ABSTRACT

CANTARA, WILLIAM ANTHONY. Arginyl-tRNA Modifications Modulate Anticodon Domain Structure, Function and Dynamics in *Escherichia coli.* (Under the direction of Paul F. Agris)

During ribosome-mediated protein synthesis, non-initiator tRNAs act as translating chaperones by transferring the correct amino acid to the ribosome through sequence-specific interactions with mRNA codons in the ribosomal A-site. To achieve uniformity in codon recognition ability between different tRNAs, the anticodon domains of all must adopt a singular conformation in the ribosomal decoding center. Since there is a necessary sequence variation for the three anticodon residues, innovative ways of attaining a constant loop structure are required. Nature has resolved this issue by introducing functional groups to the loop residues, changing the chemical environment and, thus, the conformation. In fact, there is a large diversity and number of these modifications found in tRNAs of all known life.

Escherichia coli (*E.coli*) arginyl-tRNAs offer the opportunity to study these modifications in the context of six-fold degenerate codons that vary widely in their cellular usage. A single set of very similar isoacceptors, $tRNA^{Arg1}_{ICG}$ and $tRNA^{Arg2}_{ICG}$, have anticodon domains that differ only in the presence or absence of a very rare modification, 2-thiocytidine (s^2C_{32}). These tRNAs are particularly interesting not only because of the rare s^2C_{32} modification, but also through the wobble position modification I_{34}, they are both responsible for decoding two very common codons CGU and CGC as well as the very rare CGA codon. Typically, the codon usage correlates with the tRNA content of the cell; however, in this instance, the tRNA content is high which does not match what would be expected for an isoacceptor that recognizes a rare codon. Here we have shown conclusively that modifications can act to restrict the decoding capability of the tRNA to only the two common codons, suggesting that the tRNA is present at high levels, consistent with decoding common codons, but that there is a smaller pool of tRNA in different states of modification, corresponding to decoding of the rare codon. The mechanism of this function is not completely understood, however, we show that the modifications do not accomplish this through a completely structural device. In this case, modifications act to modulate the flexibility of specific loop residues, thereby

modulating the conformational landscape that must be traversed between two disparate, but necessary, cellular conformations.

It is clear from many previous studies that modifications can perform specific functions in the anticodon stem and loop (ASL), most notably at position 34 which is known to facilitate the "wobble" recognition of multiple codons. *E.coli* $tRNA^{Arg4}_{UCU}$ offers a novel function of modification at this position that is not fully understood. In a very similar isoacceptor, partially modified $tRNA^{Lys}_{UUU}$, a very different decoding function is seen. Both of these tRNAs contain the same loop sequence with the exception that $tRNA^{Arg4}_{UCU}$ has a C at position 35, whereas $tRNA^{Lys}_{UUU}$ has U_{35}. Interestingly, they both contain the same loop modifications, mnm^5U_{34} and t^6A_{37}, with the exception that $tRNA^{Arg4}_{UCU}$ has an additional s^2C_{32} modification; however $tRNA^{Lys}_{UUU}$, in this particular modification state can decode both AAA and AAG and $tRNA^{Arg4}_{UCU}$ decodes only AGA. Although s^2C_{32} enhances codon discrimination in $tRNA^{Arg1}_{ICG}$, codon-specific ribosome binding assays clearly refute this role in $tRNA^{Arg4}_{UCU}$. The modifications, however, appear to impart similar biophysical and thermodynamic properties to the ASL as in $tRNA^{Lys}_{UUU}$, therefore a full structural characterization is underway using NMR spectroscopy to ascertain whether modifications drive the conformation toward a canonical U-turn conformation as with $tRNA^{Lys}_{UUU}$. Preliminary evidence indicates a role for modifications in altering the stability of the non-canonical C_{32}•A_{38} base pair through stable stacking interactions and elevation of A_{38}H1 pKa.

Arginyl-tRNA Modifications Modulate Anticodon Domain Structure, Function and Dynamics in *Escherichia coli*

by
William Anthony Cantara

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

Biochemistry

Raleigh, North Carolina

2012

APPROVED BY:

Paul Wollenzien Committee Chair	Paul F. Agris
E. Stuart Maxwell	James W. Brown

DEDICATION

To Shraddha, my loving wife, who keeps me on the right path through life and is someone that I can always count on. Also, to my mother and other family members for their sacrifices throughout my life and for nurturing my curiosity. Finally, to my friends and labmates for their support and advice.

BIOGRAPHY

William Anthony Cantara spent part of his childhood in York, Pennsylvania until moving to Grassflat, Pennsylvania during kindergarten. He spent his childhood and high school years enjoying rural life and participating in athletics. He attended Juniata College, where he met his wife and graduated with a bachelor's degree in Biochemistry and Physics. During college, his lab experiences with Dr. Jill Keeney showed him that research was his true calling. He then moved to Raleigh, NC with his future wife to work under the guidance of Dr. Paul Agris at North Carolina State University. Will looks forward to using the skills and experiences he has gained over the past five years to lead a successful career as an independent researcher.

ACKNOWLEDGEMENTS

First, I would like to express gratitude to my advisor, Paul F. Agris, for helping me to develop into a researcher. His mentoring and discussions about the inner workings of a lab and research facility have opened my eyes to the nuances that must be considered when starting a lab and the joys associated with the process. Also, my advisory committee has offered meaningful insights and knew how to offer constructive criticism to ensure that my projects were comprehensive. In particular, I acknowledge the willingness of E. Stuart Maxwell for opening his lab to me during my first year and offering his graduate student, Keith Gagnon, to teach me cloning techniques. I extend gratitude to the faculty, staff and graduate students of the Department of Molecular and Structural Biochemistry at North Carolina State University and the Life Sciences Building at the University at Albany for their friendship, advice and willingness to offer their expertise.

No acknowledgement would be complete without noting the contributions of current Agris lab members and alumni. I would like to specifically thank William Graham, Franck Vendeix, Estella Gustilo, Yann Bilbille and Kun Lu for their hard work and contributions to my projects. I would also like to thank the undergraduates that I have had the opportunity to mentor and have assisted with various aspects of my projects: Jia Kim, Minhal Makshood and Erick Harr. My friendships and experiences over the past five years with members of the Agris lab in both Raleigh and Albany have shaped who I am as a researcher and I owe a great deal of my successes to them.

Finally, I thank my loving wife. Not only has she supported my research and acted as a constant inspiration to me, but she has been unwavering in her willingness to help me overcome all obstacles during my graduate career. From staying by my side during a brief medical scare to leaving her job and moving with me to New York, she has gone above and beyond what I could have hoped for.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

APPENDICES

CHAPTER 1. Introduction and literature review

1.1 Transfer ribonucleic acid (tRNA): functions and modification

Protein synthesis is the culmination of a process known as the central dogma of molecular biology [1]. In this activity, the information stored within 2'-deoxyribonuclic acid (DNA) is transcribed into messenger ribonucleic acid (mRNA). The mRNA is then loaded onto a ribosome composed of ribosomal ribonucleic acid (rRNA) and protein. Once on the ribosome, triplet nucleotide sequences, or codons, on the mRNA are recognized by the anticodon residues of transfer ribonucleic acid (tRNA), which bring the correct amino acid to the ribosome for addition to the growing peptide chain. Since there are 64 possible triplet combinations of the four canonical RNA residues in the genetic code (61 that code for amino acids) and only 20 common amino acids, there is degeneracy in the code (Figure 1). In addition to this, there are often fewer tRNAs than codons. The main cellular role of tRNA is to bring the correct amino acid to the ribosome for protein synthesis through base pairing between the three anticodon residues in the tRNA and the codon residues of the mRNA. Therefore, to properly navigate the degenerate genetic code, tRNAs with many different anticodon sequences must conform to a common structure when bound to the ribosome and have the ability to specifically recognize multiple codons in certain situations.

The tRNA secondary structure is in the shape of a cloverleaf [2] consisting of the aminoacyl acceptor stem, a variable loop and three stem and loop domains: the dihydrouridine stem and loop (D-loop), the thymidine stem and loop (T-loop or TψC-loop) and the anticodon stem and loop (ASL) domain (Figure 2a). As shown in early crystal structures of yeast $tRNA^{Phe}$ [3,4] and confirmed a few years later with yeast $tRNA^{Asp}$ [5,6], the tRNA folds into an L-shaped tertiary structure in which the aminoacyl acceptor stem is at one end and the ASL is at the other (Figure 2b; [7]). Interestingly, tRNAs with unusual secondary structure characteristics such as the eukaryotic selenocysteine $tRNA^{Sec}$ [8] show that the highly canonical L-shaped tertiary structure is not a necessity for tRNA function.

However, the high resolution structures show that the locations and rotational displacement of the aminoacyl acceptor stem and ASL are highly conserved even in this unusual structure [9]. It should also be noted that the decoding region, the ASL, is not stabilized or structured by any tertiary interactions from the other tRNA domains. Therefore, it must independently conform to the geometry necessary for decoding in the ribosomal A-site as well as carry the chemistry required for base pairing with multiple codons. Nature accomplishes this feat through chemical modification of the tRNA residues.

During maturation of a tRNA transcript, specific nucleoside residues are chemically modified via reactions catalyzed by modification enzymes [10]. The high propensity for tRNAs to be modified is inherent in the fact that two of the stem loops are named by the highly conserved modified residues contained within them, dihydrouridine in the D-loop and thymidine and pseudouridine in the T-loop. In fact, of the 109 currently known naturally occurring post-transcriptional RNA modifications, 92 are found in tRNA [11]. At least twelve of these, including the common ASL modifications N^6-threonylcarbamoyl adenosine (t^6A) and inosine (I), can be found in tRNA from all three phylogenetic domains [11-13]. While the diversity of tRNA modification chemistries highlights the probability that specific modifications evolved to perform specific functions based on the needs of individual species, the phylogenetic conservation of a subset of them suggests that tRNAs and their modifications may have even been important to the progenitor prior to the divergence of phylogenetic domains.

The cellular roles of many tRNA modifications have yet to be deciphered; however, their high importance can be inferred from the energetic cost and the amount of genetic information devoted to modifications as well as their abundant usage in tRNAs. In a "back of the napkin" calculation in 1995, Glenn Bjork estimated that tRNA genes make up ~0.25% of the *Escherichia coli* genome, while at least four times that amount, ~1%, is devoted to tRNA-modifying enzymes [12]. While the energy consumption required for generating small amounts of these modification enzymes is relatively low, one must consider that many of these enzymes require *S*-adenosylmethionine (SAM), for which the synthesis of just one

molecule requires 12 molecules of ATP [14-16]. The highly modified nature of tRNAs adds to the energy requirement. In fact, up to 25% of the tRNA residues are modified in some eukaryotes [13]. The large amount of modifications present in tRNAs and the energetic devotion required highlights their importance to all cells.

The tRNA anticodon domain contains a number of conserved sequence properties (Figure 3). In addition to being the most heavily and variously modified RNA domain, the ASL is always characterized by a five base-pair stem and seven nucleotide loop. Within the stem, there is very little sequence conservation; however, the nucleotides co-vary to allow for a highly stable stem consisting of A-form RNA with Watson-Crick base pairing. More pronounced sequence conservation occurs in the loop, with the exception of the anticodon residues, which must co-vary with cognate and synonymous codons. Within the loop, the most common sequence is 5'-$C_{32}U_{33}N_{34}N_{35}N_{36}A_{37}A_{38}$-3', which is found in 38.5% of all known tRNA sequences. In fact, U_{33} is nearly universal, occurring in all cytoplasmic, plastid and viral tRNAs and in all but seven known mitochondrial tRNAs [13]. While position 37 is nearly always a purine (>99%), the great majority of the time (>70%), this comes in the form of an adenosine [13]. Finally, other than the anticodon, the least conserved of the loop residues are at positions 32 and 38 where C_{32} (~68%) and A_{38} (~69%) are most common [13]. Interestingly, this sequence conservation does not necessarily cause the loop to occupy a conserved conformation. Here, modifications can cause alterations in both conformation and structural dynamics in order to overcome the conformational limitations that sequence conservation cannot overcome.

1.2 Anticodon domain modifications

1.2.1 Position 34 modifications

Within the subset of modifications found in tRNAs, the largest number and greatest chemical diversity can be found within the ASL domain [11,13]. The most commonly

modified residues in the ASL occur at the first anticodon residue, position 34, and 3' adjacent to the anticodon, position 37 [11,13]. Position 34, known as the "wobble" position (see nucleoside numbering in Figures 3,5,6), is so called because of Francis Crick's famous "Wobble Hypothesis" in 1966, which notes the unusual ability of uridine (U), guanosine (G) or inosine (I) at this position to "wobble" base pair with two or three different codons [17]. This ability to "wobble" explains how tRNAs with these position 34 residues can decode two or three codons differing in their third residue; thus, accounting for a certain amount of degeneracy within the genetic code.

The extensive study and the large diversity of modification chemistries located at position 34, most specifically in modified uridines, prompted the "Modified Wobble Hypothesis" in 1991 [18]. This hypothesis encompassed the evidence that while U, G and I can expand the decoding ability of the anticodon, modifications can modulate the expansion or restriction of wobble decoding to fine-tune these interactions [18]. In the case of many mitochondria [19,20] and plastids [21,22], an unmodified wobble position uridine is sufficient to expand the decoding ability of the isoacceptor to read all four codons in a box; however, cytosolic tRNAs with wobble position uridines are rarely able to efficiently decode any codons in the absence of modification. Indeed, 84% of the uridines found at the wobble position of cytosolic tRNAs are modified [23] to either expand [24-27] or restrict [28-31] the decoding ability of these uridines. Interestingly, in contrast to the differences in the Watson-Crick faces of the wobble nucleosides that prompted the original Wobble Hypothesis, the largest and most diverse of these uridine modifications occur at the 5-position, opposite the Watson-Crick face [11]. It is clear that for these 5-modified uridines to regulate wobble decoding, the structure and geometry of the nucleoside is just as important as the chemistry of the Watson-Crick face. In fact, a strong correlation has been found between the conformation of the ribose moiety and the tRNAs decoding ability. When a 5-position modification of uridine promotes either a C2'-*endo* or C3'-*endo* sugar pucker, there tends to be an expansion or restriction of the decoding capacity, respectively [18,32,33]. However, it should also be noted that the 5-modified uridines that restrict wobble decoding also tend to have a 2-

thiolation, which does change the chemistry of the Watson-Crick face, but also further stabilizes the C3'-*endo* sugar conformation [34-37].

Position 34 adenosines (A) are nearly universally modified to the guanosine analong inosine. In fact, only four instances [13] have been found where a wobble position adenosine is not modified to inosine (*Mycoplasma capricolum* [38,39] and *Mycoplasma mycoides* [40,41] $tRNA^{Thr}_{AGU}$ and the mitochondrial $tRNA^{Arg}_{ACG}$ of *Saccharomyces cerevisiae* [42] and *Ascaris suum* [43]). In comparison to bacteria where only $tRNA^{Arg}_{ICG}$ contains inosine, in eukaryotes all fourfold degenerate codon boxes are decoded by at least one isoacceptor that contains I_{34}. As mentioned above, inosine at position 34 (I_{34}) in the ASL allows recognition of codons ending in U, C or A. Although the positioning of the I_{34}•A_3 anticodon•codon base pair suggested in the original wobble hypothesis was in the traditional Watson-Crick geometry [17], this does not account for the increased size of the two purine residues within the ribosomal A-site. It was theorized that the geometry of this purine•purine base pair required that the inosine must adopt a *syn N*-glycosidic bond conformation, exposing the Hoogsteen face for base pairing with the Watson-Crick face of A_{38} [44]. A crystal structure of the I_{34} modified *E. coli* ASL^{Arg}_{ICG} in the ribosomal A-site revealed that the I_{34}•A_3 base pair is indeed in the Watson-Crick geometry, and this is accomplished by a compression of the ribose and phosphate moieties between C1′ and P, most drastically in the β (P-O5′-C5′-C4′) torsion angle, which changes from about ±180˚ to -37.8˚ [45].

1.2.2 Position 37 Modifications

Another commonly modified nucleoside in the ASL occurs with the purine residue at position 37. This purine residue is so highly conserved and highly modified that >99% of the known tRNA sequences contain a purine at position 37 and >70% are modified [13,46]. Modifications at this position have a correlation to the identity of the residue at position 36 [47,48]. A notable example of this occurs where tRNAs with U_{36}, with very few exceptions, also contain N^6-threonylcarbamoyl adenosine (t^6A) or its derivatives at position 37 [13]. These hypermodified t^6A_{37} derivatives have been shown, through structural determination, to

form a base stacking interaction over the A_1•U_{36} codon•anticodon base pair, thereby stabilizing the low enthalpy of this base pair [47,49]. In addition to stabilization, structural studies of the human $tRNA^{Lys3}_{UUU}$ ASL indicate a role for t^6A_{37} and $ms^2t^6A_{37}$ in negating intraloop hydrogen bonding in the pyrimidine-rich loop, resulting in an opening of the tri-nucleotide loop into a canonical 7-membered loop [50,51]. This loop rearrangement ability appears to be a pre-requisite for the formation of a U-turn motif and further stabilize this conformation.

For ASLs that have G_{36}, 78% of the position 37 residues are methylated purines [13]. In most cases (83%) this methylated purine is 1-methylguanosine (m^1G_{37}); however, in some cases (17%) there is either a 2-methyladenosine (m^2A_{37}) or 6-methyladenosine (m^6A_{37}) [13]. The increased stability of C_1•G_{36} codon•anticodon base pairs reduces the need for a stabilizing stacking interaction from the position 37 nucleoside. Instead, the small hydrophobic modifications prevent translational frameshifting [52,53] and significantly increase affinity of the cognate codon•anticodon interaction [54]. The positioning of the methylations on the Watson-Crick face of these purines also underscores the possibility that these modifications may effect frameshifting or misreading by preventing either base pairing of these residues to the nucleoside 5' adjacent to the codon [53] or to the invariant U_{33} 5' adjacent to the anticodon, disrupting the possibility of forming a tri-nucleotide loop structure.

1.2.3 Position 32 Modifications

Finally, although modifications at position 32 are not usually discussed with commonly modified anticodon domain modifications, ~32% of known tRNA sequences contain a modification at this position [13]. This nucleoside, at the base of the stem is nearly always a pyrimidine (>99%) [13] and is located directly across the loop from the commonly unmodified and less conserved residue at position 38 (Figure 3). The geometry of the non-canonical interaction between position 32 and position 38 residues (including the modified residues pseudouridine and 2'-*O*-methylcytidine at position 32) has been thoroughly studied and shown to significantly affect the ability of the tRNA to discriminate between codons

[55]. In all of these structures, the geometry of this base pair is stabilized by stacking interactions with the stem residues, requiring a stabilized A-form-like C3'-*endo* sugar pucker [56-61]. Interestingly, 83% of nucleoside modifications that occur at position 32 are either 2-thiocytidine or 2'-*O*-methylated pyrimidines [13], both of which stabilize the C3'-*endo* conformation. Specifically, the increased size of the thioketone and the large, hydrophobic nature of the 2'-*O*-methyl group each cause steric clashes in the C2'-*endo* conformation resulting in a significant preference for the C3'-*endo* pucker [33,62-64]. Additionally, the 2-thiolated pyrimidines experience greater stacking interactions caused by the greater polarizability of sulfur in comparison to oxygen [37]. Therefore the highly modified nature of this residue may have important implications for the tRNAs ability to decode.

1.3 Modulation of anticodon stem and loop structure, stability and function

We have now established that modification of tRNA is done at great cost in terms of genetic material, precursors, modifications enzymess and cellular energy, and that modifications are prominent aspects of the anticodon domain. Functionally, ASL modifications are important for accuracy and efficiency of protein synthesis. These modifications have been shown to be important for many different features of translation such as thermal stability of the ASL [28,65-70], decoding specificity [70-74], ribosomal codon binding affinity [28,65,75-77], proper translocation [77] and reading frame maintenance [78-81]. Additionally, pre-structuring of the ASL toward a canonical U-turn structure in solution for proper recognition in the ribosomal A-site is a common function of anticodon domain modifications [70,78,82,83]. The formation of a U-turn-like structure in solution is beneficial to the cell because it reduces the energy barrier of binding to synonymous mRNA codons in the ribosomal A-site.

The U-turn motif has many characteristics that make it ideal for proper function in the decoding center of the ribosomal A-site (Figure 4). Structurally, there is an abrupt change in direction of the RNA backbone at the universal U_{33} residue and a well-ordered base stacking

network along the 3' side of the loop that results in the three anticodon residues being exposed in the proper geometry for codon binding [45,49]. Also, a non-Watson-Crick base pair is typically formed between the first two residues of the loop (N_{32}•N_{38}), resulting in a pseudo-5nt loop. Finally, the highly stacked nature of the 3' side of the loop results in a highly stabilized stem loop structure. To achieve this ideal structure in all tRNAs, a high amount of sequence conservation has evolved in the non-anticodon residues of the anticodon loop (Figure 3). However, the need for differences in the sequence of the three nucleotide anticodon can significantly alter the structure of the loop. Indeed it has been shown that the highly dynamic nature of the pyrimidine rich anticodon of unmodified ASL^{Lys}_{UUU} causes the formation of stable cross-loop base pairs, resulting in a very rigid tri-nucelotide loop conformation [84,85].

Since differences in anticodon sequence can cause disruption of the U-turn conformation, chemical modifications to the anticodon loop residues can restore a more canonical structure. In the case of ASL^{Lys}_{UUU}, modifications at positions 34 and 37 not only disrupt the cross-loop interactions that cause the tri-nucleotide loop conformation, but also contribute to a prestructuring of the ASL into a canonical U-turn motif [51,84,85]. A fully canonical U-turn conformation is not a necessity, however, for binding to the ribosomal A-site. As seen with ASL^{Val}_{UAC} [24,86] and ASL^{Lys}_{UUU} [51], small deviations are seen in solution structures of the modified ASL to that of the crystal structures on the ribosome. The energy barrier associated with the conformational transition between the solution structure and the structure on the ribosome is a determining factor for fine tuning of induced-fit remodeling and, therefore, of translational speed and accuracy. The fine tuning capability associated with the strength and geometry of the N_{32}•N_{38} cross-loop interaction [55,87] indicates a possible role for chemical modifications at positions 32 and 37 to have either subtle or drastic effects on the energy barrier associated with this structural transition. Additionally, as noted above, modifications at position 34 can alter the sugar pucker at this position, significantly affecting the structural characteristics of the anticodon loop [18,32,33]. The effects of these modifications on loop structure and flexibility must be more pronounced in isoacceptors that require similar

conformational abilities for recognition by a single aminoacyl-tRNA synthetase, but have multiple anticodon nucleotide sequences such as the case of six-fold degenerate codons.

In addition to their structural function, ASL modifications have recently been shown to be dynamic and can regulate gene expression. It has been suggested that tRNA modification levels may be used as a regulatory device in response to such factors as antibiotics, iron levels and aerobic *vs* anaerobic growth [88]. Indeed, recent evidence shows alteration in the state of overall tRNA modification during stress conditions in *Saccharomyces cerevisiae* [89]. Also, the efficient synthesis of 5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U) in *Saccharomyces pombe* is required for proper transition through the cell cycle in a modification dependent manner [90]. While there is substantial sequence data for tRNAs and their modifications [13], during purification and determination of modified residues present in many isoacceptors, wild type cells were typically grown under only one environmental circumstance (ie. in rich media at optimal growth temperature) to obtain a suitable amount of material; therefore it remains unclear as to whether modifications respond to specific environmental or stress conditions as a common regulatory device.

1.4 Six-fold degenerate decoding of arginine in *Escherichia coli*

Arginine shares the distinction (with leucine and serine) of being one of only three amino acids that is decoded by six codons (Figure 1). In *Escherichia coli*, these six codons are decoded by five tRNA isoacceptors (Figure 5; [13,91,92]). Interestingly, the usage of these six codons in *E. coli* varies wildly from > 2% of all codons for CGU and CGC to < 0.6% for CGA, CGG, AGA and AGG (Table 1; [93-95]). Three of the four less common codons (CGG, AGA and AGG) are decoded by individual isoacceptors, whereas the two most common codons (CGU and CGC) are decoded by the same isoacceptors ($tRNA^{Arg1,2}_{ICG}$) as the rare CGA codon. This marks a rare instance in which a single tRNA isoacceptors is solely responsible for recognition of both very common and very rare codons.

While the primary sequences of their anticodon domains are quite different, commonalities can be found in arginyl-isoacceptors. In addition to the less notable invariant U_{33}, highly conserved A_{38} and middle anticodon nucleotide C_{35}, the rare 2-thiocytidine modification is present at position 32 in four of the five isoacceptors. This modification is only found in $tRNA^{Arg1,3,4,5}$ and $tRNA^{Ser2}_{GCU}$ in select bacteria and archaea [13]. Extensive studies have been performed to elucidate the function of 2-thiolated uridines at the wobble position [28,49,51,71,84]; however, relatively little is known about s^2C_{32}. Chemical similarities between the two pyrimidines suggest that a 2-thiolation of cytidine would elicit similar structural properties as seen in s^2U. In comparison to oxygen, the sulfur atom has a lower electronegativity, resulting in greater stacking interactions for s^2U_{34} [37]. In addition, as mentioned above, the increased size of sulfur over oxygen in s^2U_{34} causes steric effects forcing the ribose to adopt a C3'-*endo* pucker conformation and enhancing the *anti*-conformation of the *N*-glycosidic (χ) dihedral angle [34-37]. Empirical evidence of this was shown in the altered accessibility of the $tRNA^{Arg1,2}_{ICG}$ anticodon loop to nuclease S1 upon chemical conversion of s^2C_{32} to C_{32} [96]. Functionally, s^2C_{32} promotes reading frame maintenance [96] and reduces the rate of tRNA selection into the ribosomal A-site [97]. Interestingly, however, a *Salmonella enterica* mutant without s^2C_{32} exhibits wild type growth under optimal growth conditions, thus it has been proposed that the s^2C_{32} modification could be important for improving accuracy by reducing the speed of translation [97], a function that may only be necessary during suboptimal growth conditions.

Three of the six arginine codons (CGU, CGC and CGA) are decoded by the $tRNA^{Arg1}_{ICG}$ and $tRNA^{Arg2}_{ICG}$ isoacceptors. In *E. coli*, these two isoacceptors are expressed from a common gene and, thus, their primary sequences are nearly identical. There are only two differences between these two tRNAs: (1) $tRNA^{Arg2}_{ICG}$ contains an additional adenosine nucleotide at position 20 (A_{20}) in the dihydrouridine loop and (2) $tRNA^{Arg1}_{ICG}$ contains the rare modification 2-thiocytidine at position 32 (s^2C_{32}), whereas $tRNA^{Arg2}_{ICG}$ has an unmodified C_{32} (Figure 5; [13]). Interestingly, both A_{20} and s^2C_{32} are present in $tRNA^{Arg3,4,5}$. In addition to the s^2C_{32} found only in $tRNA^{Arg1}_{ICG}$, both $tRNA^{Arg1}_{ICG}$ and $tRNA^{Arg2}_{ICG}$ contain

modifications at position 34 (inosine, I_{34}) and position 37 (2-methyladenosine, m^2A_{37}). Wobble position inosines (I_{34}) are a very common adenosine modification. In fact, as mentioned above there are only four known instances where this position contains an unmodified adenosine [13,38-43]. Inosine has the well-known function of enabling tRNAs to decode codons ending in U, C or A, whereas an unmodified adenosine and guanosine recognize only U, and U or C, respectively. In a majority of cases, I_{34} is accompanied by a modification at position 32 or a modification at position 37 (~67% and ~82%, respectively), and ~56% in both positions [13]. While, the function of I_{34} can be assumed to be expansion of the decoding capability to three codons, the roles of the additional modifications at positions 32 and 37 remain unclear. More specifically, the roles of the rare $tRNA^{Arg1,2}_{ICG}$ modifications s^2C_{32} and m^2A_{37} (present in only five and 13 known tRNA isoacceptors, respectively [13]), in the context of I_{34} in tRNA structure and function are not understood.

Another interesting isoacceptor is *E. coli* $tRNA^{Arg4}_{UCU}$. Of particular interest is that in this tRNA the 5-methylaminomethyluridine at position 34 (mnm^5U_{34}) causes this isoacceptor to recognize only one of the rare codons AGA, but not AGG [30], whereas, in *E. coli* $tRNA^{Lys}_{UUU}$, it confers the ability to read both AAA and AAG codons [28]. Ribosome binding studies directed toward understanding each chemical modification moiety using the anticodon domain of $tRNA^{Lys}_{UUU}$ confirm that the mnm^5 modification alone is sufficient to allow recognition of both codons [28,75]. Indeed, a crystal structure of this ASL bound to AAG in the A-site of the 30S ribosomal subunit clearly shows that the recognition is allowed by the specific positioning of the mnm^5U_{34} residue which positions it in just the right geometry for a bifurcated hydrogen bond between the O2 of mnm^5U_{34} and the N1H and N2H of the codon G_3 [49]. The only differences in the loop sequence of *E. coli* $tRNA^{Lys}_{UUU}$ and $tRNA^{Arg4}_{UCU}$ is the identity of the second anticodon nucleoside and the s^2C_{32} modification in $tRNA^{Arg4}_{UCU}$ (Figure 6). If both anticodon loops are so similar, how is it that they have different decoding functions when it comes to wobble codon discrimination?

