ABSTRACT

GUSTILO, ESTELLA MAILUM. Post-transcriptional Modifications of the tRNA Anticodon Stem and Loop (ASL) Affect the Ability of tRNA to Bind Synonymous Codons. (Under the direction of Dr. Paul F. Agris).

The Genetic Code is arranged into sixteen codon boxes, where the four codons in each box are similar in their first two letters but differ at the third position (the wobble position). In the universal Genetic Code, each amino acid, except for Tryptophan and Methionine that have one codon each, is encoded by two to six codons (two to six-fold degenerate). There are fewer tRNA species than codons; therefore, a tRNA species can recognize more than one codon. This flexibility in recognition resides in the third position of the codon:anticodon pairing, the wobble position.

Codon recognition by tRNA is dependent on the anticodon loop. The sequence of the anticodon (tRNA positions 34, 35, and 36) does not necessarily predict codon binding according to Watson-Crick rules. In all organisms, post-transcriptional modifications occur quite extensively and of great variety at the anticodon loop. These modifications, usually found on the nucleosides in tRNA position 34 (the wobble position) and position 37, direct the tRNA’s ability to read codons accurately and efficiently. Just as the types of modifications are diverse, the abilities of modifications to recognize codons also vary. When a particular amino acid is encoded by an entire codon box, such as four-fold degenerate Valine, we show that a tRNA species with a specific modification can read all four codons of that box. The modification 5-oxyacetic acid (cmo5) at the wobble position of tRNAVal (tRNAVal-cmo5U34) allows cmo5U34 to recognize U, A, G, and perhaps C. In instances where the difference between the codes of two amino acids resides only in the third letter of their codons (2-fold degenerate codons in a split box), modifications at the wobble position of the
anticodon restrict codon recognition to the two codons specific for that 2-fold degenerate amino acid. For example, Lysine has two codons (AAA and AAG) that share a codon box with Asparagine codons (AAU and AAC). The modified nucleoside 5-methoxycarbonylmethyl-2-thiouridine at the wobble position of human tRNA (tRNA^{Lys mcm^5s^2U_{34}}) confers this tRNA’s ability to restrict codon recognition to the two Lysine codons only.

Similar to the cytoplasmic tRNAs, mitochondrial tRNAs also contain post-transcriptional modifications. The mitochondria deviate from the universal Genetic Code in that it uses the universal Isoleucine codon AUA to decode Methionine. In all organisms, there are two Methionine tRNAs: an initiator tRNA^{Met} and an elongator tRNA^{Met}. Mitochondria, however, have but one tRNA^{Met} that acts as both initiator and elongator, has characteristics of both types of tRNA^{Met}, and decodes AUG and AUA in the aminoacyl-(entry or A)-site and the peptidyl (P)-site of the ribosome. The human mitochondrial tRNA^{Met} is modified with a 5-formyl-group at the wobble position cytidine-34 (hmtRNA^{Met-f^5C_{34}}). This modification allows the hmtRNA^{Met-f^5C_{34}} to expand codon recognition to include AUA in translating Methionine. At times, the mitochondria use Isoleucine codons AUU and AUC to initiate translation. Surprisingly, the 5-formyl modification of hmtRNA^{Met-f^5C_{34}} also allows codon-reading expansion at the P-site to include the entire codon box AUN.
Post-transcriptional Modifications of the tRNA Anticodon Stem and Loop (ASL) Affect the Ability of tRNA to Bind Synonymous Codons

by
Estella M. Gustilo

A dissertation submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy

Molecular and Structural Biochemistry

Raleigh, North Carolina
June 5, 2009

APPROVED BY:

________________________________  ______________________________
Dr. Paul F. Agris     Dr. E. Stuart Maxwell
Committee Chair

_______________________________  ______________________________
Dr. Linda Spremulli     Dr. Paul Wollenzien
DEDICATION

To my father who showed me the magic in books and the wonders of learning. To my mother who expressed unconditional love. To my sister Marietta who offered unconditional friendship. The beauty in my life I owe to you, my dear family.
BIOGRAPHY

Estella M. Gustilo was born in the Philippines and immigrated to the United States with her family at a young age. She grew up in Charleston, SC, where she attended college. Upon graduation from college she worked in molecular biology research, where she found her calling. She then attended graduate school in Raleigh, North Carolina.
ACKNOWLEDGMENTS

The process of earning my Ph.D. has blessed me with wonderful opportunities of learning. I will be forever grateful for those who contributed to my growth during these years. First, I extend my deepest thanks to my mentor, Dr. Paul F. Agris, whose brilliance and dedication to his craft has been inspiring. Dr. Agris has shown me amazing possibilities. I thank each of the other members of my graduate committee; Dr. E. Stuart Maxwell, Dr. Linda Spremulli, and Dr. Paul Wollenzien have given me valuable advice, collaborated in these studies, and expanded my horizons.

I acknowledge my fellow lab members, especially William Darnell Graham, Dr. Franck A. P. Vendeix, and William A. Cantara for providing me with dependable support. My fellow biochemistry graduate students made this experience so much fun. Dr. Tatjana Shapkina and Dr. Keith Gagnon have shared technical knowledge that has helped make this study become more successful. Dr. Simpson Joseph has been extremely kind and opened his lab and resources to me for several weeks to learn precious methods used in this work. This study wouldn’t have been possible without technical training provided by Dr. Steven S. Phelps. Steve has been such a wonderful, supportive friend who has taught me so much.

As always, my close family and friends have been instrumental in all the good things that have come from my life. Even as science fills my mind, they fill my heart. My parents and my sister Marietta have never failed me. Finally, but not least, I must thank the Creator, whose mind I’ve seen only glimpses of in my study of science but whose heart I know fully. Thank you all.
# TABLE OF CONTENTS

**LIST OF TABLES** ........................................................................................................ vii

**LIST OF FIGURES** ........................................................................................................ viii

**CHAPTER 1. Post-transcriptional modifications are a trademark of tRNA**........ 1  
1.1 The roles of transfer ribonucleic Acid (tRNA) and the significance of post-transcriptional modifications to tRNA functions ......................................................... 1  
1.2 The Genetic Code .................................................................................................. 7  
1.3 The effects of post-transcriptional modifications of tRNA on synonymous-codon reading ................................................................................................................. 9  
REFERENCES ............................................................................................................. 16

**CHAPTER 2. Modifications of the anticodon stem and loop of human tRNA\textsuperscript{\text{Lys}} (ASL\textsubscript{\text{Lys}}\text{UUU}) restrict codon recognition to Lysine codons AAA and AAG** ...... 24  
2.1 Introduction ............................................................................................................ 24  
2.2 Materials and Methods .......................................................................................... 26  
2.3 Results .................................................................................................................... 28  
2.4 Conclusions .......................................................................................................... 28  
REFERENCES ............................................................................................................. 32

**CHAPTER 3. Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding** ................................................................................. 35  
ABSTRACT ..................................................................................................................... 36  
INTRODUCTION .......................................................................................................... 39  
EXPERIMENTAL PROCEDURE .................................................................................. 42  
RESULTS ....................................................................................................................... 47  
DISCUSSION .................................................................................................................. 62  
CONCLUSION .............................................................................................................. 70  
ACKNOWLEDGEMENTS ............................................................................................. 71  
SUPPORTING INFORMATION AVAILABLE ................................................................. 71  
REFERENCES ............................................................................................................. 72

**CHAPTER 4. Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and of the human mitochondrial tRNA\textsuperscript{Met}** ................................................................................. 94
LIST OF TABLES

CHAPTER 1

Table 1. Distribution of modified nucleosides.......................................................... 11
Table 2. The recognition of tRNA wobble position-34 for the wobble position (N3) of the
codon.......................................................................................................................... 15

CHAPTER 3

Table 1. Thermal Parameters of Unmodified and Modified ASL Val3 UAC.................. 77
Table 2. Affinity of Unmodified and Modified ASL Val3 UAC for the Valine Codons .... 77

SUPPLEMENTARY MATERIALS:

Table 1. Local base step parameters of the unmodified ASL Val3 UAC...................... 93

CHAPTER 4

Table 1. Thermodynamic contributions of f5C34..................................................... 125

CHAPTER 5

Table 1. Affinity of Unmodified and f5C34-Modified hmtASL Met CAU-Ψ27 for the
Mitochondrial Methionine Codons at E. coli Ribosomal P or A-site......................... 161

CHAPTER 6

Table 1. Contribution of modifications to codon recognition, frameshifting, and as
aminoacyl-tRNA synthetase identity determinants.................................................. 181
LIST OF FIGURES

CHAPTER 1

Figure 1. Bacterial tRNAs on the ribosome during translation .......................... 2
Figure 2. The tRNA ......................................................................................... 3
Figure 3. The Universal Genetic Code ............................................................ 8

CHAPTER 2

Figure 1. The anticodon stem and loop of human tRNA\textsuperscript{Lys3}(\textit{ASL\textsuperscript{Lys3} mcm\textsuperscript{s2}U\textsubscript{34}, ms\textsuperscript{2}6A\textsubscript{37}}) .............................................................. 30
Figure 2. Ribosomal, equilibrium binding curves of the fully-modified human \textit{ASL\textsuperscript{Lys3 uuU-mcm\textsuperscript{s2}U\textsubscript{34}, ms\textsuperscript{2}6A\textsubscript{37}}} or the unmodified \textit{ASL\textsuperscript{Lys3 uuU}} ........................................ 31

CHAPTER 3

Figure 1. Nucleotide sequence of the \textit{Escherichia coli tRNA\textsuperscript{Val3 uuU} anticodon stem and loop (ASL\textsuperscript{Val3 uuU}) and its modified nucleosides} ......................................................... 82
Figure 2. UV-monitored, thermal transitions and circular dichroism spectra of unmodified and modified \textit{ASL\textsuperscript{Val3 uuU}} ................................................................. 83
Figure 3. Binding of unmodified and modified \textit{ASL\textsuperscript{Val3 uuU}} to valine codons at the A-site of \textit{E. coli} 30S ribosomal subunits ........................................................................ 84
Figure 4. Detection of the base-paired imino protons of the ASLs by NMR ............. 85
Figure 5. Superimposed 1H-1H COSY spectra of unmodified and modified \textit{ASL\textsuperscript{Val3 uuU}} ................................................................. 86
Figure 6. Determination of ribose conformation from the 1H DQF-COSY spectra .... 87
Figure 7. Solution structures of ASL$^{Val}_UAC$ -cmo$^5U_{34}$; m$^6A_{37}$ and ASL$^{Val}_UAC$ and the rmsd for each of their nucleotides. 88

Figure 8. Superimposition of the average structures of the loop residues of ASL$^{Val}_UAC$ -cmo$^5U_{34}$; m$^6A_{37}$ with that of ASL$^{Val}_UAC$. 89

Figure 9. Loop structure of ASL$^{Val}_UAC$-cmo$^5U_{34}$; m$^6A_{37}$ in solution compared to that of the crystallographic structure on the ribosome. 90

SUPPLEMENTARY MATERIALS:

Figure 1. Superposition of the $^1$H-$^3$1P HETCOR spectra of ASL$^{Val}_UAC$ and ASL$^{Val}_UAC$-cmo$^5U_{34}$; m$^6A_{37}$. 91

Figure 2. Anomeric to aromatic connectivity for ASL$^{Val}_UAC$ and ASL$^{Val}_UAC$-cmo$^5U_{34}$; m$^6A_{37}$. 92

CHAPTER 4

Figure 1. Human mitochondrial tRNA$^{Met}_{CAU}$. 126

Figure 2. Synthesis of the 5-formylcytidine phosphoramidite. 127

Figure 3. NMR spectra of the hmASL$^{Met}_{CAU}$. 128

Figure 4. HPLC nucleoside composition and NMR analyses of hmASL$^{Met}_{CAU}$ and f$^5$C. 129

Figure 5. Thermal denaturations and circular dichroism spectra of the hmASL$^{Met}_{CAU}$-Ψ$^{27}$ and the hmASL$^{Met}_{CAU}$-Ψ$^{27}$; f$^5$C$^{34}$. 130

Figure 6. Codon binding by hmASL$^{Met}_{CAU}$-Ψ$^{27}$ and the hmASL$^{Met}_{CAU}$-Ψ$^{27}$; f$^5$C$^{34}$. 131

Figure 7. Analysis of the pKa of cytidine in comparison to that of 5-formylcytidine. 132

Supplemental Figure. Heavy atoms root mean square deviation (r.m.s.d.) variations as a function of the molecular dynamics simulation time for the hmASL$^{Met}_{CAU}$-Ψ$^{27}$; f$^5$C$^{34}$ and for the hmASL$^{Met}_{CAU}$-Ψ$^{27}$. 133

CHAPTER 5

Figure 1. The mammalian mitochondrial Genetic Code. 165
CHAPTER 6

Figure 1. Accommodation of tRNA on the ribosome…………………………………….174

Figure 2. Modifications of the anticodon stem and loop (ASL) of tRNA order the loop, prevent frameshifting and allow accurate codon selection………………………………….176

Figure 3. The Universal Genetic Code…………………………………………………….179
CHAPTER 1

Post-transcriptional modifications are a trademark of tRNA

1.1 The roles of transfer ribonucleic acid (tRNA) and the significance of post-transcriptional modifications to tRNA functions.

Transfer ribonucleic acid (tRNA) is a crucial participant in protein biosynthesis. The tRNA’s primary role occurs on the ribosome during translation of genetic information, where the tRNA reads the codes for amino acids imbedded in messenger RNA (mRNA) and places the appropriate amino acid into a growing protein (Figure 1). The Genetic Code is the language of protein synthesis; it is stored in deoxyribonucleic acid (DNA), transcribed into mRNA, and translated into proteins in the ribosome by the tRNA. Therefore, the tRNA is a critical translator of the Genetic Code in gene expression. tRNA is the faithful keeper of the Genetic Code, keeping fidelity from the start of protein synthesis with aminoacylation to completion with the production of viable proteins. Inherent in its primary, secondary, and tertiary structure is information that gives the tRNA its ability to function (Figure 2). The tRNA’s function is dependent on its interactions with a variety of cellular molecules such as aminoacyl-tRNA synthetases (aaRSs), translation initiation factors, translation elongation factors, mRNAs, ribosomes, and peptidylhydrolases. Post-transcriptional modifications of tRNA nucleosides are a hallmark of the chemistry, structure, and function of tRNA [1-4].
Figure 1. **Bacterial tRNAs on the ribosome during translation.** During initiation, the 30S ribosomal subunit (transparent grey circle) binds the mRNA (light blue strip) and the initiator tRNA at the ribosomal peptidyl (P)-site with the aid of initiation factors (not shown). The 50S subunit (dark grey) then joins the complex to form the complete 70S ribosome. Amino acids are added to the nascent protein during elongation, where an elongator aminoacyl-tRNA binds its codon on the mRNA at the aminoacyl (A)-site with the aid of elongation factors. A peptide bond is formed between the amino acid bound to the tRNA at the P-site and the amino acid of the tRNA at the A-site, and the entire protein chain is transferred to the tRNA at the A-site. The P-site tRNA then moves to the exit (E)-site (not shown), and the tRNA with the growing protein at the A-site is translocated to the P-site. Another elongator aminoacyl-tRNA binds its codon at the A-site until the termination stage, where release factors terminate protein synthesis. Accuracy and efficiency are achieved when the appropriate tRNAs bind their corresponding codon.
Figure 2. The tRNA. A. The secondary cloverleaf structure of tRNA. B. The tertiary L-shaped structure of tRNA. The anticodon triplet consists of positions 34, 35, and 36. The tRNA consists of four main domains: the dihydrouridine stem and loop (DSL), the anticodon stem and loop (ASL), the ribothymidine (T), pseudouridine (Ψ), cytidine (C) stem and loop (TSL). The tRNA also consists of a segment of a variable number of nucleosides (VL). Positions 34 and 37 are often post-transcriptionally modified.
Approximately 100 different modified nucleosides have been found in RNA [5]. tRNA is the most highly modified nucleic acid in the cell and consists of the greatest variety of modification chemistries [6]. Post-transcriptional modifications decorate nucleosides all along the tRNA and are integral to its purpose [1-4]. Up to thirty percent of tRNA nucleosides at more than sixty different positions are modified [6]. Modified nucleosides are derivatives of the four ribonucleic acid monomers Adenosine (A), Guanosine (G), Uridine (U), and Cytidine (C), and therefore add additional chemistries to the limited number of only four main nucleosides. The significance of modified nucleosides in the tRNA is evident in their conservation among species and the amount of genetic material assigned to their synthesis. In bacteria, approximately one percent of the genome codes for tRNA modifying enzymes - this is four times more than the approximately 0.25 percent of bacterial genes designated to tRNA [7]. Therefore, more genetic information is allocated to the modifying of tRNA than to the tRNA themselves [7]. Modifications help the tRNA fold into its secondary cloverleaf structure and tertiary L-shaped conformation mandatory for function and may aid in the tRNA’s many interactions with proteins [2, 8-12]. Modifications are important to the stability of tRNA and thus, affect its half-life [13-14].

Aminoacylation of the tRNA is the first step of protein synthesis. Aminoacyl-tRNA synthetase charges the tRNA with the correct amino acid. Fidelity of translation begins at aminoacylation. Modifications on the tRNA are specific identity determinants by some aminoacyl-tRNA synthetases and therefore important to the aminoacylation of these tRNAs [15-19]. Modifications may also affect aminoacylation via the modifications’ contributions to the tRNA’s proper conformation [20-21]. Although most in-vitro transcribed tRNAs (thus
can be aminoacylated, there are small differences in kinetics, suggesting that modified nucleosides can be of some influence [11]. This is likely due to the inability of tRNA to fold into its functional conformation. Unmodified tRNAs also have a greater probability of being mischarged by a noncognate aminoacyl-tRNA synthetase [21, 22].

In the succeeding steps to the production of functional proteins, tRNA interacts with many proteins such as initiation factors or elongation factors, and the proteins of the ribosome. In the ribosome, it decodes the sequence of mRNAs into sequences of protein. The tRNA is not a passive adaptor in translation but an active player during its function by undergoing conformational changes [23-25]. Being the linker between genetic language and end product (protein), the tRNA’s accuracy is necessary if the cell is to produce functional proteins. And this fidelity must be balanced with the high speed of translation (over 20 peptide bonds per second). Post-transcriptional modifications at or near the anticodon of tRNA are often required for accurate and efficient translation of the codons of mRNA [1, 11, 25-35].

The tRNA is an ancient molecule that is a remnant of the RNA World, where cellular processes were carried out solely by RNA [36]. Thus, the tRNA has evolved into a multifunctional molecule in the cell and is involved in a variety of processes in the cell. While its main job is the adaptor molecule in protein synthesis, the tRNA has been found to have other roles outside of translation on the ribosome [37]. The tRNA has been found to function in viral proliferation [38-40], gene expression regulation [41, 42], cell division/DNA replication [42-45], cell wall regulation [46, 47], protein degradation [48, 49], and
chlorophyll synthesis [50]. Post-transcriptional modifications of tRNA may be directly or indirectly involved in the functions of tRNA [11, 51].

The tRNA has been shown to be an important player in retroviral replication [38-40]. A specific tRNA isoacceptor, tRNA\textsuperscript{Lys3}, is recruited by infected cells as the primer for the reverse transcription of human immunodeficiency virus (HIV-1) [39, 52-55]. This has placed specific tRNAs in study as potential drug targets [38]. The tRNA\textsuperscript{Lys3} consists of specific modifications at the anticodon loop [30]. These modifications play a role in HIV-1’s specificity for tRNA\textsuperscript{Lys3} and stabilize the tRNA\textsuperscript{Lys3}’s interaction with the HIV-1 genome [56-58].

The tRNA is also involved in DNA replication and/or cell division. Specific Serine and Arginine tRNA isoacceptors have been shown to affect the cell cycle [59, 60]. The modifications of tRNA may be involved in cell cycle regulation in that certain tRNA isoacceptors with specific modifications are apparent in certain cell cycle stages [61-63].

The degree of tRNA modifications’ importance in many cellular processes has yet to be fully appreciated. tRNA modifications are most studied in the context of tRNA’s central role of translation. The modifications at tRNA’s anticodon stem and loop (ASL) are especially important, as they have been shown to be significant factors in codon: anticodon pairing, reading frame maintenance, translocation of tRNA from the ribosomal aminoacyl (A)-site to the peptidyl (P)-site, and the balance between decoding accuracy and speed [64-66]. The tRNA position 34 (wobble position) and position 37 at the ASL are often modified and consist of the greatest variety of modifications of any position of any nucleic acid [1].
Modifications at these two specific positions affect tRNA’s main function of reading codons on the ribosome.

1.2 The Genetic Code

Proteins are produced by the cell using instructions compiled in the Genetic Code (Figure 3). The Genetic Code is stored in DNA, transcribed into mRNA, and translated on the ribosome by tRNA. The elucidation of the Genetic Code is arguably the greatest accomplishment in the biomedical research of the 1960s. Upon its discovery, the Genetic Code was considered universal in that there appeared to be little variation among organisms. The Genetic Code consists of 64 codons: 61 sense codons that code for any of the 20 amino acids and three termination codons. In the universal Genetic Code, most amino acids have two to six codons. The different codons that code for the same amino acid are synonymous codons. Only Tryptophan and Methionine have no synonymous codons, as these two amino acids only have one codon each.

The Genetic Code is arranged into sixteen sets of four codons, portrayed in a table of sixteen codon boxes (Figure 3). The four codes of a codon box are similar to each other in the first two letters and differ only in the third letter. Some amino acids are encoded in synonymous codons all residing in the same codon box. Other amino acids share the codes of a codon box; these codons are termed to be in a “split box.” For example, Valine is encoded by four codons that reside in the same codon box, while Asparagine and Lysine each...
have two codons that share a codon box. Thus, the codons of Asparagine and Lysine are very similar and differ only in the third position, the wobble position of the codon.

**Figure 3. The universal Genetic Code.** There are 61 sense codons that code for the 20 amino acids and 3 stop codons. The Genetic Code is arranged into 16 codon boxes. Each codon box consists of four codons that have the same first two letters. The four codons of a box differ only in their third letter (the wobble position).
The Genetic Code was once thought to be “frozen” in that it was considered universal among all protein-producing systems [67]. However, the sequencing of the human and bovine mitochondrial DNA in 1981 and 1982, respectively, has proven that the Genetic Code can deviate from the Code established from *Escherichia coli* in the 1960s [68, 69]. Soon after the sequencing of the mammalian mitochondrial DNA, other deviations from the universal Genetic Code were found [70-72].

In the ribosome, the tRNA must decipher the Genetic Code with great accuracy and efficiency. tRNA modifications at the anticodon loop have been shown to facilitate accurate and efficient reading of codons. ASL modifications aid in the recognition of synonymous codons by restricting codon recognition of split boxes or expanding codon recognition of whole boxes [1]. Certain ASL modifications also allow the reading of unconventional codons in the alteration of the universal Genetic Code [73]. Thus, modifications at the ASL modulate the tRNA’s decoding capacity.

1.3 The effects of tRNA post-transcriptional modifications on synonymous-codon recognition

The Genetic Code is redundant or degenerate in that there are 61 codons for only 20 amino acids. Codons also outnumber the tRNA species that decode them. In *E. coli*, there are 45 tRNA species [74] that must translate the 61 codes; therefore, on average, there are only some three tRNAs for four codons. Francis Crick’s Wobble Hypothesis explained that a tRNA can read more than one codon by practicing base-pairing flexibility at the third
position (the wobble position) of codon:anticodon binding [75]. Crick hypothesized that the first two positions of codon:anticodon pairings followed Watson-Crick nucleoside base pairing rules of Adenosine (A) pairing with Uridine (U) and Guanosine (G) pairing with Cytosine (C). However, the third position (wobble position) is more flexible in that a G can pair with its cognate C or “wobble” to a U. Also, Crick explained that Inosine (I), a modified A, can exhibit a great deal of flexibility by pairing with U, C, or A.

The forty-plus years of research since Crick’s hypothesis has revealed that, although Crick was fundamentally correct in his assumption of wobble-pairing flexibility, wobble-pairing can be more complicated than just G:U and Inosine pairings. Oftentimes, the wobble position of tRNA (position 34, first position of the anticodon) is modified, and these modifications are significant to tRNA’s ability to wobble to a particular codon (Table 1) [76]. tRNA position 37 is also often modified and aid in codon recognition (Table 1).

Modified nucleosides at the anticodon stem and loop (ASL) of tRNA often give the tRNA its ability to recognize synonymous codons. Some modifications at the ASL restrict codon recognition to a few synonymous codons while other modifications serve to extend codon recognition to include synonymous codons (Table 2). The ability of ASL modifications to restrict or expand codon reading is due to the extent of degeneracy of the codons. For example, Lysine (Lys) is two-fold degenerate in that it has two codons, AAA and AAG. The two Lys codons share a codon box with Asparagine (Asn). There are two Asn codons – AAU and AAC. Thus, Lys and Asn codons are similar, with the only difference being the third letter, the wobble position, of the codon.
<table>
<thead>
<tr>
<th>Name and no.</th>
<th>Symbol</th>
<th>Code</th>
<th>N. of iRNAs</th>
<th>Relative %</th>
<th>Absolute %</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosine</td>
<td>A</td>
<td></td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>B, O</td>
</tr>
<tr>
<td>Unknown A</td>
<td>H</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>m'H</td>
<td></td>
<td>13</td>
<td>3</td>
<td>2</td>
<td>B, E, O</td>
</tr>
<tr>
<td>3</td>
<td>m'A</td>
<td></td>
<td>24</td>
<td>6</td>
<td>4</td>
<td>B, O</td>
</tr>
<tr>
<td>4</td>
<td>m'A</td>
<td></td>
<td>29</td>
<td>7</td>
<td>5</td>
<td>B, E, O</td>
</tr>
<tr>
<td>5</td>
<td>m'U</td>
<td></td>
<td>94</td>
<td>6</td>
<td>4</td>
<td>B, F, O</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td></td>
<td>5</td>
<td>1</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>11</td>
<td>m'tU</td>
<td></td>
<td>114</td>
<td>29</td>
<td>21</td>
<td>A, B, E, O</td>
</tr>
<tr>
<td>12</td>
<td>E</td>
<td></td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>B, E, O</td>
</tr>
<tr>
<td>13</td>
<td>m'tA</td>
<td></td>
<td>3</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>B, E</td>
</tr>
<tr>
<td>14</td>
<td>m'tG</td>
<td></td>
<td>46</td>
<td>94</td>
<td>8</td>
<td>B, E, O</td>
</tr>
<tr>
<td>15</td>
<td>m'tC</td>
<td></td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>16</td>
<td>m'tU</td>
<td></td>
<td>101</td>
<td>100</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Guanosine</td>
<td>G</td>
<td></td>
<td>116</td>
<td>65</td>
<td>21</td>
<td>A, B, E, O</td>
</tr>
<tr>
<td>Unknown G</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>Gm</td>
<td></td>
<td>3</td>
<td>2</td>
<td>&lt;1</td>
<td>A, B</td>
</tr>
<tr>
<td>2</td>
<td>m'G</td>
<td></td>
<td>25</td>
<td>14</td>
<td>4</td>
<td>B, E</td>
</tr>
<tr>
<td>14</td>
<td>m'G</td>
<td></td>
<td>116</td>
<td>83</td>
<td>21</td>
<td>A, B, E, O</td>
</tr>
<tr>
<td>16</td>
<td>Gm</td>
<td></td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>17</td>
<td>G</td>
<td></td>
<td>11</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>m'G</td>
<td></td>
<td>20</td>
<td>12</td>
<td>4</td>
<td>B, E, O</td>
</tr>
<tr>
<td>20</td>
<td>m'G</td>
<td></td>
<td>4</td>
<td>7</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>70</td>
<td>m'G</td>
<td></td>
<td>149</td>
<td>100</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>152</td>
<td>100</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Cytidine</td>
<td>C</td>
<td></td>
<td>116</td>
<td>73</td>
<td>20</td>
<td>A, B, E, O</td>
</tr>
<tr>
<td>Unknown C</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>Cm</td>
<td></td>
<td>5</td>
<td>3</td>
<td>&lt;1</td>
<td>A, E</td>
</tr>
<tr>
<td>4</td>
<td>m'C</td>
<td></td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>A, B</td>
</tr>
<tr>
<td>6</td>
<td>m'C</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>f'C</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>B, E</td>
</tr>
<tr>
<td>10</td>
<td>F'cm</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>F'cG</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>152</td>
<td>100</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Uracil</td>
<td>U</td>
<td></td>
<td>56</td>
<td>36</td>
<td>10</td>
<td>A, B, E, O</td>
</tr>
<tr>
<td>Unknown U</td>
<td>N</td>
<td></td>
<td>46</td>
<td>29</td>
<td>8</td>
<td>A, B, E, O</td>
</tr>
<tr>
<td>1</td>
<td>Um</td>
<td></td>
<td>4</td>
<td>2</td>
<td>&lt;1</td>
<td>E, O</td>
</tr>
<tr>
<td>2</td>
<td>m'U</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>12</td>
<td>m'U</td>
<td></td>
<td>5</td>
<td>3</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>16</td>
<td>m'mU</td>
<td></td>
<td>1</td>
<td>3</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>18</td>
<td>m'mU</td>
<td></td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>F</td>
</tr>
<tr>
<td>22</td>
<td>m'mU</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>E</td>
</tr>
<tr>
<td>24</td>
<td>m'mU</td>
<td></td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>25</td>
<td>m'mU</td>
<td></td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>E, O</td>
</tr>
<tr>
<td>27</td>
<td>m'mU</td>
<td></td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>E, O</td>
</tr>
<tr>
<td>28</td>
<td>m'mU</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>29</td>
<td>m'mU</td>
<td></td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>1'</td>
<td>S</td>
<td></td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>B</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>156</td>
<td>100</td>
<td>29</td>
<td>1</td>
</tr>
</tbody>
</table>

* For the numbering, name, and chemical structure of the modified nucleosides, see Appendix 1. Only the nucleotides detected at position 34 and 37 are indicated.

* See also Appendix 1.

* One-letter or one-symbol code used in the iRNA database by Spindel et al. (1999).

* Number of nucleotides of a specific type relative to the total number of nucleotides of its class (i.e., A, G, C, or U).

* Number of nucleotides of a specific type relative to the total number of nucleotides in their position (the 1996 release of the iRNA database contains 346 iRNA sequences).

* Origins of the RNA sequences in which the modified nucleotides are reported (A, archaea; B, bacteria; F, fungi; E, eukaryotes; O, others (filamentous, mitochondria, and virus [RNA or DNA]))

* The sequence of *Nocardiopsis dassonvillei* strain RNA3 containing the rare 5'AP adenosine (Lazebnikova-Rutkowska et al., 1994) has been added to the present compilation. See also Fig. 2 of Chapter 6.
The human tRNALys$_{UUU}$ is modified with 5-methoxycarbonylmethyl-2-thiouridine at position 34 (wobble position of tRNA) and 2-methylthio-N6-threonylcarbamoyladenosine at position 37 (tRNALys$_{UUU}$mcm$_5^s$U$_{34}^s$, ms$^2$t$_6$A$_{37}$). Chapter 2 describes how the specific modifications at the ASL of tRNALys$_{UUU}$ (ASL tRNALys$_{UUU}$mcm$_5^s$U$_{34}^s$, ms$^2$t$_6$A$_{37}$) serve to limit codon recognition to only the two Lys codons, AAA and AAG.

