I, Susan P. Russell, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Chemistry.

It is entitled: Characterizing Modified Nucleosides in RNA by LC/UV/MS

Student's name: Susan P. Russell

This work and its defense approved by:

Committee chair: Patrick Limbach, PhD
Committee member: Bruce Ault, PhD
Committee member: Joseph Caruso, PhD
I, Susan P Russell, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Chemistry.

It is entitled:
Characterizing Modified Nucleosides in RNA by LC/UV/MS

Student's name: Susan P Russell

This work and its defense approved by:

Committee chair: Patrick Limbach, PhD
Committee member: Bruce Ault, PhD
Committee member: Joseph Caruso, PhD
Characterizing Modified Nucleosides in RNA by LC/UV/MS

A dissertation submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy (PhD)

in the Department of Chemistry
of the McMicken College of Arts and Sciences

by

Susan Pitsenberger Russell

B.S. Chemistry, Centre College. 1983
B.S. Economics, Centre College, 1983

December 2012

Committee Chair: Patrick A. Limbach, PhD
Abstract

Posttranscriptional nucleoside modification is an important characteristic in all types of ribonucleic acids (RNA) and increases the functional diversity of RNA, exceeding that of the four canonical bases. In ribosomal RNA (rRNA) and transfer RNA (tRNA) these modifications enhance RNA structure and contribute to the accuracy and efficiency of protein translation. These posttranscriptional modifications are dynamic and function cooperatively within the cell and also exert regulatory influences. The goal of this dissertation is to identify and quantify modified nucleosides in RNA using optimized LC/MS methods of analysis. This method was shown to be sufficient for characterizing changes in the abundance of modified nucleosides as a function of varying cellular conditions and can be used to explore the dynamics of RNA modification. These methods were optimized and characterized, then used to compare the global modification status of tRNA and rRNA under normal and heat shock conditions.

It was also demonstrated how this LC/UV/MS method can be used to further understand the functional significance and cellular dynamics of RNA modification by applying these methods as an assay for functional characterization of modifying enzymes. A survey and comparison of RNA modifications present in unfractionated tRNAs from bacterial and archaeal organisms is also presented in this dissertation. Studying modification differences among organisms, especially extremophiles that live in harsh conditions, provides clues to how modifications function to enhance structural stability of RNA. RNA modification guides the decoding process in tRNA and organisms of different kingdoms use slightly different modifications for decoding. A new nucleoside, agmatidine, was structurally characterized using LC/UV/MS methods. This nucleoside modification, found only in Archaea, is responsible for specific decoding properties of a rare tRNA and present in very low abundance.
The methods presented in this dissertation can be used for quantitative comparison of changes in modified nucleoside composition. By establishing the overall method and sample variability for measuring modified nucleosides from tRNA and rRNA, without resorting to stable isotope labeling methods, we are now able to define the statistical significance of changes in modification status when comparing biological systems of varying environmental or physiological cellular conditions, which should enable future studies aimed at measuring relative changes in modification profiles under these experimental conditions.
Acknowledgements

To my advisor, I appreciate tremendously your expert guidance and unending patience during my graduate school experience. It was truly a privilege to work with you. I would also like to thank my committee members Dr. Ault and Dr. Caruso for their advice and guidance. To Stephen and Larry, who guided much of my development in the field of mass spectrometry by teaching me so much about instrumentation. Thank you, to the members of the Limbach group for the fun and support.

To my husband and family, I thank you for the love and support throughout my graduate school experience and beyond.
# TABLE OF CONTENTS

Abstract                                                                     ii  
Acknowledgements                                                            v  
List of Figures                                                              3  
List of Tables                                                               7  
**Chapter 1 Introduction**  
1.1 Research Goal                                                          9  
1.2 Introduction                                                            9  
1.3 Posttranscriptional Modifications                                       11  
1.4 Modifying Enzymes                                                      17  
1.5 Cooperative Effects Among Modifications                                 18  
1.6 Influence of Growth Temperature on Posttranscriptional Modification    19  
1.7 Heat Shock and Posttranscriptional Modifications                      20  
1.8 Why Determine Changes in Global Modification                          21  
1.9 Analytical Methods of Analysis                                         22  
**Chapter 2 Total Census of Posttranscriptional Modifications in tRNA from Selected Bacterial and Archaeal Organisms**  
2.1 Introduction                                                            27  
2.2 Experimental                                                            28  
2.3 Results                                                                 30  
2.4 Discussion                                                              42  
2.5 Conclusions                                                             76  
**Chapter 3 Identification of an Unknown Modified Nucleoside by LC/UV/MS**  
3.1 Introduction                                                            78  
3.2 Experimental                                                            79  
3.3 Results                                                                 80
# List of Figures

**Figure 1.1**  RNA structure.  
12

**Figure 1.2**  Structure of modified nucleosides.  
13

**Figure 2.1**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Escherichia coli* tRNA.  
33

**Figure 2.2**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Bacillus subtilis* tRNA.  
33

**Figure 2.3**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Lactococcus lactis* tRNA.  
34

**Figure 2.4**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Thermus thermophilus* tRNA.  
34

**Figure 2.5**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Haloarcula marismortui* tRNA.  
35

**Figure 2.6**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Methanococcus maripaludis* tRNA.  
35

**Figure 2.7**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Sulfolobus sofataricus* tRNA.  
36

**Figure 2.8**  UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Haloquadratum walsbyi* tRNA.  
36

**Figure 2.9**  Cloverleaf structure.  
41

**Figure 2.10**  Structures of modified nucleosides found in Dihydrouridine stem and loop domain in tRNA.  \( \Psi = \text{pseudouridine}, \; D = \text{dihydrouridine}, \; m^1A = 1\text{-methyladenosine}, \; \text{Gm} = 2'\text{-O-methylguanosine}, \; s^4U = 4\text{-thiouridine}. \)  
47

**Figure 2.11**  Structures of modified nucleosides found in the variable loop domain.  \( m^7G = 7\text{-methylguanosine}, \; \text{acp}^3U = 3\text{-}(3\text{-amino-3-carboxypropyl)uridine}. \)  
52

**Figure 2.12**  Structures of modified nucleosides found in T\( \Psi \)C loop domain in tRNA.  \( m^3U = 5\text{-methyluridine}, \; m^5s^2U = 5\text{-methyl-2-thiouridine}. \)  
53

**Figure 2.13**  Structure of modified nucleoside found in the acceptor stem domain.  \( m^2G = N^2\text{-methylguanosine}. \)  
56

**Figure 2.14**  Structures of modified nucleosides present at position 32 in tRNA.  \( \text{Cm} = 2'\text{-O-methylcytidine}, \; \text{Um} = 2'\text{-O-methyluridine}, \; \Psi = \text{pseudouridine}, \; s^2C = 2\text{-thiocytidine}. \)  
59
Figure 2.15 Structures of two types of modified uridines commonly found at position 34 in tRNA of bacterial species. a) $\gamma^2$-substituted uridines ($\gamma^2$U), $\gamma^2\gamma^2$U = 5-methylaminomethyluridine, $\gamma^2\gamma^5$U = 5-methylaminomethyl-2-thiouridine, $\gamma^2\gamma^5\gamma^2$U = 5-carboxymethylaminomethyl-2-thiouridine, $\gamma^2\gamma^5\gamma^2\gamma^2$Um = 5-carboxymethylaminomethyl-2'-O-methyluridine and b) $\gamma^4$-methoxy uridines ($\gamma^4$U), $\gamma^4$U = 5-methoxyuridine, $\gamma^4\gamma^2$U = 5-carboxymethylaminomethyl-2'-O-methyluridine.

Figure 2.16 Structure of other modified nucleosides at the Wobble position in tRNA. I = inosine, Q = queuosine, $k^2$C = lysidine, ac$^4$C = 4-acetylcytidine.

Figure 2.17 Structure of modified nucleoside often present at position 37 in tRNA. $m^1$G = 1-methylguanosine.

Figure 2.18 Structures of $N^6$-threonylcarbamoyladenosine ($t^6$A) and derivatives found in tRNA. $m^2t^6$A = $N^6$-threonylcarbamoyladenosine-2-methylthioadenosine, $m^6t^6$A = $N^6$-methyl-$N^6$-threonylcarbamoyladenosine.

Figure 2.19 Structures of $N^6$-isopentenyladenosine ($i^6$A) and derivatives. $m^2i^6$A = 2-methylthio-$N^6$-isopentenyladenosine, $i^6$A = $N^6$-(cis-hydroxyisopentenyl)adenosine, $m^2i^6$A = 2-methylthio-$N^6$-(cis-hydroxyisopentenyl)adenosine.

Figure 2.20 Modified nucleosides unique to Archaea. G+ = archaeosine, $m^1$I = 1-methylpseudouridine, $m^1$I = 1-methylinosine, $m^2$G = $N^2$, $N^2$-dimethylguanosine.

Figure 2.21 Structures of wyosine and derivatives. imG = wyosine, mimG = methylwyosine, imG-14 = 4-demethylwyosine, imG2 = isowyosine.

Figure 3.1 Sequence of tRNA$^{Ile}_2$ from *H. marismortui*. C* denotes unknown modified cytidine at position 34.

Figure 3.2 LC/MS analysis of nucleosides present in *H. marismortui* tRNA$^{Ile}_2$.

Figure 3.3 Mass spectral analysis of unknown modified cytidine (C*). Selected ion chromatograms of (A) the molecular ion at $m/z$ 356, [MH]$^+$ and (B) the base ion at $m/z$ 224, [BH2]$^+$. (C) Mass spectrum of the modified cytidine, C*.

Figure 3.4 High resolution mass spectral analysis of unknown modified cytidine (C*). Selected ion chromatograms of (A) the molecular ion at $m/z$ 356, [MH]$^+$ and the base ion at $m/z$ 224, [BH2]$^+$. High resolution mass spectra of of (B) the molecular ion at $m/z$ 356, [MH]$^+$ and (C) the base ion at $m/z$ 224, [BH2]$^+$.

Figure 3.5 Collision induced dissociation of 5-methylcytidine. (A) No fragmentation of base ion, $m/z$ 126. CE = 23%, Q = 0.25. (B) Fragmentation of base ion, $m/z$ 126 but ion yield low, CE = 35%, Q = 0.25. (C) Fragmentation of base ion, $m/z$ 126 with high ion yield, CE = 35%, Q = 0.38.
Figure 3.6  Identification of unknown modified cytidine (C*).  (A) Accurate mass analysis of CID MS/MS product ions of unknown base ion, m/z 224.  (B) Low resolution CID MS/MS of unknown base ion, m/z 224.  Minor ions of m/z 114, m/z 153, m/z 182 likely due to side chain cleavages of agmatidine moiety.

Figure 3.7  Low resolution CID MS/MS of agmatine standard.  Both unknown and agmatine exhibit loss of NH$_3$ and CH$_3$N$_3$.

Figure 3.8  Structure of modified nucleoside at C34 in tRNAlle2.

Figure 3.9  Selected ion chromatograms and mass spectra; Sulfolobus solfataricus, A-C and Methanococcus maripaludis, D-E.

Figure 3.10  Predicted base pairing between adenosine: agmatidine and adenosine: lysidine.

Figure 4.1  Structures of hypermodified adenosines in B. subtilis tRNA.  (t$^6$A, N$^6$-threonylcarbamoyladenosine.  ms$^2$t$^6$A, 2-methylthio-N$^6$-threonylcarbamoyladenosine.  i$^6$A, N$^6$-isopentenyladenosine.  ms$^2$i$^6$A, 2-methylthio-N$^6$-isopentenyladenosine.)

Figure 4.2  Modifying enzymes involved in the biosynthesis of hypermodified adenosines of B. subtilis.

Figure 4.3  UV chromatogram of wild type B. subtilis, total tRNA digested to nucleosides.

Figure 4.4  (a) Extracted ion chromatograms of molecular and base ions and (b) mass spectra of the adenosine derivatives t$^6$A, ms$^2$t$^6$A, i$^6$A, and ms$^2$i$^6$A from digested wild type B. subtilis total tRNA.

Figure 4.5  UV Chromatograms of B. subtilis total tRNA nucleoside digests, highlighting hypermodified adenosines t$^6$A, ms$^2$t$^6$A, i$^6$A and ms$^2$i$^6$A.  A) Wild type strain 168, B) strain BSF2608 and C) strain YQEvd.  Absence of peak is denoted in parentheses.  Non-nucleoside peak denoted with asterisk.

Figure 5.1  UV chromatogram of nucleoside test mix, 4.6 x 250 mm / LC-18S column, 30 °C, 250 mM ammonium acetate pH 6.0 (A) 40% acetonitrile in water (B), 2.0 mL/min.  UV detection @ 260 nm.

Figure 5.2  UV chromatogram of nucleoside test mix, 2.1 x 250 mm / LC-18S column, 30 °C, 5 mM ammonium acetate pH 5.3 (A) 40% acetonitrile in water (B), 0.3 mL/min.  UV detection @ 260 nm.  (a) 20 °C and (b) 30 °C.

Figure 5.3  Modified nucleosides in Escherichia coli total tRNA.

Figure 5.4  (a) UV chromatogram of E. coli 16S RNA nucleoside digest, 4.6 x 250 mm LC-18S column, 30°C, 250 mM ammonium acetate pH 6.0 (A), 40% acetonitrile in water (B), 2.0 mL/min, UV detection @ 260 nm.  (b) UV chromatogram of E. coli 16S RNA nucleoside digest, 2.1 mm x 250 mm, LC-18S column, 30°C, 5 mM ammonium acetate pH 5.3 (A), 40% acetonitrile in water (B), 0.3 mL/min, UV detection @ 260 nm.

Figure 6.1  Experimental Outline.
Figure 6.2  Growth Curve for *E. coli* grown at 30 °C.

Figure 6.3  Growth Curves for Normal (NHS) and Heat Shocked (HS) *E.coli* culture. Arrow indicates half the culture transferred to 42 °C.

Figure 6.4  Cellular proteins in cell lysate from *E. coli* cultures. A) Polyacrylamide gel (10% PAGE) of cell lysates. Lane 1: cell lysate of normal culture. Lane 2: cell lysate of heat shocked culture. Lane 3: protein marker. B) Western blot for DnaK. Lane 1 and 2: cell lysate of normal culture. Lane 3 and 4: cell lysate of heat shocked culture. Lane 5: protein marker.

Figure 6.5  Purity of rRNA and tRNA isolated from *E. coli* cell cultures. A) 1% Agarose Gel. Lane 1: 23S RNA isolated from 50S ribosomal subunit. Lane 2 and 3: 16S RNA isolated from 30S ribosomal subunits. Lane 2, normal culture; Lane 3, heat shocked culture. B) 2% Agarose Gel. Lane 1: total tRNA marker. Lane 2 and 3: tRNA isolated from cell culture. Lane 2, normal culture; Lane 3, heat shock culture. Lane 4: total tRNA marker.

Figure 6.6  UV Chromatogram from LC/MS analysis of an enzymatic digest of *E.coli* 16S rRNA.

Figure 6.7  UV Chromatogram from LC/MS analysis of an enzymatic digest of *E.coli* tRNA.

Figure 6.8  Comparison of average relative amounts of modified nucleosides present in 16S rRNA from heat shocked and normal cell cultures.

Figure 6.9  Comparison of average relative amounts of modified nucleosides present in tRNA from heat shocked and normal cell cultures.
List of Tables

Table 1.1  Modifications present in 16S rRNA, with sequence location and modifying enzyme noted.  

Table 1.2  Modifications present in 23S rRNA, with sequence location and modifying enzyme noted.  

Table 1.3  Modifications in *E. coli* transfer RNA, with sequence location and enzyme noted.  

Table 2.1  Organisms used in this study.  

Table 2.2  HPLC Gradient used for analysis of nucleoside digest.  

Table 2.3  Comparison of nucleosides identified in tRNA from selected organisms.  

Table 2.4  Modifications in Tertiary Core of tRNA.  

Table 2.5  Modifications in the Anticodon Stem Loop Domain of tRNA.  

Table 4.1  Distribution of hypermodified adenosines *B. subtilis* tRNAs. (Q, queuosine. moU, 5-methoxyuridine. Gm, 2′-O-methylguanosine. cmnm'sU, 5-carboxymethylaminomethyl-2-thiouridine.)  

Table 4.2  Modified nucleosides observed in *B. subtilis* strains. For the genotypes, (+) indicates intact genes, (-) indicates insertion mutants, (+P) is plasmid encoded gene copy and (?P) indicates plasmid encoded heterologous gene copy that may or may not complement gene in question. Phenotypes indicate the presence (+) or absence (-) of modified nucleosides based on LC/UV/MS analysis.  

Table 5.1  Reproducibility of UV peak area and retention time for analysis of modified nucleosides in standard test mix, 4.6 x 250 mm column used.  

Table 5.2  Reproducibility of UV peak areas and retention times for modified nucleoside in standard test mix, 2.1 x 250 mm column used, 20 °C.  

Table 5.3  Reported reproducibilities for analysis of modified nucleosides in urine from other studies using UV or MS detection. MRM denotes multiple reaction monitoring, SIM denotes selected ion monitoring and CNL denotes constant neutral loss.  

Table 5.4  Reproducibility of UV and MS peak areas and retention times for modified nucleosides (both molecular ion [MH+] and base ion [BH2+]) in standard test mix, 2.1 x 250 mm column used, 30°C.  

Table 5.5  Reproducibility of UV and MS peak areas for modified nucleosides (both molecular ion [MH+] and base ion [BH2+]) in *E. coli* tRNA. Pooled percent relative standard deviations are for 3 or 4 replicate analyses within three separate cultures. For MS peak area %RSD 4 replicate analyses were performed on tRNA isolated from a single culture. (* denotes isobaric molecular ion and “nd” denotes base ion not detected in MS analysis)
**Table 5.6** Biological Reproducibility for analysis of modified nucleosides in *E. coli* 16S rRNA, 4.6 x 250 mm column used. Amounts of modified nucleosides present are based on UV peak areas relative to m^6^2A. Percent relative standard deviations in parentheses are for three replicate analyses within each separate culture.

**Table 5.7** Biological Reproducibility for analysis of modified nucleosides in *E. coli* 16S RNA, 2.1 x 250 mm column used. Amounts of modified nucleosides present are based on UV peak areas relative to m^6^2A. Percent relative standard deviations in parentheses are for four replicate analyses within each separate culture.

**Table 6.1** Comparison of relative amounts of modified nucleosides in rRNA in three different cell cultures.

**Table 6.2** Comparison of relative amounts of modified nucleosides in unfractionated tRNA in three different cell cultures.
CHAPTER 1 Introduction

1.1 Research Goal

The goal of this dissertation is to identify and quantify modified nucleosides in RNA by using optimized Liquid Chromatography/Mass Spectrometry (LC/MS) methods of analysis. The posttranscriptional modifications present in total unfractionated tRNA were identified and compared for selected bacterial and archaeal species. Liquid chromatography tandem mass spectrometry (LC/MS/MS) was used to identify a previously uncharacterized modified nucleoside, agmatidine, which was structurally very different from currently known modifications. LC/MS/MS was used to correlate the presence or absence of modified nucleosides in total transfer RNA (tRNA) to identify specific modifying enzymes responsible for methylthio modifications present in Bacillus subtilis tRNA. An LC/MS method for determining the global changes in modification status and the degree of modification in rRNA and tRNA was developed and applied to a bacterial system under heat shock conditions. Investigating how the posttranscriptional modifications change in response to cellular stress may lead to a greater understanding of the structural and functional significance of these modifications in the ribosome and in tRNA.

1.2 Introduction

RNA (ribonucleic acid) plays a pivotal role in the transformation of genetic information from DNA (deoxyribonucleic acid) to functioning protein as dictated in the central dogma of molecular biology proposed by Crick. [1, 2] Both DNA and RNA are biopolymers composed of nucleotides joined together by sugar-phosphate linkages. Nucleotides are phosphate esters of a five carbon sugar with a nitrogenous base attached to the sugar. The sugar is a deoxyribose in DNA and ribose in RNA. The four canonical or nitrogenous bases found in nucleic acids are adenine, cytosine, guanine and thymine in DNA (uracil in RNA). A nucleotide without the phosphate group is a nucleoside. RNA structure is shown below. (Figure 1.1)
DNA exists primarily in a double-stranded helical form where adenine base pairs with thymine through the formation of two hydrogen bonds and guanine base pairs with cytosine forming three hydrogen bonds in opposite DNA strands. [3] In contrast, RNA usually occurs as a single-stranded polynucleotide resulting in a flexible backbone that allows formation of a variety of different structures.

The sequence or order of nucleotides encodes the genetic information. During transcription the genetic information or nucleotide sequence contained in DNA is used as a template to synthesize RNA by the action of RNA polymerase. The RNA produced is either a protein coding messenger RNA (mRNA), or is non-coding RNA. Proteins are synthesized according to the sequence encoded in mRNA on the ribosome during the process of translation. Non-coding RNAs are not translated into proteins but play many other important cellular roles. Ribosomal RNA (rRNA) and transfer RNA (tRNA), for example, form the core of the translation apparatus. An important characteristic of RNA is the presence of post-transcriptional modifications in all types of RNA, which increases the structural and functional diversity
of RNA beyond the four canonical bases. The ability of single-stranded RNA to form a variety of secondary and tertiary structures and the presence of posttranscriptional modifications that allow alternative hydrogen bonding and base stacking interactions enable RNA to complete its myriad of cellular functions. Additionally, posttranscriptional modifications are well-conserved, a further indication of their importance to the cell.

1.3 Posttranscriptional Modifications

Naturally-occurring chemical modification of the four major RNA nucleosides are found in all types of RNA and are formed posttranscriptionally. More than 81 different modified nucleosides have been identified in tRNA and more than 30 in rRNA. [4] Some of the most common types of modifications are methylations of the base, such as 5-methyluridine (Figure 1.2). Another common modification is pseudouridine, which is produced by isomerization of uridine to pseudouridine (Figure 1.2). The 2’-OH of the ribose is often methylated. Other modifications are produced by deamination, thiolation, alkylation, hydroxylation of the standard four bases. Complex hypermodifications also occur where structural changes are substantial and often require a multi-step enzymatic process, such as queuosine. (Figure 1.2)

Figure 1.2 Structure of modified nucleosides.
1.3.1 Posttranscriptional Modification in Ribosomal RNA

Ribosomal RNA is the most abundant noncoding RNA in the cell and comprises approximately 80% of the RNA in rapidly growing bacterial cells. The bacterial ribosome is a 2.5 M Da ribonucleoprotein complex composed of both proteins and RNA and is the site of protein synthesis in the cell. Sizes of ribonucleoproteins and large RNA are expressed in Svedberg units (S). A Svedberg is a sedimentation coefficient and relates to particle size during a sedimentation process such as centrifugation. The bacterial ribosome, referred to as the 70S particle, is composed of a large 50S subunit and a smaller 30S subunit. In the bacterium, Escherichia coli, the 50S subunit is composed of a 23S rRNA (2904 nucleotides), 5S rRNA (120 nucleotides) and 34 proteins. The 30S subunit contains 16S rRNA (1542 nucleotides) and 21 proteins. The ribosome functions as a ribozyme with the RNA catalyzing the formation of peptide bonds within the peptidyl transferase center, the active site of the ribosome where the peptide bonds are formed between amino acids. [5, 6] The ribosomal proteins are not present near the peptidyl transferase center and are not involved directly in protein synthesis but, rather, provide a structural role and stabilize large-scale folding during ribosome assembly.

In E.coli the 5S rRNA is unmodified. The 16S rRNA contains the 11 modifications shown in Table 1.1. The modifications are all methylations to the base or ribose, with the exception of pseudouridine, an isomerization. The base position refers to location within the 16S RNA sequence. The modifying enzyme responsible for the biosynthesis of the modification is also listed. \( N^4-2'-O\)-dimethylcytidine contains methylations on both the base and the ribose, and two separate enzymes are required for each of these methylations.
Table 1.1. Modifications present in 16S rRNA, with sequence location and modifying enzyme noted.

<table>
<thead>
<tr>
<th>Base Position</th>
<th>Modified Nucleoside</th>
<th>Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>516</td>
<td>Pseudouridine</td>
<td>RsuA</td>
</tr>
<tr>
<td>527</td>
<td>7-methylguanosine</td>
<td>RsmG</td>
</tr>
<tr>
<td>966</td>
<td>N\textsuperscript{2}-methylguanosine</td>
<td>RsmD</td>
</tr>
<tr>
<td>967</td>
<td>5-methylcytidine</td>
<td>RsmB</td>
</tr>
<tr>
<td>1207</td>
<td>N\textsuperscript{2}-methylguanosine</td>
<td>RsmC</td>
</tr>
<tr>
<td>1402</td>
<td>N\textsuperscript{4}, 2\textprime-O-dimethylcytidine</td>
<td>RsmH/RsmI</td>
</tr>
<tr>
<td>1407</td>
<td>5-methylcytidine</td>
<td>RsmF</td>
</tr>
<tr>
<td>1498</td>
<td>3-methyluridine</td>
<td>RsmE</td>
</tr>
<tr>
<td>1516</td>
<td>N\textsuperscript{2}-methylguanosine</td>
<td>RsmJ</td>
</tr>
<tr>
<td>1518</td>
<td>N\textsuperscript{6}, N\textsuperscript{6}-dimethyladenosine</td>
<td>RsmA</td>
</tr>
<tr>
<td>1519</td>
<td>N\textsuperscript{6}, N\textsuperscript{6}-dimethyladenosine</td>
<td>RsmA</td>
</tr>
</tbody>
</table>

The modifications present in 23S RNA are primarily methylations of the base and ribose or pseudouridylation, with the exception of one hydroxylation at C2501 and dihydrouridine at U2449. (Table 1.2)

The posttranscriptional modifications in ribosomal RNA cluster in the functionally important regions of the ribosome and are generally thought to be important in both the structure and function by fine tuning structure to maintain the integrity and efficiency of the translational process as it occurs on the ribosome. [7-9]

The 30S subunit is involved in proper decoding of the message by mediating the correct mRNA codon and tRNA anticodon interactions so that the appropriate amino acid is added to the growing polypeptide chain. The 50S subunit contains the peptidyl transferase center where the RNA catalyzed peptide bond formation occurs. Posttranscriptional modifications have been shown to be important in 1) subunit assembly and maturation [10-15], 2) intersubunit bridging [16], 3) antibiotic resistance [17, 18], 4) decoding fidelity [19, 20] and 5) translation initiation or termination. [21, 22] The precise functional
significance and structural role of the individual, well-conserved, modifications remains poorly understood.

**Table 1.2** Modifications present in 23S rRNA, with sequence location and modifying enzyme noted.

<table>
<thead>
<tr>
<th>Base Position</th>
<th>Modified Nucleoside</th>
<th>Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>745</td>
<td>1-methylguanosine</td>
<td>RlmA</td>
</tr>
<tr>
<td>746</td>
<td>pseudouridine</td>
<td>RluA</td>
</tr>
<tr>
<td>747</td>
<td>5-methyluridine</td>
<td>RlmC</td>
</tr>
<tr>
<td>955</td>
<td>pseudouridine</td>
<td>RluC</td>
</tr>
<tr>
<td>1618</td>
<td>N(^6)-methyladenosine</td>
<td>RlmF</td>
</tr>
<tr>
<td>1835</td>
<td>N(^2)-methylguanosine</td>
<td>RlmG</td>
</tr>
<tr>
<td>1911</td>
<td>pseudouridine</td>
<td>RluD</td>
</tr>
<tr>
<td>1915</td>
<td>3-methylpseudouridine</td>
<td>RluD/RlmH</td>
</tr>
<tr>
<td>1917</td>
<td>pseudouridine</td>
<td>RluD</td>
</tr>
<tr>
<td>1939</td>
<td>5-methyluridine</td>
<td>RlmD</td>
</tr>
<tr>
<td>1962</td>
<td>5-methylcytidine</td>
<td>RlmI</td>
</tr>
<tr>
<td>2030</td>
<td>N(^6)-methyladenosine</td>
<td>RlmJ</td>
</tr>
<tr>
<td>2069</td>
<td>7-methylguanosine</td>
<td>RlmKL</td>
</tr>
<tr>
<td>2251</td>
<td>2'-O-methylguanosine</td>
<td>RlmB</td>
</tr>
<tr>
<td>2445</td>
<td>N(^2)-methylguanosine</td>
<td>RlmKL</td>
</tr>
<tr>
<td>2449</td>
<td>dihydrouridine</td>
<td>unknown</td>
</tr>
<tr>
<td>2457</td>
<td>pseudouridine</td>
<td>RluE</td>
</tr>
<tr>
<td>2498</td>
<td>2'-O-methylcytidine</td>
<td>RlmM</td>
</tr>
<tr>
<td>2501</td>
<td>5-hydroxyuridine</td>
<td>unknown</td>
</tr>
<tr>
<td>2503</td>
<td>2'-methyladenosine</td>
<td>RlmN</td>
</tr>
<tr>
<td>2504</td>
<td>pseudouridine</td>
<td>RluC</td>
</tr>
<tr>
<td>2552</td>
<td>2'-O-methyluridine</td>
<td>RlmE</td>
</tr>
<tr>
<td>2580</td>
<td>pseudouridine</td>
<td>RluC</td>
</tr>
<tr>
<td>2604</td>
<td>pseudouridine</td>
<td>RluF</td>
</tr>
<tr>
<td>2605</td>
<td>pseudouridine</td>
<td>RluB</td>
</tr>
</tbody>
</table>
1.3.2 Posttranscriptional Modification in Transfer RNA

All organisms contain many different tRNAs, at least one for each amino acid. *Escherichia coli* contains 48 different tRNAs. Transfer RNA is much smaller than rRNA, containing 75 to 95 nucleotides and is heavily modified. (Table 1.3)

Transfer RNA functions in translation as the “adaptor” molecule by delivering the correct amino acid to the ribosome for incorporation into the elongating polypeptide. The tRNA is covalently linked to the corresponding amino acid through the action of specific aminoacyl synthetases that recognize both the amino acid and the tRNA. The anticodon of this tRNA, a trinucleotide sequence, base pairs with the complementary trinucleotide codon present on the mRNA in the ribosome, where the amino acid is transferred to the growing polypeptide.

Post-transcriptional modifications present in tRNA are intimately involved in maintaining the fidelity and efficiency of the translation process. First, modifications are important in maintaining a specific overall tRNA tertiary structure [23] and are also important in prestructuring the anticodon loop region necessary for accurate and efficient decoding. Second, modifications play a role in aminoacyl synthetase recognition, so that the tRNA is aminoacylated with the correct amino acid. [24] Third, the numerous and diverse modifications in the anticodon loop modulate recognition and binding to the cognate codon thereby ensuring proper decoding. [25, 26] Fourth, post-transcriptional modifications are involved in reading frame maintainence, further ensuring proper decoding. [27, 28]

Modifications in tRNA have been more extensively studied than those of rRNA mainly due to difference in size and thus complexity of analysis. The most numerous and varied tRNA modifications occur at position 34 in tRNAs, the wobble position, and at position 37, 3’ adjacent to the anticodon and are clearly responsible for proper decoding. Structural characterization and identification of modified nucleosides in tRNA, especially those involved in decoding interactions, remains an active area of research.
Table 1.3 Modifications in *E. coli* transfer RNA, with sequence location and enzyme noted.