Three projects are presented herein in which we ask the questions: (1) What roles, if any, do anticodon domain modifications play in the ability of $tRNA^{Arg1,2}_{ICG}$ to distinguish between

the common CGU and CGC codons and the rare CGA codon? (2) Are modifications important for modulating flexibility of the loop and, thus, altering the energy barrier between different necessary conformations? (3) How do s^2C_{32} and/or C_{35} cause a restricted decoding functionality for $tRNA^{Arg4}_{UCU}$ when the very similar anticodon loop of $tRNA^{Lys}_{UUU}$ promotes wobble decoding? To study these tRNAs in a modification dependent manner, a smaller model system consisting of the ASL was utilized. As mentioned previously, this portion of the tRNA does not make any contact with any other tRNA domains (Figure 2b). Also, many NMR solution and X-ray crystal structures show a very high degree of conformational similarity between ASLs and their corresponding tRNAs [3,24,51,66,86]. This information, coupled with corroborating ribosome binding data, indicates that the ASL is a sufficient model system to test the effects of anticodon domain modifications. With this system, it is possible to select and site specifically incorporate modified residues into the heptadecamer hairpin using standard RNA synthesis chemistry. Using synthesized ASLs in different states of modification, comprehensive biochemical analyses such as functional binding assays, thermodynamic characterization and structure determination have given key insights into answering the questions posed above. Additionally, computational approaches have been utilized where empirical experiments were not feasible.

1.5 References

1. Crick F: Central dogma of molecular biology. *Nature* 1970, **227**(5258):561-563.
2. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A: Structure of a Ribonucleic Acid. *Science* 1965, **147**:1462-1465.
3. Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A: Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature* 1974, **250**(467):546-551.
4. Suddath FL, Quigley GJ, McPherson A, Sneden D, Kim JJ, Kim SH, Rich A: Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. *Nature* 1974, **248**(443):20-24.
5. Moras D, Comarmond MB, Fischer J, Weiss R, Thierry JC, Ebel JP, Giege R: Crystal structure of yeast tRNAAsp. *Nature* 1980, **288**(5792):669-674.

6. Giege R, Moras D, Thierry JC: Yeast transfer RNAasp: a new high-resolution x-ray diffracting crystal form of a transfer RNA. *J Mol Biol* 1977, **115**(1):91-96.

7. Bénas P, Bec G, Keith G, Marquet R, Ehresmann C, Ehresmann B, Dumas P: The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA* 2000, **6**(10):1347-1355.

8. Sturchler C, Westhof E, Carbon P, Krol A: Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic Acids Res* 1993, **21**(5):1073-1079.

9. Itoh Y, Chiba S, Sekine S, Yokoyama S: Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res* 2009, **37**(18):6259-6268.

10. Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K: MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res* 2009, **37**(Database issue):D118-121.

11. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FA, Fabris D, Agris PF: The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 2011, **39**(Database issue):D195-201.

12. Björk G: Biosynthesis and function of modified nucleosides. In: *tRNA : structure, biosynthesis, and function.* Edited by Söll D, RajBhandary U. Washington, D.C.: ASM Press; 1995: 165-205.

13. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J: tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 2009, **37**(Database issue):D159-162.

14. Stock JB, Simms S: Advances in post-translational modifications of proteins and aging. In: *Advances in experimental medicine and biology.* Edited by Zappia V, Galletti P, Porta R, Wold F. New York: Plenum Press; 1988: 201-212.

15. Atkinson DE: Cellular energy metabolism and its regulation. New York: Academic Press; 1977.

16. Bakin A, Lane BG, Ofengand J: Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* 1994, **33**(45):13475-13483.

17. Crick FH: Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* 1966, **19**(2):548-555.

18. Agris PF: Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* 1991, **73**(11):1345-1349.

19. Lagerkvist U: "Two out of three": an alternative method for codon reading. *Proc Natl Acad Sci U S A* 1978, **75**(4):1759-1762.

20. Lagerkvist U: Unorthodox codon reading and the evolution of the genetic code. *Cell* 1981, **23**(2):305-306.

21. Wakasugi T, Ohme M, Shinozaki K, Sugiura M: Structures of tobacco chloroplast genes for tRNAIle (CAU), tRNALeu (CAA), tRNACys (GCA), tRNASer (UGA) and tRNAThr (GGU): a compilation of tRNA genes from tobacco chloroplasts. *Plant Mol Biol* 1986, **7**(5):385-392.

22. Rogalski M, Karcher D, Bock R: Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* 2008, **15**(2):192-198.

23. Gehrke CW, Kuo KCT: Chromatography and modification of nucleosides. New York, NY, U.S.A.: Elsevier; 1990.

24. Weixlbaumer A, Murphy FV, Dziergowska A, Malkiewicz A, Vendeix FAP, Agris PF, Ramakrishnan V: Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature Struct Biol* 2007, **14**(6):498-502.

25. Ishikura H, Yamada Y, Nishimura S: Structure of serine tRNA from Escherichia coli. I. Purification of serine tRNA's with different codon responses. *Biochim Biophys Acta* 1971, **228**(2):471-481.

26. Mitra SK, Lustig F, Akesson B, Axberg T, Elias P, Lagerkvist U: Relative efficiency of anticodons in reading the valine codons during protein synthesis in vitro. *J Biol Chem* 1979, **254**(14):6397-6401.

27. Samuelsson T, Elias P, Lustig F, Axberg T, Folsch G, Akesson B, Lagerkvist U: Aberrations of the classic codon reading scheme during protein synthesis in vitro. *J Biol Chem* 1980, **255**(10):4583-4588.

28. Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, Agris PF: Modified nucleoside dependent Watson-Crick and wobble codon binding by $tRNA^{Lys}_{UUU}$ species. *Biochemistry* 2000, **39**(44):13390-13395.

29. Sakamoto K, Kawai G, Niimi T, Satoh T, Sekine M, Yamaizumi Z, Nishimura S, Miyazawa T, Yokoyama S: A modified uridine in the first position of the anticodon of a minor species of arginine tRNA, the argU gene product, from Escherichia coli. *Eur J Biochem* 1993, **216**(2):369-375.

30. Spanjaard RA, Chen K, Walker JR, van Duin J: Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg). *Nucleic Acids Res* 1990, **18**(17):5031-5036.

31. Johansson MJ, Esberg A, Huang B, Bjork GR, Bystrom AS: Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. *Mol Cell Biol* 2008, **28**(10):3301-3312.

32. Agris PF: The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. *Prog Nucleic Acid Res Mol Biol* 1996, **53**:79-129.

33. Yokoyama S, Watanabe T, Murao K, Ishikura H, Yamaizumi Z, Nishimura S, Miyazawa T: Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc Natl Acad Sci USA* 1985, **82**(15):4905-4909.

34. Agris PF, Guenther R, Ingram PC, Basti MM, Stuart JW, Sochacka E, Malkiewicz A: Unconventional structure of tRNA(Lys)SUU anticodon explains tRNA's role in bacterial and mammalian ribosomal frameshifting and primer selection by HIV-1. *RNA* 1997, **3**(4):420-428.

35. Agris PF, Soll D, Seno T: Biological function of 2-thiouridine in Escherichia coli glutamic acid transfer ribonucleic acid. *Biochemistry* 1973, **12**(22):4331-4337.

36. Lustig F, Elias P, Axberg T, Samuelsson T, Tittawella I, Lagerkvist U: Codon reading and translational error. Reading of the glutamine and lysine codons during protein synthesis in vitro. *J Biol Chem* 1981, **256**(6):2635-2643.

37. Bilbille Y, Vendeix FAP, Guenther R, Malkiewicz A, Ariza X, Vilarrasa J, Agris PF: The structure of the human $tRNA^{Lys3}$ anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs. *Nucleic Acids Res* 2009, **37**(10):3342-3353.

38. Andachi Y, Yamao F, Iwami M, Muto A, Osawa S: Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in Mycoplasma capricolum. *Proc Natl Acad Sci U S A* 1987, **84**(21):7398-7402.

39. Andachi Y, Yamao F, Muto A, Osawa S: Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in Mycoplasma capricolum. Resemblance to mitochondria. *J Mol Biol* 1989, **209**(1):37-54.

40. Guindy YS, Samuelsson T, Johansen TI: Unconventional codon reading by Mycoplasma mycoides tRNAs as revealed by partial sequence analysis. *Biochem J* 1989, **258**(3):869-873.

41. Samuelsson T, Guindy YS, Lustig F, Boren T, Lagerkvist U: Apparent lack of discrimination in the reading of certain codons in Mycoplasma mycoides. *Proc Natl Acad Sci U S A* 1987, **84**(10):3166-3170.

42. Sibler AP, Dirheimer G, Martin RP: Codon reading patterns in Saccharomyces cerevisiae mitochondria based on sequences of mitochondrial tRNAs. *FEBS Lett* 1986, **194**(1):131-138.

43. Watanabe Y, Tsurui H, Ueda T, Furusihima-Shimogawara R, Takamiya S, Kita K, Nishikawa K, Watanabe K: Primary sequence of mitochondrial tRNA(Arg) of a nematode Ascaris suum: occurrence of unmodified adenosine at the first position of the anticodon. *Biochim Biophys Acta* 1997, **1350**(2):119-122.

44. Yokoyama S, Nishimura S: Modified nucleosides and codon recognition. In: *tRNA: Structure, biosynthesis and function.* Edited by Schimmel PR, Söll D, RajBhandary UL. Washington D.C.: ASM Press; 1995: 207-223.

45. Murphy FV, Ramakrishnan V: Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 2004, **11**(12):1251-1252.

46. Nierhaus KH, Wilson DN: Protein synthesis and ribosome structure: translating the genome. Weinheim, Germany: Wiley-VCH; 2004.

47. Nishimura S: Modified nucleosides in tRNA. In: *Transfer RNA: Structure, properties and recognition.* Edited by Schimmel PR, Söll D, Abelson J. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1979: 59-79.

48. Yokoyama S, Nishimura S: tRNA Structure, Biosynthesis and Function. In. Edited by Söll D, RajBhandary UL. Washington D.C.: ASM Press; 1995: 207-223.

49. Murphy FV, Ramakrishnan V, Malkiewicz A, Agris PF: The role of modifications in codon discrimination by $tRNA^{Lys}_{UUU}$. *Nature Struct Biol* 2004, **11**(12):1186-1191.

50. Stuart JW, Gdaniec Z, Guenther R, Marszalek M, Sochacka E, Malkiewicz A, Agris PF: Functional anticodon architecture of human $tRNA^{Lys3}$ includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t^6A. *Biochemistry* 2000, **39**(44):13396-13404.

51. Vendeix FA, Murphy FV, Cantara WA, Leszczynska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF: Human tRNA(Lys3)(UUU) Is Pre-Structured by Natural Modifications for Cognate and Wobble Codon Binding through Keto-Enol Tautomerism. *J Mol Biol* 2011.

52. Hagervall TG, Tuohy TM, Atkins JF, Bjork GR: Deficiency of 1-methylguanosine in tRNA from Salmonella typhimurium induces frameshifting by quadruplet translocation. *J Mol Biol* 1993, **232**(3):756-765.

53. Bjork GR, Wikstrom PM, Bystrom AS: Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science* 1989, **244**(4907):986-989.

54. Ashraf SS, Guenther RH, Ansari G, Malkiewicz A, Sochacka E, Agris PF: Role of modified nucleosides of yeast tRNA(Phe) in ribosomal binding. *Cell Biochem Biophys* 2000, **33**(3):241-252.

55. Olejniczak M, Uhlenbeck OC: tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 2006, **88**(8):943-950.

56. Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, Ramakrishnan V: Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 2001, **292**(5518):897-902.

57. Shi H, Moore PB: The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* 2000, **6**(8):1091-1105.

58. Westhof E, Dumas P, Moras D: Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr A* 1988, **44 (Pt 2)**:112-123.

59. Bullock TL, Uter N, Nissan TA, Perona JJ: Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J Mol Biol* 2003, **328**(2):395-408.

60. Nissen P, Thirup S, Kjeldgaard M, Nyborg J: The crystal structure of Cys-tRNACys-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. *Structure* 1999, **7**(2):143-156.

61. Yaremchuk A, Tukalo M, Grotli M, Cusack S: A succession of substrate induced conformational changes ensures the amino acid specificity of Thermus thermophilus prolyl-tRNA synthetase: comparison with histidyl-tRNA synthetase. *J Mol Biol* 2001, **309**(4):989-1002.

62. Smith WS, Sierzputowska-Gracz H, Sochacka E, Malkiewicz A, Agris PF: Chemistry and structure of modified uridine dinucleosides are determined by thiolation. *J Amer Chem Soc* 1992, **114**(21):7989-7997.

63. Kawai G, Yamamoto Y, Kamimura T, Masegi T, Sekine M, Hata T, Iimori T, Watanabe T, Miyazawa T, Yokoyama S: Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 1992, **31**(4):1040-1046.

64. Agris PF, Sierzputowska-Gracz H, Smith W, Malkiewicz A, Sochacka E, Nawrot B: Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J Amer Chem Soc* 1992, **114**(7):2652-2656.

65. Lusic H, Gustilo EM, Vendeix FA, Kaiser R, Delaney MO, Graham WD, Moye VA, Cantara WA, Agris PF, Deiters A: Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNAMet. *Nucleic Acids Res* 2008, **36**(20):6548-6557.

66. Stuart JW, Koshlap KM, Guenther R, Agris PF: Naturally-occurring modification restricts the anticodon domain conformational space of $tRNA^{Phe}$. *J Mol Biol* 2003, **334**(5):901-918.

67. Davis DR, Durant PC: Nucleoside modifications affect the structure and stability of the anticodon of tRNA(Lys,3). *Nucleosides Nucleotides* 1999, **18**(6-7):1579-1581.

68. Cabello-Villegas J, Winkler ME, Nikonowicz EP: Solution conformations of unmodified and A(37)N(6)-dimethylallyl modified anticodon stem-loops of Escherichia coli tRNA(Phe). *J Mol Biol* 2002, **319**(5):1015-1034.

69. Cabello-Villegas J, Nikonowicz EP: Solution structure of psi32-modified anticodon stem-loop of Escherichia coli tRNAPhe. *Nucleic Acids Res* 2005, **33**(22):6961-6971.

70. Denmon AP, Wang J, Nikonowicz EP: Conformation effects of base modification on the anticodon stem-loop of Bacillus subtilis tRNA(Tyr). *J Mol Biol* 2011, **412**(2):285-303.

71. Ashraf SS, Sochacka E, Cain R, Guenther R, Malkiewicz A, Agris PF: Single atom modification (O-->S) of tRNA confers ribosome binding. *RNA* 1999, **5**(2):188-194.

72. Li J, Esberg B, Curran JF, Bjork GR: Three modified nucleosides present in the anticodon stem and loop influence the in vivo aa-tRNA selection in a tRNA-dependent manner. *J Mol Biol* 1997, **271**(2):209-221.

73. Chiari Y, Dion K, Colborn J, Parmakelis A, Powell JR: On the possible role of tRNA base modifications in the evolution of codon usage: queuosine and Drosophila. *J Mol Evol* 2010, **70**(4):339-345.

74. Ninio J: Multiple stages in codon-anticodon recognition: double-trigger mechanisms and geometric constraints. *Biochimie* 2006, **88**(8):963-992.

75. Yarian C, Townsend H, Czestkowski W, Sochacka E, Malkiewicz AJ, Guenther R, Miskiewicz A, Agris PF: Accurate translation of the genetic code depends on tRNA modified nucleosides. *J Biol Chem* 2002, **277**(19):16391-16395.

76. Kurata S, Weixlbaumer A, Ohtsuki T, Shimazaki T, Wada T, Kirino Y, Takai K, Watanabe K, Ramakrishnan V, Suzuki T: Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J Biol Chem* 2008, **283**(27):18801-18811.

77. Phelps SS, Malkiewicz A, Agris PF, Joseph S: Modified nucleotides in $tRNA^{Lys}$ and $tRNA^{Val}$ are important for translocation. *J Mol Biol* 2004, **338**(3):439-444.

78. Gustilo EM, Vendeix FA, Agris PF: tRNA's modifications bring order to gene expression. *Curr Opinion in Microbiol* 2008, **11**(2):134-140.

79. Urbonavicius J, Qian Q, Durand JM, Hagervall TG, Björk GR: Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J* 2001, **20**(17):4863-4873.

80. Björk GR, Durand JM, Hagervall TG, Leipuviene R, Lundgren HK, Nilsson K, Chen P, Qian Q, Urbonavicius J: Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett* 1999, **452**(1-2):47-51.

81. Brierley I, Meredith MR, Bloys AJ, Hagervall TG: Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J Mol Biol* 1997, **270**(3):360-373.

82. Agris PF: Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* 2008, **9**(7):629-635.

83. Agris PF, Vendeix FAP, Graham WD: tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* 2007, **366**(1):1-13.

84. Durant PC, Bajji AC, Sundaram M, Kumar RK, Davis DR: Structural effects of hypermodified nucleosides in the *Escherichia coli* and human $tRNA^{Lys}$ anticodon loop: The effect of nucleosides s^2U, mcm^5U, mcm^5s^2U, mnm^5s^2U, t^6A, and ms^2t^6A *Biochemistry* 2005, **44**(22):8078-8089.

85. Durant PC, Davis DR: Stabilization of the anticodon stem-loop of $tRNA^{Lys,3}$ by an A+-C base-pair and by pseudouridine. *J Mol Biol* 1999, **285**(1):115-131.

86. Vendeix FAP, Dziergowska A, Gustilo EM, Graham WD, Sproat B, Malkiewicz A, Agris PF: Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry* 2008, **47**(23):6117-6129.

87. Olejniczak M, Dale T, Fahlman RP, Uhlenbeck OC: Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding. *Nat Struct Mol Biol* 2005, **12**(9):788-793.

88. Winkler ME: Genetics and Regulation of Base Modification in the tRNA and rRNA of Prokaryotes and Eukaryotes. In: *Modification and Editing of RNA.* Edited by Grosjean H, Benne R. Washington, D.C.: ASM Press; 1998: 441-470.

89. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ: A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet* 2010, **6**(12):e1001247.

90. Bauer F, Matsuyama A, Candiracci J, Dieu M, Scheliga J, Wolf D, Yoshida M, Hermand D: A coordinated codon dependent control of translation by Elongator. *EMBO J* 2012, **In Press**.

91. Söll D, Cherayil JD, Bock RM: Studies on polynucleotides. LXXV. Specificity of tRNA for codon recognition as studied by the ribosomal binding technique. *J Mol Biol* 1967, **29**(1):97-112.

92. Söll D, RajBhandary UL: Studies on polynucleotides. LXXVI. Specificity of transfer RNA for codon recognition as studied by amino acid incorporation. *J Mol Biol* 1967, **29**(1):113-124.

93. Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985, **2**(1):13-34.

94. Nakamura Y, Gojobori T, Ikemura T: Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 2000, **28**(1):292.

95. Wada K, Wada Y, Ishibashi F, Gojobori T, Ikemura T: Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res* 1992, **20 Suppl**:2111-2118.

96. Baumann U, Fischer W, Sprinzl M: Analysis of modification-dependent structural alterations in the anticodon loop of Escherichia coli tRNAArg and their effects on the translation of MS2 RNA. *Eur J Biochem* 1985, **152**(3):645-649.

97. Jager G, Leipuviene R, Pollard MG, Qian Q, Bjork GR: The conserved Cys-X1-X2-Cys motif present in the TtcA protein is required for the thiolation of cytidine in position 32 of tRNA from Salmonella enterica serovar Typhimurium. *J Bacteriol* 2004, **186**(3):750-757.

1.6 Tables and Figures

Table 1. Codon usage for arginine codons in *Escherichia coli*.

Strain	Arg1,2 CGU	Arg1,2 CGC	Arg1,2 CGA	Arg5 CGG	Arg4 AGA	Arg3 AGG
O157:H7 EDL933	20.2	20.8	3.8	6.2	2.9	1.8
O157:H7 str. Sakai	20.3	20.9	3.9	6.4	3.0	1.9
CFT073	20.3	21.0	3.9	6.3	2.9	1.9
536	20.8	21.1	3.7	5.6	2.4	1.5
UT189	20.4	21.2	3.8	6.2	2.8	1.9
APEC 01	20.3	20.8	4.0	6.2	3.0	2.0
Average	20.4 ±0.2	21 ±0.2	3.9 ±0.1	6.2 ±0.3	2.8 ±0.2	1.8 ±0.2

*All values are the frequency per 1000 codons in strains with CDS > 4500 from the Codon Usage Database [94]. The errors are calculated as one standard deviation. The isoacceptor responsible for decoding each codon is noted above the codon.

First Letter	Second Letter: U		Second Letter: C		Second Letter: A		Second Letter: G		Third Letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

Figure 1. The universal genetic code. The universal genetic code consists of 64 triplet codons, 61 code for amino acids and three signal the stop of translation (stop codons are highlighted red).

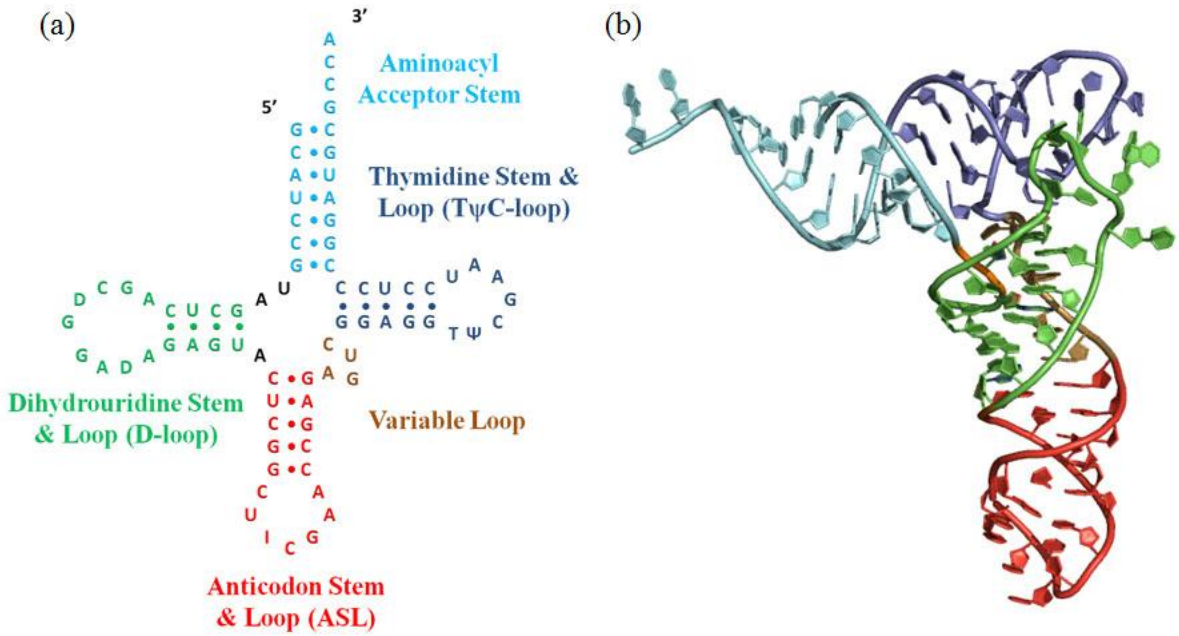

Figure 2. Representative examples of tRNA secondary and tertiary structures. (a) The $tRNA^{Arg1}_{ICG}$ isoacceptor from *Escherichia coli* secondary, or cloverleaf, structure has three characteristic stem and loop domains in addition to an aminoacyl acceptor stem and a variable loop. These domains are common to most tRNA isoacceptors. For clarity, the tRNA is shown with only the modifications that are characteristic to the domains described (two dihydrouridine residues in the D-loop, a thymidine and pseudouridine in the TψC-loop and the anticodon inosine in the ASL domain). **(b)** The crystal structure of the human $tRNA^{Lys3}_{UUU}$ isoacceptor at 3.3 Å resolution [7] represents the canonical L-shaped tertiary structure common to all tRNAs. The domain coloring for **(a)** and **(b)** are the same.

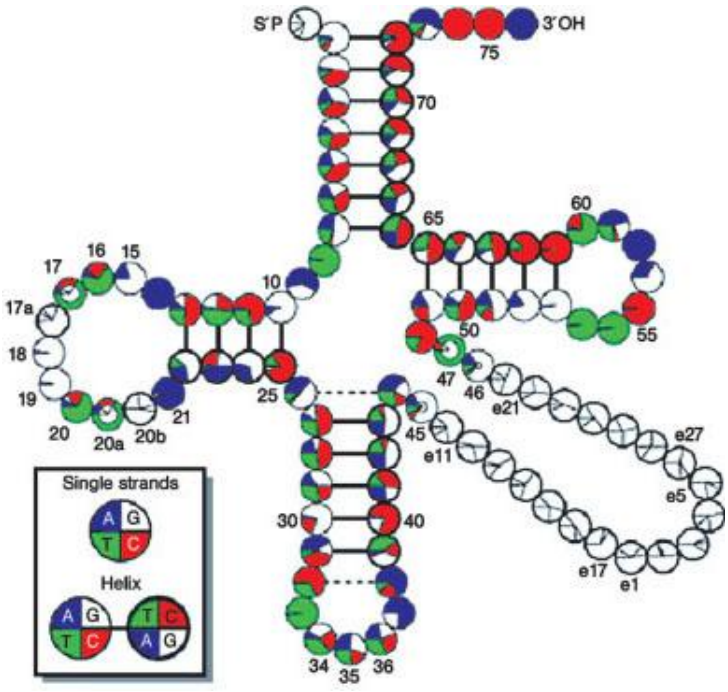

Figure 3. tRNA secondary structure with nucleoside conservation noted. This figure is reproduced from [46].

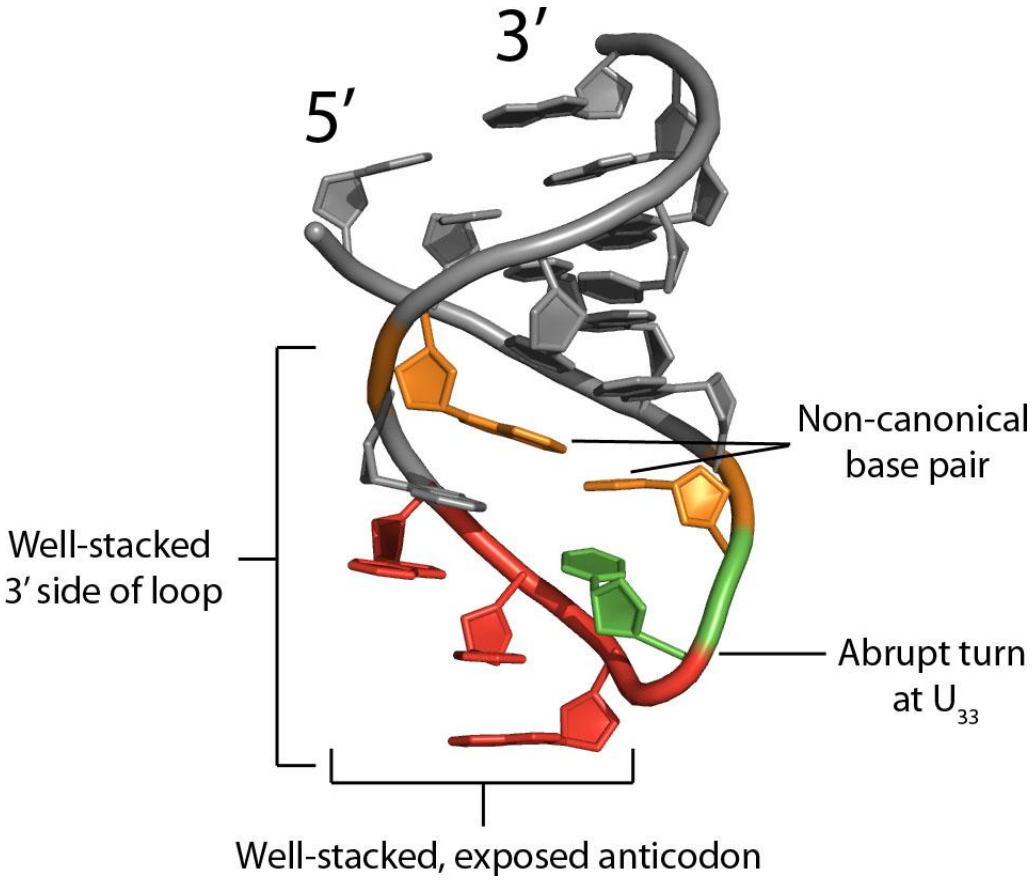

Figure 4. Characteristics of the canonical U-turn anticodon stem and loop conformation. The U-turn is characterized by an abrupt backbone turn at U_{33} (green), a non-canonical closing base pair (orange) that forms a pseudo-five-membered loop, well-stacked and solvent exposed anticodon residues (red) and a stable and energetically favorable base stacking network along the 3' side of the loop. This figure is adapted from PDB entry 1XNR [45].

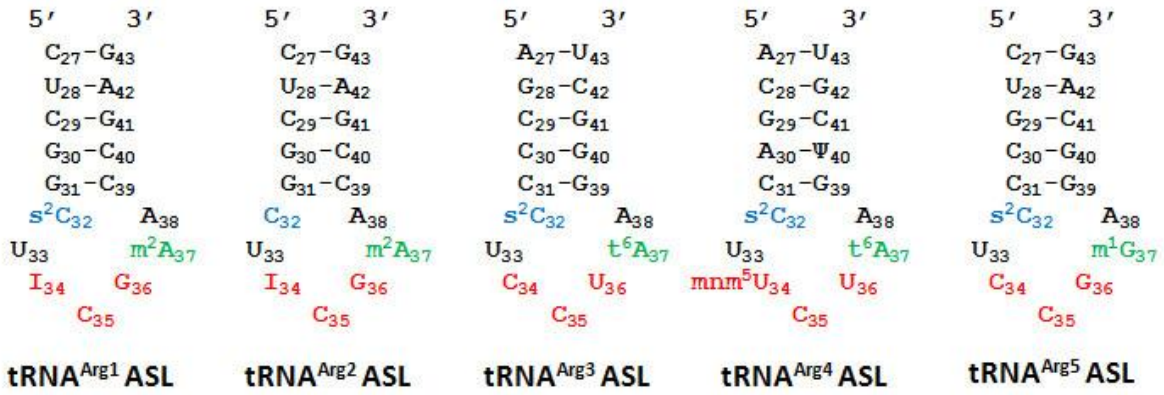

Figure 5. The anticodon domain primary sequences of the five tRNA^{Arg} isoacceptors from *E. coli*. Abbreviated modifications shown are: Ψ = pseudouridine; s^2C = 2-thiocytidine; I = inosine; m^2A = 2-methyladenosine; mnm^5U = 5-methylaminomethyluridine; t^6A = *N*^{6}-threonylcarbamoyladenosine; m^1G = 1-methylguanosine. The anticodon residues (red) and the position 32 (blue) and 37 (green) loop residues are colored for clarity.

