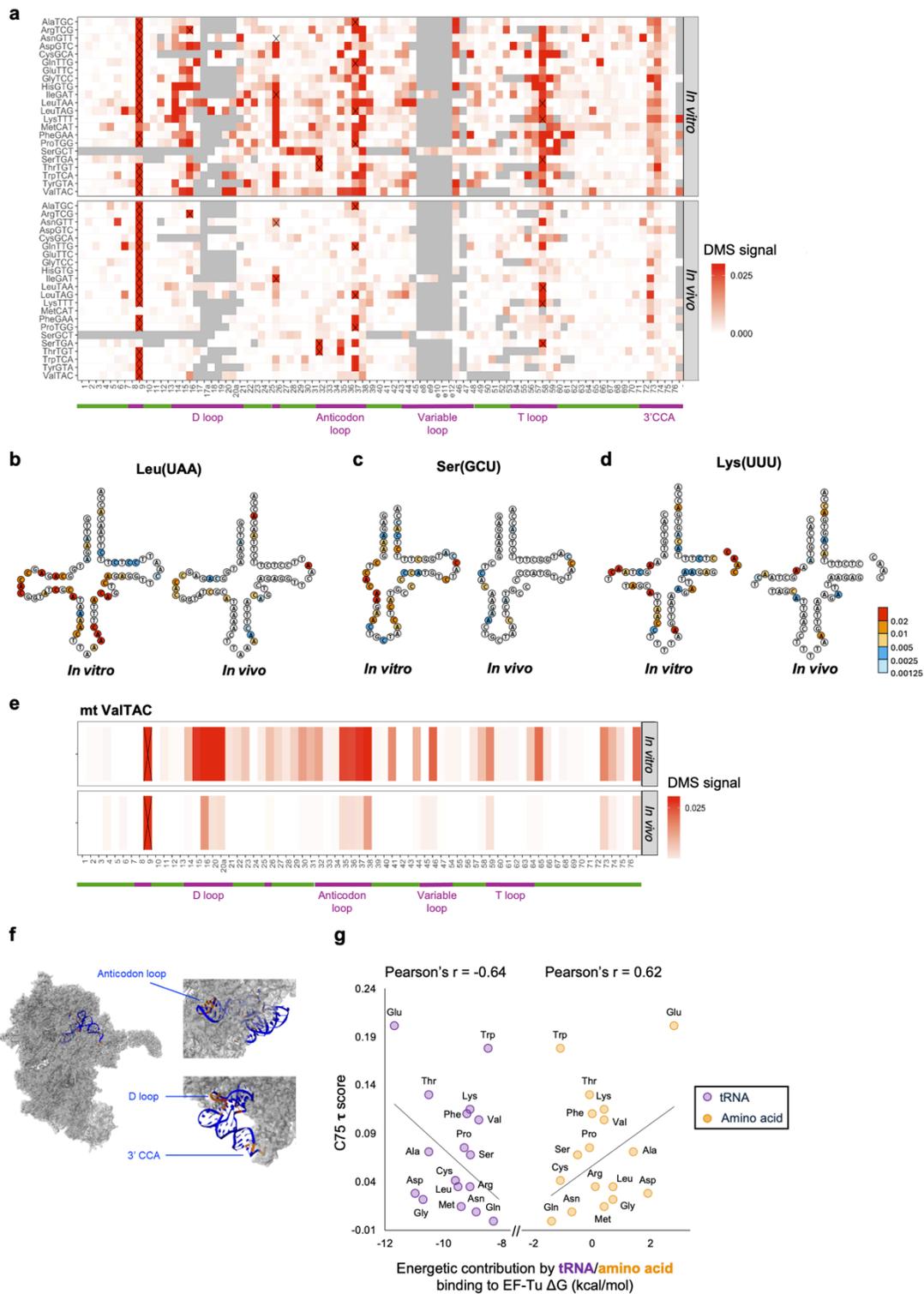
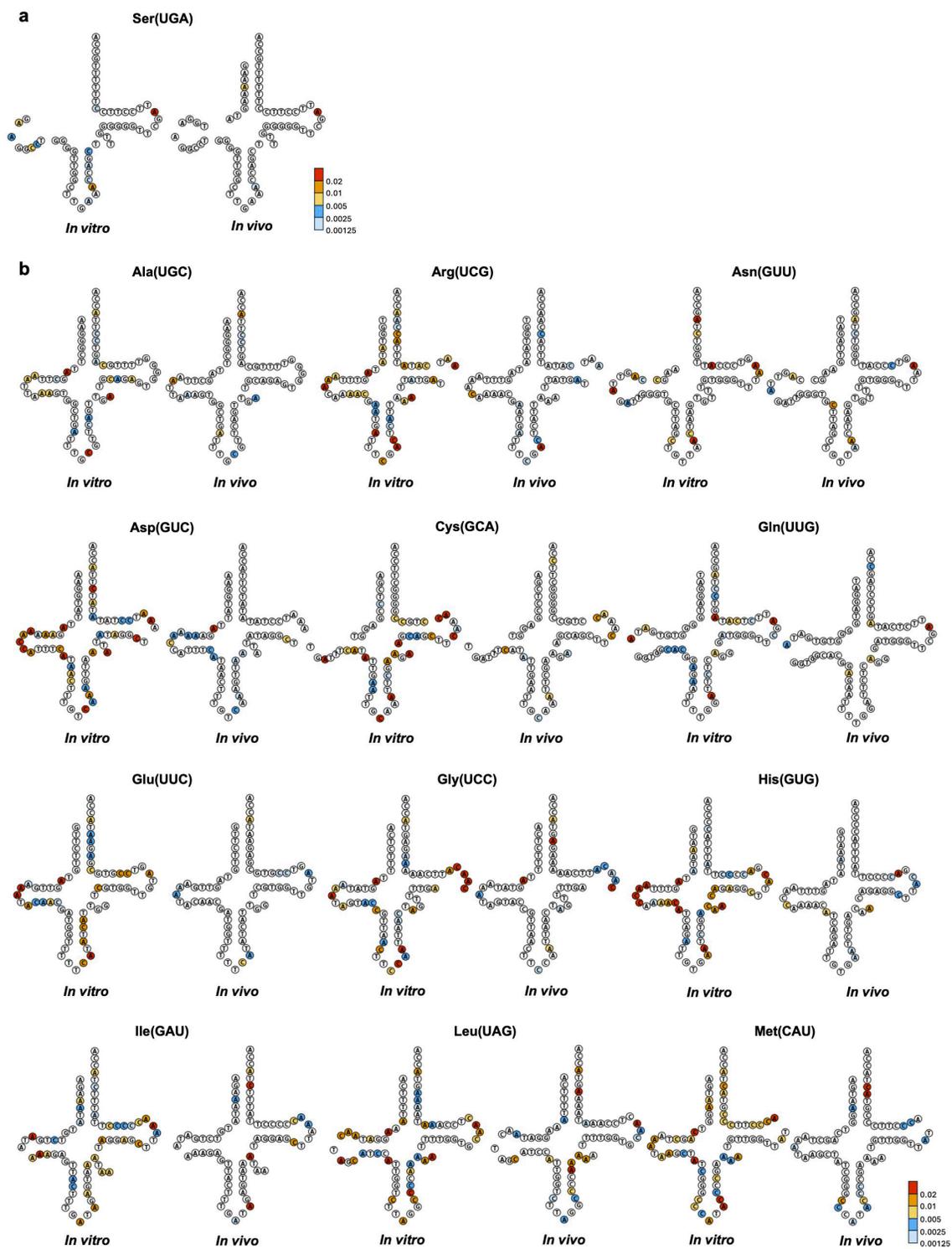


Figure 3.2.1 DMS mapping of mitochondrial-encoded tRNAs reveals in vitro versus in vivo differences and in vivo interaction. (a) Heatmap showing the in vitro and in vivo positional 798

DMS signals on the mitochondrial (mt) tRNAs. Native modifications that are incompletely removed by demethylases are marked with “X”. (b) The secondary structure of mt-tRNA^{Leu}(UAA) is overlaid with in vitro and in vivo DMS signals. (c) The secondary structure of mt-tRNA^{Ser}(GCU) is overlaid with in vitro and in vivo DMS signals. (d) The secondary structure of mt-tRNA^{Lys}(UUU) is overlaid with in vitro and in vivo DMS signals. (e) Heatmap showing the in vitro and in vivo positional DMS signals on the mt-tRNA^{Val}. Native modifications³⁰ that are incompletely removed by demethylases are marked with “X”. (f) The Cryo-EM structure of mt-tRNA^{Val} in human mt-ribosome (PDB: 6ZM683) overlaid with DMS signals of high (red), medium (orange), and low (yellow). (g) Scatter plot showing energetic contribution by mt-tRNA binding or by amino acid binding to EF-Tu ΔG (kcal/mol) and τ scores at C75 averaged across all isoacceptors for each tRNA anticodon family. The x-axis scale was from Asahara et al. 2002 and 2005.



(Continue to the next page)

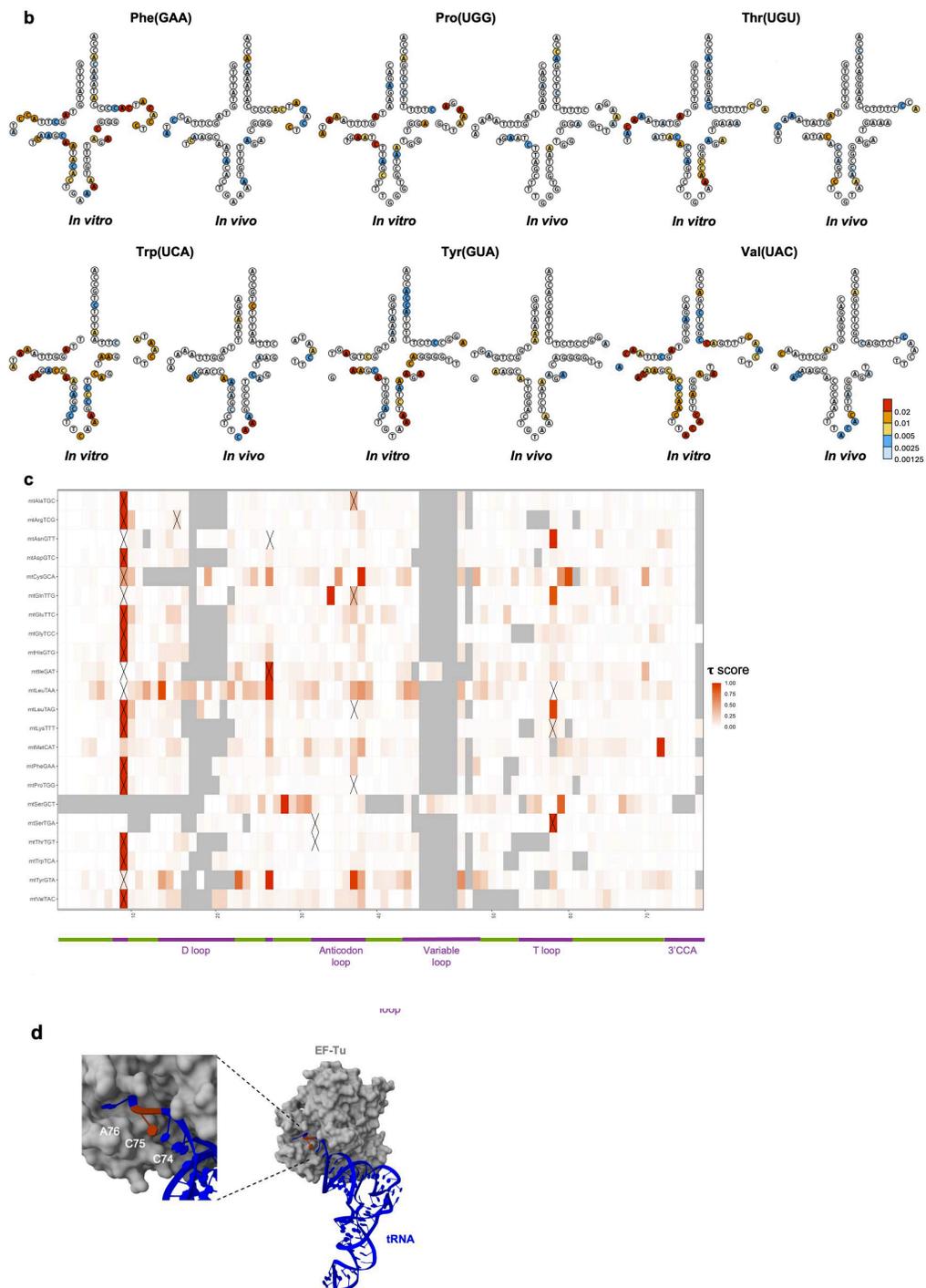


Figure 3.2.S1 Secondary structures of mt tRNA, and interactions analysis of mt tRNA. (a)

The secondary structure of mt-tRNA^{Ser}(UGA) overlaid with *in vitro* and *in vivo* DMS signals.