On the other hand, Valine (Val) is encoded by four codons. Thus, Val is four-fold degenerate. All four Val codons reside in the same codon box; therefore, Val codons are unique from the codons of the other amino acids. Chapter 3 explains how a 5-oxyacetic acid modification at the wobble Uracil and N$^6$-methyladenosine at position 37 of the anticodon stem and loop of tRNAVal$_{UAC}$ (ASL tRNAVal$_{UAC}$cmo$_5^{U}U_{34}$, m$^6$A$_{37}$) are not required to bind the tRNA’s cognate codon GUA but required to read the other Val codons of the GUN codon box. Thus, the modifications of ASL tRNAVal$_{UAC}$cmo$_5^{U}U_{34}$, m$^6$A$_{37}$ allow expansion of codon recognition. The tRNAVal$_{UAC}$cmo$_5^{U}U_{34}$, m$^6$A$_{37}$ is an example of one ASL-modified tRNA that can read the synonymous codons from the same codon box.

tRNA modifications can also serve to expand codon recognition to include unconventional codons. The Genetic Code varies to a certain extent among certain organisms and semiautonomous organelles. Various codon reassignments can be found in the translational systems of the semiautonomous organelles, mitochondria and chloroplasts, of different organisms. The mammalian mitochondrial Genetic Code deviates from the universal Genetic Code in that the universal termination codon UGA codes for Tryptophan; the two universal Arginine codons AGA and AGG are stop codons, and the universal Isoleucine (Ile) codon AUA is reassigned to decode Methionine (Met) [77].
In the universal Genetic Code, AUG is the start codon and the sole codon for Methionine. Genetic translation occurs in three phases: initiation, elongation, and termination. Initiation is when the mRNA binds the small ribosomal subunit and the initiator tRNA reads the start codon at the ribosome’s peptidyl (P)-site. Initiation is aided by initiation factors. After initiation, elongation causes the growth of the polypeptide chain until completion with the termination step. In the cytoplasm, two tRNA species recognize AUG in accordance to the codon’s placement on the ribosome. At initiation, AUG is at the P-site of the ribosome and is recognized by the initiator tRNA\textsuperscript{Met CAU} with the aid of initiation factors. During elongation, elongator tRNA\textsuperscript{Met CAU} reads AUG at the aminoacyl (A)-site of the ribosome, facilitated by elongation factors. Mammalian mitochondrial translation differs from cytoplasmic translation in that 1) there are two Met codons AUG and AUA, 2) only one tRNA\textsuperscript{Met CAU} recognizes both codons AUG and AUA, and 3) this one tRNA\textsuperscript{Met CAU} acts in both initiation and elongation. Interestingly, this one mitochondrial tRNA\textsuperscript{Met CAU} is modified at the wobble position with a 5-formyl group (mtRNA\textsuperscript{Met CAU-f5C34}) [73, 78, 79]. Cytoplasmic initiator tRNA\textsuperscript{Met CAU} is unmodified at the wobble position. Cytoplasmic elongator tRNA\textsuperscript{Met CAU} is modified at the wobble position with a 2’O-methyl group at the ribose moiety. Bacterial initiator tRNA\textsuperscript{Met CAU} is unmodified at the wobble position, while the bacterial elongator tRNA\textsuperscript{Met CAU} is wobble-modified with N\textsuperscript{4}-acetylcytidine (tRNA\textsuperscript{Met CAU-ac4C34}). Chapter 4 reports the first chemical synthesis of modified human mitochondrial ASL\textsuperscript{Met} (hmtASL\textsuperscript{Met CAU-f5C34}) and the significance of the 5-formyl modification in reading the unconventional Met codon AUA at both the P- and A-sites.
For some mRNAs in the mitochondria, the universal Isoleucine codons AUU and AUC are used as initiation codons. Chapter 5 shows how the 5-formyl group enhances binding to the unconventional Met codon AUA at the ribosomal P- and A-sites and expands codon recognition to include the unconventional start codons AUU and AUC at the P-site. Thus, the 5-formyl modification of hmtRNA\textsuperscript{Met}_{CAU-f^5C_{34}} allows codon-reading extension to include the entire codon box AUN. Furthermore, these results demonstrate how a modification at the ASL enhances the reading of unconventional codons in the reassignment of the universal Genetic Code.
Table 2. The recognition of tRNA wobble position-34 for the wobble position (N3) of the codon.  $x\text{m}^5\text{U}$ includes all 5-methylene uridine derivatives, excluding 2-thio; $s^2\text{U}$ includes all 2-thiouridine derivatives, and $x\text{o}^5\text{U}$ consists of all 5-oxyuridine derivatives.

\(^a\) $f^5\text{C}$ refers to the wobble-modified Cytidine found in mitochondrial tRNA$^{\text{Met}}$.

\(^b\) Recognition of codon N3 at the P-site.

<table>
<thead>
<tr>
<th>Anticodon N\textsubscript{3}\textsubscript{4}</th>
<th>Codon N3</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C, U</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>I</td>
<td>U, C, A</td>
</tr>
<tr>
<td>$x\text{m}^5\text{U}$</td>
<td>G</td>
</tr>
<tr>
<td>$s^2\text{U}$</td>
<td>A, G</td>
</tr>
<tr>
<td>$x\text{o}^5\text{U}$</td>
<td>A, G, U, C</td>
</tr>
<tr>
<td>$f^5\text{C}^a$</td>
<td>A, G, U, C$^b$</td>
</tr>
</tbody>
</table>
REFERENCES


mitochondrial transfer RNA\textsubscript{Met} to decode AUG and AUA codons. *Biochimie*, 77, 104–108.

CHAPTER 2

Modifications of the anticodon stem and loop of human tRNA^{Lys} (ASL^{Lys}_{UUU}) restrict codon recognition to Lysine codons AAA and AAG.

2.1 Introduction

The universal Genetic Code consists of 64 codons arranged into sixteen codon boxes. Each codon box is composed of four codons that are similar in their first two letters and differ only in their third letter. In the ribosome, this third letter of the codon binds to the first letter of the tRNA anticodon to form the wobble pair. The two modified nucleosides 5-methylaminomethyl-2-thiouridine or a derivative at the wobble position-34 (mnm^5s^2U_{34}) and N^6-threonylcarbamoyladenosine or a derivative at position 37 (t^6A_{37}) are found in almost all tRNAs that read A or G in the wobble (third) letter of Glutamine (Gln), Lysine (Lys), and Glutamate (Glu) codons [1-5]. Modifications at the wobble position and position 37 of tRNA^{Gln}, tRNA^{Lys} and tRNA^{Glu} are also required for efficient aminoacylation by the tRNAs’ cognate synthetases [6-10]. In bacteria, these two modifications within the anticodon loop of tRNA^{Lys} (ASL^{Lys}_{UUU}- mnm^5s^2U_{34}, t^6A_{37} or derivatives) have been found to be crucial to the structure of the anticodon loop, and therefore credited to the tRNA’s ability to perform its function on the ribosome [4]. The 5-methylaminomethyl modification of tRNA^{Lys}’s wobble position 34 (tRNA^{Lys}_{UUU}- mnm^5U_{34}) has been implicated in the recognition of the Lys codons AAA and AAG, the maintenance of the reading frame, and translocation from the aminoacyl (A)-site to the peptidyl (P)-site on the ribosome [4-5, 11-12]. The thio group (s^2) contributes
significantly to the stability of the U:A base pairs, whereas the mnm\(^5\)-group enhances recognition of G. *E. coli* has but one gene for tRNA\(^{Lys}\) and it has the mnm\(^5\)\(\_2\)U\(_{34}\)UUt\(_6\)A\(_{37}\) anticodon loop modification. The codon preference is for AAA.

The human tRNA\(^{Lys}\) consists of the modifications 5-methoxycarbonylmethyl-2-thiouridine at the wobble position and 2-methylthio-N\(_6\)-threonylcarbamoyladenosine at position 37 (tRNA\(^{Lys}\)\(_{UUU-mcm}^{5\_2}\)U\(_{34}\), ms\(^2\)t\(_6\)A\(_{37}\)). Humans have two other Lys tRNAs, isoacceptors 1 and 2. The anticodon sequence and modifications for tRNA\(^{Lys1, 2}\) are C\(_{34}\)UUt\(_6\)A\(_{37}\), which decode the codon AAG. Codon preference in human cells is for AAG. However, tRNA\(^{Lys3}\) has come under scrutiny for its ability to act as the primer for reverse transcription of HIV-1 and HIV-2 genomes [13-16]. The details to why tRNA\(^{Lys3}_{UUU-mcm}^{5\_2}\)U\(_{34}\), ms\(^2\)t\(_6\)A\(_{37}\) is specifically chosen by the virus are not entirely clear. The modifications of tRNA\(^{Lys3}_{UUU-mcm}^{5\_2}\)U\(_{34}\) are important in HIV’s ability to utilize tRNA\(^{Lys3}\) for annealing to the viral genome and for reverse transcription [17-18].

Derivatives of the 5-methylaminomethyl-2-thio modification at the wobble position (xm\(^5\)\(\_2\)U\(_{34}\)) and N\(_6\)-threonylcarbamoyladenosine (xt\(_6\)U\(_{37}\)) at position 37 of tRNA are highly conserved among organisms of Archea, Bacteria, and Eukarya [1]. Most of our current understanding of the functions of tRNA modifications is a result of original studies with bacterial translational systems. Far fewer studies have focused on the effects of modifications on the functions of tRNAs of higher organisms. Recently, the 5-methoxycarbonylmethyl-2-thiouridine at the wobble position of tRNA\(^{Lys3}_{UUU-mcm}^{5\_2}\)U\(_{34}\) has been shown to be required for viability in yeast [3]. Here, we report binding characteristics of the anticodon stem and loop domain of human tRNA\(^{Lys3}_{UUU-mcm}^{5\_2}\)U\(_{34}\),
ms²t⁶A₃⁷ (ASL Lys₃ UUU-mcm⁵s²U₃₄, ms²t⁶A₃⁷; Figure 1). As in bacteria, the modifications at the anticodon loop of tRNA restrict codon binding to the Lys codons AAA and AAG. This function prevents binding of the tRNA¹⁷₃ to the similar Asn codons in the split codon box.

2.2 Materials and Methods

Ribosomal Binding Assay

The ribosomal binding assays consisted of reaction mixtures of purified E. coli (MRE 600) 70S ribosomes, chemically synthesized mRNAs, and chemically synthesized ASLs. The primary sequences of the 27-nt mRNA oligonucleotides were derived from T4 gp32 mRNA [19] and were purchased from Dharmaco (ThermoFisher, Lafayette, CO). In order to study binding at the ribosomal A-site, the P-site needs to be blocked or occupied. Therefore, we designed the mRNA with the Methionine (Met) codon AUG at the P-site. The E. coli tRNA⁷Met was then used to saturate the P-site. The mRNA sequences were tested for a low probability of any secondary structure using the program RNA Structure 4.2 [20]. The mRNA sequences synthesized for our studies were as follows (Lys codons AAA and AAG are in bold):

1) 5’-GGCAAGGAGGUAAAAAUGAAAGCAGCGU-3’;
2) 5’-GGCAAGGAGGUAAAAAUGAAAGGCACGU-3’.

The 70S ribosomal subunits were isolated as previously described [21]. The ASLs were 5’-end-labeled using [γ-³²P] ATP (MP Biomedicals). Unlabeled ASLs in a range of concentrations (0-5µM) were mixed with insignificant amounts, but radiochemically
detectable (2,000-5,000 CPM) of 5'-end, $^{32}$P-labeled ASLs in a fixed ratio of unlabeled ASL to labeled ASL, in order to maintain radiochemical-specific activity. The assay was performed in ribosomal binding buffer [50 mM HEPES, pH 7.0; 30 mM KCl; 70 mM NH$_4$Cl; 1 mM DTT; 100 μM EDTA; 20 mM MgCl$_2$]. Ribosomes (0.25 μM) were activated by heating to 42 °C, incubating for 10 minutes and then slowly cooled to 37 °C. The ribosomes were programmed with 2.5 μM mRNA for 15 minutes at 37 °C. To experiment with the A-site, the P-site was saturated with *E. coli* tRNA$^{\text{fMet}}$ (Sigma-Aldrich) for 15 minutes at 37 °C. tRNA$^{\text{fMet}}$ binds to the Met codon AUG; see underlined codons of the mRNA sequences above. Binding of ASL$^{\text{Lys}}$ to the A-site was allowed to proceed for 30 min at 37 °C. The reaction mixtures (20 μL each) were then placed on ice for 20 minutes, diluted with 100 μL buffer per reaction mixture, and filtered through nitrocellulose in a modified Whatman Schleicher and Schuell (Brentford, U.K.) 96-well filtration apparatus [22]. Prior to filtration of experimental samples, the nitrocellulose filter was equilibrated in binding buffer at 4 °C for at least 20 min and each well of the filtration apparatus was washed with 100 μL of cold binding buffer. After filtration of reaction samples, each well was then washed twice with 100 μL of cold ribosomal binding buffer. The nitrocellulose was blotted dry with kim wipes, and the radioactivity was measured using a phosphorimager (Molecular Dynamics, GE Healthcare). Data were measured for radioactive intensity using the program ImageQuant (Amersham). Nonspecific binding was determined by the binding of ASLs to ribosomes without mRNA and subtracted from the experimental data. The final data is a result of at least two separate experiments, each done with samples in triplicate, i.e. at least six results for each binding point.
2.3 Results

We have previously assessed the significance of modifications at the anticodon stem and loop of the bacterial tRNA$^{\text{Lys}}$ (ASL$^{\text{Lys}}_{UUU}$-mnn$^{5}$U$_{34}$, t$^{6}$A$_{37}$) in binding the Lys codons AAA and AAG [4-5]. While either modification at the wobble position or position 37 enhanced codon binding over the unmodified ASL$^{\text{Lys}}$ approximately seven-fold, the modifications in combination increased binding by over ten-fold that of the unmodified ASL$^{\text{Lys}}$. Our experiments with the anticodon stem and loop of human tRNA$^{\text{Lys}}_{3}$ (ASL$^{\text{Lys}}_{3UUU}$-mcm$^{5}s^{2}$U$_{34}$, ms$^{2}t^{6}$A$_{37}$) revealed similar results. In comparison to the modified ASL$^{\text{Lys}}_{3UUU}$-mcm$^{5}s^{2}$U$_{34}$, t$^{6}$A$_{37}$ in \textit{in vitro} ribosomal binding assays, the unmodified ASL$^{\text{Lys}}_{UUU}$ showed very poor binding to the Lys codons AAA and AAG (Figure 2). The binding of the unmodified ASL$^{\text{Lys}}$ to either Lys codon AAA or AAG was nearly undetectable unless modifications at positions 34 and 37 were present (Figure 2). The modified ASL$^{\text{Lys}}_{UUU}$-mcm$^{5}s^{2}$U$_{34}$, ms$^{2}t^{6}$A$_{37}$ bound AAA with a $K_d$ of 3.1 ± 0.4 μM and AAG with a $K_d$ of 3.9 ± 0.8 μM.

2.4 Conclusions

tRNA modifications are highly conserved among all three kingdoms. In split boxes, the difference between codons of two different amino acids resides in the wobble position (third letter) of the codon. Because codons in the same box are very similar, recognition of split boxes must be stringent. tRNAs that decode A or G at the wobble position of split boxes tend to have the xm$^{5}s^{2}$U$_{34}$ at the tRNA’s wobble position (position 34) and a modification at position 37. In \textit{E. coli}, the tRNA’s ASL is modified with mnm$^{5}s^{2}$U$_{34}$ and
In humans, the tRNA’s ASL is modified with derivatives of the bacterial modifications: mcm⁵s²U₃₄ and ms²t⁶A₃₇. These modifications are found to be highly significant to this tRNA’s ability to decode the Lys codons AAA and AAG. Here we report that the modifications at the anticodon stem and loop of human tRNA^{Lys₃}^{UUU} (ASL^{Lys₃}^{UUU}-mcm⁵s²U₃₄, ms²t⁶A₃₇) are required for the anticodon to bind to the Lys codons AAA and AAG. According to Crick’s Wobble Hypothesis, the unmodified ASL^{Lys₃}^{UUU} should bind AAA and AAG because the anticodon UUU is a cognate or Watson-Crick (W-C) pair for codon AAA and the wobble pair for codon AAG [23]. However, we found that the unmodified ASL^{Lys₃} showed very low levels of binding in in vitro ribosomal binding assay. Only with the addition of the naturally-occurring modifications 5-methoxycarbonylmethyl-2-thiouridine to the wobble position and 2-methylthio-N6-threonylcarbamoyladenosine to position 37 (ASL^{Lys₃}^{UUU}-mcm⁵s²U₃₄, ms²t⁶A₃₇), does the ASL bind both codons AAA and AAG. Considering their significance to tRNA^{Lys₃}’s central role of decoding, modifications at the ASL of human tRNA^{Lys₃} may be have implications on other functions of the tRNA^{Lys}, such as its involvement in retroviral replication.

In addition to the quantitative results of codon binding reported here we now have the NMR solution structure of the ASL^{Lys₃} and the X-ray crystallographic structures of the ASL^{Lys₃}^{UUU}-mcm⁵s²U₃₄, ms²t⁶A₃₇ bound to both Lys codons AAA and AAG. A comparison of the solution structure to that of the structure of the ribosomal X-ray indicates that the modifications have restricted motional dynamics that directs the structure toward that required for codon binding. These results are comparable to structures of bacterial ASL^{Lys} [4].
Figure 1. The anticodon stem and loop of human tRNA^{Lys3} (ASL^{Lys3}_{UUU-mcm^5s^2U_{34}} ms^{2,6}A^3_{37}). The ASL is modified at the wobble position-34 with 5-methoxycarbonylmethyl-2-thiouridine, at position 37 with 2-methylthio-N6-threonylcarbamoyladenine, and at position 39 with a pseudouridine.
Figure 2. Ribosomal, equilibrium binding curves of the fully-modified human ASL$^{\text{Lys}_3}$UUU$^{\text{mcm}^5\text{s}^2\text{U}_{34}}$, ms$^2$t$^6$A$_{37}$ (▲) or the unmodified ASL$^{\text{Lys}_3}$UUU (■). *E. coli* 70S ribosomes were programmed with the Lysine codon A. AAA or B. AAG at the A-site. The P-site was saturated with *E. coli* tRNA$^{\text{fMet}}$, which binds to its cognate Methionine codon AUG.
REFERENCES


Chapter 3

Anticodon Domain Modifications Contribute Order to tRNA for Ribosome-Mediated Codon Binding

Franck A. P. Vendeix, Agnieszka Dzierungowska, Estella M. Gustilo, William D. Graham, Brian Sproat, Andrzej Malkiewicz, and Paul F. Agris.

Department of Molecular and Structural Biochemistry, North Carolina State University, 128 Polk Hall, Raleigh, North Carolina 27695-7622, Institute of Organic Chemistry, Technical University, Zeromskiego 116, 90-924 Łódź, Poland, and Integrated DNA Technologies BVBA, Provisorium 2, Minderbroedersstraat 17-19, B-3000 Leuven, Belgium

Published in Biochemistry, 2008, 47, 6117–6129
Anticodon Domain Modifications Contribute Order to tRNA for Ribosome-Mediated Codon Binding†‡

Franck A. P. Vendeix‡, Agnieszka Dziergowska§, Estella M. Gustilo‡, William D. Graham‡, Brian Sproat‖, Andrzej Malkiewicz§, and Paul F. Agris*‡

Department of Molecular and Structural Biochemistry, North Carolina State University, 128 Polk Hall, Raleigh, North Carolina, USA 27695-7622, Institute of Organic Chemistry, Technical University, Żeromskiego 116, 90-924, Łódź, Poland, and Integrated DNA Technologies BVBA, Provisorium 2, Minderbroedersstraat 17-19, B-3000 Leuven, Belgium

Running Title: Wobble Modifications Order tRNA's Anticodon

†This work was supported by grants from the National Institutes of Health (Grant 2-RO1-GM23037 to PFA), the National Science Foundation (MCB-0548602 and 2-RO1-GM23037 to PFA), RNA-TEC NV, and the Polish Ministry of Science and Education (to AM).

*To whom correspondence should be addressed.

Phone: 1-919-515-6188.

E.Mail: Paul_Agris@ncsu.edu

FAX: 1-919-515-2047.

‖North Carolina State University

§Technical University
Abbreviations: ASL, anticodon stem and loop domain; ASL\textsuperscript{Val\textsubscript{UAC}}\textsuperscript{3}, ASL of \textit{E. coli}'s valine tRNA isoaccepting species 3 with anticodon UAC; CD, circular dichroism spectropolarimetry; cmo\textsubscript{5}U\textsubscript{34}, uridine-5-oxyacetic acid; HPLC, high performance liquid chromatography; m\textsuperscript{6}A\textsubscript{37}, N\textsuperscript{6}-methyladenosine-37; NOE, nuclear Overhauser effect; Tm, temperature at the mid-point in the UV-monitored, major thermal transition.
ABSTRACT

The accuracy and efficiency with which tRNA decodes genomic information into proteins require posttranscriptional modifications in or adjacent to the anticodon. The modification uridine-5-oxyacetic acid (cmo$^5$U$_{34}$) is found at wobble position 34 in a single isoaccepting tRNA species for six amino acids, alanine, leucine, proline, serine, threonine, and valine, each having 4-fold degenerate codons. cmo$^5$U$_{34}$ makes possible the decoding of 24 codons by just six tRNAs. The contributions of this important modification to the structures and codon binding affinities of the unmodified and fully modified anticodon stem and loop domains of tRNA$^{Val3}_{UAC}$ (ASL$^{Val3}_{UAC}$) were elucidated. The stems of the unmodified ASL$^{Val3}_{UAC}$ and that with cmo$^5$U$_{34}$ and N6-methyladenosine, m$^6$A$_{37}$, adopted an A-form RNA conformation (rmsd 0.6 Å) as determined with NMR spectroscopy and torsion-angle molecular dynamics. However, the UV hyperchromicity, circular dichroism ellipticity, and structural analyses indicated that the anticodon modifications enhanced order in the loop. ASL$^{Val3}_{UAC}$-cmo$^5$U$_{34}$;m$^6$A$_{37}$ exhibited high affinities for its cognate and wobble codons GUA and GUG, and for GUU in the A-site of the programmed 30S ribosomal subunit, whereas the unmodified ASL$^{Val3}_{UAC}$ bound less strongly to GUA and not at all to GUG and GUU. Together with recent crystal structures of ASL$^{Val3}_{UAC}$-cmo$^5$U$_{34}$;m$^6$A$_{37}$ bound to all four of the valine codons in the A-site of the ribosome’s 30S subunit, these results clearly demonstrate that the xo$^5$U$_{34}$ -type modifications order the anticodon loop prior to A-site codon binding for an expanded codon reading, possibly reducing an entropic energy barrier to codon binding.
INTRODUCTION

Transfer RNA is one of the most understood biological macromolecules. The relationship of its nucleoside chemistry and oligonucleotide structure to its functions in protein synthesis has been studied extensively (1, 2). Those studies have led to the identification of more than 70 different posttranscriptional modifications present in tRNAs (3). tRNA modifications increase stability (4), enhance decoding (5), restore ribosomal binding (6), and influence reading frame maintenance (7, 8). In the course of translation, anticodon domain modifications in particular play important roles in the accuracy and efficiency of protein synthesis. Therefore, the modifications that occur at anticodon wobble position 34, and at the conserved purine 37, 3'-adjacent to the anticodon, are of obvious interest in tRNA’s decoding of mRNA (Figure 1). These modifications play critical and distinctive roles in tRNA’s accurate and efficient binding of cognate and wobble codons within the ribosome’s A-site (9).

Some 40 years ago, Francis Crick explained how a limited number of tRNAs could decode the 61 amino acid codons (10). Our view of codon recognition by tRNA was then altered in the modified wobble hypothesis to accommodate new information about modifications (11). The limited number of tRNAs requires most tRNAs to read more than one codon. Some tRNAs respond to codons in “mixed” codon boxes where distinction of the third codon base (the most degenerate of coding positions) is important for discriminating between the correct cognate or wobble codons and the incorrect but near-cognate codons. For example, a wobble position, 2-thiouridine 34 (s^2U_34), enables recognition of the canonical third codon nucleotide A3 and will wobble to G3 but will not recognize either U3 or C3 (10,
In contrast, other modification chemistries expand wobble codon recognition, such as tRNA’s wobble position inosine 34 (I$_{34}$) that will bind codons ending in A3, C3, or U3 (12). Modifications expand tRNA’s codon recognition beyond that envisioned in the wobble hypothesis. Crick suggested that a U$_{34}$ would pair with an A and wobble to a G in the third position of the codon (A3 or G3) but would not base pair with a U3 or C3. He argued that a U$_{34}$ paired with a U3 or C3 would markedly distort the anticodon-codon minihelix (10). However, wobble position U34 of some tRNA species will base pair with a U3 and even a C3 in recognizing all of the synonymous codons of a 4-fold degenerate codon box (9, 13–16). For tRNAs to recognize codons ending with a U3 or C3, as well as codons ending with A3 and G3, modification of wobble position U34 is essential (9, 13–16). The modified nucleoside capable of binding to A, G, U, and C is uridine-5-oxyacetic acid, cmo5U341 (Figure 1). Each of the six amino acids, alanine, leucine, proline, serine, threonine, and valine, having 4-fold degenerate codons, has a single isoaccepting tRNA species containing the modification cmo$^5$U$_{34}$ (17), but would a tRNA species with cmo$^5$U$_{34}$ be sufficient for cell viability in the absence of all other isoaccepting tRNAs in vivo? The function of cmo$^5$U$_{34}$ in tRNA$^{Pro}$ was analyzed by introducing null copies of the two genes (cmoA and cmoB) identified as part of the cmo$^5$U$_{34}$ synthetic pathway into a strain of *Salmonella* having only one of three tRNA$^{Pro}$ isoacceptors, that with the wobble cmo$^5$U$_{34}$ modification (18). Growth of this and other mutant strains demonstrated that all four proline codons (CCU/C/A/G) were read by the cmo$^5$U-containing tRNAPro, and that the complete modification was critical for reading codons ending in U and C (18). However, the physicochemical contributions of cmo$^5$U$_{34}$ to anticodon-codon recognition were not known, though a study of mononucleotide
conformation suggested that cmo\textsuperscript{5}U\textsubscript{34} favored the C\textsuperscript{2\textprime}-endo sugar conformation due to an interaction between the modification’s free acid and the 5\textprime-phosphate backbone (15). We had determined that the cmo\textsuperscript{5}U\textsubscript{34} modification was required for translocation of tRNA\textsuperscript{Val\textsubscript{3}UAC} from the ribosome’s A-site to its P-site when bound to the GUU codon but was not required when it was bound to the cognate codon GUA (19). Surprisingly, the insertion of the single cmo\textsuperscript{5}U\textsubscript{34} modification allowed translocation. The absence of N6-methyladenosine, m\textsuperscript{6}A\textsubscript{37}, which is also a modification found in native tRNA\textsuperscript{Val\textsubscript{3}UAC}, did not alter the translocation of unmodified tRNA\textsuperscript{Val\textsubscript{3}UAC} on its cognate codon and confirmed that modifications of purine 37 are not required for translocation on the cognate codon. Thus, results from studies \textit{in vivo} and \textit{in vitro} confirmed the importance of the cmo\textsuperscript{5}U\textsubscript{34} modification in tRNAs for the six amino acids, and the decoding of the corresponding 18-24 codons. However, the chemical and structural contributions of cmo\textsuperscript{5}U\textsubscript{34} to tRNA\textsuperscript{Val\textsubscript{3}UAC} function were unknown. To understand how cmo5U34 could affect function, we report the solution structures of the unmodified and fully modified anticodon stem and loop domains of tRNA\textsuperscript{Val\textsubscript{3}UAC} (ASL\textsubscript{Val\textsubscript{3}UAC}) and compare them to the crystal structures of the ASL\textsubscript{Val\textsubscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37} bound to its four valine codons in the decoding site of the ribosome’s 30S subunit. In addition, we report the thermodynamic and A-site codon binding properties of the unmodified ASL\textsubscript{Val\textsubscript{3}UAC} and modified ASL\textsubscript{Val\textsubscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37}. We observed that cmo\textsuperscript{5}U\textsubscript{34} and m\textsuperscript{6}A\textsubscript{37} in ASL\textsubscript{Val\textsubscript{3}UAC} confer order and stability to prestructure the tRNA before binding to the codon in the ribosome’s A-site.
EXPERIMENTAL PROCEDURES

RNA Sample Preparation. The unmodified Escherichia coli ASL Val3 UAC used in this study was chemically synthesized by Dharmacon (ThermoFisher, Lafayette, CO) using “ACE” chemistry (20). The modified ASL Val3 UAC construct was synthesized by Integrated DNA Technologies BVBA (formerly RNA-Tec). For the solid-phase synthesis of the modified ASL Val3 UAC, the unmodified and modified phosphoramidite monomers were protected by 5'-O-dimethoxytrityl and 2'-O-tert-butyldimethylsilyl groups, whereas the exoamine groups were protected with 4-tert-butyphenoxycetetyl groups. The cmo5U carboxylic acid moiety was protected as its p nitrophenylethyl ester. The deprotection of the oligomers has been achieved (21). The ASLs were purified by HPLC using a Nucleogen 60-7 DEAE (250 mm × 10 mm) column. Desalting was accomplished with Waters Corp. Sep-pak cartridges. For purposes of crystallography and NMR spectroscopy, the ASLs were further purified by extensive dialysis (using Slide-A-Lyzer MINI Dialysis Units, 3.5K MWCO from Pierce) against 20 mM phosphate buffer (pH 6.2), 50 mM Na+, and 50 mM K+. After the pH had been adjusted to 6.2 (pH/Ion Analyzer MA 235, Mettler Toledo), the oligonucleotides were heated to 80 °C followed by slow cooling to form a solution homogeneous in ASL conformation. After cooling, the samples were lyophilized using a freeze-dryer (Thermo Savant SPD Speed Vac, Thermo Scientific) and then dissolved in 99.996% 2H2O or a 90% 1H2O/10% 2H2O mixture to give a final volume of 300 μL. Samples in 2H2O were redissolved in 2H2O at least twice more and lyophilized. NMR samples of the ASLs were generally at concentrations of 2 mM (in 1H2O) and 1.5 mM (in 2H2O). In addition to HPLC nucleoside
composition analyses (22), mass spectroscopy (MALDI-TOF), and NMR analyses, the results of X-ray crystallography unambiguously confirmed the incorporation of \( m^6A_37 \) and \( cmo^5U_{34} \) into the ASLs.

Ribosomal Binding Assay. The 27-mer mRNA sequences used in codon binding assays were designed from that of T4 gp32 mRNA (23), purchased from Dharmaco, chemically deprotected, and HPLC-purified. Each mRNA sequence was entered into mFold (33) to ensure that there was a low probability it would fold into a stable conformation. The mRNA sequences were 5′-GGCAAGGAGGUAANAAUUGGUAGCAGCU-3′,
5′-GGCAAGGAGGUAANAAUUGGUGGCACGU-3′,
5′-UGCAAGGAGGUAANAAUUGGUCGCACGU-3′, and
5′-UUGAAGGAGGGUUUUAUGGUUGCACGU-3′.