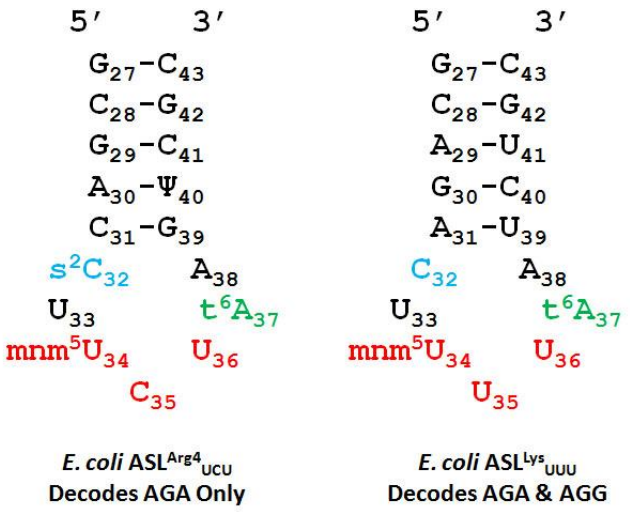

Figure 6. The anticodon domain primary sequences of *E. coli* ASL^{Arg4} and ASL^{Lys}. The wild type ASL$^{Arg4}_{UCU}$ (left) and the tested ASL$^{Lys}_{UUU}$ (right, [28]) contain the same position 34 and 37 anticodon loop modifications. Although these ASLs are very similar, they differ in their ability to discriminate between A and G at the third codon position. The anticodon residues (red) and the position 32 (blue) and 37 (green) loop residues are colored for clarity. Abbreviated modifications shown are: Ψ = pseudouridine; s^2C = 2-thiocytidine; mnm^5U = 5-methylaminomethyluridine; t^6A = N^6-threonylcarbamoyladenosine.

CHAPTER 2. Modifications modulate anticodon loop dynamics and codon recognition of *E. coli* $tRNA^{Arg1,2}$

William A. Cantara, Yann Bilbille, Jia Kim, Rob Kaiser, Grażyna Leszczyńska, Andrzej Malkiewicz and Paul F. Agris

Published in Journal of Molecular Biology, 2012, Vol. 416 (4), p. 579-597.

2.1 Statement of project contribution

The comprehensive detail with which this project was composed hinged on coordinated efforts in three main areas. For this project, I, along with Dr Paul F. Agris, acted in a supervisory role for the entire effort. The first component, NMR, was mainly performed by Dr. Yann Bilbille. My role in this aspect was limited to interpretation of the structural data as well as relating this data to biophysical and functional characteristics of the RNA. I was responsible for biophysical and functional characterization using UV-monitored thermal denaturation and circular dichroism spectroscopies and functional binding assays. These assays were also optimized to account for non-specific binding events and calculate picomoles bound. I also mentored an undergraduate, Jia Kim, to assist with the analysis of the thermal denaturation data as well as to perform native polyacrylamide gel electrophoresis.

Abbreviations: ASL, anticodon stem and loop; CD, circular dichorism spectroscopy; COSY, COrrelation SpectroscopY; HSQC, Heteronuclear Single Quantum Correlation; I_{34}, inosine in position 34; mRNA, messenger RNA; NMR, nuclear magnetic resonance; NOE, Nuclear Overhauser Effect; NOESY, Nuclear Overhauser Enhancement Spectroscopy; ppm, part per million; RMSD, root mean square deviation; s^2C_{32}, 2-thiocytidine in position 32; m^2A_{37}, 2-methyladenosine in position 37; TOCSY, TOtal Correlation Spectroscopy.

Keywords: 2-thiocytidine, inosine, 2-methyladenosine, RNA structure, RNA function

2.2 Abstract

Three of six arginine codons are read by two $tRNA^{Arg}$ isoacceptors in *Escherichia coli*. The anticodon stem and loop ($ASL^{Arg1,2}$) of these isoacceptors differ only in that the position 32 cytidine of $tRNA^{Arg1}$ is posttranscriptionally modified to 2-thiocytidine (s^2C_{32}). The $tRNA^{Arg1,2}$ are also modified at positions 34 (inosine, I_{34}) and 37 (2-methyladenosine, m^2A_{37}). To investigate the roles of modifications in the structure and function, six $ASL^{Arg1,2}$ constructs differing in their array of modifications were analyzed by spectroscopy and codon binding assays. Thermal denaturation and circular dichroism spectroscopy indicated that modifications contribute thermodynamic and base stacking properties, resulting in more order but less stability. NMR-derived structures of the $ASL^{Arg1,2}$ showed that the solution structures of the ASLs were nearly identical. Surprisingly, none possessed the U-turn conformation required for effective codon binding on the ribosome. Yet, all $ASL^{Arg1,2}$ constructs efficiently bound the cognate CGU codon. Three ASLs with I_{34} were able to decode CGC, whereas only the singly modified $ASL^{Arg1,2}_{ICG}$ with I_{34} was able to decode CGA. The dissociation constants for all codon bindings were physiologically relevant (0.4 – 1.4 μM). However, with the introduction of s^2C_{32} or m^2A_{37} to $ASL^{Arg1,2}_{ICG}$ the maximum amount of ASL bound to CGU and CGC was significantly reduced. These results suggest that by allowing loop flexibility the modifications modulate the conformation of the $ASL^{Arg1,2}$ which takes one structure free in solution, and two others when bound to the cognate arginyl-tRNA synthetase, or to codons on the ribosome where modifications reduce or restrict binding to specific codons.

2.3 Introduction

Transfer RNA (tRNA) molecules translate the genetic code by recognizing triplet codons on messenger RNA (mRNA) during protein synthesis. The ribosome-mediated interaction of the mRNA codons with the anticodon of the tRNA results in discrimination of synonymous versus non-synonymous codons. No less than 93 different naturally-occurring

posttranscriptional modifications are found in tRNAs, each with unique hydrophobic or hydrophilic properties [1-3]. These modifications play an important role in the accuracy and efficiency of protein synthesis. Modifications occurring in the anticodon stem and loop (ASL) domain of tRNAs are the most studied and the best understood. The majority of modifications and the most chemically complex occur at position 34 or at position 37, immediately 3' adjacent to the anticodon triplet. These anticodon domain modifications alter conformation and thermal stability [4-10], enhance decoding specificity [10-14], enhance ribosomal binding [4,6,15-17], promote proper translocation [17] and maintain the translational reading frame [18-21]. Anticodon domain modifications have been shown to be essential for pre-structuring the ASL toward a canonical U-turn structure in solution for proper recognition in the ribosomal A-site [2,10,18,22].

Escherichia coli (*E. coli)* has five $tRNA^{Arg}$ isoacceptors, four of which have the modification 2-thiocytidine at position 32, s^2C_{32} (Figure 1a [23]). The only other *E. coli* tRNA having s^2C_{32} is $tRNA^{Ser2}_{GCU}$ [23]. As such, the s^2C_{32} modification is highly interesting. It is one of the rarest of modifications [1], but found within a very common anticodon domain consensus sequence element, $C_{32}U_{33}NNNA_{37}A_{38}$. A C_{32} or s^2C_{32} appears in 56% of all *E. coli* sequences along with A_{38} that appears in 69% of the sequences. The only other known nucleoside thiolations are s^2U_{34} and its derivatives, s^4U_8, and four different 2-methylthio-derivatives of adenosine [1]. In contrast to what is known about these modifications, the contribution of s^2C_{32} to the structure and function of the $tRNA^{Arg}$ isoacceptors and their abilities to decode the six arginine codons has not been explored in detail [24,25].

The unmodified, primary nucleotide sequence of the ASL of the $tRNA^{Arg1}$ isoacceptor and that of $tRNA^{Arg2}$ are identical [26,27]. Therefore, to avoid confusion they are denoted as $ASL^{Arg1,2}$ for the purposes of this study (Fig. 1a). The two isoacceptors are derived from a common gene and only differ at two positions in the entire tRNA sequence. $tRNA^{Arg1}$ has the rare modification s^2C_{32} whereas $tRNA^{Arg2}$ has an unmodified C_{32} (Fig. 1a). In addition, $tRNA^{Arg1}$ lacks A_{20} of the dihydrouridine loop. A_{20} is invariant in the other four isoacceptors

[23,26,27] and has been shown to be an important identity determinant for aminoacylation [28]. Indeed, mutant $tRNA^{Arg2}$ transcripts in which the A_{20} is either substituted or deleted result in a 370-fold decrease in aminoacylation activity suggesting that $tRNA^{Arg1}_{ICG}$ is aminoacylated with a lower efficiency than $tRNA^{Arg2}$ [29]. The anticodon domains of the two isoacceptors also have an inosine modification at the "wobble" position 34 (I_{34}) and a 2-methyladenosine at position 37 (m^2A_{37}), 3' adjacent to the anticodon (Fig. 1b,c). Thus, the $tRNA^{Arg1,2}_{ICG}$ isoacceptors have the anticodon of ICG (Fig. 1a). In agreement with the "wobble" hypothesis [30], the guanosine analog inosine, derived from adenosine and located at position 34 allows $tRNA^{Arg1,2}_{ICG}$ to decode three of the four-fold degenerate codons (CGU, CGC and CGA). In comparison, $tRNA^{Arg5}_{CCG}$ with a cytidine at position 34 reads only the fourth codon CGG.

The functions of s^2C_{32} and m^2A_{37} are not fully understood. Both exclusively occur in anticodon loops containing a pyrimidine at position 35 [23,31]. The s^2C_{32} is found in four of the five *E. coli* $tRNA^{Arg}$ isoacceptor species, $tRNA^{Arg1,3,4,5}$, and $tRNA^{Ser2}_{GCU}$. All s^2C_{32} containing tRNA isoacceptors in *E. coli* have common sequence characteristics. In addition to the s^2C_{32}, they all have a modified purine at position 37 (Fig. 1 [23,31]). Additionally, they all contain an A_{38} at the 3'-terminus of the anticodon loop [23,31], and thus a mismatch base pair between A_{38} and C_{32}/s^2C_{32} can be formed (Fig. 1). The classic bifurcated hydrogen bond observed between O2 (C_{32}) and N6 (A_{38}) in tRNA [32] has been shown to be stable over the nanosecond time scales of molecular dynamics simulations [33,34]. All contain a conserved C_{35} at the second position of the anticodon [31] and the invariant U_{33} responsible for the ubiquitous U-turn in tRNA anticodons. C_{35} has been shown to be an important identity element for aminoacylation [27,28].

Although a mutant without s^2C_{32} exhibited wild-type growth, the A-site selection rate for the $tRNA^{Arg3}$ in response to the AGG codon was dependent on this thiolated nucleoside in *Salmonella enteric* [35]. The sulfur atom at position 2 of pyrimidines is known to stabilize stacking interactions by its greater polarizability than oxygen and to promote the *C3'-endo, anti*-conformation of the nucleoside in RNA [36]. Since the bases of nucleosides 31, 32 and

33 should be stacked in the classic U-turn conformation, one wonders how an increase in stacking induced by the 2-thiocytidine may influence the base pairing of C_{32} with A_{38}. The structure of the *E. coli* $tRNA^{Arg1,2}_{ICG}$-s^2C_{32} free in solution may be influenced by the s^2C_{32} modification because a change from C_{32} to s^2C_{32} results in an altered accessibility of the anticodon loop to nuclease S1 to a pattern more indicative of an initiator than an elongator tRNA [37]. The s^2C_{32} modified nucleoside of the *E. coli* $tRNA^{Arg1,2}_{ICG}$-s^2C_{32},m^2A_{37} maintains the translation reading frame, preventing formation of an incorrect protein, the product of a -1 frameshift, during *in vitro* synthesis of the T4 phage coat protein [37]. In this case, the s^2C_{32} modification promotes the correct reading of the mRNA. In fact, chemical alteration of the s^2C_{32} to C_{32} results in a decrease in the ability of the tRNA to maintain the reading frame [37]. However, in single point binding experiments using the fully modified $tRNA^{Arg1,2}_{ICG}$-s^2C_{32},m^2A_{37}, there does not seem to be a significant change in the ribosome binding characteristics in comparison to a tRNA without the s^2C_{32} modification [38]. In fact, it has been suggested that the s^2C_{32} modification could be important to improve accuracy by reducing the speed of translation [35].

We have investigated the importance of all three modifications in $tRNA^{Arg1,2}_{ICG}$ by biochemical, thermodynamic and structural techniques using six differentially modified, heptadecamer ASLs corresponding to the unmodified $ASL^{Arg1,2}_{ACG}$, singly modified $ASL^{Arg1,2}_{ACG}$-s^2C_{32}, $ASL^{Arg1,2}_{ACG}$-m^2A_{37} and $ASL^{Arg1,2}_{ICG}$, and the doubly modified $ASL^{Arg1,2}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}_{ICG}$-m^2A_{37} (Fig. 1b). Structures of four of the ASLs, $ASL^{Arg1,2}_{ACG}$, $ASL^{Arg1,2}_{ACG}$-s^2C_{32}, $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, have been solved by NMR spectroscopy and restrained molecular dynamics. Combinations of anticodon domain modifications modulate the structural flexibility of the anticodon loop and responsiveness to the three degenerate codons. The ASLs lack the classical U-turn and present a singular conformation in solution that is incongruous with the canonical anticodon structure thought to be required for ribosome-mediated codon binding. The results of this study indicate a possible regulatory role for these modifications in decoding of the genome. By modulating the structural dynamics of the loop nucleosides, the modifications alter the codon reading

capacity of the tRNA to a preference for the more commonly used CGU and CGC codons over the rarer CGA codon.

2.4 Results

2.4.1 Modifications remodel thermodynamic properties and conformational dynamics

Constructs of the anticodon stem and loop domain of *E. coli* $tRNA^{Arg1,2}$ ($ASL^{Arg1,2}$) were chemically synthesized with site-selected introduction of naturally-occurring, post-transcriptional modifications. Standard protection and phosphoramidite chemistries, as well as 5'-silyl-2'-acetoxyethyl orthoester ('ACE') chemistries, were employed for introduction of the modified nucleosides [39]. Six different $ASL^{Arg1,2}$ constructs were synthesized, differing in the state of modification (Fig. 1c). Thermodynamic properties and base stacking contributions of the modifications were assessed using UV-monitored thermal denaturation and renaturation, and circular dichroism, CD (Fig. 2). In general, the addition of modifications reduced thermal stability while modestly adding to structural order through base stacking (Table 1). When compared to the unmodified ASL, neither the introduction of inosine at position 34 (I_{34}) nor 2-methyladenosine at position 37 (m^2A_{37}) alone changed the melt temperature, the temperature at which half of the ASL molecules were fully denatured, T_m (Fig. 2a,b; Table 1). Introduction of I_{34} resulted in a modest increase in the Gibbs free energy, ΔG_{37}. In contrast, a single 2-thiocytidine at position 32 (s^2C_{32}) of the ASL caused the most significant reduction in T_m by 4.8 °C and a correspondingly large increase in ΔG_{37} (Fig. 2a; Table 1). The addition of s^2C_{32} to the inosine-modified $ASL^{Arg1,2}{}_{ICG}$ resulted in a less drastic, but significant, destabilization compared to that of s^2C_{32} alone (ΔT_m = -3.7 °C). Though the introduction of m^2A_{37} did not alter the thermodynamic properties of an otherwise unmodified $ASL^{Arg1,2}{}_{ACG}$, m^2A_{37} produced a modest destabilization of the $ASL^{Arg1,2}{}_{ICG}$ relative to the thermodynamic properties of the unmodified $ASL^{Arg1,2}{}_{ACG}$ and $ASL^{Arg1,2}{}_{ICG}$.

Differences in hyperchromicity monitored during thermal denaturations, and differences in ellipticity of circular dichroism (CD) spectra are indicative of changes in base stacking that occur with the introduction of a modified nucleoside [18,22]. However, hyperchromicity measurements of an ASL composed of a five base paired stem and a seven nucleoside loop are monitored at ~260 nm and influenced mostly by the ten nucleosides of the RNA's base paired stem. In comparison, ellipticity measurements, conducted over a large range of wavelengths and in the native state of the RNA when observed at ambient temperatures, reflect the nucleosides of the seven membered loop as well as the ten nucleosides of the stem. Thus, the two measures of base stacking do not always agree [4]. The CD spectra of the ASLs exhibited the positive ellipticity ($\Delta\varepsilon$) that occurs between 250-290 nm ($\lambda_{max} \approx 268$ nm) and is indicative of A-form RNA (Fig. 2b). The two constructs containing m^2A_{37} also exhibited the minor positive ellipticity (290-320 nm) characteristic of this modification.[40,41]. The signature ellipticity of m^2A_{37} was more pronounced in the doubly modified $ASL^{Arg1,2}{}_{ICG}$-m^2A_{37}. The ASLs containing s^2C_{32} did not exhibit the positive ellipticity characteristic of the thio group of s^2U at 310 nm (Fig. 2 [42]). This ellipticity was not observed in spectra of the mononucleoside s^2C under the same conditions employed for observation of the ASLs (data not shown). Thus, lack of the 2-thio ellipticity at 310 nm for s^2C appears to be a difference in the CD properties of the two pyrimidine ribonuclesides, rather than a conformational difference caused by the modification.

Introduction of the individual nucleosides I_{34} and m^2A_{37} into the $ASL^{Arg1,2}{}_{ACG}$ altered the base stacking as evidenced by a reduction in hyperchromicity during thermal denaturations (Table 1). The addition of s^2C_{32} to the $ASL^{Arg1,2}{}_{ACG}$ did not appreciably change the hyperchromicity (Fig. 2a; Table 1). The hyperchromicity of both doubly modified constructs $ASL^{Arg1,2}{}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}{}_{ICG}$-m^2A_{37} was reduced by approximately 4% compared to that of the unmodified $ASL^{Arg1,2}{}_{ACG}$ and could be attributed to the introduction of I_{34} on the 3'-side of the anticodon loop (Table 1). However, the CD ellipticity of the $ASL^{Arg1,2}{}_{ICG}$-m^2A_{37} increased relative to that of the unmodified $ASL^{Arg1,2}{}_{ACG}$ indicating that the m^2A_{37} modification ordered the 3'-side of the loop (Fig. 2b(ii)). The increase in order on the 3'-side

of the anticodon is indicative of an enhanced base stacking by m^2A_{37} and a concomitant reduction in the normally large amount of conformational space in the loop. This increase in ellipticity could also be attributed to the altered chemical properties inherent with the methylated residue; however, studies have shown that the methyl group of a 2'-deoxy-5-methylcytidine does not impart any intrinsic CD character [43]. Our observations of changes in thermal stability, hyperchromicity and CD ellipiticity for the variously modified $ASL^{Arg1,2}$ are similar to effects reported for modifications that occur on the 5'- vs. 3'-sides of anticodon loops [44]. Modifications of nucleosides on the 5'-side of the anticodon loop such as those at wobble position 34 tend to disorder and destabilize the structure. Modifications on the 3'-side of the loop at positions 37, 38 and 39 tend to enhance the 5'-3' base stacking found in tRNAs and ASLs, though these very same modifications negate intra-loop hydrogen bonding [5,44,45]. In summary, the introduction of anticodon domain modifications on the 5'-side of the loop, s^2C_{32} and I_{34} reduced the thermal stability of the $ASL^{Arg1,2}{}_{ACG}$, and modification on the 3'-side of the loop, m^2A_{37}, increased order through base stacking. Thus, the expected result from modification would be an anticodon domain with a canonical U-turn and the Watson-Crick face of the anticodon nucleosides exposed for codon binding.

2.4.2 Determination of $ASL^{Arg1,2}$ structures

The results of the thermal stability and base stacking studies suggested that the modified nucleoside contributions to the anticodon loop structure and conformational dynamics of $ASL^{Arg1,2}$ were qualitatively similar to that of other tRNAs. In order to determine if the modifications of $ASL^{Arg1,2}$ were driving the anticodon domain structure to that of the canonical conformation, four of the six differentially modified ASLs were studied by NMR and full, high resolution structures were calculated for $ASL^{Arg1,2}{}_{ACG}$, $ASL^{Arg1,2}{}_{ACG}$-s^2C_{32}, $ASL^{Arg1,2}{}_{ICG}$ and $ASL^{Arg1,2}{}_{ICG}$-s^2C_{32}. A considerable number of spectra were collected for $ASL^{Arg1,2}{}_{ICG}$-m^2A_{37}, but insufficient material was available for a full structure determination. All six differently modified $ASL^{Arg1,2}{}_{ACG}$ constructs were subjected to non-denaturing polyacrylamide gel electrophoresis, PAGE (Supplemental Information, Fig. S1). The results demonstrated that the ASLs were monomolecular species of potentially single conformation

at the concentrations tested. Surprisingly, there were no noticeable differences in the rate of migration due to the presence or identity of the modifications. All six ASLs were found to migrate within ±1% of the mobility of unmodified $ASL^{Arg1,2}_{ACG}$. The $ASL^{Arg1,2}$ constructs migrated faster than the standard ASL, the unmodified yeast ASL^{Phe}, even though their molecular weights were all ~100 Da larger, perhaps suggesting a more compact structure in solution than a typical ASL with a U-turn. The results of pulsed field-gradient spin-echo NMR diffusion experiments also indicated that the ASLs were monomers in solution even at the higher concentrations required for determination of their structures by NMR (Supplemental Information, Table S1). Here we have focused our analysis on the structures of $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32} because of the interest in understanding the effect of the rare modification s^2C_{32} on structure-function relationships.

NMR resonances assignments. Exchangeable ^{1}H NMR resonances were identified and assigned as accomplished previously for unlabeled, unmodified and modified ASLs [5,9,45-47]. The 1D spectrum recorded in water exhibited three sharp resonances in the imino region which were assigned to the imino resonances of the three guanines of the stem, G_{41}, G_{31} and G_{30}, all involved in canonical G•C base pairing. Two other resonances, broader than the previous signals, were assigned to G_{43}H1 and U_{28}H3. A very broad exchangeable resonance was observed near 11 ppm, but was not assigned. No exchangeable proton resonances were observed near 15.5 ppm, characteristic of a protonated cytosine. Non-exchangeable resonances were assigned using well established techniques [48]. The sequential pathway of nuclear Overhauser enhancement, NOE, from the aromatic protons to the H1' of the 3' nucleoside could be followed from C_{27} to C_{35} and again from G_{36} to G_{43} with a break appearing between C_{35} and G_{36} (Fig. 3a). Though there was a break in the conventional NOE walk of A-form RNA, the G_{36}H8 had a very weak NOE with C_{35}H2'. This is indicative of C_{35} not being stacked with G_{36}. The C_{35}/G_{36} stack would be observed in the classical U-turn, and thus the canonical U-turn would not be expected to be present. The chemical shifts of H5, H6 and H8 are affected by the 5'-neighboring base when in an A-form helix.[49] The C_{35}H5 chemical shift was not in agreement with a pyrimidine experiencing a 5'-purine ring

current [49], and thus was not supportive of the stacking of C_{35} with A_{34}/I_{34} that would be observed in a classical A-RNA or a U-turn.

Over 100 inter-nucleotide, sequential and non-sequential NOEs were observed for each of the $ASL^{Arg1,2}{}_{ICG}$, approximately half of these were for the anticodon loop and the adjacent nucleosides from 31 to 39 (Table 2; Fig. 4). On the 5' side of the loop, the s^2C_{32}/C_{32} was stacked with G_{31} and U_{33}, as shown by the inter-aromatic NOEs between the G_{31} and the unmodified or modified C_{32} bases, and the latter with U_{33} ($G_{31}H8$ - $s^2C_{32}/C_{32}H5$ and $G_{31}H8$ - $s^2C_{32}/C_{32}H6$; $s^2C_{32}/C_{32}H6$ – $U_{33}H6$ and $s^2C_{32}/C_{32}H6$ – $U_{33}H5$). The s^2C_{32} H1' downfield chemical shift (~6.1 ppm) was characteristic of the 2-thio-modification, as exemplified by that of s^2U [50]. The $I_{34}H2$ was assigned by its C2 specific chemical shift, upfield from the main C2 resonances. Even in the 400 ms NOESY spectrum, the $I_{34}H2$ resonance did not exhibit an NOE with another aromatic or sugar proton resonance. Nevertheless, a weak NOE was observed between the $U_{33}H6$ and the $I_{34}H8$ in the long mixing time NOESY experiments. An inter-aromatic NOE was observed between the $I_{34}H8$ and $C_{35}H6$. However, the NOEs expected of U_{33} and I_{34}/A_{34} that would be attributable to a U-turn were not observed.

In the middle of the anticodon loop, the ribose resonances of $C_{35}H2'$, H4', H5' and H5'' were strongly shifted upfield from the other ribose resonances. The $C_{35}H4'$ and H5' resonances overlapped, but their assignments were confirmed by the natural abundance ^1H-^{13}C HSQC (Fig. 3b). It is interesting to note that in $ASL^{Arg1,2}{}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}{}_{ACG}$-s^2C_{32}, the $C_{35}H4'$, H5' and H5'' upfield resonances were shifted slightly less than that of $ASL^{Arg1,2}{}_{ICG}$ and $ASL^{Arg1,2}{}_{ACG}$ (0.1 to 0.2 ppm). Identical upfield chemical shifts have been observed previously [46,51] and are characteristic of a sugar directly located over a base, thus experiencing a base ring current. A strong connectivity was observed between the $C_{35}H5$ and $U_{33}H2'$, still observable at 50 ms (Fig. 3c), as well as a medium to weak NOE connectivity between the $C_{35}H6$ and $U_{33}H2'$. The $C_{35}H5$ proton exhibited strong dipolar connectivities to the A_{34}/I_{34} H3', H5' and H5'' protons. At 400 ms, a very weak NOE could be observed between the $C_{35}H6$ and $U_{33}H5$ of the $ASL^{Arg1,2}{}_{ICG}$-s^2C_{32}.

On the 3'-side of the anticodon loop, the G_{36} exhibited a strong intra-nucleotide H8 - H1' NOE observed in the 50 ms NOE spectrum (Fig. 3c). Sequential G_{36}H1' - A_{37}H8 and G_{36}H8 - A_{37}H8 NOEs connectivities were observed in the 400 ms NOESY spectrum. More interestingly, non-sequential NOEs connectivities were observed between the G_{36}H8 and A_{38}H8, as well as between the G_{36}H1' and A_{38}H8 and between the G_{36}H8 and A_{38}H1' (Fig. 3a). G_{36}H8 exhibited NOEs with C_{35}H5'' and either or both of the C_{35}H4', H5' degenerated resonances. A protonated A_{38} facilitates and strengthens the potential for C_{32}-A_{38} hydrogen bonding. The A_{38}C2 connectivity was not upfield in comparison to other C2 resonances, indicating that A_{38} was not protonated under our conditions [52]. The NOE connectivity observed between A_{38}H8 and A_{37}H1' was relatively strong and still observable in the 50 ms NOESY spectrum. However, the A_{37}H2, assigned with the help of the natural abundance ^{1}H-^{13}C HSQC spectrum, exhibited no NOEs with other aromatics and ribose resonances, hinting at this base's orientation in regard to other nucleosides within the structure. The $ASL^{Arg1,2}_{ACG}$-s^2C_{32} spectrum exhibited two very weak NOEs between A_{37}H2 and A_{38}H1' and between A_{37}H2 and A_{37}H1', but these NOEs were only observable at a 400 ms mixing time. Finally, A_{38} and C_{39} were stacked as demonstrated by the C_{39}H5 - A_{38}H8 and C_{39}H6 - A_{38}H8 NOE connectivities. The A_{38}H2 exhibited a strong NOE connectivity with C_{39}H1', as well as a medium inter-loop NOE connectivity with U_{33}H1'.

Sugar conformations and backbone geometry. The COSY, DQF-COSY and TOCSY spectra indicated that all the nucleosides of the stem (C_{27} - G_{31}, C_{39} - G_{43}), except G_{43}, adopted the *C3'-endo* sugar pucker ($^{3}J_{H1'-H2'} < 2$Hz) characteristic of an A-form RNA helix. Among the loop nucleosides, the s^2C_{32}/C_{32} and A_{38} took the *C3'-endo* conformation. U_{33} and G_{36} exhibited a mixture of *C2'-endo/C3'-endo* conformation and I_{34}/A_{34}, C_{35} and A_{37} adopted the *C2'-endo* conformation, as suggested by their $^{3}J_{H1'-H2'} > 5$Hz. The G_{36} sugar pucker was *C3'-endo.* The G_{36} has a χ dihedral angle of 91.3 ± 0.6° and 91.8 ± 7.4° for $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, respectively, which corresponded to the transition between the *+sc* or *+ac* conformation. The distances between G_{36}H8 and G_{36}H1' were 2.60 ± 0.03 Å and 2.58 ± 0.04 Å for $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, respectively. The ^{31}P phosphorous spectrum

showed that all the resonances are grouped together within ±0.5 ppm except for two resonances belonging to the ^{31}P resonances of A_{38} and the A_{34}/I_{34} which were downfield from the center of the main ^{31}P resonances by ~0.8 and ~0.4 ppm, respectively. All the other resonances, including the G_{36} P, C_{35} P and U_{33} P, remain in the main cluster, suggesting that the α and ζ angles are typical of A-RNA form.