(b) The secondary structures of 16 type II mt-tRNAs overlaid with *in vitro* and *in vivo* DMS signals. (c) A heatmap of the positional τ score on each mt-tRNA. Native modifications that are incompletely removed by demethylases are marked with "X". (d) EF-Tu-mt-tRNA structure showing C75 interaction (PDB: 1TTT)

3.2.2 Arsenite stress induces mt-tRNA structural changes

Arsenite stress is known to significantly reduce mitochondrial activities, including translation. (Huber et al 2022) Indeed, mt-tRNA^{Met}, which works in both mitochondrial translational initiation and elongation, shows a large increase in DMS signal in the anticodon loop upon arsenite treatment, consistent with its decreased participation in mitochondrial translation (**Figure 3.2.2a**). Analysis of previously published data under the same arsenite stress condition shows a reduction in mt tRNA^{Met} charging as well (**Figure 3.2.S2a**), which may account for its loss of multiple types of protein interactions *in vivo* under arsenite stress. We also observe a reduction of the DMS signal for the equivalent of C71A70 residues in mt-tRNA^{Met}, reminiscent of the C71 DMS signal reduction of cytosolic tRNA^{iMet}. It remains to be seen whether this acceptor stem response to arsenite stress may also relate to mt-tRNA^{Met} structural change and its contribution to the reduction of mitochondrial translation under stress. Arsenite stress also induces universal and notable alterations for all mt-tRNAs (**Figure 3.2.2b**). Overall, mt-tRNAs show far more drastic changes throughout the tRNA bodies, whereas cytosolic tRNA structural changes are more concentrated in loop regions. This likely reflects a combined effect of lower structure stability of mt-tRNAs and a more acute loss of their interactions with protein/ribosome interaction under arsenite stress. Both effects may exacerbate the reduction of mitochondrial translation.

We also examined the polysome association of mitochondrial tRNAs with and without arsenite stress from our previous publication. (Watkins et al 2022) Strikingly, mt-tRNA^{Asp} shows the highest enrichment in the polysome fraction, followed by several other mt-tRNAs (**Figure 3.2.2c**). Asp accounts for only 1.85% of all codons in human mitochondrial proteins; therefore, the mt tRNA^{Asp} enrichment in the polysome fraction is unlikely to be attributed to Asp codon usage in mitochondrial translation. Mt-tRNA^{Asp} shows an increased DMS signal in the D-loop and anticodon loop and a reduced DMS signal in the acceptor stem under arsenite stress (**Figure 3.2.2d**). Paradoxically, aside from mt-tRNA^{Met}, mt-tRNA^{Asp} is the only other mt-tRNA with reduced charging under arsenite stress (**Figure 3.2.S2a**). Uncharged tRNA may accumulate in the E-site of the ribosome. One possible explanation would be a unique inability of mt-tRNA^{Asp} to be released from the ribosome under arsenite stress, thereby contributing to a reduction of mitochondrial translation through its accumulation in the polysome. Finally, we compared the Δ DMS signals of the polysome-enriched mt-tRNAs versus the polysome-depleted mt-tRNAs in different stem and loop regions (**Figure 3.2.2e**). In contrast to cytosolic tRNAs in polysomes, the anticodon and D loop regions of the polysome-enriched mt tRNAs show higher, the T loop region lower, and the acceptor stem no difference in DMS signals under arsenite stress compared to no stress. This result suggests that polysome-associated mitochondrial tRNAs on the polysome respond to arsenite stress differently from the polysome associated cytosolic tRNAs, perhaps through altered interactions in different ribosome binding sites in the mitochondrial and cytosolic translation systems.

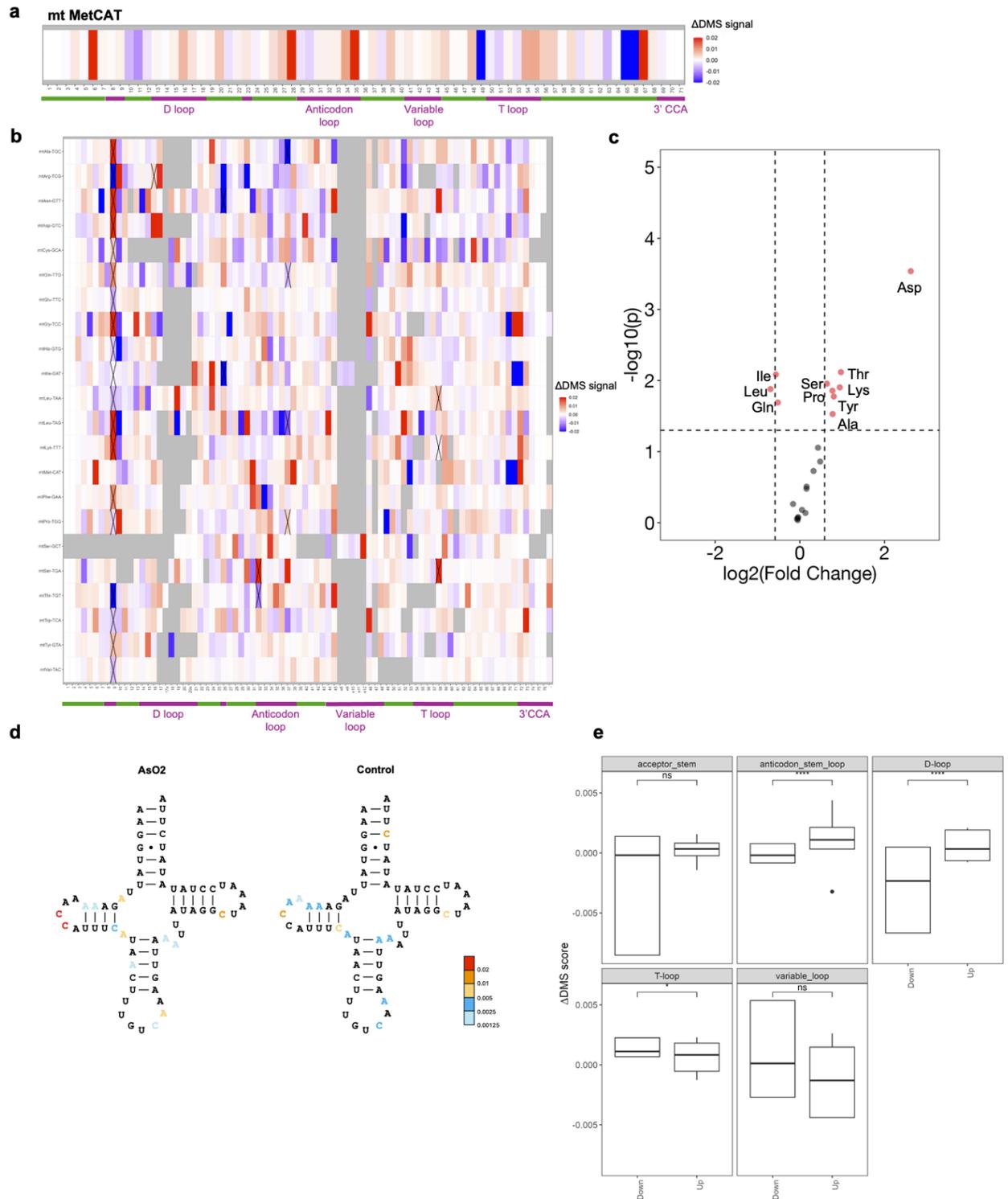


Figure 3.2.2 Mitochondrial tRNA structure and interaction in response to arsenite stress.

(a) Heatmap showing the positional Δ DMS signals on mt-tRNA^{Met}. Native modifications³⁰ that

are incompletely removed by demethylases are marked with “X”. (b) Heatmap showing the positional Δ DMS signals on each mt-819 tRNA. Native modifications that are incompletely removed by demethylases are marked with “X”. (c) Mt-tRNA abundance fold changes in polysome fractions under arsenite treatment. The dotted vertical lines indicate a 1.5-fold change (FC), and the dotted horizontal line indicates $p = 0.05$. Polysome-AsO₂ over polysome-unstressed. Data from NCBI GEO GSE198441. (d) The secondary structure of mt-tRNA^{Asp} is overlaid with in vivo DMS signals with or without arsenite treatment. (e) The Δ DMS signals in D, T, anticodon, variable loops, and the 3' unpaired CCA regions in tRNAs that are enriched (up) or depleted (down) in polysome fractions. P-values are calculated with standard Wilcox t-test; n.s.: not significant, *: $p < 0.05$, ****: $p < 10^{-4.49,50}$

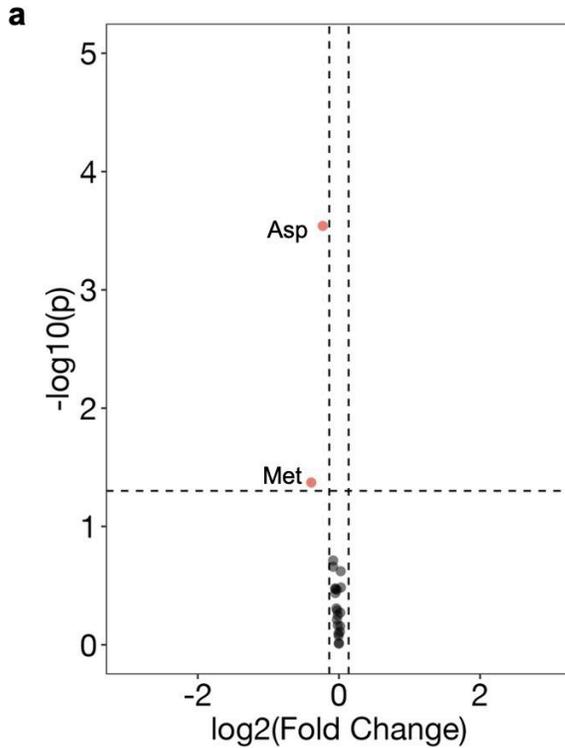


Figure 3.2.S2 mt tRNA charging in polysomes under arsenite stress. (a) Dotted vertical lines indicate 1.1-fold change (FC) , and the dotted horizontal line indicates $p = 0.05$. Data from NCBI GEO GSE198441

3.3 METHODS

The method section for this chapter is the same as those previously discussed in Chapter 2.3. The only exemption is that a pulse chase assay was not conducted for mitochondrial tRNA.

3.4 DISCUSSION

Human cells also contain 22 separate, mitochondrial-encoded tRNAs that are required for mitochondrial protein synthesis. Unlike chromosomal-encoded tRNAs, mt-tRNAs have A/U-rich sequences, and most do not have loop residues for forming canonical tertiary interactions. The

human cytosolic tRNAs contain an average of 13 modifications, whereas the human mt-tRNAs contain an average of only 5-6 modifications. This lower number of mt-tRNA modifications can also contribute to their lower stability. Using samples isolated from total cellular RNAs *both in vitro* and *in vivo*, thus containing the same set of modifications, our *in vitro* DMS mapping result shows that mitochondrial tRNAs do not fold into stable structures, but the *in vivo* DMS data shows that they do fold into canonical structures in general. This result is consistent with mt-tRNA interacting very extensively with cellular proteins/mt-ribosomes *in vivo*. It is likely that the mt tRNA only folds into a tRNA-like structure *in vivo* because of their intracellular interactions. Using the C75 τ score, we also find an inverse correlation between mt-tRNA and amino acid binding affinity to human EF-Tu, consistent with the binding paradigm originally identified from *in vitro* prokaryotic studies.

Arsenite stress induces large-scale changes in mt-tRNA DMS reaction profiles that include both stem and loop regions. This result is once again consistent with mt-tRNA folding *in vivo* undergoing widespread change once mt-tRNA is no longer engaged in translation under arsenite stress, where they likely lose protein/mt-ribosome interaction. Even for those mt-tRNAs that are enriched in the polysome fraction under arsenite stress, their DMS reactivities are also higher compared to the mt-tRNA depleted in the polysome fraction, once again pointing to an alternative mode of tRNA-ribosome interaction, in this case in the likely stalled mt-polysome. Human mt tRNA abundance at the steady state is highly uneven. This abundance is mostly derived from differential degradation rates of individual mt-tRNAs, as mitochondrial transcription synthesizes equal amounts of 11 tRNAs in the plus strand and 8 mt-tRNAs in the minus strand. How individual mt-tRNAs fold without protein/mt-ribosome interaction or upon altered mt-ribosome interaction likely contributes to their degradation rates.

CHAPTER 4

PROFILING SELECTIVE PACKAGING OF HOST TRNA AND VIRAL RNA MODIFICATIONS IN SARS-COV-2 VIRAL PREPARATIONS

Acknowledgment: This chapter is derived from an article published in *Frontiers in Cell and Development Biology* (Peña, et al., 2022). The authors of that article were Noah Peña, Wen Zhang, Christopher Watkins, Mateusz Halucha, Hala Alshammary, Matthew M. Hernandez, Wen-Chun Liu, Adolfo Garcia-Sastre, Viviana Simon, Christopher Katanski, and Tao Pan. Author contributions were as follows: NP analyzed small RNA-seq data; CK analyzed large RNA-seq data; WZ, CW, and MH built small RNA and large RNA-seq libraries, HA, MH, W-CL, RA, AG-S, and VS cultured SARS-CoV-2 viruses and extracted total RNA. VS, CK, and TP designed the experiments. NP, VS, and TP wrote the paper.