<table>
<thead>
<tr>
<th>Modifications in e. coli tRNA</th>
<th>tRNA position</th>
<th>Enzyme</th>
<th>tRNA isoacceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>dihydrouridine</td>
<td>16, 17, 20, 21</td>
<td>DusA, DusB, DusC</td>
<td>all except Arg3, Arg4, Glu1, Gly2, Try1, Try2</td>
</tr>
<tr>
<td>pseudouridine</td>
<td>55</td>
<td>TruB</td>
<td>all contain pseudo at position 55.</td>
</tr>
<tr>
<td>methylcytidine</td>
<td>65</td>
<td>TruC</td>
<td>Asp, Ile,</td>
</tr>
<tr>
<td>methylguanosine</td>
<td>40</td>
<td>TruA</td>
<td>Arg3, Leu1, Ser4</td>
</tr>
<tr>
<td>methyluridine</td>
<td>39</td>
<td>TruA</td>
<td>Asn, Cys, Gln2, His, Ile2, Ile3, Leu4, Lys, Met, Phe, Tyr1, Tyr2</td>
</tr>
<tr>
<td>methylaminomethyluridine</td>
<td>38</td>
<td>TruA</td>
<td>Gln1, Gln2, His, Leu1, Leu2, Leu3</td>
</tr>
<tr>
<td>methylguanosine</td>
<td>32</td>
<td>RiuA</td>
<td>Cys, Leu4, Phe</td>
</tr>
<tr>
<td>methyldcytidine</td>
<td>13</td>
<td>TruD</td>
<td>Glu</td>
</tr>
<tr>
<td>5-methylaminomethyluridine</td>
<td>34</td>
<td>MiaC</td>
<td>Arg3</td>
</tr>
<tr>
<td>5-carboxymethyluridine</td>
<td>34</td>
<td>CmaA/CmaB</td>
<td>Ala, Pro, Ser1, Thr4, Val1</td>
</tr>
<tr>
<td>3-(3-amino-3-carboxypropyl)uridine</td>
<td>47</td>
<td>unknown</td>
<td>Arg1, Ile1, Ile2, Ile3, Lys1, Met1, Phe1, Val2, Val3</td>
</tr>
<tr>
<td>2-thiocytidine</td>
<td>32</td>
<td>IscS</td>
<td>Arg1, Arg2, Arg3, Ser3</td>
</tr>
<tr>
<td>2'-O-methylcytidine</td>
<td>32</td>
<td>TrmJ</td>
<td>Ser1, Trp1, fMet1, fMet2</td>
</tr>
<tr>
<td>2'-O-methyluridine</td>
<td>34</td>
<td>TrmL</td>
<td>Leu5</td>
</tr>
<tr>
<td>5-carboxymethylaminomethyl-2'-O-methyluridine</td>
<td>34</td>
<td>unknown</td>
<td>Leu4</td>
</tr>
<tr>
<td>5-methyluridine</td>
<td>54</td>
<td>TrmA</td>
<td>all tRNA isoacceptors</td>
</tr>
<tr>
<td>5-methylaminomethyl-2-thioruridine</td>
<td>34</td>
<td>MmM/CmM/MmM/MmmG/MmmC</td>
<td>Glu1, Lys1</td>
</tr>
<tr>
<td>inosine</td>
<td>34</td>
<td>TsaA</td>
<td>Arg1</td>
</tr>
<tr>
<td>4-thioruridine</td>
<td>24</td>
<td>Thl/IsNS</td>
<td>present in many tRNAs</td>
</tr>
<tr>
<td>2'-O-methyluridine</td>
<td>32</td>
<td>TrmJ</td>
<td>Gin1, Gin2</td>
</tr>
<tr>
<td>7-methylguanosine</td>
<td>46</td>
<td>TrmB</td>
<td>present in 24 tRNAs</td>
</tr>
<tr>
<td>2'-O-methylguanosine</td>
<td>18</td>
<td>TrmH</td>
<td>Gin1, Gin2, Ile2, Ile3, Tyr1, Tyr2</td>
</tr>
<tr>
<td>1-methylguanosine</td>
<td>37</td>
<td>TrmD</td>
<td>Arg2, Pro1, Pro2, Pro3, Leu1, Leu2, Leu3</td>
</tr>
<tr>
<td>N4-acetylcytidine</td>
<td>34</td>
<td>TrmA</td>
<td>Met1</td>
</tr>
<tr>
<td>queuosine</td>
<td>34</td>
<td>Tgt/QerC/QerD/QerF/QerG</td>
<td>Asn1, Asp1, His1, Tyr1, Tyr2</td>
</tr>
<tr>
<td>lysidine</td>
<td>34</td>
<td>Tis</td>
<td>Ile2, Ile3</td>
</tr>
<tr>
<td>N6-threonylcarbamoyladenosine</td>
<td>37</td>
<td>RimN</td>
<td>Arg3, Arg4, Asn1, Ile1, Ile2, Ile3, Lys1, Met1, Ser3</td>
</tr>
<tr>
<td>N6-methyl-N6-threonylcarbamoyladenosine</td>
<td>37</td>
<td>TsasA (gene product)</td>
<td>Thr3</td>
</tr>
<tr>
<td>2'-methyladenosine</td>
<td>37</td>
<td>TrmG (yfF gene product) ??</td>
<td>Arg1, Asp, Gin1, Glu2, Glu1, His1,</td>
</tr>
<tr>
<td>N6-methyladenosine</td>
<td>37</td>
<td>Trmx yfC (gene product)</td>
<td>Val1</td>
</tr>
<tr>
<td>N6-(3'-methyl-2-butenyl)ladenosine</td>
<td>37</td>
<td>MiaA</td>
<td>Sec1</td>
</tr>
<tr>
<td>N6-(3'-methyl-2-butenyl)-2-methylthioadenosine</td>
<td>37</td>
<td>MiaA/MiaB/iscA</td>
<td>Cys1, Leu4, Leu5, Phe1, Ser1, Ser2, Trp1, Tyr1, Tyr2</td>
</tr>
</tbody>
</table>

Determining the amounts of the various modified nucleosides to be expected in unfractionated tRNA is not a straightforward task. *E. coli* contains 26 different modified nucleosides distributed unevenly among the 46 different tRNAs. [29, 30] (Table 1.3) Some modifications, such as pseudouridine and 5-methyluridine, are common to all tRNAs. Other modifications are present in the majority, but not all tRNAs, for example, dihydrouridine (D), 4-thioruridine (s4U), 7-methylguanosine (m7G), and 2'-O-methylguanosine (Gm). Several modifications, 5-methylaminomethyluridine (mm5U), 2'-O-methylcytidine (Cm), 5-methylaminomethyl-2-thioruridine (mm5'sU), inosine (I), N4-acetylatedine (ac4C), N6-methyl-N6-threonylcarbamoyladenosine (m6A), N6-methyladenosine (m6A) and N6-
isopentenyladenosine (i^6A), occur in only one specific tRNA. Most modifications are specific to one location within the tRNA sequence with the exception of dihydrouridine and pseudouridine, which are located at several different positions within the same tRNA. All tRNAs contain pseudouridine at position 55, but several tRNAs contain additional pseudouridines at other locations. The dihydrouridine loop (D loop) and stem domain of tRNA often contains more than one dihyrouridine.

Not surprisingly, the different tRNA species are not present in equal abundance within the cell. The relative abundance of the individual tRNAs is thought to be correlated to the frequency with which the cognate codon occurs on the mRNA. Several different codons can specify the same amino acid, i.e., synonymous codons, and these synonymous codons do not occur with equal frequency in the mRNA pool. A codon bias exists. In a set of synonymous codons, one dominates in frequency and this codon is read by the most abundant tRNA isoacceptor. Codon bias varies among species [31], varies with cellular conditions such as amino acid availability [32] and is correlated with growth rate [33-35].

Thus the complexity and large dynamic range in the analysis of modified nucleosides present in total tRNA nucleoside digests result from both the differences of frequency in occurrence of specific modifications in tRNAs and the effects of codon bias which are reflected in the differential expression of individual tRNAs.

1.4 Modifying Enzymes

Bacterial RNA is modified after transcription by site-specific modifying enzymes. These modifying enzymes can be specific for a single target in one class of RNA or have dual specificity for more than one class of noncoding RNA. FtsJ, for example, can modify both rRNA and tRNA. Some modifying enzymes are multi-site specific and can catalyze the formation of the same modification at several different locations in a given RNA. Some modifications require two or more enzymes for a single modification. For example, 2-methylthio-N^6-isopentenyladenosine (ms^2i^6A) requires the action of two different enzymes, MiaA forms the isopentenyl group and MiaB is responsible for the methylthiolation. Several of the modifying enzymes responsible for these posttranscriptional modifications have been
identified, but many remain unknown and uncharacterized, especially rRNA modifying enzymes; moreover, their regulation is not well understood. A mass spectrometry-based approach has the sensitivity and specificity to be used successfully in the characterization and identification of modifying enzymes in RNA and is demonstrated in this work.

1.5 Cooperative Effects Among Modifications

The study of post-transcriptional modifications in RNA has primarily focused on the absence or presence, location and structural characterization of individual modified nucleosides, but few studies measure quantitatively global changes in rRNA or tRNA modification. During protein synthesis, both tRNA and rRNA modifications have been shown to act cooperatively, and some modifications in tRNA have been shown to function in a regulatory role or act as sensors. Lack of a single modification is rarely reflected in any observable phenotype and only a few modifications have been shown to be essential to cell viability. However, combinations of modification deficiencies in tRNA and rRNA will decrease the efficiency and/or accuracy of translation.

Several examples of deficiencies in multiple ribosomal RNA modifications have shown defects in growth, ribosome production and function. In *E. coli*, several methylations make important contributions to the process of initiator selection at the P-site on the ribosome. [22] In *E. coli*, defects in cell growth, ribosome assembly and ribosome function occur in RluD mutants unable to form pseudouridine in three positions of helix 69 at the active site of the ribosome. [36] In yeast, deficiency of a single conserved pseudouridine in the A loop of the large subunit caused growth and translation defects and these defects were increased for multiple pseudouridine deficiencies. [37] Combinations of modification mutants demonstrate cooperative effects in terms of translation termination and reading frame maintenance among pseudouridine and 2’-O-methyl modifications in yeast. [38] Loss of two to three pseudouridine or ribose methylation modifications in the peptidyl transferase center of yeast caused multiple defects in translation and ribosome structure. [16] Modification is important in ribosomal subunit assembly and occurs with stepwise addition of modifications in the early, intermediate and late stages of
ribosome biosynthesis. [39] These examples demonstrate the cooperative nature of RNA modification and provides a possible entry point for RNA modification mediated control of ribosome biosynthesis.

In transfer RNA many examples are found where modification status varies depending on cellular conditions such as culture temperature [40, 41], nutrient availability [42], growth phase [43], growth rate [44, 45] and cell cycle progression. [46] Some modifications act as sensors for the presence of UV light [47], oxygen, [48] iron [49] and cancer. [50] Deficiency in a single modification can affect the modification status of several other modifications within tRNA. For example, lack of \( \Psi \) in tRNA *Thermus thermophilus* causes reduced growth and protein synthesis and an enhancement in amounts of \( m^1A \), \( m^1G \), Gm and \( s^2T \).[51] Similarly, deficiency of \( m^7G46 \) in *T. thermophilus* results in the reduction of the amounts of \( \Psi \), \( m^2G \), \( m^6U \), \( m^6A \), Gm and \( m^1G \). [52] Rapid tRNA degradation was found to occur in yeast double mutant strains lacking the ability to form \( m^7G46 \) in combination with either \( m^5C49 \), \( \Psi \) or D47. [53] Although other tRNA degradation pathways exist, the central role for modification in this tRNA degradation pathway may suggest possible regulatory features of this \( m^7G \) methyltransferase. Dihydrouridine synthetase can reversibly modify RNA, [54] and may be used as a modification control mechanism in RNA similar to the reversible phosphorylation in proteins. These examples support the existence of definite modification networks in the dynamic control of RNA modification and suggest a role for RNA modification in adaptation to changing cellular conditions.

The preceding examples show RNA modification is dynamically controlled within cellular systems; however, the mechanisms for this control are largely unknown. The dynamic control of RNA modification has not been explored primarily due to the lack of methods to quantitatively evaluate the global degree of modification.

### 1.6 Influence of Growth Temperature on Posttranscriptional Modification Levels

The relationship between heat-induced denaturation or loss of higher level structure in RNA and proteins resulting in loss of activity or function is commonly encountered in the study of biomolecules. One of the first observed and most frequently studied aspects of RNA modification was the influence of
temperature. As early as 1973 it was noted ribose methylation in tRNA increased with increasing growth temperature. [55] The effect of growth temperature on modification levels in extremeophiles has been most frequently studied. Experiments with the hyperthermophile *Pyrococcus furiosis* grown at 75°C, 85°C and 100°C revealed progressively increasing amounts of some modified tRNA nucleosides. [40] It was suggested that tRNA modifications account for the higher melting temperature in thermophiles and hyperthermophiles as compared to mesophiles. In bacterial psychrophiles, low levels of modification have been reported with the exception of dihydrouridine, which is present in greater amounts and is associated with enhancement of molecular flexibility at low temperature. [56] Experiments on the hyperthermophile, *Sulfolobus solfataricus*, grown at progressively higher temperatures showed a slight trend toward increased ribose methylation levels in 16S and 23S RNA, [57] although only changes greater than 20% were statistically significant. These are examples where the degree of modification varies with temperature and possibly enhances structural stabilization of RNA. This area of research requires more study and is largely unexplored. The main difficulty is lack of an assay to evaluate the degree of modification of a given nucleotide in an RNA. Many studies, especially those involving tRNA, note the absence or presence of specific modifications but few studies make rigorous quantitative measurements of global changes in all RNA modifications present at once.

1.7 Heat Shock and Posttranscriptional Modifications

The effect of heat shock on rRNA modification is investigated in this work. The heat shock response is a mechanism which allows a cell to cope with a sudden increase in temperature or other environmental stress. The initiation of the heat shock response is transcriptionally regulated by an alternative sigma factor that directs RNA polymerase to different promoter sites, which induces a rapid and transient expression of many genes that code for “heat shock proteins”. Investigating how the posttranscriptional modifications change in response to cellular stress may lead to a greater understanding of the functional significance of these modifications in the ribosome and in tRNA.
At least two modifying enzymes, FtsJ and MiaA, were shown to have significantly increased expression levels [58] under heat shock conditions and have been identified as heat shock proteins. FtsJ has been identified as a methyltransferase responsible for the formation of 2’-O-methyluridane at base position 2552 in 23S rRNA. [59, 60] Studies show this methyltransferase is active on the fully assembled 50S ribosomal subunit, not free 23S rRNA, and can also methylate tRNA. [60, 61] The yeast homolog of FtsJ also demonstrates dual substrate specificity for rRNA and tRNA. [59] The FtsJ deficient strain grows more slowly than the wild type and demonstrates translational defects at low and high temperatures. [62] The location of this modification in the A loop of 23S RNA suggests it has a role in forming the correct structure for accommodation of the A-site tRNA into the peptidyl transferase center. [63] The isopentyl-tRNA transferase, MiaA, is one of the two enzymes responsible for the formation of ms$^{2,6}$A. Most tRNAs (except tRNA serine I and V) reading codons beginning with U contain a hypermodified adenosine at residue 37, 3’ adjacent to the anticodon. The role of ms$^{2,6}$A is to improve reading frame maintenance. [28] The increased expression of miaA and ftsJ under heat shock conditions suggests these modifying enzymes participate in the cell’s response to heat shock, but how these enzymes function in response to this stress is unknown. This work seeks to provide methods of analysis so that comparisons of RNA modifications can be made under normal and heat shock conditions.

1.8 Why Determine Changes in Global Modification

An important focus here is to measure the differences in the degree of rRNA and tRNA modification in response to different physiological or environmental constraints of the bacterial cell. Many studies, especially regarding tRNA, note the absence or presence of specific modifications, but few studies make rigorous quantitative measurements of global changes in all rRNA modifications present at once. Several methods have been used for the analysis of modified nucleosides; however, HPLC-UV-MS based methods are the most well-suited for quantitative analyses and are used in this study.
For a meaningful comparison of modification status under different conditions to be made, the variability or precision of the method used must be characterized. Large variability in the method used diminishes the significance of the results and important changes could be missed in modification status. The reproducibility of this method is limited primarily by 1) the multistep isolation required to analyze modified nucleosides from a complex biological matrix and 2) the dynamic range of the analysis. In this study the instrumental variability and the culture-to-culture biological variability in the determination of modification status of rRNA and tRNA modification is assessed using an HPLC-UV/MS based method.

1.9 Analytical Methods of Analysis

The analysis of posttranscriptional modifications present in RNA is challenging due to the large size of the RNA biopolymer and the low abundance of modified nucleotides relative to unmodified major nucleotides. The nucleoside is the smallest structural unit where the modification is intact, thus controlled enzymatic digestion of the large RNA to ribonucleosides yields a complex mixture of modified and unmodified nucleosides that can be readily separated and characterized by HPLC/UV/MS. [64] The most obvious difference in modified nucleosides is the increase in mass compared to the unmodified major nucleosides, thus mass spectrometry has become the primary analytical tool used in the identification and characterization of posttranscriptionally modified nucleosides in enzymatic digests of RNA. [65] The sensitivity, specificity and accuracy of mass spectrometry combined with HPLC separation enables complete characterization and quantitation of complex mixtures of nucleosides.

In rRNA and tRNA, the extent of variation in the identity and relative quantities of modified nucleosides (i.e., a census of modifications) under different cellular conditions is a valuable measure of global modification status, which can be used to assess the dynamic nature of RNA modification. Similarly, the variation in the complete profile of modified nucleosides present in human urine can be used as a potential diagnostic biomarker for disease, especially cancer and inflammatory processes. Unlike the major nucleosides that are reused within the cell, salvage pathways for modified nucleosides
do not exist and thus they are excreted in urine. Analysis of enzymatic hydrolysates of cellular RNA is challenging due to the presence of large amounts of unmodified major nucleosides and the complex procedures required for rRNA and tRNA isolation from cells. Urinary nucleosides are easily isolated and do not contain large amounts of unmodified nucleosides.

For the analysis of modified nucleosides, a separation technique such as high performance liquid chromatography (HPLC), ultrahigh performance liquid chromatography (UPLC), gas chromatography (GC) or capillary electrophoresis (CE) is usually combined with ultraviolet (UV) or mass spectrometry based detection. [66, 67] Immunoassays have also been used for the analysis of only a limited number of nucleosides where specific antibodies have been developed. [68] Thin layer chromatography can be used for identification of known nucleosides but requires radiolabeling or large sample amounts. [69, 70] Capillary gas chromatography mass spectrometry (GC/MS) is less frequently used because derivatization with trimethylsilane is required, and GC/MS is less sensitive than other MS techniques for the analysis of modified nucleosides. [71]

In this work, electrospray ionization ion trap mass spectrometry coupled with reversed phase HPLC and UV detection is used for the identification and quantitation of modified nucleosides. The aromatic nature of nucleosides yields characteristic UV spectra, which with relative retention time, can be used to distinguish nucleosides, but requires suitable reference standards for positive identification. Coupling of this chromatographic separation technique to mass spectrometric techniques via electrospray ionization enhances the qualitative and quantitative capabilities of the analysis tremendously. The most abundant ions formed in positive mode electrospray mass spectra of nucleosides consist of two ion types, 1) the protonated molecule [MH]⁺ and 2) the protonated base [BH₂]⁺ released by cleavage of the C-N glycosidic bond and transfer of H from the ribose to the base. The modification status of the ribose is deduced by the difference in mass between the two ions; 132 u indicates a normal ribose while 146 u difference corresponds to the presence of a methyl group on the 2’ OH of the ribose. The exception is pseudouridine, which contains a stronger C-C glycosidic bond and fragments differently producing a molecular ion and
an ion due to the loss of water as the most abundant species. [72] Nucleosides generate a set of molecular and base ions that track together in the same HPLC retention time and MS ionization time.

Accurate mass measurements of molecular ions, base ions and fragments were made with a high resolution ion trap mass spectrometer to confirm nucleoside identity. The ability of the ion trap mass analyzer used in this work to perform successive MS/MS (MS^n) collision-induced dissociation experiments allows further fragmentation of the base ion and base fragments, thus providing greater structural detail. Fragmentation patterns of the collision-induced dissociation of cytidine, guanine, uracil and adenine and their substituted derivatives [73-76] have been studied and can be used to deduce structures of modified nucleosides in tandem MS^n experiments. The MS data, along with retention time and UV absorbance characteristics, further confirm nucleoside structure.

LC/MS has been used extensively in the identification and structural characterization of posttranscriptional modifications in tRNA and rRNA. The total census of modifications present in tRNA and rRNA of several different organisms have been determined and compared by LC/MS based techniques. [77-80] The modifying enzymes responsible for the synthesis of RNA modifications have been identified and characterized using LC/MS. [81] LC/MS has also been used to structurally characterize modified nucleosides excreted in urine.

Increasingly, LC/MS is being used for quantitative analysis and several studies have been reported for quantitative analysis of modified nucleosides in urine. [82, 83] Recently, quantitative studies of urinary nucleosides have used MS/MS techniques based on the known fragmentation behavior of nucleosides to monitor multiple molecular ion to base ion transitions (MRM) of several nucleosides known to be present. This type of MS/MS scan is exceptionally selective and sensitive, but somewhat limited in the number of nucleosides that can be characterized in one experiment. The reproducibility of LC-UV and various LC-ESIMS techniques for the quantitative analysis of urinary nucleosides has been reported.[82-87] Those methods, using a set of urinary nucleoside standards, yield reproducibilities in the range of 1-15 % relative standard deviations (RSD) using either UV or MS as the detector. Not unsurprisingly,
reproducibilities are significantly worse for quantifying isolated urinary nucleosides. With UV detection, reproducibility ranges from a low of 1 %RSD for well-behaved urinary nucleosides to as high as 30 – 35 %RSD for poorly behaved urinary nucleosides. Among the various MS/MS techniques used for quantification, Wang et al. characterized nine urinary nucleosides with a reproducibility ranging from 5 – 16 %RSD using selected ion monitoring MS.[86] More recently, Teichert et al. used constant neutral loss scans to quantify 35 urinary nucleosides with reproducibilities ranging from 6 – 50 %RSD.[87]

Unlike the case for urinary nucleosides, the analysis of enzymatic hydrolysates of cellular RNA is challenging due to the presence of large amounts of unmodified (major) nucleosides and the multi-step procedures required for rRNA and tRNA isolation from cells. Despite these challenges, similar LC-ESIMS methods have been used extensively in the identification and structural characterization of posttranscriptional modifications in tRNA and rRNA. The total census of modifications present in tRNA and rRNA of several different organisms have been determined using LC-ESIMS, often with on-line UV detection as well.[88-92] Among the analytical challenges faced during the quantitative analysis of modified nucleosides in tRNAs and rRNAs are dynamic range and specificity of analysis. For example, in the bacterial ribosome of *E. coli*, only 11 nucleotides are modified of the total 1542 nucleotides present in 16S RNA of the small subunit; similarly, only 23 nucleotides are modified of 2904 nucleotides in the large subunit 23S RNA. Because modification occurs at very low levels, 0.03 to 0.06% of total nucleotides, an analytical method must be capable of detecting such low levels of modified nucleosides in a background of the significantly more abundant major (unmodified) nucleosides.

In this work, the overall utility of reversed phase HPLC separation of modified nucleosides in an analytical method using LC-ESIMS and LC-UV for nucleoside identification and quantitation is presented. One particular interest of this study was to characterize the reproducibility of these methods when attempting to quantify modified nucleosides from RNA isolated from biological samples. Contrary to the situation for urinary nucleosides, to date there have been no reports describing the reproducibility of results one would obtain from isolating tRNAs and/or rRNAs and quantifying their constituent modified nucleosides generated by total nucleoside digestion. Large variability in method precision
diminishes the significance of the results, and if the variability is too great important differences could be missed. By establishing the overall method and sample variability for measuring modified nucleosides from tRNA and rRNA, the statistical significance of changes in modification status can be defined when comparing biological systems of varying environmental or physiological cellular conditions.

This work begins in Chapter 2 with a survey or census of modifications found in unfractionated tRNA from four bacterial and four archaeal organisms. A detailed comparison and discussion of the specific chemical and structural properties of each nucleoside is presented with emphasis on how modified nucleosides contribute to overall tertiary structure and decoding in tRNA. Chapter 3 demonstrates how the selectivity and sensitivity of LC/UV/MS/MS techniques are used to identify an unknown nucleoside. In Chapter 4, LC/MS techniques are used to functionally characterize modifying enzymes, specifically, two methylthiotransferases. Chapter 5 characterizes the technical and biological reproducibility encountered when using LC/UV/MS methods of analysis for modified nucleosides in tRNA and rRNA. In Chapter 6, these methods are used to characterize changes in global modification status in both rRNA and tRNA under different cellular conditions, ie, under heat shock and normal conditions. Conclusions of this dissertation and directions for future research are discussed in Chapter 7.
CHAPTER 2

Total Census of Posttranscriptional Modifications in tRNA from Selected Bacterial and Archaeal Organisms

2.1 Introduction

The complete genomes of many organisms, including all the organisms studied here, are known. Many of the tRNA genes have been identified within these DNA sequences, and some tRNA sequences can be determined from the whole genome. However, the posttranscriptional modifications of the individual tRNAs present in these organisms cannot be determined directly from DNA sequence information. The determination of individual tRNA sequences including modification status requires direct sequencing and modification characterization. As described in Chapter 1, the analysis is complex due to the presence of many isoaccepting tRNAs with different sequences. Isolation of tRNAs can be challenging due to the structural similarities and vast differences in relative abundances of the different tRNAs present in the cell. Transfer RNA sequences, including identities and location of modified nucleosides, are available for only a few bacterial and archaean organisms including Escherichia coli, Bacillus subtilis and Haloferax volcanii. [29, 30]

To gain further understanding of modifications in tRNA and their functional roles, it is useful to begin with the identification of the modified nucleosides present in the total unfractionated tRNA pool, a census of modifications, from a particular organism. Comparisons of modifications present among organisms of different phylogenetic classifications and those living in extreme conditions further our understanding of how modifications enhance the structure and function of tRNA. In this study, LC/MS was used to determine the identity of the posttranscriptional modifications present in unfractionated, total, tRNA in four different bacterial organisms and four archaean organisms. The organisms analyzed here are listed in Table 2.1. A total census of posttranscriptional modifications determined by LC/MS has been reported for Escherichia coli [81, 93], Sulfolobus soflatarius [65, 94], Methanococcus maripaludis [80] and a
partial census for *Thermus thermophilus*. [51, 52, 95] For *Haloquadratum walsbyi, Bacillus subtilis* and *Haloarcula marismortui* the total census of posttranscriptional modifications by LC/MS is reported here for the first time.

### Table 2.1 Organisms used in this study.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Kingdom</th>
<th>Optimum Growth Conditions</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>Bacteria</td>
<td>37 °C, mesophile</td>
<td>model organism</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>Bacteria</td>
<td>30 °C, mesophile</td>
<td>model organism</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em></td>
<td>Bacteria</td>
<td>37 °C, mesophile</td>
<td>gram positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>gram positive and spore forming</td>
</tr>
<tr>
<td><em>Thermus thermophilus</em></td>
<td>Bacteria</td>
<td>68 °C, thermophile</td>
<td>model organism</td>
</tr>
<tr>
<td><em>Haloarcula marismortui</em></td>
<td>Archaea</td>
<td>37 °C, mesophile</td>
<td>closely related to <em>H. walsbyi,</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.3-4.5 M NaCl, halophile</td>
</tr>
<tr>
<td><em>Haloquadratum walsbyi</em></td>
<td>Archaea</td>
<td>45 °C, mesophile</td>
<td>extremely thin square-shaped</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt; 3.3 M NaCl and MgCl₂ enriched</td>
</tr>
<tr>
<td><em>Sulfolobus solfataricus</em></td>
<td>Archaea</td>
<td>85 °C, hyperthermophile</td>
<td>sulfur metabolizing archaea</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pH 2.0-4.5, acidophile</td>
</tr>
<tr>
<td><em>Methanococcus maripaludis</em></td>
<td>Archaea</td>
<td>37 °C, mesophile</td>
<td>methane producing archaea</td>
</tr>
</tbody>
</table>

#### 2.2 Experimental

##### 2.2.1 Desalting of tRNA

A Nucleobond (Machery Nagel) anion exchange column was used to purify and desalt the tRNA prior to enzymatic digestion. The tRNA was loaded onto the column in a solution of 200 mM KCl, 100 mM tris acetate (pH 6.3) and 15% ethanol then washed with the same solution. A solution of 400 mM KCl, 100 mM tris acetate (pH 6.3) and 15% ethanol was used for a second wash, then the tRNA was eluted with 750 mM KCl, 100 mM tris acetate (pH 6.3) 15% ethanol. Isopropanol was used to precipitate the
tRNA and the pelleted RNA was rinsed with 70% ethanol. The purified tRNA pellet was resuspended in autoclaved water.

2.2.2 Digestion of tRNA to nucleosides

Prior to enzymatic digestion, the tRNA was denatured at 100 °C for 3 min then chilled in an ice water bath. To lower the pH, 1/10 volume 0.1M ammonium acetate (pH 5.3) was added. For each 0.5 AU of tRNA, 2 units Nuclease P1 was added and incubated at 45 °C for 2 h. The pH was readjusted by adding 1/10 volume of 1.0 M ammonium bicarbonate then 0.002 units of snake venom phosphodiesterase was added and incubated at 37 °C for 2 h. Finally, 0.5 units of antartic phosphatase was added and incubated at 37 °C for 1 h. The nucleoside digests were stored at -80 °C until analyzed. [64]

2.2.3 HPLC/MS conditions

The nucleoside digests were analyzed using a Hitachi D-7000 HPLC system with a Hitachi D-7400 UV detector set at 260 nm. A LC-18-S (Supelco) 2.1 x 250 mm, 5 µm particle column was used at a flow rate of 300 µL/min. Mobile phases used were as follows, A: 5 mM ammonium acetate, pH 5.3 and B: 40% acetonitrile in water. The column eluent was split immediately post column, 1/3 to the electrospray ion source and 2/3 to the UV detector. The gradient used is outlined in Table 2.2. A Thermo LTQ-XL ion trap mass spectrometer was used for the low resolution LC/MS and LC/MS/MS analyses of the bacterial organisms. A high resolution Thermo LTQ-FT ion trap mass spectrometer was used for analysis of the archael organisms. Both instruments were equipped with an ion max electrospray ionization source. Mass spectra were recorded in the positive ion mode with a capillary temperature of 275°C, spray voltage, 3.7 to 4.0 kV and sheath gas, auxillary gas and sweep gas of 45, 25 and 10 arbitrary units, respectively. Data dependent MS/MS of each of the two most intense ions were recorded throughout the LC/MS run.
Table 2.2 HPLC Gradient used to separate nucleoside digest.

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>% B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5.8</td>
<td>1</td>
</tr>
<tr>
<td>9.2</td>
<td>2</td>
</tr>
<tr>
<td>10.9</td>
<td>3</td>
</tr>
<tr>
<td>12.7</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>45</td>
<td>99</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
</tr>
</tbody>
</table>

2.3 Results

Unfractionated total tRNA from each of the eight organisms listed in Table 2.1 was enzymatically digested to nucleosides and analyzed by LC/MS. The nucleosides were separated by HPLC and identified based on relative retention times and mass spectra. The UV chromatograms of the enzymatic hydrolysates of the unfractionated tRNA with nucleosides identified are shown in Figures 2.1 - 2.8
Figure 2.1 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated Escherichia coli tRNA. 1, dihyouridine (D). 2, pseudouridine (Ψ). 3, 5-carboxymethoxyuridine (cmoU). 4, 3-(3-amino-3-carboxypropyl)uridine (acpU). 5, 5-methylaminomethyluridine (mmm5U). 6, 2-thiocytidine (s2C). 7, 5-methylcytidine (m5C). 8, 2'-O-methylcytidine (Cm). 9, 5-carboxymethylaminomethyl-2'-O-methyluridine (cmnm5Um). 10, 5-methyluridine (m5U). 11, 5-methylaminomethyl-2-thiouridine (mmm5s2U). 12, inosine (I). 13, 4-thiouridine (s4U). 14, 2'-O-methyluridine (Um). 15, 7-methylguanosine (m7G). 16, 2'-O-methylguanosine (Gm). 17, 1-methylguanosine (m1G). 18, N4-acetylcystidine (ac4C). 19, epoxyqueuosine (oQ). 20, queuosine (Q). 21, lysidine (k2C). 22, N6-threonylcarbamoyladenosine (t6A). 23, N6-methyl-N6-threonylcarbamoyladenosine (m6t6A). 24, 2-methyladenosine (m2A). 25, N6-methyladenosine (m6A). 26, N6, N6-dimethyladenosine (m62A). 27, N6-isopentenyladenosine (i6A). 28, 2-methylthio-N6-isopentenyladenosine (msi6A).

Figure 2.2 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated Bacillus subtilis tRNA. 1, D. 2, Ψ. 3, 5-carboxymethylaminomethyluridine (cmm5U). 4, C. 5, U. 6, 1-methyladenosine (m1A). 7, 5-carboxymethylaminomethyl-2-thiouridine (cmmm5s2U). 8, m7U. 9, 5-methoxyuridine (mo5U). 10, I. 11, G. 12, m7G. 13, Gm. 14, Q. 15, m7G. 16, k2C. 17, N5-methylguanosine (m5G). 18, A. 19, t6A. 20, m2A. 21, m6A. 22, N6-threonylcarbamoyl-2-methylthioadenosine (msi6A). 23, 2-methylthioadenosine (ms2A). 24, m62A. 25, i6A. 26, msi6A.
Figure 2.3 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Lactococcus lactis* tRNA. 1, D. 2, Ψ. 3, C. 4, U. 5, 2’-deoxyctydine (dC). 6, cmnm5s2U. 7, Cm. 8, m’A. 9, I. 10, m9U. 11, G. 12, m6U. 13, m7G. 14, 2’-deoxyguanosine (dG). 15, thymidine (dT). 16, m’G. 17, ac4C. 18, A. 19, 2’-deoxyadenosine (dA). 20, t6A. 21, m2A. 22, m6A. 23, ms2t6A. 24, tA.

Figure 2.4 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Thermus thermophilus* tRNA. 1, D. 2, Ψ. 3, C. 4, U. 5, Cm. 6, m1A. 7, mnm3s2U. 8, m3U. 9, G. 10, m7G. 11, dG. 12, s4U. 13, Gm. 14, m7G. 15, m5G. 16, 5-methyl-2-thiouridine (m5s2U). 17, A. 18, dA. 19, t6A. 20, m6A. 21, ms2t6A. 22, m3G. 23, N6-(cis-hydroxyisopentenyl)adenosine (io6A). 24, 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine (ms2io6A).
Figure 2.5 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated Haloarcula marismortui tRNA. 1, Ψ. 2, C. 3, U. 4, 1-methylpseduouridine (m$^5$Ψ). 5, m$^5$C. 6, m$^1$A. 7, Cm. 8, G. 9, m$^5$G. 10, dG. 11, 1-methylinosine (m$^1$I). 12, m$^1$G. 13, acG. 14, m$^2$G. 15, A. 16, dA. 17, N$^2$, N$^2$-dimethylguanosine (m$^2$G). 18, t$^6$A. 19, 5-methoxycarbonylmethyl-2-thiouridine (mcm$^5$2U). 20, archaeosine (G$^+$). 21, m$^6$A and m$^2$A. 22, 4-demethylwyosine (imG14). 23, ms$^5$t$^6$A. 24, m$^6$A.