Codon binding assays were performed as previously described (19). The 30S ribosomal subunits were prepared from *E. coli* MRE600 (25). To test that the mRNA fragments bound the 30S subunits, the mRNAs were 5′-end \(^{32}\)P-labeled, titrated into 30S subunits, and monitored for binding using a filter binding assay (described below). ASLs were 5′-end \(^{32}\)P-labeled using \([\gamma-^{32}\)P]ATP (MP Biomedicals). Unlabeled ASLs over a range of concentrations (0-7.5 μM) were mixed with insignificant amounts, but radiochemically detectable (3000-5000 cpm), of 5′-end \(^{32}\)P-labeled ASLs. A fixed ratio of unlabeled ASL to labeled ASL was maintained for each concentration in the range. The 30S ribosomal subunits (0.5 μM) were activated by incubation at 42 °C for 10 min and then slowly cooled to 37 °C. The subunits were then programmed with 50 μM mRNA for 15 min at 37 °C. The P-site was
saturated with 1-2.5 μM tRNAfMet (Sigma, St. Louis, MO) for 15 min at 37 °C before each concentration of ASL was added. A-Site binding was allowed to proceed for 30 min at 37 °C. The reaction mixtures were then placed on ice for 20 min and then filtered through nitrocellulose in a modified Whatman Schleicher and Schuell (Brentford, U.K.) 96-well filtration apparatus (26). Prior to filtration of experimental samples, the nitrocellulose filter was equilibrated in binding buffer at 4 °C for at least 20 min and each well of the filtration apparatus was washed with 100 μL of cold binding buffer [80 mM potassium cacodylate (pH 7.2), 20 mM MgCl₂, and 150 mM NH₄Cl]. Cold binding buffer (100 μL) was added to each sample, and the entire volume was quickly filtered. Each well was then washed at least twice with 100 μL of cold binding buffer. The nitrocellulose was dried out on paper towels, and the radioactivity was measured using a PhosphoImager (Molecular Dynamics, GE Healthcare). Data were measured for radioactive intensity using ImageQuant (Amersham) and were analyzed using single-site, nonlinear regression (Prism, GraphPad Software, Inc., San Diego, CA).

**UV-Monitored Thermodynamic Experiments.** The ASL samples were dissolved to obtain a concentration of 2 μM in 20 mM phosphate buffer. UV-monitored, thermal denaturations and renaturations were replicated four times and monitored by measuring UV absorbance (260 nm) using a Cary 3 spectrophotometer as previously described (27). The data points were averaged over 20 s and collected three times per minute with a temperature change of 1 °C/min from 4 to 90 °C. The data were analyzed (28), and the thermodynamic parameters were determined using Origin (Microcal).
**NMR Spectroscopy.** NMR spectrometers used for this study included the Bruker DRX500 and Varian Inova-600 instruments. NMR data were processed with either XWINNMR (Bruker Inc., Rheinstetten, Germany) or NMRPipe (29), and the analysis was conducted with SPARKY (30). For the samples dissolved in $^1$H$_2$O, the WATERGATE (31) method was used to suppress the water signal. For samples dissolved in $^2$H$_2$O, a low-power presaturation technique was used. NOESY (32) experiments of the samples in $^1$H$_2$O were performed at low temperatures (2, 5, and 10 °C) to observe the exchangeable proton resonances. To aid the NMR resonance assignments of the nonexchangeable protons, the following experiments were performed: COSY, DQF-COSY (42), TOCSY (33), natural abundance $^1$H-$^13$C HSQC, and $^1$H-$^31$P HETCOR. For the determination of structures with the ASLs dissolved in $^2$H$_2$O, NOESY experiments were conducted with different mixing times and at 22 °C. The NOESY spectra were recorded with mixing times of 50, 75, 100, 200, 300, and 400 ms, without removing the samples from the magnet. The spectra were acquired with spectral widths of 5000 Hz in both dimensions, 1024 points in $t_2$, 360 points with 64 scans per block in $t_1$, and a recycle delay of 1.5 s. The FIDs were processed with 60° phase-shifted sine bell apodization functions and third-degree polynomial baseline flattening in both dimensions. To improve the digital resolution for the cross-peak integration, the FIDs were zerofilled to 2048 × 2048 points.

**Structure Determination.** NMR spectra and the initial NMR-restrained molecular dynamics calculations of structure indicated that the 5 bp stems of the two ASLs adopted the A-form double-helix conformation. Therefore, the nucleotide base pairs were restrained by
ensuring that their hydrogen bond donor-acceptor distances were canonical. To achieve planarity of the base pairs, a pseudoenergy term was applied. In addition to the NOE-derived restraints, the A-form backbone configuration was introduced by the use of proton-proton distances taken from an ideal A-form helix model previously generated using CNS (34). This addition increased the rate of convergence but did not affect the initial conformation of the stems or of the loops obtained with only the NMR-derived restraints that were used for the structure calculation. Thus, the effect of the complementary restraints on the stems did not propagate into the loops and, therefore, did not influence the loop conformations. The backbone and the ribose torsion angles were restrained to be within 15° of the standard A-form values. NOESY mixing time studies also provided distance restraints between nonexchangeable protons in the loop region. The NOE cross-peaks were integrated by using the peak fitting Gaussian function and volume integration in SPARKY. The distance for each crosspeak was calculated and normalized to the nonoverlapped pyrimidine H5-H6 cross-peaks with a distance of 2.44 Å. Upper and lower bonds were assigned to 20% of the calculated distances. The distances involving the unresolved protons, i.e., methyl of m6A37 and methylene of cmo5U34, were assigned using pseudoatom notation to make use of the pseudoatom correction automatically computed by CNS.

High-resolution DQF-COSY spectra were used to characterize the sugar pucker based on the $^3J_{H1'-H2'}$ scalar couplings. Residues with nonobservable H1'-H2' crosspeaks or with $^3J_{H1'-H2'}$ values of less than 3Hz were restrained to the C3'-$endo$ conformation, whereas in those with $^3J_{H1'-H2'}$ values appearing to be between 4 and 5 Hz, the ribose was left unconstrained to take into account the conformational averaging between the C3'-endo and
C2'-endo conformations. The $\alpha$ and $\zeta$ torsion angles were restrained to exclude the $trans$ conformation for those residues for which 31P chemical shifts fell within the narrow range commonly seen for regular A-form RNA structures (35). The $\beta$, $\varepsilon$ and $\gamma$ angles were restrained on the basis of the $^1H$-$^{31}P$ HETCOR spectra as described previously (35, 36).

Structure calculations were performed using CNS and were based on published protocols (37) with minor modifications. Briefly, the initial structures consisted of extended strand conformations. The initial coordinates were regularized using simulated annealing and conjugate-gradient minimization to obtain good local geometry. Thus in the first stage, the regularized extended strands were subjected to 60 ps of torsion-angle molecular dynamics at 20000 K. The term $w_{vdw}$ was set to 0.1 to ease rotational barrier crossing. The structures were then subjected to a slow-cooling torsion-angle molecular dynamics step where the temperature was reduced from 20000 to 1000 K over 60 ps. During this period, $w_{vdw}$ was linearly increased from 0.1 to 1.0. In the third stage, the structures were slowly cooled from 1000 to 300 K over 6 ps by using Cartesian molecular dynamics. In the final stage, the structures were minimized by using the conjugate-gradient method. To obtain an ensemble of structures, this protocol was repeated with different initial velocities. The acceptance of the resulting structures was achieved by using the described criteria (37).

RESULTS

Modified nucleotides individually and in concert contribute to a tRNA’s function in translation by providing new chemistries and structure (9). The anticodon stem and loop
domain of tRNA\textsubscript{Val\textsubscript{3\textsubscript{UAC}}}, ASL\textsubscript{Val\textsubscript{3\textsubscript{UAC}}}, is modified at wobble position 34 and, 3'-adjacent to the anticodon, at the conserved purine 37 (Figure 1). To understand the important contributions of these modifications to the decoding function of \textit{E. coli}'s tRNA\textsubscript{Val\textsubscript{3\textsubscript{UAC}}}, we analyzed and compared the properties of the unmodified ASL to that of the ASL modified with both cmo\textsuperscript{5}U\textsubscript{34} and m\textsuperscript{6}A\textsubscript{37} (Figure 1). Synthesis of ASL\textsubscript{Val\textsubscript{3\textsubscript{UAC}}- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37} required the development of a new strategy for protecting and later deprotecting the carboxylic acid function without producing the amide. The amide is readily formed from the commonly utilized ammonia or methylamine deprotection protocol. The presence and stoichiometric quantification of the modifications were confirmed through mass spectroscopy of the ASLs and HPLC of the composite mononucleosides and later by X-ray crystallography. The two ASL\textsubscript{Val\textsubscript{3}} constructs were found to be unimolecular at concentrations for UV-monitored, thermal denaturations and NMR spectroscopy.

\textit{Modifications Order the Anticodon Loop}. The chemistry and structure of ASL modifications alter the thermal stability of the RNA (4, 27, 38, 39). To determine and compare the thermodynamic and base stacking contributions of cmo\textsuperscript{5}U\textsubscript{34} and m\textsuperscript{6}A\textsubscript{37} to ASL\textsubscript{Val\textsubscript{3\textsubscript{UAC}}}, UV-monitored, thermal transitions and circular dichroism (CD) spectra of the unmodified and doubly modified ASLs were analyzed under conditions used for structure determination by NMR. The ASLs showed a cooperative behavior during the melting process by exhibiting one major transition (Figure 2). The melting temperatures ($T_m$) for the two ASLs were high due to the G-C rich stem region but were within error of each other. ASL\textsubscript{Val\textsubscript{3\textsubscript{UAC}}- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37} had a melting temperature of 71.8 (1.3°C, whereas the $T_m$ of unmodified ASL\textsubscript{Val\textsubscript{3\textsubscript{UAC}} was 71.2 (1.1°C (Table 1). cmo\textsuperscript{5}U\textsubscript{34} is a free acid. However, the $T_m$
of ASL\textsubscript{Val\textsuperscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} was not affected by the presence of Mg\textsuperscript{2+} at a concentration 5-fold greater than that of ASL, and neither was that of unmodified ASL. However, the modifications made a significant entropy contribution (\(\Delta\Delta S\)) 25 cal K\textsuperscript{-1} mol\textsuperscript{-1}). This suggested an increase in the level of base stacking and order within the loop in the presence of modification and was confirmed by analyses of the degree of hyperchromicity for doubly modified ASL as compared to that of unmodified ASL. The hyperchromicity of ASL\textsubscript{Val\textsuperscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} was more than double that of unmodified ASL (Figure 2a and Table 1). Thus, it is inferred that the base stacking in the loop is enhanced by the modifications cmo\textsuperscript{5}U\textsubscript{34} and m\textsuperscript{6}A\textsubscript{37}.

The degree of ellipticity in the CD spectra of nucleic acids is most affected by the restricted rotation around the glycosidic bond, and base stacking interactions (40). Therefore, the contributions of modifications to base stacking and order can be ascertained by a comparison of CD spectra from the modified and unmodified ASLs (40). To further investigate the effect of modifications on the secondary structure of the loop, CD spectra of the two ASL\textsubscript{Val\textsuperscript{3}UAC} forms were collected under the same buffer conditions that were used for thermal denaturations and NMR structure determination. The resulting spectra (Figure 2b) exhibited the three expected ellipticities for an RNA hairpin at 210, 230, and 260 nm. The negative and positive ellipticities at 210 and 260 nm, respectively, were characteristic of right-handed helices (40). The third ellipticity at 230 nm could be attributed to the anticodon loop. However, it is the ellipticity at 260 nm that most accurately reflects differences in base stacking (41). The introduction of modifications significantly increased the intensity of the ellipticity at 260 nm (Figure 2). While the thermal transitions observed in UV-monitored
thermal denaturations are biased toward base stacking interruptions in stems, CD spectra monitor stems and loops more equally. Thus, our CD results indicate that the base stacking interactions in the ASLs, more precisely in the loops, were enhanced by the presence of cmo5U34 and m6A37. The increased ellipticity, together with the enhanced hyperchromicity observed during the UV-monitored melting experiments, supports the conclusion that the modifications within the ASL loop increase the level of base stacking.

**Recognition of Valine Codons at the Ribosomal A-Site.** tRNAValUAC decodes GUA and wobbles to GUG but also decodes GUU and GUC (14). The cmo5U34-U3 and cmo5U34-C3 base pairs were unexpected in light of the wobble hypothesis. However, isoaccepting species of Ala and Pro tRNAs with cmo5U34 also have been shown to read their respective codons ending in U and C (18, 42, 43). Recent studies of mutant strains of *Salmonella enterica* have shown that mutants consisting of only the cmo5U34-containing tRNAVal isoacceptor and, thus, lacking the other tRNAVal isoacceptor were viable (44). Furthermore, cmo5U34 is required for effective reading of Val, Pro, and Ala codons ending with G (44). In an effort to determine the contributions of cmo5U34 and m6A37 in the tRNA’s binding of the four codons, unmodified ASLVal3UAC and fully modified ASLVal3UAC-cmo5U34;m6A37 were assayed for binding affinities for each of the four valine codons in the A-site of the 30S ribosomal subunit. The 30S subunit was programmed with a message consisting of the initiation codon (AUG) at the P-site and one of the valine codons at the A-site. Binding of the ASL to codons in the A-site of the 30S ribosomal subunit is comparable to binding of tRNA to codons on the 70S ribosome (45). To ensure binding of ASLVal3UAC to the A-site only, the
P-site was saturated with *E. coli* tRNAfMet. ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}} and ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37}} bound the cognate codon GUA with submicromolar \(K_d\) values of 0.37 (0.06 and 0.17 (0.03 \(\mu\)M, respectively (Figure 3 and Table 2). ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}} did not bind to GUG, in comparison to its fully modified counterpart, ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37}}. ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37}} bound GUG with an affinity of 1.96 (0.32 \(\mu\)M (Table 2). These results are consistent with those of the growth experiments in vivo (44). Only the fully modified ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37} was found to bind GUU and did so with a \(K_d\) of 1.93 (0.70 \(\mu\)M. We had succeeded in determining the X-ray crystallographic structure of ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37} bound to GUC, as well as the other three valine codons (46). Therefore, ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37} will bind to GUC, but we could not determine a binding constant under our experimental conditions or using ribosome binding conditions of other investigators (23, 43). The cmo\textsuperscript{5}U\textsubscript{34}-containing tRNA\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}}, in the absence of other tRNA\textsuperscript{Val} isoacceptors, was able to read all four valine codons in vivo, although the extent of growth of cells was reduced (44). The reduction in the extent of growth was attributed mostly to the inefficiency of the cmo\textsuperscript{5}U\textsubscript{34}-modified tRNA\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}}, in comparison to isoacceptor tRNA\textsuperscript{Val\textsubscript{GAC}}, in reading the GUC codon (44). The GUC codon tends to be read by the tRNA\textsuperscript{Val\textsubscript{GAC}} isoacceptor. Therefore, our inability to detect the binding of ASL\textsuperscript{Val\textsubscript{3}U\textsubscript{AC}-cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37} to GUC may be due to the reduced efficiency of the UAC anticodon in base pairing with the GUC codon, even with the cmo\textsuperscript{5}U\textsubscript{34} modification present (44). Under experimental conditions different from those reported here, the tRNA\textsuperscript{Ala\textsubscript{UGC}} species with cmo\textsuperscript{5}U\textsubscript{34} was observed to bind weakly to GCC (43).
Resonance Assignments. UV-monitored thermal transitions and the CD spectra indicated that the cmo$^5$U$_{34}$ and m$^6$A$_{37}$ modifications contributed to the ordered structure of ASL$^{\text{Val}_3}^{\text{UAC}}$. With the purpose of understanding the degree to which the modifications influence anticodon conformation and dynamics, we determined the solution structures of the unmodified and fully modified ASL$^{\text{Val}_3}^{\text{UAC}}$. A series of one and two-dimensional NMR spectra were collected. The exchangeable, base-paired, imino proton resonances were identified and assigned using one-dimensional (1D) and Watergate-NOESY spectra. The nonexchangeable proton signals were identified and assigned using a combination of one-dimensional, COSY, DQF-COSY, $^1$H-$^3^1$P HETCOR, natural abundance $^1$H-$^3^1$C HSQC, and NOESY spectra, following well-established procedures (35, 47). The addition of 5 mM Mg$^{2+}$ did not alter the NMR spectra, and higher concentrations of Mg$^{2+}$ will broaden the NMR resonances, reducing the spectral resolution (4). Therefore, the NMR spectra were collected without Mg$^{2+}$ being added to the ASLs.

Assignment of Exchangeable Proton Resonances. The base-paired, imino proton region of the 1D spectra exhibited five low-field-shifted resonances between 12.00 and 14.50 ppm for each of the ASLs (Figure 4). The chemical shift dispersion observed in this region was characteristic of imino protons engaged in Watson-Crick base pairing (35, 47). The number of base paired imino proton peaks corresponded exactly to the number of base pairs present in the stem of the four ASLs, indicating that the stems of the ASLs adopted stable duplex secondary structures. As a result, we initially assigned sequential resonances from the NOESY spectra by assuming that the stem conformations were close to that of the right-handed, A-form RNA duplexes. Later, this was to be confirmed by the “NOE walk” of
spectra in D2O (32). The spectra of $\text{ASL}^{\text{Val}_3\text{UAC}-\text{cmo}^5\text{U}_{34};\text{m}^6\text{A}_{37}}$ exhibited two minor imino proton resonances. These small signals could be due to heterogeneity of conformation, or to the presence of failed sequences (impurities) that occurred during synthesis that could not be removed during the different steps of the oligonucleotide purifications. Also, two broad resonances were identified between the chemical shifts of 11.60 and 11.30 ppm in the spectra of modified ASL. In contrast, the same peaks were poorly resolved, as one very broad resonance, in the spectrum of unmodified $\text{ASL}^{\text{Val}_3\text{UAC}}$. These resonances were attributed to the non-hydrogen-bonded, imino protons of the $\text{U}_{33}$ and $\text{cmo}^5\text{U}_{34}$ residues that are located in the loop. Thus, $\text{U}_{33}$ and $\text{cmo}^5\text{U}_{34}$ exhibit a slower than expected exchange with the solvent in comparison to unmodified $\text{U}_{34}$. The slower exchange could be attributed to a change in their $\text{pK}_a$ values or to an ordering of the loop through base stacking. In fact, the $\text{pK}_a$ values of the hydrogens at the N3 position of U (HN3), mcm$^5$U, and cmnm$^5$U have been determined to be 9.38, 8.20, and 8.44, respectively (48–50). The $\text{pK}_a$ of HN3 of cmo$^5$U has been estimated to be 8.02 (51). Though modification of position 5 of uridine reduces the $\text{pK}_a$, it is still significantly higher than the pH of the NMR experiments (pH 6.2) and the physiological pH.

If the modifications contribute to the base stacking, then one would expect that HN3 of unmodified $\text{U}_{34}$ within $\text{ASL}^{\text{Val}_3\text{UAC}-\text{cmo}^5\text{U}_{34};\text{m}^6\text{A}_{37}}$, a homologue of $\text{ASL}^{\text{Val}_3\text{UAC}}$ containing only the $\text{m}^6\text{A}_{37}$ modification, may exchange at a rate slower than that of completely unmodified ASLVal3UAC. In a recently recorded NMR spectrum of $\text{ASL}^{\text{Val}_3\text{UAC}-\text{m}^6\text{A}_{37}}$, the proton of N3 of $\text{U}_{34}$ was found to be in a slower exchange than that of unmodified ASL (F. A. P. Vendeix et al., personal communication). Thus, the more resolved resonances of the uridine imino protons in the spectra of $\text{ASL}^{\text{Val}_3\text{UAC}-\text{cmo}^5\text{U}_{34};\text{m}^6\text{A}_{37}}$ suggest that in the
presence of modifications the imino protons of U33 and cmo5U34 were protected from rapid exchange with the solvent. These results are consistent with an increased order through base stacking in the anticodon loop.

The resonance most shifted to the low field at 14.09 ppm and observed in the NOESY spectra of the ASLs was readily assigned to U29 since only one uridine was present in the stem sequence (Figure 4). This cross-peak was a convenient starting point for the specific sequential resonance assignments of the imino-imino and imino-amino proton connectivities. Nevertheless, the resonance observed at 13.42 ppm on the 1D spectra of the ASLs could not be identified on the NOESY spectra, indicating fast exchange between the imino proton and the protons of water. The fast exchange could be explained by the fraying of the terminal stem base pair (C27-G43), and therefore, this resonance was assigned to the G43 imino proton. Analogues of unmodified ASLVal3UAC were synthesized with the reverse orientation of the C28-G42 and U29-A41 base pairs. NMR analysis of these analogues’spectra clarified the imino signal assignments and helped in the assignment of the many guanosine spin systems.

Assignment of Nonexchangeable Proton Resonances. The methylene group of cmo5U34 and the methyl group of m6A37 were identified and assigned on the basis of their distinctive high-field, sharp resonances observed in the 1D spectra. The pyrimidine H5 and H6 resonances were readily identified on the COSY and 1H-13C HSQC spectra. In the case of unmodified ASLVal3UAC, nine cross-peaks corresponding to the total number of Cs and Us present in the sequence were observed. This was not the case for ASLVal3UAC-cmo5U34;m6A37, where only eight cross-peaks were present (Figure 5). The missing cross-peak confirmed that
the H5 proton of U34 was substituted by the functional group cmo5 and also allowed us to unambiguously assign the H5 and H6 protons of U34.

The H1′, H6, H8, and H2 protons were assigned by using the NOESY spectra on which the H1′-H6/H8 NOE sequential connectivity could be followed from the 5′- to 3′-end of the ASLs (Supporting Information). Breaks in the sequence between the connectivity of U33 H1′-U34 H6 and that of C36 H1′-A37 H8 were observed on the spectra of two ASLs. ASL_{Val^3}^{UAC} exhibited a missing NOE cross-peak between A37 H8 and A38 H1′. An additional resonance was assigned to an NOE between A35 H8 and U33 H1′. With the introduction of cmo5U34 and m6A37, we observed that the H1′ protons of specific loop residues experience changes in chemical shift. The H1′ protons of U34, A35, C36, and A37 had altered chemical shifts in the F2 dimension (Supporting Information). In addition, these spectra indicated that even the unmodified ASL_{Val^3}^{UAC} had at least one U33 connectivity that is routinely associated with the canonical U-turn motif of tRNA anticodons, the U33 H1′-A35 H8 NOE.

The DQF-COSY and COSY spectra of the ASLs revealed that most of the nucleotides were completely within the expected C3′-endo conformation. However, just three of the ribose moieties, those of G43, A35, and U34/cmo5U34 having $^3J_{H1\prime-H2\prime}$ couplings between 3 and 5 Hz, exhibited almost equally proportioned C3′-endo and C2′-endo conformations (Figure 6) (35). This suggested that the conformational dynamics of the sugar pucker was not affected by the modifications. Since the ribose conformations were established for the 17 nucleotides, the identification of the H2′ resonances could be confirmed through observation of the strong cross-peaks to the H1′ protons in the short mixing time NOESY experiment (35). The H3′ protons were identified by using the 1H-31P,
HETCOR, and $^1$H-$^{13}$C HSQC experiments. A nearly complete identification of each individual nucleotide’s proton resonances was accomplished by using the sequential connectivities of the $\text{H1}'$-$\text{H8/H6}$, $\text{H2}$-$\text{H8}$, and $\text{H3}'$-$\text{H8}$ NOE cross-peaks to assign the $\text{H1}'$, $\text{H2}'$, and $\text{H3}'$ protons. The remaining anomeric protons ($\text{H4}'$ and $\text{H5}'/\text{H5}''$) were identified on the $^1$H-$^{13}$C HSQC spectra due to their distinctive $^{13}$C chemical shift dispersion. The $\text{H4}'$ protons were assigned by using the $\text{H1}'$-$\text{H4}'$ NOE crosspeaks observed with long mixing time. Due to severe resonance overlaps, only a partial assignment of the $\text{H4}'$ and $\text{H5}'/\text{H5}''$ protons was achieved.

The phosphorus chemical shifts observed in the $1\text{H}-31\text{P}$ HETCOR spectra of the ASLs were between 0.2 and 2.6 ppm (Supporting Information). A group of cross-peaks that corresponded to the loop and the stem residues was found between 0.2 and 1.5 ppm. The $\text{U}_{33}$ $\text{H3}'$-$\text{P34}$ cross-peak of all of the ASLs exhibited a 1 ppm downfield shift of the phosphorus compared to the rest of the residues which exhibited moderate chemical shift changes ($\Delta \delta$ ) 0.1-0.2 ppm). This cross-peak resonated within the chemical shift window common to regular A form structures. However, the downfield shift was more pronounced for ASL$^{\text{Val}3\text{UAC}}_{\text{UAC-cmo}5\text{U34;m6A}37}$ ($\Delta \delta$ ) 0.4-0.5 ppm) than for unmodified ASL$^{\text{Val}3\text{UAC}}_{\text{UAC}}$ ($\Delta \delta$ ) 0.2 ppm).

The spectral analyses described above and assignments lead to several conclusions. The high quality of the different spectra indicated well-behaved unmodified and modified ASLs under the NMR experimental conditions. The A-RNAtype conformation adopted by the two stems was confirmed by the NOE connectivities. Although slight differences in NOE patterns and chemical shifts were found between the ASLs, the stability and the similarity of
their respective stems were emphasized by the considerable identity of the resonances assigned to the residues of the stems. In comparison with unmodified ASL\textsuperscript{Val3\textsubscript{UAC}}, the insertion of modified bases in the loop induced noticeable changes in chemical shifts which indicated specific structural variations. The NMR results confirmed the correlation between the modification of the ASL loop and the enhanced base stacking as observed in the increased ellipticity of the CD spectrum and the hyperchromicity of the UV-monitored thermal denaturations. In addition, the interruption in the NOE sequential connectivity between U\textsubscript{33} H1’ and U\textsubscript{34} H6, the NOE cross-peak connecting A\textsubscript{35} H8 to U\textsubscript{33} H1’, and the low-field-shifted U\textsubscript{33} H3’-P34 cross-peak were among the distinctive NMR “fingerprint” resonances that indicate a canonical “U” turn in anticodon domains (52).

\textit{Structure Determination. Distance and Angle Restraints.} The distance and dihedral angle restraints required for structure determination were derived from NOESY crosspeaks and $J$ couplings (DQF-COSY and $^{31}$P HETCOR spectra). The data obtained from the UV-monitored thermal transitions, the CD spectra, and the NMR experiments clearly indicated an A-form helix conformation adopted by the stems of the ASLs studied herein. These data also showed that the modifications in the anticodon loop did not induce a conformational change or dynamics in the stem. On the basis of the observations described above, and knowing that posttranscriptional modifications do not alter the overall helical structure of tRNA\textsuperscript{Val} (53), we focused on elucidating the structural differences located in the loops. Therefore, the NMR torsion-angle and distance restraints from the two stems were complemented with distance restraints that yield the conventional A-form helix conformation, as reported and
utilized by others (52). The restraints that were added to the stems of modified and unmodified ASLs did not alter the structures of the loops but did enhance the rate of convergence (from 40 to 60%) during structure calculations. In the loop region alone, a total of 76 and 90 NOE-derived distance restraints were obtained for unmodified ASL$^{\text{Val3}}_{\text{UAC}}$ and ASL$^{\text{Val3}}_{\text{UAC}}$-cmo$^5$U$_{34}$,m$^6$A$_{37}$, respectively. (See the structure deposition entries for detailed structure determination statistics.)

We and others have explored the possibility of noncanonical hydrogen bonding across the ASL loop between loop residues C$_{32}$ and A$_{38}$ (4, 9, 54). However, the hydrogen involved has yet to be observed by NMR spectral analyses. Still, the distance between A$_{38}$N6 and C$_{32}$O2 in the two ASL$^{\text{Val3}}_{\text{UAC}}$ structures was determined to be 2.68 Å, and this is strongly suggestive of noncanonical hydrogen bonding between C$_{32}$ and A$_{38}$ occurring at the base of the stem. Protonation of N1 of A$_{38}$ would facilitate formation of such a base pair and does not require a significant reduction in pH. The protonation of N1 of A$_{38}$ has been observed in the structure determination of other ASLs under the solution conditions in which the pH was only moderately acidic (pH 5.5-7.0) (5, 54).

The alpha ($\alpha$) and zeta ($\zeta$) dihedral angles were estimated from the $^{31}$P chemical shift values and were restrained to exclude the trans conformation (35) with the exception of the U$_{33}$-P-U$_{34}$ angle, which was left unrestrained. The $\beta$ angles were estimated from the $^3J_{P-H5'}$ and $^3J_{P-H5''}$ scalar couplings (47). The absence of intense cross-peaks between the phosphorus atoms and the H5'/H5'' protons in the 1H-31P HETCOR spectra indicated that the $\beta$ angles were in the trans conformation commonly observed in A-form RNA. The $\gamma$ angles were restrained to adopt the trans conformation for residues having a weak P-H4' cross-peak.
Otherwise, the \( \gamma \) angles were left unconstrained. Since all \( ^3J_{P-H_3'} \) coupling constants were determined to be greater than 5 Hz, the dihedral \( \epsilon \) angles were restrained to exclude the gauche- (g-) conformation for all residues in the loop.

The glycosidic torsion angle \( \chi \) was restrained to the anti conformation as dictated by the weak intensity of the crosspeak observed between H1' and H6/H8 for all loop residues. The ribose pucker of each residue was restrained on the basis of the \( ^3J_{H1'-H2'} \) scalar couplings. Thus, the ribose puckers of G43, unmodified U34, cmo\(^5\)U34 and A35 were left unconstrained to take into account the conformational averaging between the C3'-endo and C2'-endo conformations.

Description of the ASL Loops. After collection of the distances and dihedral angles derived from NOE and \( J \) coupling constants, respectively, the solution structures of the four ASLs were determined by torsion-angle molecular dynamics simulation (37). Twenty-five structures were generated for each ASL, each of which contained no violations of NOE restraints greater than 0.5 Å and no dihedral angle violations greater than 5°. In addition to these selection criteria, a structure was rejected if the root-meansquare deviation (rmsd) of bonds varied from ideal values by greater than 0.02 Å or the rmsd of angles was greater than 2.0°. Among the 25 structures that converged, 10 had low total energy and no restraint violations. These 10 lowest-energy structures were in agreement with the NMR data and were selected to represent the final ensemble.

We observed interesting similarities and important distinctions between unmodified ASL\(^{Val3}_{UAC} \) and ASL\(^{Val3}_{UAC-cmo^5U34;m^6A37} \) structures. The ASL stems were identical due to the A-form RNA structure imposed by the many NMR-derived restraints, and those few
added restraints that were consistent with an RNA duplex. Structures of the unmodified and fully modified ASL$^{\text{Val3 UAC}}$ (Figure 7) featured R-helical axes of the loops that were not aligned with the respective stems. In addition, the seven nucleotides of the loop exhibited helical conformations stacked to different degrees from the 5′- to 3′-end of the loop, as previously confirmed by the sequential NOE cross-peak connectivity.