4.1 INTRODUCTION

Viral assembly is a critical stage in the viral life cycle that produces mature virus containing the viral genome and proteins needed to infect another host target cell. As early as 1980s it was shown that certain viruses also package host RNAs into their virions. Host transfer RNA (tRNA) is a major cellular RNA family, which is packaged in virions (Isaac and Keene, 1981; Jiang et al., 1993). tRNAs are the most abundant RNA in copy numbers in cells, and their small size and stable structure make them good targets for interacting with viral RNA and viral proteins.

The best studied viral packaging of host RNAs has been described for retroviruses. Retroviruses require a specific host tRNA as reverse transcriptase primers in the cDNA synthesis of the viral genomic RNA upon infection. HIV-1 uses tRNA^{Lys}(TTT) from the host cell since it has a fully complementary sequence of ~20 nucleotides to the primer binding site of the retroviral genome (Litvak et al., 1994). In addition to tRNA^{Lys}(TTT), other prominent tRNAs packaged into the virions include tRNA^{Lys}(CTT), tRNA^{Asn}(GTT) and others (Pavon-Eternod et al., 2010). Furthermore, retroviruses also package non-coding RNAs, the prominent one is the signal recognition particle (SRP) RNA (Eckwahl et al., 2015; Eckwahl et al., 2016). SRP RNA is a component of the SRP particle that is required for the co-translational synthesis of membrane proteins and secretory proteins (Keenan et al., 2001). However, aside from the tRNA serving as the primer for retroviral replication, the precise functions of the other packaged RNAs remain to be elucidated. One possibility is that the packaged RNAs are proximal to the cellular locations of viral assembly and encapsulation. If assembly is co-translational, the composition of packaged RNAs may reflect the translation machinery where the viral structural protein synthesis occurs. For example, the packaged tRNAs may be enriched for those reading the retroviral gag protein codons (van Weringh et al., 2011). Another possibility is that packaged host tRNAs may reduce innate immune recognition of viral genomic sequences by cytosolic pattern recognition receptors, as the host tRNAs may be seen as self-RNA upon infection (Karikó et al., 2005). However, viral packaging of non-retroviruses has rarely been explored using transcriptomic approaches which could potentially generate functional hypotheses on host RNA packaging in viral biology.

Here, we utilize both small RNA (<200 nt) and large RNA (>200 nt) sequencing to identify host RNAs that are present in cell free viral preparations which consist mostly of the

SARS-CoV-2 viruses cultured on VeroE6 cells. We obtain sequencing data for six distinct primary SARS-CoV-2 isolates (**Table 4.2.1**) and compare them to those from uninfected cells. We identify selective enrichment of host tRNAs and SRP RNA in the viral preparations. tRNA^{Lys}(TTT) is among the selectively packaged tRNAs, just like HIV-1. We find that a specific tRNA modification may influence tRNA packaging, and some packaged tRNAs are likely tRNA fragments. We also identify a low level of SARS-CoV-2 subgenomic transcripts in the viral preparations, as well as several candidate modification sites in the SARS-CoV-2 genomic RNA.

4.2 RESULTS

4.2.1 tRNA seq of SARS-CoV-2 Viral Particles Reveal Selective Enrichment

We performed Illumina sequencing starting with total RNA extracted from VeroE6 cells and from SARS-CoV-2 viral preparations cultured on VeroE6 cells (**Figure 4.2.1A**; **Figure 4.2.S1**). To improve efficiency and quantitative assessment of small RNA-seq, defined here as RNA of <200 nucleotides in length, we built two libraries for each sample. The first library was treated with a demethylase mixture (DM) (Zheng et al., 2015) which removed many Watson-Crick face methylations in tRNA that impede reverse transcription in library construction while the second library was left untreated. As described previously, the DM-treated libraries are useful for quantitative assessments of transcript abundance, whereas the untreated samples are useful for modification analysis (Clark et al., 2016). As expected, sequencing reads of the VeroE6 cells mostly mapped to tRNAs, followed by those from 5S to 5.8S rRNA, a small amount of SRP RNA, and others such as spliceosomal RNA (snRNA) and Y RNA (**Figure 4.2.1B**). In the viral preparations, a substantial proportion of reads mapped to viral genomic RNA as expected, and

tRNA and SRP RNA are present at almost equally high proportions, followed by a small amount of rRNA (**Figure 4.2.1B**). Although one cannot exclude RNA in exosomes or extracellular RNA not associated with vesicles, our data clearly show RNAs that are differently present in our cell free preparations as compared to those in cells. Our viral preparations used for the sequencing experiments contain high levels of infectious particles outside the cell, strongly suggesting that the sequenced RNA is derived, to a large extent, from cell free virions. For example, our results show a ~150-fold enrichment of the SRP RNA over tRNA in the viral preparation samples vs. the cell samples, which suggests that we eliminated most if not all of the cellular debris. These results indicate that SARS-CoV-2 virions also package tRNA and SRP RNA in significant proportions.

We next examined the selectivity of packaged tRNA at two levels. Mammalian genomes contain many tRNA isodecoder genes that share the anticodon but possess different body sequences; all tRNA isodecoders with the same anticodon belong to a single tRNA isoacceptor family (Goodenbour and Pan, 2006; Schimmel, 2018). Abundance of isodecoders was summed for each isoacceptor family and used to calculate the fraction of tRNA reads for each anticodon. The isoacceptor abundance fraction for each sample (three biological replicates of uninfected VeroE6 cells and six distinct viral culture supernatant preparations), was compared to the mean of Vero E6 cells. All three VeroE6 cell replicates were nearly identical, as the heat map shows close to zero values in all cases (**Figure 4.2.1C**). To avoid exaggerated representation of low abundant tRNAs by ratioed comparison, we subtracted the tRNA fraction in each viral preparation to its counterpart in the VeroE6 cells, so that the differences were readily identified for more abundant tRNAs (**Figure 4.2.1C**). We found several isoacceptor families that are significantly enriched across all six isolates. They include tRNA^{Glu}(TTC), tRNA^{Lys}(TTT),

tRNA^{Leu}(AAG), tRNA^{Ser}(AGA), tRNA^{Ser}(GCT), and tRNA^{Ser}(TGA). These results indicate that SARS-CoV2 virions selectively incorporate tRNA isoacceptors.

Code used in this manuscript	BEI #	Viral isolate name	Titer (PFU/ml) (Vero E6)	Lineage	GISAID clade	GISAID ID	Clinical presentation/COVID-19 outcome
Viral isolate #1	NR-53517	SARS-CoV-2, Isolate New York-PV09197/2020	1.5×10^4	B.1.3	GH	EPI_ISL_422552	90 years old Male; severe COVID-19 with fatal outcome
Viral isolate #2	NR-53514	SARS-CoV-2, Isolate New York-PV08410/2020	5×10^3	B.1	GH	EPI_ISL_421374	63 years old Male; severe COVID-19 with fatal outcome
Viral isolate #3	NR-52439	SARS-CoV-2, Isolate Chile/Santiago_op4d1/2020	3.25×10^4	A.2	S	EPI_ISL_415661	Patient has respiratory tract infection. History of travel to Europe
Viral isolate #4	NR-53515	SARS-CoV-2, Isolate New York-PV08449/2020	1×10^4	B.1	GH	EPI_ISL_421400	88 years old Female; severe COVID-19 with fatal outcome
Viral isolate #5	NR-52368	SARS-CoV-2, Isolate New York 1-PV08001/2020	2×10^4	B.4	O	EPI_ISL_414476	39 years old Female; history of travel to Iran
Viral isolate #6	NR-53516	SARS-CoV-2, Isolate New York-PV09158/2020	5.75×10^4	B.1.3	GH	EPI_ISL_422525	62 years old Male; severe COVID-19 with fatal outcome

Table 4.2.1 Summary of the SARS-CoV-2 isolates used in this study.

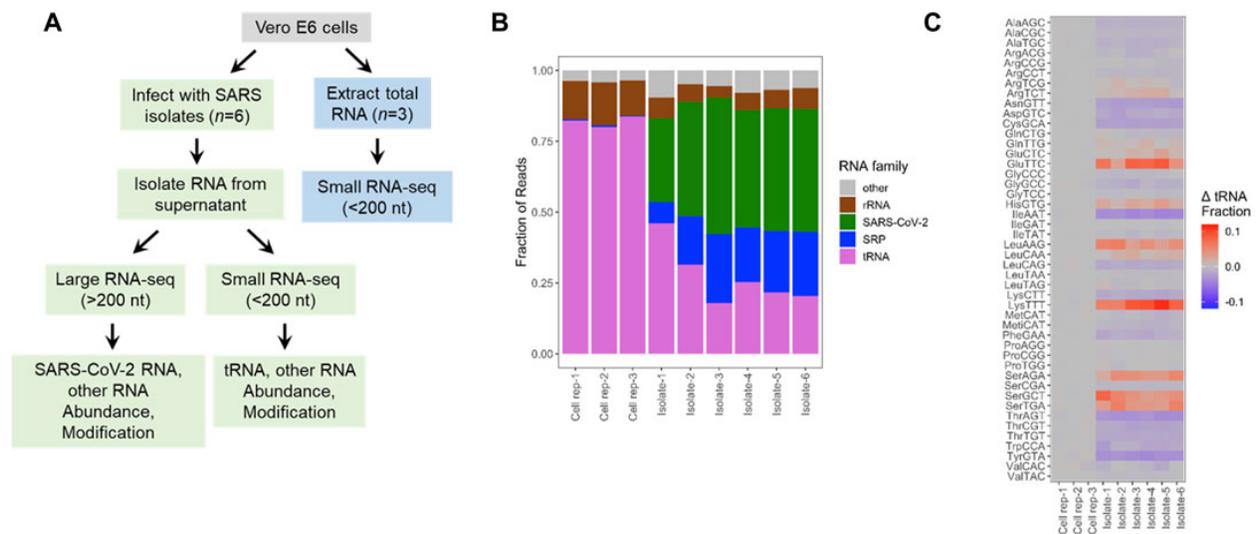


Figure 4.2.1 Selective enrichment of small RNA in SARS-CoV-2 viral preparations. (a)

Experimental scheme. Vero E6 cells were either infected with SARS-CoV-2 virus isolates from infected individuals (n = 6 biological isolates) or uninfected cultures (n = 3 biological replicates). Total RNA was extracted from the cells (blue boxes) or only from the cell free viral preparations (green boxes). Small RNA-seq was carried out using total RNA with and without demethylase treatment. Large RNA-seq was carried out with the RNA fraction after the removal of small

RNAs of <200 nt, and chemical fragmentation (Not discussed in this Chapter). (b) Small RNA-seq results. Vero cell data are mostly tRNA and 5S/5.8S rRNA. Aside from SARS-CoV-2 RNA, virions contain significant portions of tRNA, rRNA, and signal recognition particle (SRP) RNA. (c) Enrichment and depletion of specific tRNAs in the cell free viral samples. Shown are the combined reads from all tRNA isodecoders that share the same anticodon. Heatmap shows the abundance of tRNAs for each anticodon subtracted from the mean of control cultures. Subtraction emphasizes the differences among abundant tRNAs. Enriched tRNAs are in red, depleted tRNAs in blue. Top 3 enriched tRNAs are tRNA^{Lys}(TTT), tRNA^{Glu}(TTC), and tRNA^{Ser}(GCT). Top 3 depleted tRNAs are tRNA^{Ile}(AAT), tRNA^{Tyr}(GTA), and tRNA^{Asn}(GTT).

4.2.2 Selective Enrichment of tRNA Isodecoders in SARS-CoV-2 Viral Preparations

Our downstream analysis, thus, focused on those six tRNAs enriched in the viral preparations. First, we analyzed the tRNA at the isodecoder level for all six tRNAs. Among the seven tRNA^{Glu}(TTC) isodecoders, four could be detected in the viral preparations. However, only two isodecoders represent almost all tRNA^{Glu}(TTC) in the viral preparations, even though neither is the most abundant isodecoder in VeroE6 cells (**Figure 4.2.2A**). In contrast, the single dominant tRNA^{Leu}(AAG), tRNA^{Lys}(TTT), tRNA^{Ser}(AGA) isodecoders in cells are also the ones in the viral preparations (**Figures 4.2.2B–D**). For tRNA^{Ser}(GCT) and tRNA^{Ser}(TGA), two isodecoders each are present at appreciable levels, in each case, the isodecoder at the highest level is also the one in the viral preparations (**Figures 4.2.2E,4.2.2F**).