Figure 2.6 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated Methanococcus maripaludis tRNA. 1, Ψ. 2, C. 3, U. 4, m$^1$Ψ. 5, s$^2$C. 6, m$^5$C. 7, m$^1$A. 8, Cm. 9, G. 10, s$^6$U. 11, m$^1$I. 12, m$^5$G. 13, m$^2$G. 14, A. 15, dA. 16, m$^2$G. 17, t$^6$A. 18, 5-methoxycarbonylmethyluridine (mcm$^5$U). 19, G$^+$. 20, m$^6$A. 21, N$^6$-hydroxynorvalycarbamoyl-adenosine (hn$^6$A). 22, imG14. 23, ms$^5$t$^6$A. 24, wyosine (isomer unknown). 25, m$^6$A.
Figure 2.7 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Sulfolobus* *sofataricus* tRNA. 1, Ψ. 2, C. 3, U. 4, m^5^-C. 5, m^1^-A. 6, Cm. 7, G. 8, m^7^-G. 9, Um. 10, dG. 11, 5,2'–O-dimethylcytidine (m^5^-Cm). 12, m^1^-I. 13, Gm. 14, m^1^-G. 15, ac^4^-C. 16, m^2^-G. 17, A. 18, dG, 19, m^2^-G. 20, t^6^-A. 21, 2'–O-methyladenosine (Am). 22, N^4^-acetyl-2'–O-methylcytidine (ac^4^-Cm). 23, agmatidine (C+). 24, G+. 25, m^6^-A. 26, ms^2^-t^6^-A. 27, N^2^, N^2^, 2'–O-trimethylguanosine (m^2^-Gm). 28, m^6^-A. 29, methylwyosine (mimG).

Figure 2.8 UV Chromatogram from LC/MS analysis of an enzymatic digest of unfractionated *Haloquadratum walsbyi* tRNA. 1, Ψ. 2, C. 3, U. 4, m^-1^-Ψ. 5, m^-5^-C. 6, Cm. 7, m^-1^-A. 8, G. 9, m^-7^-G. 10, m^-1^-I. 11, m^-G. 12, ac^-4^-C. 13, m^-2^-G. 14, A. 15, m^-2^-G. 16, t^6^-A. 17, mcm^5^-5^-G. 18, C+. 19, G+. 20, m^-6^-A. 21, imG14. 22, threonylcarbamoyladenosine (ms^2^-t^6^-A). 23, m^-6^-A.
No UV peak is observed for dihydrouridine because it does not possess a chromophore and thus has no UV absorbance at 260 nm. Dihydrouridine is identified by its characteristic mass spectra and very early elution time. The glycosidic bond in a nucleoside fragments easily in the electrospray process, thus the major ions present in the mass spectra are the protonated base ion and protonated molecular ion. The relative abundance of these two ions depends on the nature of the nucleoside and the electrospray conditions. In the selected ion chromatograms of the molecular and base ions, the peaks will track together at the same retention time as the UV peak, providing evidence of the presence of a nucleoside. The mass difference between the molecular ion and the base ion provides information on the attached sugar. A mass difference of 132 u indicates a normal ribose, a mass difference of 146 u is a methylated ribose and mass difference of 116 u is a deoxyribose. MS/MS data were collected from the two most intense ions continuously during the analysis. This low energy MS/MS is helpful in distinguishing the correct molecular/base ion pairs and can also be used to confirm the identity of sodium, potassium or ammonia adducts. MS/MS is an especially diagnostic feature in identifying nucleosides that may coelute chromatographically and where base ion intensity is low. Isomers with identical molecular and base ion masses can usually be distinguished by their different UV retention times. The nucleosides identified for each of the eight organisms studied are listed for comparison in Table 2.3.
### Table 2.3 Comparison of nucleosides identified in tRNA from selected organisms.

<table>
<thead>
<tr>
<th>E. coli</th>
<th>B. subtilis</th>
<th>L. lactis</th>
<th>T. thermophilus</th>
<th>H. marismortui</th>
<th>M. maripaludis</th>
<th>S. solfatarius</th>
<th>H. walsbyi</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>(\psi)</td>
<td>(\psi)</td>
<td>(\psi)</td>
<td>(\psi)</td>
<td>(\psi)</td>
<td>(\psi)</td>
<td>(\psi)</td>
<td>(\psi)</td>
</tr>
<tr>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
<td>cmnm(^1)U</td>
</tr>
<tr>
<td>Cm</td>
<td>Cm</td>
<td>Cm</td>
<td>Cm</td>
<td>Cm</td>
<td>Cm</td>
<td>Cm</td>
<td>Cm</td>
</tr>
<tr>
<td>m(^3)U</td>
<td>m(^3)U</td>
<td>m(^3)U</td>
<td>m(^3)U</td>
<td>m(^3)U</td>
<td>m(^3)U</td>
<td>m(^3)U</td>
<td>m(^3)U</td>
</tr>
<tr>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
<td>mo(^5)U</td>
</tr>
<tr>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
</tr>
<tr>
<td>Um</td>
<td>Um</td>
<td>Um</td>
<td>Um</td>
<td>Um</td>
<td>Um</td>
<td>Um</td>
<td>Um</td>
</tr>
<tr>
<td>s(^4)U</td>
<td>s(^4)U</td>
<td>s(^4)U</td>
<td>s(^4)U</td>
<td>s(^4)U</td>
<td>s(^4)U</td>
<td>s(^4)U</td>
<td>s(^4)U</td>
</tr>
<tr>
<td>Gm</td>
<td>Gm</td>
<td>Gm</td>
<td>Gm</td>
<td>Gm</td>
<td>Gm</td>
<td>Gm</td>
<td>Gm</td>
</tr>
<tr>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
</tr>
<tr>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
<td>m(^1)G</td>
</tr>
<tr>
<td>oQ</td>
<td>oQ</td>
<td>oQ</td>
<td>oQ</td>
<td>oQ</td>
<td>oQ</td>
<td>oQ</td>
<td>oQ</td>
</tr>
<tr>
<td>k(^5)C</td>
<td>k(^5)C</td>
<td>k(^5)C</td>
<td>k(^5)C</td>
<td>k(^5)C</td>
<td>k(^5)C</td>
<td>k(^5)C</td>
<td>k(^5)C</td>
</tr>
<tr>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
<td>ac(^6)C</td>
</tr>
<tr>
<td>m(^2)G</td>
<td>m(^2)G</td>
<td>m(^2)G</td>
<td>m(^2)G</td>
<td>m(^2)G</td>
<td>m(^2)G</td>
<td>m(^2)G</td>
<td>m(^2)G</td>
</tr>
<tr>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
<td>m(^2)s(^4)U</td>
</tr>
<tr>
<td>Am</td>
<td>Am</td>
<td>Am</td>
<td>Am</td>
<td>Am</td>
<td>Am</td>
<td>Am</td>
<td>Am</td>
</tr>
<tr>
<td>t(^6)A</td>
<td>t(^6)A</td>
<td>t(^6)A</td>
<td>t(^6)A</td>
<td>t(^6)A</td>
<td>t(^6)A</td>
<td>t(^6)A</td>
<td>t(^6)A</td>
</tr>
<tr>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
<td>mcm(^3)s(^4)U</td>
</tr>
<tr>
<td>G+</td>
<td>G+</td>
<td>G+</td>
<td>G+</td>
<td>G+</td>
<td>G+</td>
<td>G+</td>
<td>G+</td>
</tr>
<tr>
<td>m(^5)A</td>
<td>m(^5)A</td>
<td>m(^5)A</td>
<td>m(^5)A</td>
<td>m(^5)A</td>
<td>m(^5)A</td>
<td>m(^5)A</td>
<td>m(^5)A</td>
</tr>
<tr>
<td>m(^6)A</td>
<td>m(^6)A</td>
<td>m(^6)A</td>
<td>m(^6)A</td>
<td>m(^6)A</td>
<td>m(^6)A</td>
<td>m(^6)A</td>
<td>m(^6)A</td>
</tr>
<tr>
<td>imG14</td>
<td>imG14</td>
<td>imG14</td>
<td>imG14</td>
<td>imG14</td>
<td>imG14</td>
<td>imG14</td>
<td>imG14</td>
</tr>
<tr>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
</tr>
<tr>
<td>io(^5)A</td>
<td>io(^5)A</td>
<td>io(^5)A</td>
<td>io(^5)A</td>
<td>io(^5)A</td>
<td>io(^5)A</td>
<td>io(^5)A</td>
<td>io(^5)A</td>
</tr>
<tr>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
</tr>
<tr>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
</tr>
<tr>
<td>imG (W)</td>
<td>imG (W)</td>
<td>imG (W)</td>
<td>imG (W)</td>
<td>imG (W)</td>
<td>imG (W)</td>
<td>imG (W)</td>
<td>imG (W)</td>
</tr>
<tr>
<td>mimG</td>
<td>mimG</td>
<td>mimG</td>
<td>mimG</td>
<td>mimG</td>
<td>mimG</td>
<td>mimG</td>
<td>mimG</td>
</tr>
<tr>
<td>AgmC</td>
<td>AgmC</td>
<td>AgmC</td>
<td>AgmC</td>
<td>AgmC</td>
<td>AgmC</td>
<td>AgmC</td>
<td>AgmC</td>
</tr>
<tr>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
<td>ms(^5)t(^6)A</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


All of the *E. coli* modified nucleosides reported here (Figure 2.1) were found in at least one of the 47 *E. coli* sequences listed in the tRNA database. In addition, 5-carboxymethylaminomethyl-2′-O-methyluridine (cmm\(^5\)Um) and epoxyqueuosine (oQ), not listed in the database, are found here and reported elsewhere. Both 2′-O-methyluridine and cmm\(^5\)Um were not found in two similar LC/MS analyses of modified nucleosides present in *E. coli* (3). One of those studies reported the presence of 5-methoxycarbonylmethoxyuridine (mcmo\(^5\)U), which was not found here or listed in the tRNA sequence database.

The tRNA modifications present in *B. subtilis* are consistent with those reported in the tRNA sequence database and those found in previous studies. An LC/MS-based total census of modifications present in unfractionated tRNA from *B. subtilis* has not been reported elsewhere. The only modified nucleoside known to be present in *B. subtilis* total tRNA and not detected in this analysis was 4-thiouridine (s\(^4\)U). It is likely that the 4-thiouridine co-eluted with the major guanosine peak and was difficult to detect due to this difference in relative abundance. Another possibility is that the uridine is incompletely thiolated because cells were harvested and the tRNA extracted for this analysis during late exponential growth phase when the growth rate is very high and thiolation is incomplete. The thiolation of some tRNA species, but not all, varies dramatically at different growth rates.

Little information on the modified nucleoside content is available for *L. lactis*, and this is the first report of the census of modifications by LC/MS in unfractionated tRNA from *L. lactis* (Figure 2.2). For the bacterial thermophile, *T. thermophilus*, only 6 tRNAs have been experimentally sequenced, which include locations of modified nucleosides as listed in the tRNA database. The total census of modifications of the unfractionated tRNA of *T. thermophilus* presented here includes several modified adenosines and a modified uridine not found in previous studies.

The RNA sequences including identity and location of modified nucleotides are known for very few archaeal organisms. Only 76 of the total 1405 tRNA sequences reported in the tRNA sequence
database belong to archaeal organisms. The most complete set of archaeal tRNA sequences is reported for the mesophilic halophile, *Haloferax volcanii.* [29] A total census of modifications in tRNA of *H. marismortui* and *H. walsbyi* has not been reported. For both *S. solfataricus* [65, 94] and *M. maripaludis* [80], a census of tRNA modifications has been determined by LC/MS analysis. A comparison of the *S. solfataricus* tRNA modifications present in this analysis to previous studies reveals the presence of two additional modifications, agmatidine (agmC) and N'-acetyl-2'-O-methyleytidine (ac^4Cm); however 2-thio-2'-O-methyluridine (s^2Um) was not found here but reported previously. [65, 94] Several modifications found in previous analyses of *M. maripaludis* (7) were not found here, m^2^Gm, mnm^5^s^2^U, I, Um, Gm, Am and m^2^Gm. However, mcm^5^U was found in *M. maripaludis* in this analysis and not reported earlier.

The total census of modifications present in unfractionated tRNA cannot give information on location of modified nucleosides within tRNA. However, the sequence location can be inferred from previous studies compiled in the tRNA sequence database from similar organisms. For example, the hypermodified adenosines are assumed to be present at position 37, adjacent to the anticodon, and the hypermodified uridines are likely present at position 34, the first position of the anticodon. The complete set of tRNA sequences and location of modified nucleosides are known for only a few model organisms such as *E. coli, B. subtilis* and *H. volcanii.* For comparison, the location of the modified nucleosides identified in this study are shown in secondary structure cloverleaf representations for the bacterial species *E. coli, B. subtilis,* and *H. volcanii* in Figure 2.9. Modification locations are known for a few *T. thermophilus* tRNAs and those modification locations are also noted in Figure 2.9. The sequence information available for archaeal tRNAs is very limited and *H. volcanii* is the only archaeal species where all the tRNAs have been sequenced and modification locations known. The *H. volcanii* modification locations can be used to predict locations of the archaeal tRNA modifications identified in the organisms studied here.
Figure 2.9 Secondary structure of tRNA with type and sequence location of modified nucleosides. *E. coli* modifications are in bold. (A) denotes locations of modifications in the archaea, *H. volcanii*. (B) denotes sequence location of modifications in *B. subtilis*. (T) denotes sequence location of modifications in *T. thermophilus*.

All of the organisms here contain the universal modifications, pseudouridine (Ψ), 1-methylguanosine (m1G), N6-threonylcarbamoyladenosine (t6A) and N6-methyladenosine (m6A). In addition to these, all of the bacteria studied here contain dihydrouridine (D), 7-methylguanosine (m7G), and 5-methyluridine (m5U). Modifications common to all of the archaea studied here but not found in bacteria include 1-methylinosine (m1I), archaeosine (G+) and N2, N2'-dimethylguanosine (m2G). Modifications unique to archaea and found in at least one of the archaea studied here include 1-methylpseudouridine (m1Ψ), 5, 2'-
O-dimethylecytidine (m5Cm), N4-acetyl-2'-O-methylecytidine (ac4Cm) and 4-demethylwyosine (imG14). A few of the archaeal modifications found are not present in bacteria but occur in eukarya, including N2, N2-dimethylguanosine (m2G), wyosine (imG), 5-methoxycarbonylmethyluridine (mcm5U) and 5-methoxycarbonylmethyl-2-thiouridine (mcm5s2U).

In every organism, post-transcriptional modifications in tRNA serve to maintain tertiary structure and provide for accurate and efficient translation. Characterizing the variations in these modification patterns between domains and among different species can assist in determining the functional significance of modifications.

Within the tertiary core, several distinct differences in modifications exist among the four archaeal species and four bacterial species studied here. Archaeosine, a distinctly archaeal modification, is found in all archaea analyzed here. Archaeosine is present in all archaea analyzed to date and is important in maintaining tertiary structure. Some of these bacterial organisms contain the structurally similar queuosine. Queuosine and archaeosine share initial biosynthesis pathways, but are located at very different positions and thus have different functions in tRNA. Inosine is present in all these bacterial organisms but absent in the archaeal organisms. However, archaea contain methylated inosine, m1I, which is characteristically a eukaryal modification and not found in bacteria. Inosine from bacterial and eukaryal organisms is usually present at position 34, while the m1I modification in archael tRNA is located in the tertiary core at position 57. The uridine present at position 54 is always methylated to form m5U54 in most all bacteria and eukarya, but this position in archaea is modified with a methylated pseudouridine, m1Ψ. Both bacterial m5U and archaeal m1Ψ are involved in critical conserved tertiary folding interactions. The dimethylated guanosine, m2G, is present in these archaeal species but not in the bacterial species. Dimethylated guanosine (m2G) is also present in eukarya and is important in tertiary interactions.
Dihydrouridine is not usually found in archaea. Dihydrouridine contributes flexibility to the tRNA core structure. Archaea are usually extremeophiles living under harsh conditions, especially high temperature and likely do not require the flexibility in structure dihydrouridine imparts. Archaea, especially thermophiles, contain many ribose methylated or doubly methylated nucleosides. *S. sofatarius* analyzed here exhibits this with the presence of Cm, Um, Gm, Am, m^5^Cm, ac^4^Cm, and m^2^2Gm. Ribose methylated nucleosides impart structural rigidity to tRNA and are frequently present in thermophilic archaea and associated with increased thermal stability. [40]

A variety of modifications are present in the anticodon stem loop domain and some are phylogeneticaly unique. Differences in modifications present in the first position of the anticodon (position 34) and the position directly adjacent to the anticodon (position 37) directly affect decoding accuracy and efficiency. Notable differences exist in the modified uridines usually present at position 34 and the hypermodified adenosines of position 37 in archaeal and bacterial species. N^6^-isopentenyladenosine and derivatives, present at position 37, are absent in these archaea, but occur in these bacterial species. The wyosine family of tricyclic nucleosides are present in archaea and eukarya at position 37 and, indeed, are present in the archaea analyzed here. Wyosine derivatives are absent in these bacterial species.

The modified uridines usually associated with position 34 identified in these archaeal species are common to eukarya and do not include any 5-substituted methoxyuridines. However, other archaea contain 5-substituted methyluridines common to bacteria. Inosine is often present at position 34 in bacteria and is absent in these archaea. Agmatidine (AgmC+) was found in all the archaeal species analyzed here and is not found in bacteria. Lysidine, with the same decoding properties as agmatidine, is found in bacterial systems. Queuosine is also found in bacteria at position 34 and absent in archaeal species. Curiously, its archaeal counterpart, archaeosine, is not present at position 34 in archaea, but associated with maintaining tertiary structure at position 15.
Modification differences within each domain appear more subtle for the organisms studied here. *T. thermophilus* and *S. sofataricus* are the only thermophiles of their respective domains and each contains modifications known to be important in thermostability. In *S. sofataricus*, several ribose methylated and double methylated modifications are present, but absent in the archaeal mesophiles analyzed. *T. thermophilus* contains $m^5s^2U$, known to contribute to thermal stability. The gram positive bacteria, *B. subtilis* and *L. lactis*, are distinguished by the presence of 5-methoxyuridines. *B. subtilis* and *L. lactis* contain mo$^5U$ in contrast to cmo$^5U$ often found in gram negative bacteria such as *E. coli* and *T. thermophilus*. *B. subtilis* and *L. lactis* contain the carboxylated modified uridines, cmmn$^5U$ and cmmn$^5s^2U$ where *E. coli* and *T. thermophilus* contain mnm$^5U$ and mnm$^5s^2U$. Differences in the modified adenosines include the absence of ms$^2t^6A$ in *E. coli* and the presence of hydroxylated hypermodified adenosines, io$^6A$ and ms$^2io^6A$, found only in *T. thermophilus*. *E. coli* contains m$^2t^6A$ and not the methylthiolated derivative found in the other bacteria.

Specific similarities and differences among the modifications found in these organisms are discussed in the next section.

### 2.4 Discussion

#### 2.4.1 Bacterial Organisms

The detailed comparison of posttranscriptional modifications present in tRNA begins with *E. coli*, the most studied bacteria and the primary model system in biochemistry and molecular biology. The complete genome sequence is known and the sequences of all the tRNAs are known including the identity and location of posttranscriptional modifications. As early as 1926 microbiologist, Albert Jan Kluyver, articulated that “from the elephant to butyric acid bacterium—it’s all the same.” Later, in 1954 Jacques Monod published the phrase “anything found to be true of *E. coli* must also be true of elephants” when commenting on the unity of biochemistry. Of course, the elephant is a significantly more complex
organism but, regardless, the use of *E. coli* as a model system builds knowledge and guides experimentation in other organisms. [99]

Here, we begin with a look at the posttranscriptional modifications present in *E. coli* and then compare those with the modifications present in the bacteria *B. subtilis, T. thermophilus* and *L. lactis*. Several modifications are universally conserved in bacteria and all are, indeed, found in the four bacterial tRNA studied here: 5-methyluridine (m\(^5\)U54), pseudouridine (Ψ55), dihydrouridine (D20), 7-methylguanosine (m\(^7\)G46), N\(^6\)-threonylcarbamoyladenosine (t\(^6\)A37), and N\(^6\)-methyladenosine (m\(^6\)A37). Distinct differences exist among these organisms especially modifications present in the anticodon domain. When comparing modifications in tRNA it is useful to note the similarities and differences in each of the tRNA domains: the aminoacyl stem, the D stem-loop, the anticodon stem and loop (ASL), the variable loop and the TΨC stem-loop. In general, modifications in the acceptor stem, D stem-loop, variable loop and TΨC stem-loop serve to maintain the tertiary structure. The conserved L-shape tertiary structure, formed from folding of D stem-loop into the TΨC stem-loop and variable loop, is maintained through specific conserved Watson-Crick and triple base pairing interactions. Nucleotide modification, divalent metal ion binding, temperature and RNA backbone interactions also influence the tertiary structure. [100, 101]

Modifications in the anticodon stem-loop are important in maintaining proper structure for accurate and efficient decoding.

The identity and location of modifications in each domain of tRNA are discussed below with emphasis on the structural and chemical properties of each nucleoside in relation to how that modification may contribute to the overall tertiary structure or decoding in tRNA. Information regarding the modifying enzymes responsible for the synthesis of these modifications are included, if known.

**2.4.1.1 Posttranscriptional Modifications in the tertiary core of tRNA**

The modifications present in the tertiary core of tRNA in the bacterial organisms studied here are listed in Table 2.4 by domain. Domain and sequence location of the modifications is based on data from
the tRNA sequence database. *L. lactis* is not included in this table because no sequence information is currently available for this organism. Because *B. subtilis* and *L. lactis* are both mesophilic gram positive bacteria, they likely have very similar modification patterns.

**Table 2.4** Modifications in Tertiary Core.

<table>
<thead>
<tr>
<th></th>
<th><em>E. coli</em></th>
<th><em>B. subtilis</em></th>
<th><em>T. thermophilus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>D Loop</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>ψ</td>
<td>ψ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>m1A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gm</td>
<td>Gm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s4U</td>
<td>s4U</td>
<td>s4U</td>
</tr>
<tr>
<td>Variable Loop</td>
<td>m7G</td>
<td>m7G</td>
<td>m7G</td>
</tr>
<tr>
<td></td>
<td>acp3U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TΨC Loop</td>
<td>m3U</td>
<td>m3U</td>
<td>m3U</td>
</tr>
<tr>
<td></td>
<td>ψ</td>
<td>ψ</td>
<td>ψ</td>
</tr>
<tr>
<td></td>
<td>m1A</td>
<td>m1A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>m5s2U</td>
<td></td>
</tr>
<tr>
<td>Acceptor Stem</td>
<td></td>
<td></td>
<td>m2G</td>
</tr>
</tbody>
</table>

2.4.1.1.1 Dihydrouridine Stem and Loop

The Dihydrouridine stem and loop (D loop) of *E. coli* contains pseudouridine (Ψ), dihydrouridine (D), 2’-O-methylguanosine (Gm) and 4-thiouridine (s4U). (Figure 2.10) Dihydrouridine is found in seven different positions in *E. coli* and in three different positions in both *T. thermophilus* and *B. subtilis* within the D loop and stem. Dihydrouridine is usually only present in the D loop and stem of most organisms, but is occasionally present in the variable loop at position 47. Pseudouridine is present in *E. coli* and *T. thermophilus* at position 13 but completely absent in the D loop of *B. subtilis*. However, *B. subtilis* contains 1-methyladenosine at position 22 in the D loop that is not reported in *E. coli* or *T. thermophilus*. The m1A22 base pairs with base position 13 to form part of the D stem. In *E. coli* and *T. thermophilus*, position 13 is pseudouridine. The base pair (U13-A22) forms a base triplet with residue 46 and is an
important interaction in the folding of the D and TΨC loops to form the L-shaped tertiary structure of tRNA.

In the D loop, the ribose methylated 2’-O-methylguanosine (Gm) is present at position 18 in both *E. coli* and *T. thermophilus* but absent in *B. subtilis* at base position 18. *E. coli* contains 4-thiouridine (s^4^U) in the D stem at position 24. Although there are no reported D stem-loop 4-thiouridines in *B. subtilis* and *T. thermophilus*, both do contain s^4^U at position 8 in the single-stranded region between the aminoacyl stem and D stem. This s^4^U8 is part of a conserved U8-A15 base pair and triple base pair U8-A14-A21 important in maintaining the L-shaped tertiary structure of tRNA. [100] To further explore how these nucleosides may contribute to the overall structure and function of tRNA, the structural and chemical characteristics of the individual nucleosides present in the D loop and stem are considered below.

![Figure 2.10](image)

**Figure 2.10** Structures of modified nucleosides found in Dihydrouridine stem and loop domain in tRNA. Ψ = pseudouridine, D = dihydrouridine, m^1^A = 1-methyladenosine, Gm = 2’-O-methylguanosine, s^4^U = 4-thiouridine.

In a ribonucleoside the relative spatial orientation of the ribose and the base directly affects RNA stability. The ribose can exist in two conformations: 1) C2’-endo where the nucleobase is attached to the ribose in the equatorial position, or 2) C3’-endo where the base is attached in the axial position. Modification of the base and/or sugar can shift the conformational equilibrium to either the C2’-endo or
C3’-endo conformation. The ribose is in the C3’-endo conformation in the most stable A form RNA helix. So, for any ribose or base modification or combination of modifications, if the predominate ribose conformation is C3’-endo, the modification lends stability to the A form RNA helix.

**Pseudouridine (Ψ).** Pseudouridine (Ψ) is the most abundant modified nucleoside, found in all three domains, and contains a C-C glycosidic bond between the ribose and the nitrogenous base rather than a C-N glycosidic bond. This introduces an extra hydrogen at the N1 base position available for hydrogen bonding. Pseudouridine contributes to structural stability of RNA in two ways: 1) through base stacking, [102-104] and 2) by coordination of a water molecule. [105-107] Pseudouridine prefers the C3’-endo ribose conformation which induces greater base stacking and a stable A helix form in both the single stranded and duplex RNA. This stabilization induced by the ribose conformation is propagated to neighboring nucleotides resulting in regional stabilization. This effect is greater on the 5’ side of the nucleotide and may be due to the coordination of the water molecule. The N1H of pseudouridine coordinates a water molecule with the 5’ phosphate oxygen of the pseudouridne and the 5’ phosphate oxygen of the preceding (5’) nucleotide. This further stabilizes the RNA restricting the conformation and mobility by linking the base and RNA backbone.

Pseudouridine is present in many different locations in tRNA, however there is only one pseudouridine present in the D loop of *E. coli* and *T. thermophilus*. In *E.coli*, Ψ13 only occurs in tRNA^Glu^*. This pseudouridine is located at position 13 and likely stabilizes tertiary structure when base pairing with residues at postion 22 to form the triple base pair C13-G22-m7G46. The *E. coli* pseudouridine synthase enzyme (TruD) has been identified and has substrate specificity only for pseudouridine at position 13 and is not essential for cell viability. [108] TruD was the last pseudouridine synthase to be identified in *E. coli* and, surprisingly, does not share sequence homology with the other known pseudouridine synthases. It does, however, share sequence homology with a group of 58 proteins with unknown function. [108]
**Dihydrouridine (D).** Dihydrouridine (D) is the second most abundant modification and found in all three domains and is located primarily within the D loop at positions 16, 17, 20 and 21 in tRNA and occasionally in the variable loop. In *E. coli* only two isoacceptors, tRNA\(^{Tyr}\) and tRNA\(^{Gln}\), do not contain D. Dihydrouridine synthases site-specifically reduce uridines to form dihydrouridine, thus forming the only known nonaromatic modified nucleoside. [109] Interestingly, this enzymatic reduction reaction is reversible; dihydrouridine can be oxidized to uridine. [54]

In contrast to pseudouridine, NMR structural studies in dihydrouridine containing oligonucleotides show dihydrouridine adopts the C2’-endo ribose conformation which disrupts the A form helix causing base destacking. The destabilization of the C2’-endo ribose conformation is propagated to the 5’ neighboring nucleoside in oligonucleotides, but not to the 3’ neighboring residues. [110] This results in a more flexible oligonucleotide, which may be required for loop formation and in tertiary interactions in the “elbow” region where the D loop and TΨC loops fold together. The dihydrouridine content in tRNA of organisms that live in extreme cold (psychrophiles) is 40 to 70% higher than in mesophiles. Dihydrouridine in tRNA may be necessary to maintain conformational flexibility and dynamic motion that would otherwise be restricted at low temperatures. [56] The instability of dihydrouridine due to hydrolytic ring opening at high temperature may explain why hyperthermophiles often do not contain dihydrouridine. [111] Recenly, a bioinformatics analysis revealed three genes (*yjbN, yhdG, yohI*) thought to be responsible for the synthesis of dihydrouridine in *E. coli*. [109]

**1-methyladenosine (m\(^1\)A).** Although not found in *E. coli*, m\(^1\)A at position 58 (m\(^1\)A58) is highly conserved and present in all three domains of life and here is present in *T. thermophilus, B.subtilis* and *L. lactis*. The 1-methyladenosine modification is also found at position 22 in tRNA\(^{Ser}\) and tRNA\(^{Tyr}\) of *B. subtilis*. [112] 1-methyladenosine is one of the few charged nucleosides, with a pK\(_a\) of 8.2 and a positive charge on N1, resulting in different hydrophobic and electrostatic character than that of adenosine [113] (33). The methylation prevents the formation of Watson-Crick type base pairs with other residues. The stacking properties are also altered, as studies show that 1-methyladenosine stacks less extensively than
unmodified adenosine residues. The ability of the methyl group to block base formation with other residues is important for maintaining the correct structure or preventing formation of alternative structures. Specific examples of this are found in mitochondrial tRNA, where alternative structures are prevented by the methylation at N1 in T-armless tRNAs. [115, 116]

TrmI is the enzyme responsible for the synthesis of m1A at position 58 in bacteria [117], archaea [118] and eucarya. [119] TrmI was found to be essential in yeast and is required for growth at high temperatures in T. thermophilus. [120] A different enzyme (TrmK), responsible for site specific formation of m1A at position 22 in B. subtilis, has been identified, but is not essential for growth. [112] Interestingly, orthologues of TrmK have not been found in archaea or eukarya.

2'-O-methylguanosine (Gm). E. coli and T. thermophilus both contain 2'-O-methylguanosine at position 18 in the D loop. B. subtilis does not contain 2'-O-methylguanosine at position 18, however, Gm is present at the wobble position 34 in some B. subtilis tRNAs. In this study, Gm was not found in L. lactis. Methylation to the 2' hydroxyl on the ribose makes the nucleoside resistant to alkaline hydrolysis and digestion by some nucleases. Due to steric constraints involving the 2'-O-methyl on the ribose, the base and the 3'-phosphate, the C3'-endo ribose configuration is preferred thus, as in pseudouridine, confers rigidity to the local RNA structure by stabilizing the A type helix. [121, 122] This 2'-O-methylation of the guanosine likely contributes to the stability in the D loop by stabilizing structure between the flexible dihydrouridine residues often present nearby. The 2'-O-methylation of guanosine at position 18 is also involved in an interloop base pair with Ψ55 in the TΨC loop, which is important in maintaining the L-shaped tRNA tertiary structure. [100]

The gene responsible for this methylation at G18 has been identified in E. coli (TrmH) and in T. thermophilus. [123] TrmH alone does not appear to be essential for growth [124] in E. coli, but mutants deficient in the synthesis of Gm18, m5U54 and Ψ55 were found to exhibit growth defects. [125] The
amounts of both Gm and m^1A were found to increase when *T. thermophilus* was grown at high temperatures (80 °C). [52]

**4-thiouridine (s^4U).** Many tRNAs of *E. coli, T. thermophilus* and *B. subtilis* contain 4-thiouridine (s^4U) at position 8 and, in *E. coli*, it is also present at position 24 in the D loop and stem. The structure of 4-thiouridine (s^4U) is similar to uridine except the keto group at base position 4 is replaced with a sulfur. The large size, polarizability and position of the sulfur are the primary contributors to differences in structure and stability as compared to uridine. The 3’-endo ribose configuration is more stabilized in s^2U than s^4U because the steric clash between the ribose 2’ OH on ribose and sulfur is greater in s^2U than s^4U due to base position of the sulfur. The sulfur alters the imino proton pKa and effects hydrogen bonding. The s^4U modification is stabilizing when involved in reverse Hoogsten s^4U-A base pairing, which occurs in the conserved s^4U8-A14 tertiary interactions. The s^4U modification results in increased base stacking interactions due to the more polarizable sulfur.

One of the unique properties of 4-thiouridine is photosensitivity. 4-thiouridine will form covalent crosslinks to bases in close proximity (3-4 Å) when exposed to near UV light. For example, in *E. coli*, UV light induces this cycloaddition reaction at the 5,6 double bond of the cytidine at tRNA position 13 with the elimination of sulfur to form a stable covalent adduct between s4U8 and C13. [126] This cytidine is part of the conserved triple base pair C13-G22-m7G46 that occurs near the elbow region of the L-shape tertiary structure of tRNA. The tertiary structure is altered and in *E. coli* this induces the stringent response and growth slows dramatically. [127-129] Additionally, the tRNA structure changes such that some aminoacyl synthetases no longer recognize tRNA. [130] Two genes are required for the synthesis of s^4U, *IscS* and *thil*. [131] *IscS* is the cysteine desulfurase which transfers sulfur from cysteine to ThiL. [132]
2.4.1.1.2 The variable loop

The variable loop, as its name implies, varies in the number of bases located there, and, except for the well conserved m7G46acp3UC modified sequence in *E. coli*, other modifications are not usually present here in bacterial and archael species. (Figure 2.11)

![Figure 2.11](image)

**m7G** ![Structure of m7G](image)

**acp3U** ![Structure of acp3U](image)

**Figure 2.11** Structures of modified nucleosides found in the variable loop domain in tRNA. m7G = 7-methylguanosine, acp3U = 3-(3-amino-3-carboxypropyl)uridine.