In contrast to the stems, there were distinctive differences between anticodon loops. The anticodon loop of ASL$^{\text{Val3 UAC}}$ was less structured than that of ASL$^{\text{Val3 UAC-cmo5U34;m6A37}}$ (Figure 7). The rmsd values for each of the loop nucleotides of ASL$^{\text{Val3 UAC-cmo5U34;m6A37}}$ were consistently lower than those of the unmodified ASL (Figure 7c). The unmodified U$^{34}$ was the most dynamic of the loop residues. cmo$^5$U$^{34}$ appears to be as dynamic as that of the unmodified U$^{34}$ (Figure 7c). However, the methylene protons of cmo$^5$U$^{34}$ had sequential NOE cross-peaks to A$^{35}$ H8 and an intraresidue NOE to its H6 and H1′ protons. We suspect that the free rotation around the carbon of the methylene group of the cmo$^5$U$^{34}$ residue produced different orientations that are responsible for the higher rmsd, even though the base and 5-oxy group were fixed by NOE restraints. The N6-methyl of A$^{37}$ exhibited NOE cross-peaks to A$^{38}$ H1′ and H2, yet the methyl group of m6A$^{37}$ was observed to have two distinct orientations. The orientation of the methyl group toward the plane defined by the aromatic ring of A$^{38}$ was observed in 7 of 10 structures (Figure 7). The three structures with the methyl group oriented away from the plane of A$^{38}$ exhibited a higher rmsd for cmo$^5$U$^{34}$ (Figure 8a,c). Stereochemical hindrance of the methyl group constrained the m$^6$A$^{37}$ nucleotide to a position translated outward from the loop (1.6 Å) relative to that of A$^{37}$ of the unmodified ASL$^{\text{Val3 UAC}}$ (Figure 8). As a consequence, the loop of ASL$^{\text{Val3 UAC-cmo5U34;m6A37}}$
cmo$^5$U$_{34}$;m$^6$A$_{37}$ is broader than that of unmodified ASL$^{Val_{3}}_{UAC}$ (Figures 8 and 9). With the exception of cmo$^5$U$_{34}$, the rmsd values for each of the loop nucleotides of the seven aforementioned structures of ASL$^{Val_{3}}_{UAC}$-cmo$^5$U$_{34}$;m$^6$A$_{37}$ were consistently lower than those of the unmodified ASL (Figure 7c). This indicated that the anticodon loop of ASL$^{Val_{3}}_{UAC}$-cmo$^5$U$_{34}$;m$^6$A$_{37}$ was more highly ordered than that of ASL$^{Val_{3}}_{UAC}$. Thus, the loops of the modified ASL exhibited restricted dynamics caused by a steric hindrance to the positioning of the N6-methyl of A$_{37}$ and to the geometry of the cmo$^5$ moiety due to hydrogen bonding by the base and that of the 5-oxy group (Figures 8 and 9). The differences in loop architecture and dynamics were supported by the numbers of restraints for the loop nucleotides. The NMR based molecular dynamics simulations were conducted with an average of 16 restraints per loop residue for the two ASLs. Thus, the differences were due to the NOE cross-peaks generated by the methyl and the methylene groups of m$^6$A$_{37}$ and cmo$^5$U$_{34}$, respectively. However, more dependable information about the distinguishing dynamics of these loops could be obtained from a detailed study of relaxation measurements using stable samples isotopically labeled with $^{13}$C and $^{15}$N.

Although the DQF-COSY spectrum of the unmodified ASL$^{Val_{3}}_{UAC}$ indicated that some loop nucleotides were in a C3’-endo-C2’-endo sugar equilibrium, the furanose rings were found to adopt preferentially the C3’-endo conformation. A$_{35}$ of the doubly modified ASL exhibited a C2’-endo sugar conformation. The $\chi$ angles of m$^6$A$_{37}$ and cmo$^5$U$_{34}$ were confirmed to be in the anti conformation. However, analysis of the unmodified ASL family of structures revealed that the $\chi$ angle of U$_{34}$ was mostly syn but oscillated between the syn and anti conformations.
The folding of the ASLs is marked by a turn of the backbone between the invariant U₃₃ and the wobble position U₃₄. This folding is allowed by the high degree of rotation of the angle R. The P-O₅′ torsion angle of residue 34 (of the two ASLs) was in the gauche+(g+ or +sc) conformation instead of the g- orientation commonly found in the helical conformation. These turns are generally defined as π-turns (48) and are typically found at the invariant U₃₃ of tRNAs, where the predominant form exhibited is the “U-turn”. They are also known for stabilizing and ordering the anticodon loop by base-phosphate stacking interaction and hydrogen bonding. Thus, the 5’-phosphate attached to O3’ of residue U₃₄ is engaged to some extent in stacking interactions with U₃₃. The hydrogen bond between U₃₃ O2’ and A₃₅ N7 is characteristic of U-turns (52) and was observed in the loop of the two ASLs. This bond was found to be shorter (2.85 Å) in the case of ASLVal⁵UAC-cmo₅U₃₄;m₆A₃₇. Hydrogen bonds between U₃₃ H₃ and C₃₆ PO₂ and between A₃₅ N₆ and U₃₃ O2’ were not observed.

**DISCUSSION**

The six prokaryotic and eukaryotic tRNA species for alanine, leucine, proline, serine, threonine, and valine that have the posttranscriptional oxy modification of wobble position uridine 34 (xo₅U₃₄) (4) are able to read fully 18-24 of the 61 amino acid codons (9). Previously, we found that modifications at wobble position U₃₄ are directly correlated to the need of certain tRNAs to discriminate at the third position of the codon (6). The lysine and asparagine codons are each 2-fold degenerate and share a codon box. Therefore, the codons for lysine and asparagine differ only in the third position. We have shown that modifications at positions 34 and 37 of ASL₆₃LUUU (ASL₆₃UUU-s²U₃₄;t₆A₃₇) are required for the recognition...
of both lysine codons, even the cognate AAA (6). Valine codes, however, are 4-fold degenerate and are translated by tRNAs that do not discriminate among the third base of the codon. We have shown here that modifications are not required for ASL_{Val3}^{Val3} to bind the cognate codon GUA. However, the cmo5_{U34} and m6{A37} modifications are necessary for the ASL to bind GUG and GUU. In addition, the fully modified ASL_{Val3}^{Val3} - cmo5_{U34}, m6{A37} binds GUC, for we have determined the structure in the A-site of the small subunit of the ribosome (46). Unfortunately, we were not able to determine a binding constant. However, the cmo5_{U34}-modified tRNA_{Ala}^{Ala} UGC was shown to bind weakly to GCC (43). In the binding of GCC by tRNA_{Ala}^{Ala} UGC, the first and second base pairs are C_{36}-G1 and G_{35}-C2, respectively, whereas in the binding of GUC by ASL_{Val3}^{Val3} - cmo5_{U34}, m6{A37}, the first and second base pairs are C_{35}-G1 and A_{36}-U2, respectively. The stronger base pairing and stacking of the two G-C pairs formed when tRNA_{Ala}^{Ala} UGC binds GCC may facilitate the cmo5_{U34}-C pairing at the third position. Furthermore, studies in vivo have shown that the cmo5_{U34}-modified tRNA_{Val3}^{Val3} recognizes the GUC codon, although poorly (44).

An understanding of the physicochemical contributions of cmo5_{U34} to anticodon structure and dynamics in solution, and on the ribosome, will provide critical insights into the role of modifications in reading of the universal codes. The results obtained from the UV-monitored thermal transition, the degree of ellipticity observed in the CD spectra, and the NOE connectivities of NMR spectra allowed us to confirm that the secondary structure of the stems of the unmodified and modified E. coli ASL_{Val3}^{Val3} adopts the A-RNA conformation. Moreover, these studies demonstrated that the insertion of modified bases at positions 34 (wobble position) and 37 did not induce any conformational changes in the stem. However,
significant differences in anticodon loop conformations between unmodified and modified ASLs were revealed by the variations in UV hyperchromicity, CD ellipticity, and NMR spectra.

The NMR-determined structures indicated a more highly ordered loop for the modified ASL $^{\text{Val3}}_{\text{UAC}}$, as compared to that of the unmodified ASL. The ASL $^{\text{Val3}}_{\text{UAC}}$-cmo $^{5}$U$_{34}$; m$^{6}$A$_{37}$ structure appears to be conformationally restrained. Furthermore, the free energy difference ($\Delta\Delta G^\circ$) between modified and unmodified ASL $^{\text{Val3}}_{\text{UAC}}$ was appreciably influenced by the entropy term ($\Delta\Delta S$), 20 cal mol$^{-1}$ K$^{-1}$. The hyperchromicity of ASL $^{\text{Val3}}_{\text{UAC}}$-cmo $^{5}$U$_{34}$; m$^{6}$A$_{37}$ also a measure of order, was 2-fold greater than that of the unmodified form. This clearly showed a contribution of the modifications to thermodynamic stability through ordering of the anticodon loop. Numerous posttranscriptional modifications have been found at the conserved purine 37, 3′-adjacent to the anticodon, as well as at wobble position 34. It was predicted that modification of purine 37 would maintain the all-important open anticodon loop structure for codon binding by negating intraloop hydrogen bonding but at the same time contribute order through enhanced base stacking (9). The presence of the modification N6-isopentenyladenosine, i$^{6}$A$_{37}$, in *E. coli* ASL$^{\text{Phe}}$ reduced the $T_m$ and enhanced a low-temperature transition as opposed to its unmodified counterpart (38). Similar changes in thermal stability were observed in constructs of yeast ASL$^{\text{Phe}}$ with three and four modifications (4, 27). The $T_m$ value decreased due to the presence of m$^{1}$G$_{37}$. These studies confirmed that i$^{6}$A$_{37}$ and m$^{1}$G$_{37}$ negated intraloop hydrogen bonding and increased the level of base stacking. The anticodon loop of ASL $^{\text{Val3}}_{\text{UAC}}$-cmo $^{5}$U$_{34}$; m$^{6}$A$_{37}$ was wider than that of the unmodified ASL, indicating that
either m\textsuperscript{6}G\textsubscript{37}, cmo\textsuperscript{5}U\textsubscript{34}, or both were maintaining the open loop. Considering what we know about other purine 37 modifications, m\textsuperscript{6}G\textsubscript{37} may negate intraloop hydrogen bonding and may contribute to the base stacking in the loop of ASL\textsuperscript{Val\textsubscript{3}UAC}. A statistical analysis of the local base step parameters (i.e., the tilt, the roll, and the twist angles) (55) of all the nucleotides of the two ASLs revealed that only the nucleotides of the loops differed significantly (see the Supporting Information). In particular, the angles of base tilt for the loop nucleotides of ASL\textsuperscript{Val\textsubscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} were considerably smaller than that of the unmodified ASL. The modified loop tilts averaged less than half that of the unmodified loop. The reduced step parameters of the loop nucleotides indicate again that the modifications of ASL\textsuperscript{Val\textsubscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} have an ordered loop through base stacking. This increase in order was not due solely to the added NOEs provided by the methyl group of m\textsuperscript{6}G\textsubscript{37} since no such order was observed in the step parameters of the loop nucleosides of the singly modified ASL\textsuperscript{Val\textsubscript{3}UAC- m\textsuperscript{6}G\textsubscript{37}} (F. A. P. Vendeix et al., personal communication). Though the enhanced base stacking could result from the steric hindrance of the hydrophobic methyl group of m\textsuperscript{6}G\textsubscript{37}, cmo\textsuperscript{5}U\textsubscript{34} appears to be the major contributor. The increase in the level of base stacking may be achieved by \(\pi-\pi\) interactions and hydrogen bonding of cmo\textsuperscript{5}U\textsubscript{34}, the modification of which constrained its sugar pucker toward the C3\textsuperscript{'}-endo conformation, and supported a \(\pi\)-like turn.

**Anticodon \(\pi\)- or U-Turn Motif.** Torsion angles, distance restraints, and NOE connectivities implicated the involvement of U33 in a \(\pi\)-like turn, though not all NMR criteria were met. However, the turn at U\textsubscript{33} was more distinct in the ASL\textsuperscript{Val\textsubscript{3}UAC-cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} structure than in the unmodified ASL. The NOE cross-peak between U\textsubscript{33} H1\textsuperscript{'

65
T loop regions of yeast tRNA$^{\text{Phe}}$ (48). In this anticodon loop region, the $\pi$-turn occurred at the 3'-phosphate of the invariant nucleotide, U$_{33}$, which was stacked with the semi-invariant 2'-O-methylcytidine 32 (Cm$_{32}$). At the wobble position, Gm$_{34}$ was displaced by the $\pi$-turn with angles $\alpha$ and $\zeta$ adopting the trans and g conformations, respectively. This distortion induced G$_{34}$ P to stack with U$_{33}$. The imino proton of U$_{33}$ was bound to the oxygen of its 3'-phosphate. As in tRNA$^{\text{Phe}}$ and other tRNAs, the modified ASL$^{\text{Val3 UAC}}$ structure exhibited most, but not all, of the features characterizing the canonical U-turn. In particular, the expected NOE connectivities between U$_{33}$ H2' and the aromatic protons of U$_{34}$ and A$_{35}$ could not be observed. The addition of 5 mM Mg$^{2+}$ did not alter the NMR spectra, indicating that the absence of the canonical U-turn was not due to the lack of the divalent ion. The presence of modifications at positions 37 and/or 34 per se does not induce a U-turn, although the chemical shift of the U$_{33}$ 3'-phosphate was shifted downfield (4, 38). The anticodon loop of the elongator tRNA$^{\text{Met CAU}}$ (52) has a base composition similar to that of ASL$^{\text{Val3 UAC}}$. The only difference is the anticodon base sequence. Although modifications were absent from the loop of ASL$^{\text{Met CAU}}$, a U-turn was clearly identified (52). Thus, canonical turns could be observed in the absence or presence of modified bases in the loop of an ASL. However, specific combinations of modified bases in the loop generate a conformational restriction that could compel the ASL toward a U-turn conformation, for example, the fully modified ASL$^{\text{Lys}}$ (39, 56).

**Structural and Functional Contribution of m6A37.** The highly conserved modification of purine 37 has expanded our concept of codon recognition beyond that of the three codon-anticodon interactions (9), invoking the “expanded anticodon” hypothesis (3, 26, 57, 58). The
degree of hydrophobicity of the purine 37 modifications is associated with the stability of anticodon-codon base pairing (9). Analysis of the NMR and X-ray-derived structures showed that m^6G_{37} is probably negating intraloop base pairing and, thus, widening the anticodon loop. This modification reinforced the stability of the anticodon-codon minihelix by cross-strand stacking above the first adjacent codon base pair in the crystal structure (46), as did N6-threonylcarbamoyladenosine, t^6A_{37} in ASL^{Lys} (12). However, the geometry of m^6G_{37} differs in cross-strand base stacking from that of t^6A_{37}, as seen in a comparison of the crystal structure of ASL^{Val}_{UAC-cmo5U_{34}; m^6A_{37}} (46) with that of ASL^{Lys}_{UUU-mnm5U_{34}; t^6A_{37}} (12) bound to their respective codons. The two modifications are markedly different in size, and the t^6A_{37} ureido group acts as a third ring expanding its availability to stack with the adjacent base. The tRNA^{Lys}_{UUU} anticodon-codon interaction is composed entirely of U-A base pairs which are less thermodynamically stable than C-G pairs (48). The unmodified ASL^{Lys}_{UUU} binds its cognate codon poorly as compared to ASL^{Lys}_{UUU-t^6A_{37}}, whereas the unmodified tRNA^{Val}_{UAC} binds its cognate codon in which the first base pair is C_{36}-G_{1} with an affinity similar to that of the modified ASL.

**Modifications Order tRNA’s Anticodon Domain.** In his wobble hypothesis, Crick postulated that the ribosome would conform, remodel in today’s terms, the anticodon to the codon (10). In contrast, the modified wobble hypothesis proposed that modifications restrain the anticodon domain conformation for accurate and effective reading of cognate and wobble codons (11). The solution structures presented here together with our recently published crystal structures of ASL^{Val}_{UAC-cmo5U_{34}; m^6A_{37}} in complex with its four valine codons in the decoding site of the 30S ribosomal subunit (46) (resolution of 2.8-3.1 Å) provide us with
the opportunity to test the two hypotheses. The doubly modified ASL$_{Val}^{Val3}_{UAC}$ employed in the solution studies and that used in the ribosome crystallization processes originated from the identical sample and at the same purity. The presence of the modified bases cmo$^5$U$_{34}$ and m$^6$G$_{37}$ in ASL$_{Val}^{Val3}_{UAC}$ was detected by the unbiased Fourier difference maps of ASL$_{Val}^{Val3}_{UAC}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ bound to the cognate codon GUA, the wobble codon GUG, and GUU and GUC. The crystal structures of ASL$_{Val}^{Val3}_{UAC}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ bound to the four codons were essentially identical. In addition, the superposition of the NMR and the crystal structure of ASL$_{Val}^{Val3}_{UAC}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ exhibited considerable similarities with few significant noticeable differences. For the most part, the bases in the loop adopted the same orientation and stacking patterns with a rmsd of 2.30 (0.02 Å (Figure 9). As observed in the solution structure of the doubly modified ASL$_{Val}^{Val3}_{UAC}$, the crystal structure on the ribosome displayed a high level of base stacking and order. At the top of the anticodon loop, A$_{38}$ NH$_2$ was only 3.04 (0.22 Å from C$_{32}$ O$_2$, suggesting the probable formation of an intraloop A$_{38}$-C$_{32}$ noncanonical base pair as observed in the solution structure (Figure 9). The backbone of the ASL loop in solution appears to be slightly wider than that in the crystal structure. This discrepancy in loop size could be attributed to bulk solvent differences between the solution and ribosome crystal states that influence the backbone dynamics. Also, the Mg$^{2+}$ ion concentrations used during crystallization may contribute to phosphate backbone shielding. As a result, a more compact loop is obtained. This is not the case in the bulk solvent where effective ion shielding was achieved mostly by monovalent Na$^+$ and K$^+$.

Both the solution and crystal structures of the ASL had an abrupt turn at the wobble position (Figure 9). The ASL in the four crystal structures exhibited the canonical U-turn but
also displayed an additional hydrogen bond that must be unique to $\text{xoyo}^5\text{U}_{34}$ modifications. All three hydrogen bonds defining the presence of a canonical U-turn, $\text{U}_{33}\text{N3H} \cdots \text{U}_{36}\text{O2P}$, $\text{U}_{33}\text{O2'} \cdots \text{A}_{35}\text{N7}$, and $\text{A}_{35}\text{NH2} \cdots \text{U}_{33}\text{O2'}$, were observed in the crystals (46). However, the invariant $\text{U}_{33}\text{O2'}$ within all four crystal structures was also within hydrogen bonding distance (3.04 (0.20 Å) of the ether oxygen (O5) of cmo$^5\text{U}_{34}$). Though the $\text{U}_{33}\text{O2'}$-cmo$^5\text{U}_{34}$ O5 distance in the solution structure exceeded that of hydrogen bonding, the observation in the crystals is of particular interest with regard to the expanded decoding capabilities of tRNAs with $\text{xoyo}^5\text{U}_{34}$ (46). The U-turn and $\pi$-like turn hydrogen bonding within the crystal and solution structures of the modified ASL provided more stability to the turn by restraining the anticodon dynamics. A low variation of standard deviation was found for $\text{U}_{33}\text{N3} \cdots \text{U}_{36}\text{O2P}$, $\text{U}_{33}\text{O2'} \cdots \text{A}_{35}\text{N7}$, and $\text{A}_{35}\text{N6} \cdots \text{C}_{32}\text{O2}$ hydrogen bonds, indicating a positional restraint of these bases as opposed to cmo$^5\text{U}_{34}$ for which a slight repositioning is observed between the unbound and codon-bound structures. Obviously, only a subtle remodeling of the anticodon wobble position cmo$^5\text{U}_{34}$ relative to $\text{U}_{33}$ has occurred upon A-site binding.

The conformation of cmo$^5\text{U}_{34}$ whether in the ASL$^{\text{Val3}_{\text{UAC}}}$-cmo$^5\text{U}_{34}$; m$^6\text{A}_{37}$ structure in solution or in the crystal contrasted sharply with the conformation found for the mononucleoside 5'-phosphate in solution. cmo$^5\text{U}_{34}$ adopted the C3'-$\text{endo}$ pucker conformation in the solution structure of the ASLs and in all four crystal structures of ASL$^{\text{Val3}_{\text{UAC}}}$-cmo$^5\text{U}_{34}$; m$^6\text{A}_{37}$. In contrast, the mononucleoside 5'-phosphate of cmo$^5\text{U}$ (and other oxy$^5$ derivatives, pxo$^5\text{U}$) in solution adopted the C2'-$\text{endo}$ form which was not observed when the 5'-phosphate group was removed (15). Neither AASL$^{\text{Val3}_{\text{UAC}}}$-cmo$^5\text{U}_{34}$ in solution nor ASL$^{\text{Val3}_{\text{UAC}}}$-cmo$^5\text{U}_{34}$; m$^6\text{A}_{37}$ in the four crystals exhibited a hydrogen bond
between cmo\(^5\) and the nucleoside’s 5’-PO\(_4\) group. In contrast to the solution and crystal structures of the ASLs, the conformations of the xo\(^5\)U mononucleotides indicated that the 5-substituent was interacting with the 5’-phosphate (15). Thus, a cautious deduction of modified nucleoside/nucleotide contributions to RNA structure is warranted in interpreting modified mononucleoside and mononucleotide conformations.

**CONCLUSION**

After four decades of investigation, we are beginning to establish the critical chemical and physical contributions of modified nucleotides to the biology of tRNA’s decoding of the genome (9). Crick postulated in the wobble hypothesis that the first two base pairs, and in some cases the third base pair, between tRNA’s anticodon and the mRNA codon would be canonical, that pyrimidine-pyrimidine base pairs would not occur, and that the ribosome would conform the anticodon to the demands of the codon (10). He also proposed that G34 was able to form a noncanonical base pair with U3, and vice versa, and that the purine I34 would pair with U, C, and A, thus expanding the ability of tRNA to read multiple codons. Following this hypothesis, studies *in vitro* and *in vivo* (14–16, 18, 44) demonstrated that tRNAs with cmo\(^5\)U\(_{34}\) were able to pair with A and G, but also with U and C. The enhanced hyperchromicities derived from the UV-monitored thermal transitions processes, the increased ellipticities of the CD spectra, the base stacking and base step parameters observed in the solution structures, and their comparison to X-ray analyses all indicate that modifications order and, thus, prestructure the anticodon domain. NMR relaxation studies of stable isotope-labeled ASLs should corroborate this conclusion. The A-site entry of a tRNA
with an unmodified U$_{34}$ and an unstructured anticodon domain would impart an energetic penalty to codon binding that the ribosome would need to overcome to yield a productive pairing. In contrast, a modified U$_{34}$ in prestructuring the anticodon would not require the ribosome to invest in remodeling the anticodon. These and previous observations clearly suggest that the distinctive chemistry, structure, and dynamics of the modification at position 34 together with those at purine 37 achieve either a restricted or an expanded codon reading through an entropically derived loop structure that relieves the ribosome from having to remodel the anticodon.

ACKNOWLEDGEMENTS

We thank Dr. Glenn Björk for the HPLC analyses, Dr. Hanna Gracz for her technical support at NCSU NMR facility, and Mr. Song-oh Han and Ms. Virginia Moye for their contribution to the UV and CD experiments, and Drs. Peter Moore and Jody Puglisi for suggestions to the manuscript.

SUPPORTING INFORMATION AVAILABLE

Superposition of the 1H-31P HETCOR spectra of ASL$^{\text{Val3}_{UAC}}$ and ASL$^{\text{Val3}_{UAC-cmo5U34}; \text{m6A37}}$ (Figure 1), anomeric to aromatic NOE connectivity for ASL$^{\text{Val3}_{UAC}}$ and ASL$^{\text{Val3}_{UAC-cmo5U34}; \text{m6A37}}$ (Figure 2), local base step parameters of unmodified ASL$^{\text{Val3}_{UAC}}$ (Figure 3), and local base step parameters of ASL$^{\text{Val3}_{UAC-cmo5U34}; \text{m6A37}}$ (Figure 4). This material is available free of charge via the Internet at http://pubs.acs.org.
REFERENCES


### Table 1: Thermal parameters of unmodified and modified ASL\textsuperscript{Val3\textsubscript{UAC}}

<table>
<thead>
<tr>
<th>ASL\textsuperscript{Val\textsubscript{UAC}}</th>
<th>Tm (ºC)</th>
<th>ΔG (Kcal/mole)</th>
<th>Hyperchromicity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodified</td>
<td>71.2 ± 1.1</td>
<td>-6.0 ± 0.4</td>
<td>5.0 ± 0.2</td>
</tr>
<tr>
<td>cmo\textsuperscript{5}U\textsubscript{34}</td>
<td>72.3 ± 1.3</td>
<td>-5.4 ± 0.3</td>
<td>10.0 ± 0.2</td>
</tr>
<tr>
<td>m\textsuperscript{6}A\textsubscript{37}</td>
<td>71.7 ± 0.9</td>
<td>-5.6 ± 0.2</td>
<td>10.1 ± 0.2</td>
</tr>
<tr>
<td>cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37}</td>
<td>71.8 ± 1.3</td>
<td>-5.3 ± 0.3</td>
<td>11.0 ± 0.2</td>
</tr>
</tbody>
</table>

### Table 2: Affinity of unmodified and modified ASL\textsuperscript{Val3\textsubscript{UAC}} for the valine codons.

<table>
<thead>
<tr>
<th>A-site Codon</th>
<th>Unmodified</th>
<th>m\textsuperscript{6}A\textsubscript{37}</th>
<th>cmo\textsuperscript{5}U\textsubscript{34};m\textsuperscript{6}A\textsubscript{37}</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUU</td>
<td>&gt;2000</td>
<td>&gt;2000</td>
<td>1.93 ± 0.70</td>
</tr>
<tr>
<td>GUA</td>
<td>0.50 ± 1.1</td>
<td>0.42 ± 0.05</td>
<td>0.17 ± 0.03</td>
</tr>
<tr>
<td>GUG</td>
<td>3.65 ± 1.28</td>
<td>2.46 ± 0.58</td>
<td>1.96 ± 0.32</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1: Nucleotide sequence of the *Escherichia coli* tRNA\textsuperscript{Val3\textsubscript{UAC}} anticodon stem and loop (ASL\textsuperscript{Val3\textsubscript{UAC}}) and its modified nucleosides. The nucleotide sequence and secondary structure of the ASL\textsuperscript{Val3\textsubscript{UAC}} has two modifications, uridine-5-oxyacetic acid at wobble position 34 (cmo\textsuperscript{5}U\textsubscript{34}) and N6-methyladenosine at position 37 (m\textsuperscript{6}A\textsubscript{37}). The oxyacetic acid is dissociated at physiological pH. The oxyacetic acid and methyl groups are colored red.

Figure 2: UV-monitored, thermal transitions and circular dichroism specta of unmodified and modified ASL\textsuperscript{Val3\textsubscript{UAC}}. (a) Thermal transitions of modified and unmodified ASL\textsuperscript{Val3\textsubscript{UAC}}. Unmodified ASL\textsuperscript{Val3\textsubscript{UAC}} (red line) and ASL\textsuperscript{Val3\textsubscript{UAC}} doubly modified with both cmo\textsuperscript{5}U\textsubscript{34} and m\textsuperscript{6}A\textsubscript{37} (blue line) were subjected to repeated thermal denaturations and renaturations monitored at 260 nm and under conditions identical to that for NMR analysis of structure. The melting temperatures for the ASLs were similar and high (71.5°C) due to the G-C rich stem region. This also demonstrated that the stems were very stable and, therefore, well-structured. The hyperchromicity measurements, denoted for unmodified ASL\textsuperscript{Val3\textsubscript{UAC}}, differed considerably (Table 1). (b) Circular dichroism (CD) spectra of unmodified and modified ASL\textsuperscript{Val3\textsubscript{UAC}}. CD spectra of unmodified ASL\textsuperscript{Val3\textsubscript{UAC}} (red line) and ASL\textsuperscript{Val3\textsubscript{UAC}}- cm\textsuperscript{o5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37} (blue line) were collected under conditions identical to those used for structure determination by NMR. The introduction of modifications increased the intensity of the ellipticity at 260 nm: unmodified ASL\textsuperscript{Val3\textsubscript{UAC}} < ASL\textsuperscript{Val3\textsubscript{UAC}}- cm\textsuperscript{o5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}. The green line is that for buffer alone.
Figure 3: Binding of unmodified and modified ASL\textsuperscript{Val3}\textsubscript{UAC} to valine codons at the A-site of \textit{E. coli} 30S ribosomal subunits. Codons in the A-site of the 30S subunit were titrated with increasing concentrations of radiochemically, 5'-end \textsuperscript{32}P-labeled ASLs (0-7.5 \textmu M). The 30S subunits (0.5 \textmu M) were programmed with mRNA (50 \textmu M) and saturated with \textit{E. coli} tRNA\textsuperscript{fMet} (1-2.5 \textmu M). The binding curves were analyzed with single-site, nonlinear regression and the dissociation constants (\(K_d\)) calculated (see Table 2). (a) Binding to the GUA codon: unmodified ASL\textsuperscript{Val3}\textsubscript{UAC} (■) and ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} (▲). (b) Binding to the GUG codon: unmodified ASL\textsuperscript{Val3}\textsubscript{UAC} (■) and ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} (▲). Only ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} was found to bind the GUG codon. (c) Binding to the GUU codon: unmodified ASL\textsuperscript{Val3}\textsubscript{UAC} (■) and ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}}. Only ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}} was found to bind the GUU codon (▲).

Figure 4: Detection of the base-paired imino protons of the ASLs by NMR. The base-paired, imino proton region of the 1D \textsuperscript{1}H NMR spectra of unmodified and modified ASL\textsuperscript{Val3}\textsubscript{UAC} exhibited five lowfield-shifted resonances between 12.00 and 14.50 ppm: (a) ASL\textsuperscript{Val3}\textsubscript{UAC} and (b) ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}}. The chemical shift dispersion observed in this region is characteristic of imino protons engaged in Watson-Crick base pairing. As expected, the number of peaks identified in this region corresponded exactly to the number of base pairs present in the stem of the four ASLs. The imino proton resonance of U\textsubscript{34} is more resolved in the spectrum of ASL\textsuperscript{Val3}\textsubscript{UAC- cmo\textsuperscript{5}U\textsubscript{34}; m\textsuperscript{6}A\textsubscript{37}}, indicating that on the NMR time scale the rate of exchange with the solvent is decreased and suggesting that the decreased exchange rate is induced by an increase in the level of base stacking.
Figure 5: Superimposed $^1$H-$^1$H COSY spectra of unmodified and modified ASL$_{UAC}^{Val^3}$. Spectra were collected under identical conditions for ASL$_{UAC}^{Val^3}$ (red) and ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ (blue) (2H2O, 500 MHz). The spectral region corresponding to the throughbond correlation between the pyrimidine H5 and H6 protons demonstrates significant homologies, with the notable exceptions of U$_{34}$ for ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$. Each peak is labeled with its corresponding residue where the H5-H6 correlation is observed.

Figure 6: Determination of ribose conformation from the $^1$H DQFCOSY spectra. The H2'-H1' regions of DQF-COSY spectra of ASL$_{UAC}^{Val^3}$ (red) and ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ (blue) are superimposed and were used to assess the sugar pucker of the individual nucleotides. The C2'-endo conformation is adopted by residues G$_{43}$, A$_{35}$, and U$_{34}$/cmo$^5$U$_{34}$.