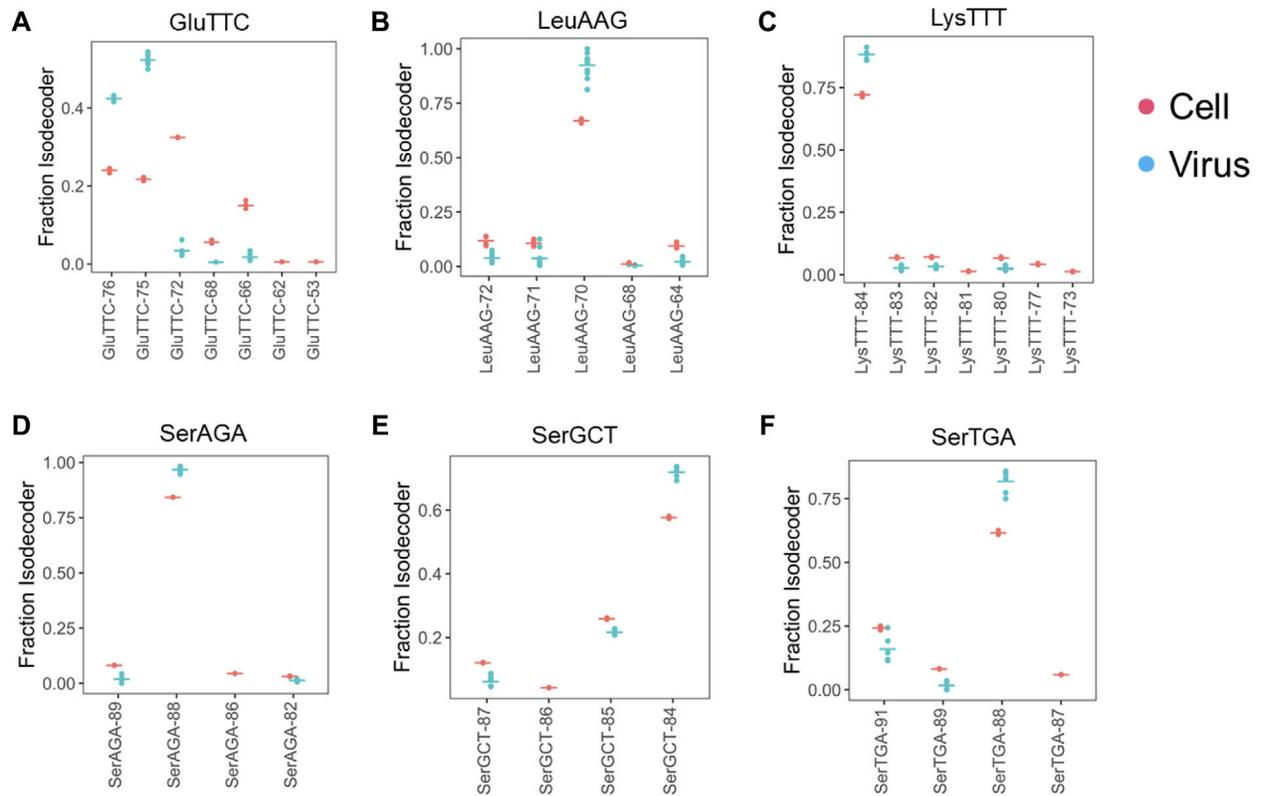


Figure 4.2.2 Selective Enrichment of tRNA isodecoders in SARS-CoV-2 viral preparations.

tRNA isodecoder fractions from uninfected Vero cell ($n = 3$, red) or cell free viral preparations ($n = 6$, blue) are shown. Mean values are shown as a horizontal bar. Isodecoder nomenclature is according to the tRNAScan score of the *Chlorocebus sabaeus* tRNA genes identified in Rfam database. (a) tRNA^{Glu}(TTC). (B) tRNA^{Leu}(AAG). (c) tRNA^{Lys}(TTT). (d) tRNA^{Ser}(AGA). (e) tRNA^{Ser}(GCT). (f) tRNA^{Ser}(TGA).

4.2.3 Selective Enrichment of tRNA Isodecoders in SARS-CoV-2 Viral Preparations

We examined the read pileup of the most abundant isodecoder in the viral preparations. By experimental design, our tRNA-seq results always start from the 3' end of the tRNA and show a decline toward the 5' end with sharp drops at certain tRNA modifications, an expected behavior for full-length tRNAs (Zheng et al., 2015). Three types of results are observed: first, the

pileup decreases faster in the viral preparation tRNA compared to the cellular tRNA, this group includes tRNA^{Glu}(TTC) (**Figure 4.2.3A**). The pronounced drop of the viral preparation tRNA in the anticodon loop region is consistent with GluTTT-75 in the virion being a 3' half tRNA fragment with the 5' end in the anticodon loop, because there is no known RT stopping modifications in this tRNA. In the second type, the read pileup decreases at about the same rate, this group includes tRNA^{Leu}(AAG) and tRNA^{Lys}(TTT) (**Figures 4.2.3B,4.2.3C**). The similar drop off is consistent with the tRNA in the viral preparations as the full-length tRNA like those in cells, and the sharp drop offs corresponds to the N^{2,2}-dimethyl-G at position 26 (m²²G26) in tRNA^{Leu}(AAG) which is difficult to remove by the demethylase because it is buried in the tRNA structure (Dai et al., 2017) and 2-methylthio-6-carbamoylthreonine at position 37 (ms²t⁶A37) in tRNA^{Lys}(TTT) (Machnicka et al., 2013) which does not react with the demethylase. In the third type, the read pileup decreases slower in the viral preparation tRNA, this group includes all three tRNA^{Ser}, and the sharp drop offs correspond to the m²²G26 modification (**Figures 4.2.3D–F**). This result is consistent with the tRNA^{Ser} in the viral preparations having lower modification levels in the anticodon stem-loop region which can include N⁶-methyl-N⁶-threonylcarbamoyladenosine at position 37 (m^{6t}A37) in tRNA^{Ser} (Machnicka et al., 2013).

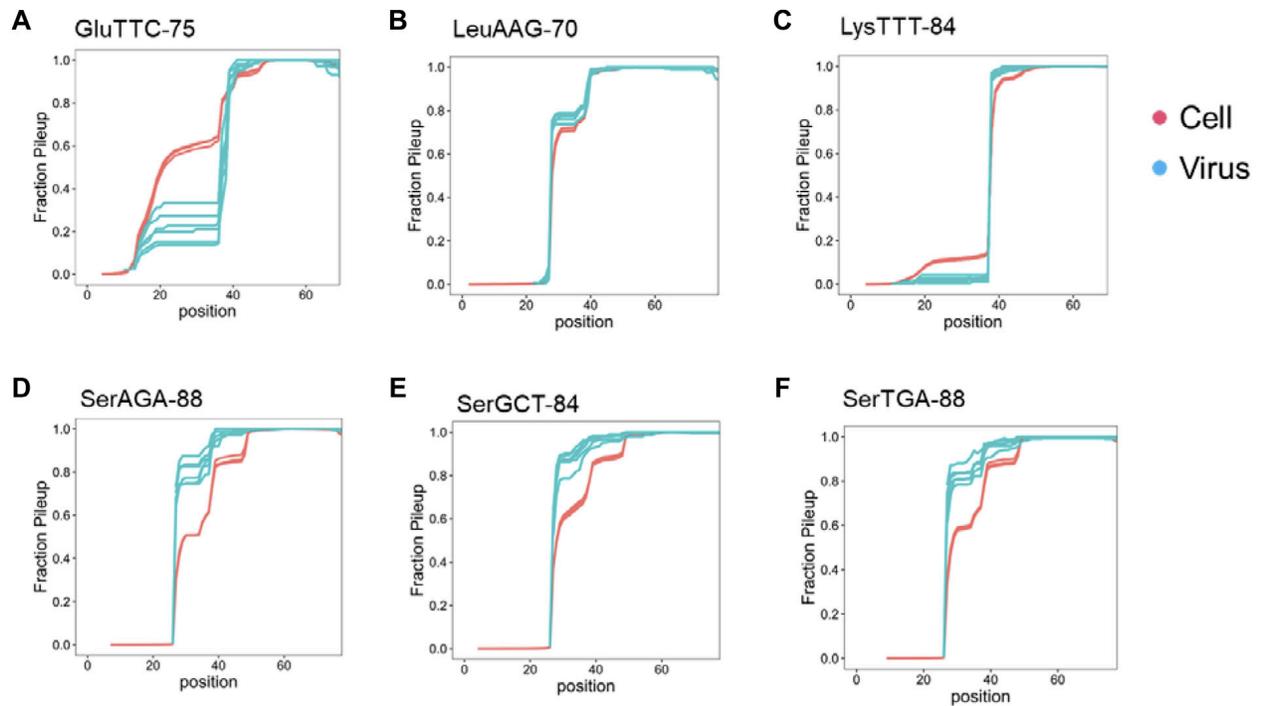


Figure 4.2.3 Read pileup of the enriched tRNA fragments in SARS-CoV-2 viral

preparations. Shown are read pileups of the most abundant tRNA isodecoders in viral isolates ($n = 6$, blue) and their counterparts in uninfected Vero cell ($n = 3$, red). Isodecoder nomenclature is according to the tRNAScan score of the *Chlorocebus sabaeus* tRNA genes identified in Rfam database. (a) tRNA^{Glu}(TTC). This result is consistent with 3' tRNA fragment being the dominant form in the viral preparations. (b) tRNA^{Leu}(AAG). This result is consistent with full-length tRNA in the viral preparations. (c) tRNA^{Lys}(TTT). This result is consistent with full-length or 3' tRNA fragment with 5' end exactly at position 39 in the viral preparations. (d) tRNA^{Ser}(AGA). The tRNA^{Ser} results are consistent with full-length tRNA in the viral preparations. (e) tRNA^{Ser}(GCT). (f) tRNA^{Ser}(TGA).

4.2.4 Modification Profile of Enriched tRNA in SARS-CoV-2 Viral Particles

We next looked for RNA modification differences among the tRNAs from cells and viral preparations. In tRNA-seq, certain modifications can be identified by their “mutation” signatures in the sequencing data. Vero cells are derived from *Chlorocebus sabaesus* kidney, their tRNA modifications have not been reported in the literature. We analyzed the mutation signatures of the cellular tRNAs in the libraries without demethylase treatment and validated the methylations using the results from demethylase treatment (Clark et al., 2016) to provide a comprehensive analysis of Vero cell tRNA modifications (**Table 4.2.2**). Overall, the *C. sabaesus* tRNA modification patterns detected by sequencing are very similar to those from the human HEK293T cells (Table 1 in Clark et al. (2016)). A minor difference is the m²²G²⁶ modification which is present in tRNA^{Val} in *C. sabaesus* Vero cells but not in HEK293T cells. We detected inosine modification at the wobble anticodon position (I34) in all tRNAs that are A34 in the genome which include tRNA^{Leu}(AAG) and tRNA^{Ser}(AGA) (**Figure 4.2.4A; 4.2.S1A; Table 4.2.2**).

We then compared the mutation levels between the tRNAs enriched in viral preparations and their counterparts in cells. For high confidence analysis we applied a filter of ≥ 50 read coverage at nucleotide positions of interest. The modification sites passing this filter among the tRNAs in viral preparations only include m¹A⁵⁸ (tRNA nomenclature) in tRNA^{Leu}(AAG), tRNA^{Lys}(TTT), and tRNA^{Glu}(TTC), and I34 in tRNA^{Leu}(AAG) and tRNA^{Ser}(AGA). For tRNA^{Leu}(AAG) and tRNA^{Lys}(TTT), the mutation fraction at m¹A⁵⁸ is higher in the tRNA from the viral preparations than the VeroE6 cell tRNA (**Figure 4.2.4B; Figure 4.2.S1B**), suggesting preferential packaging of m¹A modified tRNA. Among the tRNA^{Glu}(TTC) isodecoders, m¹A⁵⁸ level is variable in cells. Only two of the five abundant tRNA^{Glu}(TTC) isodecoders have high

modification levels in VeroE6 cells, but only isodecoders with low modification fractions are present in the viral preparations (**Figure 4.2.4C**). Since tRNA^{Glu}(TTC) in the virions are likely tRNA fragments, this result is consistent with low m¹A modified tRNA^{Glu}(TTC) being the preferred source of tRNA^{Glu}(TTC) fragments in cells. tRNAs in cells and in the viral preparations are >90% modified with I34 in both tRNA^{Leu}(AGA) and tRNA^{Ser}(AGA) (**Figure 4.2.4A; Figure 4.2.S1A**).