2.4.3 Structure of the tRNA^{Arg1,2} ASLs

The structures of the four differently modified ASLs, $ASL^{Arg1,2}_{ACG}$, $ASL^{Arg1,2}_{ACG}$-s^2C_{32}, $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, were surprisingly similar to each other, and not reflective of the conformation that would be required of their codon binding on the ribosome. The structures were determined by restrained molecular dynamics from the NMR data and were calculated using NOE restraints based on 200 and 250 ms mixing times. The ten selected lowest-energy structures for $ASL^{Arg1,2}_{ICG}$ and for $ASL^{Arg1,2}_{ICG}$-s^2C_{32} were superimposed (Fig. 5). Relative to the mean structure, the average structures for $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32} have an overall heavy atom RMSD of 1.2 ± 0.4 Å and 1.6 ± 0.6 Å, respectively (Table 2). The stems formed the classical A-RNA helix. In the determination of hairpin RNA structures by NMR, loop nucleosides are usually less restrained than that of the stem due to a lack of NOEs. However, the RMSD of the loop nucleosides s^2C_{32}/C_{32} to A_{38} are 0.6 ± 0.2 Å and 1.3 ± 0.6 Å for $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, respectively. The somewhat higher RMSD values and errors observed for $ASL^{Arg1,2}_{ICG}$-s^2C_{32} were probably due to the quality of the NMR spectra for the doubly modified ASL, resulting in a lower and less accurate number of NOEs distance restraints that could be used for the structure calculation. Nevertheless, we cannot rule out the possibility that the s^2C_{32} modification induced local dynamics and thus, would correlate to results of the thermodynamic studies. The anticodon loops of $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32} (s^2C_{32}/C_{32}-A_{38}) exhibited the same global folding with minor differences that could be attributed to differences in RMSD (Fig. 6a). Interestingly, G_{36} was stacked nicely on A_{38} consistent with the NOEs observed between these two nucleosides. The position of G_{36} was stabilized by one hydrogen bond between G_{36}N2 and A_{37}OP1 with an average acceptor donor distance of 3.29 ± 0.26 Å and 3.50 ± 0.38

Å for $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}. A_{37} did not participate in what is the canonical 3'-stack of anticodon loop nucleosides. A_{37} was flipped out and directly in contact with the solvent; the position of A_{37}H2 was consistent with its lack of NOE connectivities (Fig. 6a). An intra-nucleoside hydrogen bond could be detected between A_{37}O2' and A_{37}N3 (Fig. 6b). The average distance between the acceptor and donor atoms was 3.11 ± 0.10 Å for $ASL^{Arg1,2}_{ICG}$ and 3.09 ± 0.11 Å for $ASL^{Arg1,2}_{ICG}$-s^2C_{32}.

The sugar of C_{35} was stacked below the base of G_{36}, explaining the unusual upfield chemical shifts observed for the ribose protons of C_{35} (Fig. 6a). Moreover, the C_{35}H5 and H6 were close to I_{34}H5', H5'' and H3', consistent with the NMR data. The base of C_{35} stacked loosely below U_{33}, in agreement with the weak NOE observed between C_{35}H5 and U_{33}H6 for $ASL^{Arg1,2}_{ICG}$-s^2C_{32}. The C_{35} position was stabilized probably by a hydrogen bond between C_{35}N4 and A/I_{34} O2P with an average distance between the acceptor and donor atoms of 3.67 ± 0.66 Å for $ASL^{Arg1,2}_{ICG}$ and 4.00 ± 0.85Å for $ASL^{Arg1,2}_{ICG}$-s^2C_{32} (Fig. 6b). The positions of I_{34}, C_{35} and G_{36} exposed the non-bridging phosphate oxygen atoms from C_{35} at the top of the loop (Fig. 6a). The exposure of this oxygen atom may have important functional consequences, especially in the critical recognition of C_{35} by the arginyl-tRNA-synthetase [53]. The s^2C_{32}/C_{32} was well stacked under U_{33} and over the G_{31} of the stem. The I_{34}, as well as A_{34} in $ASL^{Arg1,2}_{ACG}$ and $ASL^{Arg1,2}_{ACG}$-s^2C_{32}, was bulged out and this position explains the absence of NOEs observed from $I_{34}H_2$. The loop was closed by the non-canonical A_{38}•C_{32} base pair (Fig. 6c). However, hydrogen bonding seen here with the N6 of A_{38} acting as a proton donor for N3 of C_{32} (Fig. 6b,c) differs from the classically bifurcated network typically seen in which the N6 of A_{38} acts as a dual donor for the O2 of C_{32} only [54]. We observed one hydrogen bond between A_{38}N6 and C_{32}/s^2C_{32} N3 with an acceptor donor distance of 3.35 ± 0.43 Å for $ASL^{Arg1,2}_{ICG}$ and 3.73 ± 0.48 Å for $ASL^{Arg1,2}_{ICG}$-s^2C_{32} (Fig. 6b). The intra-loop hydrogen bonds stabilized the loop conformation (Fig. 6d) contributing to the relative inside/outside orientation of each base (Fig. 6d).

The four differently modified $ASL^{Arg1,2}$, $ASL^{Arg1,2}_{ACG}$, and $ASL^{Arg1,2}_{ACG}$-s^2C_{32}, $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, had similar spectral characteristics with only minor

differences, suggesting that they exhibited the same overall conformation. Indeed a superimposition of spectra for the singly modified $ASL^{Arg1,2}_{ICG}$ and the doubly modified $ASL^{Arg1,2}_{ICG}$-m^2A_{37} showed only slight differences which may be caused by local environmental variations associated with the methylated adenosine (Supplemental Information; Fig. S2). Yet, the resulting conformation of the anticodon loop is at odds with the conventional wisdom for anticodon structures required for tRNA's binding to cognate and synonymous codons in the A-site of the ribosome. None of the constructs exhibited the NMR characteristics of the canonical U-turn [47]. tRNA's U-turn functions to facilitate long-range tertiary interactions exposing the Watson-Crick faces of the anticodon bases to the solvent for codon binding.

2.4.4 Modifications modulate binding to non-cognate synonymous codons

The unexpected and unconventional architecture of the anticodon of the $ASL^{Arg1,2}_{ICG}$ constructs brought into question their abilities to bind to their synonymous codons. To investigate the contribution of post-transcriptional modifications to the decoding function of the tRNA, ribosome A-site binding was assessed *in vitro* with both the cognate (CGU) and non-cognate, synonymous codons (CGC, CGA) that $tRNA^{Arg1,2}_{ICG}$ is known to decode. All six $ASL^{Arg1,2}$ constructs bound to the A-site cognate CGU codon with physiologically relevant binding constants (Fig. 7; Table 3). The ASLs with A_{34}, $ASL^{Arg1,2}_{ACG}$, $ASL^{Arg1,2}_{ACG}$-s^2C_{34}, and $ASL^{Arg1,2}_{ACG}$-m^2A_{37} and $ASL^{Arg1,2}_{ICG}$ bound to CGU at approximately the same levels of saturation, 4-5 pmoles (Fig. 7a,b). The uniformity of the dissociation constants and amounts of ASL bound indicated that modifications were not necessary for efficient codon binding to the cognate codon, CGU. As would be expected, the $ASL^{Arg1,2}_{ACG}$, $ASL^{Arg1,2}_{ACG}$-s^2C_{34}, and $ASL^{Arg1,2}_{ACG}$-m^2A_{37} did not bind to the CGC or CGA codons even at levels of ASL that saturated the mRNA-programmed ribosomes (data not shown).

$ASL^{Arg1,2}_{ICG}$ bound CGC and CGA, and $ASL^{Arg1,2}_{ICG}$-s^2C_{34} and $ASL^{Arg1,2}_{ICG}$-m^2A_{37} bound CGC with physiologically relevant dissociation constants (K_ds between 0.4 and 1.4 µM; Table 3). However, the amount of $ASL^{Arg1,2}_{ICG}$ maximally bound to CGU, CGC and CGA

varied with the modification under identical assay conditions (Fig. 7b-d). The maximal amount of $ASL^{Arg1,2}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}_{ICG}$-m^2A_{37} that bound to the cognate codon CGU was ~2 pmoles, ~50 % of that bound to CGU. Further reductions in the amount bound were observed for the binding of $ASL^{Arg1,2}_{ICG}$, $ASL^{Arg1,2}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}_{ICG}$-m^2A_{37} to the non-cognate synonymous codon CGC, 1.0, 0.8 and 0.4 pmoles bound, respectively. Although $ASL^{Arg1,2}_{ICG}$ was the only ASL to bind to CGA, the maximum amount bound was reduced to 1.6 pmol (Fig. 7d). The introduction of either s^2C_{32} or m^2A_{37} completely abrogated binding of the $ASL^{Arg1,2}_{ICG}$ to the non-cognate synonymous codon CGA, as evidenced by their inability to bind to the ribosome above the levels of the control mRNA codon GCG. In observing what appears to be a negative function for the s^2C_{32} and m^2A_{37} modifications of $ASL^{Arg1,2}_{ICG}$, we hypothesized that the modifications may play a role in restricting the anticodon G_{36} from binding erroneously in a wobble pair manner to the codons beginning with U_1. However, none of the six ASLs were able to bind the UGU, UGC or UGA codons showing that G_{36} is unable to decode a uridine at the first position of the codon for any of the ASLs tested (data not shown).

2.5 Discussion

2.5.1 Modifications alter tRNA's decoding of specific codons

Previous studies have shown the importance of position 34 modifications in both expanding [4,46] and restricting [9] wobble discrimination at the third codon base. Indeed, Crick hypothesized correctly that I_{34} would expand the decoding ability of a single tRNA isoacceptor to codons ending in U, C and A [30]. Here, codon-specific ribosomal binding assays confirm this function for I_{34} in *E. coli* $ASL^{Arg1,2}$. The singly modified $ASL^{Arg1,2}_{ICG}$ bound all three synonymous codons with low micromolar binding constants and the unmodified ASL was unable to bind either of the two non-cognate synonymous codons CGC or CGA. Free energy calculations of base pairing with inosine suggest that the purine•purine pair of I_{34}•A3 should be significantly less favorable than I_{34}•U3 or I_{34}•C3 [55]. The binding constants demonstrated that the $ASL^{Arg1,2}_{ICG}$ constructs had no preference among the three

bases, U, C and A, at the third codon position. However, the maximal amount of ASL bound varied. In binding CGU in which the I_{34}•U3 pair is formed, the amount of ASL bound exceeded the binding of the CGC and CGA codons by ~1.7 fold. These results confirmed the physiological relevance of the crystal structures of a singly modified $ASL^{Arg1,2}_{ICG}$ bound to CGC and CGA in the ribosomal A-site [56] and support *in vivo* assays showing the inefficiency of the I_{34}•A3 wobble base pair in recognition of CGA [57].

The $ASL^{Arg1,2}_{ICG}$-s^2C_{34} and the $ASL^{Arg1,2}_{ICG}$-m^2A_{37} were capable of decoding only CGU and CGC. A-site binding of CGA was undetectable for these doubly modified $ASL^{Arg1,2}_{ICG}$. The lack of codon binding by $ASL^{Arg1,2}_{ICG}$-m^2A_{37} was in accord with S1 nuclease cleavage experiments that showed that the ASL exhibits characteristics unlike that of a canonical U-turn-containing isoleucyl-tRNA [37]. The presence of s^2C_{32} caused a significant change in nuclease accessibility to the anticodon domain, albeit in the direction of a cleavage pattern more similar to $tRNA^{Ile}$, but still lacked the signature pattern consistent with a U-turn conformation [37]. The $ASL^{Arg1,2}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}_{ICG}$-m^2A_{37} exhibited a significant reduction in the amount of ASL maximally bound to CGU and CGC in comparison to that of the singly modified $ASL^{Arg1,2}_{ICG}$ construct. Taken together, these results could be interpreted as modifications causing a shift in the conformational equilibrium to a more populated, but unfavorable, structure for codon binding. However, non-denaturing PAGE, NMR structure determination and NMR gradient diffusion results conclusively showed a highly populated singular, monomolecular conformation in solution for all constructs.

In order for the ASL to adopt the necessary geometry for the purine•purine (I_{34}•A3) pair during CGA decoding, the anticodon bases need to have a larger amount of allowed conformational space than is required for binding to CGU or CGC. Also, the anticodon must be able to overcome the elevated energy barrier between the unpaired and paired states [55]. The higher level of loop rigidity indicated by circular dichroism for doubly modified $ASL^{Arg1,2}_{ICG}$-m^2A_{37} would most likely reduce the conformational space of the anticodon, similarly to how a methylated guanosine, in combination with a modified cytidine at position 32 and a modified purine at position 34, in ASL^{Phe} restricts the ASL conformational space

[44]. Chemical synthesis of the fully modified $ASL^{Arg1,2}_{ICG}$-s^2C_{34};m^2A_{37} was problematic, but perhaps if tested, it would be able to recognize CGA in its fully modified state.

2.5.2 $ASL^{Arg1,2}$ constructs form a similar but unconventional solution structure

The anticodon loops of the variously modified $ASL^{Arg1,2}$, had the same overall, but unexpected structural characteristics (Fig. 8a). Chemical shift differences between constructs did not appear significant. A very slight change in base stacking would have been observed as a dramatic change in chemical shift [58]. The anticodon loops of the doubly modified ASLs may be more dynamic than that of the singly modified ASLs. The unmodified and modified $ASL^{Arg1,2}$ constructs bound cognate and synonymous codons in solution with physiologically relevant binding constants. However, the $ASL^{Arg1,2}$ solution structures adopted a $^{5'}$UNCG$^{3'}$ tetraloop conformation [59,60], and not that of the canonical U-turn (Fig. 6b). In a tetraloop conformation, the 3'-terminal G adopts the *syn* conformation with a *C3'-endo* sugar pucker. The tetraloop G is base paired with uridine and the majority of time adopts the classical wobble G•U base pair. The interaction is characterized by a very intense NOE between the GH1 and the UH3 imino protons as observed in the water NOESY spectra (distance ~ 2.5 Å). The N nucleoside of the $^{5'}$UNCG$^{3'}$ motif is pulled out and adopts the *C2'-endo* sugar conformation. Finally, this tetraloop is closed by a Watson-Crick base pair. The $ASL^{Arg1,2}$ constructs have the $^{5'}U_{33}(A/I_{34})C_{35}G_{36}{}^{3'}$ primary sequence of the $^{5'}$UNCG$^{3'}$ motif and are 'closed' by the non-canonical s^2C_{32}/C_{32}•A_{38} base pair (Fig. 7c). The A/I_{34} is pulled outside (Fig. 7d) and adopts the *C2'-endo* sugar conformation usually observed in the $^{5'}$UNCG$^{3'}$ motif [59]. Nevertheless, the strong GH1-UH3 NOE connectivity of the classical $^{5'}$UNCG$^{3'}$ tetraloop was not observed due to distances between the G_{36}H1 and U_{33}H3 being on average 5.65 ± 0.84 Å and 5.60 ± 0.51 Å for $ASL^{Arg1,2}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}_{ICG}$, respectively. In the solution structure of the $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, G_{36} was stacked with A_{38} and A_{37} was turned out emphasizing the potential importance of the m^2A_{37} and other position 37 modified nucleotides on the structure of the anticodon loop. However, NMR spectra of the $ASL^{Arg1,2}_{ICG}$-m^2A_{37} reflected the same unusual structural characteristics as $ASL^{Arg1,2}_{ICG}$-s^2C_{32} and $ASL^{Arg1,2}_{ICG}$ (Supplemental Information, Fig. S2). Thus, the m^2A_{37} modification alone

was unable to change the loop conformation. It is tempting to speculate that both modified nucleotides in position 32 and 37 could play an important role in destabilizing the pseudo $^{5'}$UNCG$^{3'}$ conformation observed and thereby induce formation of the U-turn. The results reported in this work favor such an interpretation because of the unusual destabilization by s^2C_{32}. Indeed, the S1 nuclease digestion experiments of fully modified $tRNA^{Arg1}$ showed a pattern more indicative of a U-turn. However, the accessibility between positions 36 and 37 indicated that this would still not conform to the fully canonical conformation [37]. This nuclease digestion pattern could be indicative of the tRNA's conformational dynamics rather than that of a single dominant structure at equilibrium.

The lack of NMR resonances that were indicative of a conventional U-turn, and the unique presence of the thio-modification of s^2C_{32} in the $tRNA^{Arg}$ isoaccepting species and in $tRNA^{Ser}$ prompted us to investigate the possible contribution of divalent metal ions to the anticodon loop structure. The metal-ion binding properties of the $ASL^{Arg1,2}{}_{ACG}$ and $ASL^{Arg1,2}{}_{ICG}$ were investigated using Mg^{2+} and cobalt hexamine, $[Co(NH_3)_6]^{3+}$, as a model for the biologically relevant $[Mg(H_2O)_6]^{2+}$ ligand (Supplemental Information, Fig. S3 [61-63]). Titration of the RNA with magnesium or cobalt ions broadened most of the ^{1}H-NMR resonances. The ^{31}P spectra indicated that the metal ions induced some conformational changes in the ASL, but not toward the U-turn conformation. The chemical shift dispersion of the ^{31}P resonances narrowed around the A-form RNA (Supplemental Information, Fig. S3), whereas in the U-turn conformation there is a strong downfield signal separated from the other A-form ^{31}P resonances [46,47].

The solution structures of the variously modified $ASL^{Arg1,2}$ can be compared to the crystal structures of the yeast $tRNA^{Arg1,2}{}_{ICG}$ with its cognate arginyl-tRNA synthetase [53], and that of $ASL^{Arg1,2}{}_{ICG}$ on the ribosome bound to A-site codons (Fig. 8b [56]). The yeast $tRNA^{Arg1,2}{}_{ICG}$ has the same anticodon loop sequence as that of the *E. coli* $tRNA^{Arg1,2}$. In the crystal structure in complex with the synthetase, the anticodon loop is completely distorted by the protein's interaction (Fig. 8b [53]). A_{38} is pulled outside of the loop while C_{32} forms a cross-loop stack with A_{37} and thus, there is no possibility for a base pair between A_{38} and C_{32}

[53]. The anticodon's C_{35} is an identity determinant and is recognized by the synthetase through the protein's backbone interactions of histidine residues 22 and 23, and a stacking interaction with tryptophan 569. The structure of the $ASL^{Arg1,2}_{ICG}$ bound to codon in the ribosomal A-site [56] differs significantly from that bound to the synthetase. All the $ASL^{Arg1,2}_{ICG}$ loop nucleosides adopted the *C3'-endo* conformation, and the loop adopted the classical U-turn conformation expected for codon recognition with the stacking of the anticodon $I_{34}C_{35}G_{36}$ (Fig. 8b). Additionally, A_{38} is positioned inside the loop in close enough proximity to C_{32} to allow for a cross loop interaction. It is interesting to note that the geometry of the C_{32}•A_{38} interaction seen in these crystal structures shifts depending on whether CGC or CGA is present in the ribosomal A-site. When bound to CGC, the interaction consists of a single hydrogen bond between C_{32}N3 and A_{38}N6, whereas binding to CGA results in a conformational shift of C_{32} such that a hydrogen bond forms between C_{32}O2 and A_{38}N6 [56]. A very similar, although not as drastic, conformational shift is seen in our solution structures, albeit, not from the presence of a codon, but from the introduction of s^2C_{32} to $ASL^{Arg1,2}_{ICG}$ (Fig. 7c). Indeed, it has been shown that the nature and geometry of the C_{32}•A_{38} base pair can directly alter the affinity for both cognate and wobble codons [64]. When the adenosine is protonated, C_{32}•A^{+}_{38}, even at a pH of 6.5, the ASL is stabilized [52]. The presence of a U_{32} allowed a tRNA to discriminate between different codons, reducing the ability of the tRNA to recognize codons with the third position wobble base [64]. However, mutating U_{32} to a C_{32} restored the wobble decoding ability. In our structures, the introduction of s^2C_{32} to $ASL^{Arg1,2}_{ICG}$ shifted the conformation toward a geometry more similar to the U_{32}•A_{38} pair [65]. This modification modulated change in conformation is in agreement with our binding data and indicative of a role for s^2C_{32} in discriminating wobble base pairing to CGA.

The NMR structure of the free $ASL^{Arg1,2}_{ICG}$ differed considerably from the crystal structures where the ASL of $tRNA^{Arg1,2}_{ICG}$ was bound to the cognate synthetase [53] or to the mRNA codon in the ribosome [56]. The solution structures of the variously modified $ASL^{Arg1,2}_{ICG}$ are more like each other (Fig. 8a) than the disparate crystal structures with the

arginyl-tRNA synthetase and bound to codon on the ribosome (Fig. 8b). Therefore, three significantly different conformations have been observed for the anticodon domain of the $tRNA^{Arg1,2}$ isoacceptors in recognition of the identity determinant of C_{35} by the arginyl-tRNA synthetase, in solution and with various modified nucleosides, and on the ribosome in response to the codons CGC and CGA. Since the ASL adopts a U-turn in the ribosomal A-site and highly distorted conformation when bound to the synthetase, these macromolecules must induce structural changes in the loop requiring a certain degree of conformational space. In this case, the modifications may imbue a certain degree of flexibility or deformability to the loop which would either allow or disallow the adoption of each of the necessary conformations. Structural differences also exist between the solution and ribosome-bound structures of fully modified $hASL^{Lys3}$; however, these differences are small due to the inherent U-turn solution structure induced by the loop modifications (F.A.P. Vendeix and P.F. Agris, personal communication).

2.5.3 Modifications modulate anticodon domain dynamics for most effective codon usage

The Modified Wobble Hypothesis suggests that a key role for anticodon domain modifications is to alter or pre-structure the anticodon loop conformation into a U-turn, permitting proper geometry of the anticodon bases for codon recognition in the ribosomal A-site [66]. Indeed, the combination of modifications at positions 32, 34 and 37 have been shown to pre-form the anticodon loop into a U-turn conformation in many cases [4,45,46,67,68]. While examples of unmodified ASLs that can bind to cognate codons can be found, they all show spectral characteristics of a U-turn conformation [46]. The solution structures of the $ASL^{Arg1,2}$ constructs, coupled with the ribosome binding results signify the first instance of an ASL with a conformation showing no U-turn indicators that has the ability to efficiently bind codons in the ribosomal A-site.

Codon binding assays and structural characterizations of $ASL^{Arg1,2}$ demonstrated that there are clear functional differences between differently modified ASLs though their solution structures are quite similar (Fig. 8a). Surprisingly, structural analysis showed that none of the

three modifications altered the solution structure significantly from the highly stable $^{5'}$UNCG$^{3'}$ motif seen in the unmodified ASL. Also, the singly modified ASL$^{Arg1,2}_{ICG}$ efficiently decodes CGA, whereas ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$ and ASL$^{Arg1,2}_{ICG}$-m^{2}A$_{37}$ cannot. This functional difference cannot be explained by structural differences as these ASLs contain the same chemical modification at position 34 (inosine) and have nearly identical solution structures. Also, the structural results cannot explain the differences in S1 nuclease accessibility seen with the addition of s^{2}C$_{32}$; however, this may be due to different solution conditions (25 mM sodium acetate, pH 4.5, 5 mM $MgCl_2$, 50 mM KCl and 1 mM zinc acetate) used during those studies [37]. There is precedence for structural features that do not explain functional differences in the anticodon domain of tRNA. Biophysical and structural characterization indicate that the methylated guanosine (m^{1}G$_{37}$) in yeast ASL^{Phe} causes a slight destabilization of the stem and coordinates the loop, allowing for a more defined Cm$_{32}$•A$_{38}^+$ cross loop pair, restricting the conformational space of the anticodon [5,37].

Here, we have shown that ASL^{Arg1,2} modifications do not contribute appreciably to the anticodon's conformation at equilibrium, and that the modifications do not alter the equilibrium solution structure toward either of the two conformations seen in the crystal structures. However, the modifications caused significant differences in the dynamics and stability of the molecule. Thermal denaturation of the ASLs showed that s^{2}C$_{32}$ alone caused a considerable reduction in thermal stability as indicated by a low T_m (ΔT_m = -5 °C) and increased free energy of folding ($\Delta\Delta G$ = +1.93 kcal/mol) compared to the unmodified ASL. The s^{2}C$_{32}$ and m^{2}A$_{37}$ decreased base stacking interactions, as determined from a reduction in the ASLs hyperchromicity. The small differences in ellipticity in the CD spectra and yet large reductions in thermal stability could be explained by the proximity of the s^{2}C$_{32}$ loop modification to the stem. As suggested previously, this is a clear example that the structural and dynamic contributions of anticodon modifications cannot be ascertained by summing up the contributions of the individual nucleosides [44]. The differently modified ASL$^{Arg1,2}_{ICG}$, ASL$^{Arg1,2}_{ACG}$-s^{2}C$_{32}$ and ASL$^{Arg1,2}_{ACG}$-m^{2}A$_{37}$ and the doubly modified ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$ and ASL$^{Arg1,2}_{ICG}$-m^{2}A$_{37}$ constructs differed in thermal stability and in their abilities to bind

cognate and synonymous codons. These ASLs may have a large energy barrier to overcome during a shift from the $^{5'}$UNCG$^{3'}$ conformation seen in solution to the U-turn motif needed for codon binding or the odd conformation in binding to the arginyl-tRNA synthetase. A higher energy barrier between functional conformations would be in agreement with the observation that reduced amounts of the $ASL^{Arg1,2}_{ACG}$-s^2C_{32} and $ASL^{Arg1,2}_{ACG}$-m^2A_{37} were found bound to CGU and CGC. This coupled with the unusually high energy barrier for the I_{34}•A3 base pair would sufficiently explain the inability to bind to the CGA codon. Thus, the anticodon domain modifications in $tRNA^{Arg1,2}$ may function by lowering an energy barrier between the solution structure and the two drastically different conformations needed for binding to the synthetase [53] and to the ribosome [56]. The inability of the ASL to efficiently bind to CGA codons may be a side effect of this need to modulate the energy barrier between conformations, causing CGA to be used less often in *E. coli* coding sequences.

Codon usage in *E. coli* and *Saccharomyces cerevisiae* has been directly related to the relative content of individual tRNA species within the cell [56,69]. In the case of *E. coli* $tRNA^{Arg1,2}$, the recognition of the codons CGU, CGC and CGA, differing in usage from between 21.0 for CGC to 3.9 for CGA per 1000 codons [70], is accomplished using a single set of isoacceptors. Interestingly, we found that the ability of the differentially modified $ASL^{Arg1,2}_{ICG}$ to bind these three codons mimicked the codon usage in *E. coli* (Supplemental Information, Table S2). For the two codons, CGU and CGC, that are more frequently used, 20.4 and 21.0 per 1000 codons respectively [70], all three of the inosine-containing ASLs were able to bind the codons efficiently *in vitro*. However, for CGA (3.9 per 1000 codons) only the singly modified $ASL^{Arg1,2}_{ICG}$ bound efficiently. This suggests a role for s^2C_{32} and m^2A_{37} in restricting codon recognition by $ASL^{Arg1,2}_{ICG}$ to CGU and CGC. A modification-dependent mechanism of selective codon binding, as suggested by these results, would allow a single tRNA species to have vastly different codon usage dependent on the proportion of anticodon domain modifications at nucleosides 32, 34 and 37. The modification state of the tRNA acts as the regulator of codon recognition. This would suggest a role for these

modifications in $ASL^{Arg1,2}_{ICG}$ in particular in regulating the expression of the genes which may be necessary for certain environmental, stress or growth conditions. Precedence for tRNA modification as a regulator of decoding in response to cellular stress has been shown for tRNA methylation in yeast [71]. Further studies, including a characterization of fully modified $ASL^{Arg1,2}_{ICG}$, would provide evidence to decipher the mechanisms by which s^2C_{32} and m^2A_{37} negate CGA binding, while allowing an induced U-turn conformation in the ribosomal A-site for the binding of CGU and CGC. Post-transcriptional nucleoside modifications in the anticodon loop of *E. coli* $ASL^{Arg1,2}$ contribute functional significance to the decoding of the cognate and synonymous codons during protein synthesis.

2.6 Materials and methods

2.6.1 Oligonucleotide preparation

The modified nucleoside 2-thiocytidine (s^2C) was synthesized (Trilink Biotechnologies, San Diego, CA) and then derivatized to the 5‘-*O*-[benzhydryloxybis(tri-methylsilyloxy)silyl]-2‘-*O*-[bis(2-acetoxyethoxy)methyl]-2-thiouridine-3‘-(methyl-*N*,*N*-diisopropyl)- phosphoramidite, the ‘ACE’ protected phosphoramidite (Thermo Fisher, Dharmacon Products, Lafayette, CO). The s^2C_{32}-containing heptadecamer constructs were synthesized with or without I_{34} (Thermo Fisher, Dharmacon Products, Lafayette, CO). The chemically synthesized unmodified and inosine containing ASLs also were produced by ACE chemistry.

Chemical synthesis of 2-methyladenosine was accomplished using a previously described procedure [72-74]. Peracetylated guanosine was chlorinated to give 2-amino-6-chloro-9-(2’,3’,5’-tri-*O*-acetylribofuranosyl)purine and then converted to 2-iodoadenosine. Following methylation with trimethylaluminium, in the presence of palladium catalyst gave 2-methyladenosine [75]. The exocyclic amine function of m^2A was protected with a benzoyl group using transient protection methodology [73]. N^6-benzoyl-2-methyladenosine was then

protected with DMTr and TBDMS on the 5' and 2' hydroxyls and phosphitylated to give the fully protected 3'-*O*-phosphoramidite [76]. Modified 2-methyladenosine units were supplied to IDT (formerly RNA-Tec, Belgium) for synthesis of 2-methyladenosine-modified ASLs with and without I_{34}.