Figure 7: Solution structures of ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ and ASL$_{UAC}^{Val^3}$ and the rmsd for each of their nucleotides. Families of structures were derived from torsion-angle-restrained, molecular dynamics calculations and then averaged to yield the structures of (a) ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ and (b) ASL$_{UAC}^{Val^3}$ with no violations. The modified residues (cmo$^5$U$_{34}$ and m$^6$A$_{37}$) of ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ are labeled. The alternate position of the m$^6$A$_{37}$ is depicted with the nucleotide colored dark gray. (c) Individual nucleotide average rmsd for unmodified ASL$_{UAC}^{Val^3}$ (red) and ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$ (blue). The rmsds of all heavy atoms (angstroms) from the mean are plotted for each of the nucleotides of the two ASLs. The blue dashed curve corresponds to the rmsd of seven ASL$_{UAC}^{Val^3}$-cmo$^5$U$_{34}$; m$^6$A$_{37}$
structures in which one orientation of the methyl group of m\(^6\)A\(_{37}\) is taken into account, above the plane of A\(_{37}\).

**Figure 8:** Superimposition of the average structures of the loop residues of ASL\(^{\text{Val3}}\)\(^{\text{UAC-cmo5U34;m6A37}}\) with that of ASL\(^{\text{Val3}}\)\(^{\text{UAC}}\). The seven nucleotides of the loop differ in the degree of stacked helical conformation and base orientation from the 5' to 3'-end of the loops. The stereochemical hindrance of the methyl group constrained the m\(^6\)A\(_{37}\) nucleotide to a position translated outward from the loop (1.6 Å) relative to that of A\(_{37}\) of the unmodified ASL\(^{\text{Val3}}\)\(^{\text{UAC}}\) (see the dashed arrow).

**Figure 9:** Loop structure of ASL\(^{\text{Val3}}\)\(^{\text{UAC-cmo5U34;m6A37}}\) in solution compared to that of the crystallographic structure on the ribosome. The restrained molecular dynamics structure of ASL\(^{\text{Val3}}\)\(^{\text{UAC-cmo5U34;m6A37}}\) in solution (red) is superimposed upon the X-ray crystallographic structure (pink). ASL residues C\(_{32}-G_{40}\) are shown. The crystal structures of the ASL bound to codon GUA and the other three valine codons were determined at a resolution of 2.8 Å.
**FIGURES**

**Figure. 1:** Nucleotide sequence of the *Escherichia coli* tRNA\textsubscript{Val}\textsuperscript{UAC} anticodon stem and loop (ASL\textsubscript{Val}\textsuperscript{UAC}) and its modified nucleosides.
Figure 2: UV-monitored, thermal transitions and circular dichroism spectra of unmodified and modified \( \text{ASL}^{Val3}_{UAC} \).
Figure 3: Binding of unmodified and modified ASL $^\text{Val}_3$ to valine codons at the A-site of $E.\ coli$ 30S ribosomal subunits.
Figure 4: Detection of the base-paired imino protons of the ASLs by NMR. The base-paired, imino proton region of the 1D $^1$H NMR spectra of unmodified and modified ASL$_{Val^3UAC}^{Val^3UAC}$ exhibited five lowfield-shifted resonances between 12.00 and 14.50 ppm: (a) ASL$_{Val^3UAC}^{Val^3UAC}$ and (b) ASL$_{UAC^3cmo^5U_{34};m^6A_{37}}^{Val^3UAC}$.
Figure 5: Superimposed $^1$H-$^1$H COSY spectra of unmodified and modified ASL$^{\text{Val}_3}_{\text{UAC}}$. 
Figure 6: Determination of ribose conformation from the $^1$H DQFCOSY spectra.
Figure 7: Solution structures of ASL\textsuperscript{Val\_UAC} \textsuperscript{cmo5}U\textsubscript{34}; \textsuperscript{m6}A\textsubscript{37} and ASL\textsuperscript{Val\_UAC} and the rmsd for each of their nucleotides.
Figure 8: Superimposition of the average structures of the loop residues of ASL$^\text{Val}_3^\text{UAC}$-cmo$^5^\text{U}_{34}$; m$^6^\text{A}_{37}$ with that of ASL$^\text{Val}_3^\text{UAC}$. 
Figure 9: Loop structure of ASL$^{\text{Val}_{\text{UAC}^{-\text{cmo}}{5}U_{34}m^6A_{37}}}$ in solution compared to that of the crystallographic structure on the ribosome.
Figure 1: Superposition of the $^1$H-$^31$P HETCOR spectra of ASL$^{Val_3}_{UAC}$ (red) and ASL$^{Val_3}_{UAC}$-cmo$^5$U$^{m_6}$A$^{37}$ (blue). The arrow indicates that the phosphorus of the U$_{33}$ H$^3'$phosphate is downfield shifted which is characteristic of a U-turn.
Figure 2: Anomeric to aromatic connectivity for ASL$^{Val}_{UAC}$ and ASL$^{Val}_{UAC}$-cmo$^5$U$_{34}$;m$^6$A$_{37}$. The two dimensional, $^1$H NOESY spectra of (a) ASL$^{Val}_{UAC}$ and (b) ASL$^{Val}_{UAC}$-cmo$^5$U$_{34}$;m$^6$A$_{37}$ were collected to determine the intra-H1'-aromatic sequence connectivity for the ASLs. The cross-peak identification is designated in the figure. The U$_{33}$H1'-A$_{35}$H8 NOE, a hallmark of the canonical U-turn, is highlighted in dashed green line. Introduction of modifications at positions 34 and 37 caused the H1' protons of the loop residues 34, 35, 36 and 37 to alter their chemical shifts in the F2 dimension (vertical black dashed lines).
Table 1: Local base step parameters of the unmodified ASL$^{\text{Val}_{\text{UAC}}}$. The loop nucleosides are shaded. The absolute values of the angles were used to determine the average (aver.) and the standard deviation (s.d.).

<table>
<thead>
<tr>
<th>Step</th>
<th>Shift</th>
<th>Slide</th>
<th>Rise</th>
<th>Tilt</th>
<th>Roll</th>
<th>Twist</th>
</tr>
</thead>
<tbody>
<tr>
<td>C27/C28</td>
<td>-0.27</td>
<td>-1.61</td>
<td>3.38</td>
<td>-1.93</td>
<td>12.39</td>
<td>30.67</td>
</tr>
<tr>
<td>C28/U29</td>
<td>-0.30</td>
<td>-1.71</td>
<td>3.59</td>
<td>-3.85</td>
<td>8.63</td>
<td>29.11</td>
</tr>
<tr>
<td>U29/C30</td>
<td>0.41</td>
<td>-1.69</td>
<td>3.69</td>
<td>-1.18</td>
<td>-3.05</td>
<td>34.90</td>
</tr>
<tr>
<td>C30/C31</td>
<td>-0.27</td>
<td>-1.67</td>
<td>3.24</td>
<td>0.22</td>
<td>10.76</td>
<td>30.51</td>
</tr>
<tr>
<td>C31/C32</td>
<td>4.84</td>
<td>-0.31</td>
<td>4.54</td>
<td>-0.06</td>
<td>-14.40</td>
<td>64.05</td>
</tr>
<tr>
<td>C32/U33</td>
<td>3.20</td>
<td>-2.09</td>
<td>6.07</td>
<td>-32.18</td>
<td>29.84</td>
<td>61.57</td>
</tr>
<tr>
<td>U33/U34</td>
<td>-10.02</td>
<td>-3.75</td>
<td>-7.94</td>
<td>65.39</td>
<td>138.29</td>
<td>-37.34</td>
</tr>
<tr>
<td>U34/A35</td>
<td>0.11</td>
<td>-1.15</td>
<td>4.48</td>
<td>-13.66</td>
<td>4.87</td>
<td>88.73</td>
</tr>
<tr>
<td>A35/C36</td>
<td>-0.22</td>
<td>-1.53</td>
<td>2.75</td>
<td>18.39</td>
<td>4.73</td>
<td>30.59</td>
</tr>
<tr>
<td>C36/A37</td>
<td>5.79</td>
<td>-0.63</td>
<td>2.85</td>
<td>10.30</td>
<td>11.52</td>
<td>64.58</td>
</tr>
<tr>
<td>A37/A38</td>
<td>-0.79</td>
<td>-0.95</td>
<td>2.56</td>
<td>20.36</td>
<td>-10.94</td>
<td>27.06</td>
</tr>
<tr>
<td>A38/G39</td>
<td>2.93</td>
<td>-1.22</td>
<td>4.50</td>
<td>-11.05</td>
<td>10.30</td>
<td>48.54</td>
</tr>
<tr>
<td>G39/G40</td>
<td>0.04</td>
<td>-1.75</td>
<td>3.38</td>
<td>-3.11</td>
<td>11.40</td>
<td>33.07</td>
</tr>
<tr>
<td>G40/A41</td>
<td>-0.13</td>
<td>-1.52</td>
<td>3.67</td>
<td>-0.33</td>
<td>-4.06</td>
<td>29.78</td>
</tr>
<tr>
<td>A41/G42</td>
<td>0.06</td>
<td>-1.92</td>
<td>3.46</td>
<td>-3.42</td>
<td>11.75</td>
<td>32.06</td>
</tr>
<tr>
<td>G12/G43</td>
<td>0.12</td>
<td>-1.80</td>
<td>3.30</td>
<td>-1.35</td>
<td>9.10</td>
<td>31.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ASL aver.</th>
<th></th>
<th>s.d.</th>
<th></th>
<th>loop aver.</th>
<th></th>
<th>stem aver.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.41</td>
<td>1.58</td>
<td>3.96</td>
<td>11.67</td>
<td>18.50</td>
<td>42.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.34</td>
<td>0.75</td>
<td>1.37</td>
<td>17.01</td>
<td>32.53</td>
<td>18.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.58</td>
<td>1.62</td>
<td>4.45</td>
<td>24.47</td>
<td>30.07</td>
<td>51.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>1.55</td>
<td>3.50</td>
<td>1.72</td>
<td>9.50</td>
<td>35.12</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4

Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNA\textsuperscript{Met}

HRVOJE LUSIC,\textsuperscript{1*} and ESTELLA M. GUSTILO,\textsuperscript{2*} FRANCK A.P. VENDEIX,\textsuperscript{2} ROB KAISER,\textsuperscript{3} MICHAEL O. DELANEY,\textsuperscript{3} WILLIAM D. GRAHAM,\textsuperscript{2} VIRGINIA A. MOYE,\textsuperscript{2} WILLIAM A. CANTARA,\textsuperscript{2} PAUL F. AGRIS,\textsuperscript{2,*} and ALEXANDER DEITERS\textsuperscript{1,*}

\textsuperscript{1Department of Chemistry, \textsuperscript{2Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA, and \textsuperscript{3Dharmacon, 2650 Crescent Drive #100, Lafayette, CO 80026, USA.}

Published in Nucleic Acids Research, 2008, Vol. 36, p. 6548–6557
Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNA^Met

HRVOJE LUSIC,1† and ESTELLA M. GUSTILO,2† FRANCK A.P. VENDEIX,2 ROB KAISER,3 MICHAEL O. DELANEY,3 WILLIAM D. GRAHAM,2 VIRGINIA A. MOYE,2 WILLIAM A. CANTARA,2 PAUL F. AGRIS,2,* and ALEXANDER DEITERS1,*

1Department of Chemistry, 2Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA, and 3Dharmacon, 2650 Crescent Drive #100, Lafayette, CO 80026, USA.

Running Head: 5-Formylcytidine modified mitochondrial tRNA^Met.

Keywords: tRNA, 5-formylcytidine, RNA synthesis, translation, wobble position, mitochondria

†HL and EMG contributed equally to this research.

*To whom correspondence should be addressed:
Paul F. Agris, Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA, 919-515-6188; Paul_Agris@ncsu.edu; and

Alex Deiters, Department of Chemistry, North Carolina State University, Raleigh, NC 27695, USA, 919-513-2958; Alex_Deiters@ncsu.edu

95
ABSTRACT

Human mitochondrial methionine transfer RNA (hmtRNA<sup>Met<sub>CAU</sub></sup>) has a unique posttranscriptional modification, 5-formylcytidine, at the wobble position-34 (f<sup>5</sup>C<sub>34</sub>). The role of this modification in hmtRNA<sup>Met<sub>CAU</sub></sup> for the decoding of AUA, as well as AUG, in both the peptidyl- and aminoacyl-sites of the ribosome in either chain initiation or chain elongation is still unknown. We report the first synthesis and analyses of the tRNA’s anticodon stem and loop domain containing the 5-formylcytidine modification. The modification contributes to the tRNA’s anticodon domain structure, thermodynamic properties, and its ability to bind codons AUA and AUG in translational initiation and elongation.

INTRODUCTION

Mitochondria generate over 90% of the energy used by mammalian cells through oxidative phosphorylation. Thirteen proteins, components of the electron transfer chain and the ATP synthase, are the products of mitochondrial DNA. The synthesis of these proteins is carried out by a specific protein synthesizing machinery within this organelle. In some three decades since the sequencing of the human mitochondrial genome (1), the mitochondrial genetic code has been found to differ significantly from the universal code. The human mitochondrial gene for the one methionine specific tRNA (hmtRNA<sup>Met<sub>CAU</sub></sup>, where CAU is the anticodon) plays a unique role since it must provide the tRNA used for both the initiation of protein synthesis and the elongation of the protein chain by responding to the codon AUA, normally an isoleucine codon in the cytoplasm, as well as the universal methionine code,
AUG. This is highly unusual since all cytoplasmic protein biosynthetic systems employ two different tRNA\textsuperscript{Met} species, one for initiation and one for elongation, and both respond to the single methionine codon, AUG. Maternally inherited mutations in the gene of this tRNA, including an A\textsubscript{37} to G\textsubscript{37} mutation adjacent to the anticodon nucleosides that read the two codons (Figure 1), are responsible for some devastating diseases (2-5). Moreover, the hmtRNA\textsuperscript{Met\_CAU} has a unique modification, 5-formylcytidine (Figure 1), at the wobble position-34 (f\textsuperscript{5}C\textsubscript{34}) seen only in one other tRNA, a bovine liver, cytoplasmic tRNA\textsuperscript{Leu} with a f\textsuperscript{5}C\textsubscript{34} further modified with a 2’O-methyl (6). Nothing, however, is known about the decoding characteristics of tRNA\textsuperscript{Leu\_f 5\textsubscript{CmAA}}. Since its discovery in bovine and nematode mitochondrial tRNA\textsuperscript{Met} in 1994 (7;8), f\textsuperscript{5}C\textsubscript{34} also has been found in the mitochondrial tRNA\textsuperscript{Met} of squids, frogs, chickens, rats, and Drosophila (9-11). The contribution of the f\textsuperscript{5}C\textsubscript{34} modification to the structure of the hmtRNA\textsuperscript{Met\_CAU}, its role in the decoding of AUG and AUA, and its possible participation in either chain initiation or chain elongation by this unique tRNA\textsuperscript{Met} is still unknown.

We speculate that mitochondria have a unique mechanism to partition this single hmtRNA\textsuperscript{Met\_CAU} species between initiation and elongation. While a tRNA\textsuperscript{Met\_CAU} unmodified at wobble position-34 can read AUG, we hypothesize that the 5-formyl modification allows one tRNA\textsuperscript{Met} to expand codon reading to include AUA in the mitochondria. Towards proving the hypothesis that this wobble modification affords the single tRNA\textsuperscript{Met} the ability to decode AUG and AUA, we are reporting the first synthesis of a 5-formylcytidine-modified RNA and the initial structural and biological investigations. Previously, bovine mitochondrial tRNA\textsuperscript{Met\_CAU} had been shown to translate AUG and AUA in an E. coli translational system in
vitro, where AUG coded for methionine and AUA for isoleucine (12). However, site specific binding was not investigated. Here we compare the codon binding affinities of the f5C34-modified anticodon stem and loop of human mitochondrial tRNA<sup>Met</sup> (hmtASLMet-f5C34) with that of the unmodified ASL. Our results for both AUA and AUG codons at both the A and P-site of *E. coli* ribosomes increase our understanding of the modification’s contributions to decoding and are consistent with previous results from the translation of poly(AUA) by bovine mtRNA<sup>Met</sup><sub>CAU</sub> on *E. coli* ribosomes (12).

**MATERIALS AND METHODS**

**Experimental procedures and analytical data for the synthesis of the f5C phosphoramidite (9)**

All reagents used in the following experiments are of the highest purity and dryness possible. Before use, glassware was thoroughly cleaned and dried (oven at 110°C for 30 minutes). NMR analysis of intermediates were conducted in the appropriate deuterated solvent (referenced accordingly for CDCl<sub>3</sub>: 1H 7.24 ppm, 13C 77.23 ppm; and CD<sub>3</sub>CN: 1H 1.94 ppm) using a Bruker Avance Ultrashield 300 MHz spectrophotomer. Phosphorus, 31P, NMR experiments were referenced according to an external H<sub>3</sub>PO<sub>4</sub> standard (0.00 ppm). Mass-spec analysis of the samples was performed on a Micromass LCT ESI-TOF or an Agilent LC-TOF. Analytes were dissolved in acetonitrile and flown against a Leucin Enkephalin lock mass standard. Chemical and physical properties of the intermediates were those of the expected compounds (Supplementary data).
2′,3′-O-isopropylidenecytidine (2). To a suspension of cytidine (1, Figure 2, 1 g, 4.1 mmol) in 50 mL of acetone was added 2,2-dimethoxypropane (6 mL, 5.1 g, 49 mmol). HClO₄ was then added dropwise until the solution turned clear. The reaction mixture was stirred at room temperature for 12 h, after which it was neutralized by addition of Ca(OH)₂, filtered and evaporated. Purification by silica gel chromatography using CHCl₃:MeOH (80:20) with 2% TEA, afforded 966 mg of 2 as a white foam (83% yield). The analytical data obtained matched known literature data for 2 (13).

5-(Hydroxymethyl)-2′,3′-O-isopropylidenecytidine (3). To a solution of 2 (Figure 2, 1 g, 3.5 mmol) in 15 mL of 0.5 M KOH was added paraformaldehyde (1.05 g, 35 mmol). The reaction was stirred at 55 ºC for 36 h, after which it was cooled to room temperature and neutralized with 6M HCl. The solution was filtered and evaporated. The oily residue was dissolved in MeOH:DCM (40:60), filtered and evaporated again. Purification by silica gel chromatography using MeOH:DCM (gradient 3:97, 8:92, 12:88) with 2% TEA, afforded 416mg of 5-(hydroxymethyl)-2′,3′-O-isopropylidenecytidine (3) as a white foam (38% yield, 60% yield based on 357mg of recovered starting material 2′,3′-O-isopropylidenecytidine (2).

5-Formyl-2′,3′-O-isopropylidenecytidine (4). To 3 (Figure 2, 100 mg, 0.32 mmol) in 3 mL dioxane was added 500 mg ruthenium dioxide hydrate (5 wt.eq.). The reaction mixture was refluxed for 12 h and filtered. Purification by silica gel chromatography MeOH:DCM (5:95) with 2% TEA, afforded 81mg of the 5-formyl-2′,3′-O-isopropylidenecytidine (4) as a white solid (82% yield).
5-Formylcytidine (5). The acetonide, 5-formyl-2’,3’-O-isopropylidene-2’-O-isopropylidene-cytidine (4, Figure 2, 500mg, 1.61 mmol) was suspended in 1M HCl (15mL) at room temperature. The reaction progress was monitored by TLC. Upon disappearance of the starting material, the solution was neutralized with TEA. Water was subsequently evaporated. Recrystallization from MeOH afforded 414mg of 5FC (5) as a white solid (95% yield).

Nº-[Diisobutylamino)methylidene]-3’,5’-O-(1,1,3,3-tetraisopropyl-1,3-disiloxanediyl)-5-formylcytidine (6). 5-Formylcytidine (5, Figure 2, 0.47 g, 1.72 mmol) was dissolved in a mixture of 20 mL of pyridine and 2 mL of DMF. The solution was cooled to 0 ºC and TIPDSCl₂ (0.59 g, 1.89 mmol) in 2 mL of pyridine was added dropwise over a period of one hour. The reaction was allowed to gradually warm to room temperature overnight. The following morning the reaction was quenched with 5 mL of MeOH and evaporated to dryness. The resulting paste was coevaporated twice with 20 mL of toluene and the crude material was purified by flash chromatography on 30 mL of silica gel using a gradient of MeOH in DCM (3-4%). Product fractions were pooled and evaporated to afford the TIPDS protected intermediate (0.88 g, 100%) as light yellow oil that is contaminated with residual pyridinium salts. The above compound was used as is without further purification to remove the residual pyridinium salts. TIPDS protected 5FC (5) (0.88 g, 1.72 mmol) was dissolved in 20 mL DMF and N,N-diisobutylformamidine dimethyl acetal (14) (0.70 g, 3.44 mmol) was added. The reaction was stirred for 16 h and evaporated under high vacuum. The resulting loose oil was coevaporated twice with 20 mL of toluene and the crude material was purified
by flash chromatography on 30 mL of silica gel using a gradient of MeOH in DCM (1-2%). Product fractions were pooled and evaporated to afford 0.80 g of 6 as light yellow oil in 71% overall yield from f5C (5).

2′-O-[Bis(2-acetoxyethoxy)methyl]-N4-[(diisobutylamino)methylidene]-5-formylcytidine (7). A mixture of 6 (Figure 2, 0.80 g, 1.23 mmol), pyridinium para-toluenesulfonate (0.31 g, 1.23 mmol), and Tris(2-acetoxyethoxy)methyl orthoformate (1.98 g, 6.15 mmol) was dissolved in 5 mL of DCM and stirred at room temperature. After 2 days, TBDMS-pentanedione (0.53 g, 2.46 mmol) was added and the reaction was stirred at ambient temperature. After stirring for an additional day, the reaction was quenched with 1 mL of TEMED. The crude material was separated from excess reagents by flash chromatography on 50 mL silica gel using a gradient of 25% ethyl acetate in hexanes with 0.1% TEMED to 50% ethyl acetate in hexanes with 0.1% TEMED. This material was concentrated to near dryness and taken directly onto the desilylation reaction.

A freshly made solution of TEMED (0.71 g, 6.15 mmol) in 10 mL of acetonitrile at 0 ºC was added 48% HF (0.15 mL, 4.30 mmol). This solution was allowed to stir for 5 min and added to the foregoing material from above at room temperature. The reaction was stirred for 2 h and concentrated to dryness. The crude material was purified by flash chromatography on 50 mL silica gel using a gradient of 20% hexanes in ethyl acetate with 0.1% TEMED to 1% methanol in ethyl acetate with 0.1% TEMED. Product fractions were pooled and evaporated to leave the 2′-O-protected compound [7] as a light yellow oil (0.41 g) in 53% yield from the nucleobase protected compound 6.
5′-O-[Benzydryloxy-bis(trimethylsilyloxy)silyl]-2′-O-[bis(2-acetoxyethoxy)-methyl]-N4-[(diisobutyl-amino)methylidene]-5-formylcytidine (8). Diisopropylamine (0.07 g, 0.65 mmol) was added to a solution of the 2′-O- and N4-protected nucleoside, 7 (Figure 2, 0.41 g, 0.65 mmol) in 7 mL of DCM and the solution was cooled to 0 °C. In a separate flask BZHCl (0.34 g, 0.81 mmol) was diluted in 5 mL of DCM. Diisopropylamine (0.10 g, 0.98 mmol) was added to the silylating solution and the solution was allowed to stir for 2 min before being added dropwise to the nucleoside solution. The addition was completed within 30 min and the reaction was allowed to for 3 h and the reaction was quenched with 1 mL of MeOH and evaporated to dryness. The crude material was purified by flash chromatography on 30 mL silica gel using a gradient of 10% acetone in hexanes containing 0.1 % (v/v) TEA to 20% acetone in hexanes containing 0.1 % (v/v) TEA. Product fractions were pooled and evaporated to afford the 5′-O- protected compound 8 as a colorless oil. The yield was 0.56 g (84%).

5′-O-[Benzydryloxy-bis(trimethylsilyloxy)silyl]-2′-O-[bis(2-acetoxyethoxy)-methyl]-N4-[(diisobutyl-amino)methylidene]-5-formylcytidine-3′-(methyl-N,N-diisopropyl) phosphoramidite (9). Bis(diisopropyl-amino) methoxy phosphine (0.21 g, 0.82 mmol) was dissolved in 3 mL of DCM and a 0.5 M solution of 5-ethylthio-1-H-tetrazole in anhydrous acetonitrile (0.08 mL, 0.55 mmol) was added. Diisopropylamine (0.06 g, 0.55 mmol) was then added and the phosphine solution was allowed to stir for 5 min at ambient temperature. In a separate flask, the 2′-O-, 5′-O- and N4-protected f5C, compound 8 (Figure 2, 0.56 g, 0.55
mmol) and diisopropylamine (0.06 g, 0.55 mmol) were dissolved in 5 mL of DCM. The activated phosphine solution was added into the nucleoside solution and the reaction was stirred at room temperature. After 16 h the reaction was quenched with 2 mL of absolute ethanol and concentrated to dryness. The resulting white paste was purified by flash chromatography on 30 mL of silica gel using a mixture of DCM in hexanes (5:95 (v/v) containing 2% (v/v) TEA followed by acetone in hexanes (1:9 (v/v) to 2:8 (v/v) containing 0.5% (v/v) TEA. Product fractions were pooled and evaporated to afford the protected fC phosphoramidite, 9 (Figure 2), as a colorless oil.

**Polyribonucleotide synthesis of 5′-ΨCGGGCC-fC-AUCCCCGA-3′**

The above sequence was synthesized on a 1 μmol scale using a ABI 394 DNA synthesizer using previously published procedures (15;16). The fC phosphoramidite (9, 0.067 M in anhydrous acetonitrile) was coupled to the growing polyribonucleotide chain for 3.5 min using 5-ethylthio-1H-tetrazole (0.5 M in anhydrous acetonitrile) as the activator. Once the synthesis of the polyribonucleotide chain was completed, the phosphate protecting groups were removed from the immobilized polyribonucleotide by treatment with disodium 2-carbamoyl-2-cyanoethylene-1,1-dithiolate trihydrate in DMF for 10 minutes. The support was washed excessively with water for 5 min and then flushed with Argon gas for 5 min to dry the support. The support was then transferred to a 2 mL Eppendorf tube and the polyribonucleotide was cleaved from the support and the exocyclic amine protecting groups were removed with 1:3 (v/v) tert-butylamine:water for 6 hours at 60 °C. The sample was cooled to room temperature, filtered, and lyophilized to obtain the crude polyribonucleotide.
The hmASL Met CAU-Ψ_{27}\textsuperscript{f5C}_{34} was deprotected with acetate/TEMED according to standard Dharmacon protocols, purified by ion exchange HPLC (17), and dialyzed extensively against H\textsubscript{2}O. The hmASL Met CAU-Ψ_{27} was synthesized and deprotected under standard conditions (18).

**Confirmation of nucleoside composition by nucleoside HPLC and NMR of the hmASL Met CAU constructs**

Incorporation of f5C modification within the hmASL Met was confirmed by NMR, including the two-dimensional NOESY (Figures 3 and 4). The nucleoside composition of the hmASL Met products were confirmed by enzymatic hydrolysis of the RNA to its constituent nucleosides (17) and then subjected to HPLC monitored by diode array UV spectrometry on the fly, the peaks identified, integrated and quantified (19) (Figure 4).

**Analysis of thermodynamic stability, circular dichroism and molecular dynamics simulations**

The ASL samples were dissolved to obtain a concentration of ~4 µM in 20 mM Na-K phosphate buffer (pH 6.8). UV-monitored, thermal denaturations and re-naturations were replicated five times and monitored by measuring UV absorbance (260 nm) using a Cary 3 spectrophotometer as published (20;21). The data points were averaged over 20 seconds and collected four times a minute with a temperature change of 0.5 °C per minute from 4 - 90 °C. The data was analyzed (22), and the thermodynamic parameters were determined (Origin software, Microcal, Inc.) (Figure 5A). CD spectral ellipticity data were collected using a
Jasco 600 spectropolarimeter and an interfaced computer (Jasco, Inc., Easton, MD). hmASL$_{\text{CAU}}$-$\Psi$$_{27}$ or the hmASL$_{\text{CAU}}$-$\Psi$$_{27};$f$^{6}$C$_{34}$ (0.2 A260/mL, 20 mM Na-K phosphate buffer, pH 6.8) was placed in a temperature-regulated, 1 cm path-length quartz cell. Each sample was scanned 10 times at 25 °C. The final data is an average of the 10 scans (Figure 5B). The molecular dynamics simulation (MDS) were performed by following standard published protocol (23) with the exception of using a truncated octahedral TIP3P water box (24).

**Ribosomal Binding Assay**

The 27-mer mRNA oligos used in codon binding assays were designed from that of T4 gp32 mRNA (25) and purchased (Dharmacon RNA Technologies). They were chemically deprotected and HPLC-purified in our lab. Each mRNA sequence was entered into the program RNA Structure 4.2 (26) and was found to have low probability of folding into any stable conformation. The mRNA sequences are as follows (mitochondrial methionine codons AUA and AUG are in bold):

1) 5’-GGCAAGGAGGUAAGGAUUAGUAGCAGCU-3’;
2) 5’-GGCAAGGAGGUAAGGAUGGUAGCAGCU-3’;
3) 5’-GGCAAGGAGGUAAGGAAGUAAGAUAUGCACGU-3’;
4) 5’-GGCAAGGAGGUAAGGAAGUAUGGCACGU-3’.
The 70S ribosomal subunits were prepared from *E. coli* MRE600 (27). The ASLs were 5'-end $^{32}$P-labeled using [$\gamma$-$^{32}$P] ATP (MP Biomedicals). Unlabeled ASLs (5 $\mu$M) were mixed with 10000 CPM of 5'-end $^{32}$P-labeled ASL. The assay was performed in ribosomal binding buffer [50 mM HEPES, pH 7.0; 30 mM KCl; 70 mM NH$_4$Cl; 1 mM DTT; 100 $\mu$M EDTA; 20 mM MgCl$_2$]. The ribosomes, activated at 42 °C for 10 min and then slowly cooled to 37 °C, were then programmed with 2.5 $\mu$M mRNA for 15 min at 37 °C. The ribosomal site not in observation was saturated with ASL$^{\text{Val}}_{\text{UAC}}$ (unmodified) for 15 min at 37 °C. P-site binding was performed prior to A-site binding. ASL$^{\text{Val}}_{\text{UAC}}$ binds to the Val codon GUA; see underlined codons of the mRNA sequences above. Binding of ASL$^{\text{Met}}_{\text{CAU}}$ in either A or P-site was allowed to proceed for 30 min at 37 °C. The reaction mixtures (20 $\mu$L each) were then placed on ice for 20 min and filtered through nitrocellulose in a modified Whatman Schleicher and Schuell (Brentford, U.K.) 96-well filtration apparatus (28). Prior to filtration of experimental samples, the nitrocellulose filter was equilibrated in binding buffer at 4 °C for at least 20 min and each well of the filtration apparatus was washed with 100 $\mu$L of cold binding buffer. Cold binding buffer (100 $\mu$L) was added to each sample, and the entire 120 $\mu$L volume was quickly filtered. Each well was then washed twice with 100 $\mu$L of cold binding buffer. The nitrocellulose was dried out on ‘kim’ wipes, and the radioactivity was measured using a phosphorimager (Molecular Dynamics, GE Healthcare). Data were measured for radioactive intensity using ImageQuant (Amersham). Nonspecific binding was determined by the binding of ASLs to ribosomes without mRNA and subtracted from the experimental data. The final data is a result of at least three separate experiments, each done with samples in triplicate, ie. minimally nine determinations for each binding (Figure 6).
Analysis of the f5C pKa

UV spectra were compiled (220nm – 320nm) using a Varian Cary3 Spectrophotometer at different pH values for cytidine and 5-formylcytidine. The spectra were normalized to 0.2 O.D. at 260 nm. Entire spectra were collected to ensure that they all intersected at 260 nm at an OD of ~0.2. However, the absorbance maximum at 280 nm was plotted against the pH, a previously published method of assessing the pKa of nucleosides (29). A pH range of 2.2 – 7.0 (citrate-phosphate buffer) was used for cytidine, and a pH range of 1.1 – 5.0 was used for f5C (KCl-HCl buffer for pH values between 1.1 – 2.0 and citrate-phosphate buffer for pH values between 2.2 – 5.0). The line fitting and data analysis was conducted with Graphpad Prism v3.00 (Figure 7).