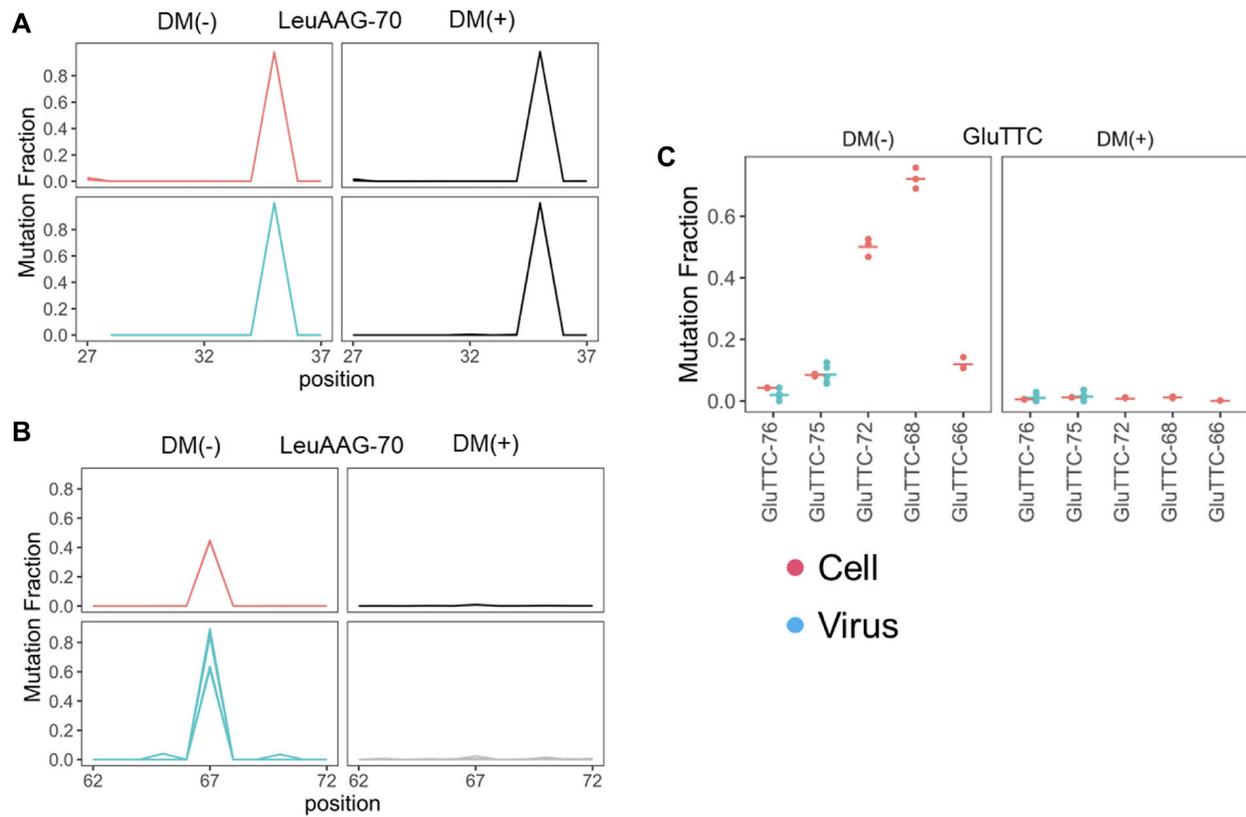


Figure 4.2.4 Selective enrichment of tRNA with m¹A profiles. Mutation fractions from uninfected Vero cell (n = 3, red) or cell free viral preparations (n = 6, blue) are shown. (a) Mutation fractions of tRNA^{Leu}(AAG) residues around the wobble anticodon position (35 for this tRNA) without (DM-) and with (DM+) demethylase treatment showing the I34 modification. (b) Mutation fractions of tRNA^{Leu}(AAG) around the residues at position 67 which corresponds to m¹A58 in the tRNA nomenclature. tRNA^{Leu}(AAG) shows higher mutation fraction in the viral

preparations, consistent with SARS-CoV-2 selectively packaging m¹A modified tRNA^{Leu}(AAG).

(c) Mutation fractions of the top five abundant tRNA^{Glu}(TTC) isodecoders at position 57 (DM-) which is validated as m¹A in the T loop upon removal by demethylase treatment (DM+).

Isodecoder nomenclature is according to the tRNAScan score of each tRNA^{Glu}(TTC) gene. The two isodecoders enriched in the viral preparations are nearly unmodified, corresponding to their counterparts in the Vero cells.

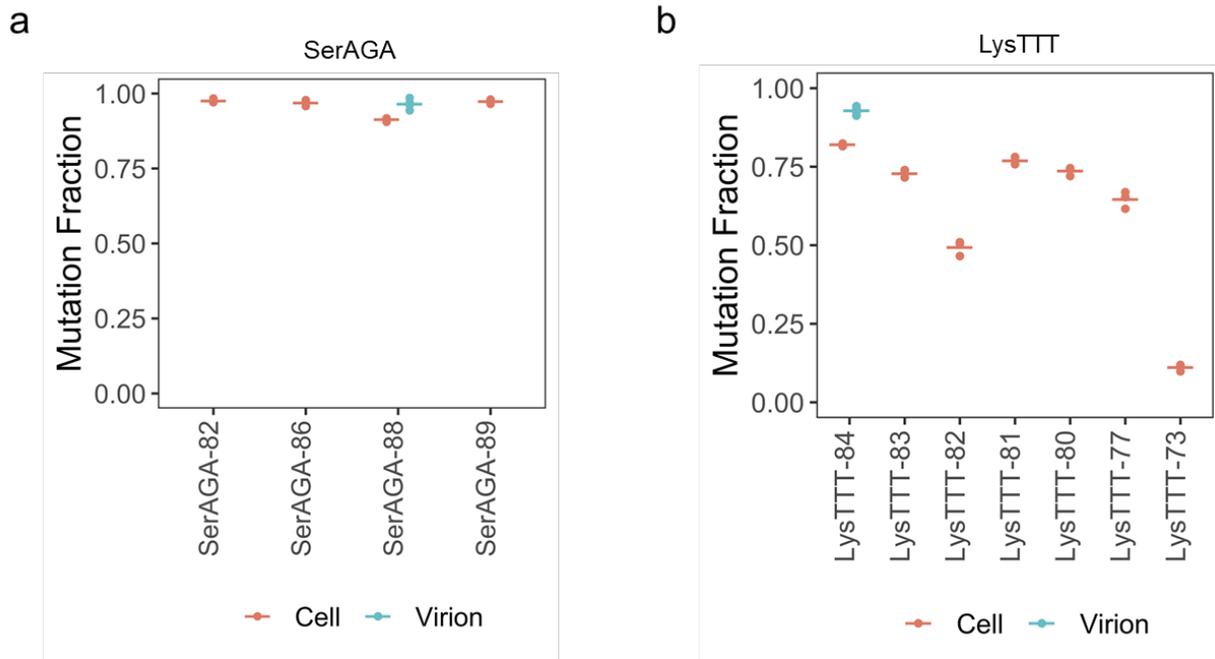


Figure 4.2.S1 Selective enrichment of other tRNA with modification profiles. (a) Mutation fractions of tRNA^{Ser}(AGA) isodecoders from uninfected Vero cell (n=3, red) or viral preparations (n=6, blue) of the wobble anticodon position (I34). (b) Mutation fractions of tRNA^{Lys}(TTT) isodecoders of position m¹A58.

tRNA ^a	m ¹ A58	m ¹ G37	I34 ^b	m ² G26	m ¹ G9	m ³ C
AlaAGC	X	X ^c	X	X		
AlaCGC	X	X ^c		X		
AlaTGC	X	X ^c		X		
CysGCA	X					
AspGTC	X				X ^d	
GluCTC	X					
GluTTC	X					
PheGAA	X			X		
GlyCCC	X					
GlyGCC	X					
GlyTCC	X					
HisGTG	X	X				
IleAAT	X		X	X		
IleTAT	X			X		
LysCTT	X					
LysTTT	X					
LeuAAG	X	X	X	X		X ^e
LeuCAG	X	X		X		
LeuTAG	X	X		X		
LeuCAA	X	X		X		
LeuTAA	X	X		X		
Met-i	X			X	X	X ^f
Met-e	X					
AsnGTT	X			X	X	
ProAGG	X	X	X		X	
ProCGG	X	X			X	
ProTGG	X	X			X	
GlnCTG	X					
GlnTTG	X					
ArgACG	X		X	X		
ArgCCG	X			X		
ArgTCG	X			X		
ArgCCT	X				X	X ^g
ArgTCT	X				X	X ^g
SecTCA	X					
SerAGA	X		X	X		
SerCGA	X			X		X ^h
SerTGA	X			X		X ^h
SerGCT	X			X		X ^h
ThrAGT	X		X	X	X	X ^g
ThrCGT	X			X	X	X ^g
ThrTGT	X			X	X	X ^g
ValAAC	X		X	X		
ValCAC	X			X		
ValTAC	X			X		
TrpCCA	X	X		X	X	
TyrGTA	X	X		X		

^am¹A, m¹G, m²G, and m³C mutations in sequencing are demethylase sensitive. X = present.

^bI34 is not sensitive to demethylase treatment.

^cm¹I37 in tRNA^{Ala}.

^dm¹A9 in tRNA^{Asp}.

^em³C47d (in variable loop of type II tRNA) in tRNA^{Leu}(CAG).

^fm³C20 in tRNA^{Met}.

^gm³C32 in tRNA^{Arg} and tRNA^{Thr}.

^hm³C32 and m³C47d in tRNA^{Ser}.

Table 4.2.2 Modifications identified in *C. sabaesus* tRNAome

4.3 METHODS

4.3.1 SARS-CoV-2 Isolates

Residual nasopharyngeal swab specimens were collected after completion of the diagnostic process as part of the Mount Sinai Pathogen Surveillance Program. To culture SARS-CoV-2 isolates, .1 ml of viral transport media was inoculated into one well of a six-well plate seeded with a confluent monolayer of VeroE6 cells. Culture supernatants were harvested when CPE (cytopathic effect) became visible, aliquoted, and stored at -80°C . All work related to SARS-CoV-2 culture was performed in a BSL3 biocontainment facility by trained personnel and in accordance with the research registration approved by the Institutional Biosafety Committee (IBC).

We cultured six distinct SARS-CoV-2 isolates representing the early lineages of the pandemic (**Table 4.2.1**). After isolation of the clinical isolates on VeroE6 cells, we determined the infectious viral titers for each of the viral culture supernatants by plaque assay (**Table 4.2.1**). All six isolates displayed medium plaque phenotypes. We shared aliquots of the viral stocks analyzed in this study with the NIH BEI repository in the early summer of 2020.

4.3.2 RNA Isolation

The viral culture supernatants were spun at 3,000 rpm for 10 min to remove particulates. vRNA from each viral preparation and total RNA from VeroE6 cells was performed using QIAamp Viral RNA Kits (Qiagen) following the manufacturer's instructions. We did not filter the supernatants as that could result in some viruses being absorbed to the membrane and/or causing a loss of viral infectivity.

4.3.3 RNA Library Construction and Sequencing

The procedure was adapted from DM-tRNA-seq (Zheng et al., 2015) with the following modifications for tRNA deacylation: 10 μ l of total RNA from viral culture supernatants or uninfected cells (containing up to a maximum of 1 μ g of total RNA, as measured by NanoDrop) were deacylated by adding 5 μ l of 100 mM Na₂B₄O₇, pH 9.5 (final concentration: 33.3 mM) and incubated at 37°C for 30 min. To the deacylated samples, 5 μ l of a 3'-end clean-up mixture (200 mM Tris-HCl, pH 6.8, 40 mM MgCl₂, and 4 U/ μ l T4 PNK [NEB]) were added and incubated at 37°C for 20 min, and then heat inactivated at 65°C for 10 min. We used superscript IV RT in this work.

4.3.4 Sequencing Data Analysis

Reference RNA sequences from *Chlorocebus sabaesus* that included non-coding RNA and tRNA were downloaded from Rfam database (<https://rfam.xfam.org/>, Kalvari et al., 2021). *C. sabaesus* tRNA sequences from Rfam were processed through tRNAScan-SE (<http://lowelab.ucsc.edu/tRNAScan-SE/>, Lowe and Chan, 2016), only sequences with high confidence (i.e., tRNAScan score ≥ 50) were used as reference. Following this, tRNA sequences were appended by adding CCA at their 3' end as well as removing intron sequences. These processed *C. sabaesus* cytosolic tRNA, mitochondrial tRNA, and non-coding RNA sequences such as SRP from Rfam were combined with the Wuhan SARS-CoV-2 genome sequence (MN908947.3) to generate a custom reference database.

Raw reads following sequencing were designated reads 1 and reads 2 and were merged together using `bbmerge.sh` present within the `bbmap` package (<https://github.com/BioInfoTools/BBMap>), which results in merged fastq files. These merged

fastq were converted to fasta file format using reformat.sh present within the bmap package. These fasta files were aligned to our custom reference genome using bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>, Langmead et al., 2009) with the following parameters: -f -p 10 --local --no-unal. The aligned reads were then used to determine RNA sequence abundance using custom python script. RNA modifications were detected based on aligned reads using samtools sort (<http://www.htslib.org/>, Li et al., 2009) feature sort the reads in a bam file format. Then IGVtools count (<https://software.broadinstitute.org/software/igv/igvtools>) count feature was utilized to output a wig files using the following parameters: -z 5 -w 1 -e 250 — bases. The resulting wig files were processed using a custom python script to identify nucleotide mutations as well as coverage of aligned reads.

4.3.5 Data Availability Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below:

<https://www.ncbi.nlm.nih.gov/>, GSE182883.

4.4 DISCUSSION

In this work, we identified selective enrichment of host tRNAs and SRP RNA in cell free SARS-CoV-2 viral preparations, differences in tRNA modification between the tRNA in viral preparations and in cells, and candidate modification sites in the packaged SARS-CoV-2 genomic RNA. We estimate that a SARS-CoV-2 virion contains up to one molecule of SRP RNA. Given the roughly equivalent read counts of SRP and tRNA in the viral preparation

(Figure 4.2.1B), and full-length tRNAs are approximately one fourth in size to SRP RNA, we estimate that a SARS-CoV-2 virion also contains up to four tRNA molecules.

How and why specific tRNAs and the SRP RNA are enriched in virions remains unclear. Packaging of the HIV primer tRNA^{Lys}(TTT) is facilitated by the host lysyl-tRNA synthetase (LysRS) and gag protein interaction (Cen et al., 2002). As a consequence, tRNA^{Lys}(CTT) is also packaged at similar levels. In our case, however, tRNA^{Lys}(CTT) was not enriched in the virions. On the other hand, the enrichment of multiple tRNA^{Ser} isoacceptors may go through this mechanism of facilitating tRNA packaging through host seryl-tRNA synthetase (SerRS) and SARS-CoV-2 protein interactions. Retroviruses also package a large amount of SRP RNA into virions (Eckwahl et al., 2015; Eckwahl et al., 2016). Our results add SARS-CoV-2 to this list.