After deprotection via standard protocol, RNA was purified by preparative anion-exchange HPLC, desalted and dialyzed extensively against 20 mM sodium phosphate buffer, pH 6.8 and 0.05 mM EDTA using a 3500 Da cutoff membrane. For observations of the exchangeable protons, the samples were resuspended in 300 µL of 90 % H_2O/ 10 % D_2O. For experiments involving the non-exchangeable protons, the samples were exchanged twice with 99.9 % D_2O. The NMR sample concentrations were between 0.8 and 1.2 mM.

2.6.2 Native polyacrylamide gel electrophoresis

Global conformational differences between the variously modified ASLs were determined by native polyacrylamide gel electrophoresis (PAGE). An 18 % (w/v) polyacrylamide gel was prepared in pH 8.3 tris borate (TB) buffer (89 mM Tris base, 89 mM boric acid). RNA samples (30 µM in 12 µL total volume) were in loading buffer (5 % [w/v] glycerol, 0.04 % [w/v] bromophenol blue and 0.04 % [w/v] xylene cyanol FF in TB buffer). All samples were heated to 85 °C for 2 minutes and allowed to cool at room temperature for 30 minutes prior to loading. Electrophoresis was performed at 4 °C using a temperature controlling apparatus (Novex Mini-cell Thermoflow, Invitrogen) followed by subsequent staining in 0.5 µg/mL ethidium bromide. Gel results were recorded with a BioRad Imager running Quantity One software.

2.6.3 Thermal denaturation spectroscopy

Thermal denaturation and renaturation of the ASLs was monitored by UV absorbance at 260 nm using a Varian Cary 3 UV-visible spectrophotometer running Thermal software [6,44]. All samples were analyzed at an optical density of ~0.2 A_{260}/ml (ambient temperature) in 20 mM sodium phosphate buffer at pH 6.8. The temperature was ramped at 1

°C/min from 7 - 92 °C. Absorbance data was collected at a rate of 4 data points per minute. Thermodynamic properties were determined with MeltWin v3.5 software. All experiments were performed simultaneously with a control cell containing buffer only and this data was subtracted as the background prior to analysis. Data from a minimum of five denaturations and renaturations were averaged and the error was calculated as the standard error of the mean. Optical melting experiments were performed at different pH values (pH = 6.0, 6.8 and 7.2) for the unmodified $ASL^{Arg1,2}_{ACG}$ and results were found to be independent of pH within that range (data not shown).

2.6.4 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were collected on a Jasco J600 spectropolarimeter. All samples were adjusted to ~0.2 A_{260}/ml (ambient temperature) in 20 mM sodium phosphate buffer (pH 6.8) prior to CD analysis and analyzed in a 1 cm path length quartz cuvette. Samples were temperature controlled to 5 ± 1 ˚C during data collection. All data were baseline corrected using a control containing buffer only. Data was analyzed by normalizing the spectra to molar circular dichroic absorbance using simultaneously collected absorbance data to calculate the concentrations of each sample ($\Delta\varepsilon = \theta/32980 \bullet C \bullet L \bullet N$ [77]). Spectra shown are the averaged results of three independent experiments each consisting of triplicate data collection.

2.6.5 Ribosomal A-site codon binding assays

Sucrose density gradient centrifugation was used to purify tight-coupled 70S ribosomes from *E. coli* MRE 600 cells [78]. The mRNAs were designed as 27-mer derivatives of T4 gp32 mRNA [79]. Each mRNA sequence was analyzed by UNAFold [80] to ensure a low probability of stable secondary structure. The mRNA oligos were then purchased from Dharmacon (Thermo Scientific), deprotected and purified by dialyzing extensively against 20 mM sodium phosphate buffer (pH 6.8) using a 3500 Da cutoff membrane. The following

three mRNA oligos were used for this study with the adjacent P-site and A-site codons bolded and underlined:

(i) 5' – GAAAAGGAGGUAAAA**AUGCGU**GCACAU – 3'

(ii) 5' – GAAAAGGAGGUAAAA**AUGCGC**GCACAU – 3'

(iii) 5' – GAAAAGGAGGUAAAA**AUGCGA**GCACAU – 3'

The 17-mer ASL constructs were 5'-end ^{32}P-radiolabeled using ^{32}P-γ-ATP (MP Biomedical) and polynucleotide kinase (New England Biolabs). Radiolabeled ASLs were then purified using preparative 15 % PAGE with 7 M urea as a denaturant. Assays were performed using a constant amount of ribosome (250 nM, 5 pmoles per reaction) and increasing concentrations of unlabeled ASL (0, 0.25, 0.5, 1.0, 1.5, 2.5, 5 μM) spiked with a corresponding amount of radiolabeled ASL to a maximum of 2000 CPM at the 5 μM level [46]. All reaction mixtures were prepared in ribosome binding buffer (50 mM HEPES, pH 7; 30 mM KCl; 70 mM NH_4Cl; 1 mM DTT; 100 μM EDTA; 20 mM $MgCl_2$ adjusted to pH 7 with 2 M NaOH). Ribosome mixtures were activated at 42 ˚C for 10 min, cooled to 37 ˚C and programmed with 2.5 μM mRNA for 15 min at 37 ˚C. Mixtures were then incubated for 15 min with $tRNA^{fMet}$ to saturate the P-site and ensure that ASL binding corresponds exclusively to A-site codon binding. ASL was then added to each reaction and incubated at 37 ˚C for 60 minutes, then placed directly on ice for 30 minutes. Ice cold ribosome binding buffer (100 μL) was then added to the 20 μL reaction mixtures and immediately filtered through a nitrocellulose filter with vacuum using a modified Micro-Sample Filtration Manifold (Whatman Schleicher & Schuell Minifold) 96-well dot blot apparatus [81]. The filter was then washed twice with 150 μL ribosome binding buffer in each well, removed from the apparatus and allowed to dry. A small strip of nitrocellulose was spotted with known amounts of ASL mixture to generate a standard curve for calculating pmoles bound from the intensity measurements. Filters were exposed to a phosphor screen and scanned using a Molecular Dynamics phosphor imager (GE Healthcare). The radioactive density of the spots was measured using ImageQuant software (Amersham). Mismatch mRNA (GCG

A-site codon) was used to determine nonspecific binding for each experimental data point. These values were subtracted from the experimental data to generate binding curves. Dissociation constants were then determined using the one-site specific binding equation in Prism v.3 (Graphpad). All results are the compilation of at least two independent experiments performed in triplicate with an internal standard binding assay using ASL^{Phe} with polyU as the mRNA performed in parallel. Ribosome activity and mRNA programming were determined by 5' ^{32}P-labeling the mRNAs and testing the efficiency by filter ribosome binding. All mRNAs showed similar binding indicative of high ribosome activity.

2.6.6 NMR spectroscopy

Solution and temperature conditions used for spectra collection with the RNA samples were those required and widely used for determination of RNA hairpin structures [10,33,46]. NMR spectra were collected on a DMX Bruker 500 MHz instrument equipped with a triple-resonance ^{1}H, ^{13}C, and ^{15}N probe and three-axis pulsed field gradient capabilities and on a Varian INOVA 600 MHz instrument equipped with a cryoprobe. All NMR data were processed using NMRPIPE [82]. Spectra were displayed and analyzed using SPARKY software [83]. NOESY spectra in D_2O were recorded at 10, 25 and 30°C with mixing time of 50, 100, 150, 200, 250, 300 and 400 ms. TOCSY, COSY and DQF-COSY experiments were acquired at 25°C. Natural abundance ^{1}H-^{13}C HSQC and ^{1}H-^{31}P HETCOR experiments were recorded in D_2O at 25ºC. For the study of exchangeable protons, 2D 1H-1H NOESY spectra were recorded at different temperatures (5, 10, 25 and 30 °C) with mixing times of 125, 150 and 250 ms. To determine if, in solution, under NMR conditions, we have an hairpin or a duplex conformation, the translational diffusion constants of the ASLs were determined and compared to a 17mer ASL of the *E. coli* $tRNA^{Val}$, a 24mer RNA and a 56mer double stranded RNA (112mer total). This was accomplished with the pulsed field-gradient spin-echo technique [84,85] performed on the NMR sample. The translational diffusion coefficients of the three different $ASL^{Arg1,2}$ are comparable to the ASL^{Val} and larger than the translational diffusion coefficient for the 24mer and the 56mer double stranded RNAs (Supplemental Information, Table S1). Our results are in good agreement with published

values of the translational diffusion coefficients of nucleic acids with similar sizes [86] and demonstrated that, under the conditions of our study, we have a hairpin conformation and a monomer in solution.

2.6.7 NMR restraints and structure calculation

NMR data collected at 25 and 30 °C were used for structure calculations. The NOE cross-peaks were integrated using the peak fitting Gaussian function and volume integration in SPARKY software. The distance for each cross peak was calculated and normalized to the non-overlapped pyrimidine H5-H6 cross-peaks (2.42 Å). Upper and lower bonds were assigned to ± 20 % of the calculated distances. For the overlapped peak, the cross-peaks were qualitatively classified as strong (3.0, 1.2, 0.0), medium (4.0, 2.0, 0.0), weak (5.0, 3.0, 0.0) or very weak (6.0, 3.0, 0.0 or 7.0, 3.0, 0.0). No internucleotide restraints were used from the H5' and H5'' proton connectivities. For restraints involving imino protons, the intervals (3.0, 1.2, 0.0) and (5.5, 3.5, 0.0) were used. A total of 253 and 231 distance restraints were used for the structure generation of $ASL^{Arg1,2}_{ICG}$ and $ASL^{Arg1,2}_{ICG}$-s^2C_{32}, respectively. Consistent with the NMR data, the five base pair stem formed Watson-Crick base pairs and they were subjected to six and four hydrogen-bond restraints for G•C and U•A base pairing, respectively, with 2.7 - 3.5 Å between the acceptor and donor atoms. Dihedral constraints on the ribose ring and backbone were derived from semi-quantitative measurement of $^3J_{H-H}$ and $^3J_{H-P}$. Unobserved NOEs or "unoes" were used to separate protons for which NOEs were not observed [87].

Sugar pucker conformations were derived from the observation of the COSY and DQF-COSY spectra. Nucleosides with H1'-H2' couplings >5 Hz were constrained to the *C2'-endo* (δ: 160 ± 20°), whereas nucleosides with absent or very weak H1'-H2' connectivity on the COSY spectra were constrained to the *C3'-endo* conformation (δ: 85 ± 20°). Nucleosides with $^3J_{H1'-H2'}$ between 2 and 5 Hz were left unrestrained. An independent confirmation of the sugar pucker conformation was derived from the ^1H-^{13}C HSQC spectra and the C1' resonance shift upfield to the main cluster. The sugar pucker conformation was also

confirmed from the C3' and C4' resonances downfield from the respective main cluster. The γ torsion angle was derived from the $^{3}J_{H4'-H5'}$ and $^{3}J_{H4'-H5''}$ peaks in the COSY and DQF-COSY. For those nucleosides not exhibiting H4'-H5' and H4'-H5'' peaks (J <5 Hz), γ was constrained to the g^{+} conformation (54 ± 20°). For all other nucleosides, γ was left unrestrained.

The β dihedral angles were derived from the observation of $^{3}J_{P-H5'}$ and $^{3}J_{P-H5''}$ couplings on the ^{1}H-^{31}P HETCOR NMR spectrum. When the $^{3}J_{P-H5'}$ and $^{3}J_{P-H5''}$ were clearly absent, β was constrained to the *trans* conformation (178 ± 20°). When $^{3}J_{P-H5'}$ and $^{3}J_{P-H5''}$ could be observed, β was left unrestrained. All H3'-P were clearly identified on the ^{1}H-^{31}P HETCOR (^{3}J > 5Hz), therefore the ε dihedral angle was loosely constrained to avoid the sterically forbidden g^{+} conformation (-120 ± 120°). Dihedral angle restraints for α and ζ were derived from the observation of the ^{31}P chemical shifts. All ^{31}P resonances, except $A_{38}P$ and to a lesser extent A/I_{34}, were in a defined range of chemical shift. Thus, the α and ζ were constrained to avoid the *trans* conformation that result in strongly downfield ^{31}P resonances (0 ± 120°). $A_{38}P$ α and ζ were left unrestrained. Each of the intra-aromatic to H1' NOE connectivities of the different ASLs nucleosides had an intensity significantly less than that of the H5-H6 NOE connectivity showing that all the bases, with the exception of G_{36}, adopted an *anti*-glycosidic torsional angle. Thus, the χ glycosidic dihedral angles were constrained to the *anti*-conformation for all nucleosides (-158 ± 20°), except for G_{36} which was left unrestrained.

Molecular modeling of the $ASL^{Arg1,2}$ anticodon stem and loop structures was achieved using CNS 1.2 [88,89]. One hundred structures were calculated using the standard NMR restrained annealing protocol [90]. The 10 structures of the lowest total energy, without dihedral or distance restraint violations, were chosen for further analysis. The helical parameters were analyzed using X3DNA software [91]. The output structures were visualized with MOLMOL [92] and PYMOL [93].

2.6.8 Accession numbers

The atomic coordinates and structural factors for the ASLs were deposited in the protein data bank (PDB) under the accession numbers 2KRP ($ASL^{Arg1,2}_{ACG}$), 2KRQ ($ASL^{Arg1,2}_{ICG}$), 2KRV ($ASL^{Arg1,2}_{ICG}$-s^2C_{32}) and 2KRW ($ASL^{Arg1,2}_{ACG}$-s^2C_{32}).

2.7 Acknowledgements

The authors would like to express gratitude to William Graham for HPLC purification of ASL construct, as well as Drs. Kun Lu, Jessica Spears and Caren Stark for productive conversations regarding the project and critical reading of the manuscript. This work was supported in part by a National Science Foundation grants to P.F.A. (MCB-0548602; MCB-1101859), and by Poland's National Science Centre to G.L. (1306/B/2011/40).

2.8 References

1. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF: The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 2011, **39**(Database issue):D195-201.

2. Agris PF, Vendeix FAP, Graham WD: tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* 2007, **366**(1):1-13.

3. Agris PF: The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. *Prog Nucleic Acid Res Mol Biol* 1996, **53**:79-129.

4. Lusic H, Gustilo EM, Vendeix FA, Kaiser R, Delaney MO, Graham WD, Moye VA, Cantara WA, Agris PF, Deiters A: Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNAMet. *Nucleic Acids Res* 2008, **36**(20):6548-6557.

5. Stuart JW, Koshlap KM, Guenther R, Agris PF: Naturally-occurring modification restricts the anticodon domain conformational space of $tRNA^{Phe}$. *J Mol Biol* 2003, **334**(5):901-918.

6. Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, Agris PF: Modified nucleoside dependent Watson-Crick and wobble codon binding by $tRNA^{Lys}_{UUU}$ species. *Biochemistry* 2000, **39**(44):13390-13395.

7. Davis DR, Durant PC: Nucleoside modifications affect the structure and stability of the anticodon of tRNA(Lys,3). *Nucleosides Nucleotides* 1999, **18**(6-7):1579-1581.

8. Cabello-Villegas J, Winkler ME, Nikonowicz EP: Solution conformations of unmodified and A(37)N(6)-dimethylallyl modified anticodon stem-loops of Escherichia coli tRNA(Phe). *J Mol Biol* 2002, **319**(5):1015-1034.

9. Cabello-Villegas J, Nikonowicz EP: Solution structure of psi32-modified anticodon stem-loop of Escherichia coli tRNAPhe. *Nucleic Acids Res* 2005, **33**(22):6961-6971.

10. Denmon AP, Wang J, Nikonowicz EP: Conformation effects of base modification on the anticodon stem-loop of Bacillus subtilis tRNA(Tyr). *J Mol Biol* 2011, **412**(2):285-303.

11. Ashraf SS, Sochacka E, Cain R, Guenther R, Malkiewicz A, Agris PF: Single atom modification (O-->S) of tRNA confers ribosome binding. *RNA* 1999, **5**(2):188-194.

12. Li J, Esberg B, Curran JF, Bjork GR: Three modified nucleosides present in the anticodon stem and loop influence the in vivo aa-tRNA selection in a tRNA-dependent manner. *J Mol Biol* 1997, **271**(2):209-221.

13. Chiari Y, Dion K, Colborn J, Parmakelis A, Powell JR: On the possible role of tRNA base modifications in the evolution of codon usage: queuosine and Drosophila. *J Mol Evol* 2010, **70**(4):339-345.

14. Ninio J: Multiple stages in codon-anticodon recognition: double-trigger mechanisms and geometric constraints. *Biochimie* 2006, **88**(8):963-992.

15. Yarian C, Townsend H, Czestkowski W, Sochacka E, Malkiewicz AJ, Guenther R, Miskiewicz A, Agris PF: Accurate translation of the genetic code depends on tRNA modified nucleosides. *J Biol Chem* 2002, **277**(19):16391-16395.

16. Kurata S, Weixlbaumer A, Ohtsuki T, Shimazaki T, Wada T, Kirino Y, Takai K, Watanabe K, Ramakrishnan V, Suzuki T: Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J Biol Chem* 2008, **283**(27):18801-18811.

17. Phelps SS, Malkiewicz A, Agris PF, Joseph S: Modified nucleotides in $tRNA^{Lys}$ and $tRNA^{Val}$ are important for translocation. *J Mol Biol* 2004, **338**(3):439-444.

18. Gustilo EM, Vendeix FA, Agris PF: tRNA's modifications bring order to gene expression. *Curr Opinion in Microbiol* 2008, **11**(2):134-140.

19. Urbonavicius J, Qian Q, Durand JM, Hagervall TG, Björk GR: Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J* 2001, **20**(17):4863-4873.

20. Björk GR, Durand JM, Hagervall TG, Leipuviene R, Lundgren HK, Nilsson K, Chen P, Qian Q, Urbonavicius J: Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett* 1999, **452**(1-2):47-51.

21. Brierley I, Meredith MR, Bloys AJ, Hagervall TG: Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J Mol Biol* 1997, **270**(3):360-373.

22. Agris PF: Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* 2008, **9**(7):629-635.

23. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J: tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 2009, **37**(Database issue):D159-162.

24. Söll D, Cherayil JD, Bock RM: Studies on polynucleotides. LXXV. Specificity of tRNA for codon recognition as studied by the ribosomal binding technique. *J Mol Biol* 1967, **29**(1):97-112.

25. Söll D, RajBhandary UL: Studies on polynucleotides. LXXVI. Specificity of transfer RNA for codon recognition as studied by amino acid incorporation. *J Mol Biol* 1967, **29**(1):113-124.

26. Chakraburtty K: Primary structure of tRNA Arg II of E. coli B. *Nucleic Acids Res* 1975, **2**(10):1787-1792.

27. Murao K, Tanabe T, Ishii F, Namiki M, Nishimura S: Primary sequence of arginine transfer RNA from Escherichia coli. *Biochem Biophys Res Commun* 1972, **47**(6):1332-1337.

28. Sissler M, Giege R, Florentz C: Arginine aminoacylation identity is context-dependent and ensured by alternate recognition sets in the anticodon loop of accepting tRNA transcripts. *EMBO J* 1996, **15**(18):5069-5076.

29. Tamura K, Himeno H, Asahara H, Hasegawa T, Shimizu M: *In vitro* study of *E.coli* $tRNA^{Arg}$ and $tRNA^{Lys}$ identity elements. *Nucleic Acids Res* 1992, **20**(9):2335-2339.

30. Crick FH: Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* 1966, **19**(2):548-555.

31. Sprinzl M, Vassilenko KS: Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* 2005, **33**(Database issue):D139-140.

32. Shi H, Moore PB: The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* 2000, **6**(8):1091-1105.

33. Durant PC, Bajji AC, Sundaram M, Kumar RK, Davis DR: Structural effects of hypermodified nucleosides in the *Escherichia coli* and human $tRNA^{Lys}$ anticodon loop: The effect of nucleosides s^2U, mcm^5U, mcm^5s^2U, mnm^5s^2U, t^6A, and ms^2t^6A *Biochemistry* 2005, **44**(22):8078-8089.

34. Auffinger P, Louise-May S, Westhof E: Molecular dynamics simulations of solvated yeast tRNA(Asp). *Biophys J* 1999, **76**(1 Pt 1):50-64.

35. Jager G, Leipuviene R, Pollard MG, Qian Q, Bjork GR: The conserved Cys-X1-X2-Cys motif present in the TtcA protein is required for the thiolation of cytidine in position 32 of tRNA from Salmonella enterica serovar Typhimurium. *J Bacteriol* 2004, **186**(3):750-757.

36. Bilbille Y, Vendeix FAP, Guenther R, Malkiewicz A, Ariza X, Vilarrasa J, Agris PF: The structure of the human $tRNA^{Lys3}$ anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs. *Nucleic Acids Res* 2009, **37**(10):3342-3353.

37. Baumann U, Fischer W, Sprinzl M: Analysis of modification-dependent structural alterations in the anticodon loop of Escherichia coli tRNAArg and their effects on the translation of MS2 RNA. *Eur J Biochem* 1985, **152**(3):645-649.

38. Kruse TA, Clark BF: The effect of specific structural modification on the biological activity of E. coli arginine tRNA. *Nucleic Acids Res* 1978, **5**(3):879-892.

39. Hartsel SA, Kitchen DE, Scaringe SA, Marshall WS: RNA oligonucleotide synthesis via 5'-silyl-2'-orthoester chemistry. *Methods Mol Biol* 2005, **288**:33-50.

40. Hattori M, Ikehara M, Miles HT: Poly(2-methyl-N6-methyladenylic acid): synthesis, properties, and interaction with poly(uridylic acid). *Biochemistry* 1974, **13**(13):2754-2761.

41. Ikehara M, Hattori M, Fukui T: Synthesis and properties of poly(2-methyladenylic acid). Formation of a poly(A)-poly(U) complex with Hoogsteen-type hydrogen binding. *Eur J Biochem* 1972, **31**(2):329-334.

42. Watanabe K, Oshima T, Nishimura S: CD spectra of 5-methyl-2-thiouridine in tRNA-Met-f from an extreme thermophile. *Nucleic Acids Res* 1976, **3**(7):1703-1713.

43. Guenther RH, Hardin CC, Sierzputowska-Gracz H, Dao V, Agris PF: A magnesium-induced conformational transition in the loop of a DNA analog of the yeast tRNA(Phe) anticodon is dependent on RNA-like modifications of the bases of the stem. *Biochemistry* 1992, **31**(45):11004-11011.

44. Ashraf SS, Guenther RH, Ansari G, Malkiewicz A, Sochacka E, Agris PF: Role of modified nucleosides of yeast tRNA(Phe) in ribosomal binding. *Cell Biochem Biophys* 2000, **33**(3):241-252.

45. Stuart JW, Gdaniec Z, Guenther R, Marszalek M, Sochacka E, Malkiewicz A, Agris PF: Functional anticodon architecture of human $tRNA^{Lys3}$ includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t^6A. *Biochemistry* 2000, **39**(44):13396-13404.

46. Vendeix FAP, Dziergowska A, Gustilo EM, Graham WD, Sproat B, Malkiewicz A, Agris PF: Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry* 2008, **47**(23):6117-6129.

47. Schweisguth DC, Moore PB: On the conformation of the anticodon loops of initiator and elongator methionine tRNAs. *J Mol Biol* 1997, **267**(3):505-519.

48. Cromsigt J, van Buuren B, Schleucher J, Wijmenga S: Resonance assignment and structure determination for RNA. *Methods Enzymol* 2001, **338**:371-399.

49. Cromsigt JA, Hilbers CW, Wijmenga SS: Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation. *J Biomol NMR* 2001, **21**(1):11-29.

50. Sierzputowska-Gracz H, Sochacka E, Malkiewicz A, Kuo K, Gehrke CW, Agris PF: Chemistry and structure of modified uridines in the anticodon, wobble position of transfer RNA are determined by thiolation. *J Am Chem Soc* 1987, **109**:7171-7177.

51. Puglisi JD, Wyatt JR, Tinoco I, Jr.: Solution conformation of an RNA hairpin loop. *Biochemistry* 1990, **29**(17):4215-4226.

52. Durant PC, Davis DR: Stabilization of the anticodon stem-loop of $tRNA^{Lys,3}$ by an A+-C base-pair and by pseudouridine. *J Mol Biol* 1999, **285**(1):115-131.

53. Delagoutte B, Moras D, Cavarelli J: tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J* 2000, **19**(21):5599-5610.

54. Auffinger P, Westhof E: Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J Mol Biol* 1999, **292**(3):467-483.

55. Vendeix FAP, Munoz AM, Agris PF: Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA* 2009, **15**(12):2278-2287.

56. Murphy FVt, Ramakrishnan V: Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 2004, **11**(12):1251-1252.

57. Curran JF: Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res* 1995, **23**(4):683-688.

58. Giessner-Prettre C, Pullman B, Borer PN, Kan LS, Ts'o PO: Ring-current effects in the Nmr of nucleic acids: a graphical approach. *Biopolymers* 1976, **15**(11):2277-2286.

59. Cheong C, Varani G, Tinoco I, Jr.: Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* 1990, **346**(6285):680-682.

60. Moore PB: Structural motifs in RNA. *Annu Rev Biochem* 1999, **68**:287-300.

61. Kieft JS, Tinoco I, Jr.: Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine. *Structure* 1997, **5**(5):713-721.

62. Colmenarejo G, Tinoco I, Jr.: Structure and thermodynamics of metal binding in the P5 helix of a group I intron ribozyme. *J Mol Biol* 1999, **290**(1):119-135.

63. Gonzalez RL, Jr., Tinoco I, Jr.: Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *J Mol Biol* 1999, **289**(5):1267-1282.

64. Olejniczak M, Uhlenbeck OC: tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 2006, **88**(8):943-950.

65. Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, Ramakrishnan V: Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 2001, **292**(5518):897-902.

66. Agris PF: Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* 1991, **73**(11):1345-1349.

67. Cabello-Villegas J, Tworowska I, Nikonowicz EP: Metal ion stabilization of the U-turn of the A37 N6-dimethylallyl-modified anticodon stem-loop of Escherichia coli tRNAPhe. *Biochemistry* 2004, **43**(1):55-66.

68. Lescrinier E, Nauwelaerts K, Zanier K, Poesen K, Sattler M, Herdewijn P: The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res* 2006, **34**(10):2878-2886.

69. Wada K, Wada Y, Ishibashi F, Gojobori T, Ikemura T: Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res* 1992, **20 Suppl**:2111-2118.

70. Nakamura Y, Gojobori T, Ikemura T: Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 2000, **28**(1):292.

71. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ: A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet* 2010, **6**(12):e1001247.

72. Robins MJ, Uznanski B: Nucleic acid related compounds. 33. Conversions of adenosine and guanosine to 2,6-dichloro, 2-amino-6-chloro, and derived purine nucleosides. *Can J Chem* 1981, **59**:2601-2607.

73. Ti GS, Gaffney BL, Jones RA: Transient protection: efficient one-flask syntheses of protected deoxynucleosides. *J Am Chem Soc* 1982, **104**(5):1316-1319.

74. Nair V, Young DA: A new synthesis of isoguanosine. *J Org Chem* 1985, **50**:406-408.

75. Adah SA, Nair V: Synthesis of complex ethynyladenosines using organic triflic enolates in palladium-catalyzed reactions: Potential agonists for the adenosine A2 receptor. *Tetrahedron* 1997, **53**:6747-6754.

76. Agris PF, Malkiewicz A, Kraszewski A, Everett K, Nawrot B, Sochacka E, Jankowska J, Guenther R: Site-selected introduction of modified purine and pyrimidine ribonucleosides into RNA by automated phosphoramidite chemistry. *Biochimie* 1995, **77**(1-2):125-134.

77. Sosnick TR: Characterization of tertiary folding of RNA by circular dichroism and urea. *Curr Protoc Nucleic Acid Chem* 2001, **Chapter 11**:Unit 11 15.

78. Powers T, Noller HF: Dominant lethal mutations in a conserved loop in 16S rRNA. *Proc Natl Acad Sci USA* 1990, **87**(3):1042-1046.

79. Fahlman RP, Dale T, Uhlenbeck OC: Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol Cell* 2004, **16**(5):799-805.

80. Markham NR, Zuker M: DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* 2005, **33**(Web Server issue):W577-581.

81. Wong I, Lohman TM: A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci USA* 1993, **90**(12):5428-5432.

82. Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A: NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995, **6**(3):277-293.

83. Goddard TD, Kneller DG: SPARKY3. In. San Franscisco, California, USA; 2007.

84. Stejskal EO, Tanner JE: Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J Chem Phys* 1965, **42**:288-292.

85. Tanner JE: Use of the stimulated echo in NMR diffusion studies. *J Chem Phys* 1970, **52**(2523-6).

86. Lapham J, Rife JP, Moore PB, Crothers DM: Measurement of diffusion constants for nucleic acids by NMR. *J Biomol NMR* 1997, **10**(3):255-262.

87. Stallings SC, Moore PB: The structure of an essential splicing element: stem loop IIa from yeast U2 snRNA. *Structure* 1997, **5**(9):1173-1185.

88. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al*: Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crysallogr D Biol Crystallogr* 1998, **54**(Pt 5):905-921.

89. Brünger AT: Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2007, **2**(11):2728-2733.

90. Stein EG, Rice LM, Brünger AT: Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J Magn Reson* 1997, **124**(1):154-164.

91. Lu XJ, Olson WK: 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 2003, **31**(17):5108-5121.

92. Koradi R, Billeter M, Wuthrich K: MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996, **14**(1):51-55, 29-32.

93. DeLano WL: Pymol. In. San Carlos, California, USA: Delano Scientific; 2006.

2.9 Tables and figures

Table 1. Thermal stability and thermodynamic properties[a] of the six $ASL^{Arg1,2}$ constructs.