*N7*-methylguanosine (m7G). *N7*-methylguanosine (m7G) is present at position 46 in tRNA from many bacterial, eukaryal and a few archael species. All of the bacterial and archael organisms studied here contain m7G. The structural characteristics of m7G that contribute to tRNA structure are the positive charge and the methylation at base position N7. The methylation of N7 produces a zwitterion and blocks the N7 from any hydrogen bonding. The negative zwitterionic charge is eliminated by H-bonded base pairing of m7G46 with the G22-G13 tertiary base pair, forming one of the base triplets known to be important in maintaining the L-shape tertiary structure of tRNA. This leaves the positive charge localized on the solvent exposed imidazole ring of m7G in the tRNA. Analogous to m1A, the methylation blocks base pairing at N7 and may also affect stacking interactions. In contrast to base methylated nucleosides such as m2G, m2G, m5U and m5C, where methylation increases hydrophobicity, the methylation on m7G clearly decreases hydrophobicity. Agris has suggested that because the positive
charge of both m\(^1\)A and m\(^7\)G is localized to the solvent accessible side of the base, this may be important in creating areas of positive charge on the RNA important in protein interactions. [133]

The enzyme responsible for formation of m\(^7\)G46 in *E. coli* is TrmB and does not appear to be essential for growth. [134] However, in *T. thermophilus*, TrmB is found to be necessary for cell viability at temperatures > 80 °C and several other modifications (Gm18, m\(^1\)G37, m\(^1\)A58) were decreased in the m\(^7\)G deficient mutants cultured above 70 °C. [52]

**3-(3-amino-3-carboxypropyl)uridine (acp\(^3\)U.)** A modification found in *E. coli* but not present in *B. subtilis*, *T. thermophilus* or *L. lactis* is acp\(^3\)U located at position 47 in the variable loop. This modified nucleoside always appears in the sequence m\(^7\)G46acp\(^3\)U47C48 and is only present in a few tRNAs from bacterial species. Residue 47 is not known to be involved in important tertiary interactions. Few studies have investigated the role of acp\(^3\)U modification in tRNA structure and function, likely because the enzyme responsible for its synthesis has not been identified, thus suitable mutants cannot be created.

### 2.4.1.1.3 The TΨC loop

Present in the TΨC loop in all the bacteria studied here are the well conserved modifications 5-methyluridine (m\(^5\)U54) and pseudouridine (Ψ55). (Figure 2.12) Both *T. thermophilus* and *B. subtilis* also contain m\(^1\)A58, which is not found in *E. coli*. *E. coli* contains an additional pseudouridine at position 65. Structural features of m\(^1\)A and Ψ were discussed in section 2.4.1.1.

![Figure 2.12](image_url) Structures of modified nucleosides found in TΨC loop domain in tRNA. m\(^5\)U = 5-methyluridine, m\(^5\)s\(^2\)U = 5-methyl-2-thiouridine.
5-methyluridine (m$^5$U). The modification, 5-methyluridine (m$^5$U54 or T54), occurs once in almost every tRNA and is highly conserved in bacterial and eukaryal organisms, suggesting the importance of its presence in function. NMR and melting studies suggest the m$^5$U stabilizes local and tertiary structure primarily by enhancing base stacking. The methyl group increases polarizability in the base and this increases the van der Waals interactions among neighboring bases. [135, 136] Both Ψ55 and T54 play major roles in the folding of tRNA into its functional L-shaped tertiary structure. [137] Two of the major interactions that allow the D and T loops to fold correctly are the base pairing between T54-A58 and G18-Ψ55 located in the “elbow” region of the L-shaped tRNA. The combination of deficiency of these important tertiary structural modifications Gm18, m$^5$U54 and Ψ55 does indeed result in defects in growth, translation efficiency and altered metabolism. [125]

The m$^5$U methyltransferase in *E. coli* is encoded by the *trmA* gene and this gene is essential for cell viability. [138] However, it is not the methyltransferase activity that is essential, suggesting *trmA* has multiple functions in the cell. *B. subtilis* does not contain *trmA*; the methyltransferase gene responsible for formation of m$^5$U54 is *trmFO* in *B. subtilis*. Most methyltransferases, including TrmA, use S-adenosylmethionine as the methyl donor; TrmFO uses N$^5$-N$^{10}$-methylene tetrahydrofolate for incorporation of the methyl group. [139]

5-methyl-2-thiouridine (m$^5$s$^2$U or s$^2$T). One particular modification not present in the mesophillic bacteria studied here but present in thermophiles such as *T. thermophilus* is 5-methyl-2-thiouridine (m$^5$s$^2$T or s$^2$T). [140] This modification is located at position 54 in thermophiles, whereas mesophilic bacteria contain the well-conserved m$^5$U54. Whether modified with a methyl group alone or a methyl and a thionyl group, the position 54 uridine is critical in folding of D and T loops into the L-shape tRNA structure at the elbow region through the s$^2$T54-m$^1$A58 base pair. Early studies have demonstrated s$^2$T has a role in the thermostability of tRNA where NMR melting profiles show that s$^2$T stabilizes the L-shaped structure of tRNA. [136] The melting temperature of tRNA was shown to increase with increasing s$^2$T content in *T. thermophilus*. [41] Structurally, the stability of s$^2$T can be attributed to the
preference for the C3’-endo ribose conformation induced by the steric clash between the large sulfur and the ribose 2’-OH. Stabilizing base stacking interactions can also contribute through the increase in van der Waals interactions as described for 5-methyluridine. NMR and CD studies show the more rigid C3’-endo ribose structural conformation is preferred in s^2T as compared to m^5U and is further evidence of the role of structural stabilization for s^2T. [136] The relative amounts of s^2T present in unfractionated tRNA of the hyperthermophile *Pyrococcus furiosus* also increase with increasing culture temperature. [40] Additionally, studies found that s^2T modified tRNA functions more efficiently at higher temperature in the translational process. [141]

The biosynthetic pathway for the synthesis of s^2T is not yet completely elucidated but involves two cysteine desulfurases (IscS or SufS), two tRNA-two-thiouridine synthesizing proteins (TtuA and TtuB) and an ATPase (tRNA-two-thiouridine C, TtuC). [142] Another modification of the T loop, m^1A58, is required for efficient synthesis of the s^2T modification in *T. thermophilus*. [95] No growth defects were observed in TruA and TtuB mutant strains of *T. thermophilus* when grown below 75 °C. However, both mutants failed to grow above 80 °C, indicating the s^2T modification is required for cell growth at high temperatures. [143].

### 2.4.1.1.4 The acceptor stem

The acceptor stem consists of a seven base paired stem formed by pairing the first seven residues of the 5’ terminus with the last seven residues of the 3’ terminus (not including the conserved CCA tail and residue 73) of the tRNA. This region of tRNA is rarely modified in bacterial species and infrequently modified in eukaryotes. The only modified nucleoside found in the acceptor stem of the bacterial species studied here was N^2-methylguanosine (Figure 2.13) found at position 6 in *T. thermophilus* tRNA. [144]

Yeast contains a 2’-O-methylcytidine at position 4 and its function is unknown. [145] The structure and sequence of the aminoacyl stem loop often serves as an identity element in the recognition of a
particular isoacceptor for its corresponding aminoacyl-tRNA synthetase. The residue at position 73 is also an identity element for aminoacyl synthetases.

![Figure 2.13](image.png)

**Figure 2.13** Structure of modified nucleoside found in the acceptor stem domain. $m^2G = N^2$-methylguanosine

$N^2$-methylguanosine ($m^2G$). $N^2$-methylguanosine contains a single methylation on the exocyclic amine (N2) and can exist in a cis or trans rotamer with unrestricted rotation. Methylation does not interfere with the hydrogen bonding ability of $m^2G$ and base pairs form with C, A and U. Thermodynamic characterization using melting studies and NMR of oligonucleotide tetraloops containing $m^2G$ modified and unmodified G revealed there is no significant stabilizing effect on secondary structure and equal preference for the cis and trans forms. [146] However, in a study of self-complementary oligoribonucleotides, $m^2G$ was found to contribute significantly to duplex stabilization. [147] Thus, $m^2G$ has a role in stabilization of stem or duplex regions of tRNA. In *T. thermophilus*, $m^2G6$ located in the acceptor may contribute stability to this stem region or be involved in aminoacyl tRNA synthetase recognition. The absence of $m^2G6$ in bacterial mesophiles may indicate this modification is required for stability at high temperatures.

*T. thermophilus* is the only currently known bacterial organism where $N^2$-methylguanosine is present at position 6 [29] and the corresponding enzyme is unknown. However, an archaeal methyltransferase,
Trm14, has recently been identified that is responsible for the formation of m$^2$G at position 6 in archaean tRNA. [148]

### 2.4.1.2 Posttranscriptional Modifications in the Anticodon Stem and Loop (ASL)

The posttranscriptional modifications found in the anticodon stem loop domain (ASL) serve to pre-structure the ASL for accurate and efficient codon binding and those found in the bacteria studied here are shown in **Table 2.5.** For *E.coli* and *B.subtilis*, the sequence location is known for all modified nucleosides and a few sequences are known for *T. thermophilus*. [29] Sequence information is not known for modified nucleosides found in *L. lactis* but locations are likely similar to *B. subtilis* as both are mesophilic gram positive bacteria. These modifications give the aminoacyl tRNA anticodon domain the correct structure and base pairing properties to base pair with the corresponding codon on the mRNA in the A site of the ribosome. Once this pairing is verified as structurally correct, then the peptidyl transferase reaction can take place. The ASL consists of a five base-pair A-form helix capped by a seven nucleotide loop (residues 32-38) which contains the anticodon (residues 34,35,36). Several characteristic features and tertiary interactions of the tRNA ASL are known to contribute to the structure of the ASL. Position 32 is usually a pyrimidine (C or U) and position 33 is almost always a uridine. The first anticodon position, 34, is frequently modified with a diverse set of modifications. Positions 37 and 38 are usually purines (G, A) and adenosines at position 37 are frequently hypermodified. The U-turn structural motif is common, but not always present, in the ASL and is formed by a sharp turn at U33 which results in an abrupt change in direction of the RNA phosphodiester backbone. [101] The anticodon bases stack on the 3’ side of the loop and are well-positioned to base pair with the codon. Typically, the three anticodon nucleotides base pair with the three codon nucleotides to form an A-type double helical RNA structure in the A site of the ribosome and make contact with ribosomal residues, A1493, A1492 and G530. Modifications at position 37 immediately 3’ to the anticodon stabilize this tRNA-mRNA interaction by providing the proper ASL structure and dynamics. Wobble modifications (position 34) in the anticodon expand or restrict codon base pairing with alternate hydrogen bonding or steric interactions.
In the organisms studied here, distinct differences are noted in the variety of modified uridines present at position 34 (wobble position) and the different hypermodified adenosines occurring at position 37.

**Table 2.5** Modifications in the Anticodon Stem Loop Domain.

<table>
<thead>
<tr>
<th>Position</th>
<th><em>E. coli</em></th>
<th><em>B. subtilis</em></th>
<th><em>T. thermophilus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>ψ</td>
<td>ψ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s^2C</td>
<td>s^2C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cm</td>
<td>Cm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Um</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td></td>
<td>k^2C</td>
<td>k^2C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cm</td>
<td>Gm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ac^6C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cmo^5U</td>
<td>mo^5U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mnm^5U</td>
<td>cmnm^5U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mnm^5s^2U</td>
<td>cmnm^5s^2U</td>
<td>mnm^5s^2U</td>
</tr>
<tr>
<td></td>
<td>cmnm^5Um</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>m^1G</td>
<td>m^1G</td>
<td>m^1G</td>
</tr>
<tr>
<td></td>
<td>m^3G</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>m^6A</td>
<td>m^6A</td>
<td>m^6A</td>
</tr>
<tr>
<td></td>
<td>t^6A</td>
<td>t^6A</td>
<td>t^6A</td>
</tr>
<tr>
<td></td>
<td>m^6t^6A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ms^2t^6A</td>
<td>ms^2t^6A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>i^6A</td>
<td>i^6A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ms^2i^6A</td>
<td>ms^2i^6A</td>
<td>io^6A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ms^2io^6A</td>
</tr>
</tbody>
</table>
2.4.1.2.1 Position 32

Position 32 at the 5' end of the anticodon is always a pyrimidine, C or U. Modifications known to be present at position 32 in the ASL are pseudouridine (Ψ), 2-thiocytidine (s\textsuperscript{2}C), 2'-O-methylcytidine (Cm) and 2'-O-methyluridine (Um). (Figure 2.14) All four modifications are found in the *E. coli* tRNA analyzed in this study. These position 32 modifications, in concert with modifications at the wobble position and position 37, have been shown to stabilize the ASL and introduce the correct structure and dynamics for codon recognition. [149, 150] Position 32 can form hydrogen bonds with position 38, which is often a pseudourine resulting in a closing of the ASL loop or restricting the loop size. [104]

![Figure 2.14](image)

*Figure 2.14. Structures of modified nucleosides present at position 32 in tRNA. Cm = 2'-O-methylcytidine, Um = 2'-O-methyluridine, Ψ = pseudouridine, s\textsuperscript{2}C = 2-thiocytidine.*

2'-O-methylcytidine (Cm), 2'-O-methyluridine (Um) and Pseudouridine (Ψ). Methylations to the 2'-OH of the ribose cause the ribose to adopt primarily the 3'-endo ribose configuration, which results in the most stable A form RNA helical structure. In pyrimidines, the 3'-endo ribose configuration minimizes the steric clash between the ribose methyl group and the oxo group at base position C2. In *E. coli*, methylation of C and U at position 32 is carried out by the same enzyme, TrmJ, the *yfhQ* gene product. [151]
Pseudouridine also imparts stabilization to RNA structure as discussed in section 2.4.1.1.1 of this chapter. The dual function enzyme in *E. coli* responsible for pseudouridylation at position 32 in tRNA and position 746 in 23S RNA is RluA, the gene product of *yabO*. [152]

2-thiocytidine (s²C). In *E. coli*, 2-thiocytidine (s²C) is present in tRNA^{Arg} with anticodons ICG, CCG and mnm^{5}UCU. 2-thiocytidine is also present in tRNA^{Ser} with anticodon GCU. [29] 2-thiocytidine is found in bacterial and archaeal organisms but thus far unknown in eukaryal organisms. Replacement of the oxygen at the 2 position on the cytidine with the larger and more polarizable sulfur to form s²C alters both hydrogen bonding and metal binding properties. The 3′-endo ribose configuration is favored to minimize the steric interactions between the ribose 2′-OH and the very large sulfur on the cytosine. The sulfur is a better nucleophile and will bind divalent metals more strongly. Modification also alters the hydrogen bonding properties of 2-thiocytidine; the pKa of cytidine is 4.2 and that of 2-thiocytidine is 3.4. [153, 154] [150, 155]

IscS is required for the biosynthesis of all thionucleosides in *E. coli*, including s²C. [153] The product of the *ttcA* gene is also required for thiolation of C32 to form s²C32. [154] The lack of s²C causes no apparent growth defects or differences in sulfur metabolism. The impact of s²C depends on the codon being read. Jager investigated the coding properties of wild type and s²C modification deficient strains and found a lack of s²C32 in tRNA^{Arg}_{ICG} did not affect selection rate for codons CGU, CGC or CGA but s²C32 increases selection of AGG codon in tRNA^{Arg}_{mmn5UCU}. Jager concluded s²C32 participates in the decoding process and may modulate flexibility of the ASL for optimum decoding. Cantara studied the conformational dynamics and stability of tRNA^{Arg}_{ICG} ASL constructs with and without s²C32. These constructs showed no U-turn structure but efficiently bound CGU, CGC and CGA codons in the ribosomal A site. As compared to the unmodified ASL constructs, the s²C32 modified constructs showed reduced thermal stability and the s²C32 and m²A37 modified constructs had decreased base stacking interactions. Cantara suggested by allowing loop flexibility s²C32 and m²A37 modifications modulate
ASL conformation in these constructs to restrict codon recognition to CGU and CGC, not the rare CGA codon. [150]

2.4.1.2.2 Position 34; The Wobble Position

Most tRNAs read or decode more than one triplet mRNA codon, and modification is an important part of this ability. The 1st and 2nd bases of the codon (NI, NII) pair with the 3rd and 2nd base of the tRNA anticodon (N36, N35) according to normal Watson Crick base pairing rules. The base pairing between the 1st base of the anticodon (N34) and the 3rd base (NIII) of the codon can be normal Watson-Crick base pairing, but often these bases form alternative, nonstandard base pairs, referred to as wobble and this is the key to expanding the decoding ability of tRNAs and minimizing the number of tRNAs required for translation. These wobble modifications modulate recognition of the 61 different codons by either restricting, expanding or altering the decoding properties of the tRNAs so that only a limited set of tRNAs is needed for decoding all 61 codons. In most tRNAs the wobble position (position 34) contains U, C, G or I and these bases are often modified. The bacterial modifications identified in this study and known to be located at the wobble position are listed in Table 2.4. The structural characteristics of these modifications are discussed below in terms of their contribution to decoding and ASL loop conformation.

2.4.1.2.2.1 Modified Uridines at the Wobble Position

Two types of modified uridines at the wobble position, 34, are frequently encountered to ensure the correct anticodon-codon base pairing for optimum accuracy and efficiency of translation, 1) 5-substituted uridines (xm^5U) and 2) methoxy 5-uridines (xmo^5U). (Figure 2.15) The 5-substituted uridines (xm^5U) are often base thiolated (xm^s^2U) or ribose methylated (xm^5Um) and occur in tRNAs that read codons ending in G or A. The 5-methoxy uridine derivatives read codons ending in A, G, C and U. These modifications to U34 tend to be domain specific; mnms^5^2U and cmnm^5^2U derivatives occur in bacterial tRNAs, mcm^5^2U derivatives occur in eukaryotes and tm^5^2U derivatives occur in mammalian mitochondrial tRNAs. Modified 5-substituted uridines in archaeal tRNA appear much less diverse in the
few organisms studied to date. The modified nucleosides present in archaea may include uridine derivatives characteristic of both bacterial and eukaryal organisms. The modified uridines found in *B. subtilis* were mo<sup>5</sup>U, cmmm<sup>5</sup>s<sup>2</sup>U and cmnm<sup>5</sup>s<sup>2</sup>U, those in *E. coli* were cmo<sup>5</sup>U, mnm<sup>5</sup>U, mmm<sup>5</sup>s<sup>2</sup>U and cmnm<sup>5</sup>Um. In this study, cmnm<sup>5</sup>s<sup>2</sup>U was found in *L. lactis* and mnm<sup>5</sup>s<sup>2</sup>U in *T. thermophilus*.

a) 5'-substituted uridines (xm<sup>5</sup>U)

![Structures of two types of modified uridines commonly found at position 34 in tRNA of bacterial species](image)

b) 5'-methoxy uridines (xmo<sup>5</sup>U)

![Structures of two types of modified uridines commonly found at position 34 in tRNA of bacterial species](image)

**Figure 2.15** Structures of two types of modified uridines commonly found at position 34 in tRNA of bacterial species a) 5'-substituted uridines (xm<sup>5</sup>U), mnm<sup>5</sup>U = 5-methylaminomethyluridine, mnm<sup>5</sup>s<sup>2</sup>U = 5-methylaminomethyl-2-thiouridine, cmnm<sup>5</sup>s<sup>2</sup>U = 5-carboxymethylaminomethyl-2-thiouridine, cmnm<sup>5</sup>Um = 5-carboxyethylaminomethyl-2'-O-methyluridine and b) 5'-methoxy uridines (xmo<sup>5</sup>U), mo<sup>5</sup>U = 5-methoxyuridine, cmo<sup>5</sup>U = 5-carboxymethoxyuridine.

5'-methylene substituted Uridines (xm<sup>5</sup>U, xm<sup>5</sup>s<sup>2</sup>U, xm<sup>5</sup>s<sup>2</sup>Um) The xm<sup>5</sup>U, xm<sup>5</sup>s<sup>2</sup>U, xm<sup>5</sup>s<sup>2</sup>Um modifications are present in *E. coli* tRNA<sup>lys</sup>, tRNA<sup>Glu</sup>, tRNA<sup>Glu</sup> and tRNA<sup>Arg</sup> that recognize purine ending
codons, G and A. NMR studies show the ribose adopts predominately the more stable C3'-endo ribose conformation in the xmn5U and xms5U type nucleosides. The 5-substituent at least as long as the methylaminomethyl group contributes to the stability of the 3’ endo ribose conformation, primarily through steric effects. [156] A smaller 5-methyl substitution actually stabilizes the C2'-endo ribose configuration. The largest contribution to stability of the C3'-endo ribose form is through the steric interactions between the large sulfur at the 2 position of the base and the 2'-OH of the ribose. Additionally, NMR studies found the 5’-phosphate group also contributes to the stability of the xmn5U modification. [157] Due to the preference for the C3'-endo ribose configuration, which is stabilized largely by steric effects, this type of modified U can only form Watson-Crick base pairs. The formation of non Watson-Crick base pairs requires a more flexible nucleoside conformation. [158] These modified uridines also will bind to G but with much less efficiency. A model has been proposed, but not experimentally verified, that suggests for binding to G the modified U is in protonated form. In the xmn5U type modifications, the positively charged amino nitrogen withdrawals electrons from the uracil ring, lowering the pKa of the N3 of uracil. This forms a partially ionized xmn5U+, which can form two base pairs with G(III) in the A site while still maintaining the rigid C3'-endo ribose configuration. For the thiolated derivatives, the ionization could give the sulfur a negative charge and allow it to become a proton acceptor. [159]

Modified uridines at position 34 in concert with position 37 modifications also influence the structure of the ASL necessary for efficient and accurate codon binding. The effects on ASL structure were compared for mcm5s2U and mnm5s2U in tRNA5'vs anticodon stem loops by NMR. The mcm5 modification rotates freely and is not involved in any hydrogen bonding interactions and appears to have no effect on loop conformation. Addition of the sulfur to the modification increases the formation of the stable 3’-endo ribose conformers, increases base stacking and promotes a more defined U-turn. The sulfur in mnm5s2U also stabilizes the C3'-endo ribose configuration. The mnm5 moiety stabilized the C3'-endo ribose configuration to a greater extent than mcm5. Bases 34-35-36-37 are stacked resulting in an even
more definite U-turn motif for the mnm\textsuperscript{5}U modified ASL. The amino group within the mnm\textsuperscript{5}U side chain is positively charged at pH 7.5 and may contribute to the stability of the anticodon-codon further through charge neutralization as base pairs are formed with the codon. [160]

**Methoxy 5-substituted Uridines (xmo\textsuperscript{5}U)** In contrast to the restrictive base pairing properties of the xm\textsuperscript{5}U type modifications, the xmo\textsuperscript{5}U type modification expands base pairing to include all four bases allowing reading of synonymous codons. While the xm\textsuperscript{5}U type modification exists predominately in the 3'-endo ribose configuration, the xmo\textsuperscript{5}U modification can adopt both the the C2' and 3'-endo ribose conformations. NMR studies show that the \(-\text{OCH}_2\)- group interacts with the 5’ phosphate and contributes to the stabilization of the C2'-endo ribose conformation. The small 5-substituent of xo\textsuperscript{5}U also stabilizes the C2'-endo ribose conformation.

In the cmo\textsuperscript{5}U-A base pair, the predominate ribose conformation is C3'-endo and in this configuration Watson-Crick base pairs are formed. The 2'-endo ribose conformation predominates in the cmo\textsuperscript{5}U-U base pair and enables non Watson-Crick base pairing. Conformation of the cmo\textsuperscript{5}U ribose interconverts between C3'-endo and C2'-endo forms in the cmo\textsuperscript{5}U-\text{G} base pair. [158] These solution studies show how, in solution, the cmo\textsuperscript{5}U modification is stable when the conformation of the ribose is C2'-endo but can also exist as the C3'-endo form. The ability of cmo\textsuperscript{5}U to exist in both forms allows base pairing with all bases, thus the modification is flexible and this contributes to the efficiency of decoding. In contrast, the xm\textsuperscript{5}U modification is restricted to the C3'-endo ribose conformation and can only pair well with A and somewhat with G. This rigid modification contributes to the accuracy of decoding.

Crystal structures of the modified ASL of *E. coli* tRNA\textsuperscript{Val}_{cmo5UAC} in the decoding site of the 30S ribosomal subunit were determined and the molecular interactions of the cmo\textsuperscript{5}U34-A(III), cmo\textsuperscript{5}U34-G(III), cmo\textsuperscript{5}U34-U(III) and cmo\textsuperscript{5}U34-C(III) base pairs were described. [161] Based on these crystal structures of the tRNA\textsuperscript{Val} ASL containing cmo\textsuperscript{5}U34 in the decoding site of the 30S ribosomal subunit, the cmo\textsuperscript{5}U34 ether oxygen stabilizes an ASL conformation by forming a hydrogen bond with the 2'OH of the
conserved U33 ribose. This orientation allows base pairing with the A, G, U and C of the third position of the codon. Two hydrogen bonds are formed in Watson-Crick type base pairing between cmo5U34 and A(III). Three hydrogen bonds are formed between the cmo5U34 and G(III) base pairs. The base pairing of cmo5U34 to G induces tautomeration and the enol form is stabilized. The role of the –CH3 or –CH2COOH groups is thought to be to stabilize the enol form of the nucleoside. There is one hydrogen bond formed in each of the cmo5U34 base pairs formed with U(III) and C(III). In the cmo5U34-C(III) base pair fewer base stacking interactions are observed than for mo5U34-U (III) base pairs, which indicates the cmo5U34-C base pairing interactions are weaker. [29, 161]

Clearly, xmo5U type modifications expand the decoding ability of tRNA through direct base pairing with codon bases. Additionally, in tRNA^Val_{cmoUAC} ASL constructs the cmo5U modification functions to maintain a stable, open and ordered loop conformation necessary for codon binding by enhancing base stacking. [162]

In B. subtilis mo5U occurs in tRNAs specific for alanine, proline, serine, threonine and valine. [29] The cmo5U34 modification in E. coli is used to decode the same codons. The genes necessary for the synthesis of these modifications have been identified in the bacterium, Salmonella enterica serovar Typhimurium. [163] In salmonella, tRNAs specific for proline, alanine, and valine with cmo5UGG at the wobble position 34 were found to read codons ending in U, A, G and C with the presence of cmo5U being especially necessary when decoding U and C ending codons. [163] [164] In B. subtilis, tRNA(mo5UGG) was also able to read all 4 codons, CCU, CCC, CCA and CCG that are specific for proline. [165] However, B. subtilis valine specific tRNA(mo5UAC) can read GUA, GUG and GUU but not GUC. [166] Interestingly, in B. subtilis, tRNA^Pro_{mo5UGG} contains pseudouridine at position 32 which is not found in valine specific tRNA, and it is postulated that Ψ32 may enable the ASL to adopt a structure that assists in mo5U(43) and C(III) base pairing. [165]
2.4.1.2.2.2 Other Posttranscriptional Modifications at the Wobble Position

**Figure 2.16.** Structure of other modified nucleosides at the Wobble position in tRNA. I = inosine, Q = queuosine, k\(^2\)C = lysidine, ac\(^4\)C = 4-acetylcytidine.

**Inosine (I).** Inosine occurs at position 34 in tRNA\(^{Arg} \) (ICG) in E. coli for reading codons CGU, CGC and CGA. Inosine is formed by deamination of adenosine and is structurally similar to guanosine but without the 2-amino group and forms two hydrogen bonds similar to guanosine with C and U (Figure 2.16). Inosine also forms two hydrogen bonds with adenosine which guanosine cannot. This I-A, purine-purine, base pair is wide and thus was predicted to destabilize the decoding center on the ribosome. [167] However, crystal structures show that within the decoding center of the 30S ribosome the base pair width is not highly constrained and the geometry of the anticodon changes only slightly to accommodate this wide pair. The I-A pair is recognized by the decoding center as cognate and this wide pair does not cause disruption within the ribosomal A site. [168]

Inosine is formed by hydrolytic deamination of adenosine. In *E. coli*, TadA is the enzyme responsible for the formation of inosine and it appears to be an encoded by an essential gene. However, whether it is the absence of I34 or the absence of the TadA protein that will not allow cell growth is not known. [169] In yeast, I34 is necessary for cell cycle progression. [46]
**Queuosine (Q).** Queuosine occurs at position 34 in the anticodon of tRNAs specific for histidine, asparagine, aspartate and tyrosine and recognizes NAU or NAC codons. Queuosine base pairs with both C and U, forming three Watson-Crick type hydrogen bonds to C and U. Queuosine seems to be involved in codon choice and may prefer binding with U, but experimental examples are limited. [170].

Queuosine is a positively charged hypermodified guanosine analogue and is one of the most complex modifications with a cyclopentenediol side chain and a 7-deazaguanine ring. The cyclopentenediol side chain and positively charged linker is likely solvent exposed and away from the base pairing side, thus it may be involved with interactions other than decoding. The formation of queuosine involves the only known base-exchange reaction where the queuosine is exchanged for a genomically encoded guanine. In higher eukaryotes queuosine is further modified with the addition of sugars such as mannose and galactose to manQ and galQ. In E. coli, amino acids such as glutamate can also be attached to give glutamyl-queuosine. [171] Queuosine has been implicated in frameshifting in the A site. [28] Queuosine modification does not appear to be essential to the cell, however modification deficient strains of E. coli do show growth defects [172] and queuosine is important in maintaining cell viability under unfavorable growth conditions. [173] Much of the interest in queuosine is its unique and complicated biosynthesis and its involvement in other cellular processes such as, virulence in Shigella flexneri [174], cell differentiation [175] and cancer processes. [50]

**Lysidine (k²C) and N⁴-acetylcytidine (ac⁴C).** In E. coli, tRNA\textsuperscript{Ile²} and tRNA\textsuperscript{Met} have very similar sequences including the same anticodon sequence, CAU, yet decode different codons. The precise and correct codon recognition of these two tRNAs is determined by the different modifications present on the “C” of the anticodon. The wobble modification N⁴-acetylcytidine (ac⁴C) ensures precise recognition of AUG codons for tRNA\textsuperscript{Met}; and the similar codon AUA is decoded by tRNA\textsuperscript{Ile²} containing a lysidine (k²C) at the wobble position.
\(N^4\)-acetylcytidine in tRNA\(^{\text{Met}}\) ensures precise decoding of AUG and prevents misreading of AUA codon by stabilizing the C-G base pair. Because the C3'-endo ribose configuration is predominate in \(N^4\)-acetylcytidine the C-G base pairing is stabilized thus ensuring precise decoding of AUG and not AUA. [176]

Theoretical calculations predict the most stable conformation of lysidine is with the lysidine moiety trans to the N1 of cytidine. [177] In this orientation lysidine cannot form Watson-Crick base pairs with guanosine, thus preventing misreading of AUG codons. Only the tautomeric form of lysidine provides compatible donor acceptor sites to pair with adenosine. The N3 of the lysidine tautomer donates one hydrogen to the N1 of adenosine and the N4 of lysidine accepts one hydrogen from N6 of adenosine.

These two tRNAs (Ile, Met) are very similar in sequence including other modifications present. The only differences lie in one or two base pairs in the acceptor stem and in the anticodon stem. The anticodon sequence is the same in each tRNA, although the modification on cytidine is different. The enzymes responsible for the synthesis of these modifications ac\(^4\)C and k\(^2\)C, are TmcA and Tils, respectively. [178, 179] It is interesting that the enzymes use these few sequence differences to discriminate between the two RNAs. Lysidine synthetase (TilS) recognizes the two different base pairs in the acceptor stem and cytidine acetyltransferase uses the two different base pairs in the anticodon stem as positive determinates. [178] The base pairs that function as positive determinates in one enzyme, function as negative determinates in the other.