Cells release tRNA fragments into extracellular medium in many ways (Torres and Martí, 2021). tRNA modifications influence tRNA fragment biogenesis, and the secreted tRNA fragments often match the abundance profiles of those fragments in cells. For the tRNA^{Glu}(TTC) fragment, its levels and m¹A58 modification profile match in the viral preparations and in cells. However, the full-length tRNA^{Leu}(AGA) and tRNA^{Lys}(TTT) in the viral preparations have higher m¹A58 levels than their counterparts in cells. M¹A58-modified tRNA can interact differently with cellular components such as eEF1A compared to the hypo-modified tRNA (Liu et al., 2016). The higher tRNA m¹A58 level in the viral preparations may be related to their enhanced interaction with viral proteins. We also identified candidate modifications in the SARS-CoV-2 genomic RNA. It is surprising that none of the five modification sites described here could be readily assigned to well characterized modifications, just like those sites reported by nanopore sequencing (Kim et al., 2020). Viral RNA modification studies have only taken off

recently and future studies will be needed to reveal the chemical nature and the functional consequences of these modifications.

This proof-of-concept study was performed using the viral isolates cultured early in the pandemic (March/April 2020). In future work, we will produce larger amounts of SARS-CoV-2 isolates representative of the early circulating strains as well as the viral variants of concern that have dominated most of the pandemic in 2021. We will culture these isolates on ACE2-expressing human cells rather than on African green monkey VeroE6 cells. These studies will firmly establish the dependency of host RNA packaging on SARS-CoV-2 variants and on host cell source.

CHAPTER 5

DESIGN, CONSTRUCTION, AND FUNCTIONAL CHARACTERIZATION OF A tRNA NEOCHROMOSOME IN YEAST

Acknowledgment: This chapter is derived from an article that was published in *Cell*. (Schindler, Walker, et al 2023) The authors of this article were Daniel Schindler, Roy S.K. Walker, Shuangying Jiang, Aaron N. Brooks, Yun Wang, Carolin A. Müller, Charlotte Cockram, Yisha Luo, Alicia García, Daniel Schraivogel, Julien Mozziconacci, Noah Peña, Mahdi Assari, María del Carmen Sánchez Olmos, Alba Ballerini Yu Zhao, Benjamin A. Blount, Jitong Cai, Lois Ogunlana, Wei Liu, Katarina Jönsson, Dariusz Abramczyk, Eva Garcia-Ruiz, Tomasz W. Turowski, Reem Swidah, Tom Ellis, Tao Pan, Francisco Antequera, Yue Shen, Conrad A. Nieduszynski, Romain Koszul, Junbiao Dai, Lars M. Steinmetz, Jef D. Boeke, and Yizhi Cai. Conceptualization, J.D.B. and Y.C.; coordination, J.D.B. and Y.C.; design and construction of the tRNA neochromosome, D. Schindler, R.S.K.W., Y.L., J.D.B., and Y.C.; design of experiments, D. Schindler, R.S.K.W., F.A., Y.S., C.A.N., R.K., J.D., L.M.S., J.D.B., and Y.C.; conducting experiments, D. Schindler, R.S.K.W., S.J., A.N.B., Y.W., C.A.M., C.C., A.G., D. Schraivogel, J.M., M.A., M.d.C.S.O., Y.Z., A.B., L.O., W.L., K.J., D.A., E.G.-R., T.W.T., and R.S.; data analysis, D. Schindler, R.S.K.W., A.N.B., Y.W., C.A.M., C.C., A.G., D. Schraivogel, J.M., N.P., J.C., B.A.B., T.W.T., T.E., T.P., F.A., Y.S., C.A.N., R.K., J.D., L.M.S., J.D.B., and Y.C.; writing original draft, D. Schindler, R.S.K.W., and Y.C.; writing – review & editing, all authors.

5.1 INTRODUCTION

Synthetic biology is a field that focus on developing organisms or *de novo* genomes that can be easily engineered to take on new functions. (Liu, et al 2019) This can be achieved by having organism or genomes are standardized and designed to contain modular building blocks. An example of this modular and standardized designed building block is the development of minimal synthetic promoters that reduce promoter lengths by at least 80% that are still able to express at high levels while not leading to toxicity within the cells. (Redden and Apler, 2015) This has been accomplished in yeast *Saccharomyces cerevisiae*, which is an organism that is important for genetic and biological research but also a necessary component for production of industrial pharmaceuticals and food ingredients. However, though advances have been made in synthetic biology in well-studied biological systems like *S. cerevisiae*, the major impendent to achieve standardization and modularity is the high complexity and unpredictability of biological systems. (Szymanski and Calvert 2018) As discussed by Szymanski and Calvert, the development of the recent smallest genome that can support life by the J. Center Venture Institute was done via trial-and-error experiments because rational design attempts based on existing microbial functioning knowledge did not work. For this reason, directed evolution, the process of having cells adapt to challenging environments and using the resulting genomic solution, is suggested to have more practical value for synthetic biology. It is this strategy that is the basis of the Synthetic Yeast Genome Project (Sc2.0), which is project that is aimed at creating a designer version of the yeast genome.

The major goals of the Sc2.0 project are the create a yeast genome that has the same fitness as wild-type, has increased genetic flexibility, and has removed genetically unstable components like transposons and tRNA genes using Synthetic Chromosome Rearrangement and

Modification by LoxPsym-mediated Evolution (SCRaMbLE), a directed evolution strategy. (Szymanski and Calvert 2018) Yeast tRNA, like mammals, contains high number of tRNA genes (isodecoders) that encode for each tRNA family (isoacceptors). Some of these tRNA genes also contain introns, which are removed during processing of tRNA to mature tRNA forms. However, to comply with Sc2.0 guidelines, the yeast tRNA genes should be on a synthetic chromosome (tRNA neo-chromosome) that contains the tRNA genes that can express at wild-type abundance levels while also removing any introns. (Richardson et al 2017). We report that a yeast tRNA neochromosome has been designed, synthesized, and characterized for *S. cerevisiae*.

In brief, I will review the *de novo* design and synthesis of the yeast tRNA neochromosome. Yeast contains 275 nuclear tRNA genes in addition to one pseudogene. (Han et al 1998) These genes are known to also be genomic instability hotspots. This is due to polar replication fork collisions between transcription machinery of tRNA genes that are highly transcribed, resulting in replication stalling. This stalling leads to DNA breakage and overall chromosome instability. (Admire et al 2006, Hamperl et al 2017) In addition, retrotransposons and solo long terminal repeats (LTR), which contain highly repetitive sequences, are known to bind and interact with the upstream regions of tRNA genes. (Ji et al 1993, Mularoni et al 2012). This binding leads to increase chromosomal recombination and chromosomal instability. Because of these reasons, placing tRNA genes on a synthetic chromosome that has removed these repetitive elements and has the tRNA genes oriented in tandem to ensure a single direction of transcription will alleviate the major sources for chromosomal instability. To achieve this, Schindler and Walker designed a tRNA neochromosome that had 275 tRNA gene cassettes (~680 bp) that make up 16 tRNA arrays (2.6-2.3 kB). One to three tRNA arrays were used to make mega arrays. Linearization and telomere seed sequence release using a telomerator system,

finally converting the tRNA neochromosome into a linear chromosome. (**Figure 5.1.1a**) Native tRNA genes were retained for the tRNA neochromosome to maintain endogenous expression levels, except introns were removed when necessary. To limit host homology, the flanking regions were derived from non-*S. cerevisiae* yeast species, *Ashbya gossypii* or *Eremothecium cymbalariae*. These species do not contain LTRs, or retrotransposons however are still closely related to *S. cerevisiae*. Due to multiple replication sites being present within the tRNA neochromosome, there remained a possibility that the replication fork may expand beyond its designated site leading to fork collision and resulting chromosomal instability. To alleviate this possibility, Fob1 mediated termination sites were added. These termination sites are similar to those used by ribosomal genes to prevent head-on collisions by RNA Polymerase I and replication machinery. (Brewer et al 1992) (**Figure 5.1.1b**)

Following the design of this tRNA neochromosome, it was inserted in yeast cells to be expressed and characterized. This was conducted using tRNA-seq, with characterization of native expression further validated by MSR-seq. Using this analysis, we determined that neochromosome expression of native tRNA genes did not impeded wild-type yeast expression at the global level and that tRNA from the neochromosome was highly expressed. Due to this characterization of tRNA expression being my contribution to this project, I will only discuss this result further in this chapter. The MSR-seq characterization of tRNA expression shows that a *de novo* designed tRNA neochromosome that meets the major aims of Sc2.0 has been achieved. With the presence of *rox* recombination sites, this tRNA neochromosome can be utilized for future studies further optimize tRNA neochromosome expression and study features of tRNA chromosome using SCRaMbLE.

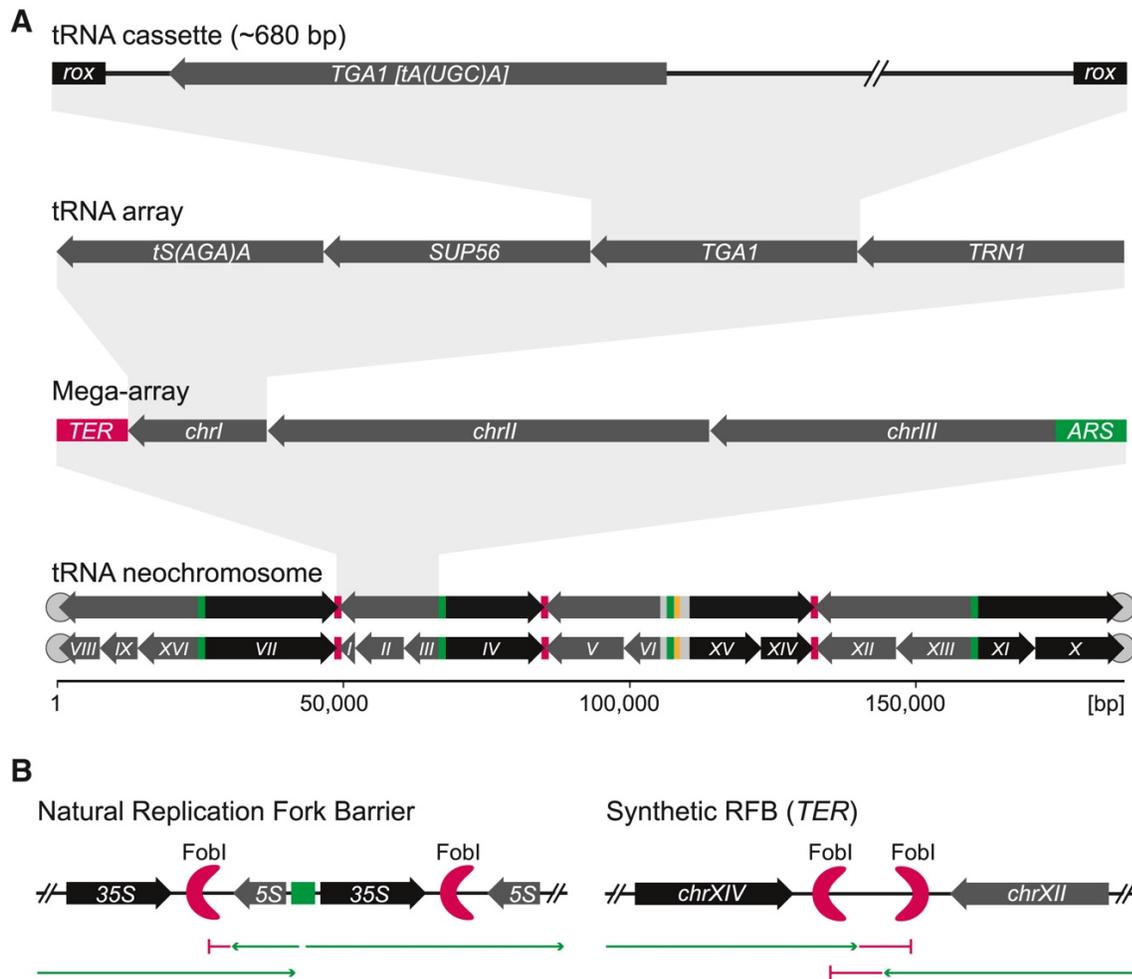


Figure 5.1.1 Design and hierarchy of the tRNA neochromosome. (a) tRNA cassettes have native tRNA genes, without introns as necessary, of orthogonal 500-bp 5' and 40-bp 3' sequences flanking each tRNA gene. *rox* recombination sites sequences also flank each tRNA cassette. Each tRNA cassette of the tRNA arrays are oriented so that transcription follows in the same direction. The tRNA arrays form mega arrays that are each of the approximate same size, with replication sites and transcription also oriented in same direction. Autonomous replicating sequences (ARS) are indicated in green, synthetic bidirectional termination sites (*TER*) are in red, and the pRS413 centromere site is in orange. (b) Synthetic bidirectional terminal sites contain Fob1 blocking sites, which reduce the collision events. This design is based on the native replication fork of ribosomal genes.