RESULTS AND DISCUSSION

A 5-formylcytidine (f5C) has previously been synthesized from 5-(hydroxymethyl)cytosine (13) and from 5-methyluridine (30), but not incorporated into an RNA sequence. First, we developed a short (4 steps) and facile synthesis of f5C (compound 5, Figure 2) from commercially available cytidine (1, Figure 2), starting by protecting cytidine as the acetonide 2 under standard acid catalysis with an 83% yield. Installation of the hydroxymethylene unit occurred through an assisted Baylis-Hillman-type reaction with formaldehyde (3, 38% yield, 60% yield based on recovered starting material; Supplementary data). Selective oxidation of the allylic alcohol with RuO2 to the aldehyde 4 proceeded with an 82% yield. The acetonide protecting group was subsequently removed to deliver f5C (5) in 95% yield. The comparison between the NMR signals of C and those of f5C clearly
demonstrated that the C-5 position of $\tilde{f}^5$C was substituted (Figures 3B and 4C). This substitution was further confirmed to be the formyl group by the presence of a low field shifted signal at ($F_1 = 185$ ppm; $F_2 = 9.40$ ppm) corresponding to the CH group of the $\tilde{f}^5$C modification. A complete and unambiguous assignment of the non-exchangeable protons was achieved by using the two dimensional (2D) $^1$H-$^{13}$C Heteronuclear Multiple Quantum Correlation (HMQC) method (31;32). As expected, the NMR peaks observed between ($F_1 = 50.00$-$85.00$ ppm; $F_2 = 4.50$-$6.00$ ppm) corresponded to the protons (H1’-H5’/H5’’) of the ribose moiety (Figure 3B) (33). Conversion of $\tilde{f}^5$C into the 2’,5’-protected $\tilde{f}^5$C phosphoramidite (9, Figure 2) commenced with the protection of the 3’ and 5’ hydroxyl groups as a disiloxane followed by protection of the 4-NH$_2$ group in the formamidine 6 (71% yield). Installation of a 2’-ACE orthoester [2’-O-bis(acetoxyethoxy)methyl-] followed by fluoride treatment delivered the diol 7 in 53% yield over 2 steps (15). The synthesis of the $\tilde{f}^5$C phosphoramidite 9 was completed through 5’-BZH (5’-O-benzhydroxy-bis(trimethylsiloxy)silyl-) protection (8, 84% yield) and phosphor-amidite formation (93% yield). A major concern for the incorporation of $\tilde{f}^5$C phosphoramidite into synthetic RNA oligomers was the formation of imine adducts with the formyl group under resin cleavage and deprotection conditions. However, we decided not to protect the formyl group since the final deprotection of the 2’-ACE groups under mildly acidic conditions would potentially hydrolyze any imine formation that results during base-deprotection back to the formyl group.
In order to investigate the contribution of f\textsuperscript{5}C\textsubscript{34} to the structure of the anticodon loop and to the decoding of both the AUG and AUA codons at both the A- and P-sites, f\textsuperscript{5}C was incorporated into the anticodon stem and loop domain of hmtRNA\textsuperscript{Met}\textsubscript{CAU} (hmASL\textsuperscript{Met}\textsubscript{CAU}) at the wobble position 34, along with pseudouridine, Ψ\textsubscript{27}. The oligonucleotide was synthesized with Ψ\textsubscript{27} and with and without f\textsuperscript{5}C\textsubscript{34} (hmASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsuperscript{5}C\textsubscript{34}, and hmASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}}) using care not to oxidize the formyl group. Incorporation of f\textsuperscript{5}C into RNA was accomplished by activating with S-ethyl tetrazole and coupling of the activated species for 3.5 minutes to the growing polyribonucleotide on the solid-support. Cleavage from the support and deprotection of the exocyclic amines was tested using NH\textsubscript{4}OH at room temperature for 24 h, methylamine at room temperature for 6 h, and t-butyl amine in water (1:3, v/v) at 60 °C for 6 h. Only the t-butyl amine conditions resulted in the correct mass upon MALDI-TOF analysis of the crude products. There was no indication of any t-butyl-amine adducts present from the MALDI-TOF results. Successful incorporation of f\textsuperscript{5}C was confirmed by NMR measurements (Figures 3 and 4) and HPLC of constituent nucleosides (Figure 4). The proton resonance of the formyl group is observed in the low field region of the 1D spectrum of hmASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsuperscript{5}C\textsubscript{34} and absent from that of the hmASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}} (Figure 3). The formyl proton in hmASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsuperscript{5}C\textsubscript{34 resonates at the same chemical shift as that of the mononucleoside f\textsuperscript{5}C\textsubscript{34}, as observed in the superimposed \textsuperscript{1}H-\textsuperscript{13}C HMQC spectra of cytidine and 5-formyleytidine (Figure 3B). The HPLC nucleoside composition analysis confirms the presence of the f\textsuperscript{5}C (Figure 4A).
The structure of the resulting hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> was characterized by 1D <sup>1</sup>H and 2D <sup>1</sup>H NOESY NMR experiments conducted in H<sub>2</sub>O at 2 °C (34), and by determining the thermodynamic contributions of f<sup>5</sup>C to the RNA. The formyl proton resonance was found at 9.45 ppm corresponding almost exactly to that of the nucleoside alone (Figure 3A). The imino protons of the stem of hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> were found to resonate between 12 and 13.5 ppm on the <sup>1</sup>H 1D NMR spectrum (Figure 3A). The NMR spin systems that involve the exchangeable imino protons of hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> were identified by conducting NMR experiments in H<sub>2</sub>O at 2 °C (Figure 3A and Figure 4B). The identification and assignment of the exchangeable protons were indicative of the overall stability the hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> in solution, and the comparison with hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup></sub> (Figure 3A) demonstrated the successful incorporation of f<sup>5</sup>C<sub>34</sub> into the sequence of hmASL<sup>Met</sup><sub>CAU</sub>.

The modified RNA synthesis has allowed us to begin examining the role of f<sup>5</sup>C<sub>34</sub> in thermal stability and decoding activity of hmtRNA<sup>Met</sup>. Thermodynamic parameters were extracted from the repeated denaturations and renaturations of both hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> and hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup></sub> (Table 1 and Figure 5A). Introduction of f<sup>5</sup>C<sub>34</sub> lowered the melting temperature and standard free energy (ΔG°<sub>37</sub>) considerably, but did not alter the ASL’s hyperchromicity. The circular dichroism spectrum of the hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> exhibited a greater ellipticity at 270 nm than that of the hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub>. The lower degree of ellipticity of hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> is indicative of a decrease in base stacking. The decreased base stacking must be attributed to the anticodon loop nucleosides because of the location of the modification. These differences in thermodynamics and circular dichroism ellipticity between hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup>f<sup>5</sup>C<sub>34</sub></sub> and hmASL<sup>Met</sup><sub>CAU-Ψ<sup>27</sup></sub> indicated that f<sup>5</sup>C<sub>34</sub> may
enhance the motional dynamics of the loop. This difference in motional dynamics was observed by a molecular dynamics simulation (MDS) performed on the hmASL$^{\text{Met}_{\text{CAU}}-\Psi_{27}}$ and the hmASL$^{\text{Met}_{\text{CAU}}-\Psi_{27}, \text{f}^{\text{C}_{34}}}$ using AMBER 9 (35). The hmASL$^{\text{Met}_{\text{CAU}}-\Psi_{27}}$ displayed an average root mean square deviation (r.m.s.d.) from the starting structure of $2.18 \pm 0.23$ as opposed to hmASL$^{\text{Met}_{\text{CAU}}-\Psi_{27}, \text{f}^{\text{C}_{34}}}$ for which higher fluctuations of $2.60 \pm 0.60$ were detected (Supplemental Material). The enhanced motional dynamics may be important for the decoding of AUA, as well as AUG.

The tRNA$^{\text{Met}}$ anticodon CAU is a cognate pair for the Met codon AUG. According to Crick’s Wobble Hypothesis (36), the binding of anticodon CAU to codon AU would be unlikely due to the C-A mismatch at the wobble position (wobble pair nucleosides in bold). However, the mitochondrial ribosome decodes both AUG and AUA using one tRNA with the anticodon CAU. This one tRNA consists of the modification f$^{\text{C}_{34}}$. In contrast, two tRNAs decode the one Met codon AUG in the cytoplasm (37). One of the tRNAs is an initiator tRNA that decodes AUG in the ribosome’s peptidyl- or P-site at the initiation of translation, where AUG is the first codon to be translated on the mRNA. This initiator tRNA$^{\text{Met}_{\text{CAU}}}$ consists of an unmodified CAU anticodon. The second cytoplasmic tRNA$^{\text{Met}_{\text{CAU}}}$ is responsible for elongation and recognizes AUG located within the mRNA, and thus responds only to the aminoacyl- or A-site codon. In E. coli, this elongator tRNA$^{\text{Met}}$ is modified with N4-acetylcytidine at the wobble position (ac$^{4}\text{C}_{34}$) (38). Thus, at the anticodon, one of the main distinguishing factors between the cytoplasmic initiator and elongator tRNA$^{\text{Met}_{\text{CAU}}}$ is the modification at the wobble position. We used a codon binding assay to observe the affinity of the hmASL$^{\text{Met}_{\text{CAU}}-\Psi_{27}, \text{f}^{\text{C}_{34}}}$ and hmASL$^{\text{Met}_{\text{CAU}}-\Psi_{27}}$ for the codons AUA and AUG.
at either A-site or P-site of *E. coli* 70S ribosomes. To ensure binding of the two ASL$_{\text{Met}}$s to the A- or the P-site, the ribosomal site not in observation (P- or A-site, respectively) was saturated with the unmodified *E. coli* ASL$_{\text{Val}}^\text{Val3}_\text{UAC}$ in response to its cognate codon GUA. The unmodified ASL$_{\text{Val}}^\text{Val3}_\text{UAC}$ binds its cognate codon with high affinity and specificity (39).

The hmASL$_{\text{CAU}}^\text{CAU-Ψ}^27$ bound AUG in the A-site and the P-site with an affinity comparable to what have observed previously for certain ASLs with unmodified wobble positions responding to cognate codons (Figure 6) (21;39). In contrast, the hmASL$_{\text{CAU}}^\text{CAU-Ψ}^27$ bound poorly to AUA in both the A- and P-sites. Surprisingly, introduction of fC$_34$ enhanced binding to AUA by two-fold (Figure 6). Our results indicated that of the two codons at either of the two ribosomal sites, the fC$_34$ modification appears to be most important for reading AUA.

Both the hmASL$_{\text{CAU}}^\text{CAU-Ψ}^27$ and the fully modified hmASL$_{\text{CAU}}^\text{CAU-Ψ}^27$;fC$_34$ exhibited considerable affinity for AUG, and at both the A-site and the P-site. However, only the hmASL$_{\text{CAU}}^\text{CAU-Ψ}^27$;fC$_34$ exhibited significant affinity for AUA. There was a doubling in the affinity of ASL$_{\text{Met}}$ for the AUA codon when fC$_34$ was present. This increase in affinity of the fC$_34$-modified ASL in comparison to that of the hmASL$_{\text{CAU}}^\text{CAU-Ψ}^27$, unmodified at the wobble position, was not observed on AUG and may therefore contribute to the solely to the efficient translation of AUA codons. A two-fold increase in affinity of tRNA towards a codon has been shown to be significant in translation (40). Although some ASL modifications cause small increases in codon-binding affinity, others can dramatically increase affinity to codons (39).
Of particular interest is the chemical and stereochemical mechanism by which a stable, but non-canonical base pair occurs between $\text{f}^5\text{C}_{34}$ and the third base of the AUA codon, an adenosine, on the ribosome. C-A base pairs are extremely unusual. Although the C-A pairing has been found in the folded structure of some RNAs such as ribosomal RNAs (rRNAs), it is rarely found in codon:anticodon pairs. C-A anticodon:codon mismatch has been detected when $C_{34}$ of tRNA$^{\text{Leu}}_{\text{CAU}}$, modified with lysidine ($k^2\text{C}_{34}$) at the wobble position, is paired to codon AUA. The lysine moiety of $C_{34}$ on the anticodon provides an amino group which hydrogen bonds to A of the codon, thus allowing the wobble position C-A mismatch to occur. One could imagine that the 5-formyl modification raises the $pK_a$ of cytidine’s N3 to the physiological range where an additional hydrogen bond could be formed to AUA. However, the $pK_a$ of $\text{f}^5\text{C}$ determined by UV spectral analysis was lower than that of C (2.3 and 4.2, respectively; Figure 7), corresponding well with previous determinations (41) including those for df$^5\text{C}$ (42) and for $\text{f}^5\text{U}$ (43). Thus, $\text{f}^5\text{C}$ must contribute to the decoding of the mitochondrial genome through a different mechanism. Another C-A anticodon:codon mismatch may occur at the wobble position when tRNA$^{\text{Leu}}_{\text{CMAA}}$ pairs with the leucine codon UUA. Like hmtRNA$^{\text{Met}}$, this tRNA$^{\text{Leu}}$ isoacceptor has a 5-formylated, 2-O-methylated C ($\text{f}^5\text{Cm}_{34}$) at the wobble position. The wobble modifications are thought to be a general characteristic of mammalian cytoplasmic tRNA$^{\text{Leu}}$ that may aid in the decoding of leucine codons UUG and UUA and prevent the miscoding of the similar codons of phenylalanine, UUU and UUC. However, there is a lack of information on the decoding properties of RNA$^{\text{Leu}}_{\text{CMAA}}$ and therefore, the possibility that an isoacceptor other than tRNA$^{\text{Leu}}_{\text{CMAA}}$ is responsible for specifically reading UUA (6;38).
Eighty percent of the methionine codons internal to mitochondrial mRNA are the AUA codon. Thus, the enhanced affinity of hmASL<sup>Met</sup><sub>CAU</sub>-Ψ<sub>27</sub>;f<sup>5</sup>C<sub>34</sub> for AUA in the A-site of the <i>E. coli</i> ribosome has important implications for the affinity and kinetics of decoding AUA during elongation. The enhanced A-site binding of AUA by the f<sup>5</sup>C-modified, hmASL<sup>Met</sup><sub>CAU</sub> may be even more evident on the mitochondrial ribosome, a concept not studied here. Also, the disease-related A37-G37 (A4435G) mutation, associated with an increased penetrance and expression of the primary Leber hereditary optic neuropathy mutation (G11778A), LHON (44), may critically alter the anticodon architecture such that either or both decoding events do not occur. This has yet to be examined. The synthesis of the wild type modified and unmodified anticodon stem and loops of the hmASL<sup>Met</sup><sub>CAU</sub> and their physical, chemical characterizations will be important to understanding the contributions of the modification to biological function and to characterization of the human disease-relevant mutant tRNA.

**FUNDING**

This work was supported in part by the North Carolina State University RNA Biology Group to AD and a grant to PFA (United Mitochondrial Disease Foundation, jointly with Dr. Linda Spremulli of the University of North Carolina-Chapel Hill) [grant number 05-20 to PFA] and NSF [grant number MCB0548602 to P.F.A.]. EMG is the recipient of an NSF Graduate Fellowship. Dharmacon RNA Technologies, Inc. (ThermoFisher, Inc.) provided support through synthesis of the modified anticodon stem and loop heptadecamer.
ACKNOWLEDGEMENTS

We are very grateful to Dr. Glenn Björk (University of Umeå, Sweden) for the HPLC analysis of the synthesized RNA. AD is a Beckman Young Investigator and a Cottrell Scholar. We thank Mr. Antonio M. Munoz for his contribution to the molecular dynamics simulations.
REFERENCES


FIGURE LEGENDS

Figure 1. Human mitochondrial tRNA^{Met}_{CAU}. The sequence and secondary structure of hmtRNA^{Met}_{CAU} are shown with the heptadecamer, anticodon stem and loop domain in bold. Nucleosides 27 and 50 are modified to pseudouridine, Ψ. The wobble position-34 modification is 5-formylcytidine, f5C34.

Figure 2. Synthesis of the 5-formylcytidine phosphoramidite. The starting compound cytidine is numbered compound 1, intermediates are numbered 2-8, and the protected f5C phosphor-amidite is compound 9. The synthetic reactions are: a) acetone, dimethoxypropane, cat HClO4 (83%); b) paraformaldehyde, 0.5M KOH, 55°C (38%, 60% brsm); c) RuO2 xH2O, dioxane, reflux (82%); d) 1M HCl (95%); e) TIPDSCl2, Pyr, DMF; f) DBF-CH(OMe)2, DMF (71% over 2 steps); g) ACE-orthoester, PPTS, TBDMS-pentanediene, DCM; h) HF-TEMED, CH3CN (53% over 2 steps); i) BZH-Cl, DIA, DCM (84%); j) P(OMe)(DIA)2, DIA, S-Et Tetrazole, DCM (93%).

Figure 3. NMR spectra of the hmASL^{Met}_{CAU}. (A) One-dimensional 1H-NMR spectrum (in H2O) of hmASL^{Met}_{CAU}-Ψ_{27};f5C_{34} (top) is compared to that of the unmodified hmASL^{Met}_{CAU}-Ψ_{27} (bottom). The formyl proton’s chemical shift in the RNA is almost identical to that of the mononucleoside f5C. *Denotes impurities. (B) Superimposed 1H-13C HMQC spectra of cytidine (blue) and 5-formylcytidine (red).
Figure 4. HPLC nucleoside composition and NMR analyses of hmASL$^{\text{Met}}_{\text{CAU}}$ and f$^\delta$C. (A) The upper section of the figure depicts the HPLC separation of standard ribonucleosides where C, G and A (solid line) had been injected together, and U, G and f$^\delta$C (dotted line) had been injected in a separate control experiment. The lower section depicts the chromatography of nucleoside composing the hmASL$^{\text{Met}}_{\text{CAU}}$ to include Ψ, as well as f$^\delta$C. (B) The NOESY connectivities between imino protons at 500 MHz in a 2D $^1$H NOESY NMR spectrum (mix = 250 ms) of hmASL$^{\text{Met}}_{\text{CAU}}$ (90% $^1$H$_2$O + 10% $^2$H$_2$O; 20 mM PO$_4$$^{3-}$, 50 mM NaCl, pH 6.2; 2°C) with water suppression using the WATERGATE sequence. (C) 1D NMR spectra of (a) cytidine (1) and (b) 5-formylcytidine (5) (500 MHz; 100% $^2$H$_2$O; 20 mM PO$_4$$^{3-}$, 50 mM Na$^+$, 50 mM K$^+$; pH = 6.2; 25°C) using the presaturation NMR sequence to suppress the water peak. The spectral regions and peaks corresponding to the various proton types in the molecules are labeled.

Figure 5. Thermal denaturations and circular dichroism spectra of the hmASL$^{\text{Met}}_{\text{CAU}}$-Ψ$_{27}$ and the hmASL$^{\text{Met}}_{\text{CAU}}$-Ψ$_{27}$;f$^\delta$C$_{34}$. (A) Thermodynamic stability of the ASLs. UV-monitored, thermal data were averaged from three denaturations and two renaturations for the hmASL$^{\text{Met}}_{\text{CAU}}$-Ψ$_{27}$;f$^\delta$C$_{34}$ (thin gray line) and for the hmASL$^{\text{Met}}_{\text{CAU}}$-Ψ$_{27}$ (thick black line). (B) Circular dichroism spectra. Spectra of the hmASL$^{\text{Met}}_{\text{CAU}}$-Ψ$_{27}$;f$^\delta$C$_{34}$ (thin gray line) and that of the hmASL$^{\text{Met}}_{\text{CAU}}$-Ψ$_{27}$ (thick black line) at the approximately equal concentrations of 2 µM were collected over the wavelength range of 200 to 300 nm.
**Figure 6.** Codon binding by hmASL<sub>CAU</sub><sup>Ψ<sub>27</sub></sup> and the hmASL<sub>CAU</sub><sup>Ψ<sub>27;f<sub>5</sub>C<sub>34</sub></sup></sup>. The equilibrium binding of the two ASLs to the cognate and non-cognate codons, AUG and AUA respectively, was assessed using programmed *E. coli* ribosomes. (The ASL unmodified and the wobble position is designated “Ψ<sub>27</sub>; ASL modified at the wobble position is designated as “Ψ<sub>27;f<sub>5</sub>C<sub>34</sub>”).) The ASLs were bound to the A-site with *E. coli* ASL<sub>Val3</sub>UAC bound to its cognate codon in the P-site. The ASLs were bound to the P-site with the ASL<sub>Val3</sub>UAC bound to its cognate codon in the A-site.

**Figure 7.** Analysis of the pKa of (■) cytidine in comparison to that of (▲) 5-formylcytidine. UV spectra were collected for the two nucleosides over a range of pH values, pH 2.2 – 7.0 for cytidine, and a pH 1.1 – 5.0 for f<sub>5</sub>C. The absorbance at 280 nm was normalized and plotted against the pH.

**TABLE 1:** Thermodynamic contributions of f<sub>5</sub>C<sub>34</sub>.

<table>
<thead>
<tr>
<th>hmASL&lt;sub&gt;CAU&lt;/sub&gt;&lt;sup&gt;Met&lt;/sup&gt;</th>
<th>Tm (°C)</th>
<th>ΔG°&lt;sub&gt;37&lt;/sub&gt; (Kcal/mol)</th>
<th>ΔH (Kcal/mol)</th>
<th>ΔS (cal/K·mol)</th>
<th>Hyperchrom. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ψ&lt;sub&gt;27&lt;/sub&gt;</td>
<td>67.7 ± 1.8</td>
<td>-2.8 ± 0.1</td>
<td>-31.1 ± 2.0</td>
<td>-91.3 ± 6.3</td>
<td>16 ± 1</td>
</tr>
<tr>
<td>Ψ&lt;sub&gt;27;f&lt;sub&gt;5&lt;/sub&gt;C&lt;sub&gt;34&lt;/sub&gt;</td>
<td>60.1 ± 0.8</td>
<td>-1.6 ± 0.1</td>
<td>-22.8 ± 1.2</td>
<td>-68.4 ± 3.4</td>
<td>16 ± 1</td>
</tr>
</tbody>
</table>
Figure 1. Human mitochondrial tRNA$^{\text{Met}}_{\text{CAU}}$. The sequence and secondary structure of hmtRNA$^{\text{Met}}_{\text{CAU}}$ are shown with the heptadecamer, anticodon stem and loop domain in bold. Nucleosides 27 and 50 are modified to pseudouridine, $\Psi$. The wobble position-34 modification is 5-formylcytidine, $f^5C_{34}$. 
**Figure 2.** Synthesis of the 5-formylcytidine phosphoramidite. The starting compound cytidine is numbered compound 1, intermediates are numbered 2-8, and the protected f5C phosphor-amidite is compound 9. The synthetic reactions are: a) acetone, dimethoxypropane, cat HClO4 (83%); b) paraformaldehyde, 0.5M KOH, 55ºC (38%, 60% brsm); c) RuO2 xH2O, dioxane, reflux (82%); d) IM HCl (95%); e) TIPDSCl2, Pyr, DMF; f) DBF-CH(OMe)2, DMF (71% over 2 steps); g) ACE-orthoester, PPTS, TBDMS-pentanedione, DCM; h) HF-TEMED, CH3CN (53% over 2 steps); i) BZH-Cl, DIA, DCM (84%); j) P(OMe)(DIA)2, DIA, S-Et Tetrazole, DCM (93%).
**Figure 3.** NMR spectra of the hmASL\textsuperscript{Met\textsubscript{CAU}}. \textbf{(A)} One-dimensional 1H-NMR spectrum (in H\textsubscript{2}O) of hmASL\textsuperscript{Met\textsubscript{CAU}-Ψ\textsubscript{27};f\textsuperscript{5}C\textsubscript{34}} (top) is compared to that of the unmodified hmASL\textsuperscript{Met\textsubscript{CAU}-Ψ\textsubscript{27}} (bottom). The formyl proton’s chemical shift in the RNA is almost identical to that of the mononucleoside f\textsuperscript{5}C. *Denotes impurities. \textbf{(B)} Superimposed \textsuperscript{1}H-\textsuperscript{13}C HMQC spectra of cytidine (blue) and 5-formylycytidine (red).

---

128
Figure 4. HPLC nucleoside composition and NMR analyses of hmASL$^{\text{Met}}_{\text{CAU}}$ and $f^5$C. (A) The upper section of the figure depicts the HPLC separation of standard ribonucleosides where C, G and A (solid line) had been injected together, and U, G and $f^5$C (dotted line) had been injected in a separate control experiment. The lower section depicts the chromatography of nucleoside composing the hmASL$^{\text{Met}}_{\text{CAU}}$ to include $\Psi$, as well as $f^5$C. (B) The NOESY connectivities between imino protons at 500 MHz in a 2D $^1$H NOESY NMR spectrum (mix = 250 ms) of hmASL$^{\text{Met}}_{\text{CAU}}$ (90 % $^1$H$_2$O + 10 % $^2$H$_2$O; 20 mM PO$_4^{3-}$, 50 mM NaCl, pH 6.2; 2°C) with water suppression using the WATERGATE sequence. (C) 1D NMR spectra of (a) cytidine (1) and (b) 5-formylcytidine (5) (500 MHz; 100 % $^2$H$_2$O; 20 mM PO$_4^{3-}$, 50 mM Na$^+$, 50 mM K$^+$; pH = 6.2; 25°C) using the presaturation NMR sequence to suppress the water peak. The spectral regions and peaks corresponding to the various proton types in the molecules are labeled.
Figure 5. Thermal denaturations and circular dichroism spectra of the hmASL$^{\text{Met}}_{\text{CAU}-\Psi_27}$ and the hmASL$^{\text{Met}}_{\text{CAU}-\Psi_27;f^5C_34}$. (A) Thermodynamic stability of the ASLs. UV-monitored, thermal data were averaged from three denaturations and two renaturations for the hmASL$^{\text{Met}}_{\text{CAU}-\Psi_27;f^5C_34}$ (thin gray line) and for the hmASL$^{\text{Met}}_{\text{CAU}-\Psi_27}$ (thick black line). (B) Circular dichroism spectra. Spectra of the hmASL$^{\text{Met}}_{\text{CAU}-\Psi_27;f^5C_34}$ (thin gray line) and that of the hmASL$^{\text{Met}}_{\text{CAU}-\Psi_27}$ (thick black line) at the approximately equal concentrations of 2 µM were collected over the wavelength range of 200 to 300 nm.
Figure 6. Codon binding by hmASL$^{\text{Met}}_{\text{CAU}}$Ψ$_{27}$ and the hmASL$^{\text{Met}}_{\text{CAU}}$Ψ$_{27}$;f$^5$C$_{34}$. The equilibrium binding of the two ASLs to the cognate and non-cognate codons, AUG and AUA respectively, was assessed using programmed *E. coli* ribosomes. (The ASL unmodified and the wobble position is designated “Ψ$_{27}$”; ASL modified at the wobble position is designated as “Ψ$_{27}$;f$^5$C$_{34}$.”) The ASLs were bound to the A-site with *E. coli* ASL$^{\text{Val}}_{\text{UAC}}$ bound to its cognate codon in the P-site. The ASLs were bound to the P-site with the ASL$^{\text{Val}}_{\text{UAC}}$ bound to its cognate codon in the A-site.
Figure 7. Analysis of the pKa of (■) cytidine in comparison to that of (▲) 5-formylcytidine. UV spectra were collected for the two nucleosides over a range of pH values, pH 2.2 – 7.0 for cytidine, and a pH 1.1 – 5.0 for 5C. The absorbance at 280 nm was normalized and plotted against the pH.
Supplemental Figure:

Heavy atoms root mean square deviation (r.m.s.d.) variations as a function of the molecular dynamics simulation time for the hmASL^{Met}_{CAU-Ψ_{27}; t^C_{34}} (gray line) and for the hmASL^{Met}_{CAU-Ψ_{27}} (black line).
Chapter 5

The wobble position modified nucleoside of human mitochondrial tRNA\textsuperscript{Met} decodes unconventional codons in the mitochondrial Genetic Code

ESTELLA M. GUSTILO,\textsuperscript{1} CHRISTIE JONES,\textsuperscript{2} LINDA SPREMULLI,\textsuperscript{2,*} and PAUL F. AGRIS,\textsuperscript{1,*}

\textsuperscript{1}Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA and \textsuperscript{2}Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599, USA

Running Head: Unconventional decoding by human mitochondrial tRNA\textsuperscript{Met}

Keywords: tRNA, 5-formylcytidine, translation, wobble position, mitochondria

*To whom correspondence should be addressed:
Paul F. Agris, Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA, 919-515-6188; Paul_Agris@ncsu.edu; and

Linda Spremulli, Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599, USA, 919-966-1567; linda_spremulli@unc.edu
ABSTRACT

The mammalian mitochondrial Genetic Code deviates from the universal code in that the universal Isoleucine code AUA is used to code for Methionine (Met) in the mitochondria. Thus, mitochondrial mRNAs utilize two codons for Met, the universal AUG and the unconventional AUA. Furthermore, while translation in the cytoplasm employs an initiator tRNA$^{Met}_{CAU}$ and an elongator tRNA$^{Met}_{CAU}$ to read the Met codon AUG in the peptidyl- and aminoacyl- sites of the ribosome, respectively, the mitochondria use one species of tRNA$^{Met}_{CAU}$ to read both Met codons at both sites. This one species of mitochondrial tRNA$^{Met}_{CAU}$ consists of a unique 5-formyl-modification of Cytidine at the wobble position-34 (hmtRNA$^{Met}_{CAU}$-f$^{5}$C$^{34}$). Here, we report the codon binding properties of the anticodon stem and loop domain of hmtRNA$^{Met}_{CAU}$ with f$^{5}$C$^{34}$ and the naturally occurring pseudouridine at position-27 (hmtASL$^{Met}_{CAU}$-Ψ$^{27}$, f$^{5}$C$^{34}$). Codon binding studies with both E. coli and bovine mitochondrial ribosomes reveal how this single modification to the anticodon wobble position allows hmtRNA$^{Met}_{CAU}$-Ψ$^{37}$, f$^{5}$C$^{34}$ to read AUG and AUA at the two ribosomal sites. Interestingly, the mitochondrial translation of many organisms, including that of one human protein, sometimes initiates translation using the universal Isoleucine codons AUU and AUC. We report that the f$^{5}$C$^{34}$-modification of hmtRNA$^{Met}_{CAU}$-Ψ$^{27}$, f$^{5}$C$^{34}$ permits codon-reading expansion to include the rare initiation codons AUU and AUC. In this way, codon recognition by hmtRNA$^{Met}_{CAU}$ is extended to the entire codon box AUN. Thus, this one modification at the wobble position of hmtRNA$^{Met}_{CAU}$-f$^{5}$C$^{34}$ allows recognition of non-traditional codons in the reassignment of universal codons in the mitochondria.
INTRODUCTION

An interesting feature of tRNA is its ability to read synonymous codons. In every organism, codons outnumber tRNA species; therefore, many tRNAs must read more than one codon. This ability of a tRNA to read more than one codon is most apparent in the mitochondria. In the mammalian mitochondrion, just 22 tRNAs must read all of the codons of the 20 amino acids (Scheffler 2008). Thus, except for Serine and Leucine which each have six codons, there is one tRNA species that reads all the codons of each amino acid. Since each amino acid in the mitochondria consists of 2-6 codons, each tRNA species in the mitochondria must read at least two codons.