**2'-O-methylcytidine (Cm).** 2'-O-methylcytidine is found at position 34 in tRNA\(^{\text{CmAA}}\) specific for Leucine. Methylation of the 2'-OH of the ribose is a common modification that enhances the stability of the RNA structure, and at position 34 in the anticodon loop the Cm modification induces conformational rigidity in the anticodon-codon complex. The C3’-endo ribose configuration of 2’-O-methylcytidine is stabilized primarily by the steric interaction between the 2-carbonyl group of cytidine, the 2’-methoxy group of the ribose and the 3’-phosphate. These interactions destabilize the C2’-endo
ribose configuration, thus stabilizing the C3'-endo ribose configuration. The methylation of the 2'-OH of the ribose also inhibits RNA hydrolysis by blocking the ability of the 2'-OH to function as a proton donor. [122] Conformational rigidity is required to restrict codon recognition of tRNA^{Leu}_{CmAA} and tRNA^{Leu}_{cmmm5U/mAA} to leucine codons UUA and UUG and prevent reading of phenylalanine codons UUU and UUC. Modification at position 37 is usually present in these tRNAs to contribute to the stability and correct ASL structure. The enzyme responsible for this site-specific methylation in E. coli has recently been identified. It was shown to require a pyrimidine at position 34, adenosine at position 35, and fails to methylate without the presence of ms^{2,6}A at position 37. [180]

2.4.1.2.3 Posttranscriptional Modifications at Position 37

Hypermodified adenosines with hydrophobic side chains are often found at postion 37 directly adjacent to the 3’ side of the anticodon in tRNAs that read codons beginning with A or U. Approximately 80% of tRNAs contain an A at position 37 and 20% have a G at position 37. The most common modifications at this position are t^6A, i^6A and m^1G. These modifications maintain the structure of the anticodon-codon loop that otherwise is destabilized by the intrinsically weak A-U pairing. Transfer RNAs with G and C bases in the anticodon usually do not contain modifications at position 37.

m^1G

Figure 2.17 Structure of modified nucleoside often present at position 37 in tRNA. m^1G = 1-methylguanosine.
**1-methylguanosine (m\(^1\)G).** In Bacteria, Eukarya and Archaea, guanosine at position 37 in tRNA is frequently modified by methylation at N1. 1-methylguanosine (m\(^1\)G37) is present in *E.coli* tRNA\(^{Pro}\), tRNA\(^{Arg}\) and tRNA\(^{Leu}\) which read codons CCN(Pro), CGG(Arg) and CUN(Leu), respectively. The N1 methyl modification blocks the N1 of guanosine from H bonding to N3 of cytidine, thus the classic G-C Watson-Crick base pair containing three hydrogen bonds cannot be formed. Methylation also produces a more hydrophobic base and increases base stacking. At position 37, 3’-adjacent to the anticodon in the stem loop, the m\(^1\)G37 does not form Watson-Crick hydrogen bonds, thus disrupts intra-loop base pairing. Formation of stable intra-loop base pairs adversely affects anticodon-codon binding. NMR studies show m\(^1\)G37, indeed, does not form intra-loop base pairs or stack with other bases in the interior of the ASL loop. m\(^1\)G37 induces an open loop conformation for specific codon binding. The m\(^1\)G37 base extends into the solution on the outer side of the ASL loop with parallel orientation to the positively charged adenosine at position 38 (A\(^+\)38) and was found to increase the stability of the Cm32-A\(^+\)38 base pair located at the top of the ASL loop possibly through base stacking interactions. [181] These NMR studies show the structure that m\(^1\)G37 imposes on the ASL is ordered and exhibits reduced conformational dynamics when compared to unmodified ASL. The ASL loop structure modulated by the m\(^1\)G37 residue prevents frameshifting [182, 183] and contributes to codon recognition on the ribosome. [149]

The enzyme, tRNA( guanine-N1-methyltransferase, TrmD), specifically methylates the N1 position of G37 using S-adenosyl methionine as the methyl donor. TrmD is essential for growth in gram positive bacteria such as *Streptococcus pneumonia*. [184] Modification deficient strains in the gram negative bacterium, *Salmonella typhimurium*, show a reduction in growth rate, but TrmD does not appear to be essential for growth. [185]
Figure 2.18 Structures of \( N^6 \)-threonylcarbamoyladenosine (\( t^6A \)) and derivatives found in tRNA. \( ms^2t^6A = N^6 \)-threonylcarbamoyladenosine-2-methylthioadenosine, \( m^6t^6A = N^6 \)-methyl-\( N^6 \)-threonylcarbamoyladenosine.

**Derivatives of \( N^6 \)-threonylcarbamoyladenosine (\( t^6A \)).** \( N^6 \)-threonyladenosine (\( t^6A \)) is universally conserved and found in all organisms at position 37 in tRNAs that read codons beginning with A. \( N^6 \)-threonylcarbamoyl contains both a carbonyl group and a threonine attached to the amino group of adenosine. The methylthio derivative, 2-methylthio-\( N^6 \)-threonylcarbamoyladenosine (\( ms^2t^6A \)), is also found throughout the three domains of life and is found in all of the organisms studied here except \( E. coli \). In \( E. coli \) the closely related \( N^6 \)-methyl-\( N^6 \)-threonylcarbamoyladenosine (\( m^6t^6A \)) is found. The \( N^6 \)-threonylcarbamoyladenosine modification and its methylthiolated derivatives stabilize codon-anticodon binding and maintain the reading frame. [28, 160, 168, 186-189]

Crystal structures and NMR studies give clues as to how this modification enhances the structure and stability of the anti-codon/codon duplex. [160, 168, 186, 190] \( N^6 \)-threonyladenosine and derivatives do not change the conformation of the anticodon stem but do pre-structure the loop into an open loop conformation required for codon-anticodon binding. The bulky threonylcarbamoyl moiety maintains this open loop structure by sterically preventing base pairing between U33 and A37.
The ureido group of the \( \text{t}^6 \text{A} \) modification (HN-CO-NH) attached to the N6 of the adenine ring forms a stable ring structure by forming a hydrogen bond between the N1 of the adenine and the N11 of the modification. The delocalization of charge within the hydrogen bound ureido group acts as a third heterocycle and is coplanar with the adenine ring. The threonyl moiety is not coplanar with the adenine ring but rotates freely about the N11-C12 bond. The threonyl moiety sterically prohibits stacking of A37 with U36. This allows the adenine ring of \( \text{t}^6 \text{A} \) to stack with the base of the first codon (A1) position and the ring-like ureido group to stack with A38 forming a cross-strand base stacking interaction. Thus, \( \text{t}^6 \text{A} \) stabilizes the codon-anticodon interaction through both the anticodon intrastrand base stacking interactions with A38 and inter-cross strand base stacking interactions with the first position of codon (A1).

Methylations or methylthiolations occur on the 2 position of the adenine base in some \( \text{t}^6 \text{A} \) derivatives (m\( ^2 \text{t}^6 \text{A} \), m\( ^6 \text{t}^6 \text{A} \)). It is postulated that the methyl and methylthio groups may enhance the interaction with the first base of the codon. [160] In fact, a recent study showed the 2-methylthio moiety enhances base stacking over the first codon position. [191] The 2-methylthio groups have been shown to enhance thermostability of N\( ^6 \)-alkyladenosines in RNA hairpins. [192] At least four proteins are involved in the biosynthesis of N\( ^6 \)-threonylcarbamoyladenosine in \( \text{E. coli} \), YrdC, YgiD, YeaZ [193] and YjeE. [194] YrdC and YgiD are part of universally conserved protein families and YjeE and YeaZ are specific to \( \text{E. coli} \). In \( \text{E. coli} \), yrdC appears to be essential for growth. [193] It is unclear whether the lack of cell viability is due to the absence of \( \text{t}^6 \text{A} \) or yrdC has another critical function in the cell. In \( \text{E. coli} \), N\( ^6 \)-methyl-N\( ^6 \)-threonylcarbamoyladenosine (m\( ^6 \text{t}^6 \text{A} \)) is present in tRNA(GGU) in two threonine species. The \( \text{tsaA} \) gene is involved in the biosynthesis of ms\( ^2 \text{t}^6 \text{A} \) and the presence of the methyl group is important in efficient codon recognition in tRNA\( ^{\text{Thr}}_{\text{GGU}} \). [195]
Figure 2.19  Structures of $N^6$-isopentenyladenosine ($i^6A$) and derivatives. $ms^2i^6A = 2$-methylthio-$N^6$-isopentenyladenosine, $io^6A = N^6$-(cis-hydroxyisopentenyl)adenosine, $ms^2io^6A = 2$-methylthio-$N^6$-(cis-hydroxyisopentenyl)adenosine.

**Derivatives of $N^6$-isopentenyladenosine ($i^6A$).**  $N^6$-isopentenyladenosine ($i^6A$) or the thiolated derivative 2-methylthio-$N^6$-isopentenyladenosine ($ms^2i^6A$), are present at position 37 in tRNAs that read codons that begin with U (Figure 2.19)  The hydroxylated versions, $io^6A$ and $ms^2io^6A$, are also found at position 37 in some bacteria and eukarya. These modifications structure the anticodon loop structure to ensure proper positioning and geometry for interactions with the A-site on the ribosome. The structural and dynamic effects, measured by NMR spectroscopy and melting experiments, of $N^6$-isopentenyladenosine on the anticodon stem loop reveal several characteristic features. [195-197]  The bulky isopentenyl moiety disrupts hydrogen bonding in the anticodon loop between C32-A38$^+$ and A37-U33, primarily by steric hindrance. This opens the helical stem region adjacent to the loop and increases the mobility of the nucleotides in the loop region. This relaxed and open loop structure facilitates anticodon-codon interaction. In the presence of the $N^6$-isopentenyladenosine, the anticodon bases appear to be stabilized by stacking interactions and the addition of the methylthio group at base position 2 may further improve base stacking. [192]
MiaA (dimethylallyl (Δ^2-isopentenyl) diphosphate:tRNA transferase) is the enzyme responsible for the synthesis of i^6A by catalyzing the addition of the dimethylallyl moiety from dimethylallyl phosphate (DMAPP) to the exocyclic N6 amino group of the adenosine at position 37.

*T. thermophilus* and *E. coli* tRNA both contain N^6-threonylcarbamoyladenosine (t^6A), N^6-isopentenyladenosine (i^6A) and 2-methylthio-N^6-isopentenyladenosine (ms^2i^6A). N^6-methyl-N^6-threonylcarbamoyladenosine occurs in *E. coli* and is not found in *T. thermophilus, B. subtilis* or *L. lactis*. *T. thermophilus* contains 2-methylthio-N^6-threonylcarbamoyladenosine (ms^2t^6A), N^6-(cis-hydroxyisopentenyl)adenosine (io^6A) and 2-methylthio-N^6-(cis-hydroxyisopentenyl) adenosine (ms^2io^6A) which are not found in *E. coli*. The hydroxylated derivatives, io^6A and ms^2io^6A, are also found in the gram negative bacterium *Salmonella typhimurium*. [198][199, 200]. The hydroxylase involved in the synthesis of these hydroxylated derivatives of ms^2i^6A and i^6A is encoded by the miaE gene which is absent in *E. coli*. The hydroxylation reaction requires the presence of molecular oxygen and it was thought that ms^2io^6A may be important in adaption to aerobic conditions. [48] However, no difference in growth, as compared to wild type cells, was observed when miaE mutants lacking ms^2io^6A were grown under anaerobic conditions and shifted to aerobic conditions. Interestingly, ms^2io^6A modification appears to have a role in metabolism by regulating the ability of *S. typhimurium* to use the citric acid cycle intermediates, succinate, fumarate and malate, as carbon sources. The ability of this organism to grow on succinate, fumarate or malate is specifically correlated to the presence of the hydroxyl group on ms^2io^6A. [42]

### 2.4.2 Archaeal Organisms

Although Organisms from the kingdom Archaea are found everywhere on earth in moderate and harsh environments, they are often found living in extreme conditions such as high temperature, high acidity, high salinity, high sulfur content, anoxic conditions and in deep sea methane seeps. They share some characteristics with both Eukarya and Bacteria yet maintain characteristics unique to only the Archaeal
kingdom. Archaea are similar to Bacteria in their metabolism, membrane transport systems and genetic organization. The archaeal DNA replication, transcription and translation machinery are similar to Eukarya. The archaeal, bacterial and eukaryal domains can be most accurately distinguished at the molecular level and the pattern of posttranscriptional modification is an important part of that molecular classification scheme. Several modified nucleosides are considered unique to archaea and those identified in this study are discussed below (Figure 2.20).

![Figure 2.20 Modified nucleosides unique to Archaea. G+ = archaeosine, m1Ψ = 1-methylpseudouridine, m1I = 1-methylinosine, m2G = N2,N2-dimethylguanosine.](image)

**Archaeosine (G+).** Archaeosine is unique to archaea and present at position 15 in the D loop in all archaeal tRNAs specific for alanine, aspartic acid, glutamine, glutamic acid and histidine. Archaeosine is the only known modification present at position 15 in the D loop and has not been found in any bacteria or eukarya. Archaeosine is a 7-deazaguanosine derivative and is neither a pyrimidine nor purine and contains an unusual charged imidino side chain at the C7 base position. Other 7-Deazaguanosine derivatives in bacteria and eukarya such as queuosine, epoxyqueuosine, and manosyl-queuosine occur at position 34, the wobble position, and are important in decoding in bacterial and eukaryal systems. Clearly, the role of archaeosine is in stabilizing the L-shape tertiary structure through forming of a reverse Watson/Crick base pair with C48 which joins the D loop and variable domains, the Levitt pair G15-C48. In the unmodified G15-C48 pair, frequently encountered in bacteria and eukarya, magnesium ion binds at
the N7 of guanosine [203] and induces electron density redistribution on the nucleic acid base which stabilizes the reverse [204] Watson Crick base pair. The imidino side chain at C7 of archaeosine prevents magnesium from binding to N7. However, the positively charged imidino group functions similarly to magnesium and stabilizes the G-C base pair. [204]. Additionally, C48 is often modified to m5C and may contribute additional stabilization of tertiary structure.

Archaeosine and queuosine share the same biosynthesis pathway leading to the synthesis of preQo, the precursor to both queuosine and archaeosine. The enzymes involved include GTP cyclohydrolyase I, QueD, QueE and QueC. [205, 206] In archaea, the insertion of preQo into the tRNA is accomplished by arcTGT [207, 208] (140-141). The final step in formation of archaeosine is the addition of ammonia to the nitrile of PreQ0 and is accomplished by arcTGT2. [209]

1-methylpseudouridine (m1Ψ). The modified nucleoside 1-methylpseudouridine (m1Ψ) is unique to archaeal tRNAs and frequently found at position 54 in the TΨC loop domain. The corresponding position in bacteria and eukarya almost always contains the structurally similar 5-methyluridine (m5U54). A few archaea contain other modified uridines known to impart structural stability at position 54 such as Um, Ψ or s2m5U at position 54. Solution and NMR studies indicate the conformation of pseudouridine is not significantly influenced by the substitution of a methyl group at the N1 base position. [210] The role of 1-methylpseudouridine is likely for maintenance of tertiary structure by stabilizing the m1Ψ54-A58 base pair where the D loop and TΨC loops fold together to form the L-shaped tRNA.

Two enzymes are required for biosynthesis of m1Ψ, a pseudouridine synthase and a methyltransferase. The RNA guided ribonucleoprotein pseudouridine synthase, Pus10, is responsible for the synthesis of pseudouridine at position 55 and 54 in archaeal tRNAs. [211] The methyltransferase that catalyzes the addition of a methyl group to the N1 position of pseudouridine at position 54 in tRNA has been described for Methanocaldococcus jannaschii and Haloferax volcanii. [212]
1-methylinosine (m^1I). Archaea often contain 1-methylinosine at position 57, and in some hyperthermophiles the ribose is also methylated to form m^1Im. All tRNAs of the model archaea, *H. volcanii*, contain the sequence m^1Ψ54-Ψ55-Cm56-m^1I57 the TΨC loop domain with the exception of tRNA histidine. In bacteria and eukarya the TΨC loop domain contains the sequence m^5U54-Ψ55-C56 -A or G57. The role of m^1I is in maintaining tertiary structure of the L-shaped tRNA through hydrogen bonding with G19 and stacking interactions between the m^1I57-G19 and A58-G18 base pairs in the elbow region of tRNA. The biosynthesis route for m^1I is different in archaea than in other organisms. In archaea A57 is methylated to m^1A57 then deaminated to form m^1I. In bacteria and eukarya A57 is first deaminated to I then methylated to m^1I. [213] In both archaea and eukarya m^1A58 is also frequently found in the TΨC domain. The archaeal enzyme TrmI catalyzes the formation of m^1A at both position 57 and 58. [214]

*N^2, N^2*-dimethylguanosine  
*N^2, N^2*-dimethylguanosine (m^2^2G) is found at position 10 and 26 in tRNA from Archaea and Eukarya. The two methyl groups on the exocyclic N2 of guanosine alter the base pairing ability of m^2^2G by preventing base pairing with C but allowing pairing with A, U or G. Early solution studies indicate that m^2^2G enhances base stacking. [215] This modification may be important in preventing or controlling folding of tRNA into alternate structures and is important for tertiary structural stabilization, especially in extremeophiles. [216, 217] The tRNA (*N^2,N^2*-guanine 10)-dimethyltransferase (EC 2.1.1.32) responsible for m^2^2G formation at position 10 in tRNA from the hyperthermophilic archaea, *Pyrococcus abyssi*, has been identified. [218] In the hyperthermophilic archaea, *Pyrococcus furiosus*, tRNA (*N^2,N^2*-guanine 26)-dimethyltransferase methylates guanosine at position 26 to form m^2^2G26. [219] One hyperthermophilic bacterial organism, *Aquifex aeolicus*, has been found to contain m^2^2G at positions 26 and 27 in tRNA^Cys^ and the bacterial tRNA (*N^2,N^2*-guanine)-dimethyltransferase encoded by the bacterial *trm1* gene methylates both positions. [220]
Figure 2.21 Structures of wyosine and derivatives. imG = wyosine, mimG = methylwyosine, imG-14 = 4-demethylwyosine, imG2 = isowyosine.

Wyosine derivatives The wyosine family of fluorescent tricyclic nucleosides is present in tRNA from Archaea and Eukarya, but absent in Bacteria. In Eukarya, the wyosine derivatives are characterized by the presence of long complex side chains and in Archaea simple methyl side chains are present at varying positions in the wyosine derivatives (Figure 2.21). Wyosine and derivatives have been identified in many archaeal tRNA, however sequence location was not determined. [94] The eukaryal derivative, wybutosine, is found in tRNA\(^{Phe}\) at position 37 and has similar function to the bacterial hypermodified \(t^6A\) and \(i^6A\) derivatives also found exclusively at position 37 in tRNA. [221] Accordingly, the hydrophobic wyosine derivatives identified in archaeal species are predicted to occur at position 37, adjacent to the anticodon. The tricyclic ring is formed from \(m^1G37\) as the first step in the biosynthetic pathway in both archaea and eukarya. The biosynthetic pathway for the synthesis for wyosine and derivatives in archaerial tRNA has been predicted. [222]

2.5 Conclusions

The overall similarities and differences in tRNA modification patterns and distribution of modifications among organisms as determined here by LC/MS analysis can be used to help elucidate the functional role of modifications in tRNA. Clearly, modifications play a role in maintaining tertiary structure, especially at high temperature. Comparison of modifications can also highlight differences in
decoding strategies among organisms. The dynamic nature of tRNA modification can also be explored by comparison of modification profiles under different cellular conditions.
CHAPTER 3  Identification of an Unknown Modified Nucleoside by LC/UV/MS

3.1 Introduction

In the process of translation, accurate base pairing is required between the codon on the mRNA and the anticodon of the corresponding aminoacylated tRNA so that the correct amino acid is added to the elongating polypeptide. Because there are only 20 standard amino acids and 64 possible codons (4 bases, triplet codon gives $4^3 = 64$ possible codons) many tRNA isoacceptors decode more than one codon. For example, there are 3 codons that specify isoleucine, AUC, AUU, and AUA. Through strict Watson-Crick base pairing, tRNA$^\text{Ile}_1$ reads the AUC codon and the formation of the G-U wobble pair allows tRNA$^\text{Ile}_1$ to read AUU codons. In the archaea, Haloarcula marismortui, the AUA codon occurs infrequently and is read by tRNA$^\text{Ile}_2$ with the sequence CAU in the anticodon [223] and this tRNA$^\text{Ile}_2$ is known to be aminoacylated with isoleucine in vivo. [224] This is surprising, as CAU anticodon would be expected to read the AUG, methionine codon, based on usual Watson-Crick base pairing. However, as is the case for many tRNAs that read rare codons, post transcriptional modifications in the wobble position of the anticodon often are found to modulate codon/anticodon binding, thus expanding base pairing interactions beyond the standard canonical bases. Kohrer demonstrated the cytidine at the wobble position 34 in tRNA$^\text{Ile}_2$ H. marismortui (Figure 3.1) was indeed modified, but the precise identity of this modification was unknown, before the work described in this chapter. [224]

![Figure 3.1](image)

Figure 3.1 Sequence of tRNA$^\text{Ile}_2$ from H. marismortui. C* denotes unknown modified cytidine at position 34.
Our lab collaborated with Professor RajBhandary (MIT) [223] to identify this modified cytidine. We used LC/MS/MS RNase mapping and sequencing to confirm the C34 wobble position was modified with an unknown modification of 112 u. This chapter describes how the structure of this modified nucleoside was elucidated using LC/UV/MS/MS.

3.2 Experimental

3.2.1 Desalting of tRNA

A Nucleobond (Machery Nagel) anion exchange column was used to purify and desalt the tRNA prior to enzymatic digestion. The tRNA was loaded onto the column in a solution of 200 mM KCl, 100 mM tris acetate (pH 6.3) and 15% ethanol then washed with the same solution. A solution of 400 mM KCl, 100 mM tris acetate (pH 6.3) and 15% ethanol was used for a second wash then the tRNA was eluted with 750 mM KCl, 100 mM tris acetate (pH 6.3) 15% ethanol. Isopropanol was used to precipitate the tRNA and the pelleted RNA was rinsed with 70% ethanol. The purified tRNA pellet was resuspended in autoclaved water.

3.2.2 Digestion of tRNA to nucleosides

Prior to enzymatic digestion of the tRNA to nucleosides, the nucleic acid was denatured at 100 ºC for 3 min then chilled in an ice water bath. To lower the pH, 1/10 volume 0.1 M ammonium acetate (pH 5.3) was added. For each 0.5 AU of tRNA, 2 units Nuclease P1 was added and incubated at 45 ºC for 2 h. The pH was readjusted by adding 1/10 volume of 1.0 M ammonium bicarbonate then 0.002 units of snake venom phosphodiesterase was added and incubated at 37 ºC for 2 h. Finally, 0.5 units of antartic phosphatase was added and incubated at 37 ºC for 60 min. The nucleoside digests were stored at -80 ºC. [64]
3.2.3 HPLC/MS conditions

The nucleoside digests were analyzed using a Hitachi D-7000 HPLC system with a Hitachi D-7400 UV detector set at 260 nm. A LC-18-S (Supelco) 2.1 x 250 mm, 5 µm particle column was used at a flow rate of 300 µL/min. Mobile phases used were as follows, A: 5 mM ammonium acetate, pH 5.3 and B: 40% acetonitrile in water. The gradient used was as described in Chapter 2 and is outlined in Table 2.2. The column eluent was split immediately post column, 1/3 to the electrospray ion source and 2/3 to the UV detector.

3.2.4 MS Conditions

A Thermo LTQ-XL ion trap mass spectrometer was used for the low resolution LC/MS and LC/MS/MS analyses. A Thermo LTQ-FT mass spectrometer was used for the high resolution, accurate mass LC/MS and LC/MS/MS experiments. Both instruments were equipped with an ion max electrospray ionization source. Mass spectra were recorded in the positive ion mode with a capillary temperature of 275 °C, spray voltage, 3.7 to 4.0 kV and sheath gas, auxiliary gas and sweep gas of 45, 25 and 10 arbitrary units, respectively. Data dependent MS/MS of each of the two most abundant ions were recorded throughout the LC/MS run. For the MS/MS experiments, the normalized collision energy was 35%, isolation width was 1.5 u, the activation Q was 0.38 and activation time was 30 ms.

3.3 Results

3.3.1 Structural Characterization of an Unknown Modified Nucleoside

To determine the structure of the unknown modification in *H. marismortui* tRNA^Ile^2 at C34, purified tRNA^Ile^2 was digested completely to nucleosides and analyzed by LC/UV/MS. The modifications found are shown in the UV chromatogram in Figure 3.2. The identity of each nucleoside is based on relative retention times and mass spectra. Modifications found to be present were pseudouridine, 1-methylpseudouridine, 2'-O-methylecytidine, 1-methylinosine, N^2^-methylguanosine, N^2^, N^2^-
methylguanosine, $N^6$-threonylcarbamoyladenosine, archaeosine and an unknown peak with mass 355 u. The same modified nucleosides, except the unknown, were found previously in *Halobacterium volcanii* tRNA$^{\text{Ille}}$. [225] Because *H. volcanii* is an archaeal organism that thrives under similar conditions as *H. marismortui* and the tRNA sequences are very similar, it is not surprising that the same modifications are detected in both tRNAs. The presence of the 2'-deoxyribonucleosides, deoxyguanosine and deoxyadenosine, can be attributed to small amounts of DNA present, most likely from the DNA oligonucleotide used to purify the tRNA$^{\text{Ille}}$.

Figure 3.2  LC/MS analysis of nucleosides present in *H. marismortui* tRNA$^{\text{Ille}}$. (1) pseudouridine; (2) cytidine; (3) uridine; (4) 1-methylpseudouridine; (5) 5-methylcytidine; (6) 2'-O-methylcytidine; (7) guanosine; (8) deoxyguanosine; (9) 1-methylinosine; (10) $N^2$-methylguanosine; (11) adenosine; (12) deoxyadenosine; (13) $N^2$, $N^2$-dimethylguanosine; (14) $N^6$-threonylcarbamoyladenosine; (15) C* (agmatidine); (16) archaeosine.

Additional sequencing and RNase mapping experiments performed by other researchers in our lab revealed the presence of an unknown modified cytidine at position C34 in the *H. marismortui* tRNA$^{\text{Ille}}$ with a modification mass of 112 u. The mass spectra of the peak eluting at 31.6 min in the nucleoside digest yielded a molecular ion of $m/z$ 356 with a corresponding base ion of $m/z$ 224. The mass difference of 132 mass units between the molecular and base ions indicates the ribose of this nucleoside is unmodified and the modification is on the base. These data are consistent with the sequencing data that places the modification on a cytidine base, with a modification mass of 112 u. The selected ion chromatograms for the molecular and base ions of the unknown nucleoside are shown in Figure 3.3 a, b. Because the unknown nucleoside coelutes with the leading edge of the archaeosine peak, the mass
The spectrum of the unknown nucleoside also includes a strong signal at m/z 325, which arises from the molecular ion of archaeosine. (Figure 3.3c)

Figure 3.3  Mass spectral analysis of unknown modified cytidine (C*). Selected ion chromatograms of (A) the molecular ion at m/z 356, [MH]^+ and (B) the base ion at m/z 224, [BH_2]^+. (C) Mass spectrum of the modified cytidine, C*.

High resolution Fourier transform ICR mass spectrometry was used to obtain accurate mass measurements so that an elemental composition could be proposed. Assuming the unknown nucleoside was a modified cytidine, the most probable elemental composition for the molecular ion would contain at least two nitrogens, a minimum of four oxygens with a degree of unsaturation of at least four. The
measured mass of the protonated molecular ion was found to be $m/z$ 356.2038 (0.6 ppm error) and yielded a most likely elemental composition of $C_{14}H_{26}N_7O_4$. For the protonated base ion, the accurate mass was $m/z$ 224.1617 (0.4 ppm error) and the most likely elemental composition was $C_9H_{18}N_7$.

The elemental formula for the base ion is consistent with the loss of a ribose ($-C_5H_{10}O_4$). As a ribose contains four oxygens and the most likely elemental formula for the base ion contains no oxygens and the proposed molecular formula contains only four oxygens, this indicates the C2 oxo group is absent in this modified cytidine. The C2 oxo group is absent in some other modified cytidines such as 2-thiocytidine, which contains sulfur at the C2 position, and lysidine with lysine conjugated to the C2 position of the base. In fact, sulfur or selenium is often present at the C2 position of uridines instead of the usual oxo group. These C2 sulfur-containing modified uridines are often found at the wobble position in the anticodon of tRNA and directly impact codon choice. However, no elemental formula could be proposed that contained sulfur that was consistent with the accurate mass value and the elemental formula.
constraints for the modified cytidine. To verify the absence of sulfur, a diode array detector was placed in series with the UV detector to record the UV spectra between 200 and 400 nm. The UV spectra collected during the retention time of the unknown nucleoside did not show an increased absorbance profile in the range of 275 to 285 nm, which is characteristic of C2 sulfur containing modified cytidines and uridines. This UV data alone, does not rule out the possibility of the presence of sulfur in the unknown nucleoside, but this result is consistent with other data indicating the absence of sulfur. [226]

Based on the data obtained, it appeared that an unknown modified cytidine was present. This modified cytidine did not possess an exocyclic C2 oxygen or sulfur. To identify a putative structure for this modified nucleoside, accurate mass measurements of the collision induced dissociation products of the protonated base ion were obtained. The glycosidic bond breaks easily in a nucleoside within the electrospray source region, and the relative abundance of molecular and base ion can be controlled by varying the conditions within the electrospray source region. The bonds in the aromatic base do not break easily and fragmentation of the aromatic base requires an input of additional energy. In the ion trap, collision induced dissociation is produced by resonant excitation of the particular isolated ion, in this case, the base ion of m/z 224. Excitation is accomplished by applying a supplementary rf voltage with a frequency that matches the axial secular frequency of the isolated base ion. The base ion becomes resonantly excited and gains kinetic energy; this gain in kinetic energy can cause the base ion to be ejected from the trap or the kinetic energy can be converted, via collisions with He buffer gas, to internal energy and base ion dissociation occurs resulting in fragment ions. These two processes, ion ejection and ion dissociation, are competitive processes and the extent to which each occurs depends greatly upon the rf excitation frequency, amplitude, duration, type and pressure of buffer gas. [227-231] The goal is to dissociate most of the isolated base ions and maximize the trapping efficiency of the resulting fragments, which requires careful optimization of ion trap conditions.

The amount of tRNA^{ile}_{2} available was limited and the unknown nucleoside was only present as a minor component in the nucleoside digest of this tRNA. Thus, optimization of MS/MS experimental conditions
was done using 5-methylcytidine, which was readily available. At low rf amplitude (or collision energy, CE) no fragmentation was observed (Figure 3.5a). As collision energy is increased, more energy deposition occurs which leads to greater fragmentation (Figure 3.5b). However, due to the competition between ion ejection and ion dissociation, trapping efficiency is lower as collision energy is increased. While the number of base ions isolated was approximately $10^4$, after increasing the collision energy only 10 fragment ions were detected. An increase in the rf frequency (activation Q) results in a greater number of fragment ions remaining in the trap and then being detected (Figure 3.5c). The optimal activation time was found to be 30 ms for these experiments.

![Figure 3.5](image)

**Figure 3.5** Collision induced dissociation of 5-methylcytidine. (A) No fragmentation of base ion, $m/z$ 126. CE = 23%, Q = 0.25. (B) Fragmentation of base ion, $m/z$ 126 but ion yield low, CE = 35%, Q = 0.25. (C) Fragmentation of base ion, $m/z$ 126 with high ion yield, CE = 35%, Q = 0.38.

The high resolution accurate mass analysis of the CID MS/MS fragments of the unknown base ion, $m/z$ 224, yielded two intense product ions which were consistent with the loss of NH$_3$ ($m/z$ 207.1352 C$_9$H$_{15}$N$_4$) and the loss of CH$_5$N$_3$ ($m/z$ 165.1135 C$_8$H$_{13}$N$_4$). Considering the known fragmentation patterns of substituted cytidines, [73] the nitrogen rich nature of the modification, the absence of oxygen, (normally a C2-oxo group on the cytidine) and the similarity to the modified nucleoside lysidine, the modification is suggested to be agmatine attached to the C2 of cytidine. Agmatine is decarboxylated arginine. For comparison, the collision induced dissociation products of a solution of commercially
available agmatine were determined. The major product ions for both the unknown modification base ion and the agmatine solution were the loss of ammonia ($\text{NH}_3$) and a guanidino group ($\text{CH}_5\text{N}_3$). Minor fragments were observed in low resolution CID of the unknown base ion, $m/z$ 224, at $m/z$ 182, $m/z$ 114, $m/z$ 153 and are likely due to side chain cleavages (Figure 3.6). Similarly, the minor CID product of agmatine at $m/z$ 60 is also consistent with side chain cleavage (Figure 3.7).

Figure 3.6  Identification of unknown modified cytidine (C*). (A) Accurate mass analysis of CID MS/MS product ions of unknown base ion, $m/z$ 224. (B) Low resolution CID MS/MS of unknown base ion, $m/z$ 224. Minor ions of $m/z$ 114, $m/z$ 153, $m/z$ 182 likely due to side chain cleavages of agmatine moiety.
Figure 3.7  Low resolution CID MS/MS of agmatine standard. Both unknown and agmatine exhibit loss of NH$_3$ and CH$_3$N$_3$.

The data presented here suggests the unknown modified nucleoside eluting at 31.6 min has the structure shown below (Figure 3.8).

Figure 3.8  Structure of modified nucleoside at C34 in tRNA$^{Ile}_2$. 
3.3.2 Presence of Agmatidine in other Archaea

The same unknown nucleoside of $355 \ u$ was found in tRNA from two other archaeal species, *Sulfolobus solfataricus* and *Methanococcus maripaludis*. The selected ion chromatograms and mass spectra are shown in Figure 3.9 a-c for total tRNA isolated from *S. solfataricus* and clearly show the presence of the protonated molecular ion, $m/z$ 356, and the protonated base ion, $m/z$ 224, of agmatidine. Agmatidine was also detected in partially purified tRNA$^\text{Ile}$ from *M. maripaludis* as shown in Figure 3.9 d-f. The conditions used for the analysis of *M. maripaludis* favored the detection of the protonated base ion. The charged nature of the agmatidine and the presence of several different imino-amino tautomers (Figure 3.8) likely contribute to the characteristic wide chromatographic (SIC/UV) peak shape.

![Figure 3.9](image)

**Figure 3.9** Selected ion chromatograms and mass spectra; *Sulfolobus solfataricus*, A-C and *Methanococcus maripaludis*, D-E.
3.4 Discussion

Previous studies in the RajBhandary lab showed a tRNA with anticodon CAU, annotated as a methionine tRNA in *H. marismortui*, was not aminoacylated with methionine. This tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) was instead aminoacylated with isoleucine. [224] The rare AUA isoleucine codon is decoded by the tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) only when the cytidine at position 34 (C34), the wobble position, is modified. The unmodified tRNA (CAU) decodes AUG, the methionine codon. It was determined that the change in aminoacylation and decoding specificity was due to modification of the wobble position for this archaeal organism. This previously unknown modified cytidine was structurally characterized by LC/MS here and identified as agmatidine, a modified cytidine with a decarboxylated arginine attached at the C2 position of the cytidine base.