5.2 RESULTS

The characterize the expression of tRNA from the tRNA neochromosome while in a native yeast context is difficult. This is because the tRNA neochromosome contains native tRNA genes, with only the 5' and 3' regions differing from native yeast chromosomes. Therefore, tRNA expression analysis would need to include tRNA sequences that still retain 5' and 3' regions to accurately determine which tRNA is expressed from the neochromosome and the native chromosomes. To conduct this analysis, we utilized MSR-seq to look at precursor tRNA genes from both the tRNA neochromosome and native chromosome. This will allow for the determination of whether tRNA neochromosome expression occurs as well as whether there is an impact on endogenous tRNA expression in the presence of the tRNA neochromosome.

To conduct this analysis, we utilized MSR-seq because of its ability to capture small RNA sequences and high-throughput ability as previously described. For this analysis it should be noted that there are five major samples: 1) S1 which is a haploid native yeast cell; 2) S2 which is a circular tRNA neochromosome; 3) S3 which is the diploid native yeast cell; 4) S4 which is the linear tRNA neochromosome. MSR-seq was able to get good coverage of 5' and 3' precursors, indicating that it can detect both native and neochromosome precursor tRNA.

(Figure 5.2.2b) Precursor analysis of all samples shows that the tRNA abundance of yeast that have tRNA neochromosome compared to native yeast are correlated, **(Figure 5.2.2a)**. Further comparison of upstream and downstream tRNA precursors between native and neochromosome showed a correlation of $r^2 = 0.89-0.92$ and $r^2 = 0.55-0.69$, respectfully. In summary, this indicates that tRNA expression global expression is not disrupted by the tRNA neochromosome expression. Moving to specific tRNA genes, it was previously shown by northern blot analysis that 5' precursor tRNA^{SerCGA} may be deficient in synthetic chromosomes within yeast. In this

study, using MSR-seq, no difference of tRNA^{SerCGA} expression was observed. Due to lack of greatly increased precursor tRNA being observed, the tRNA neochromosome does not seem to possess maturation deficiencies.

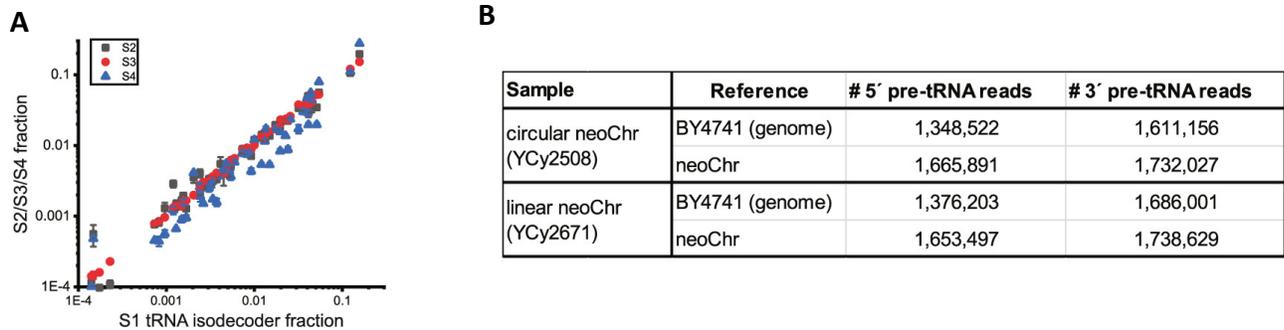


Figure 5.2.1 MSR-seq tRNA expression analysis. (a) MSR-seq shows similar results for pre-tRNA abundances. Correlation of mature tRNA expression at the isoacceptor level. S1: BY4741 (YCy2409); S2: circular tRNA neochromosome (YCy2508); S3: 2n BY4741 (YCy2649); S4: linear tRNA neochromosome (YCy2671). R values of the linear fits are 0.991 (S1-S2), 0.997 (S1-S3), 0.888 (S1-S4), 0.988 (S2-S3), 0.920 (S2-S4), and 0.872 (S3-S4). (b) Mapping statistics of reads containing tRNA precursors, either the upstream sequence (up) or downstream sequence (down) of the tRNA gene. Mean read counts from 3 biological replicates for each sample are shown.

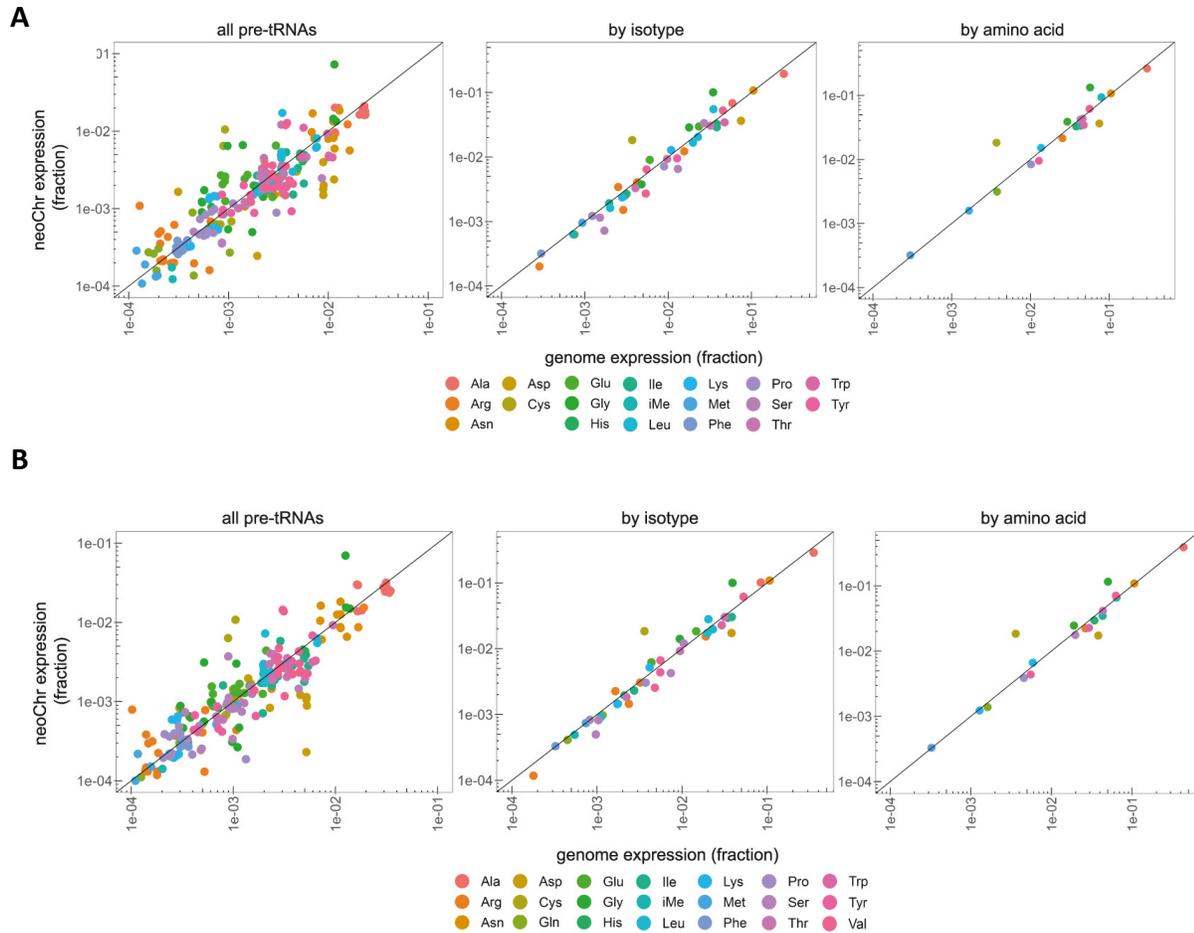


Figure 5.2.2 Correlation of tRNA precursors of native chromosomes and tRNA neochromosomes

(a) Correlation of tRNA precursors of identical tRNA sequences containing the upstream sequence from the native chromosomes (x axis) and the circular tRNA neochromosome (y axis) at the level of individual tRNA genes (left), grouped by isoacceptors (middle), and amino acids (right). For clarity, only upstream (and not downstream) mapped genes are shown. Genes are

colored by amino acids. R values of linear fits are 0.889 (S2 gene), 0.967 (S2 isoacceptor), and 0.956 (S2 amino acid). (b) Correlation of tRNA precursors of identical mature tRNA sequences containing the upstream sequence from the native chromosomes (x axis) and the linear tRNA neochromosome (y axis) at the level of individual tRNA genes (left), grouped by isoacceptors (middle), and amino acids (right). For clarity, only upstream (and not downstream) mapped genes are shown. Genes are colored by amino acids. R values of linear fits are 0.917 (S4 gene), 0.975 (S4 isoacceptor), and 0.961 (S4 amino acid).

5.3 METHODS

5.3.1 MSR-Seq

Yeast total RNA was extracted as described before.^{105,106} Briefly, 5 mL yeast cells were cultured in SC–His medium to OD₆₀₀ of 1.0 prior to harvesting by centrifugation. Cell pellets were resuspended in lysis buffer (10 mM Tris-HCl pH 8.5, 5 mM EDTA, 2% SDS), and mixed with acetate-saturated phenol/CHCl₃ at pH 4.5 (Ambion cat# AM9720). The tubes were vortexed using MP FastPrep-24 5G Homogenizer, with 6 cycles of 30 seconds shaking and 30 seconds cool down. Following cold centrifugation at 18,600 g for 10 minutes, the upper aqueous layers were collected to a new tube and mixed with equal volume of cold acetate-saturated phenol/CHCl₃ solution for another round of extraction. Then, the aqueous layers were collected and mixed with 1/10 volume of 3M sodium acetate pH 5 and 2.5 volume of cold 100% ethanol. RNA was precipitated at -80°C for 30 minutes, then centrifuged at 18,600 g for 10 minutes. RNA pellets were washed with 500 µL of cold 80% ethanol, dried in air, and finally resuspended in water. One µg total RNA from each sample was used to build multiplex small RNA sequencing (MSR-seq) libraries as previously reported. Sequencing was done using Illumina NEXT-seq, 80

bp read1, 150 bp read2. MSR-seq data analysis was performed as described previously.⁴¹ Briefly, starting from index demultiplexed fastq data, paired end reads were split by internal barcode sequence using Je81 demultiplex with options BPOS = BOTH BM = READ 1 LEN = 4:6 FORCE = true C = false 6. Barcode sequences are available on Github at https://github.com/ckatanski/Q_paper.107 Next read 2 files were used to map with bowtie2⁶⁹ with the following parameters: -q -p 10 -local --no-unal. Reads were mapped to curated list of non-redundant tRNA genes using custom yeast gene references from both BY4741 native chromosomes and neo-chromosome. Bowtie2 output sam files were converted to bam files, then sorted using samtools. Next, IGV⁷⁹ was used to collapse reads into 1 nt window. IGV output.wig files were reformatted using custom python scripts (available on GitHub at https://github.com/ckatanski/Q_paper). The bowtie2 output Sam files were also used as input for a custom python script using PySam, a python wrapper for SAMTools to sum all reads that mapped to each gene. Data were visualized with custom R scripts. All custom scripts are available on GitHub (<https://github.com/ckatanski/CHRIS-seq>).

5.4 DISCUSSION

Within this study, a tRNA neochromosome was designed, expressed, and characterized within *S. cerevisiae* cells. By using MSR-seq to analyze precursor tRNA from either the tRNA neochromosome or native yeast chromosomes, we were able to characterize the expression of tRNA from the neochromosome as well as any impact on endogenous tRNA expression. Precursor tRNA abundance from the neochromosome had a positive correlation with native tRNA precursors, suggesting that the tRNA neochromosome did not negatively impact global expression levels. Continuing this analysis of, we saw that there was not significant increase of

precursor tRNA from the tRNA neochromosome, indicating that a global processing defect was not observed. This is contrary to previous analysis looking at synthetic genome expression, which showed a increase of 5' tRNA^{SerCGA}. This may be due to the 5' and 3' regions of the tRNA neochromosome are influencing RNAse P processing, with previous work suggesting that complementary pairing between 5' and 3' sequences affecting RNAase P cleavage (Ziehler et al 2000)

In summary, this study showed that a fully synthetic tRNA neochromosome is functional and tolerable within yeast while meeting the major goals of Sc2.0. With the presence of *rox* recombination sites, this tRNA neochromosome can be utilize to further study how tRNA genomes function within yeast, such as the impact of nucleosome positioning based on tRNA gene expression. This tRNA neochromosome also shows as an excellent proof-of-concept that Sc2.0 designer structures can aid the dramatic re-designing of cellular machinery. This greatly increase the possibility of use in basic biological research as well as industrial applications.

CHAPTER 6

FUTURE DIRECTIONS AND PERSPECTIVES

6.1 *IN VIVO* STRUCTURAL MAPPING OF OTHER SMALL RNA

In Chapters 2 and 3, I discussed DM-DMS-MaPseq, a method to study small RNA structure *in vivo*, and its major results. This work allowed us to determine the tRNA structure and interactome *in vivo*. We found supporting evidence to that eukaryotic EF1A-aa-tRNA maintains a binding paradigm like bacterial EF-Tu, which was previously unknown. In addition, we validated the *in vivo* structures of mitochondrial tRNA, which were known to use non-canonical tRNA tertiary interactions to form tRNA-like structures *in vitro*. In addition to EF1A interaction, we characterized aaRS and ribosome interactions *in vivo*, which supports previous models of aaRS anticodon recognition and ribosome usage. Upon oxidative stress, we characterized global tRNA structure changes. Unexpectedly, we found that ribosome usage is dependent on the sequence and modification state of nucleotide 37 located 3' adjacent to the anticodon nucleotides upon oxidative stress.