A tRNA’s ability to read more than one codon is explained in Francis Crick’s Wobble Hypothesis (Crick, 1966) where Crick proposed that, during codon:anticodon pairing, the first two positions base-pair in strict Watson-Crick fashion – G:C or A:U. However, the third position (wobble position with the third codon base bound to the first base of the anticodon) is more flexible in that a G can bind its canonical pair C or “wobble” to a U, and a U would bind an A and wobble to a G. Crick also hypothesized that the modified nucleoside Inosine at the first position of the anticodon, I₃₄, would bind U, C, or A. The forty-plus years of codon-binding research has proven the validity of Crick’s Wobble Hypothesis by revealing the ribosome’s flexibility in accepting many base pairings other than Watson-Crick (W-C) pairs at tRNA’s wobble position (Agris et al. 2007). However, wobble pairings are more complicated than Crick’s predicted G:U/U:G and Inosine pairings. Oftentimes, the wobble position-34 of tRNA consists of post-transcriptional modifications more complex than I₃₄. The wobble position and position 37 (which is directly adjacent to the anticodon triplet
N_{34}N_{35}N_{36}) are highly modified and consist of the greatest variety of modifications of any position of any RNA (Agris et al. 2007). Modifications within the anticodon loop are important for wobble pairings (see reviews by Agris 2008; Agris et al. 2007; Björk 1995).

The sequencing of mitochondrial genomes from various organisms has revealed that the established Genetic Code is not quite as universal as previously thought (see review by Santos et al. 2004). There are many instances of codon reassignment in the mitochondria, where codons assigned to decode a certain amino acid or terminate translation in the universal Genetic Code are used to decode a different amino acid. For example, in the mammalian mitochondria, the universal stop codon UGA decodes Tryptophan, the two universal Arginine codons AGA and AGG terminate translation, and the universal Isoleucine codon AUA is used to decode Methionine (Fig. 1; Jukes and Osawa 1990).

In the cytoplasm of all organisms, there is one codon translated as Methionine – AUG. AUG is also the codon that initiates translation. There are two tRNA^{Met} isoaccepting species that decode this one AUG codon on cytoplasmic ribosomes. The isoaccepting tRNA^{Met} used during translation depends on the location of AUG. For instance, during prokaryotic translation initiation, the messenger RNA (mRNA) is aligned on the 30S subunit of the ribosome with AUG at the peptidyl-site (P-site). An initiation factor then brings the initiator tRNA^{Met} to the P-site to bind AUG. The initiator tRNA^{Met} is the only tRNA recognized by initiation factors to bind the codon at the ribosome’s P-site. All other tRNAs are elongator tRNAs and enter the ribosome via the aminoacyl-site (A-site) aided by elongation factors. AUG codons occurring in the mRNA after initiation are read by the elongator tRNA^{Met}. The one Methionine (Met) codon (AUG) and two tRNA^{Met} isoacceptors
(initiator and elongator tRNA\textsuperscript{Met}) are universal among prokaryotic and eukaryotic, cytoplasmic translational systems. However, the mitochondria deviate from the universal translational system in that there are two Met codons, AUG and AUA, and there is only one tRNA\textsuperscript{Met} species that reads both codons at both the P- and A-sites of the ribosome (Young and Anderson 1980; Anderson et al. 1981; Anderson et al. 1982; Fearnley and Walker 1987; Takemoto et al. 1995). This one mitochondrial tRNA\textsuperscript{Met} is a hybrid of the mammalian cytoplasmic initiator and elongator tRNA\textsuperscript{Met} species. Its ability to read both AUG and AUA is not fully understood. Our studies focus on the codon binding properties of the human mitochondrial tRNA\textsuperscript{Met}, hmtRNA\textsuperscript{Met}, in particular that of tRNA’s anticodon stem and loop domain, hmtASL\textsuperscript{Met}. The heptadecamer hmtASL\textsuperscript{Met} has a five base-paired stem consisting of three G:C base pairs that are proximal to the loop - a hallmark of all initiator tRNA\textsuperscript{Met} that increases its affinity for the P-site (Mayer et al. 2001). As with the mammalian initiator tRNA\textsuperscript{Met}, the hmtASL\textsuperscript{Met} has a Cytidine at position-33. tRNA’s position-33 is usually occupied by the invariant U\textsubscript{33}, responsible for the U-turn in the backbone that presents the anticodon for codon binding. Other than in the hmtASL\textsuperscript{Met}, C\textsubscript{33} has been found only in the cytoplasmic initiator tRNA\textsuperscript{Met} of insects, plants, and vertebrates (RajBhandary and Chow 1995). The hmtASL\textsuperscript{Met} shares one important similarity to elongator tRNA\textsuperscript{Met} – a post-transcriptional modification of the wobble position-34. In the bacterial and eukaryotic cytoplasm, initiator tRNA\textsuperscript{Met} is unmodified at the wobble position. However, the elongator tRNA\textsuperscript{Met} has a modified wobble position C\textsubscript{34}. In \textit{Escherichia coli}, the modification is N\textsuperscript{4}-acetylcytidine, while in humans it is the 2’-O-methylcytidine. In contrast, the post-transcriptional wobble position modification of hmtRNA\textsuperscript{Met} is the 5-formylcytidine, f\textsuperscript{5}C\textsubscript{34}.

138
We previously reported that the 5-formyl group appeared to be significant to the binding of the unconventional Met codon AUA (Lusic et al., 2008). We have sought to understand the role of the 5-formyl group in the efficiency of binding of the hmtRNA^{Met} to the mitochondrial codons AUA or AUG at the ribosomal P- or A-sites. Here, we report that the wobble position f^{5}C_{34} of hmtRNA^{Met}_{CAU} is important to the binding of the anticodon stem and loop (ASL) to both mitochondrial Methionine codons AUA and AUG at the A-site and is critical to its binding at the P-site of ribosome. Interestingly, some organisms use the Isoleucine codons AUU and AUC in initiation of translation at the P-site. Humans, giant pandas and fruit flies have been found to use AUU for initiation; honey bees use AUC, and mice use both AUU and AUC to initiate translation (Montoya et al. 1981; Bibb et al. 1981; Fearnley and Walker 1987; Crozier and Crozier 1993). The wobble-formylated tRNA^{Met} has been found in humans, bovines, and rodents and considered universal in mammalian mtRNA^{Met} (Moriya et al. 1994; Takemoto et al. 1995; Takemoto et al. 1999) The 5-formyl-modification at the wobble of mtRNA^{Met}_{CAU} has also been found in nematodes, squids, frogs, chickens, and fruit flies (Watanabe et al. 1994; Tomita et al. 1997; Moriya et al. 1994; Tomita et al. 1999). Therefore, there is a high probability that the mtRNA^{Met}_{CAU} recognizes the Isoleucine codons AUU and AUC at the initiation site of the ribosome, the P-site. We hypothesized that this one mtRNA^{Met}_{CAU-f^{5}C_{34}} that functions in both translational initiation and elongation in mitochondria can recognize AUU and AUC when these codons are used for initiation. We report here that, just as the 5-formyl-modification allows the mtRNA^{Met}_{CAU} to recognize the unconventional Methionine codon AUA, this modification also expands
codon reading to include the Isoleucine codons AUU and AUC at initiation. Therefore, the wobble modification expands codon recognition at the P-site to include the entire codon box AUN.

RESULTS

The anticodon wobble position modification, 5-formylcytidine, significantly enhances AUA codon recognition by the hmtASL$^{\text{Met}}_{\text{CAU-Ψ}27;f^5\text{C}_{34}}$. The mammalian mitochondrial translational system begins each of its thirteen proteins with Methionine (Met) when the codon AUG or AUA initiates translation at the ribosomal P-site. It also inserts Met into each growing polypeptide chain when either codon, AUG or AUA, appears at the ribosomal A-site. To study the effect of the naturally-occurring, wobble position modification, 5-formylcytidine ($f^5\text{C}_{34}$), on the codon binding characteristics of the human mitochondrial tRNA$^{\text{Met}}$, we determined the equilibrium binding properties of its anticodon stem and loop domain, hmtASL$^{\text{Met}}_{\text{CAU-Ψ}27;f^5\text{C}_{34}}$, with the naturally occurring pseudouridine-27, Ψ$^{\text{Ψ}27}$, and the $f^5\text{C}_{34}$ (Fig. 2).

The affinity of the hmtASL$^{\text{Met}}_{\text{CAU-Ψ}27;f^5\text{C}_{34}}$ and that of the hmtASL$^{\text{Met}}_{\text{CAU-Ψ}27}$ for the codons AUG and AUA at the P- and A-sites was determined by titrating mRNA-programmed E. coli 70S ribosomes with increasing amounts of the hmtASL$^{\text{Met}}_{\text{CAU-Ψ}27;f^5\text{C}_{34}}$ (Fig. 2; Table 1). In order to observe binding at the P-site, ribosomes were programmed with a 27mer oligonucleotide mRNA with a Shine-Dalgarno sequence at the 5’-leader, an AUG or AUA at the P-site, and the Valine codon GUA at the A-site (see Materials and Methods). Since ASLs have a greater affinity for codons in the P- rather than the A-site and both the
unmodified and hmtASL$_{\text{CAU}^\Psi_{27};\text{f}^\text{C}34}^\text{Met}$ have a very low affinity for the Val codon GUA, similar to that of nonspecific ribosome binding (data not shown), the A-site was left empty. Control experiments saturating the GUA-programmed A-site with ASL$_{\text{UAC}}^\text{Val}$ after P-site binding revealed similar results as those in the absence of ASL$_{\text{UAC}}^\text{Val}$ (data not shown). As we expected, the 5-formyl modification does not appear to affect equilibrium binding of hmtASL$_{\text{CAU}^\Psi_{27};\text{f}^\text{C}34}^\text{Met}$ to AUG at the P-site, as both the f$^\text{C}34$-modified and the hmtASL$_{\text{CAU}^\Psi_{27}}^\text{Met}$ bound AUG with similar $K_d$s of 0.9 ± 0.2 μM and 1.2 ± 0.3 μM, respectively (Table 1).

In order to study the binding of the hmtASLs to the A-site codons AUG or AUA, the _E. coli_ 70S ribosome was programmed with an mRNA of 27 nucleotides in length consisting of a 5'-leader with the Shine-Dalgarno sequence, the Val codon GUA at the P-site, and either AUG or AUA at the A-site. The P-site was then saturated with the unmodified ASL$_{\text{UAC}}^\text{Val}$ prior to the addition of hmtASL$_{\text{CAU}^\Psi_{27};\text{f}^\text{C}34}^\text{Met}$ or the hmtASL$_{\text{CAU}^\Psi_{27}}^\text{Met}$ (see Materials and Methods). The unmodified ASL$_{\text{UAC}}^\text{Val}$ has an affinity for its cognate codon GUA that is near that of the fully modified ASL$_{\text{UAC}}^\text{Val}$ (Vendeix et al. 2007). The f$^\text{C}34$ modification had little effect on the equilibrium binding of the hmtASL$_{\text{CAU}^\Psi_{27}}^\text{Met}$ to AUG at the A-site, as had been observed for the binding of AUG at the P-site. The hmtASL$_{\text{CAU}^\Psi_{27};\text{f}^\text{C}34}^\text{Met}$ exhibited a $K_d$ of 1.2 ± 0.1 μM, while the wobble-unmodified hmtASL$_{\text{CAU}^\Psi_{27}}^\text{Met}$ had a $K_d$ of 1.3 ± 0.1 μM. However, there appeared to be differences in the levels of A-site saturation. The saturation level of the hmtASL$_{\text{CAU}^\Psi_{27};\text{f}^\text{C}34}^\text{Met}$ was some 30% higher than that of the hmtASL$_{\text{CAU}^\Psi_{27}}^\text{Met}$ counterpart. (Fig. 3). We attributed this difference to possible differences in the on and off rates of the 5-formyl-modified hmtASL$_{\text{CAU}^\Psi_{27};\text{f}^\text{C}34}^\text{Met}$ and its hmtASL$_{\text{CAU}^\Psi_{27}}^\text{Met}$ counterpart.
In contrast to their binding of AUG, the two ASLs exhibited a significant difference in their binding to the AUA codon. In fact, the f5C34 modification enhanced the binding of the hmtASLMet_CAU to AUA at both the P- and A-sites. In binding AUA at the P- and A-sites, the hmtASLMet_CAU-Ψ27;f5C34 exhibited steady-state dissociation constants of approximately 2.5 μM (Kd of 2.5 ± 0.1 μM and 2.7 ± 0.4 μM at the P- and A-sites, respectively). However, the hmtASLMet_CAU-Ψ27 exhibited a significantly lower binding affinity that that of the hmtASLMet_CAU-Ψ27;f5C34 (Table 1). The hmtASLMet_CAU-Ψ27 bound AUA at the P-site with a Kd of 8.5 ± 1.5 μM, and at the A-site with a Kd of 11.0 ± 1.3 μM (Table 1).

**Kinetics of hmtASLMet_CAU-Ψ27;f5C34 and the wobble-unmodified hmtASLMet_CAU-Ψ27 binding AUG and AUA at the P- and A-site.** The unmodified CAU anticodon of tRNA^Met is a canonical match for the codon AUG. Therefore, we did not expect the 5-formyl modification of hmtASLMet_CAU-Ψ27;f5C34 to have a significant effect on the ASL’s binding to AUG. As expected, steady-state binding assays of hmtASLMet_CAU-Ψ27;f5C34 or the hmtASLMet_CAU-Ψ27 to AUG at the P- or A-sites resulted in similar KdS of approximately 1 μM. However, there were obvious differences in the levels of ASL concentration at which saturation was attained (Fig. 3). Thus, we explored the possibility that the two ASLs had significant differences in the kinetics of their binding to AUG and AUA, at the P- and A-sites. At a low concentration of ASL (only two-fold that of the ribosome), a significant difference in the kinetics of binding became apparent (Fig. 4). The amount of hmtASLMet_CAU-Ψ27;f5C34 bound to AUG in the P-site was five-fold greater than that of the ASL^Met-Ψ27 after only 30 secs. While the
hmtASL_{Met}^{CAU-Ψ_27} maintained a steady amount of low binding over time, the amount of fully-modified hmtASL_{Met}^{CAU-Ψ_27; f^6C_{34}} bound to AUG continued to increase, reaching a ten-fold greater amount bound than that of the ASL_{Met}^{CAU-Ψ_27} at the final time point of 30 min (Fig. 4). These results were repeated for the binding of the two ASLs to AUG at the A-site, and AUA at both the P- and A-sites (Fig. 4).

**Codon binding of hmtASL_{Met}^{CAU-Ψ_27; f^6C_{34}} and the hmtASL_{Met}^{CAU-Ψ_27} on the bovine 55S mitochondrial ribosome.** The codon binding experiments thus far described were conducted with the model *E. coli* 70S ribosome. However, previous studies of translation *in vitro* using bovine mitochondrial 55S ribosomes succeeded in demonstrating low levels of incorporation of Methionine into protein using bovine mtRNA_{Met}^{CAU-Ψ_27; f^6C_{34}} and mitochondrial ribosomes programmed with poly(AUA) (Takemoto et al. 2009). Therefore, we thought it appropriate to attempt the binding of hmtASL_{Met} to programmed, purified bovine mitochondrial 55S ribosomes. Due to the limited amount of resources and the difficulty of obtaining highly purified and active mitochondrial 55S ribosomes, we were unable to titrate the ribosomes with the ASLs. However, we have obtained reproducible results in the binding the hmtASL_{Met}^{CAU-Ψ_27; f^6C_{34}} and the hmtASL_{Met}^{CAU-Ψ_27} at one concentration. The bovine mitochondrial 55S ribosomes were programmed with mRNA, and then either hmtASL_{Met}^{CAU-Ψ_27; f^6C_{34}} or the hmtASL_{Met}^{CAU-Ψ_27} was added and incubated at 37 °C (see Materials and Methods). Mitochondrial mRNAs tend to be leaderless. Therefore, to study the binding of hmtASL_{Met} to the initiation site (P-site), we used a 30mer oligonucleotide mRNA with either AUG or AUA at the 5’ terminal end followed by the Val codon GUA at the A-site. To
observe an A-site binding, the 55S ribosomes were programmed with a 30mer mRNA construct consisting of a 5'-leader of twelve uridines, followed by GUA at the P-site and AUG or AUU at the A-site, and ending with another stretch of twelve Us (see Materials and Methods). Although the level of codon binding was considerably lower than that on the *E. coli* ribosome, binding was still detectable. The lower total amount of binding was probably due to an inability to extract highly active mitochondrial ribosomes in comparison to the *E. coli* 70S ribosomes, as has been experienced by others (Takamoto et al., 2009; O’Brien 1971). Though the magnitude of codon binding on the 55S ribosome was considerably less than that on the *E. coli* ribosome, the relative affinities of the f5C34-modified and unmodified hmtASLMetCAU for AUG were similar. Both hmtASLMetCAU-Ψ27;f5C34 and the hmtASLMetCAU-Ψ27 clearly bound to AUG at the initiation site, the P-site (Fig. 5, first two bars). However, the presence of the 5-formyl-group enhanced binding by approximately 40%, i.e. hmtASLMetCAU-Ψ27;f5C34, 0.80 pmoles bound compared to the unmodified hmtASLMetCAU, 0.55 pmoles bound (Fig. 5).

The largest affect of the 5-formyl-modification on codon binding was evident in binding of the hmtASLMet to AUG at the A-site and AUU at the P-site of the mitochondrial 55S ribosome. There was no detectable binding to either codon unless the modification was present. Only the modified hmtASLMetCAU-Ψ27;f5C34 resulted in detectable binding to AUG at the A-site and AUU at the P-site (Fig. 5). We were unable to observe the binding of either hmtASLMetCAU-Ψ27;f5C34 or the unmodified hmtASLMetCAU to AUU at the A-site. This may have been due to the very low levels of binding to mitochondrial ribosomes experienced by us (approximately 800 nmoles of hmtASLMetCAU-Ψ27;f5C34 bound to AUG at the P-site;
Ability of the 5-formyl-modification to expand codon recognition within the codon box AUN. Mitochondria use not only AUG and AUA, but also have been found to use the two mitochondrial Isoleucine codons AUU and AUC as initiation codons (Sheffler 2008; Fearnley and Walker 1987; Bibb et al. 1981). AUU and AUC are rare initiation codons and
may be used to regulate gene expression. Since the mitochondrial tRNA\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27;f5C34} initiates mitochondrial translation, we decided to study the ability of its anticodon stem and loop domain to bind AUU and AUC. We found that the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27;f5C34} bound to AUU and to AUC at the P-site with an affinity similar to that for AUA (Fig. 6). Furthermore, as it did in binding to AUA, the modified hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27;f5C34} bound to AUU and AUC with an affinity that was two- to three-fold greater than that of the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27} (Fig. 6). The wobble-unmodified hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27} bound to AUU, AUC, and AUA with an affinity comparable to that of the negative-control codon for Valine GUA (Fig. 6). The hmtASLMet\textsubscript{CAU-Ψ}\textsubscript{27} exhibited a significantly lower binding affinity that of the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27;f5C34} (Table 1). The hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27;f5C34} bound AUC at the P-site with a $K_d$ of $9.1 \pm 1.7 \mu M$, while the wobble-unmodified hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27} bound with a low affinity of $36.4\pm14.3 \mu M$ (Table 1). On AUU, the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27;f5C34} bound at the P-site with a $K_d$ of $5.8 \pm 0.8 \mu M$, which is approximately 30% better binding than the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ}\textsubscript{27} with a $K_d$ of $8.2 \pm 1.7 \mu M$ (Table 1).

**DISCUSSION**

The mitochondrion is responsible for functions that are critical to cell viability. It is responsible for supplying the cell with its main source of energy, and it regulates apoptosis and senescence. The mitochondrion’s own translational machinery allows it to produce proteins central to its function. The 22 tRNAs encoded in mammalian mitochondrial DNA comprise approximately nine percent of the mitochondrial genome; however, mutations of these genes constitute over fifty percent of human mitochondrial diseases (Scaglia and Wong
Some serious diseases resulting from malfunctioning mitochondria have been linked to a missing modification at the wobble position-34 of mitochondrial tRNA or position 37 directly adjacent to the wobble position (Kirino and Suzuki 2005).

The efficiency of the decoding of the Genetic Code in the mammalian mitochondria is interesting in that there are only 22 tRNAs for 60 sense codons (Scheffler 2008). Crick had predicted that a minimum of 32 tRNAs are required to read 61 sense codons (Crick 1966). Thus, the mitochondrial translational system must fully utilize flexibility at the wobble position during codon:anticodon pairing. This is accomplished by having most of its tRNAs read whole codon boxes. Except for Leucine and Serine, each of the twenty amino acids is decoded by a single tRNA. Leucine and Serine each have six codons, four in a whole codon box and two in a split-box; therefore, two tRNAs must decode the two different boxes for each Leu and Ser. The other eighteen amino acids have two to four codons and are decoded by eighteen tRNAs.

The mammalian mitochondrial tRNA^fMet is resourceful in that it does the job of the two cytoplasmic tRNA^Met by functioning in both translational initiation and elongation. Furthermore, while the two cytoplasmic tRNA^Met read only the one Methionine codon, AUG, the one mitochondrial tRNA^Met must read two to four codons - the mitochondrial Methionine codons AUG and AUA at initiation and elongation, and the two Isoleucine codons AUU and AUC in initiation. Details as to how the mitochondria partition translational initiation and elongation are not clear, but we show that the wobble-modified hmtRNA^{Met}_{CAU-\text{a}C_{34}} enhances binding to the four codons AUG, AUA, AUC, and AUU at the P-site for translation initiation and on Met codons AUG and AUA at the A-site. By comparing the codon binding
properties of the ASL$_{CAU}^{\text{Met}}$$^{\Psi_{27};\hat{f}\delta C_{34}}$ to that of the wobble-unmodified ASL$_{CAU}^{\text{Met}}$$^{\Psi_{27}}$, we have demonstrated that the 5-formyl modification is directly responsible for enhancing binding efficiency to the mitochondrial Methionine codons, three of which (AUU, AUC, and AUA) are unconventional Methionine codons.

The interaction between $\hat{f}\delta C_{34}$ of the anticodon with A3 of the codon AUA is particularly interesting in that it is rare base pair and not engendered by Crick in his Wobble Hypothesis. Though A:C pairs are evident in RNA structures (Xin and Olson, 2008), the structure of a codon:anticodon pair is not known. The probability of the rare A:$\hat{f}\delta C$ base pair at the wobble position of codon:anticodon pairing during translation of Methionine is high considering that AUA codon-usage in human mitochondria (Nakamura et al. 2000) is high and that the hmtRNA$_{CAU}^{\text{Met}}$$^{\hat{f}\delta C_{34}}$ is the tRNA that decodes AUA. AUA is the third most highly used codon in the human mitochondria, and AUA is used more than AUG by approximately four to one (Nakamura et al. 2000). We have previously demonstrated that the 5-formyl modification does not facilitate the protonation of the N3 of C$_{34}$ at physiological pH, and therefore, no imino hydrogen bond is likely formed between C$_{34}$ of the anticodon and A3 of the codon AUA (Lusic et al., 2008).

There are two more examples where a codon:anticodon mismatch of an A:C pair may occur. In both cases, the C$_{34}$ is modified. The bacterial tRNA$^{\text{Ile}}_{CAU}$ has Lysine-modified C (lysidine; $k^2$C) at the wobble position (tRNA$^{\text{Ile}}_{CAU}$-$k^2$C$_{34}$) and is known to be responsible for reading the Isoleucine codon AUA. The 2-lysyl modification of C$_{34}$ on the tRNA$^{\text{Ile}}_{CAU}$-$k^2$C$_{34}$ contributes an amino group that hydrogen bonds to A3 of the codon AUA, allowing the A:$k^2$C wobble to occur. Another example where the A:C mismatch may occur is the pairing
of eukaryotic tRNA^{Leu}_{CAA-f^5Cm34} to the Leucine codon UUA. As with hmtRNA^{Met}_{CAU-f^5C34},
the wobble position of tRNA^{Leu}_{CAA-f^5Cm34} is modified to a 5-formyl Cytosine, and also
methylated at the 2′-OH (Pais de Barros et al. 1996). However, the decoding properties of
tRNA^{Leu}_{CAA-f^5Cm34} are not known. The isoaccepting tRNA^{Leu}_{IAA} with a wobble position I
may be responsible for decoding UUA.

Our studies of the codon binding and structure of the ASL of the modified
tRNA^{Val}_{UAC} in solution (Vendeix et al., 2008) and within the crystal of the 30S ribosomal
subunit (Weixlbaumer et al. 2007) revealed interesting and novel hydrogen bonding that may
provide an understanding of how the hmtRNA^{Met} decodes AUA. When bound to each and
everyone of the four synonymous Valine codons, the anticodon domain of tRNA^{Val}_{UAC} has a
hydrogen bond between the 2′-OH of U_{33} and the O^5 of the wobble position 5-
methoxymethyluridine, cmo^5U_{34} (Weixlbaumer et al. 2007). This hydrogen bond rigidly
restricts the motion of the cmo^5U_{34} thereby reducing the energetic cost of binding the codons
in the ribosome’s A-site (Agris, 2008; Weixlbaumer et al., 2007). We are conducting an
investigation of the mtASL^{Met-Ψ}{^5f^5C34} structure both in solution by NMR and bound to
AUG and AUA in crystals of the ribosome’s 30S subunit. In lieu of empirical date, we have
completed a molecular dynamics simulation (MDS) of the 5′-monophosphates of the A:f^5C
pairing in explicit H_2O under neutral conditions. We hypothesized that the hmtRNA^{Met} could
use the formyl oxygen of f^5C_{34} similarly to the O^5 of the cmo^5U_{34} of the tRNA^{Val}. Therefore,
we restrained the f^5C with a hydrogen bond between the C_{33} 2′-OH and the O^5 of the formyl
group of f^5C_{34} and observed formation of a bond between f^5C_{34} of the anticodon and A3 of
the codon. In the lowest energy structure, the nucleosides paired in the C3′-endo, anti
conformation. The structure also showed one hydrogen bond formed between the f^5C_34 and A3 in a sheared or off-centered base pairing (Fig. 7). Though not evident from the MDS modeling, a second hydrogen bond may very well occur via a H_2O bridge between the f^5C_34 and A3.

In the mitochondria of some higher organisms, the Isoleucine codons AUU and AUC, along with Methionine codons AUA and AUG, are used as start codons. Thus, the entire codon box AUN is used in the translation initiation of some organisms. The formyl-modification has been found in the mitochondrial tRNA^{Met} of the higher eukaryotes fruit flies, nematodes, squids, frogs, chickens, rats, bovine and humans. Although the hmtRNA^{Met}_{CAU-f^5C_34} has previously been shown to decode the universal Isoleucine codon AUA, we report here for the first time that the anticodon stem and loop of hmtRNA^{Met}_{CAU-f^5C_34} (hmtASL^{Met}_{CAU-Ψ_27;f^5C_34}) is also capable of reading the universal Ile codons AUU and AUC when these codons are used for translation initiation. We have shown that the 5-formyl modification is responsible for enabling efficient binding to the codons AUU and AUC. Therefore, the formyl modification at the wobble position of hmtRNA^{Met}_{CAU-f^5C_34} expands codon recognition of the codon box AUN during translation initiation. Yeast is unable to use AUU, AUC, nor AUA for initiation (Bonnefoy and Fox 2000). We suspect that the lack of the 5-formyl modification at the wobble position of yeast mitochondrial tRNA^{Met}_{CAU} (Canaday et al. 1980) may limit its ability to expand codon recognition to read AUA, AUC, and AUU.
The Genetic Code is not static, but evolving. Its evolution may occur within the confines of tRNA. Thus, the structural features of tRNA, such as modified nucleosides within the anticodon loop, may influence codon reassignment (Yokobari et al. 2001). Modifications at the ASL modulate codon recognition and may aid in the rearrangement of the Genetic Code. Some modifications, such as $s^2U_{34}$ and its various 5-position derivatives, restrict codon recognition to a few codons (Yarian et al. 2000; Phelps et al. 2004; Murphy et al. 2004); other modifications, such as the $xO^5U_{34}$ derivatives expand codon recognition to more codons (Yarian et al. 2002; Vendeix et al. 2007). The 5-formyl modification, as we report here, expands codon recognition to the unconventional Methionine codons AUA, AUC, and AUU.

**MATERIALS AND METHODS**

*Preparation of hmtASL$^\text{Met}_{\text{CAU}}f^5C_{34}$ and the unmodified hmtASL$^\text{Met}_{\text{CAU}}*

The nucleoside $f^5C$ was synthesized as previously described (Lusic et al. 2008), and site-specifically inserted into the chemically-synthesized hmtASL$^\text{Met}_{\text{CAU}}\Psi_{27};f^5C_{34}$ using “ACE” chemistry (Scaringe et al. 2004; Dharmacon ThermoFisher, Lafayette, CO). The hmtASL$^\text{Met}_{\text{CAU}}\Psi_{27}$ was also produced by Dharmacon. The ASLs were then chemically deprotected following Dharmacon’s procedure, HPLC purified by ion exchange chromatography, desalted and analyzed for nucleoside composition by HPLC (Lusic et al., 2008).
Ribosomal Binding Assay

The ribosomal binding assays were performed using purified *E. coli* 70S ribosomes, chemically synthesized mRNAs, and chemically synthesized ASLs. The 27-nucleotide mRNA oligomers mimic a portion of T4 gp32 mRNA (Fahlman et al. 2004) and were purchased from Dharmacron (ThermoFisher, Lafayette, CO). Each of the mRNA sequences, entered into the program RNA Structure 4.2 (Mathews et al. 2007), was found to have a low likelihood of forming a stable secondary structure. Their sequences were as follows (mitochondrial Methionine codons AUA and AUG are in bold type):

1) 5’-GGCAAGGAGGUAAAAAUAUGUAGCAGCU-3’;
2) 5’-GGCAAGGAGGUAAAAAUGGUAGCAGCU-3’;
3) 5’-GGCAAGGAGGUAAGUAAUAGCAGCU-3’;
4) 5’-GGCAAGGAGGUAAGUAAUAGGCACGU-3’.