To accurately characterize this modification by LC/MS analysis it was necessary to purify tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) from *H. marismortui* total tRNA, confirm that indeed this particular tRNA was aminoacylated with isoleucine and verify the codon reading properties of tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU). Isolation of a low abundant tRNA is challenging, thus for the purification of the specific tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) a biotinylated DNA probe immobilized on a streptavidin-coated resin was used which hybridizes to the specific tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU). While this is a very selective technique, sometimes the hybridization efficiency can be less than 100% and some nonspecific binding of other tRNAs can occur so additional PAGE (polyacrylamide gel electrophoresis) purification was also used. Ribosome binding experiments confirmed that isolated tRNA\textsuperscript{\text{Ile}}\textsubscript{1} (GAU) binds to AUC but neither AUA or AUG. Isolated tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) binds to AUA and not AUC or AUG. The agmatidine modification characterized here clearly originated from the rare tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) at the wobble position (C34) and is required for base pairing with adenosine in reading the AUA codon. The agmatidine modification at C34 does not base pair with G, C or U.

The structure of the unknown modified nucleoside characterized here, agmatidine, is very similar to the modified nucleoside, lysidine, which occurs in tRNA\textsuperscript{\text{Ile}}\textsubscript{2} (CAU) of bacterial species. [179, 232] In
both lysidine and agmatidine the C2 of the cytidine base is the site of attachment of the primary amine. The difference between lysidine and agmatidine lies in the nature of the terminal group, a guanidine group for agmatidine and a carboxyl and amino group for lysidine. An agmatine linkage at C5 or C6 on the cytosine, although structurally possible, is not considered likely here because base pairing of adenosine with that structure would not be possible. (Figure 3.10)

![Agmatidine and Adenosine](image)

**Figure 3.10**  Predicted base pairing between adenosine and agmatidine and adenosine and lysidine. (Adapted from Ikeuchi 2010)

Both agmatidine and lysidine are present at C34, the wobble position in tRNA\(^{\text{Ile}_2}\) (CAU), and are responsible for the ability of C34 to base pair with adenosine on the mRNA codon instead of the usual guanosine. They ensure correct codon recognition and differentiation between the AUA codon and the AUG codon. The exact nature of the base pairing between agmatidine and adenosine requires additional structural studies, but the ability of agmatidine to interchange from amino to imino forms results in the
formation of several tautomeric structures (Figure 3.8) each with different H bonding properties and is likely critical to base pair formation.

Agmatine (L-amino-4-guanidinobutane), a polyamine, is an endogenous amine synthesized from the decarboxylation of arginine and is an intermediate in the synthesis of other polyamines such as putrescine and spermidine. These polyamines are common to all organisms and are thought to function as modulators of cellular functions including adaptation to cellular stress. [233-235] In thermophilic organisms such as Thermus thermophilus, polyamines contribute to growth at high temperatures. [236] Long branched polyamines are common in thermophilic organisms from both archaea and bacteria and play a role in stabilizing DNA and RNA structure. [237, 238]

The enzymes responsible for the synthesis of lysidine and agmatidine, TilS (tRNA\(^{\text{ile}}\) lysidine synthetase) and TiaS (tRNA\(^{\text{ile}}\) agmatidine synthetase), respectively, have been identified. TiaS requires the presence of agmatine and ATP (Ikeuchi 2010) while TilS requires lysine and ATP as substrates. [179] Although similar in substrate requirements, these two enzymes do not share sequence homology, suggesting convergent evolution.

In this work agmatidine was found in tRNA of H. marismortui, M. maripaludis and S. sofataricus. Agmatidine has also been detected in Sulfolobus tokodaii, Methanosarcina acetivorans and Haloferax volcanii. [239] It is likely agmatidine is responsible for correct decoding of the AUA codon throughout the Crenarchaeal and Euryarchaeal phyla of the archaeal domain. The mechanism for specifically decoding the AUA codon in archaea is now known and it is clear there are three different tRNA modification strategies used in the three domains of life. In bacteria, C34 is modified to lysidine to specifically decode AUA in tRNA\(^{\text{ile}}\)\(_2\) (CAU). The unmodified C34 in tRNA\(^{\text{ile}}\) (CAU) decodes AUG, the methionine codon. In eukaryotes, a pseudouridine at the wobble position specifically decodes the AUA codon. A less restrictive strategy can also be used in eukaryotes when inosine is the modification at position 34 allowing decoding of AUU, AUC and AUA codons. Inosine at position 34 recognizes U, C and A; although A is recognized poorly. [240] Agmatidine and lysidine are examples of modifications...
that serve to restrict decoding to specific codons while inosine expands the decoding ability of tRNAs. Interestingly these three different tRNA modification strategies are used in the three domains of life for specifically decoding the AUA codon.

3.5 Conclusion

In this chapter LC/UV/MS methods were used to determine the structure of a previously unknown nucleoside important in specific decoding properties of tRNA. The relative abundance of this rare tRNA made isolation challenging and resulted in only small amounts of tRNA available for characterization of the unknown nucleoside. However, the sensitivity and specificity of LC/UV/MS techniques including high resolution accurate mass of both nucleoside molecular ions and collision induced fragment ions enabled complete structural characterization.
CHAPTER 4 Functional Characterization of Methylthiotransferases of *Bacillus subtilis*

4.1 Introduction

The complete genome of *Bacillus subtilis* was published in 1997, and *B. subtilis* was the first gram positive bacterium to be sequenced. [241] *B. subtilis* is an aerobic, endospore-forming bacterium that lives primarily in soil and water. *B. subtilis* has been extensively used as a model organism in Biochemistry and Molecular Biology because it undergoes natural genetic DNA-mediated transformation easily making it valuable for genetic manipulation and study. Because it can secrete heterologous proteins directly into culture medium, it is used in industry to produce enzymes such as amylases and proteases. *B. subtilis* is also one of the few organisms where most of the tRNA sequences including identity and location of modified nucleotides are known. [29]

Both *B. subtilis* and *Escherichia coli* contain many of the same tRNA posttranscriptional modifications, but notable differences do occur as outlined in Chapter 2. For example, the hypermodified adenosine m$^6$t$^6$A (N$^6$-methyl-N$^6$-threonylcarbamoyladenosine) is found at position 37 in two threonine tRNAs with GGU anticodons in *E. coli* but this modification is not present in any *B. subtilis* tRNA. In *B. subtilis* tRNA threonine, the non-methylated t$^6$A (N$^6$-threonylcarbamoyladenosine) is present. *B. subtilis* contains the thiolated derivative ms$^2$t$^6$A (2-methylthio-N$^6$-threonylcarbamoyladenosine) in tRNA lysine with anticodon UUU, but ms$^2$t$^6$A is not found in *E. coli*.

Adenosine at Position 37 directly adjacent to the 3’ side of the anticodon in tRNA is often modified with bulky hydrophobic groups, especially in tRNAs that decode codons that begin with A or U and, in fact, t$^6$A is conserved throughout the three domains of life. These modifications maintain the structure of the anticodon stem loop (ASL) that otherwise is destabilized by the intrinsically weak A-U base pairing in the anticodon. Modifications at position 37 in tRNA stabilize the ASL by improving intrastrand base stacking within the tRNA ASL. These modifications also stabilize stacking and interaction between bases of the mRNA codon and tRNA anticodon. [186, 189, 195, 221, 242] The bulky hydrophobic groups of these adenosines prevent base pair formation between the invariant U33 and A37 thereby maintaining an
open stable loop structure. Specifically, the positioning of modified A37 over the first anticodon/codon base pair to form a cross-strand stack strengthens codon binding. [160]

The presence of these hypermodified adenosines has been shown to improve translational efficiency and accuracy. [45, 243-245] Modified adenosines are also important in binding tRNA to the ribosome during translation. [187, 189, 246] The methyl group of m^6^A has been shown to increase the efficiency in reading cognate codons of tRNA threonine in E. coli. [247] Several RNA modifications, including i^6^A (N^6^-isopentenyladenosine) and ms^2^i^6^A (2-methylthio-N^6^-isopentenyladenosine) have been shown to improve reading frame maintenance during translation. [28] The individual tRNAs in B. subtilis containing hypermodified adenosines at position 37 are listed below (Table 4.1). [29] Structures of these hypermodified adenosines are also shown (Figure 4.2).

<table>
<thead>
<tr>
<th>t^6^A</th>
<th>i^6^A</th>
<th>ms^2^t^6^A</th>
<th>ms^2^i^6^A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg 2 (UCU)</td>
<td>Cys1 (GCA)</td>
<td>Lys (cmnm^5^s^2^UUU)</td>
<td>Phe (GmAA)</td>
</tr>
<tr>
<td>Arg 3 (CCU)</td>
<td>Leu 3 (CCA)</td>
<td>Leu4 (UAA)</td>
<td></td>
</tr>
<tr>
<td>Asn 1 (GUU)</td>
<td>Ser 2 (mo^5^UGA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile 1 (GAU)</td>
<td>Trp 1 (CCA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile 2 (CAU)</td>
<td>Ser 1 (GCU)</td>
<td>Tyr 1 (QUA)</td>
<td></td>
</tr>
<tr>
<td>Thr 1 (mo^5^UGU)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr 2 (GGU)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1 Distribution of hypermodified adenosines B. subtilis tRNAs. (Q, queuosine. mo^5^U, 5-methoxyuridine. Gm, 2'-O-methylguanosine. cmnm^5^s^2^U, 5-carboxymethylaminomethyl-2-thiouridine.)
Figure 4.1  Structures of hypermodified adenosines in *B. subtilis* tRNA.
(t<sup>6</sup>A, N<sup>6</sup>-threonylcarbamoyladenosine. ms<sup>2</sup>t<sup>6</sup>A, 2-methylthio-N<sup>6</sup>-threonylcarbamoyladenosine. i<sup>6</sup>A, N<sup>6</sup>-isopentenyladenosine. ms<sup>2</sup>i<sup>6</sup>A, 2-methylthio-N<sup>6</sup>-isopentenyladenosine.)

The biosynthetic pathway for the formation of methylthiolated adenosine derivatives, ms<sup>2</sup>t<sup>6</sup>A and ms<sup>2</sup>i<sup>6</sup>A, requires the initial formation of t<sup>6</sup>A and i<sup>6</sup>A, respectively. (Figure 4.2) The t<sup>6</sup>A biosynthetic pathway has yet to be completely characterized and likely involves several modifying enzymes and other proteins. [248] The formation of t<sup>6</sup>A requires the presence of threonine and carbonate and is an ATP-dependent process. The gene product of *yrdC* is involved in the biosynthesis of t<sup>6</sup>A, and *yrdC* appears to be essential for cellular viability in *E. coli*. [193] However, whether this phenotype is due to the presence or absence of t<sup>6</sup>A is unknown at this time. Recently, another universally conserved protein, YgiD, has been shown to be involved in t<sup>6</sup>A biosynthesis. [249] In *E. coli*, YgiD was also appears essential and this function is dependent on the presence of another protein, yeaZ. [249]

![Diagram showing the biosynthetic pathway of hypermodified adenosines](image)

**Figure 4.2** Modifying enzymes involved in the biosynthesis of hypermodified adenosines of *B. subtilis*. ( ) denotes *E. coli* enzymes.
MiaA (tRNA isopentenyl transferase) catalyzes the addition of an isopentenyl group from dimethylallyl pyrophosphate to the exocyclic amine group of A37 in tRNA. Although MiaA does not appear to be essential, miaA mutants demonstrate errors in codon context sensitivity, elongation rate, translation efficiency and fidelity causing slow growth and temperature sensitivity. [243, 250-252] Formation of i^6A is required before the methylthiolation can occur to form ms^2i^6A. [200, 253]

Methylthiotransferases are the enzymes responsible for the addition of the methylthio group (SCH₃) to t^6A and i^6A in the formation of ms^2t^6A and ms^2i^6A, respectively. Based on phylogenetic analyses and sequence similarities, methylthiotransferases can be subdivided into four subfamilies of proteins:

1) The first includes MiaB, which forms ms^2i^6A in E. coli and Thermotoga maritima [254], and other MiaB homologues in bacteria and eukarya.

2) The second includes RimO, which modifies ribosomal protein S12 in E. coli.

3) The third includes yqeV from B.subtilis, predicted to methylthiolate t^6A.

4) The fourth includes proteins such as MJ0867 from Methanococcus jannaschii thought to methylthiolate t^6A in archaea and eukaryotes.

Before this work, proteins from families 3 and 4 above had yet to be characterized. [255]

The methylthiotransferases of all four subfamilies contain the same types of structural domains, including an N-terminal domain, the central radical S-adenosylmethionine (SAM) domain and the C-terminal TRAM (tRNA-methylase) domain. Methylthiotransferases are members of the radical SAM protein superfamily. Initially, over 600 members of the radical SAM superfamily of proteins were identified by bioinformatic techniques. [256] Currently, over 2800 proteins are considered to belong to the radical SAM protein superfamily. [257] The central radical SAM domain harbors the catalytic core where methylthiolation is accomplished by a free radical mechanism that requires S-adenosylmethionine and an Fe-S cluster chelated by conserved cysteines. The Fe-S cluster initiates a radical reaction by mediating the reductive cleavage of AdoMet to yield a 5'-deoxyadenosyl radical.
This radical abstracts a hydrogen from the C2 of the adenosine substrate then the activated C2 is thiolated and methylated. Two molecules of AdoMet are required, one used for the radical formation and substrate activation and the other is used as the methyl donor. S-adenosylmethionine is required for both sulfur insertion and methylation. MiaB is the first Fe-S containing enzyme shown to modify tRNA. [253]

The N-terminal domain of MiaB contains a second Fe-S cluster chelated by conserved cysteines that is required for activity and is postulated to be the sulfur donor. [258] The C-terminal domain of MiaB contains the TRAM domain, predicted to bind tRNA and to assume a simple β-barrel fold tertiary structure typical of RNA binding protein domains. The TRAM family describes a conserved protein family originally identified in two distinct classes of tRNA-modifying enzymes: 1) the TRM2 family of uridine methylases and 2) the methylthiolases of the MiaB family. [259] Amino acid sequence analysis of MiaB from T. maritima [254] and E. coli [260] have revealed a β-barrel RNA binding structure in the C-terminal domain.

Based on previous bioinformatics analysis of the B. subtilis genome, two putative methylthiotransferases were identified. [255] One is the gene product of ymcB, predicted to be an E. coli miaB ortholog responsible for formation of ms²i⁶A. The other predicted methylthiotransferase was the gene product of yqeV. This predicted enzyme is part of the novel MTTase subfamily with no previously characterized members and is postulated to form ms²t⁶A in B. subtilis. This chapter describes a collaborative effort with Brian Anton (Boston University) to identify the ymcB gene product as the methylthiotransferase responsible for the formation of ms²i⁶A and the yqeV gene product as the methylthiotransferase responsible for ms²t⁶A formation. [261] Protein domain switching experiments were also conducted by our collaborators in this research project to investigate which protein domains of the methylthiotransferases may be involved in the discrimination between i⁶A and t⁶A.

### 4.2 Experimental

All culturing, cloning and plasmid construction experiments were performed by our collaborators at New England Biolabs and Boston University and summarized in reference [261]. Isolation of tRNA,
tRNA purification, enzymatic digestion to nucleosides and LC/MS conditions were as described in Chapter 2.

4.3 Results

The tRNA from wild type and mutant strains of *B. subtilis* was isolated, purified and enzymatically digested to nucleosides. These nucleoside digests were analyzed by LC/UV/MS to determine the presence or absence of modified nucleosides, specifically t6A, i6A, ms2t6A and ms2i6A. The UV chromatogram of modified nucleosides present in the total unfractionated tRNA from wild type *B. subtilis* 168 is shown in Figure 4.3, with the hypermodified adenosines, t6A, ms2t6A, i6A and ms2i6A, marked with asterisks. This LC/UV/MS- based total census of modifications present in unfractionated total tRNA is the most comprehensive reported to date for *B. subtilis* strain 168. Modifications identified here are consistent with those reported in the tRNA sequence databases, [29, 30] which are largely based on sequencing studies of *B. subtilis* done by Vold. [98] The only modified nucleoside known to be present in *B. subtilis* total tRNA and not detected here is 4-thiouridine. As discussed in Chapter 2, it is likely the 4-thiouridine coelutes with guanosine and is difficult to detect due to low relative ion abundance or the uridine is incompletely thiolated under these growth conditions. In addition, lysidine (k2C) was identified which was previously only predicted in a more specialized analysis of tDNA. [262] Also detected were three nucleosides (m2G, m2A and m26A) that are typical of rRNA and therefore may be indicative of trace rRNA contamination in the tRNA preparation. One other nucleoside, ms2A, is present in trace amounts in all samples and is assumed to be a breakdown product of ms2t6A or ms2i6A created during tRNA purification.
Figure 4.3. UV chromatogram of wild type \textit{B. subtilis} 168 total tRNA digested to nucleosides. Numbered peaks are as follows: (1) D, (2) \( \Psi \), (3) \( cmnm'U \), (4) C, (5) U, (6) m\(^1\)A, (7) \( cmnm'sU \), (8) m\(^3\)U, (9) mo\(^5\)U, (10) (I), (11) G, (12) m\(^7\)G, (13) Gm, (14) Q, (15) m\(^5\)G, (16) k\(^2\)C, (17) m\(^2\)G, (18) A, (19) t\(^6\)A, (20) m\(^2\)A, (21) m\(^6\)A, (22) ms\(^1\)t\(^6\)A, (23) ms\(^2\)A, (24) m\(^2\)z\(^2\)A, (25) i\(^6\)A, (26) ms\(^1\)i\(^6\)A. The 4 peaks relevant to this work are marked with asterisks “*”.

The corresponding mass spectra shown in Figure 4.4 along with relative retention times confirm the identity of these modified adenosines. The glycosidic bond in nucleosides cleaves easily under electrospray conditions so nucleoside identification is based on the mass of the molecular ion, the mass of the base ion and the presence of a UV peak at the same retention time.

![Figure 4.3](image1)  
![Figure 4.4](image2)

\textbf{Figure 4.4} (a) Extracted ion chromatograms of molecular and base ions and (b) mass spectra of the adenosine derivatives t\(^6\)A, ms\(^1\)t\(^6\)A, i\(^6\)A, and ms\(^2\)i\(^6\)A from digested wild type \textit{B. subtilis} 168 total tRNA.
Cells were cultured from the two mutant *B. subtilis* strains, BSF2608 and YQEVd, obtained from the *B. subtilis* sequencing consortium (http://bacillus.genome.ad.jp). In these mutant strains the gene inactivation vector pMUTIN [263] was used to inactivate the genes ymcB and yqeV in strains BSF2608 and YQEVd, respectively. The UV chromatograms of the nucleoside digests from the total tRNA isolated from each of these strains are shown in Figure 4.5 and compared to the nucleoside digest from total tRNA of the wild type *B. subtilis*. The absence of ms$^{2}$t$^{6}$A was confirmed in strain BSF2608 (Figure 4.5b) and the absence of ms$^{2}$i$^{6}$A was confirmed in YQEVd (Figure 4.5c).

![UV Chromatograms of *B. subtilis* total tRNA nucleoside digests, highlighting hypermodified adenosines t$^{6}$A, ms$^{2}$t$^{6}$A, i$^{6}$A and ms$^{2}$i$^{6}$A. A) Wild type strain 168, B) strain BSF2608 and C) strain YQEVd. Absence of peak is denoted in parentheses. Non-nucleoside peak denoted with asterisk.](image-url)
In a BLASTP sequence homology search for the most similar sequences to the methylthiotransferase MiaB in *B. subtilis*, two putative methylthiotransferases were revealed, the products of *ymcB* and *yqeV* genes. The *ymcB* gene product was predicted to methylthiolate i\(^6\)A, and the *yqeV* gene product was predicted to methylthiolate t\(^6\)A. These putative methylthiotransferase genes were subcloned into plasmid pDM124c7 and reintroduced into the respective mutant *B. subtilis* strains, BSF2608 and YQEvd. Cells were cultured from these strains and the tRNA isolated, purified and enzymatically digested to nucleosides. In both cases the genes reintroduced, via plasmid, into the mutant strains resulted in the presence of ms\(^2\)i\(^6\)A in BSF2608, B(B), and ms\(^2\)t\(^6\)A in YQEvd, V(V). (Table 2) This indicates that *ymcB* is required for formation of ms\(^2\)i\(^6\)A and *yqeV* is required for ms\(^2\)t\(^6\)A modification. As a control the empty pDM124c7 plasmid (no gene added) was also incorporated into each of the mutant strains, B(124) and V(124) in Table 4.

The putative methylthiotransferase genes *mj0867* and *mmar*, predicted to be necessary for ms\(^2\)t\(^6\)A modification in the archaeal hyperthermophile, *M. jannaschii*, and in the archaeal mesophile, *M. maripaludis*, respectively, were subcloned into plasmid pDM124c7 and introduced into the YQEvd mutant strain. The *M. maripaludis* mmar gene was also introduced into the BF2608 mutant strain. These *B. subtilis* strains were cultured and the total tRNA isolated and purified. The LC/MS analysis of the nucleosides present in total tRNA from the *M. jannaschii*, V(MJ), and *M. maripaludis*, V(MM), in YQEvd strains indicates that neither gene was able to rescue the modification deficient phenotype, as indicated by the absence of ms\(^2\)i\(^6\)A in the UV chromatograms and mass spectra. The subcloned *M. maripaludis* mmar gene in the BF2608 mutant *B. subtilis* strain B(MM) also failed to complement the loss of ms\(^2\)i\(^6\)A. Cloning of these methylthiotransferases was repeated and the protein sequences were confirmed by LC/MS/MS sequencing experiments. These repeat experiments also failed to produce the desired modification rescue. These archaeal proteins appeared insoluble in the *B. subtilis* cell lysate likely indicating that these proteins are aggregating due to improper protein folding and are thus inactive.
Table 4.2. Modified nucleosides observed in *B. subtilis* strains. For the genotypes, (+) indicates intact genes, (-) indicates insertion mutants, (+P) is plasmid encoded gene copy and (?P) indicates plasmid encoded heterologous gene copy that may or may not complement gene in question. Phenotypes indicate the presence (+) or absence (-) of modified nucleosides based on LC/UV/MS analysis.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Phenotype</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ymcB</td>
<td>qeV</td>
</tr>
<tr>
<td>168</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>BSF2608</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>B(124)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>B(B)</td>
<td>+P</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>B(MM)</td>
<td>?P</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>B(B1V)</td>
<td>?P</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>B(B2V)</td>
<td>?P</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>B(V1B)</td>
<td>?P</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>B(V2B)</td>
<td>?P</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>YQEVd</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>V(124)</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>V(V)</td>
<td>+P</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>V(MJ)</td>
<td>+</td>
<td>?P</td>
<td>+</td>
</tr>
<tr>
<td>V(MM)</td>
<td>+</td>
<td>?P</td>
<td>+</td>
</tr>
<tr>
<td>V(B1V)</td>
<td>+</td>
<td>?P</td>
<td>+</td>
</tr>
<tr>
<td>V(B2V)</td>
<td>+</td>
<td>?P</td>
<td>+</td>
</tr>
<tr>
<td>V(V1B)</td>
<td>+</td>
<td>?P</td>
<td>+</td>
</tr>
<tr>
<td>V(V2B)</td>
<td>+</td>
<td>?P</td>
<td>+</td>
</tr>
</tbody>
</table>
Experiments were performed by our collaborators to determine whether the C-terminal TRAM domain, predicted to bind tRNA, of YqeV and YmcB was involved in the recognition of the specific moiety, $N^6$-isopentenyladenosine or $N^6$-threonylcarbamoyladenosine, present at the $N^6$ position of adenosine. Chimeric proteins were constructed by combining the N-terminal and radical SAM domain of one methylthiotransferase protein with the TRAM domain of the other methylthiotransferase. The boundaries of the three functional domains of YqeV and YmcB were predicted using the Pfam database [264]. Because the area of protein sequence between the radical SAM and TRAM domains was very similar, it was unclear where the functional residues of the SAM domain ended and the TRAM domain began. Thus, two different splice or breakpoints were used for the construction of the chimeric proteins. One breakpoint was close to the radical-SAM domain (breakpoint 1) and the other was closer to the boundary of the TRAM domain (breakpoint 2). The chimeric proteins generated were designated as B1V, B2V, V1B and V2B; the first letter denoting the source of the N-terminal and radical SAM domains and the second letter denoting the source of the TRAM domain (B = YmcB; V = YqeV) and the number between them refers to the location of the breakpoint. The four chimeric proteins were generated in each of the two mutant strains.

In the BSF2608 strain, the chimeric protein with N-terminal and SAM domains of the ymcB gene fused with the C-terminal TRAM domain of the yqeV gene produced the $m^2i^6$A modification, B(B1V) and B(B2V). In the corresponding domain switch, where the N-terminal and SAM domains originate from the yqeV gene and the TRAM domain originates from the ymcB gene, no $m^2i^6$A modification was found. This indicates the TRAM domain is likely not involved with discrimination of the $N^6$-isopentenyl versus the $N^6$-threonylcarbamoyl moiety of the tRNA substrate. The results were the same regardless of the splice or breakpoint, 1 or 2, used.

Complementary results were not obtained in the V1B or V2B chimeras of the YQEVd strain as modification rescue was not observed. These proteins were observed to be expressed at a much lower level than the B1V and B2V proteins, and some protein aggregation was observed indicating misfolding.
thus inactive proteins. These factors may have contributed to the inability of these chimeric proteins to catalyze formation of $\text{ms}^2\text{t}^6\text{A}$ in these experiments.

4.4 Discussion

Although numerous and diverse modified nucleosides have been identified and characterized in tRNA, the modifying enzymes responsible for their biosynthesis are less well characterized. In this work, ymcB and yqeV, two methylthiotransferases, predicted by phylogenetic comparisons and homology searches of the complete B. subtilis genome, were shown to be the modifying enzymes responsible for the formation of $\text{ms}^2\text{t}^6\text{A}$ and $\text{ms}^2\text{i}^6\text{A}$, respectively. The free radical type mechanism used by these methylthiotransferase modifying enzymes for the chemically difficult aromatic C-H to C-SCH$_3$ bond conversion is unique among known RNA modifying enzymes. In 2002, E. coli MiaB was the first Fe-S cluster containing protein shown to participate in tRNA modification.

Based on previous studies of modification status under various conditions, these methylthiotransferases may be important in cellular processes such as growth, sporulation and iron utilization in B. subtilis. The degree of methylthiolation in B. subtilis and E. coli varies with growth phase and availability of iron. During exponential growth in B. subtilis, $\text{i}^6\text{A}$ predominates and during stationary phase $\text{ms}^2\text{i}^6\text{A}$ is more abundant. [43] In B. subtilis under nutrient limited conditions sporulation begins and methylthiolation of $\text{i}^6\text{A}$ may be required for spore formation as there is evidence that tRNA lysine, which contains $\text{ms}^2\text{i}^6\text{A}$, changes during sporulation. [265, 266] Availability of iron affects the synthesis of $\text{ms}^2\text{i}^6\text{A}$ in E. coli tRNA tyrosine, resulting in undermodification. [49] Similarly the presence of iron binding proteins produced undermodified tRNA with respect to $\text{ms}^2\text{i}^6\text{A}$. Whether the effects of iron concentration on $\text{ms}^2\text{i}^6\text{A}$ synthesis are a consequence of the requirement of miaB for Fe is unknown. Now that the tRNA methylthiotransferases of B. subtilis have been identified and functionally characterized, further studies can begin to reveal regulatory or other functions.
4.5 Conclusion

In this work, LC/MS was used to identify modified adenosines for functional characterization of the methylthiotransferases of \textit{B.subtilis} tRNA. It was determined that \textit{ymcB} is required for the formation of ms$^{2}i^{6}$A and \textit{yqeV} is required for the formation of ms$^{2}t^{6}$A. The domain swapping experiments indicate the discrimination between $i^{6}$A and $t^{6}$A in the methylthiotransferases likely does not involve the C-terminal TRAM domain of the enzyme. Now that these methylthiotransferase genes have been characterized further experiments can investigate substrate recognition of these modifying enzymes.

It may be particularly interesting to further explore the role of these methylthiolases in sporulation or other regulatory processes. The reversibility and redox properties of these enzymes are also interesting and could be ways in which the modifying enzymes could exhibit dynamic control in tRNA modification.
CHAPTER 5

Quantifying Changes in Modified Nucleosides by LC-UV-MS: A Reproducibility Study

5.1 Introduction

The ability to measure the differences in the degree of RNA modification, in response to different physiological or environmental constraints of the cell, will further our understanding of the functional and structural roles and the possible dynamic regulation of post-transcriptional modification. For meaningful comparisons of global modification status under different conditions to be made, the instrumental and biological variability of a method for identification and quantitation of modified nucleosides must be characterized. Large variability in method precision diminishes the significance of the results, if the variability is too great important differences could be missed. The goal of this study is to quantitatively measure the relative amounts of modified nucleosides present in rRNA and tRNA and to characterize the reproducibility of the LC-UV-MS method used. With this information one can define the statistical significance of changes in modification status when comparing biological systems of varying environmental or physiological cellular conditions.

5.2 Experimental

5.2.1 Materials

Culture media used was Bacto tryptone and Bacto yeast extract from Becton, Dickinson & Co (Sparks MD). Ammonium chloride, ammonium acetate, acetic acid, Tris-HCl, potassium chloride, magnesium acetate, and sodium citrate used in buffers were purchased from Sigma (St. Louis, MO). Sodium chloride, ammonium bicarbonate and EDTA were obtained from Fisher Scientific (Fairlawn, NJ). Lysozyme (from chicken egg), urea, ethanol, propanol, sucrose and Tri-Reagent were purchased from Sigma. Nanopure water (18 Mohms) from a Barnstead (Dubuque, IA) filtering system was autoclaved and used in all buffers and solutions. RNase free DNase was purchased from Promega (Madison, WI).
Phenol/Chloroform/Isoamylalcohol (25:24:1) solution was obtained from Ambion (Austin, TX). Nucleobond AXR-80 gravity flow columns were purchased from Machery-Nagel (Bethlehem, PA). Nuclease P1 and nucleoside test mix were obtained from Sigma. Snake venom phosphodiesterase I was purchased from Worthington Biochemicals (Lakewood, NJ) and antartic phosphatase from New England Biolabs (Ipswich, MA).

5.2.2 Escherichia coli Culturing

An Innova 4000 incubator from New Brunswick Scientific (Edison, NJ) was used for culturing. RNA concentration and cell density measurements were made on a Shimadzu Biospec 1601 UV/Vis spectrometer (Columbia, MD) at 260 nm and 600 nm, respectively. Bacterial cells were lysed with a French Press (American Instrument Co., Silver Spring, MD). A Sorvall RC5C preparative centrifuge was used to pellet bacterial cells. A Beckman Coulter Optima L-XP ultracentrifuge (Fullerton, CA) was used for ribosome subunit separation and pelleting.

Bacterial cultures were prepared by inoculating 100 mL of Luria Broth with a small aliquot of E. coli MRE 600 stock culture and incubating at 37 °C for 16 h with agitation. This culture was then added to 900 mL of fresh Luria Broth and incubated at 30 °C until mid-log phase (OD 0.5 to 0.7 at 600 nm). Cells were harvested by centrifugation at 10,000 x g for 15 min at 4 °C and washed with buffer (20 mM Tris-HCl, 10.5 mM magnesium acetate, 0.5 mM EDTA, pH 7.5).

5.2.3 Isolation and Purification of RNA

Procedures used for isolation of ribosomes and ribosomal subunits were adapted from previously described standard protocols. The bacterial cells were physically lysed in a French press (12,000 psi) and cell debris removed via centrifugation. The cell lysate was treated with RNase-free DNase (2.5
units DNase/gram of cell). Crude ribosomes were isolated via centrifugation through a 1.1 M sucrose cushion. The resulting 70S pellet was resuspended in a low magnesium buffer (20 mM Tris-HCl at pH 7.6, 10.5 mM magnesium acetate, and 30 mM ammonium chloride) and allowed to dissociate into 50S and 30S subunits overnight at 4 °C. The subunits were separated through a 0 – 45% sucrose gradient by centrifugation on a Beckman L-XP ultracentrifuge at 19,000 rpm for 17 h in a SW28 rotor. The 30S and 50S subunit fractions were collected, pooled and pelleted via centrifugation at 48,000 rpm for 18 h. Subunits were resuspended in 20 mM Tris-HCl, 10.5 magnesium acetate, 30 mM ammonium chloride, pH 7.6 and stored at -80 °C.

Two methods were used to isolate 16S rRNA. In the first, ribosomal subunits were deproteinized using phenol/chloroform extraction followed by ethanol precipitation as previously described. [268] Alternatively, ribosomal proteins were denatured with 6 M urea and the 16S RNA was purified using a Nucleobond (Machery-Nagel) ion exchange column. Denatured subunits were loaded onto the column in a solution of 500 mM potassium chloride, 100 mM tris acetate (pH 6.3) and 15% ethanol then washed with the same solution. A solution of 800 mM potassium chloride, 100 mM tris acetate (pH 6.3) and 15% ethanol was used for a second wash followed by elution of rRNA with 1.5 M potassium chloride, 100 mM tris acetate (pH6.3) 15% ethanol. Isopropanol was used to precipitate the rRNA and the pelleted rRNA was resuspended in autoclaved water.