$ASL^{Arg1,2}$	ΔH (kcal/mol)	ΔS (cal/K*mol)	ΔG_{37} (kcal/mol, 37°C)	T_m (°C)	Hyperchromicity (%)
ACG	-45.9 ± 0.4	-135.2 ± 1.1	-3.92 ± 0.02	66.0 ± 0.1	16.9 ± 0.7%
ACG-s^2C_{32}	-27.4 ± 0.1	-82.1 ± 0.3	-1.99 ± 0.02	61.2 ± 0.2	17.1 ± 0.4%
ACG-m^2A_{37}	-45.2 ± 1.0	-133.3 ± 2.9	-3.80 ± 0.13	65.8 ± 0.6	11.2 ± 0.2%
ICG	-39.7 ± 0.4	-116.9 ± 1.2	-3.48 ± 0.04	66.8 ± 0.2	14.5 ± 0.6%
ICG-s^2C_{32}	-38.6 ± 1.7	-114.7 ± 5.1	-2.99 ± 0.10	63.1 ± 0.5	12.7 ± 0.6%
ICG-m^2A_{37}	-40.4 ± 1.6	-119.7 ± 4.6	-3.30 ± 0.15	64.5 ± 0.5	12.6 ± 0.2%

[a]Determined from curve fitting analysis of thermal denaturation curves using MeltWin v3.5 (Fig. 2). All errors are reported as standard error of the mean.

Table 2. Structural statistics for $tRNA^{Arg1,2}_{ICG}$ and $tRNA^{Arg1,2}_{ICG}$-s^2C_{32} ASLs.

Restraints	Number of empirical restraints	
	$ASL^{Arg1,2}_{ICG}$	$ASL^{Arg1,2}_{ICG}$-s^2C_{32}
Distance restraints	253	231
Intranucleotides	123	112
Internucleotides	116	106
Exchangeable	6	6
Unoes[a]	7	7
Restraints by residue	14.8	13.5
Hydrogen bonding distances restraints	28	28
Dihedral restraints	94	93
Heavy atoms RMSD[b] from mean structure	**Angstroms (Å)**	
All residues (C_{27} - G_{43})	1.2 ± 0.4	1.6 ± 0.6
Loop residues (s^2C_{32}/C_{32} – A_{38})	0.6 ± 0.2	1.3 ± 0.6

[a]Unoes refers to unobserved NOEs.
[b]Root mean square deviation (RMSD) errors are reported as one standard deviation.

Table 3. Dissociation constants of the six ASL^{Arg1,2}.

ASL^{Arg1,2}	[a]*K*d = µM		
	CGU	**CGC**	**CGA**
ACG	0.70 ± 0.2	N/A	N/A
ACG-s^{2}C$_{32}$	0.70 ± 0.1	N/A	N/A
ACG-m^{2}A$_{37}$	1.30 ± 0.2	N/A	N/A
ICG	0.88 ± 0.2	0.66 ± 0.2	0.53 ± 0.1
ICG-s^{2}C$_{32}$	0.77 ± 0.3	0.38 ± 0.2	N/A
ICG-m^{2}A$_{37}$	1.40 ± 0.2	0.53 ± 0.5	N/A

[a]Dissociation constants, *K*d, (µM) derived from ribosomal A-site codon binding assays. The notation "N/A" was used for binding curves that either contained data points or error bars that fell below the x-axis. All errors are reported as standard error of the mean.

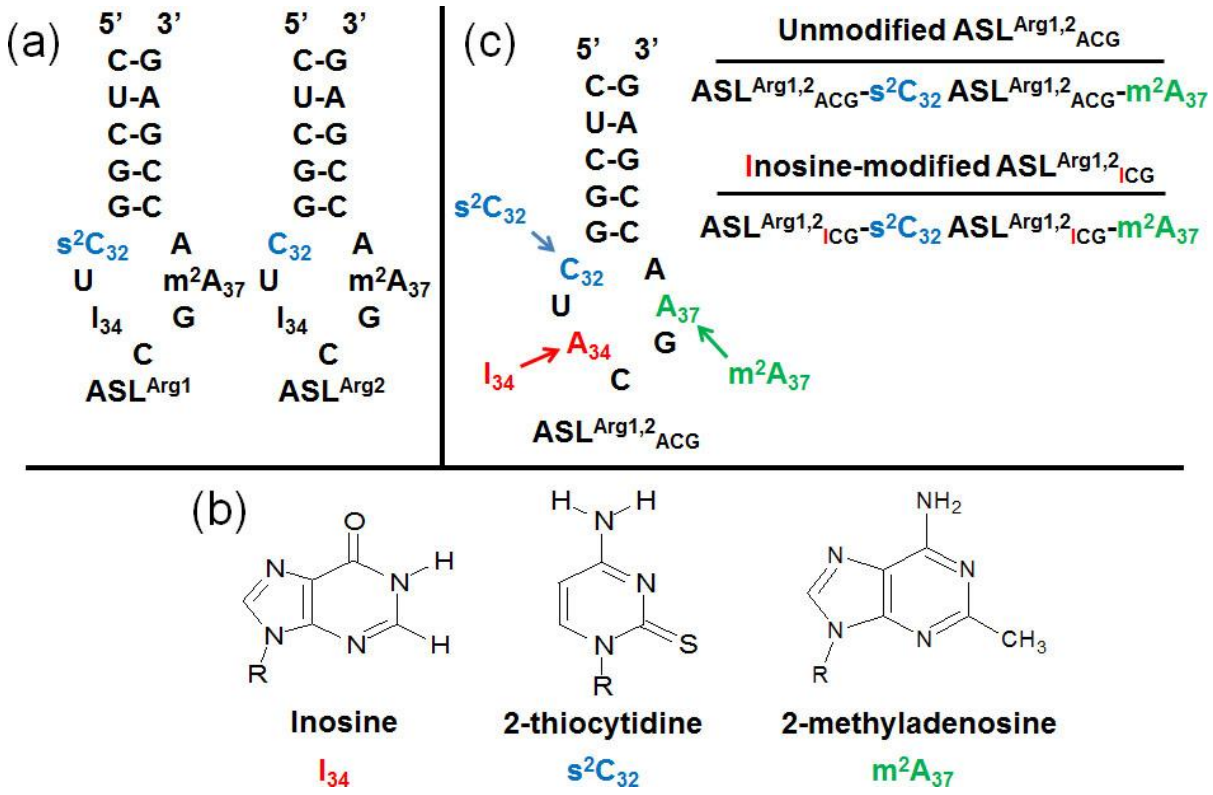

Figure 1. Anticodon stem and loop domains of *E. coli* tRNA^{Arg} (ASL^{Arg1,2}) and their modified nucleosides. (a) The primary sequence and secondary structures of the ASLs of the *E. coli* tRNA^{Arg1} and tRNA^{Arg2} with the difference between the two isoacceptors noted as a blue s^{2}C residue at position 32 of ASL^{Arg1}. **(b)** The chemical structures of the three naturally occurring modified nucleosides, inosine (I_{34}), 2-thiocytidine (s^2C_{32}) and 2-methyladenosine (m^2A_{37}). **(c)** Primary sequence and secondary structure of the ASL^{Arg1,2} showing the sites of modification that resulted in the variously modified six ASLs used in this study.

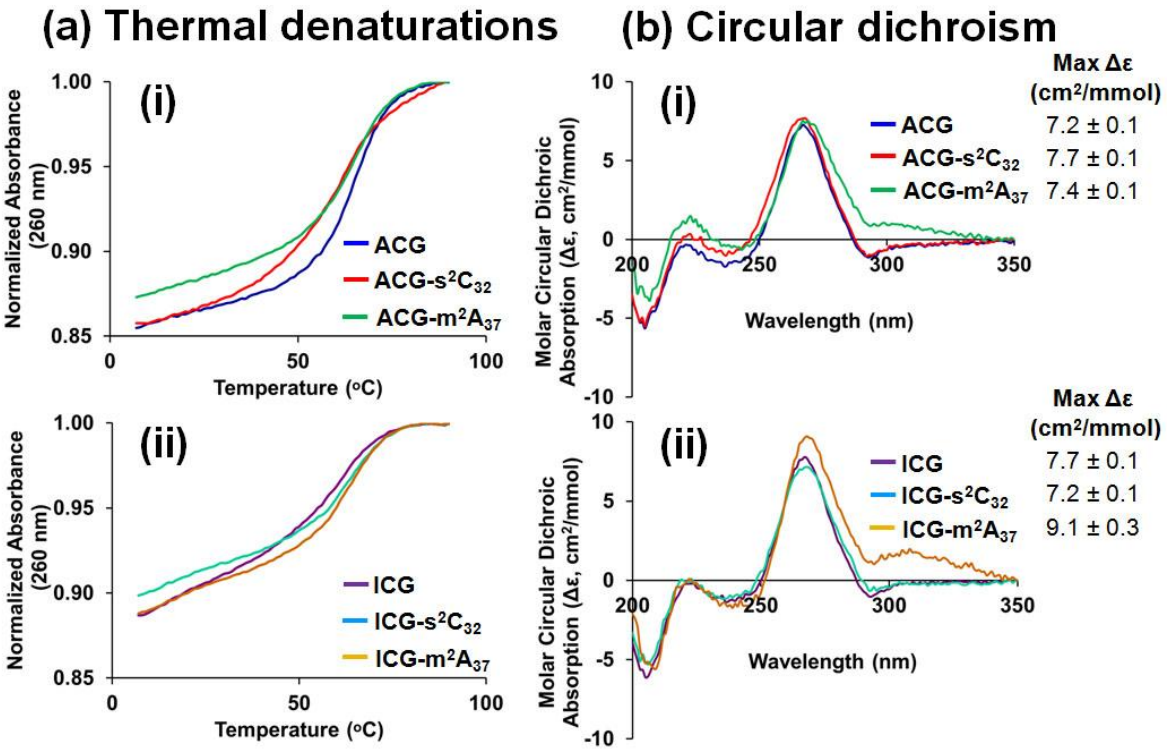

Figure 2. Thermal stability and base stacking of the ASL^{Arg1,2} constructs. (a) UV-monitored thermal denaturation and renaturation spectra of the ASL (~0.2 OD_{260}). Absorbance data was collected at 260 nm. The profiles shown are averages of at least five repeated denaturation/renaturation cycles (7-92 °C; 1 °C /min) of **(i):** ASLs containing A_{34} and **(ii):** ASLs containing I_{34}. **(b)** Circular dichroism (CD) spectra of all six ASL constructs. Spectra are the averages of nine normalized molar circular dichroic absorption spectra collected during three independent experiments with data points collected at a resolution of 1 nm for **(i):** ASLs containing A_{34} and **(ii):** ASLs containing I_{34}. Errors for maximum ellipticity are reported as one standard deviation from the mean.

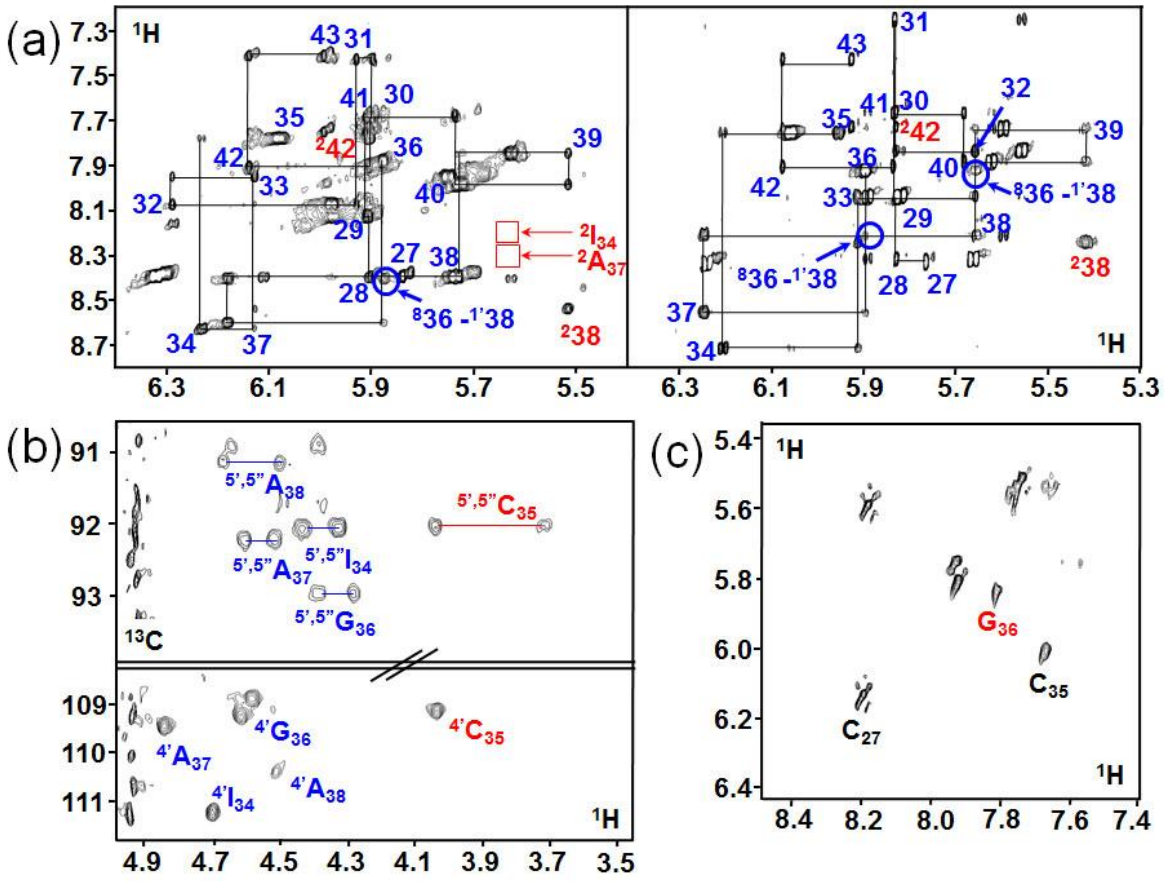

Figure 3. NOE-'walk' of NMR spectra indicate that ASL$^{Arg1,2}_{ICG}$ forms a A-form helix. **(a)** The sequential H1'-aromatic connectivities of the NOESY spectra. The H1' and H5 to aromatic region of a NOESY spectra (25 ºC; mixing time of 400 ms; D_2O) are shown for **Left:** ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$; and **Right:** ASL$^{Arg1,2}_{ICG}$. The unusual NOE connectivities between G_{36} H8 – A_{38} H1' and A_{38} H8 – G_{36} H1' are indicated. The lack of H2 resonances for I_{34} and A_{37} are denoted by red arrows and boxes showing where they would be expected. **(b)** A part of the ^{1}H-^{13}C HSQC spectrum of ASL$^{Arg1,2}_{ICG}$ shows the unusual upfield chemical shift of the C_{35}H4', H5' and H5'' resonances. **(c)** NOESY spectrum of ASL$^{Arg1,2}_{ICG}$ (50 ms) The H1' and H5 to aromatic region of a NOSEY spectrum (25 ºC; mixing time 50 ms; D_2O). The nucleoside numbering scheme reflects the numbering used in Figure 1.

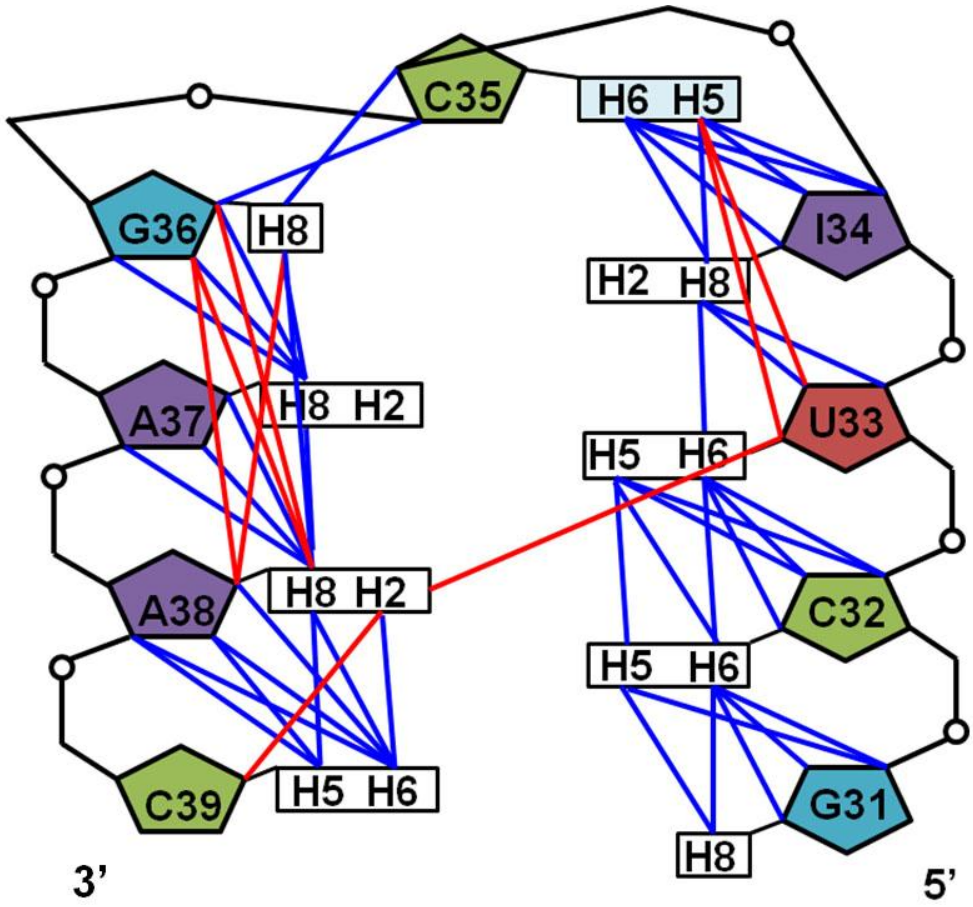

Figure 4. Inter-nucleotides NOEs of ASL$^{Arg1,2}_{ICG}$. A considerable number of NOEs were observed for the loop nucleosides C_{31} to A_{39} of all of the ASL^{Arg1,2}. The diagram illustrates some of the NOE observed for the anticodon loop nucleosides of ASL$^{Arg1,2}_{ICG}$. In blue and red are the sequential (i → i+1) and non-sequential inter-nucleotides NOEs, respectively.

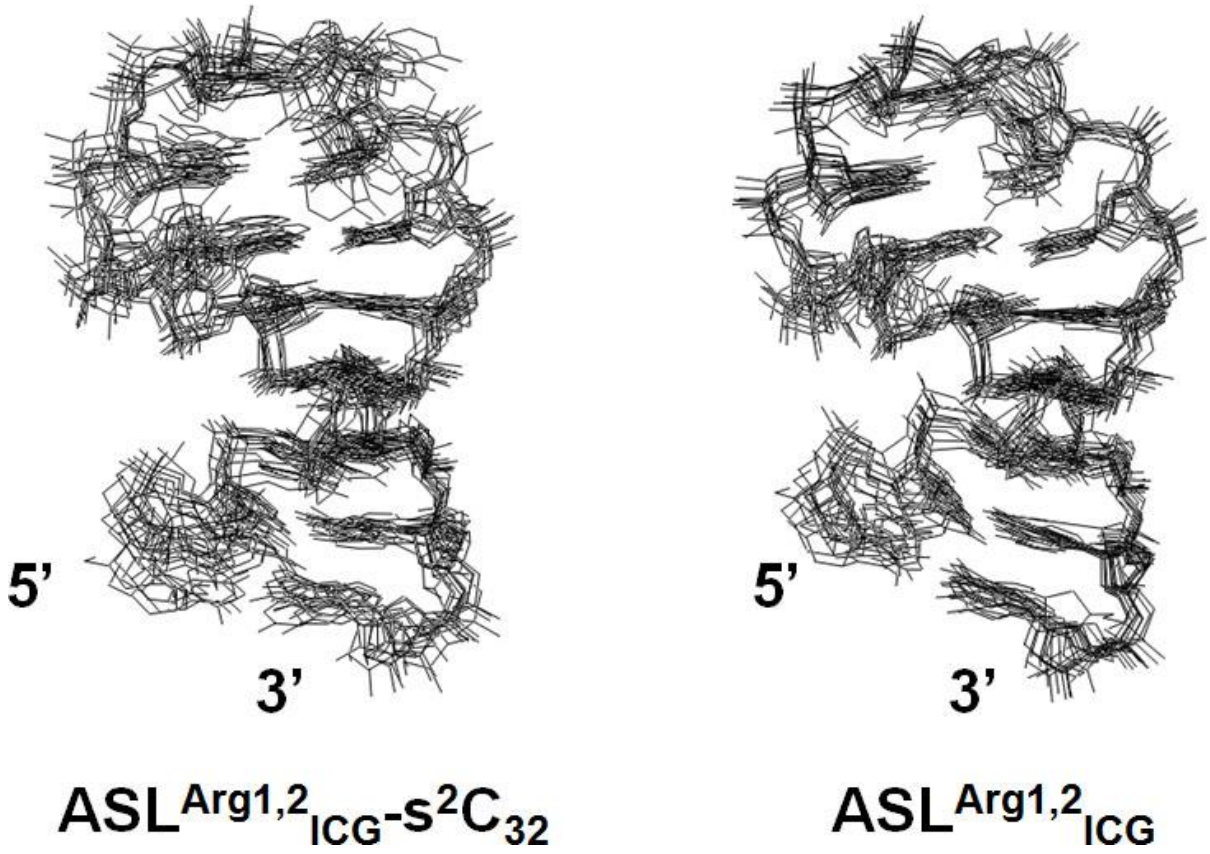

Figure 5. Lowest energy structures derived from NMR. Superposition of the 10 lowest energy structures of $ASL^{Arg1,2}_{ICG}$ (right) and $ASL^{Arg1,2}_{ICG}$-s^2C_{32} (left). For superposition, the individual structures are fitted to the average structure for nucleosides s^2C_{32}/C_{32} to A_{38} (hydrogen atoms have been removed for clarity).

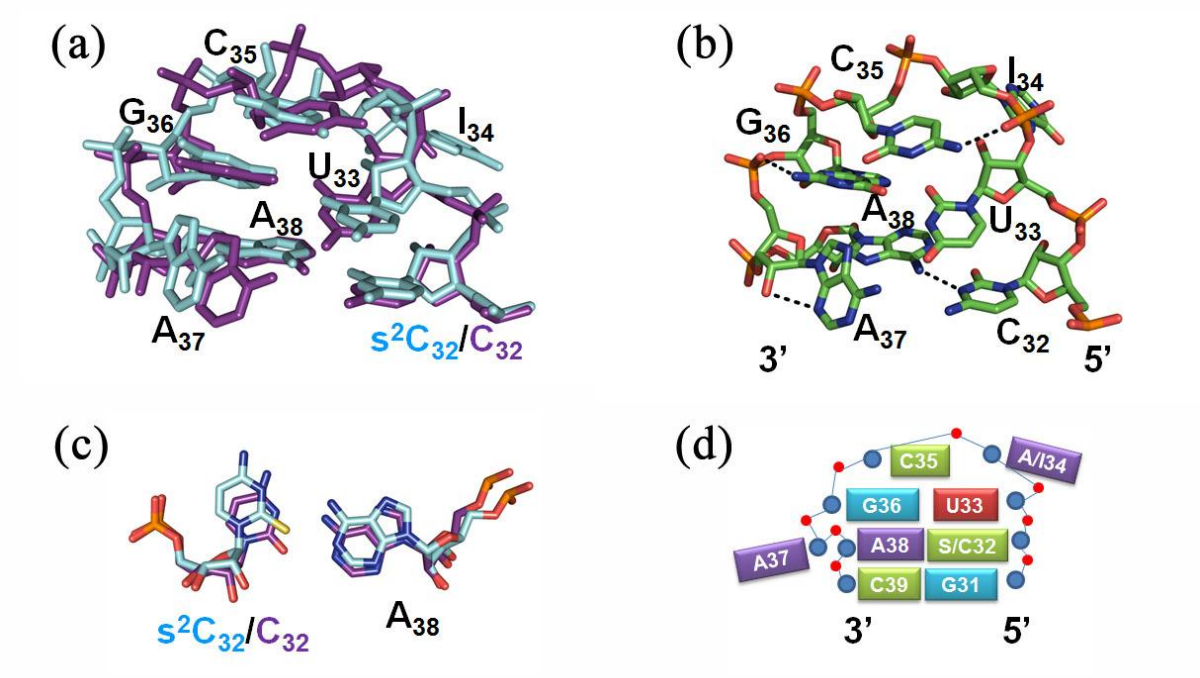

Figure 6. Structure and base pairing across the ASL$^{Arg1,2}_{ICG}$ loop; comparison of the ASL$^{Arg1,2}_{ICG}$ and ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$ loop conformations (nucleoside 32 to 38). (a) Superposition of the average structures for the nucleosides s^{2}C$_{32}$/C$_{32}$ to A$_{38}$ of ASL$^{Arg1,2}_{ICG}$ (violet) and ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$ (cyan). **(b)** The intra-loop hydrogen bond pattern stabilizing the ASL$^{Arg1,2}_{ICG}$ loop conformation (black dashes). **(c)** The s^{2}C$_{32}$/C$_{32}$•A$_{38}$ base pairs in ASL$^{Arg1,2}_{ICG}$ and ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$. **(d)** Schematic representation of the general conformations adopted by ASL$^{Arg1,2}_{ICG}$ and ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$. For clarity, the hydrogen atoms have been removed in all structures.

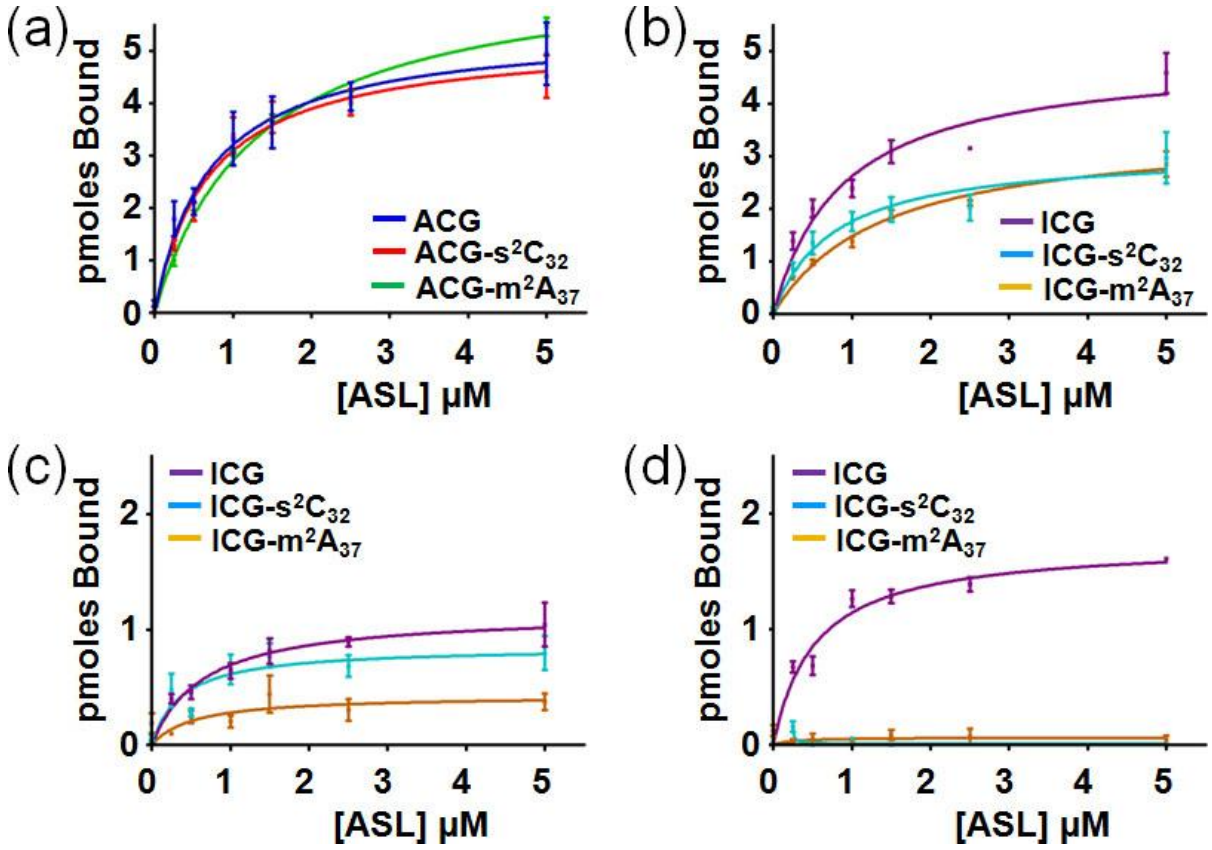

Figure 7. A-site codon binding by variously modified ASL^{Arg1,2}. The abilities of the various ASL^{Arg1,2} to bind A-site codons was assessed: **(a)** A_{34}-containing ASLs, ASL$^{Arg1,2}_{ACG}$, ASL$^{Arg1,2}_{ACG}$-s^2C_{32} and ASL$^{Arg1,2}_{ACG}$-m^2A_{37}, bound to CGU; **(b)** I_{34}-containing ASLs, ASL$^{Arg1,2}_{ICG}$, ASL$^{Arg1,2}_{ICG}$-s^2C_{32} and ASL$^{Arg1,2}_{ICG}$-m^2A_{37} bound to CGU; **(c)** I_{34}-containing ASLs, ASL$^{Arg1,2}_{ICG}$, ASL$^{Arg1,2}_{ICG}$-s^2C_{32} and ASL$^{Arg1,2}_{ICG}$-m^2A_{37}, bound to CGC; and **(d)** I_{34}-containing ASLs, ASL$^{Arg1,2}_{ICG}$, ASL$^{Arg1,2}_{ICG}$-s^2C_{32} and ASL$^{Arg1,2}_{ICG}$-m^2A_{37}, bound to CGA. Data for the binding of ASL$^{Arg1,2}_{ACG}$-s^2C_{34} and ASL$^{Arg1,2}_{ACG}$-m^2C_{37} to the non-cognate, synonymous codons CGC and CGA and the corresponding error bars fell below the x-axis and were excluded from the figure. Dissociation constants (K_d) were calculated for those ASL^{Arg1,2} constructs and codons for which there was observable binding to their synonymous codons (Table 2).