Our study primarily focused on tRNA, but DM-DMS-MaPseq data also contain other small RNA, such as Y RNA and snRNA. Y RNA was first discovered 40 years ago and is conserved amongst eukaryotes; however, its exact function remains to be elucidated. (Guila et al 2020) Recent work has determined that Y RNA structures mediate what proteins it binds and interacts with, *in vitro* structural analysis shows that it can fold into multiple distinct structures. (Kowalski et al 2015, Van Gelder et al 1994) It has been shown that Y RNA influences gene expression via its protein interactions, therefore a further characterization of its structure and interactome *in vivo* would be insightful. SnRNAs are essential spliceosome components which

are required for mRNA splicing in cells. A further direction is to use DM-DMS-MaPseq to characterize the *in vivo* interactome and structurome of these small RNA. As the RNA binding proteins (RBP) of these small RNA are well described; we could add an additional step to DM-DMS-MaPseq which involves specific RBP pulldown to only look at the small RNA interacting with these proteins prior to DMS and DM treatment. This step can also validate potentially unique structures of Y RNA and snRNA during interactions with these RBPs.

Recent work by Gao et al, gave evidence for mechanism of maintaining tRNA anticodon pool upon cell differentiation. (Gao et al 2024) Upon differentiation, select tRNA genes are repressed. Thus, tRNA isodecoder abundances changed greatly upon differentiation but at the global level the anticodon pool remained the same as undifferentiated cells. This study shows that tissue-specific tRNA regulation plays a major role and may have other mechanism besides simple repression. DM-DMS-MaPseq could provide insights of the interactome and structurome of tRNA isodecoders from different tissues and cell lines, such as stem cells that have not differentiated versus differentiated cells and determine whether specific tRNA-protein/ribosome interactions can be identified.

6.2 SELECTIVE PACKAGING OF HOST RNA IN VIRAL PARTICLES

In Chapter 4, I discussed selective RNA packaging of SARS-CoV-2 viral particles. This study was a successful proof-of-concept of utilizing MSR-seq to determine whether any host RNA, in this case Vero cells, small RNA was selectively packaged. The major result was that we identified several host cell cytosolic tRNA that increased abundance in the SARS-CoV-2 viral particles as well as the status of host tRNA modifications in selective packaging.

I propose to conduct a similar study in human cell lines instead of Vero cells, which are derived from Green Monkeys, to further determine whether human tRNAs are selectively packaged as well in SARS-CoV-2 virions. In addition, this study should be conducted on multiple SARS-CoV-2 variants to aid in the analysis of whether genetic variance of the virus impacts selective packaging of host small RNA and could aid in the determination of any potentially conserved packaging elements in host small RNA.

The role and function of these virus packaged host RNA remain a major question. Besides the priming of HIV-1 reverse transcription, well described functions for packaged host RNA are lacking. Packaged host RNA could be used by the host cell to prime innate immune response in future infected cells or be used by viruses to evade innate immune responses upon infection of other cells. For either of these models to be tested, new techniques may have to be developed to identify infected cells by viruses that bring host small RNAs obtained through previous infection, so that their associated immune response can be measured.

REFERENCES

- Asano, K., Suzuki, T., Saito, A., Wei, F. Y., Ikeuchi, Y., Numata, T., Tanaka, R., Yamane, Y., Yamamoto, T., Goto, T., Kishita, Y., Murayama, K., Ohtake, A., Okazaki, Y., Tomizawa, K., Sakaguchi, Y., & Suzuki, T. (2018). Metabolic and chemical regulation of tRNA modification associated with taurine deficiency and human disease. *Nucleic Acids Research*, *46*(4), 1565–1583. <https://doi.org/10.1093/NAR/GKY068>
- Björk, G. R., Wikström, P. M., & Byström, A. S. (1989). Prevention of Translational Frameshifting by the Modified Nucleoside 1-Methylguanosine. *Science*, *244*(4907), 986–989. <https://doi.org/10.1126/SCIENCE.2471265>
- Boccaletto, P., MacHnicka, M. A., Purta, E., Pitkowski, P., Baginski, B., Wirecki, T. K., de Crécy-Lagard, V., Ross, R., Limbach, P. A., Kotter, A., Helm, M., & Bujnicki, J. M. (2018). MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Research*, *46*(D1), D303–D307. <https://doi.org/10.1093/NAR/GKX1030>
- Cantara, W. A., Murphy IV, F. v., Demirci, H., & Agris, P. F. (2013). Expanded use of sense codons is regulated by modified cytidines in tRNA. *National Acad Sciences WA Cantara, FV Murphy IV, H Demirci, PF Agris Proceedings of the National Academy of Sciences, 2013•National Acad Sciences*, *110*(27), 10964–10969. <https://doi.org/10.1073/pnas.1222641110>
- Cech, T. R., & Steitz, J. A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, *157*(1), 77–94. <https://doi.org/10.1016/J.CELL.2014.03.008>
- Chan, C. T. Y., Pang, Y. L. J., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., & Dedon, P. C. (2012a). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications 2012 3:1*, *3*(1), 1–9. <https://doi.org/10.1038/ncomms1938>
- Chan, C. T. Y., Pang, Y. L. J., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., & Dedon, P. C. (2012b). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications*, *3*. <https://doi.org/10.1038/NCOMMS1938>
- Chan, C. T. Y., Pang, Y. L. J., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., & Dedon, P. C. (2012c). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications 2012 3:1*, *3*(1), 1–9. <https://doi.org/10.1038/ncomms1938>
- Chan, C. W., Badong, D., Rajan, R., & Mondragón, A. (2020). Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA*, *26*(3), 278–289. <https://doi.org/10.1261/RNA.073478.119/-/DC1>

- Chan, P. P., & Lowe, T. M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, *44*(D1), D184–D189. <https://doi.org/10.1093/NAR/GKV1309>
- de Crécy-Lagard, V., Boccaletto, P., Mangleburg, C. G., Sharma, P., Lowe, T. M., Leidel, S. A., & Bujnicki, J. M. (2019). Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Research*, *47*(5), 2143–2159. <https://doi.org/10.1093/NAR/GKZ011>
- Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2013). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* *2013* *505*:7485, *505*(7485), 696–700. <https://doi.org/10.1038/nature12756>
- Frazer-Abel, A. A., & Hagerman, P. J. (2008). Core flexibility of a truncated metazoan mitochondrial tRNA. *Nucleic Acids Research*, *36*(17), 5472. <https://doi.org/10.1093/NAR/GKN529>
- Giegé, R. (2008). Toward a more complete view of tRNA biology. *Nature Structural & Molecular Biology* *2008* *15*:10, *15*(10), 1007–1014. <https://doi.org/10.1038/nsmb.1498>
- Han, L., & Phizicky, E. M. (2018). A rationale for tRNA modification circuits in the anticodon loop. *RNA*, *24*(10), 1277–1284. <https://doi.org/10.1261/RNA.067736.118>
- Hopper, A. K., & Phizicky, E. M. (2003). tRNA transfers to the limelight. *Genes & Development*, *17*(2), 162–180. <https://doi.org/10.1101/GAD.1049103>
- Ibba, M., & Soll, D. (2000). Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry*, *69*(Volume 69, 2000), 617–650. <https://doi.org/10.1146/ANNUREV.BIOCHEM.69.1.617/CITE/REFWORKS>
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., & Segal, E. (2010a). *LETTERS Genome-wide measurement of RNA secondary structure in yeast*. 467. <https://doi.org/10.1038/nature09322>
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., & Segal, E. (2010b). *LETTERS Genome-wide measurement of RNA secondary structure in yeast*. 467. <https://doi.org/10.1038/nature09322>
- Korostelev, A. A. (2022). The Structural Dynamics of Translation. *Annual Review of Biochemistry*, *91*(Volume 91, 2022), 245–267. <https://doi.org/10.1146/ANNUREV-BIOCHEM-071921-122857/CITE/REFWORKS>

- Laptev, I., Dontsova, O., & Sergiev, P. (2020). Epitranscriptomics of Mammalian Mitochondrial Ribosomal RNA. *Cells* 2020, Vol. 9, Page 2181, 9(10), 2181. <https://doi.org/10.3390/CELLS9102181>
- Liu, Z., Zhang, Y., & Nielsen, J. (2019). Synthetic Biology of Yeast. *Biochemistry*, 58(11), 1511–1520. https://doi.org/10.1021/ACS.BIOCHEM.8B01236/ASSET/IMAGES/MEDIUM/BI-2018-01236T_0004.GIF
- Lorenz, C., Lünse, C. E., & Mörl, M. (2017). tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* 2017, Vol. 7, Page 35, 7(2), 35. <https://doi.org/10.3390/BIOM7020035>
- Machnicka, M. A., Olchowik, A., Grosjean, H., & Bujnicki, J. M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biology*, 11(12), 1619–1629. <https://doi.org/10.4161/15476286.2014.992273>
- Ojala, D., Montoya, J., & Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 1981 290:5806, 290(5806), 470–474. <https://doi.org/10.1038/290470a0>
- Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Research* 2018 28:4, 28(4), 395–404. <https://doi.org/10.1038/s41422-018-0013-y>
- Redden, H., & Alper, H. S. (2015). The development and characterization of synthetic minimal yeast promoters. *Nature Communications*, 6. <https://doi.org/10.1038/NCOMMS8810>
- Rocha, E. P. C. (2004). Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*, 14(11), 2279–2286. <https://doi.org/10.1101/GR.2896904>
- Sokołowski, M., Klassen, R., Bruch, A., Schaffrath, R., & Glatt, S. (2017). *Cooperativity between different tRNA modifications and their modification pathways* <https://doi.org/10.1016/j.bbagrm.2017.12.003>
- Suzuki, T., Yashiro, Y., Kikuchi, I., Ishigami, Y., Saito, H., Matsuzawa, I., Okada, S., Mito, M., Iwasaki, S., Ma, D., Zhao, X., Asano, K., Lin, H., Kirino, Y., Sakaguchi, Y., & Suzuki, T. (n.d.). *Complete chemical structures of human mitochondrial tRNAs*. <https://doi.org/10.1038/s41467-020-18068-6>
- Torres, A. G. (2019a). Enjoy the Silence: Nearly Half of Human tRNA Genes Are Silent. *Bioinformatics and Biology Insights*, 13. https://doi.org/10.1177/1177932219868454/ASSET/IMAGES/LARGE/10.1177_1177932219868454-FIG1.JPEG

- Torres, A. G. (2019b). Enjoy the Silence: Nearly Half of Human tRNA Genes Are Silent. *Bioinformatics and Biology Insights*, 13. https://doi.org/10.1177/1177932219868454/ASSET/IMAGES/LARGE/10.1177_1177932219868454-FIG1.JPEG
- Torres, A. G., Reina, O., Attolini, C. S. O., & de Pouplana, L. R. (2019). Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proceedings of the National Academy of Sciences of the United States of America*, 116(17), 8451–8456. <https://doi.org/10.1073/PNAS.1821120116/-/DCSUPPLEMENTAL>
- Wang, Y., Tao, E.-W., Tan, J., Gao, Q.-Y., Chen, Y.-X., Fang, J.-Y., & Chen, Y.-X. (2023). tRNA modifications: insights into their role in human cancers *Cell Biology. Trends in Cell Biology*, 33(12). <https://doi.org/10.1016/j.tcb.2023.04.002>
- Yamagami, R., Sieg, J. P., Assmann, S. M., & Bevilacqua, P. C. (2022). Genome-wide analysis of the in vivo tRNA structurome reveals RNA structural and modification dynamics under heat stress. *Proceedings of the National Academy of Sciences of the United States of America*, 119(25). <https://doi.org/10.1073/PNAS.2201237119/-/DCSUPPLEMENTAL>
- Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., & Miyazawa, T. (1985). Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proceedings of the National Academy of Sciences*, 82(15), 4905–4909. <https://doi.org/10.1073/PNAS.82.15.4905>
- Yoshida, M., Kataoka, N., Miyauchi, K., Ohe, K., Iida, K., Yoshida, S., Nojima, T., Okuno, Y., Onogi, H., Usui, T., Takeuchi, A., Hosoya, T., Suzuki, T., & Hagiwara, M. (2015). Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proceedings of the National Academy of Sciences of the United States of America*, 112(9), 2764–2769. https://doi.org/10.1073/PNAS.1415525112/SUPPL_FILE/PNAS.201415525SI.PDF
- Zaborske, J. M., Bauer DuMont, V. L., Wallace, E. W. J., Pan, T., Aquadro, C. F., & Drummond, D. A. (2014). A Nutrient-Driven tRNA Modification Alters Translational Fidelity and Genome-wide Protein Coding across an Animal Genus. *PLOS Biology*, 12(12), e1002015. <https://doi.org/10.1371/JOURNAL.PBIO.1002015>
- Zhang, W., Foo, M., Eren, A. M., & Pan, T. (2022). tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. *Molecular Cell*, 82(5), 891–906. <https://doi.org/10.1016/J.MOLCEL.2021.12.007>