For tRNA isolation, lysozyme (1 mg/mL) was used for cell lysis. Total tRNA was isolated by guanidinium thiocyanate-phenol-bromochloroform extraction. rRNA was separated from tRNA in a solution of high salt concentration and stepwise precipitation with isopropanol. Salts and phenol were removed using the Nucleobond ion exchange column as described above. Buffer concentrations were lower for tRNA purification: loading buffer, wash buffer and elution buffers were 200 mM, 400 mM and 750 mM potassium chloride, respectively.
5.2.4 Enzymatic Digestion of RNA

Prior to enzymatic digestion of the RNA to nucleosides, the RNA was denatured at 100 °C for 3 min then chilled in an ice water bath. To lower the pH, 1/10 volume 0.1 M ammonium acetate (pH 5.3) was added. For each 0.5 AU of RNA, 2 units Nuclease P1 was added and incubated at 45 °C for 2 h. The pH was readjusted by adding 1/10 volume of 1.0 M ammonium bicarbonate, then 0.002 units of snake venom phosphodiesterase was added and incubated at 37 °C for 2 h. Finally, 0.5 units of antarctic phosphatase was added and incubated at 37 °C for 60 min. The nucleoside digests were stored at -80 °C. [64]

5.2.5 LC-UV and LC-MS Conditions

Analysis of the nucleoside test mix and nucleoside digests of RNA by analytical HPLC were done using a Hitachi D-7000 HPLC equipped with a diode array detector. A 4.6 x 250 mm Supelcosil LC-18-S 5 µm particle reversed phase column with a 4.0 x 20 mm Supelguard LC-18-S guard column was used at a flow rate of 2.0 mL/min and temperature controlled at 30 °C. Mobile phases used were 250 mM ammonium acetate, pH 6.0 (Buffer A) and 40% aqueous acetonitrile (Buffer B). A multilinear gradient was used with only minor modification from that described previously. [65]

Analysis of the nucleoside test mix and nucleoside digests of RNA by narrow bore HPLC were done using a Hitachi D-7000 HPLC system. A 2.1 x 250 mm Supelcosil LC-18-S 5 µm particle reversed phase column was used at a flow rate of 300 µL/min and either at a temperature controlled 30°C or at a room temperature of 20 °C. Mobile phases used were 5 mM ammonium acetate, pH 5.3 (Buffer A) and 40% aqueous acetonitrile (Buffer B). The gradient used was the same as for analytical HPLC, with slight adjustments for column size differences (Table 2.2). The column eluent was split immediately post column, 1/3 to a Thermo LTQ-XL ion trap mass spectrometer and 2/3 to a Hitachi D-7400 UV detector set at 260 nm.
A Thermo LTQ-XL ion trap mass spectrometer equipped with an ion max electrospray source was used for the LC/MS identification of nucleosides. Mass spectra were recorded in the positive ion mode over an \( m/z \) range of 105-510 with a capillary temperature of 275 °C, spray voltage of 3.7 to 4.0 kV and sheath gas, auxiliary gas and sweep gas of 45, 25 and 10 arbitrary units, respectively. Data dependent MS/MS of each of the two most intense ions were recorded throughout the LC/MS run.

5.3 Results and Discussion

5.3.1 Nucleoside Quantitation: Technical Reproducibility

The LC/MS methods used here are based on several well-established methods for the identification of post-transcriptional modifications in enzymatic digests of RNA. [65, 269] In this work the instrumental and biological reproducibility were determined for the quantitative analysis of modified nucleosides in enzymatic digests of tRNA and rRNA. The reproducibility was evaluated using reversed phase analytical (4.6 x 250 mm) and narrow bore columns (2.1 x 250mm), both frequently used for the analysis of nucleosides in RNA. [39, 52, 57, 78, 79, 81]. The analytical column has greater loading capacity, does not require a low volume detector flow cell and needs less equilibration time between runs than the narrow bore column. The analytical column is useful in the analysis of nucleoside digests of large ribosomal RNA where it is necessary to load a larger amount due to the dynamic range requirements, for example, in 16S RNA where only 11 of the 1542 nucleotides are modified. The advantages of the narrow bore column are reduced solvent consumption, increased sensitivity and greater compatibility with electrospray ionization due to the lower flow rate and smaller column volume. The narrow bore column is most often used when MS-based nucleoside identification is required. The analysis conditions were the same for both columns except different HPLC instruments were used and the mobile phase flow rate and concentration of ammonium acetate were reduced as required for compatibility with the electrospray process.
The precision of the HPLC method was evaluated by performing replicate analyses of a commercially available nucleoside test mix over multiple days using different HPLC instruments in different laboratories. The test mix contained 12 nucleosides found frequently in RNA, pseudouridine, cytidine, uridine, 2-thiocytidine, 3-methylcytidine, 5-methylcytidine, 2'-O-methylcytidine, 1-methyladenosine, inosine, 5-methyluridine, guanosine, and 7-methylguanosine. Three of the major nucleosides present in RNA, cytidine, uridine and guanosine and several modified (minor) nucleosides that are present at much lower concentrations in RNA were included in the test mix. The identification of nucleosides in the test mix was based on relative retention times and either UV absorbance characteristics (ratio of absorbance at 260/280 nm) for the analytical column or ESI-MS spectral characteristics for the narrow bore column.

A representative UV chromatogram of the test mix separated on the analytical column is shown in Figure 5.1. The instrumental analysis reproducibility, determined as the variability in peak areas and retention times, is summarized in Table 5.1. Replicate injections (n=4) yielded relative standard deviations (expressed as a percentage, %RSD) for peak areas with a range of 2.9 to 4.3 % and an average of 3.4 % RSD, regardless of the amount of any individual nucleoside injected. Reproducibility of nucleoside retention times was between 1.5% and 3.2% with an average of 2.1% RSD for the analytical column.
Figure 5.1  UV chromatogram of nucleoside test mix, 4.6 x 250 mm / LC-18S column, 30°C, 250 mM ammonium acetate pH 6.0 (A) 40% acetonitrile in water (B), 2.0 mL/min. UV detection @ 260 nm.

Table 5.1  Reproducibility of UV peak area and retention time for analysis of modified nucleosides in standard test mix, 4.6 x 250 mm column used. (n = 4)

<table>
<thead>
<tr>
<th>Peak</th>
<th>Nucleoside</th>
<th>%RSD Peak Area</th>
<th>Avg. t, (min)</th>
<th>%RSD t,</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pseudouridine</td>
<td>2.9</td>
<td>3.4</td>
<td>2.8</td>
</tr>
<tr>
<td>2</td>
<td>Cytidine</td>
<td>3.6</td>
<td>4.6</td>
<td>1.7</td>
</tr>
<tr>
<td>3</td>
<td>Uridine</td>
<td>3.4</td>
<td>6.2</td>
<td>2.0</td>
</tr>
<tr>
<td>4</td>
<td>2-thiocytidine</td>
<td>3.4</td>
<td>7.3</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>3-methylcytidine</td>
<td>3.1</td>
<td>9.1</td>
<td>3.2</td>
</tr>
<tr>
<td>6</td>
<td>5-methylcytidine</td>
<td>4.3</td>
<td>9.5</td>
<td>2.7</td>
</tr>
<tr>
<td>7</td>
<td>2'-O-methylcytidine</td>
<td>3.2</td>
<td>11.5</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>5-methyluridine</td>
<td>3.1</td>
<td>12.8</td>
<td>1.8</td>
</tr>
<tr>
<td>9</td>
<td>Inosine</td>
<td>3.6</td>
<td>13.5</td>
<td>1.7</td>
</tr>
<tr>
<td>10</td>
<td>Guanosine</td>
<td>3.3</td>
<td>14.4</td>
<td>1.5</td>
</tr>
<tr>
<td>11</td>
<td>7'-methylguanosine</td>
<td>3.1</td>
<td>15.0</td>
<td>1.8</td>
</tr>
<tr>
<td>12</td>
<td>N''-methyladenosine</td>
<td>3.5</td>
<td>24.6</td>
<td>1.6</td>
</tr>
</tbody>
</table>
A similar evaluation of the reproducibility of the narrow bore column was performed. Two different temperatures, 20 °C and 30 °C, were used for the analysis of nucleosides in the RNA hydrolysates to enable separate optimization for tRNA and rRNA nucleosides, respectively. The large dynamic range between 7-methylguanosine and guanosine in rRNA digests requires the higher temperature (30 °C) for adequate separation. Improved separation of 3-methy from 5-methyl cytidine as well as 5-methyluridine from inosine was noted when the chromatography was performed at 20 °C. Because tRNA contain these modified nucleosides, subsequent experiments probing the biological reproducibility of tRNA analyses was performed at 20 °C. Figure 5.2 contains a representative UV chromatogram for the nucleoside test mix, with variability in peak areas and retention times summarized in Table 5.2 (analysis at 20 °C) and Table 5.3 (analysis at 30 °C). Relative standard deviations for peak areas at this chromatographic scale were slightly higher (1.0% - 5.8%) with an average of 3.5% RSD than for the analytical column. Retention time differences between the two columns can be attributed to the difference in dead volumes and gradient delays in the two different instruments used and the lower flow rate used (300 µL/min). The method precision reported as percent relative standard deviation in peak areas of the standard test mix here is similar to the method precision reported in studies of modified nucleosides in urine (Table 5.4).

<table>
<thead>
<tr>
<th>Peak</th>
<th>Nucleoside</th>
<th>%RSD UV Peak Area</th>
<th>%RSD ( t_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pseudouridine</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>Cytidine</td>
<td>2.7</td>
<td>3.2</td>
</tr>
<tr>
<td>3</td>
<td>Uridine</td>
<td>4.3</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>2-thiocytidine</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>5</td>
<td>5-methylcytidine</td>
<td>1.0</td>
<td>3.8</td>
</tr>
<tr>
<td>6</td>
<td>3-methylcytidine</td>
<td>5.5</td>
<td>3.6</td>
</tr>
<tr>
<td>7</td>
<td>2'-O-methylcytidine</td>
<td>4.1</td>
<td>2.7</td>
</tr>
<tr>
<td>8</td>
<td>1-methyladenosine</td>
<td>4.1</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>Inosine</td>
<td>1.9</td>
<td>1.3</td>
</tr>
<tr>
<td>10</td>
<td>5-methyluridine</td>
<td>3.2</td>
<td>1.3</td>
</tr>
<tr>
<td>11</td>
<td>Guanosine</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>12</td>
<td>7-methylguanosine</td>
<td>3.0</td>
<td>1.0</td>
</tr>
<tr>
<td>13</td>
<td>3-methyluridine</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td>14</td>
<td>N6-methyladenosine</td>
<td>0.9</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 5.2 Reproducibility of UV peak areas and retention times for modified nucleoside in standard test mix, 2.1 x 250 mm column used, 20 °C.
Figure 5.2 UV chromatogram of nucleoside test mix, 2.1 x 250 mm / LC-18S column, 30°C, 5 mM ammonium acetate pH 5.3 (A) 40% acetonitrile in water (B), 0.3 mL/min. UV detection @ 260 nm. (a) 20 °C and (b) 30 °C.
Table 5.3. Reproducibility of UV and MS peak areas and retention times for modified nucleosides (both molecular ion [MH$^+$] and base ion [BH$_2^+$]) in standard test mix, 2.1 x 250 mm column used, 30°C. (* denotes UV peak not chromatographically resolved)

<table>
<thead>
<tr>
<th>Peak</th>
<th>Nucleoside/Base</th>
<th>(m/z)</th>
<th>%RSD UV Peak Area</th>
<th>%RSD $t_r$</th>
<th>%RSD MS Peak Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pseudouridine</td>
<td>(245)</td>
<td>3.0</td>
<td>0.2</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>(209)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Cytidine</td>
<td>(244)</td>
<td>3.0</td>
<td>0.2</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>(112)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Uridine</td>
<td>(245)</td>
<td>3.4</td>
<td>0.2</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td>(113)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2-thiocytidine</td>
<td>(260)</td>
<td>4.3</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>(128)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5-methylcytidine</td>
<td>(258)</td>
<td>3.0</td>
<td>0.8</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>(126)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3-methylcytidine</td>
<td>(258)</td>
<td>3.6</td>
<td>0.9</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>(126)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2'-O-methylcytidine</td>
<td>(258)</td>
<td>1.0</td>
<td>1.1</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>(112)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1-methyladenosine</td>
<td>(282)</td>
<td>2.0</td>
<td>0.7</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>(150)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Inosine</td>
<td>(269)</td>
<td>1.9</td>
<td>0.7</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>(137)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5-methyluridine</td>
<td>(259)</td>
<td>*</td>
<td>1.4</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>(127)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Guanosine</td>
<td>(284)</td>
<td>3.2</td>
<td>0.3</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>(152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>7-methylguanosine</td>
<td>(298)</td>
<td>5.8</td>
<td>0.2</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>(166)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3-methyluridine</td>
<td>(259)</td>
<td>5.8</td>
<td>0.3</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>(127)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>N$^6$-methyladenosine</td>
<td>(282)</td>
<td>5.1</td>
<td>0.8</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>(150)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.4 Reported reproducibilities for analysis of modified nucleosides in urine from other studies using UV or MS detection. MRM denotes multiple reaction monitoring, SIM denotes selected ion monitoring and CNL denotes constant neutral loss.

<table>
<thead>
<tr>
<th>% RSD</th>
<th>Samples Evaluated</th>
<th># Nucleosides</th>
<th>Reference</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 11.4%</td>
<td>standard nucleosides</td>
<td>10</td>
<td>[82]</td>
<td>UV</td>
</tr>
<tr>
<td>&lt; 1.0 - 14.9 %</td>
<td>standard nucleosides</td>
<td>12</td>
<td>[83]</td>
<td>MS (MRM)</td>
</tr>
<tr>
<td>1.2 - 21.3 %</td>
<td>single urine sample</td>
<td>12</td>
<td>[270]</td>
<td>UV</td>
</tr>
<tr>
<td>3.3 - 31.5 %</td>
<td>single urine sample</td>
<td>18</td>
<td>[271]</td>
<td>UV</td>
</tr>
<tr>
<td>5.0 - 16.2%</td>
<td>single urine sample</td>
<td>9</td>
<td>[86]</td>
<td>MS (SIM)</td>
</tr>
<tr>
<td>2.0 - 6.0%</td>
<td>standard nucleosides</td>
<td>12</td>
<td>[272]</td>
<td>UV</td>
</tr>
<tr>
<td>5.6 - 33.1%</td>
<td>single urine sample</td>
<td>12</td>
<td>[272]</td>
<td>UV</td>
</tr>
<tr>
<td>3.0 - 14%</td>
<td>single urine sample</td>
<td>12</td>
<td>[273]</td>
<td>UV</td>
</tr>
<tr>
<td>5.5 - 48.6%</td>
<td>single urine sample</td>
<td>35</td>
<td>[274]</td>
<td>MS (CNL)</td>
</tr>
<tr>
<td>0.5 - 14.4%</td>
<td>standard nucleosides</td>
<td>9</td>
<td>[275]</td>
<td>MS (MRM)</td>
</tr>
</tbody>
</table>

To compare the precision of quantitative measurements of nucleosides using UV detection with mass spectrometry detection, MS peak areas were also determined for each of the nucleosides in the test mix by extracted ion chromatograms (XICs) for each nucleoside molecular ion (MH⁺) and base ion (BH₂⁺). Because nucleosides containing an N-C glycosidic bond can fragment easily during ESI-MS, leading to the formation of the BH₂⁺ ion, peak areas for both ions were measured. Pseudouridine contains a more stable C-C bond rather than the C-N glycosidic bond and does not fragment to yield a base ion. However, pseudouridine fragments easily forming a characteristic electrospray spectra with an abundant ion at m/z 209 which results from the elimination of 2 water molecules [72] and this ion was measured.

For MS detection the average percent RSD in peak areas for the ions measured was 5.9 % RSD (Table 5.3). The relative standard deviations are slightly higher for the MS peak areas of most nucleoside ions when compared to the UV peak areas. Mass spectra were collected in full scan mode over the m/z
range 103 to 510 in the ion trap, thus the presence of interfering ions could be contributing to variability in MS peak area. Smaller variations in peak areas are likely to be obtained in more selective analyses such as selected reaction monitoring scans where specific molecular ion to base ion transitions are recorded. The SRM (MRM) approach has been used for the analysis of urinary nucleosides resulting in reported precisions of less than 15% (Table 5.4). The use of an internal standard could also be used to decrease variability in both UV and MS peak areas. Variability in MS peak areas in this study ranged from 1.0 to 12.4 % (not including the base ion for guanosine which exhibited 20% RSD in peak area) which is slightly less than that reported for similar MS studies of modified nucleosides present in urine (Table 5.4).

Two nucleosides not listed as components of the test mix were identified by ESI-MS and found to be N<sup>6</sup>-methyl adenosine and 3-methyluridine. 1-methyladenosine undergoes a Dimroth rearrangement to form N<sup>6</sup>-methyladenosine that is 90% complete at pH 11 and 50% complete at pH 7. [113] The presence of N<sup>6</sup>-methyladenosine is most likely due to this rearrangement because 1-methyladenosine was not found in the older nucleoside test mix used in analyses using the analytical column. The formation of 3-methyluridine is likely an unintended synthesis product of 5-methyluridine.

5.3.2 Biological Reproducibility

Having established the technical reproducibility for LC-UV and LC-MS detection of modified nucleosides, I next sought to determine the biological variability associated with measuring modified nucleosides present in tRNAs and rRNAs. To conduct these measurements, separate cultures of E. coli were grown and the appropriate RNAs were isolated. Except where noted, identical growth and isolation protocols were followed, and any variability associated with the isolation steps will be reflected in the overall variability of the measured nucleoside abundances.
5.3.2.1 Transfer RNA

Determining the amounts of the various modified nucleosides to be expected in unfractionated tRNA is not a straightforward task. *E. coli* contains 26 different modified nucleosides distributed unevenly among the 46 different tRNAs. (Table 1.3) Some modifications such as pseudouridine and 5-methyluridine are common to all the tRNAs. Other modifications are present in the majority but not all tRNAs, for example D, s^4U, m^7G and Gm. Several modifications, mnm^5U, Cm, mnm^5s^2U, I, ac^4C, m^6t^6A, m^6A, i^6A occur in only one or a few specific tRNAs. Most modifications are specific to one location within the tRNA sequence with the exception of dihydrouridine and pseudouridine which are located at several different positions within the same tRNA. All tRNAs contain pseudouridine at position 55, but several tRNAs contain an additional pseudouridines at other locations. The D loop and stem domain of tRNA often contains more than one dihydouridine.

As discussed in Chapter 1, the different tRNA species are not present in equal abundance within the cell. The relative abundance of the individual tRNAs is thought to be correlated to the frequency with which the cognate codon occurs on the mRNA. A codon bias exists, which varies among species, varies with cellular conditions such as amino acid availability and is correlated with growth rate. Thus the complexity and large dynamic range in the analysis of modified nucleosides present in total tRNA nucleoside digests result from both the differences of frequency in occurrence of specific modifications in the 46 tRNAs and the effects of codon bias which is reflected in the differential expression of individual tRNAs.

Data obtained from the analysis of modified nucleosides present in *E. coli* tRNAs is presented in Table 5.5 with a corresponding representative chromatogram shown in Figure 5.3. These data were obtained using the narrow bore column following the tRNA isolation and enzymatic digestion protocols described in the Experimental section. For each analysis, 25 ug of total unfractionated tRNA was enzymatically digested to nucleosides and injected on column. All of the 25 known modified
nucleosides present in E. coli tRNA have been identified in this analysis using the combination of retention time, UV and MS characteristics (Figure 5.3). However, only a subset of modifications is adequately resolved, chromatographically, to obtain UV peak area measurements. The relative amounts of each nucleoside present were calculated based on peak area measurements using N6-threonylcarbamoyladenosine as an internal reference peak. This was used to eliminate differences in injection volume and slight variations in tRNA concentration. The variability of these relative amounts calculated as percent relative standard deviation is reported for this subset of modifications in Table 5.5 and ranged from 4.2 to 9.4 %RSD. UV peak area measurements were not obtained for dihydrouridine because it does not possess a chromophore and thus will not be detected by UV absorbance. Other nucleosides such as mnm5U, cmnm5Um and k2C are found only in rare tRNAs and thus were present in very small amounts and only detected by MS.

The variability in MS peak areas measured for both the molecular [MH+] and base ions [BH2+] are also reported and ranged from 1 to 12 %RSD. The MS peak area data is obtained from four separate tRNA isolations from one culture. For pseudouridine, the molecular ion and the characteristic ion of m/z 209 formed from water loss were measured. For a few nucleosides such as cmnm5Um, mnm5s2U, k2C and i6A base ions arising from in-source fragmentation were not detected in the XICs. However, the base ions were detected in the MS/MS analysis. For the isobaric nucleosides Gm and m1G only the MS peak area of the base ions were measured. The variability shown here for the analysis of modified nucleosides in tRNA is similar to the variability reported for modified nucleosides present in urine for both UV and MS detection (Table 5.4).
Figure 5.3. Modified nucleosides in Escherichia coli total tRNA. 1, dihydrouridine (D). 2, pseudouridine (Ψ). 3, 5-carboxymethoxyuridine (cmo U). 4, 3-(3-amino-3-carboxypropyl)uridine (acp U). 5, 5-methylaminomethyluridine (mmn U). 6, 2-thiocytidine (sC). 7, 2'-O-methylcytidine (Cm). 8, 5-carboxymethylaminomethyl-2'-O-methyluridine (cmnm 5Um). 9, 5-methylaminomethyl-2-thiouridine (mmn sU). 10, 5-methyluridine (m 5U). 11, inosine (I). 12, 4-thiouridine (s4 U). 13, 7-methylguanosine (m 7G). 14, 2'-O-methyluridine (Um). 15, 2'-O-methylguanosine (Gm). 16, 1-methylguanosine (m 1G). 17, N4-acetylcytidine (ac4 C). 18, queuosine (Q). 19, lysidine (k2 C). 20, N6-threonylcarbamoyladenosine (t6 A). 21, N6-methyl-N6-threonylcarbamoyladenosine (m6 tA). 22, 2-methyladenosine (m2 A). 23, N6-methyladenosine (m6 A). 24, N6-isopentenyladenosine (i6 A). 25, 2-methylthio-N6-isopentenyladenosine (ms2 iA). (* = unknown)
Table 5.5  Reproducibility of UV and MS peak areas for modified nucleosides (both molecular ion [MH\textsuperscript{+}] and base ion [BH\textsubscript{2}\textsuperscript{+}]) in E. coli tRNA.  Pooled percent relative standard deviations are for 3 or 4 replicate analyses within three separate cultures.  For MS peak area %RSD 4 replicate analyses were performed on tRNA isolated from a single culture.  (*denotes isobaric molecular ion and “nd” denotes base ion not detected in MS analysis)

<table>
<thead>
<tr>
<th>Peak</th>
<th>Nucleoside</th>
<th>t\textsubscript{r} (avg)</th>
<th>%RSD</th>
<th>%RSD pooled</th>
<th>[MH\textsuperscript{+}]/[BH\textsubscript{2}\textsuperscript{+}]</th>
<th>MS Peak Area m/z</th>
<th>%RSD (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D</td>
<td>4.1</td>
<td>3.8%</td>
<td>247/115</td>
<td>4.7%/4.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ψ</td>
<td>4.2</td>
<td>2.4%</td>
<td>245/209</td>
<td>6.0%/5.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>cmo\textsuperscript{5}U</td>
<td>4.5</td>
<td>3.4%</td>
<td>319/187</td>
<td>4.4%/5.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>acp\textsuperscript{3}U</td>
<td>7.6</td>
<td>4.7%</td>
<td>346/214</td>
<td>2.4%/1.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>mmn\textsuperscript{3}U</td>
<td>8.1</td>
<td>7.3%</td>
<td>288/156</td>
<td>8.7%/nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>s\textsuperscript{3}C</td>
<td>9.7</td>
<td>6.7%</td>
<td>260/128</td>
<td>3.0%/3.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Cm</td>
<td>15.7</td>
<td>7.1%</td>
<td>258/112</td>
<td>7.2%/9.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>cmnm\textsuperscript{3}Um</td>
<td>17.9</td>
<td>7.4%</td>
<td>346/200</td>
<td>3.1%/nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>mmn\textsuperscript{1}s\textsuperscript{3}U</td>
<td>18.6</td>
<td>7.2%</td>
<td>304/172</td>
<td>5.6%/nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>m\textsuperscript{5}U</td>
<td>19.3</td>
<td>5.0%</td>
<td>259/127</td>
<td>5.6%/9.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I</td>
<td>19.9</td>
<td>6.5%</td>
<td>269/137</td>
<td>9.3%/7.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>s\textsuperscript{4}U</td>
<td>22.0</td>
<td>5.2%</td>
<td>261/129</td>
<td>1.0%/2.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>m\textsuperscript{3}G</td>
<td>22.6</td>
<td>3.5%</td>
<td>298/166</td>
<td>3.5%/4.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Um</td>
<td>23.1</td>
<td>4.1%</td>
<td>259/113</td>
<td>5.2%/10.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Gm</td>
<td>26.1</td>
<td>1.9%</td>
<td>298/152</td>
<td>*/12.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>m\textsuperscript{1}G</td>
<td>26.4</td>
<td>1.8%</td>
<td>298/166</td>
<td>*/1.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>ac\textsuperscript{4}C</td>
<td>27.0</td>
<td>1.8%</td>
<td>286/154</td>
<td>10.9%/5.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Q</td>
<td>27.8</td>
<td>1.6%</td>
<td>410/295</td>
<td>5.9%/5.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>k\textsuperscript{5}C</td>
<td>28.9</td>
<td>2.0%</td>
<td>372/240</td>
<td>7.8%/nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>\textit{i}\textsuperscript{6}A</td>
<td>30.3</td>
<td>1.7%</td>
<td>413/281</td>
<td>5.2%/1.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>m\textsuperscript{4}t\textsuperscript{6}A</td>
<td>34.2</td>
<td>1.5%</td>
<td>427/295</td>
<td>5.9%/3.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>m\textsuperscript{3}A</td>
<td>34.9</td>
<td>1.5%</td>
<td>282/150</td>
<td>4.6%/7.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>m\textsuperscript{6}A</td>
<td>35.5</td>
<td>1.4%</td>
<td>282/150</td>
<td>6.3%/8.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>\textit{i}\textsuperscript{6}A</td>
<td>49.1</td>
<td>0.3%</td>
<td>336/204</td>
<td>6.7%/nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>ms\textsuperscript{1}\textit{i}\textsuperscript{6}A</td>
<td>57.6</td>
<td>0.3%</td>
<td>382/250</td>
<td>6.6%/5.0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3.2.2 Ribosomal RNA

As mentioned in the Introduction, performing quantitative measurements of modified nucleosides in rRNAs presents an analytical challenge due to the dynamic range of nucleoside quantities, especially for bacterial rRNAs. For *E. coli*, only 7 modified nucleotides are present at 11 of the 1542 sequence locations in 16S rRNA. Thus, the challenge is to obtain useful quantitative information from less than 1% of the total nucleoside content of a 16S rRNA nucleoside digest.

The biological variability was determined by analyzing multiple enzymatic digests of 16S rRNA isolated from three (or four) different cell cultures using the same conditions as for analysis of the nucleoside test mix. Dimethyladenosine was chosen as an internal reference peak for the calculation of relative amounts of modified nucleosides present since it was baseline resolved, similar in size to other modified nucleosides and present in all 16S digests. Relative amounts of the major nucleosides were not determined as they are not the subject of study and exceeded the upper range of detection. An internal reference peak for calculation of relative amounts of modified nucleosides present was used to diminish the effect of variations in the initial amount of digested RNA loaded onto the column for each replicate analysis. Although we sought to load 100 µg of digested RNA the precision and accuracy of this measurement may be limited due to the presence of varying amounts of phenol and large serial dilutions used for UV measurements.

Variability was first examined by loading 100 µg of total 16S rRNA nucleoside digest on the analytical column with UV detection (n=3). Peak area and retention time variability are summarized in Table 5.6 with a representative chromatogram shown in Figure 5.4a. The identities of the major and modified nucleosides present are based on UV spectra and comparison of retention times of nucleosides in the test mix and retention time data. Pseudouridine was not well resolved from a large solvent front in many of the digests analyzed on the analytical column, thus it was not included in the culture to culture reproducibility data. The large peak present at 25 min in the chromatogram originates from the
phenol used to extract ribosomal proteins, which is incompletely removed in the subsequent chloroform extraction and ethanol precipitation of the RNA. This interference is often reported in other HPLC analyses of 16S rRNA [39], and while not an issue in the detection of modifications in *E. coli* rRNA this peak elutes where many modified adenosines elute and may prevent their detection in RNA from other organisms.

Because of the potential interference of the phenol-associated peak and due to the success of the Nucleobond-based purification that was used for tRNA isolations, subsequent studies avoided phenol-chloroform purification of 16S rRNA. The effect of this change in protocol is easily illustrated in Figure 5.4b, which is a representative UV chromatogram of the 16S RNA enzymatic digest analyzed using the narrow bore column. Not only is the phenol-related peak absent, but the ion exchange cartridge separated the DNA more effectively than the phenol extraction thus dG and dA are not present, and the decrease in the solvent front improves resolution of pseudouridine enabling higher quality measurements of its abundance in the sample.

For both the analytical column and the narrow bore column the relative amount present of each nucleoside and corresponding % RSD was calculated for replicate analyses within each of the different cell cultures and is shown in parentheses (*Tables 5.6 and 5.7*). The overall biological variability for measurement of each nucleoside was then calculated using the calculated pooled standard deviations for the nucleoside measurements in each of the different cell cultures. The relative standard deviation in peak areas of the modified nucleosides present in 16S RNA ranged from 3.0 to 10.7% (*Table 5.6*) when analyzed using the analytical column. Retention time variability was less than 5% for all modified nucleoside peaks. The reproducibility of peak areas of modified nucleosides in 16S RNA digests as determined on the narrow bore column ranged from 1.8% to 15.6% as shown in *Table 5.7*. The variability reported here is less than that reported for similar analyses of urinary nucleosides. (*Table 5.3*) Although the dynamic range is not as large in the analysis of urinary nucleosides and fewer steps are
involved in the isolation procedures both analyses seek to compare the total profile and determine statistically significant differences in amounts of nucleosides present in different biological conditions.

**Figure 5.4** (a) UV chromatogram of E. coli 16S RNA nucleoside digest, 4.6 x 250 mm / LC-18S column, 30°C, 250 mM ammonium acetate pH 6.0 (A), 40% acetonitrile in water (B), 2.0 mL/min, UV detection @ 260 nm. (b) UV chromatogram of E. coli 16S RNA nucleoside digest, 2.1 mm x 250 mm / LC-18S column, 30°C, 5 mM ammonium acetate pH 5.3 (A), 40% acetonitrile in water (B), 0.3 mL/min, UV detection @ 260 nm. 1 = Ψ, 2 = m^5C, 3 = m^7G, 4 = m^3U, 5 = m^4Cm, 6 = m^2G, 7 = m^6_2A.
Table 5.6 Biological Reproducibility for analysis of modified nucleosides in *E. coli* 16S rRNA, 4.6 x 250 mm column used. Amounts of modified nucleosides present are based on UV peak areas relative to m^6_A. Percent relative standard deviations in parentheses are for three replicate analyses within each separate culture.

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Culture 1</th>
<th>Culture 2</th>
<th>Culture 3</th>
<th>Average</th>
<th>% RSD (pooled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ψ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>m^3C</td>
<td>0.457 (8.9%)</td>
<td>0.465 (4.1%)</td>
<td>0.424 (4.2%)</td>
<td>0.449</td>
<td>6.2%</td>
</tr>
<tr>
<td>m^7G</td>
<td>0.319 (8.9%)</td>
<td>0.339 (7.4%)</td>
<td>0.349 (8.5%)</td>
<td>0.336</td>
<td>7.5%</td>
</tr>
<tr>
<td>dG</td>
<td>0.415 (6.5%)</td>
<td>0.444 (3.6%)</td>
<td>*</td>
<td>0.430</td>
<td>5.2%</td>
</tr>
<tr>
<td>m^3U</td>
<td>0.443 (3.6%)</td>
<td>0.484 (3.4%)</td>
<td>*</td>
<td>0.464</td>
<td>3.5%</td>
</tr>
<tr>
<td>m^4Cm</td>
<td>0.293 (12.6%)</td>
<td>0.250 (9.4%)</td>
<td>0.247 (5.2%)</td>
<td>0.245</td>
<td>10.7%</td>
</tr>
<tr>
<td>m^2G</td>
<td>1.699 (3.0%)</td>
<td>1.71 (3.2%)</td>
<td>1.633 (2.6%)</td>
<td>1.681</td>
<td>3.0%</td>
</tr>
<tr>
<td>m^6_A</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5.7 Biological Reproducibility for analysis of modified nucleosides in *E. coli* 16S RNA, 2.1 x 250 mm column used. Amounts of modified nucleosides present are based on UV peak areas relative to m^6_A. Percent relative standard deviations in parentheses are for four replicate analyses within each separate culture.