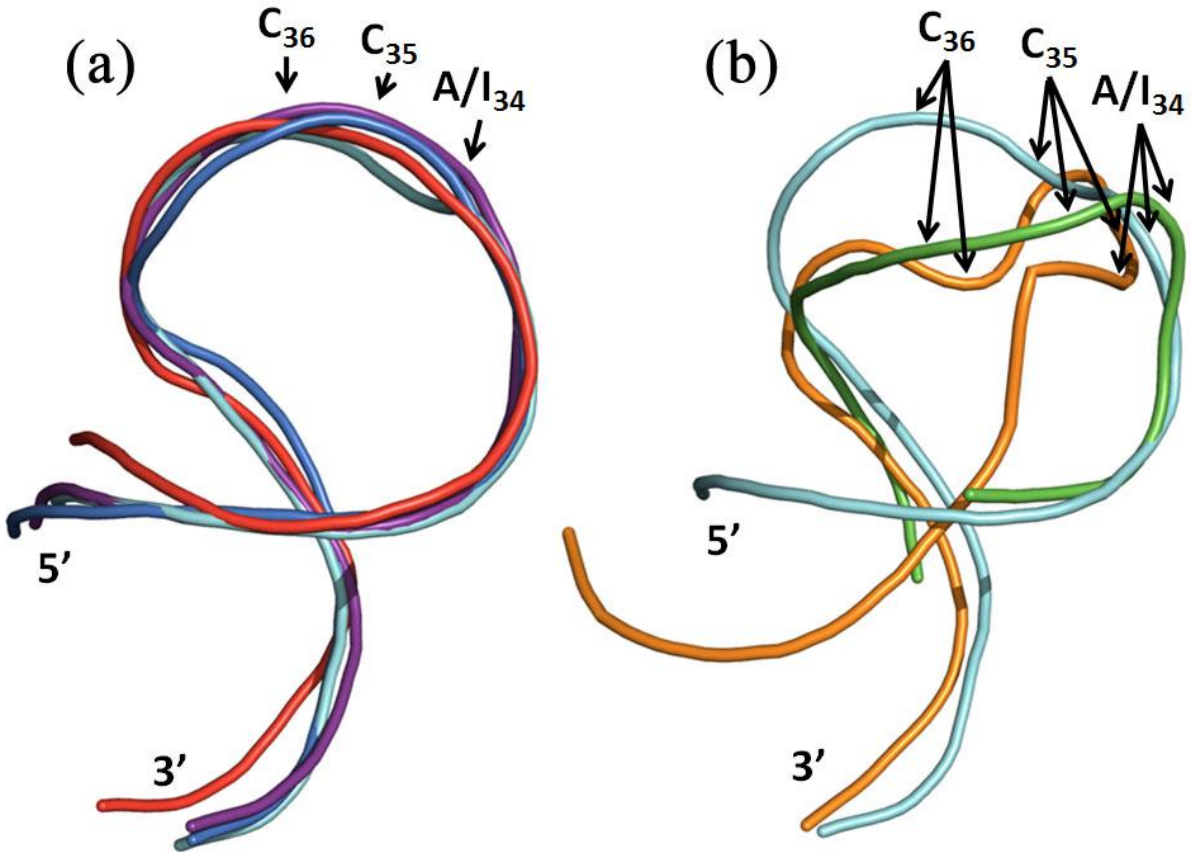

Figure 8. Solution and crystal structures of the arginyl-tRNA synthetase, and ribosome bound ASL^{Arg1,2} are very different from each other. (a) A superimposition of ASL$^{Arg1,2}_{ACG}$ (red), ASL$^{Arg1,2}_{ACG}$-s^{2}C$_{32}$ (blue), ASL$^{Arg1,2}_{ICG}$ (violet) and ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$ (cyan) reveals only minor differences in the solution structures of each respective ASL. **(b)** A superimposition of ASL$^{Arg1,2}_{ICG}$-s^{2}C$_{32}$ (cyan), ASL$^{Arg1,2}_{ICG}$ bound to the ribosomal A-site (green) and yeast ASL$^{Arg1,2}_{ICG}$ bound to the arginyl-tRNA synthetase shows drastic conformational differences among the three structures. The nucleosides were removed for clarity, and the positions of the backbone phosphates of the anticodon are shown.

2.10 Supplementary tables and figures

Table S1. Diffusion constants for ASL constructs.

RNA	Diffusion constant ($x10^{-6}cm^2/s$)
Controls[a]	
24 mer double stranded	1.81 ± 0.12
56 mer double stranded	1.39 ± 0.16
17mer $tRNA^{Val}$ ASL	2.38 ± 0.03
$ASL^{Arg1,2}$ Contructs	
$ASL^{Arg1,2}_{ACG}$	ND[b]
$ASL^{Arg1,2}_{ACG}$-s^2C_{32}	2.16 ± 0.12
$ASL^{Arg1,2}_{ICG}$	2.30 ± 0.08
$ASL^{Arg1,2}_{ICG}$-s^2C_{32}	2.27 ± 0.02

[a]The diffusion of control RNAs were tested to reflect that larger double stranded RNAs show a much lower diffusion constant. All of the tested ASLs diffuse at a rate consistent with the control 17 mer ASL^{Val} monomer.

[b]ND: not determined

Table S2. Codon usage for six *E. coli* strains.

Strain	CGU	CGC	CGA
0157:H7 EDL933	20.2	20.8	3.8
Sakai	20.3	20.9	3.9
CFT073	20.3	21.0	3.9
536	20.8	21.1	3.7
UT189	20.4	21.2	3.8
APEC 01	20.3	20.8	4.0
Codon Usage[a]	20.4 ± 0.2	21.0 ± 0.2	3.9 ± 0.1

[a]Average per 1000 codons in six *E. coli* strains with >4500 CDS's The mean is displayed in the final row with error recorded as one standard deviation.

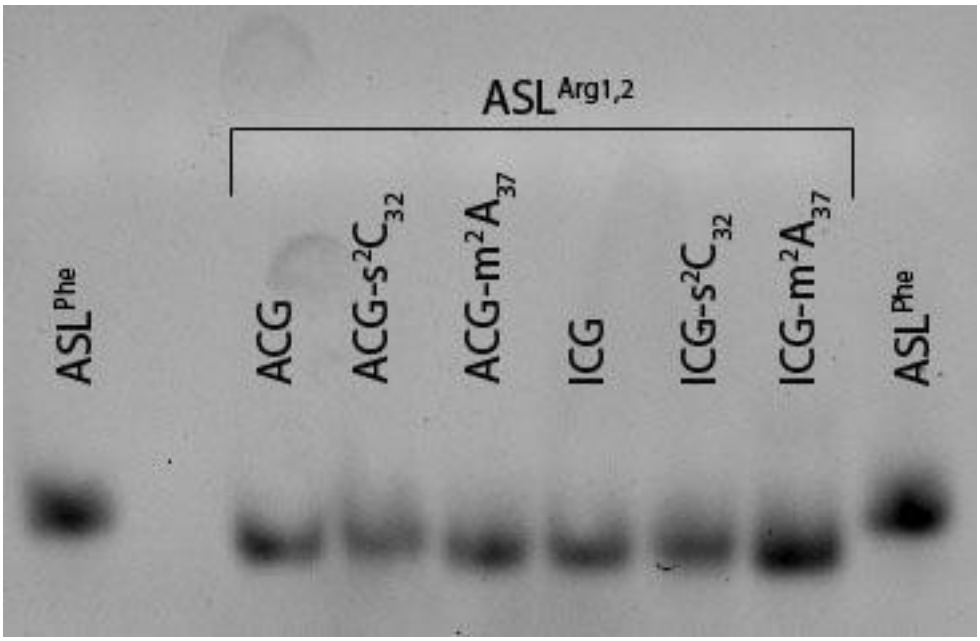

Figure S1. Non-denaturing polyacrylamide gel electrophoresis of all six ASL constructs. All samples were at a concentration of 30 μM for the gel electrophoresis (18% acrylamide-bisacrylamide). Unmodified yeast ASL^{Phe} having a canonical U-turn was run as the control.

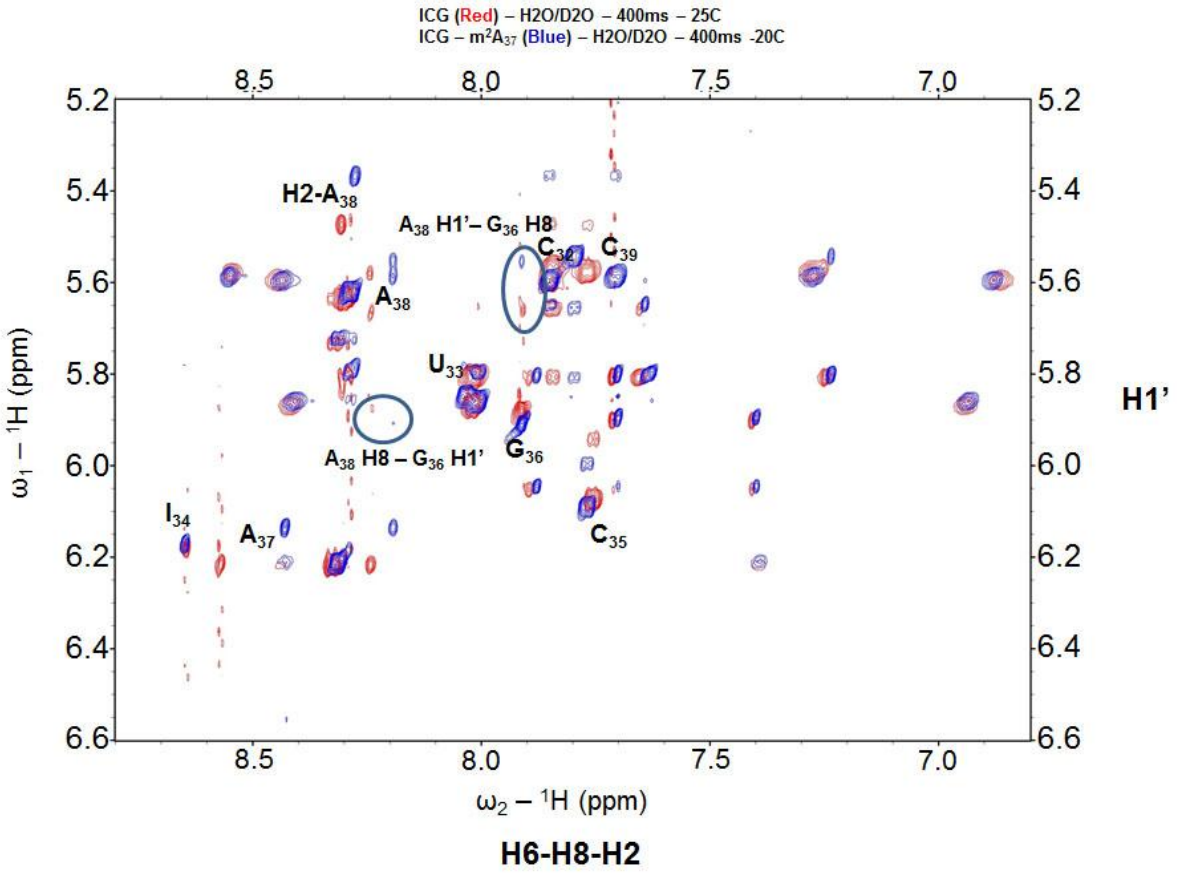

Figure S2. Superimposition of the H1' and H5 to aromatic region of the 400 ms mixing time NOESY spectra for $ASL^{Arg1,2}_{ICG}$ (red) and $ASL^{Arg1,2}_{ICG}$-m^2A_{37} (blue). Interesting chemical shift differences of the A_{38}H1' to G_{36}H8 and G_{36}H1' to A_{38}H8 are indicated by blue ovals.

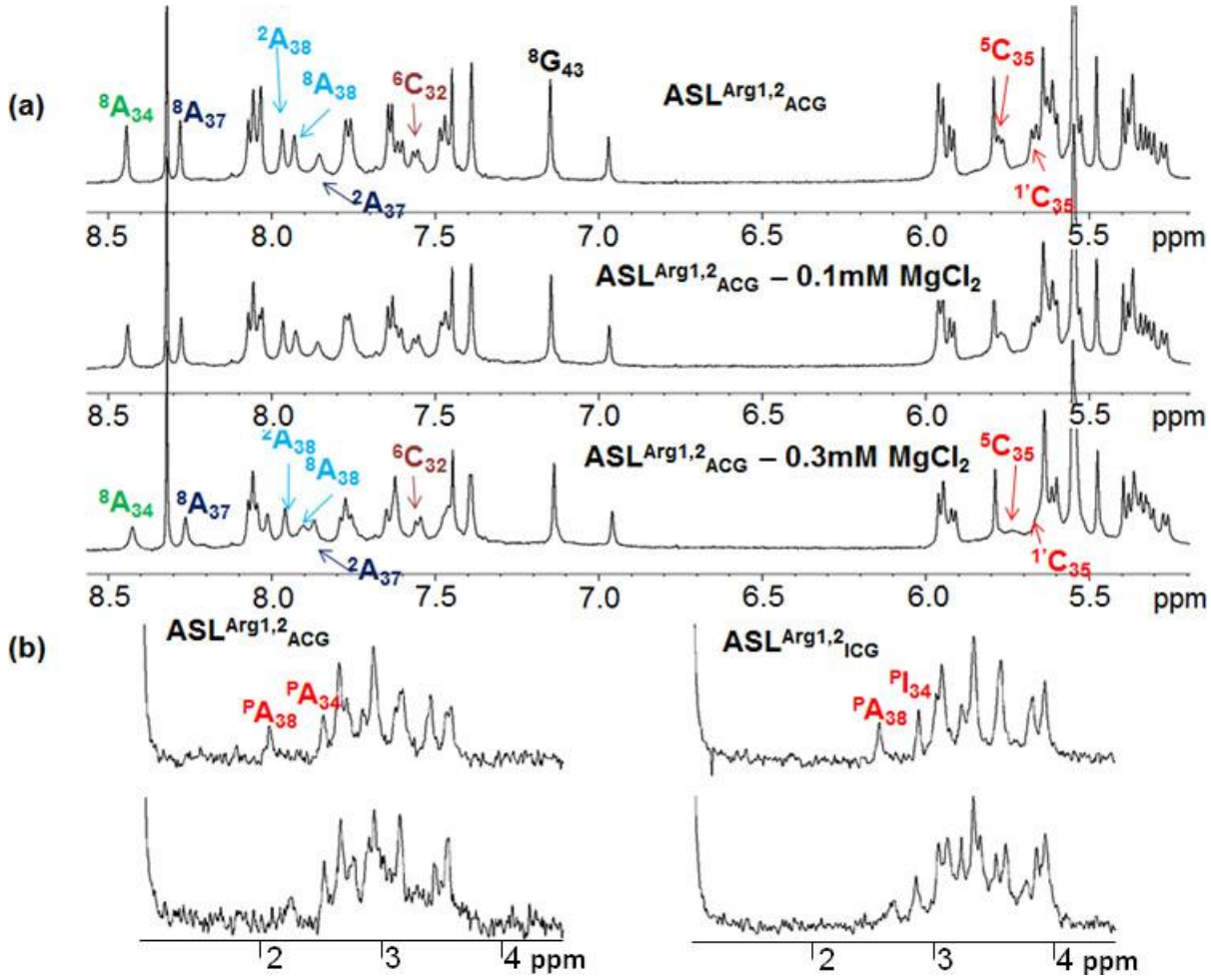

Figure S3. Effects of the magnesium and cobalt metal ions on the NMR spectra of unmodified $ASL^{Arg1,2}_{ACG}$ and singly modified $ASL^{Arg1,2}_{ICG}$, respectively. **(a)** Portion of a 1D 1H NMR spectra of $tRNA^{Arg1,2}_{ACG}$ showing the aromatic and H1' resonances after addition of $MgCl_2$. **(b)** ^{31}P spectra of the $ASL^{Arg1,2}_{ICG}$ (right) and $ASL^{Arg1,2}_{ACG}$ (left) after addition of $MgCl_2$ and $[Co(NH_3)_6]^{3+}$, respectively.

CHAPTER 3. Molecular dynamics simulations reveal modification induced alteration in *E. coli* $tRNA^{Arg1}$ conformational transitions

To be submitted once completed with the following authorship:

William A. Cantara, Erick Harr and Paul F. Agris

3.1 Statement of project contribution

The objective of this project was to determine if there were specific conformational pathways leading from the solution structure of the ASL^{Arg1} to that of the ASL on the cognate aminoacyl-tRNA synthetase, and to that on the programmed ribosome. I had the pleasure of training a very helpful undergraduate, Erick Harr, to assist with the preparation and simulation of the ribosome-bound structures as well as analysis and interpretation of the resulting structural and statistical data. I was responsible for all aspects of this project, from initial strategy to data interpretation, as well as delegation to Erick and coordination with the personnel from the Research Information Technology department.

Abbreviations: ASL, anticodon stem and loop; I_{34}, inosine in position 34; mRNA, messenger RNA; RMSD, root mean square deviation; s^2C_{32}, 2-thiocytidine in position 32; m^2A_{37}, 2-methyladenosine in position 37; MD, molecular dynamics; TMD, targeted molecular dynamics.

Keywords: 2-thiocytidine, inosine, 2-methyladenosine, RNA structure, RNA function, molecular dynamics

3.2 Abstract

Anticodon stem and loop domain (ASL) modifications in tRNA affect the conformation and dynamics of the structure at equilibrium, thus modulating the decoding function of the tRNA. It has been shown that the modifications 2-thiocytidine at position 32 (s^2C_{32}) and 2-methyladenosine at position 37 (m^2A_{37}) of *Escherichia coli* (*E. coli*) $ASL^{Arg1,2}{}_{ICG}$ can alter decoding capacity, while having no effect on the solution structure at equilibrium. Spectroscopic results suggest that modifications modulate the thermal stability and base stacking properties of the ASL. Additionally, this ASL has the uncommon property of adopting very different structures in solution, when bound to the cognate aminoacyl-tRNA synthetase and on the ribosome. It is unclear how modifications modulate decoding ability without altering the structure at equilibrium and whether they play a role in regulating adoption of the very different, but necessary conformations for aminoacylation and ribosome-mediated codon binding. To test whether modifications alter the conformational space of $ASL^{Arg1,2}{}_{ICG}$, explicit solvent all-atom molecular dynamics (MD) simulations were performed on the ASL in different states of modification using the three different conformations as starting conditions. These MD results show that modifications influence conformational dynamics in a manner that is dependent both on the state of modification and the starting conformation, suggesting that modifications alter the ability of the ASL to switch conformations. In order to investigate the pathways of the energy landscape between the different conformations, all-atom targeted MD (TMD) simulations were conducted. Preliminary investigations of the preferred conformational pathways between the solution structure and the canonical U-turn motif using TMD indicate that at least five different pathways are used by ASL^{Arg1}. The distribution of which pathways each ASL construct used was dependent on modification state. Therefore, modifications altered the residue-specific conformational space of the ASL, directing passage through certain conformational pathways and intermediates.

3.3 Introduction

Proper translation of the genetic code is highly dependent on nucleoside modifications of tRNA's anticodon domain [1,2]. There are no less than 93 different modifications that can be found in tRNAs of all species [3]. These modifications are highly enriched within the anticodon stem and loop (ASL) in both number and diversity [4,5]. ASL modifications perform many crucial translational functions such as reading frame maintenance [6,7] and promoting translocation [8]. Additionally, modifications play a critical role in pre-structuring the loop into a U-turn-like conformation commona among ASLs [1,2,9-11]. The pre-structuring is important to reduce the amount of ASL deformation necessary to adopt the canonical U-turn motif that is required for the binding of codons in the ribosomal A-site. This, in turn, reduces the requisite energy for proper codon recognition as the tRNA is used repeatedly.

In *Escherichia coli*, $tRNA^{Arg1}_{ICG}$ (Figure 1a) is responsible for decoding three of the six arginine codons. These include the two common codons CGU and CGC in addition to the rare CGA codon [12]. Interestingly, another isoacceptor, $tRNA^{Arg2}_{ICG}$, is also thought to be responsible for decoding these same codons. The two isoacceptors have nearly identical primary structure, differing only at position 20, where $tRNA^{Arg2}_{ICG}$ contains an additional adenosine in the dihydrouridine loop, and at position 32 (see Figure 1a for nucleoside numbering) in the anticodon stem and loop (ASL) where $tRNA^{Arg1}_{ICG}$ has the rare cytidine modification 2-thiocytidine (s^2C_{32}) [13-15]. In addition to the rare s^2C_{32} modification in $tRNA^{Arg1}_{ICG}$, both isoacceptors contain adenosine modifications at positions 34 and 37, inosine (I_{34}) and 2-methyladenosine (m^2A_{37}), respectively (Figure 1b). Inosine has the functional property of allowing the tRNA to decode three different codons ending in U, C or A when located at position 34, the "wobble" position [16].

The rare s^2C_{32} modification is of particular interest because of both its rarity and the fact that it is found in four out of five of the *E. coli* arginyl tRNA isoacceptors and $tRNA^{Ser2}_{GCU}$ [13]. However, relatively little is known about the 2-thiolated cytidine in the context of

structure and function of the ASL. Indeed, the only arginine isoacceptor lacking s^2C_{32} is $tRNA^{Arg2}_{ICG}$, which has an otherwise identical ASL sequence as $tRNA^{Arg1}_{ICG}$ [13-15]. The identity and geometric orientation of the residue at position 32 has been shown to regulate the ability of the tRNA to discriminate between different codons [17]. Position 32 is modified in ~32% of known tRNA sequences, 83% of which are in the form of either 2-thiocytidine or a 2'-*O*-methylated pyrimidine [13]. These types of modifications have similar structural characteristics. The 2'-*O*-methyl group causes a steric hinderance with the aromatic base of the nucleoside in the C2'-*endo* sugar pucker conformation, resulting in the energetically more favorable C3'-*endo* conformation [18-20]. The larger atomic radius of the thioketone in 2-thiolated pyrimidines as compared to that of the 2-carbonyl stabilizes both a C3'-*endo* sugar pucker and an *anti N*-glycosidic orientation. The greater polarizability of the thioketone promotes stacking interactions between adjacent nucleosides [21]. There are three crystal structures containing the singly modified version ASL^{Arg1}_{ICG} with only inosine at position 34. One contains the entire tRNA bound to the yeast arginyl-tRNA synthetase (argRS) [22]. The two other crystal structures show the ASL bound to non-cognate, synonymous codons in the ribosomal A-site, one bound to CGC and one bound to CGA [23]. Although the crystal structure of the $tRNA^{Arg1}_{ICG}$-argRS complex is from yeast [22], a crystal structure of the bacterial synthetase from *Thermus thermophilus* [24] shows a high amount of structural conservation when compared to the argRS (heavy atom RMSD difference of 2.73Å, Figure 2a). Additionally, the yeast anticodon loop primary sequence is identical to that of *E.coli* and other bacteria [13]. Therefore, it is likely that the ASL must adopt a very similar conformation in bacteria.

Surprisingly, the conformations of ASL^{Arg1}_{ICG} when bound to the ribosome and argRS are drastically different (Figure 2b). When bound in the ribosomal A-site, ASL^{Arg1}_{ICG} conforms to the canonical U-turn structure in which all loop residues are in the C3'-*endo* sugar pucker. There is highly organized base stacking on both the 5' and 3' sides of the loop. The Watson-Crick faces of the anticodon residues are solvent exposed [23]. In contrast to this conventional structure, ASL^{Arg1}_{ICG} bound to argRS reveals a highly distorted loop. C_{35}, G_{36}

and A_{38} are flipped outside of the loop, and G_{36} and A_{38} are buried in the protein. I_{34} forms a stacking interaction with tryptophan-567 [22]. The structure is stabilized by a cross-strand stacking interaction of A_{37} between the terminal stem base pair and C_{32}. This distortion of the ASL may reflect the specificity of synthetase recognition. With little room for alteration in the position 35 and 36 anticodon residues, these structures appear to hinge on the chemical properties of the nucleosides at positions 32, 37 and 38.

We recently reported the structural, thermodynamic and functional implications of the three $tRNA^{Arg1}_{ICG}$ modifications in the context of the ASL [25]. The study was, however, limited by the inability to obtain the triply modified ASL for experimentation and was therefore carried out using six differentially modified ASLs. While the ASLs showed significant differences in their thermodynamic properties and codon binding capabilities, their solution structures were nearly identical. Specifically, the modifications on the 5' side of the loop, s^2C_{32} and I_{34}, caused a destabilization of the ASL and the 3' modification, m^2A_{37}, tended to have a stabilizing effect. Functionally, I_{34}, as predicted by the "Wobble Hypothesis" [16], enabled decoding of all three synonymous codons (CGU, CGC and CGA). In combination with I_{34}, however, both s^2C_{32} and m^2A_{37} restricted codon recognition to only CGU and CGC. These thermodynamic and functional differences were not evident in the solution structures of the ASLs. All differentially modified ASLs conformed to a very stable pseudo $^{5'}$UNCG$^{3'}$ tetraloop conformation. Interestingly, this structure does not present any of the hallmark characteristics of either of the crystal structure conformations seen on the ribosome or argRS (Figure 2b).

In the present study, molecular dynamics simulations were conducted to decipher the conformational energetic and energy landscape by which anticodon domain modifications of $tRNA^{Arg1}_{ICG}$ modulate thermodynamic and functional decoding characteristics. Additionally, we sought to predict whether the fully modified ASL (ASL^{Arg1}_{ICG}-s^2C_{32};m^2A_{37}) would promote the necessary conformational transitions to enable three-fold degenerate codon recognition. Seven ASLs with different states of modification (Figure 1a) were simulated from the three known starting structures (solution, ribosome-bound and argRS-bound, Figure

2b) using standard explicit solvent molecular dynamics. Then, targeted molecular dynamics were conducted to simulate the transition from the solution structure to a canonical U-turn conformation. Here, we report that modifications can both modulate residue-specific dynamics and alter the pathways for the lowest energy transition between two biologically important conformations. The results support the hypothesis that modifications act as regulators of aminoacylation and ribosome-mediated decoding by modulating the energy requirement for adopting the necessary conformations for each activity.

3.4 Results

3.4.1 ASL and individual residue flexibility

The results of biophysical characterizations have shown modification-induced biophysical variation in differentially modified ASL^{Arg1}_{ICG} constructs [25]. To further investigate the molecular mechanisms underlying these variations, standard explicit solvent molecular dynamics simulations were performed using the SANDER module in the AMBER 11 suite of programs [26]. Explicit solvent simulations are performed with actual atomistic simulations of water molecules in potential box surrounding the solute. This type of simulation was chosen based on the inherent properties of water molecule stabilization of coordinated ions and RNA:RNA interactions such as base stacking and base pairing. Starting structures were derived from NMR solution structures previously solved [25], minimized, equilibrated and simulated at constant temperature and pressure for 6 ns. The final 5 ns of production simulation trajectories were analyzed using the ptraj module from AmberTools1.5 [26]. A plot of heavy atom root mean square deviation (RMSD) *versus* time reveals only slight variations between the differentially modified ASLs (Figure 3a). In fact, none of the plots show any significant deviation from what would be considered the average structure RMSD.

To more specifically determine the effects of modifications, B-factors (a measure of atomic fluctuation in which a higher value corresponds more conformational variability)

were calculated from the trajectories for each residue. For ASLs that contain both A_{34} and I_{34} the flexibility of the stem residues C_{27}-G_{31} and C_{39}-G_{43} was significantly less than that of the loop residues, with the exception of the terminal stem residues (Figure 3b). This is expected due to the lower stability of terminal base pairs and the high stability of internal base pairs in A-form RNA stems. Interestingly, the flexibility of residues C/s^2C_{32} and U_{33} were significantly less flexible than the other stem residues. Since the key differences in decoding were seen for the inosine-containing ASLs [25], it is interesting to note the properties of modifications in these ASLs. First, m^2A_{37} caused a significant increase in its own conformational space; but had little effect on any other residue. In contrast to m^2A_{37}, s^2C_{32} caused a slight increase in the flexibility of all three anticodon residues and contributed a significant decrease in plasticity of A_{37}. Though empirical studies on the fully modified ASL were precluded by difficult chemical synthesis, the molecular dynamics-derived B-factors predicted that this ASL would have loop deformability properties that closely resemble that of the singly modified ASL^{Arg1}_{ICG} with one exception. Residues C_{35} and G_{36} would have slightly more conformational space. In summary, molecular dynamics simulations reveal trends in loop residue flexibility that may affect the ability of the ASL to adopt different induced conformations when bound to the ribosome or argRS.