The 70S ribosomal subunits were prepared from *E. coli* MRE600 grown to mid-log phase (Shapkina et al. 2004). The ASLs were 5’-end 32P-labeled using [γ-32P] ATP (MP Biomedicals). Unlabeled ASLs in a range of concentrations (0-5µM) were mixed with insignificant amounts, but radiochemically detectable amounts of 32P, 5’-endlabeled ASLs (2,000-5,000 CPM). A fixed ratio of unlabeled ASL to labeled ASL was maintained for each concentration within the range studied. The assay was performed in ribosomal binding buffer [50 mM HEPES, pH 7.0; 30 mM KCl; 70 mM NH₄Cl; 1 mM DTT; 100 µM EDTA; 20 mM MgCl₂]. Ribosomes (0.25 µM) were activated by heating to 42 °C, incubated for 10 min and then slowly cooled to 37 °C. The ribosomes were programmed with mRNA (2.5 µM) for 15 min at 37 °C. When experimenting with the A-site, the P-site was saturated with
*E. coli* ASL$^{Val3\_UAC}$ (unmodified) for 15 min at 37 °C. ASL$^{Val3\_UAC}$ binds to the Val codon GUA; see underlined codons of the mRNA sequences above. Binding of the hmtASL$^{Met\_CAU}$ in either A or P-site was allowed to proceed for 30 min at 37 °C. The reaction mixtures (20 μL each) were then placed on ice for 20 minutes, diluted with 100 μL buffer per reaction mixture, and filtered through nitrocellulose in a modified Whatman Schleicher and Schuell (Brentford, U.K.) 96-well filtration apparatus (Wong and Lohman 1993). Prior to filtration of experimental samples, the nitrocellulose filter was equilibrated with the binding buffer at 4 °C for at least 20 min and each well of the filtration apparatus was washed with 100 μL of cold binding buffer. After the reacted samples were filtered, each well was then washed twice with 100 μL of cold binding buffer. The nitrocellulose was blotted dry with kim wipes, and the radioactivity was measured using a phosphorimager (Molecular Dynamics, GE Healthcare). Data were measured for radioactive intensity using ImageQuant (Amersham). Nonspecific binding was determined by the binding of ASLs to ribosomes without mRNA and subtracted from the experimental data. The final data is a result of at least three separate experiments, each done with samples in triplicate, ie. minimally nine determinations for each binding.

**Ribosomal Kinetic Assays**

The kinetic assays were conducted in similar fashion as the ribosomal binding assays described above with the exception that only one concentration of ASL was used (0.5 μM) and binding was monitored on a time course. Aliquots of ribosomal reactions were filtered
after: 0.0, 0.25, 0.5, 1, 5, 10, 15, 20, 30 minutes. The nitrocellulose was washed, dried, and processed as described above.

**Bovine Mitochondrial 55S Binding Assays**

Bovine 55S mitochondria were programmed with a chemically synthesized 30-mer oligonucleotide. Mammalian mitochondrial mRNAs are leaderless and translation initiation begins at the first nucleoside triplet on the mRNA. Therefore, to study binding at the P-site, we designed the mRNAs (see the following sequences) with an AUG or AUA (in bold) as the first three nucleosides, followed by the Valine codon GUA (underlined) at the A-site, and a stretch of poly-Uridines:

1) 5’-**AUG**GUAUUUUUUUUUUUUUUUUUUUUUUU-3’
2) 5’-**AUA**GUAUUUUUUUUUUUUUUUUUUUUU-3’

The mRNAs of mammalian mitochondria do not consist of a ribosomal binding sequence such as the Shine-Dalgarno sequence found in bacterial mRNAs. However, an mRNA consisting of stretches of poly(U) with the mitochondrial Methionine codon AUG has been shown to be successful in translating Methionine using an mitochondrial translational system in vitro (Takemoto et al. 1995). Therefore, this poly(U)-Methionine codon mRNA can be incorporated into the mitochondrial 55S ribosome. To study the binding of hmtASL\textsuperscript{Met} to the A-site, we incorporated the Val codon GUA (intended for the P-site; underlined in the following mRNA sequences), followed by AUG or AUA (in bold) into a string of 24 Uridines:

1) 5’-UUUUUUUUUUUUUUUUUUU**GUA**AUGUUUUUUUUUUUUUUU-3’
2) 5’-UUUUUUUUUUUUUUUUUU**GUA**AUAUUUUUUUUUUUUU-3’
Molecular Dynamics Simulations

The Molecular Dynamics Simulations (MDS) were conducted on the IBM Blade Center Linux Cluster (Henry2) at the High-Performance Computing center of North Carolina State University (Please refer to http://www.ncsu.edu/itd/hpc/main.php for more details about the configuration of the system). The methods used for MDS and base pair free energy calculations were based on published protocols (Stofer et al. 1999). The Amber 9 package (Case et al. 2005) was used to carry out molecular dynamics simulation based on X-ray structure of UUC used for initial geometries. Modifications were added by replacing individual molecules in unmodified base PDB files with those from the RNA modified parameters database. Nucleoside 5\'-monophosphates were neutralized with sodium and solvated with TIP3P water in an octahedral box, and then minimized in two steps, first the water then the whole system. MDS were then performed. First, a gradual heating step was used to raise the temperature from 0 to 300 K with positional restraints and a force constant of 50 kcal/mol-Å² for 10ps. Then a step down model was used to reduce positional restraints progressively from 50 to 10, and 5kcal/mol-Å². Each decrease of positional restraint constant was followed by 25ps of equilibration, until the final step was completely unrestrained, and the total time of simulation was equal to 100ps.

ACKNOWLEDGMENTS

We would like to thank William Darnell Graham for purifying the ASLs. We are grateful to Anthony Munoz for conducting the MDS of the structure determination of the A:f5C base pair. Funding was provided by the United Mitochondrial Disease Foundation to
L.L.S. with P.F.A. (grant number 05-20) and by NSF (grant number MCB0548602 to P.F.A.), an NSF Graduate Fellowship to E.M.G and by Dharmacon RNA Technologies, Inc. (ThermoFisher, Inc.).
REFERENCES


Crozier RH, Crozier YC. 1993. The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization. Genetics 133: 97-117.


Phelps SS, Malkiewicz A, Agris PF, Joseph S. 2004. Modified nucleotides in tRNA(Lys) and tRNA(Val) are important for translocation. *J Mol Biol* 338: 439-44.


Tomita K, Ueda T, Watanabe K. 1997. 5-formylcytidine (f5C) found at the wobble position of the anticodon of squid mitochondrial tRNA(Met)CAU. *Nucleic Acids Symp Ser* 197–198.


TABLE 1: Affinity of Unmodified and f^5C_34-Modified hmtASL^{Met}_{CAU-Ψ_{27}} for the Mitochondrial Methionine Codons at *E. coli* Ribosomal P or A-site

<table>
<thead>
<tr>
<th>hmtASL^{Met}<em>{CAU-Ψ</em>{27}}</th>
<th>Binding Affinity (K_d, μM)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-site</td>
<td>A-site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AUG</td>
<td>AUA</td>
<td>AUC</td>
<td>AUU</td>
<td>AUG</td>
<td>AUA</td>
</tr>
<tr>
<td>unmodified</td>
<td>1.2 ± 0.3</td>
<td>8.5 ± 1.5</td>
<td>36.4±14.3</td>
<td>8.2 ± 1.7</td>
<td>1.3 ± 0.1</td>
<td>11.0 ± 1.3</td>
</tr>
<tr>
<td>f^5C_34</td>
<td>0.9 ± 0.2</td>
<td>2.5 ± 0.1</td>
<td>9.1 ± 1.7</td>
<td>5.8 ± 0.8</td>
<td>1.2 ± 0.1</td>
<td>2.7 ± 0.4</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1. The mammalian mitochondrial Genetic Code. Deviations from the Universal Genetic Code are emphasized in bold. In the mammalian mitochondrial Genetic Code, the universal Stop codon UGA decodes Tryptophan, the universal Arginine codons AGA and AGG terminate translation, and the universal Isoleucine codon AUA decodes Methionine. In some mitochondrial genomes, the universal Isoleucine codons AUU and AUC, in addition to AUA and AUG, initiate translation at the peptidyl (P)-site of ribosomes. Thus, the whole codon-box AUN (shaded in blue) is used in translation initiation in the mitochondria.

Figure 2. Human mitochondrial tRNA$^{\text{Met}}_{\text{CAU}}$. The sequence and secondary structure of hmtRNA$^{\text{Met}}_{\text{CAU}}$ are shown with its heptadecamer, anticodon stem and loop domain, in bold (mtASL$^{\text{Met}}$). Nucleosides 27 and 50 are modified to pseudouridine, Ψ. The wobble position-34 modification is 5-formylcytidine, f$^5$C$_{34}$.

Figure 3. Ribosomal binding curves of the wobble-modified hmtASL$^{\text{Met}}_{\text{CAU}-\Psi_{27};f^5C_{34}}$ (▲) or the hmtASL$^{\text{Met}}_{\text{CAU}-\Psi_{27}}$ (■). *E. coli* 70S ribosomes were programmed with the mitochondrial Methionine codons AUG or AUA at the P or A-site. When observing binding at the A-site, the P-site was saturated with *E. coli* ASL$^{\text{Val}}_{\text{UAC}}$, which binds to the cognate Valine codon GUA. The equilibrium binding of the two ASLs to: A. AUG at the P-site; B. AUG at the A-site; C. AUA at the P-site; and D. AUA at the A-site.
Figure 4. Binding kinetics of the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsubscript{5}C\textsubscript{34}} (▲) and the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}} (■) to AUG or AUA at the P- and A-sites. \textit{E. coli} 70S ribosomes were programmed with the mitochondrial Methionine codons AUG or AUA at the P- or A-sites. When observing binding at the A-site, the P-site was saturated with \textit{E. coli} ASL\textsuperscript{Val}UAC, which binds to its cognate Valine codon GUA. The hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsubscript{5}C\textsubscript{34}} or hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}} (0.5\,\mu\text{M}) were added to ribosomes (0.25\,\mu\text{M}) programmed with mRNAs (2.5\,\mu\text{M}) with the codon: A. AUG at the P-site; B. AUG at the A-site; C. AUA at the P-site; or D. AUA at the A-site at progressing time points.

Figure 5. Codon binding of the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsubscript{5}C\textsubscript{34}} (black bars) and the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}} (white bars) on the bovine 55S mitochondrial ribosome. Bovine liver mitochondrial 55S ribosomes were programmed with mitochondrial Methionine codons AUG or AUA at the P- or A-sites. When observing binding at the A-site, the P-site was blocked with \textit{E. coli} ASL\textsuperscript{Val}UAC, which binds to its cognate Valine codon GUA.

Figure 6. Equilibrium binding of the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsubscript{5}C\textsubscript{34}} (dark colors) or the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}} (light colors) to the unconventional Methionine codons AUU, AUC, and AUA. \textit{E. coli} 70S ribosomes were programmed with the mitochondrial Methionine codons AUU, AUC, AUA at the P-site. The Valine codon GUA was used as the negative control. The hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27};f\textsubscript{5}C\textsubscript{34}} (5\,\mu\text{M}) bound non-specifically to the Valine codon GUA to a degree similar to that of the hmtASL\textsuperscript{Met}\textsubscript{CAU-Ψ\textsubscript{27}} (striated bars). Binding of the ASLs to specific codons, and above that of the negative control, are represented by the solid bars.
Figure 7. The molecular dynamics simulation of the A:5^C base pair. The base pairing of the 5’-monophosphates of A and 5^C was simulated in explicit H2O and under neutral conditions. The formyl oxygen was restrained by a hydrogen bond to the 2’-OH of the C33 of hmtRNA^Met (Materials and Methods).
**Figure 1.** The mammalian mitochondrial Genetic Code. Deviations from the Universal Genetic Code are emphasized in bold. In the mammalian mitochondrial Genetic Code, the universal Stop codon UGA decodes Tryptophan, the universal Arginine codons AGA and AGG terminate translation, and the universal Isoleucine codon AUA decodes Methionine. In some mitochondrial genomes, the universal Isoleucine codons AUU and AUC, in addition to AUA and AUG, initiate translation at the peptidyl (P)-site of ribosomes. Thus, the whole codon-box AUN (shaded in blue) is used in translation initiation in the mitochondria.
Figure 2. Human mitochondrial tRNA$^{\text{Met}}_{\text{CAU}}$. The sequence and secondary structure of hmtRNA$^{\text{Met}}_{\text{CAU}}$ are shown with its heptadecamer, anticodon stem and loop domain, in bold (mtASL$^{\text{Met}}$). Nucleosides 27 and 50 are modified to pseudouridine, $\Psi$. The wobble position-34 modification is 5-formylcytidine, $f^5\text{C}_{34}$. 
Figure 3

A. 

B. 

C. 

D. 

**Figure 3.** Ribosomal binding curves of the wobble-modified hmtASL<sup>Met</sup><sub>CAU-Ψ<sub>27</sub>;f</sub>C<sub>34</sub> (▲) or the hmtASL<sup>Met</sup><sub>CAU-Ψ<sub>27</sub></sub> (■). *E. coli* 70S ribosomes were programmed with the mitochondrial Methionine codons AUG or AUA at the P or A-site. When observing binding at the A-site, the P-site was saturated with *E. coli* ASL<sup>Val</sup><sub>UAC</sub>, which binds to the cognate Valine codon GUA. The equilibrium binding of the two ASLs to: **A.** AUG at the P-site; **B.** AUG at the A-site; **C.** AUA at the P-site; and **D.** AUA at the A-site.
Figure 4

Figure 4. Binding kinetics of the hmtASL$_{\text{CAU}-\Psi_{27};f^{5}C_{34}}$ (▲) and the hmtASL$_{\text{CAU-}\Psi_{27}}$ (■) to AUG or AUA at the P- and A-sites. *E. coli* 70S ribosomes were programmed with the mitochondrial Methionine codons AUG or AUA at the P- or A-sites. When observing binding at the A-site, the P-site was saturated with *E. coli* ASL$_{\text{Val-UAC}}$, which binds to its cognate Valine codon GUA. The hmtASL$_{\text{CAU-}\Psi_{27};f^{5}C_{34}}$ and hmtASL$_{\text{CAU-}\Psi_{27}}$ (0.5 μM) were added to ribosomes (0.25 μM) programmed with mRNAs (2.5 μM) with the codon: A. AUG at the P-site; B. AUG at the A-site; C. AUA at the P-site; or D. AUA at the A-site at progressing time points.
Figure 5. Codon binding of the hmtASL\textsuperscript{Met}_{CAU-Ψ_{27},f^{6}C_{34}} (black bars) and the hmtASL\textsuperscript{Met}_{CAU-Ψ_{27}} (white bars) on the bovine 55S mitochondrial ribosome. Bovine liver mitochondrial 55S ribosomes were programmed with mitochondrial Methionine codons AUG or AUA at the P- or A-sites. When observing binding at the A-site, the P-site was blocked with \textit{E. coli} ASL\textsuperscript{Val}_{UAC}, which binds to its cognate Valine codon GUA.
Figure 6. Equilibrium binding of the hmtASL$_{\text{Met}}^{\text{CAU}-\Psi_{27};f^{5}\text{C}_{34}}$ (dark colors) or the hmtASL$_{\text{Met}}^{\text{CAU}-\Psi_{27}}$ (light colors) to the unconventional Methionine codons AUU, AUC, and AUA. *E. coli* 70S ribosomes were programmed with the mitochondrial Methionine codons AUU, AUC, AUA at the P-site. The Valine codon GUA was used as the negative control. The hmtASL$_{\text{Met}}^{\text{CAU}-\Psi_{27};f^{5}\text{C}_{34}}$ (5 μM) bound non-specifically to the Valine codon GUA to a degree similar to that of the hmtASL$_{\text{Met}}^{\text{CAU}-\Psi_{27}}$ (striated bars). Binding of the ASLs to specific codons, and above that of the negative control, are represented by the solid bars.
Figure 7. The molecular dynamics simulation of the A:⋯f$\text{C}$ base pair. The base pairing of the 5'$\text{-}$monophosphates of A and f$\text{C}$ was simulated in explicit H$_2$O and under neutral conditions. The formyl oxygen was restrained by a hydrogen bond to the 2'$\text{-}$OH of the C$_{33}$ of hmtRNA$^{\text{Met}}$ (Materials and Methods).
6.1 The Evolution of tRNA and the Genetic Code

Modified nucleosides were discovered early in nucleic acid research [1,2]. The 50 plus years of tRNA studies have revealed the significance of modifications to the structure of tRNA, and therefore the functions of tRNA [3,4]. Recent work has shown evidence for the hypothesis that all tRNAs bind the ribosome with similar affinity [5]. Thus, we can speculate that tRNA chemistry and structure have evolved in concert so that the ribosome does not fully distinguish one tRNA from the other, and they bind with one mechanism. Modified nucleosides are ancient and highly conserved among all phylogenetic kingdoms [2-4]. Although there are slight variations, similar modifications are found at the anticodon loop of similar tRNA species among different organisms. Cognate codon:anticodon binding in the ribosome allows for precise and efficient decoding. Non-cognate codon:anticodon pairings are allowed in the ribosome specifically at the wobble position [5]. Interestingly, like the cognate codon:anticodon pairings, many of these non-cognate pairings proceed through the translational apparatus with accurate efficiency. Oftentimes, the wobble position of the anticodon is post-transcriptionally modified. The tRNA’s anticodon loop at position 37 (nucleoside directly adjacent to the anticodon) is also often found modified. These two modified nucleosides at the tRNA anticodon loop have been shown to enhance fidelity and efficiency in decoding [6]. We hypothesize that modifications at the anticodon loop allow the tRNA to bind non-cognate codons as it would cognate codons (Figure 1) [7].
Modifications modulate the decoding of synonymous codons by restricting or expanding tRNA’s ability to recognize codons [3].

Because tRNA modifications are so significant to tRNA chemistry and structure, they affect tRNA functions. tRNA modifications are important to tRNA’s ability to act as a primer for reverse transcription during HIV infection, to participate in cell cycle regulation, and to translate proteins accurately and efficiently. Thus, tRNA modifications have important implications in the regulation of gene expression.
Figure 1. Accommodation of tRNA on the ribosome (derived from [7]). The tRNA ternary complex (aminoacyl-tRNA-EFTu-GTP) enters the A-site of the ribosome. If the anticodon loop is left unmodified, it is disordered and dynamic and cannot bind its codon effectively. Therefore, accommodation cannot occur. However, if the anticodon loop is modified, it is more rigid and ordered, making contact with its codon at the A-site more effective. This allows the ribosome to accommodate the tRNA, and GTP is hydrolyzed.
6. 2  Implications of tRNA Modifications on Gene-Expression Regulation


The more than 100 modification chemistries are analogous to the side chains of amino acids in their contribution of hydrophobic or hydrophilic properties to RNA structure [3]. Cells devote a great deal of resources to the modification of RNA. In fact, considerably more genetic information is allocated to tRNA modifications than to tRNA genes [6]. Of the seventy-five modifications found in tRNAs, those occurring in the physically and functionally separable anticodon stem and loop domain (ASL) are best understood (Figure 2) [3]. Here, we focus on the control of gene expression associated with the physicochemical contributions of the wobble position-34 and purine-37 modifications that introduce order into tRNA’s anticodon loop. This order influences the recognition of codons and maintenance of the translational reading frame.
Figure 2. Modifications of the anticodon stem and loop (ASL) of tRNA order the loop preventing frameshifting and allowing for accurate codon selection. (a) Modifications such as $s^2\text{C}$ at (◈), $s^2\text{U}_{34}$, $\text{mmn}^5\text{U}_{34}$, cmo$^5\text{U}_{34}$ and Q at wobble position-34 (◈) and $t^6\text{A}$, $m^1\text{G}$, ms$^2\text{o}^6\text{A}$ and $y\text{W}$ at position-37 (◈) structure the anticodon toward the canonical “U-turn (at invariant U$_{33}$) and increase stability by facilitating the stacking of the aromatic nucleoside bases of positions 34-37. Purine-37 modifications negate intra-loop base pairing (-----). (b) Representative prokaryotic modified nucleosides that are involved in control of rare codon recognition and reading frame maintenance/frameshifting. The chemical structures and locations (positions 32, 34 or 37) of the modifications in tRNA’s anticodon domain are shown. Abbreviated names are defined in the text. Eukaryotic modifications with similar chemistries, such as mcm$^5s^2\text{U}$, can be found at the RNA Modification Database (URL: http://library.med.utah.edu/RNAmods/).
6.2A. Modifications select and enable codon recognition.

The genetic code is degenerate in that most amino acids have more than one codon. (Figure 3). In all organisms and organelles, codons outnumber the tRNAs required to translate them. Modifications of nucleosides within the anticodon loop modulate binding to certain codons, and therefore imbue specific tRNAs with the ability to regulate gene expression [3, 8]. One particularly interesting mechanism of gene regulation is the use of codon bias [9]. Codon bias is the disparity among synonymous codon usage. Among genomes, synonymous codons are not used in equal frequencies. Surprisingly, the use of rare codons in mRNAs is associated with a protein folding that differs in three-dimensional structure from that derived of mRNAs with common synonymous codons [10]. As would be expected, codon bias is linked to tRNA isoacceptor abundance in the cell [11]. tRNA abundance, in turn, is correlated to environmental responses [12]. In *E. coli*, usage of the six codons of arginine varies from 1.4 to 24.1 per 1000 codons [13]. The most commonly used arginine codons are CGU and CGC. The other four, CGA, CGG, AGA, AGG, are rare codons found in low frequency in *E. coli* mRNAs. Rare codon clustering or positioning near initiation or termination codons, such as with the rare arginine codons AGA or AGG, cause ribosome stalling and frameshifting related to the low abundance of the tRNAs that read these codons [14-16]. AGA and AGG are found in DNA replication and other regulatory genes [17] in *E. coli* and are responsible for fimbrie (pili) formation in *Salmonella* [18]. *E. coli* has five tRNA isoacceptors that read the six arginine codons [19]. The tRNA²Arg<sub>UCU</sub> isoacceptor decodes the two regulatory codons AGA and AGG [17]. tRNA²Arg<sub>UCU</sub> has a 2-thiocytidine (s²C₃₂) modification at position-32, 5-methylamino-methyluridine (mmn⁵U₃₄) at
wobble position 34, and $N^6$-threonylcarbamoyl-adenosine at position-37 ($t^6A_{37}$) (Table 1) [19]. These modifications may be properties of certain tRNAs responding to rare codons AGA and AGG. The $s^2C_{32}$ influences the decoding of AGG [20]. In a clear distinction from the tRNA isoacceptor that reads the rare codons AGA and AGG, the isoacceptors that read the two common codons CGU and CGC contain the modified nucleoside inosine ($I_{34}$) at the wobble position-34 (Table 1). The difference in modifications between tRNAs that read rare codons and those that recognize common codons implicates tRNA modifications in regulation via rare codon usage. Though AGA and AGG are the most used arginine codons in humans and other eukaryotes [21], the modification-dependent recognition of these codons is regulatory when the codons appear in significantly greater numbers in some proteins than in others. The yeast tRNA$^{Arg}_{UCU}$ isoacceptor with 5-methoxycarbonylmethyluridine at wobble position-34 ($mcm^5U_{34}$) is linked to regulating the translation of numerous of these codons in the DNA damage response factors (Table 1) [22]. In expression of eukaryotic proteins in E. coli, low yield from codon bias is alleviated by altering codon content of the gene or overexpressing the tRNA responsible for recognizing rare codons [23]. The fifth rarest codon in E. coli is that of isoleucine, AUA (Figure 3). Isoleucine has three isoacceptors, but only the tRNA$^{Ile}_{CAU}$ isoacceptor has a lysine modification of $C_{34}$ (lysidine) (Table 1) [3, 19]. The other two have an unmodified $G_{34}$ [19]. The unmodified tRNA$^{Ile}_{CAU}$ would incorrectly recognize the methionine codon AUG, but is modified for decoding AUA [24]. Thus, again a tRNA modification is responsible for a rare codon recognition of isoleucine. Therefore, in family codon boxes, specific tRNA wobble modifications may control expression by recognizing synonymous codons that are rarely used and only for
particular genes. In controlling gene expression, the $s^{2}$-, $\text{mnm}^{5}$-, $t^{6}\text{A}$ and other modifications of these tRNAs restrict nucleoside conformation, structure the anticodon, and reduce anticodon domain conformational dynamics in recognition of specific codons [3].

Figure 3. The Universal Genetic Code. The 61 sense codons are color shaded according to extent of degeneracy: single codes are in grey (■); twofold degenerate codes, pink (■); threefold, yellow (■); fourfold, green (■); two- plus fourfold or 6-fold degenerate codes in blue (■). The stop codons are in red. Accurate and efficient recognition of codons are dependent on the specific modifications at the anticodon stem and loop (ASL) of tRNA and the modifications’ ability to restrict or expand codon box recognition. Rare codons that are discussed are in bold.
The tRNA\textsuperscript{Lys} species of all organisms have the modifications \textit{mmn}\textsuperscript{5}s\textsuperscript{2}U\textsubscript{34} and \textit{t}\textsuperscript{6}A\textsubscript{37}, or their derivatives (Table 1) \cite{19}. The two modifications are both required for the tRNA to bind the two-fold degenerate, cognate AAA and wobble AAG, codons at the ribosomes A-site and to translocate from A- to P-site \cite{25-27}. For lysine codons, specificity resides in the third position or wobble position, because they share a ‘mixed codon box’ with the two asparagine codons (Figure 3). While they do not necessarily prevent the misreading of near-cognate codons, these modifications significantly increase the recognition of cognate lysine codons \cite{28}. In yeast, genetic deletion of enzymes responsible for modifying tRNA wobble position of tRNA\textsuperscript{Lys} is lethal to the cells \cite{28}. Thus, modifications that contribute structure and order serve to restrict codon recognition of a mixed codon box.

In contrast to lysine and asparagine codons, alanine, leucine, proline, serine, threonine and valine codons are all four-fold degenerate (‘family codon box’) (Figure 3). One isoaccepting tRNA species for each of these amino acids has the wobble position-34 modification, 5-oxyacetic acid uridine modification, cmo\textsuperscript{5}U\textsubscript{34} (Table 1). In \textit{S. enterica}, there are two tRNA\textsuperscript{Val} isoacceptors that decode the four valine codons. The ASL of \textit{E. coli} tRNA\textsuperscript{Val\textsubscript{UAC}}, with the anticodon UAC, decodes the complementary codon GUA in the absence of the modification cmo\textsuperscript{5}U\textsubscript{34} (F. A. P. Vendeix \textit{et al.}, unpublished). However, the modification is required for decoding the other valine codons, including GUG \cite{29, 30}. \textit{In vivo}, mutants having only the one tRNA\textsuperscript{Val} isoacceptor with cmo\textsuperscript{5}U\textsubscript{34} were viable, indicating that this wobble-modified tRNA\textsuperscript{Val} is capable of decoding all four valine codons \cite{29}. A genetic knockout of all tRNA\textsuperscript{Pro} isoacceptors except the isoacceptor with cmo\textsuperscript{5}U\textsubscript{34} allowed
survival of *S. enterica*. Yet, the deletion of the cmo$^{5}$U$_{34}$-modifying enzyme resulted in a significant reduction of growth [31].

6.3 Anticodon Stem and Loop Modifications Modulate the Extent of tRNA’s Codon Reading Flexibility

The tRNA’s job of decoding can be complicated in that, oftentimes, it must read multiple types of codons and must be highly accurate and efficient in its reading. Sometimes one tRNA must read several but not all codons in a codon box (split box) and must, therefore, be wary not to decode the similar codons in that box that are assigned to another tRNA. Other times, an entire codon box is devoted to one amino acid, and a tRNA must be efficient enough to read the whole box. The tRNA’s plurality in codon recognition is modulated by the modifications at the anticodon loop domain. These modifications are the key to understanding the mechanisms regulating tRNA’s ability to restrict recognition in split boxes and expand recognition in whole boxes.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contribution of modifications to codon recognition, frameshifting, and as aminoacyl-tRNA synthetase identity determinants</strong></td>
</tr>
<tr>
<td>Modification</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>$\text{cyl}$</td>
</tr>
<tr>
<td>$\text{cmo}^{5}U_{34}$</td>
</tr>
<tr>
<td>$\text{mt}^{5}U_{34}$</td>
</tr>
<tr>
<td>$\text{y}^{5}W_{34}$</td>
</tr>
<tr>
<td>$\text{f}^{5}A_{34}$</td>
</tr>
<tr>
<td>$\text{mm}^{5}U_{34}$</td>
</tr>
<tr>
<td>$\text{km}^{5}U_{34}$</td>
</tr>
<tr>
<td>$\text{s}^{5}O_{34}$</td>
</tr>
<tr>
<td>$\text{s}^{5}N_{34}$</td>
</tr>
</tbody>
</table>

ND denotes not determined.  
$\oplus$ Eukaryotic modification.
The $\text{xm}^5\text{s}^2$ modification is found at the wobble position-34 of tRNAs that read codons ending with A or G in the third column of the universal Genetic Code (codons that code for the amino acids Glutamine, Lysine, and Glutamic Acid) [3, 19]. This wobble-modification restricts tRNA’s wobbling to only A or G but not U or C. U and C at the wobble position of codons in the third column of the universal Genetic Code (codons that code for the amino acids Histidine, Aspartic acid, Asparagine and Tyrosine) are read by tRNAs with a modified G, Queuosine, at the tRNA’s wobble position-34 [19]. Thus, the $\text{xm}^5\text{s}^2\text{U}_{34}$ and Q modifications at tRNA’s wobble position restrict codon recognition in split codon boxes. Modifications at position 37 have also been shown to enhance the function of wobble-modified ASLs [3, 4, 6].

Eight of the twenty amino acids have at least four codons all in the same codon box (Leucine, Valine, Serine, Proline, Threonine, Alanine, Arginine, and Glycine). Six of these eight (Leucine, Valine, Serine, Proline, Threonine, Alanine) have tRNAs with the 5-oxy-derivative (xcmo$^5$) modification at the wobble position [3]. This one tRNA with the xcmo$^5$ modification at the wobble position is capable of decoding all four synonymous codons in a box. Thus, the xcmo$^5$ modification increases the efficiency of one tRNA by allowing it to read the synonymous codons in the same box.

The greatest extent of wobble efficiency is best exhibited by the mitochondrial translational system. The mammalian mitochondrion consists of an extremely limited amount of only 22 tRNA species, less than half of the 45 found in E. coli, yet translation of the Genetic Code into the 20 amino acids occurs accurately and efficiently in the mitochondrion’s ribosomes. One mitochondrial tRNA, the tRNA$^{\text{Met}}$, is quite efficient in that
it acts in both initiation and elongation. Furthermore, while the tRNA\textsuperscript{Met} outside of the mitochondria is responsible for only one Met codon, this one tRNA\textsuperscript{Met} reads two to four Met codons. The difference between the cytoplasmic and mitochondrial tRNA\textsuperscript{Met}s is the presence of the rare 5-formyl modification at the wobble position of mt tRNA\textsuperscript{Met}. Our comparisons of the 5-formly-modified ASL of the human mitochondrial tRNA\textsuperscript{Met} (hmtASL\textsuperscript{Met}_{CAU-Ψ27;f5C34}) with the wobble-unmodified hmtASL\textsuperscript{Met}_{CAU-Ψ27} have linked the ASL’s efficiency in codon recognition directly to the 5-formyl modification.

tRNA modifications at the anticodon loop domain are found in a great variety of organisms and have likely evolved with the ancient RNA. It is not difficult to speculate that modifications evolve the tRNA into the roles it is assigned in life, especially its central role in protein synthesis. Modifications may have facilitated in the assignment of codons, such as the splitting of codon boxes or the addition of synonymous codons in the same codon box. In the mitochondria, the Genetic Code can deviate from the universal Code in that some codons are reassigned to different amino acids (Figure 3). Wobble modifications may aid in recognizing and efficiently reading these reassigned codes.
REFERENCES