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Culture 4</th>
<th>Culture 5</th>
<th>Culture 6</th>
<th>Culture 7</th>
<th>Average</th>
<th>% RSD (pooled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ψ</td>
<td>0.497 (3.9%)</td>
<td>0.458 (2.9%)</td>
<td>0.474 (6.1%)</td>
<td>0.466 (2.6%)</td>
<td>0.474</td>
<td>4.1%</td>
</tr>
<tr>
<td>m^3C</td>
<td>0.545 (2.6%)</td>
<td>0.542 (1.0%)</td>
<td>0.533 (1.9%)</td>
<td>0.544 (1.6%)</td>
<td>0.541</td>
<td>1.8%</td>
</tr>
<tr>
<td>m^7G</td>
<td>0.762 (6.5%)</td>
<td>0.731 (6.9%)</td>
<td>0.800 (21%)</td>
<td>0.816 (16.8%)</td>
<td>0.777</td>
<td>14.9%</td>
</tr>
<tr>
<td>m^3U</td>
<td>0.554 (4.0%)</td>
<td>0.555 (4.7%)</td>
<td>0.559 (5.7%)</td>
<td>0.573 (3.6%)</td>
<td>0.560</td>
<td>4.6%</td>
</tr>
<tr>
<td>m^4Cm</td>
<td>0.510 (3.7%)</td>
<td>0.519 (4.0%)</td>
<td>0.514 (6.9%)</td>
<td>0.515 (6.7%)</td>
<td>0.515</td>
<td>5.4%</td>
</tr>
<tr>
<td>m^2G</td>
<td>1.929 (8.7%)</td>
<td>1.918 (9.1%)</td>
<td>1.761 (17.5%)</td>
<td>1.877 (13.5%)</td>
<td>1.871</td>
<td>15.6%</td>
</tr>
<tr>
<td>m^6_A</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The differences in measured amounts of nucleosides (normalized to N^6,N^6-dimethyladenosine) present reported in Tables 5.6 and 5.7, likely arise due to both instrumental considerations as well as differences in sample preparation. The rRNA digest analyzed on the 2.1 mm column and prepared using the sample purification protocol used for tRNAs was free of the interferences from phenol and DNA nucleosides. Additionally, this sample purification protocol maintains the pH at 6.3, whereas the more conventional
phenol/chloroform sample purification protocol is performed under more acidic conditions. These results illustrate that the use of solid-phase sample purification and narrow bore chromatography should be preferred for comparative studies on modified nucleosides in rRNA based on both the improved chromatography and the ability to measure more modified nucleosides as interferences such as phenol are eliminated.

Although the goal of this work was not to obtain absolute quantitative measurements of levels of modified nucleosides in rRNA, the data in Table 5.7 agree well with anticipated amounts of these modified nucleosides. A 16S rRNA contains one residue each of pseudouridine, 7-methylguanosine, 3-methyluridine and $N^4,2'$-O-dimethylcytidine. Two residues of 5-methylcytidine and $N^6,N^6'$-dimethyladenosine and three residues of $N^2$-dimethylguanosine are present in 16S rRNA. The obtained data, normalized against $N^6,N^6'$-dimethyladenosine, is consistent with these expected ratios: pseudouridine, 3-methyluridine and $N^4,2'$-O-dimethylcytidine yield peak areas approximately half that for $N^6,N^6'$-dimethyladenosine. $N^2$-methylguanosine and 7-methylguanosine are slightly higher than their expected ratios, while 5-methylcytidine is less than its expected ratio. As these ratios were obtained from UV data, these particular ratios must also incorporate differences in molar absorptivities at 260 nm to more accurately report the absolute quantities of each modified nucleoside in the sample. The relative amounts of each modification are reasonable given the assumption that the 16S rRNA analyzed here is modified stoichiometrically, and the combination of the reasonable experimental precision with realistic relative amounts of modifications allows this method to be used for future quantitative comparisons of changes in nucleoside composition.

5.4 Conclusions

LC/MS methods for the quantitative analysis of biomolecules enable a wide range of applications in a variety of fields. Surprisingly, although LC/UV and LC/MS analyses of modified nucleosides obtained from tRNAs and rRNAs are well represented in the literature, before now there have been no
investigations that establish whether the reproducibility of LC/MS would be sufficient for its use in characterizing changes to modified nucleoside abundance as a function of environmental, physiological or other factors. The results obtained here when measuring modified nucleosides from cellular RNAs are comparable to those obtained via similar LC/UV and LC/MS analyses of urinary nucleosides. (Table 5.4) Although, the dynamic range is not as large in the analysis of urinary nucleosides and fewer steps are involved in the isolation procedures, both analyses seek to compare the total profile and determine statistically significant differences in amounts of nucleosides present in different biological conditions.

Modified nucleosides in tRNAs and, especially, rRNAs may also prove useful when characterizing the extent of drug resistance for certain bacteria/antibiotic combinations. For example, it is known that 16S rRNA methylation is a common mechanism by which pathogenic bacteria develop resistance to common aminoglycosides.[276-278] Analytical methods such as LC/MS that can detect changes in methylation status, including the specific type of methylation, would be attractive options for examining the progression of antibacterial resistance within rRNA. In addition, the ability to monitor changes in tRNA modified nucleoside abundance provide new and unique insights into the regulatory pathways involved in controlling protein translation under a variety of environmental conditions.[88] Based on the findings presented here, quantitative profiling of modified nucleosides from tRNAs and/or rRNAs is feasible, and with reproducibilities of 5 %RSD or less for well-behaved modified nucleosides, this method can expand the range of biological investigations that one wishes to conduct into RNA modification.
Chapter 6  Posttranscriptional Modifications in RNA and Heat Shock

6.1  Introduction

Chapter 5 characterized both the technical and biological reproducibility of the methods used in this work to measure global modification status in tRNA and 16S rRNA. Determination of the variability present in the analysis is necessary to increase the probability of detecting significant changes in RNA modification status and decrease the probability of assigning changes in modification status that are not derived from true differences. In this chapter these analytical methods are used to compare global modification status in tRNA and ribosomal RNA (16S) under two different cellular conditions.

We have chosen to investigate the effect of heat shock on RNA modification in tRNA and rRNA. The heat shock response is a mechanism that allows a cell to cope with a sudden increase in temperature or other environmental stress. In E. coli, heat shock is induced by a non-lethal temperature increase from 30 °C to 42 °C. This sudden increase in temperature results in changes such as increased protein damage and protein denaturation. These unfolded and damaged proteins are detected by cellular systems that induce an increase in translation of heat shock proteins, such as chaperones and proteases, that enable the cell to cope with the changes and potential damage. The initiation of the heat shock response is transcriptionally regulated by an alternative sigma factor, σ32, which directs RNA polymerase to different promoter sites and induces a rapid and transient expression of many genes that code for heat shock proteins. Under normal conditions, σ32 is unstable and must associate with the protein chaperones, DnaK, DnaJ and GrpE. The σ32/chaperone complex cannot initiate transcription. When temperature increases and proteins denature, DnaK detects these damaged proteins and binds to them, thereby releasing σ32. Once released from DnaK, σ32 binds to RNA polymerase initiating increased transcription of heat shock genes for the chaperones, proteases and other proteins that will perform the necessary functions such as protein refolding and repair to enable the cell to adapt to the new environment. When there are no longer
damaged and denatured proteins detected by the cell, unbound DnaK is once again a substrate for σ32 binding, which turns off the heat shock response.

Heat shock proteins are highly conserved and control of their expression is highly variable. At least one rRNA modifying enzyme, FtsJ, the specific methyltransferase responsible for the ribose methylation of uridine at position 2552 in 23S rRNA has been identified as a heat shock protein. [60] Three universally conserved 2'O ribose methylated residues, Um2552, Cm2498 and Gm2251 are all present within the peptidyltransferase center of domain V of 23S RNA. Mutants lacking FtsJ exhibit decreased growth rate, reduced protein synthesis and an accumulation of free ribosomal subunits. Another modifying enzyme, MiaA, is also thought to be under heat shock control. MiaA is a tRNA modifying enzyme that catalyzes the addition of the dimethylallyl moiety to the N6 of adenosine at position 37 to form N6-isopentenyladenosine (i6A). As discussed in Chapter 2, this modification is important in maintaining the correct structure of the anticodon loop for decoding and is also important in maintaining the correct reading frame for decoding. FtsJ and the tRNA modifying enzyme, MiaA, are known products of the σ32 heat shock response regulon and are thought to be important for growth at high temperature. [279, 280] Given the known role of tRNA and rRNA modification in structural stabilization, especially at high temperature, and the possible involvement of RNA modifying enzymes in the heat shock response, an investigation into the modification status under heat shock conditions was performed to determine differences in modifications that may provide additional clues to the functional significance of RNA modifications.

6.2 Experimental

6.2.1 Bacterial Cultures and Heat Shock

Bacterial cultures were prepared by inoculating 100 mL of Luria Broth with a small aliquot of E. coli MRE 600 stock culture and incubating at 37 °C for 16 hours with agitation. This culture was then added to 900 mL of fresh Luria Broth and incubated at 30 °C until mid log phase (OD 0.5 to 0.7 at 600 nm).
Eight – 1 L flasks of exponentially growing cells were prepared simultaneously. Four liters of these cultures were then subjected to heat shock at 42 °C using a water bath incubator with rotary agitation. The remaining 4 L were kept at 30 °C with agitation, as a negative control. Growth phase was monitored throughout culturing including the 50 min heat shock period by turbidity measurement at 600 nm. All cultures were then cooled to 4 °C and the cells harvested by centrifugation at 10,000 x g for 15 min, at 4 °C and washed with buffer (20 mM Tris-HCl, 10.5 mM Mg(CH$_3$COO)$_2$, 0.5 mM EDTA, pH 7.5). Subsequent isolation and purification of tRNA and rRNA from *E. coli* cells was performed as described in Chapter 5.

6.2.2 Western Blotting for DnaK

The total protein concentration of the cell lysate was determined using a Bradford assay. [281] Ten micrograms of total cellular protein was applied per lane on a 10% (Tris-HCl) SDS-PAGE gel using the Laemmli method. [282] Electrophoresis was carried out at constant voltage (200 V) for 55 min. Proteins were then transferred electrophoretically onto a PVDF membrane at constant voltage (100 V) for 40 min. The PVDF membrane was incubated overnight at 4 °C in Tris buffered saline containing 5% nonfat milk to block nonspecific sites. After rinsing with Tris buffered saline containing 0.05% Tween 20 the PVDF membrane was incubated overnight at 4 °C with a 1:1000 dilution of primary DnaK antibody in Tris buffered saline with 5% nonfat milk. PVDF membrane was washed again with Tris buffered saline then incubated with a 1:3000 dilution of secondary antibody Horseradish peroxidase (HRP) conjugate at room temperature for 3 h. Blot color was developed using 4-chloro-1-naphthol and hydrogen peroxide.

6.2.3 RNA Analysis

Enzymatic digestion of RNA to nucleosides and subsequent HPLC/UV/MS analyses were performed as outlined in Chapter 5.
6.3 Results

Culturing of normal and heat shocked cultures of *E. coli* were done in parallel and the harvested cells were treated the same during the multi-step isolation of RNA and analysis of nucleosides present (Figure 6.1). Separate cultures were used for isolation of rRNA and tRNA.

![Experimental Outline](image)

Figure 6.1 Experimental Outline.

6.3.1 Growth of *E. coli* under normal and heat shock conditions

Heat shock in *E. coli* is induced by a temperature upshift from 30 °C to 42 °C. In this work heat shock was induced during log phase when the cells are growing rapidly. The maximum temperature for heat shock was chosen as 42 °C because growth rate decreases significantly at temperatures > 42 °C. [283]
Normally *E. coli* is grown at 37°C, the optimum growth temperature, where the numbers of cells are doubling in less than 20 min. Growth curves were constructed to define log and stationary phases in the 30 °C culture (Figure 6.2). Cell density was measured by monitoring the UV absorbance at 600 nm during culturing and then plotting absorbance vs. culture time to characterize the log phase and stationary phase for normal and heat shocked cultures. Half of the culture was transferred to 42 °C for 50 min, as marked by the arrow; the remaining culture is maintained at 30°C. As shown in Figure 6.3, the number of cells decrease slightly and the heat shock culture appears to enter the slow-growth stationary phase slightly earlier than the normal cultures at 30 °C. Also noted during culturing, was a reduction in gas bubbles present in the heat shocked liquid culture as compared to the cultures maintained at 30 °C. This effect is likely due to the reduction of metabolic products such as CO₂ as the cellular growth rate decreases in cells under stress.

![Escherichia Coli Growth at 30°C](image)

**Figure 6.2**  Growth Curve for *E. coli* grown at 30 °C.
6.3.2 Confirmation of Heat Shock Response

DnaK is a 70 kDa major heat shock protein present in greater amounts under heat shock conditions than normal conditions and can be used to confirm induction of the heat shock response. With over 4000 proteins possibly present at any one time within an *E. coli* cell, identifying and detecting the upregulation of DnaK requires specialized analysis. This is demonstrated in the 1-D PAGE gel of *E. coli* cell lysate from both the heat shocked and normal cultures. Due to the presence of many proteins no differences in the two cultures are observed (Figure 6.4A). The specific detection of DnaK in the normal and heat shocked cultures was accomplished using western blotting. After separation of the cell lysate proteins an antibody specific to DnaK is used to detect only DnaK. DnaK is present in greater amount in the heat shocked cell lysate as shown by the darker bands in Figure 6.4B.
6.3.3 Isolation of rRNA and tRNA

The purity of the rRNA and tRNA isolated from cells was checked by gel electrophoresis. The isolated rRNA appears intact and not degraded from both normal and heat shocked cultures. (Figure 6.5A) A small amount of 23S is present and was subsequently removed using anion exchange prior to enzymatic digestion to nucleosides. The total tRNA isolates after anion exchange treatment are shown in Figure 5B and also appear intact and pure. Heat shock treatment does not appear to alter the quality of RNA isolated.
6.3.4 LC/MS Analysis of rRNA and tRNA Modifications

The analysis of modified nucleosides by LC/MS reveal no differences in the identity of modifications present in rRNA or tRNA hydrolysates from normal and heat shocked cell cultures. (Figures 6.6 and 6.7) The nucleosides present in the 16S rRNA hydrolysates of both normal and heat shocked cells are the seven modifications known to be present in 16S rRNA from _E. coli_. The relative amounts of these modified nucleosides present in 16S rRNA from four different cultures were calculated using dimethyladenosine as an internal standard, as outlined in Chapter 5. (Table 6.1) For each separate culture, four replicate LC/MS analyses were performed. The means of these relative amounts from heat shocked and normal cultures were compared and a two-sample unpaired t-test was performed that does not assume equal variances. The resulting P values are reported (Table 6.1). Since p > 0.05 for all nucleosides compared, there is no significant difference between these means at the 95% confidence level. A graphical comparison of these means with the associated measurement variability is presented in the box plots shown in Figure 6.8.
Transfer RNA from both normal and heat shocked cultures contained the same 24 modified nucleosides as presented in Chapter 5. (Figure 6.7) Relative amounts of modified nucleosides present in both culture types were determined as outlined in Chapter 5. Only the modifications that were adequately chromatographically resolved to obtain UV peak area measurements are shown in Table 6.2. The mean relative amounts of modified nucleosides present in three different cultures of normal and heat shocked cultures were compared and the t-test was performed assuming unequal variances. The p values were > 0.05 for all tRNA nucleosides except \(N^6\)-isopentenyladenosine (\(i^6A\)) and 2-methylthio-\(N^6\)-isopentenyladenosine (\(ms^2i^6A\)). This result indicates that there exists a significant difference in the relative amounts of these two nucleosides present in the normal and heat shocked cultures but no significant differences in the relative amounts of the other subset of nucleosides determined. Comparison of mean relative amounts of these tRNA modified nucleosides are graphically represented in the box plots shown in Figure 6.9.

![UV Chromatogram from LC/MS analysis of an enzymatic digest of E.coli 16S rRNA. 1, pseudouridine (Ψ). 2, cytidine (C). 3, uridine (U). 4, 5-methylcytidine (m5C). 5, guanosine (G). 6, 7-methylguanosine (m7G). 7, 3-methyluridine (m3U). 8, N4, 2-O-methylcytidine (m4Cm). 9, N2-methylguanosine (m2G). 10, adenosine (A). 11, N6, N6-dimethyladenosine (m6_2A).](image-url)
Figure 6.7 UV Chromatogram from LC/MS analysis of an enzymatic digest of E. coli tRNA. 1, Ψ. 2, C. 3, acp3U. 4, U. 5, s2C. 6, Cm. 7, m5U. 8, G. 9, m7G. 10, Gm. 11, m1G. 12, acâC. 13, Q. 14, A. 15, t6A. 16, m6A. 17, m5A. 18, m6A. 19, i6A. 20, ms2i6A.

Table 6.1 Comparison of relative amounts of modified nucleosides in 16S RNA in four different cell cultures.

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Culture A</th>
<th>Culture B</th>
<th>Culture C</th>
<th>Culture D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NHS</td>
<td>HS</td>
<td>P value</td>
<td>NHS</td>
</tr>
<tr>
<td>Ψ</td>
<td>0.50</td>
<td>0.47</td>
<td>0.07</td>
<td>0.46</td>
</tr>
<tr>
<td>m2C</td>
<td>0.55</td>
<td>0.54</td>
<td>0.82</td>
<td>0.54</td>
</tr>
<tr>
<td>m2G</td>
<td>0.76</td>
<td>NR</td>
<td>NR</td>
<td>0.73</td>
</tr>
<tr>
<td>m2U</td>
<td>0.55</td>
<td>0.53</td>
<td>0.30</td>
<td>0.56</td>
</tr>
<tr>
<td>m4Cm</td>
<td>0.51</td>
<td>0.56</td>
<td>0.08</td>
<td>0.52</td>
</tr>
<tr>
<td>m2G</td>
<td>1.93</td>
<td>1.78</td>
<td>0.30</td>
<td>1.92</td>
</tr>
</tbody>
</table>
Figure 6.8. Comparison of average relative amounts of modified nucleosides present in 16S rRNA from heat shocked and normal cell cultures.
Table 6.2  Comparison of relative amounts of modified nucleosides in unfractionated tRNA in three different cell cultures.

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Culture 1</th>
<th>Culture 2</th>
<th>Culture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NHS</td>
<td>HS</td>
<td>P value</td>
</tr>
<tr>
<td>Ψ</td>
<td>5.77</td>
<td>6.15</td>
<td>0.25</td>
</tr>
<tr>
<td>s²C</td>
<td>0.85</td>
<td>1.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Cm</td>
<td>0.19</td>
<td>0.19</td>
<td>0.67</td>
</tr>
<tr>
<td>m²U</td>
<td>3.65</td>
<td>3.93</td>
<td>0.25</td>
</tr>
<tr>
<td>m²G</td>
<td>3.47</td>
<td>3.27</td>
<td>0.39</td>
</tr>
<tr>
<td>m¹G</td>
<td>0.57</td>
<td>0.67</td>
<td>0.10</td>
</tr>
<tr>
<td>ac⁴C</td>
<td>0.07</td>
<td>0.07</td>
<td>0.23</td>
</tr>
<tr>
<td>Q</td>
<td>0.58</td>
<td>0.56</td>
<td>0.44</td>
</tr>
<tr>
<td>m⁶t⁶A</td>
<td>0.11</td>
<td>0.11</td>
<td>0.99</td>
</tr>
<tr>
<td>m⁶A</td>
<td>1.27</td>
<td>1.42</td>
<td>0.06</td>
</tr>
<tr>
<td>m⁶A</td>
<td>0.42</td>
<td>0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>i⁶A</td>
<td>0.008</td>
<td>0.012</td>
<td>0.03</td>
</tr>
<tr>
<td>ms²i⁶A</td>
<td>0.480</td>
<td>0.409</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Figure 6.9. Comparison of average relative amounts of modified nucleosides present in tRNA from heat shocked and normal cell cultures.
Figure 6.9, (cont.). Comparison of average relative amounts of modified nucleosides present in tRNA from heat shocked and normal cell cultures.
6.4 Discussion

6.4.1 Comparison of rRNA modification in Normal and Heat Shocked Cell Cultures

Due to difficulties in the analysis of large RNAs, little definitive information is available on the functional role of these modifications, especially at the global level. Complete modification maps of rRNA are known for only a few organisms. Two lines of evidence suggest modifications may serve to enhance structural stability of rRNA, especially at higher temperatures. First, the general level of modification is higher in thermophilic than mesophilic bacteria [77, 79] and highest in hyperthermophilic organisms. In the archaeal thermophile, *Sulfolobus solfataricus*, 38 sites are modified in 16S compared to the 11 in *E. coli*. [57] Secondly, a general upward trend was observed in rRNA modification levels with culture temperature but changes in modification level reported were of uncertain significance. Additionally, at least one rRNA modifying enzyme, FtsJ, has been identified as a heat shock protein indicating it is upregulated during cellular stress.

There were no differences found in the relative amounts of nucleosides present in *E. coli* 16S. The modifications present, except for pseudouridine, were simple methylations that are typically found in rRNA. Methyltransferases can sometimes exhibit broad specificity, so it was worthwhile to characterize the differences in modification level in 16S rRNA. Analysis of 23S rRNA was not included in these experiments, but would be an interesting future experiment since FtsJ methylates uridine at position 2552 in 23S rRNA.

6.4.2 Comparison of tRNA modification in Normal and Heat Shocked Cell Cultures

In the subset of tRNA modifications compared in this study, significant differences in relative amounts of modified nucleosides were found only for \( \text{i}^* \text{A} \) and \( \text{ms} \text{i}^* \text{A} \). The relative amount of \( \text{i}^* \text{A} \) present increased < 10% in the heat shocked cultures as compared to the normal cultures. A decrease in the
methylthiolated ms\textsuperscript{2}\textit{i}\textsuperscript{6}A of approximately 20% was found to occur in the heat shocked cultures. The biosynthesis of these modifications is known to occur in a stepwise manner as discussed in Chapter 4. An accumulation of i\textit{i}\textsuperscript{6}A may occur if a deficiency in methylthiolation of i\textit{i}\textsuperscript{6}A exists. However, the decrease in ms\textsuperscript{2}\textit{i}\textsuperscript{6}A here is larger than the corresponding increase in i\textit{i}\textsuperscript{6}A suggesting there may be an additional explanation beyond undermodification of ms\textsuperscript{2}\textit{i}\textsuperscript{6}A. Alternatively, a decrease in the absolute levels of i\textit{i}\textsuperscript{6}A available for methylthiolation could ultimately result in a decrease in ms\textsuperscript{2}\textit{i}\textsuperscript{6}A levels.

Within the context of the results in this experiment, two possible explanations can be considered. First, the difference in modification level of i\textit{i}\textsuperscript{6}A and ms\textsuperscript{2}\textit{i}\textsuperscript{6}A are due to growth differences in heat shock and normal cultures (Figure 6.3). Cell growth appears to decrease and the heat shocked culture appears to enter stationary phase slightly earlier when compared to the normal cell culture. Because in \textit{E. coli} the majority of hydrophobic modifications such as ms\textsuperscript{2}\textit{i}\textsuperscript{6}A are present at all growth stages, these growth differences unlikely account for the differences in ms\textsuperscript{2}\textit{i}\textsuperscript{6}A modification levels. This result is in stark contrast to what is observed for \textit{B. subtilis} where i\textit{i}\textsuperscript{6}A predominates during exponential phase and ms\textsuperscript{2}\textit{i}\textsuperscript{6}A predominates during stationary phase. However, in \textit{B. subtilis} this growth phase transition is also related to the sporulation process, which is also under control of an alternate sigma factor.

A second explanation for differences in RNA modification levels observed is a regulation or control mechanism exists initiated by the heat shock response resulting in modification differences. Two modifying enzymes that may be involved in the heat shock response are MiaA and IscS. MiaA is part of a superoperon with heat shock promoters and is essential for growth at high temperatures. [280] Both miaA and FtsJ are known products of the sigma 32 heat shock response regulon. [279] Both MiaA and IscS are tRNA modifying enzymes identified as heat shock proteins under the control of the sigma 32 operon. [279] Few studies have addressed the putative link between heat shock and the modifying enzyme MiaA.
IscS is a cysteine desulfurase required for formation of all tRNA thiolated nucleosides. [153] Isc is part of the isc operon that includes several conserved genes, *iscR, iscS, iscU, iscA, hscA, hscB* and *fdx*. In the formation of s\(^4\)U and mnm\(^5\)s\(^2\)U, the transfer of sulfur from IscS involves several proteins but no FeS cluster proteins are involved. FeS cluster proteins which depend on IscS for formation and maintainence are required for the formation of s\(^2\)C and ms\(^2\)i\(^6\)A (ms\(^2\)i\(^6\)A). HscA and HscB are heat shock proteins associated with IscS directed FeS cluster assembly. [284] These heat shock proteins may be important in formation of s\(^2\)C and ms\(^2\)i\(^6\)A due to the involvement of FeS cluster containing proteins in the synthesis of these modified nucleosides. Iron-sulfur proteins exhibit unique redox and catalytic characteristics that could be important in dynamic control of RNA modifications.

The methylthiotransferase and radical SAM protein, MiaB, can alone form ms\(^2\)i\(^6\)A from i\(^6\)A, but with reduced efficiency. Another more efficient way to form ms\(^2\)i\(^6\)A involves a Fe-S cluster protein and depends on IscS. The sulfur is transferred from IscS to IscU; proteins IscA, HscA, HscB and/or Fdx participate in an unknown way to form Fe-S cluster loaded IscU which transfers the Fe-S cluster. [285]

Whether the differences in modification levels can be attributed to growth or heat shock effects is not clear. Additionally, a link has been suggested between growth phase and activation of the sigma 32 regulon. [286] The sigma 32 regulon is highly responsive to growth phase. The sigma 32 regulon is transiently activated during rapid cell growth (after exit from stationary phase), i.e., synthesis of heat shock proteins increases. Sigma 32 is inactivated when batch cultures enter stationary phase. Sigma S which governs transcription of genes during stationary phase increases as sigma 32 decreases.

**6.5 Conclusions**

Experiments conducted here outline an approach that can be used to compare modification status under different cellular conditions which can be used to further elucidate cellular pathways and mechanisms involved in the dynamic regulation of RNA modification. Although no significant differences were found in modification status of the 16S rRNA under heat shock conditions, a similar
analysis of 23S rRNA should be performed in future studies. The 23 S rRNA methyltransferase, FtsJ, has been identified as a heat shock protein, as noted in Chapter 1, and differences in methylations may be present under heat shock conditions.

Slight differences in growth rate were found to occur after heat shock in these experiments, which made it difficult to know whether differences in tRNA modification are due to differences in growth or are attributed to the heat shock response. It may be worthwhile to repeat this experiment and induce heat shock during stationary phase instead of during exponential growth.
Chapter 7  Conclusions and Future Work

7.1 Conclusions

In this dissertation I have optimized LC/UV/MS methods to identify and quantify modified nucleosides in RNA. I focused on the RNA modified nucleosides involved in protein translation. These methods can also be used or adapted for the analysis of other types of modifications to nucleic acids, such as xenobiotic modifications. One of the most important contributions made by this dissertation is the determination of the reproducibility of a LC/UV/MS method used to isolate, identify and quantitatively characterize modified nucleosides in tRNA and rRNA. This study shows the LC/UV/MS method is sufficient to use for characterizing changes in the abundance of modified nucleosides, as a function of varying cellular conditions. I also demonstrated how this LC/UV/MS method can be used to further understand the functional significance and cellular dynamics of RNA modification by applying these methods as an assay for functional characterization of modifying enzymes and for comparison of modifications under different cellular conditions. Also included in this dissertation is a detailed discussion based on a thorough review of the literature in how the nucleosides’ individual structural and chemical properties contribute to the tertiary structure and decoding in tRNA. This review, within Chapter 2, is a good introduction and resource for future researchers interested in developing and applying LC/MS methods to explore how modifications change dynamically at the cellular level.

This dissertation begins with a survey and comparison of RNA modifications present in unfractionated tRNA from bacterial and archaeal organisms. Although much is known about the identity of tRNA modifications and modifying enzymes for model organisms, only a few archaea have been studied. A census of modifications for several organisms presented here, including B. subtilis, L. lactis, T. thermophilus, H. marismortui and H. walsbyi, had not been previously reported. Studying modification differences among organisms, especially extremophiles, that live in harsh conditions, provides clues to how modifications function to enhance structural stability of RNA. The decoding strategies also differ
among kingdoms and it is likely, the cellular and functional dynamics of RNA modification will also vary among kingdoms.

A specific example of the differences in decoding strategies between bacteria and archaea is highlighted in Chapter 3. It was indeed, the differences in decoding properties of a single tRNA that led our collaborators to suspect differences in nucleoside modification were responsible for decoding of the AUA isoleucine codon. The sensitivity and selectivity of mass spectrometry methods enabled the structural identification of a new nucleoside, agmatidine. This nucleoside modification is responsible for specific decoding properties of a rare tRNA and present in very low abundance. Over 109 modified nucleosides have been structurally identified in RNA and there are likely other unknown modifications also present in very low abundance or with unique structures, that elude detection by present methods, especially in mixtures of unfractionated tRNA. Current mass spectrometry techniques are quite capable of achieving the necessary sensitivity and selectivity for nucleoside identification; the challenge is in the separation and isolation of the tRNA. Improvements in methods for isolation of individual tRNAs or improvements in chromatographic resolution of nucleosides would further the ability of these methods to identify new nucleosides.

Chapter 4 demonstrates how LC/UV/MS can be used as an assay for the presence of particular modifying enzymes and how this can be used to functionally characterize specific modifying enzymes. Modifying enzymes can have dual specificity and modify more than one type of RNA. Some enzymes are multi-site specific and place modifications at several different locations. The ability to quantitatively characterize differences in several different modifications, simultaneously, in mutant strains can be used to demonstrate the synergistic nature of modifications and modifying enzymes.

The focus of Chapter 5 was the determination of the reproducibility or precision of the LC/UV/MS method with an overall goal to detect statistically significant changes in rRNA or tRNA nucleoside modification when comparing two different cellular conditions. The method precision reported for this
method is comparable to similar LC/MS analyses of modified nucleosides in urine. Although the accuracy of the method was not determined, the relative amounts of modified nucleosides reported for rRNA were consistent with the known number of modified nucleosides in a stoichiometrically modified 16S rRNA. Through these experiments it became apparent a solid phase purification method was more appropriate for the analysis of rRNA and tRNA nucleosides.

Now that the precision of the method is known, the appropriate number of biological replicates can be chosen, which will enable detection of statistically significant differences at a given confidence level. This approach can be used in comparative analyses of biological samples.

The methods optimized and characterized in Chapter 5 were used to compare the global modification status of tRNA and rRNA under normal and heat shock conditions in Chapter 6. Statistically significant differences were not found in the comparisons of relative nucleoside amounts of 16S rRNA in normal and heat shocked cell cultures. Differences were noted in the relative amounts of i6A and ms2i6A in tRNA from normal cultures compared to heat shocked cell cultures. Determination of whether these differences in tRNA modifications were due to changes in growth rate, which occurred after onset of heat shock, or a function of the heat shock response itself, cannot be determined until further experiments are done.

7.2 Future Work

One limitation in quantitation of modified nucleosides in enzymatic digests is the ability to completely resolve many similar nucleosides, which is especially challenging in the analysis of unfractionated tRNA where more than 30 different modified nucleosides are present. Although the MS methods of detection do not necessarily require complete separation of nucleosides for accurate identification, MS detection of low abundant nucleosides eluting with the major nucleosides is problematic. The LC/UV method for quantitation relies on adequate separation for precise quantitation of nucleosides present. One focus of
future work in the area of nucleoside analysis should be on increasing the peak capacity in the LC separation.

In much of the previous literature that describes the analysis of nucleosides by HPLC, a Supelco C18 nucleoside column is used. This column was also used in the research described in this dissertation. Many nucleosides are very polar and for adequate reverse phase separation, mobile phases with high aqueous content are required. One of the drawbacks in using this column is its susceptibility to phase collapse, which can occur with C18 columns used under high aqueous conditions. As the column ages, the phase may be partially collapsing and this will affect resolution and reproducibility.

Ultra high performance liquid chromatography (UPLC) can be used to increase resolution and peak capacity over normal HPLC. This recently developed technique uses smaller size column particles capable of withstanding high pressure to achieve increases in chromatographic resolution, sensitivity and speed. To increase peak capacity and minimize phase collapse problems, I suggest the use of UPLC with a column designed for high aqueous conditions in future nucleoside chromatographic separations. Preliminary experiments with a Thermo Hypersil Gold column (150 x 2.1 mm, 1.9 micron particles) demonstrate good separation of nucleosides at operating pressures less than 5000 psi at 300 uL/min flow rate using an ammonium acetate/ acetonitrile mobile phase gradient similar to that used in the HPLC method used here.

Analytically, the LC/MS methods for quantitation of modified nucleosides can be improved by advances in either the chromatography or the mass spectrometry, or both. As discussed in Chapter 1, nucleosides fragment easily at the glycosidic bond with loss of the ribose moiety producing a protonated base ion fragment. In tandem MS/MS experiments, the molecular ion to base ion transition can be specifically monitored and measured in selected reaction monitoring (SRM) scanning mode. Constant neutral loss (CNL) scanning techniques can be used to selectively monitor for the loss of ribose (132 or 146 u) for identification of modified nucleosides. These techniques enhance the selectivity and lower the
limit of quantitation for specific nucleosides. These techniques can be used to extend the dynamic range of nucleoside analysis at the lower end.

Advances in our understanding of RNA modification can only come as quickly as methods of analysis are developed, particularly, quantitative methods. The methods presented in this dissertation can be used for quantitative comparison of changes in modified nucleoside composition and can be used in biological investigations for elucidation of the dynamic nature of nucleoside modification and the apparent synergy among modifications.
Bibliography


220. Awai, T., et al., Aquifex aeolicus tRNA (N2,N2-guanine)-dimethyltransferase (Trm1) catalyzes transfer of methyl groups not only to guanine 26 but also to guanine 27 in tRNA. J Biol Chem, 2009. 284(31): p. 20467-78.