While modification-induced changes in conformational space of the ASL solution structure ASL were evident, these results could not be extrapolated to that of the loop residues bound to codons in the ribosome or those of the argRS-bound conformation. Therefore, the differentially modified ASLs were simulated using both their ribosome-bound and argRS-bound structures in the absence of the bound ribosome and argRS. The same simulation conditions were applied as with the solution structure. This allowed for a macromolecule-independent determination of the relative stability and flexibility of each ASL conformation. Similar to the NMR solution structure, the RMSD *vs* time plots of the differentially modified ASLs in either starting conformation were very similar (Supplementary Figure S1). B-factors for each individual residue were calculated. In both the canonical U-turn conformation seen in the ribosomal A-site (Figure 4a) and the highly

distorted argRS-bound conformation (Figure 4b), the stem residues were very stable and show very little flexibility with the exception of the terminal stem residues. When in the U-turn conformation of the ribosome structure, there is a slight increase in the B-factors associated with the anticodon residues over that seen for the solution structure (Figure 4a). This is not surprising considering the highly stable interactions occurring between the loop residues and the flared-out nature of A_{37} in the UNCG motif of the solution structure and the solvent-exposed positioning of the anticodon residues in the U-turn. Similar to the U-turn conformation, the distorted conformation of the argRS-bound ASLs allowed for a significant increase in the B-factors associated with U_{33}, A/I_{34} and C_{35} (Figure 4b). In both the argRS and ribosome structures, A_{37} showed a much more stable dynamic than in the solution structure.

In addition to a dependence on starting structure, the residue-specific dynamics were also dependent on modification state. When modified with m^2A_{37}, this residue in the solution structure is significantly more flexible as indicated by a higher B-factor; however, this effect is more or less pronounced in the presence of I_{34} or s^2C_{32}, respectively (Figure 3b). Despite the ability to stabilize position 37, s^2C_{32} in combination with I_{34} enhances the motion U_{33}, I_{34}, C_{35} and G_{36}. With the exception of the fully modified ASL, modifications generally have a stabilizing effect on all loop residues of the U-turn structure seen on the ribosome, compared to the unmodified ASL (Figure 4a). In contrast, modifications appear to destabilize the loop residues of the argRS-bound conformation.

3.4.2 Conformational pathways between UNCG and U-turn motifs

While standard molecular dynamics simulations showed that modifications can modulate loop plasticity and deformability, using only standard molecular dynamics, it is impossible to determine whether the modifications can regulate the ability of the ASL to conform to a canonical U-turn structure. A comprehensive study of all ASL constructs is currently in progress; however, complete targeted molecular dynamics (TMD) simulations were performed on unmodified ASL^{Arg1}_{ACG} and s^2C_{32}, I_{34} and m^2A_{37} modified ASL^{Arg1}. This was

used to investigate the conformational pathways that are traversed between the solution structure and the U-turn necessary for A-site codon binding. This strategy introduced an energy penalty to the simulation that is proportional to the heavy atom RMSD between the current structure and a reference structure, in this case, the reference structure is the U-turn conformation. This energy penalty creates a bias in which the system tends toward the reference structure.

During TMD simulations, five different conformational pathways were observed for the unmodified and various singly modified ASLs (Table 1, Figure 5 and 6). The most unusual pathway, Pathway 2, only occurred during simulations involving the unmodified ASL. It is characterized by an intermediate in which A_{37} stacks between A_{38} and C_{39}. This breaks the stacking interaction between A_{37} and G_{36}, which allows G_{36} the flexibility to swing outside of the loop, where the *N*-glycosidic (χ) bond is free to rotate from *syn* to *anti*. G_{36} forms a highly stable stacking network with C_{35} and A_{34}, which flip out to stack with G_{36}. Finally, A_{37} was released from base-pairing with C_{32} and stacking between A_{38} and C_{39} to flip out and complete the highly stacked network from A_{34} to C_{39}. This allowed A_{38} to re-form a non-canonical base pair with C_{32}. In contrast to Pathway 2, Pathways 1, 3, 4 and 5 are characterized by an initial structural change in which G_{36} and A_{37} flip from the inside to the outside of the loop at the same time and stack together. During this flip, the χ dihedral angle of G_{36} either switches from *syn* to *anti* or stays *syn* for Pathways 1 and 3 or Pathways 4 and 5, respectively. In all four of these pathways, C_{35} then flips from the inside to the outside of the loop creating the characteristic U-turn with a highly stacked network of bases from the anticodon along the 3' side of the loop. In Pathway 4, it is now that the base of G_{36} rotates from *syn* to *anti*-conformation; however in Pathways 3, 4 and 5, C_{35} is not able to adopt an *anti*-conformation for the duration of the simulation.

Based on five repetitions for each ASL construct, we have divided the pathways of each ASL into a percentage chance of following each pathway (Table 2). Interestingly, only unmodified ACG was able to follow Pathway 2. Indeed, either of the modifications individually was able to completely abrogate the ability of A_{37} to stack between A_{38} and C_{39}

and form a non-canonical base pair with C_{32}. Despite an apparent preference for Pathway 2, ACG also followed Pathway 1 in 40% of the simulations. Both of these pathways resulted in the unmodified ASL adopting a canonical U-turn conformation with all loop bases in the *anti* glycosidic conformation. Similarly, the addition of m^2A_{37} still allowed for Pathway 1 to be followed 40% of the time, however, 60% of the time, Pathways 3, 4 and 5 were used resulting in either one or both of C_{35} or G_{36} adopting an unconventional *syn* conformation. For the remaining two singly modified ASLs with s^2C_{32} and I_{34}, TMD simulations revealed that all replicates followed Pathways 3, 4 or 5 indicating that there was a preference for at least one of the anticodon residues to be in the *syn N*-glycosidic conformation. To summarize the results of the TMD simulations, each differentially modified ASL was able to follow anywhere from two to four different conformational pathways from solution structure to ribosomal A-site structure. The distribution of pathways followed was dependent upon modification state. The modification-dependent pathways followed by either the unmodified ACG or singly modified ASL^{Arg1}_{ACG}-m^2A_{37} appear to be preferential due to their ability to adopt the fully canonical U-turn conformation. One commonality of all pathways is the preference for the C2'-*endo* sugar conformation of G_{36} until it rotates outside of the loop. It then adopts a C3'-*endo* conformation independent of the *N*-glycosidic rotation.

3.5 Discussion

The structure and function relationship of the anticodon domain of *E.coli* $tRNA^{Arg1}_{ICG}$ is an anomaly in the field of tRNA modification research. First, this isoacceptor is solely responsible for decoding both very common codons (CGU and CGC) and a very rare codon (CGA) [12,25]. Secondly, the structure of the ASL has been shown to be independent of up to two modifications [25]. Finally, the singular structure seen in solution for all differentially modified ASLs is more similar to highly stable 5'-UNCG-3' motif [25] rather than either the canonical U-turn conformation required for codon binding in the ribosomal A-site [23] or the highly distorted structure required for argRS recognition [22]. To investigate these unusual

characteristics and test the hypothesis that modifications alter the dynamic properties of these ASLs, molecular dynamics simulations were performed. Here, we have shown both how modifications can modulate the deformability of the loop by specifically altering the flexibility of specific loop residues and that modifications can alter the conformational pathway between the solution structure and the U-turn motif.

Each of the three conformations adopted by the singly modified ASL^{Arg1}_{ICG} has different identifying structural characteristics. In the UNCG-like motif, there is a stable stacking interaction between G_{36} and A_{38} which both stabilizes G_{36} in a *syn N*-glycosidic, C2'-*endo* orientation and causes A_{37} to be flipped out of the loop and become fully solvent exposed [25]. Further stabilizing G_{36} in this unfavorable geometry is a single hydrogen bond between G_{36}N2 and the phosphate of A_{37}. This unusual positioning of G_{36} and the consequent stacking with C_{35} results in the stabilization of an unusual positioning of C_{35}, directly below the base of U_{33}. Indeed C_{35} is also stabilized by a hydrogen bond between C_{35}N4 and the backbone phosphate of A/I_{34}. It should be noted that in this conformation, C/s^2C_{32} stacks well between G_{31} and U_{33} and makes an unusual bifurcated hydrogen bond with A_{38}.

In contrast to the UNCG motif, the U-turn is a highly ordered structure on the ribosome [23]. All of the loop residues are very well stacked and adopt a C3'-*endo* sugar pucker and *syn* glycosidic rotation. This structure is also closed by a non-canonical base pair between C_{32} and A_{38}; however, in a slightly different geometry than that seen in the UNCG motif. Contrast this highly ordered U-turn structure with the disordered conformation of the ASL when bound to argRS. The argRS-bound conformation reveals that the strongest identity determinant, C_{35}, is positioned outside of the loop where it stacks with tryptophan 569 of the protein. It is coordinated in this position by two hydrogen bonds with the backbone atoms of histidine 22 and 23. Also, A_{38} is exposed and is buried within the protein, resulting in C_{32} forming a cross-loop stack with A_{37}.

3.5.1 Modifications modulate loop flexibility

There is a common trend with ASL modifications wherein the ASL tends to be stabilized

and less flexible upon modification [27]. This function is not surprising considering a pre-structured and highly stabilized U-turn conformation is beneficial to codon binding by reducing the amount of induced-fit required prior to codon binding [10]. A logical inference can be made for situations in which the ASL is not pre-structured by modifications. In these cases, it should be beneficial for the ASL to be more flexible to reduce the energy required for deforming the loop to adopt a canonical U-turn conformation. However, the preference for flexibility of the ASL will be dependent on the conformation that the ASL is currently adopting. For instance, while ASL^{Arg1}_{ICG} is in the UNCG motif, a higher plasticity for certain residues would be more beneficial to decoding, because of the need for the loop to deform during a conformational transition to the U-turn motif.

Due to significant stabilization of the loop residues while in the UNCG conformation caused by extensive intraloop hydrogen bonding [25], most of the loop residues showed very low B-factors with the exception of the solvent exposed A_{37} (Figure 3b). When either s^2C_{32} or m^2A_{37} were present without I_{34}, neither modification caused a significant alteration in the motional dynamics of the loop. There was a slight trend in which m^2A_{37} increased its own flexibility and s^2C_{32} reduced the stability of the terminal base pair in the presence of either A_{34} or I_{34}. The 2-thio modification also has a general property of stabilizing the residue at position 37 which is more pronounced in the presence of I_{34}. While it is unclear how the s^2C_{32} modification destabilizes the terminal base pair or stabilizes position 37, this corroborates thermal melting data suggesting that s^2C_{32} causes a significant reduction in the melt temperature of the ASL (ΔTm = -4.8°C) which is less extreme in the presence of I_{34} (ΔTm = -3.7°C) [25]. The inherent flexibility of m^2A/A_{37} may play a role in altering the conformational pathway between the solution and U-turn structure, implicating both s^2C_{32} and m^2A_{37} as modulators of conformational transitions.

The very stable base stacking network of the canonical U-turn motif is responsible for very low B-factors for the loop residues (Figure 4a). Interestingly, a certain amount of conformational change would be expected for the ASL to adopt the correct geometry for each wobble base pair, but not be too flexible that it destabilizes the stacking network of the

anticodon [23]. Certain modification states (unmodified, singly modified with s^2C_{32} and fully modified with s^2C_{32}, I_{34} and m^2A_{37}) showed increased motion at position 34 and, indeed, in the other anticodon residue G_{36}. Also, both the unmodified and fully modified ASLs were more flexible at positions 35 and 37.

The loop residues of the argRS-bound conformation were much more flexible in most positions (Figure 4b). In fact, even the stem residues were slightly more flexible than the solution structure. Also, A_{37} showed much less plasticity than seen in the solution structure. The positioning of A_{37}, stacking across the loop between G_{31} and C_{32} is a characteristic property of the argRS-bound structure [22] and may be important for the ability of the ASL to be induced into this conformation. The much larger amount of flexibility seen in the anticodon residues may be the result of the distorted loop geometry.

3.5.2 Modifications alter intermediates seen during conformational changes

A previous study from our lab described the ability a singly modified ASL^{Arg1} with I_{34} to bind all three synonymous codons (CGU, CGC and CGA) whereas the addition of either s^2C_{32} or m^2A_{37} resulted in negation of binding to the less common CGA codon [25]. It was hypothesized that since the singly modified ASL^{Arg1}_{ICG} adopted a highly canonical U-turn conformation when bound in the ribosomal A-site [23], the deformability of the loop would have a significant impact on the energy required for decoding. A low deformability of the loop, caused by either s^2C_{32} or m^2A_{37} in addition to I_{34} would thus increase the energy of decoding by making the loop residues more rigid. This would explain a lower maximum amount of binding to all codons in the doubly modified state. Additionally, the increased energy barrier inherent to the I_{34}•A_1 base pair [28] may be enough to completely abrogate decoding of the rare CGA codon. It was, however, unclear whether modifications would alter how the loop is deformed during an induced fit given that they have no significant effect on the equilibrium solution structure. To investigate whether the modification state of the ASL influences the conformational pathway, targeted molecular dynamics simulations were used to bias the ASL into converting from the solution structure to a canonical U-turn

conformation.

Despite the results of the binding experiments, the simulations suggest that C_{35} and G_{36} have difficulty switching from a *syn* to an *anti*-conformation (Table 1). A reduced motion around the χ angle would inhibit the ability of these ASLs to adopt the canonical U-turn structure and, consequently, their ability to bind to codons in the ribosomal A-site. The adoption of a fully canonical U-turn conformation could be completed through interactions within the ribosomal A-site. Therefore, we hypothesize that an interaction with the codon would enable decoding, albeit at a much lower level due to the increased energy of the transition. This hypothesis fits well with the results of binding experiments which suggest that binding to the cognate codon is reduced in I_{34}-containing constructs, perhaps as the result of both a lower base pairing enthalpy and the requirement for additional restructuring of the *N*-glycosidic orientations of two of the anticodon residues. Additionally, this energy requirement leads us to predict that the adoption of a canonical U-turn motif should be slightly less favorable for the doubly modified ASLs. The energy requirement for deformation of the loop in addition to the energy requirement for the purine•purine I_{34}•A_3 base pair [28] would lead to complete abrogation of doubly modified ASL binding to the rare CGA codon.

Many of the differences seen in the intermediates of the different conformational pathways appear to be caused by differences in the initial movements of m^2A/A_{37} relative to the other loop residues. Although only subtle differences are seen in the standard MD-derived B-factors of the loop residues in the solution structure (Figure 3b), correlations are present. In the unmodified construct, which follows Pathway 2 60% of the time, A_{37} and A_{38} have less conformational space, suggesting that this state of modification may not allow A_{37} to easily stack with G_{36} during the initial stages of the conformational change. Indeed, the allowance for this initial intermediate in the singly modified constructs may cause the ASLs to get trapped in a free energy well, making it more difficult for the anticodon residues to adopt the correct *N*-glycosidic torsion angles. These results predict that methods of stabilizing the syn glycosidic rotation seen in the solution structure such as introduction of 6-methylcytidine at

position 35 should inhibit the ability of the ASL to bind to synonymous codons [29]. Similarly, the use of locked nucleic acid chemistry to alter the preference for G_{36} to either C3'-*endo* or C2'-*endo* [30] should alter the preferential conformation of the ASL to either the U-turn or the solution and argRS-bound structures, respectively. For conversion to the canonical U-turn, preference for a C3'-endo conformation should reduce the energy barrier associated with the transition to the outside of the loop, easing the *syn* to *anti* rotation.

It should be noted, however, that the energy penalty introduced to the simulations is not constant. As the target RMSD is reduced linearly with time, the simulation has the ability to get "stuck" in energy wells until the energy of the conformational change is surpassed by the energy penalty. However, the linear decrease in the target RMSD gives the simulation time to catch up. This does not allow for an unbiased calculation of the amount of time that is required for a specific conformational change. The amount of time required for a conformational change would allow for a more quantitative measure of the favorability of a conformational change. This type of simulation requires the use of a non-linear target RMSD, which would also cause a non-constant energy penalty throughout the simulation causing more unfavorable conformational changes to be allowed near the beginning of the simulation, but not at the end. While these simulations will be attempted in addition to TMD on the remaining three ASL constructs, there are caveats with both systems and careful consideration to the results of both should be made.

3.5.3 Conclusions

After thorough investigation of the structural, biophysical and functional consequences of ASL^{Arg1}_{ICG} modifications, their purpose and mechanisms had not been fully elucidated [25]. Here, we have shown computational evidence that these small modifications can modulate the residue-specific flexibility of each ASL construct. It is also the case that each necessary structure that the ASL must adopt has characteristic modification-dependent effects on structural dynamics. Additionally, the distribution of the five conserved conformational pathways that the ASL follows between the solution UNCG motif and the canonical U-turn

conformation is modification-dependent. Our results indicate that the role for modifications in this ASL to specifically regulate residue-specific conformational space and the energy landscape between different, necessary conformations.

3.6 Methods and materials

3.6.1 Initial structures and force field parameters

Standard explicit solvent molecular dynamics simulations were carried out using the AmberTools1.5 and AMBER11 program suites [26]. For simulations of ASLs using the solution structures as a starting point, structure coordinates for unmodified ASL^{Arg}_{ACG} (2KRP), ASL^{Arg}_{ACG}-s^2C_{32} (2KRW), ASL^{Arg}_{ICG} (2KRQ) and ASL^{Arg}_{ICG}-s^2C_{32} (2KRV) [25]. For all of these, the representative structure was used as the starting coordinates. Starting coordinates for the three remaining ASLs, ASL^{Arg}_{ACG}-m^2A_{37}, ASL^{Arg}_{ICG}-m^2A_{37} and ASL^{Arg}_{ICG}-s^2C_{32};m^2A_{37} were generated by editing the protein data bank files for ASL^{Arg}_{ACG} (2KRP), ASL^{Arg}_{ICG} (2KRQ) and ASL^{Arg}_{ICG}-s^2C_{32} (2KRV), respectively. Starting coordinates for the synthetase-bound and ribosome-bound structures were generated by deleting all but ASL^{Arg}_{ICG} and editing the corresponding residues of PDB entry 1F7U [22] and 1XNQ [23], respectively. Poor density of the terminal residues in the heptadecamer ASL results in the crystal structure being truncated to 11 residues. To account for this, the additional stem residues were added based on positioning of the atoms in the solution structure. Constructs were then neutralized with 16 sodium ions and solvated in an octahedral box of TIP3P water [31] with a 10.0 Å cutoff. The AMBER ff99SB force field, which uses the same nucleic acid parameters as Cornell et al [32,33], was used for the common nucleosides and force field parameters for the modified nucleosides were obtained from the Modified Nucleic Acid Parameters Database [34].

3.6.2 Standard molecular dynamics simulations

Minimization and equilibration were each conducted in two steps. During the first minimization, 1000 total steps of minimization were performed with 500 steps of steepest

descent minimization and a Particle Mesh Ewald (PME) implementation of constant volume periodic boundaries [35]. Also, the RNA was held fixed with a positional restraint of 500 kcal/mol*Å^2. In the second step of minimization, 15000 steps of minimization were performed with 1000 steps of steepest descent and the RNA was left unrestrained. The gradual heating of the system from 0 K to 300 K was carried out during the first step of equilibration for a total of 20 ps at a time step of 2 fs while holding the RNA fixed with a weak positional restraint of 10 kcal/mol*Å^2. This temperature was chosen because it is closer to physiological conditions for *E.coli* than that used for both NMR and crystallography, but not too high that it introduced significant artifacts during the equilibration. For the first equilibration the PME implementation was used for a constant volume periodic boundary and for the second equilibration and production simulations, the PME implementation was used for a constant pressure periodic boundary. The SHAKE algorithm was used to constrain all bonds involving hydrogen [36] and Langevin dynamics was used at a collision frequency of 1.0 ps^{-1} to control the temperature of the system. A second equilibration was used to relax the system for 100 ps under production conditions. Five independent production simulations were performed for a total of 6 ns (one simulation of 1 ns and a second simulation of 5 ns) using the final equilibration coordinates. All data analysis and statistical calculations were performed on the final 5 ns of each independent simulation to allow for 1 ns of further system equilibration. The use of Langevin dynamics adds the possibility of synchronization artifacts of the simulations; therefore, a random seed was used for all equilibration and production simulations [37,38]. For all steps of simulation, a non-bonded cutoff of 10.0 Å was used.

3.6.3 Targeted molecular dynamics simulations

All TMD simulations were performed with directionality starting from the fully equilibrated solution structures (after 120 ps of equilibration and 6 ns of production simulation) and ending with fully minimized ribosome-bound or synthetase-bound structures. The identity of the independent simulation that was used for TMD was determined based on convergence of the standard simulations. For TMD simulations, the total number and identity

of atoms in the initial structure and the reference structure must match exactly. While the initial structure will contain the same amount of atoms in the ASL and the same number of Na^+ ions, the number of waters will be different because the difference in conformation will require a different radius of the TIP3P octahedral water box. To account for this, a truncated octahedral box of TIP3P water was used with a radius that gave a slightly larger number of water molecules than was necessary. The unnecessary water molecules were then deleted prior to producing the input coordinates. VMD was used to determine the starting RMSD and rounded down to the nearest 0.5 Å for the starting target RMSD. Simulation times were variable such that the target RMSD was reduced linearly at a rate of 0.5 Å/ns until reaching 0.0 Å followed by a full 1 ns of simulation at 0.0 Å using a force constant of 1.0 kcal/mol* $Å^2$. Simulations were performed at a contant 300 K and 1 atm with a 2 fs time step using the SHAKE algorithm [36] using random seeds. Trajectories were analyzed using ptraj [26], visualized using VMD [39] and PyMol [40] and plotted using Grace.

3.7 Acknowledgements

The authors would like to acknowledge the contributions of Dr. F.A.P. Vendeix, Dr. A. Chen and Dr. J. Schipper for their technical expertise and helpful discussions. Also, E. Warnke and A. Shelton of the University at Albany Research IT Department for their technical assistance and software troubleshooting.

3.8 References

1. Agris PF: Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* 2008, **9**(7):629-635.
2. Gustilo EM, Vendeix FA, Agris PF: tRNA's modifications bring order to gene expression. *Curr Opinion in Microbiol* 2008, **11**(2):134-140.
3. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FA, Fabris D, Agris PF: The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 2011, **39**(Database issue):D195-201.
4. Agris PF: The importance of being modified: roles of modified nucleosides and

Mg2+ in RNA structure and function. *Prog Nucleic Acid Res Mol Biol* 1996, **53**:79-129.

5. Yokoyama S, Nishimura S: Modified nucleosides and codon recognition. In: *tRNA: Structure, biosynthesis and function.* Edited by Schimmel PR, Söll D, RajBhandary UL. Washington D.C.: ASM Press; 1995: 207-223.

6. Brierley I, Meredith MR, Bloys AJ, Hagervall TG: Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J Mol Biol* 1997, **270**(3):360-373.

7. Urbonavicius J, Qian Q, Durand JM, Hagervall TG, Björk GR: Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J* 2001, **20**(17):4863-4873.

8. Phelps SS, Malkiewicz A, Agris PF, Joseph S: Modified nucleotides in $tRNA^{Lys}$ and $tRNA^{Val}$ are important for translocation. *J Mol Biol* 2004, **338**(3):439-444.

9. Denmon AP, Wang J, Nikonowicz EP: Conformation effects of base modification on the anticodon stem-loop of Bacillus subtilis tRNA(Tyr). *J Mol Biol* 2011, **412**(2):285-303.

10. Vendeix FA, Murphy FV, Cantara WA, Leszczynska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF: Human tRNA(Lys3)(UUU) Is Pre-Structured by Natural Modifications for Cognate and Wobble Codon Binding through Keto-Enol Tautomerism. *J Mol Biol* 2011.

11. Vendeix FAP, Dziergowska A, Gustilo EM, Graham WD, Sproat B, Malkiewicz A, Agris PF: Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry* 2008, **47**(23):6117-6129.

12. Nakamura Y, Gojobori T, Ikemura T: Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 2000, **28**(1):292.

13. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J: tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 2009, **37**(Database issue):D159-162.

14. Murao K, Tanabe T, Ishii F, Namiki M, Nishimura S: Primary sequence of arginine transfer RNA from Escherichia coli. *Biochem Biophys Res Commun* 1972, **47**(6):1332-1337.

15. Chakraburtty K: Primary structure of tRNA Arg II of E. coli B. *Nucleic Acids Res* 1975, **2**(10):1787-1792.

16. Crick FH: Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* 1966, **19**(2):548-555.

17. Olejniczak M, Uhlenbeck OC: tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 2006, **88**(8):943-950.

18. Agris PF, Sierzputowska-Gracz H, Smith W, Malkiewicz A, Sochacka E, Nawrot B: Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J Amer Chem Soc* 1992, **114**(7):2652-2656.

19. Kawai G, Yamamoto Y, Kamimura T, Masegi T, Sekine M, Hata T, Iimori T, Watanabe T, Miyazawa T, Yokoyama S: Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 1992, **31**(4):1040-1046.

20. Smith WS, Sierzputowska-Gracz H, Sochacka E, Malkiewicz A, Agris PF: Chemistry and structure of modified uridine dinucleosides are determined by thiolation. *J Amer Chem Soc* 1992, **114**(21):7989-7997.

21. Bilbille Y, Vendeix FAP, Guenther R, Malkiewicz A, Ariza X, Vilarrasa J, Agris PF: The structure of the human $tRNA^{Lys3}$ anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs. *Nucleic Acids Res* 2009, **37**(10):3342-3353.

22. Delagoutte B, Moras D, Cavarelli J: tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J* 2000, **19**(21):5599-5610.

23. Murphy FV, Ramakrishnan V: Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 2004, **11**(12):1251-1252.

24. Shimada A, Nureki O, Goto M, Takahashi S, Yokoyama S: Structural and mutational studies of the recognition of the arginine tRNA-specific major identity element, A20, by arginyl-tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(24):13537-13542.

25. Cantara WA, Bilbille Y, Kim J, Kaiser R, Leszczynska G, Malkiewicz A, Agris PF: Modifications Modulate Anticodon Loop Dynamics and Codon Recognition of E. coli tRNA(Arg1,2). *J Mol Biol* 2012.

26. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM *et al*: Amber 11. In.; 2011.

27. Stuart JW, Koshlap KM, Guenther R, Agris PF: Naturally-occurring modification restricts the anticodon domain conformational space of $tRNA^{Phe}$. *J Mol Biol* 2003, **334**(5):901-918.

28. Vendeix FAP, Munoz AM, Agris PF: Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA* 2009, **15**(12):2278-2287.

29. Schweizer MP, Banta EB, Witkowski JT, Robins RK: Determination of pyrimidine nucleoside syn, anti conformational preference in solution by proton and carbon-13 nuclear magnetic resonance. *Journal of the American Chemical Society* 1973, **95**(11):3770-3778.

30. Petersen M, Nielsen CB, Nielsen KE, Jensen GA, Bondensgaard K, Singh SK, Rajwanshi VK, Koshkin AA, Dahl BM, Wengel J *et al*: The conformations of locked nucleic acids (LNA). *Journal of molecular recognition : JMR* 2000, **13**(1):44-53.

31. Jorgensen WL, Chandrasekhar J, Madura JD: Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983, **79**:926-935.

32. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell Jw, Kollman PA: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995, **117**:5179-5197.

33. Cheatham TE, 3rd, Young MA: Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* 2000, **56**(4):232-256.

34. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel H, SantaLucia J: AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Comp Theory Comput* 2007, **3**:1464-1475.

35. York DM, Darden TA, Pedersen LG: The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. *J Chem Phys* 1993, **99**:8345-8348.

36. Ryckaert JP, Ciccotti G, Berendsen HJC: Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Phys* 1977, **23**:327-341.

37. Uberuaga BP, Anghel M, Voter AF: Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J Chem Phys* 2004, **120**:6363-6374.

38. Sindhikara DJ, Kim S, Voter AF, Roitberg AE: Bad seeds sprout perilous dynamics: Stochastic thermostat induced trajectory synchronization in biomolecules. *J Chem Theory Comput* 2009, **5**:1624-1631.

39. Humphrey W, Dalke A, Schulten K: VMD: visual molecular dynamics. *J Mol Graph* 1996, **14**(1):33-38, 27-38.

40. DeLano WL: Pymol. In. San Carlos, California, USA: Delano Scientific; 2006.

3.9 Tables and figures

Table 1. Characteristics of the five conformational pathways determined by TMD.

Pathway	Conformational Steps	Final χ Orientations
1	*G_{36} & A_{37} flip together outside of the loop to stack with A_{38} and G_{36} switches from *syn* to *anti*. *C_{35} flips from the inside to the outside of the loop forming a stable 3' stacking network.	C_{35} = *anti* G_{36} = *anti*
2	*A_{37} stacks between A_{38} and C_{39} forming a C_{32}•A_{37} Watson-Crick•Hoogsteen base pair. *G_{36} flips outside of the loop, flips from *syn* to *anti* and stacks with A_{38}. *C_{35} flips from the inside to the outside of the loop forming a stable 3' stacking network. *A_{37} flips down and stacks between A_{38} and G_{36}.	C_{35} = *anti* G_{36} = *anti*
3	*G_{36} & A_{37} flip together outside of the loop to stack with A_{38} and G_{36} switches from *syn* to *anti*. *C_{35} flips from the inside to the outside of the loop forming a stable 3' stacking network.	C_{35} = *syn* G_{36} = *anti*
4	*G_{36} & A_{37} flip together outside of the loop to stack with A_{38} and G_{36}. *C_{35} flips from the inside to the outside of the loop forming a stable 3' stacking network. *After C_{35} flips from the inside to the outside of the loop, G_{36} flips from *syn* to *anti*.	C_{35} = *syn* G_{36} = *anti*
5	*G_{36} & A_{37} flip together outside of the loop to stack with A_{38} and G_{36}. * C_{35} flips from the inside to the outside of the loop forming a stable 3' stacking network.	C_{35} = *syn* G_{36} = *syn*

Table 2. Distribution of conformational pathways for each ASL construct.

Pathway	ACG	ACGm2A	ACGs2C	ICG
1	40%	40%	---	---
2	60%	---	---	---
3	---	20%	40%	40%
4	---	20%	20%	---
5	---	20%	40%	60%

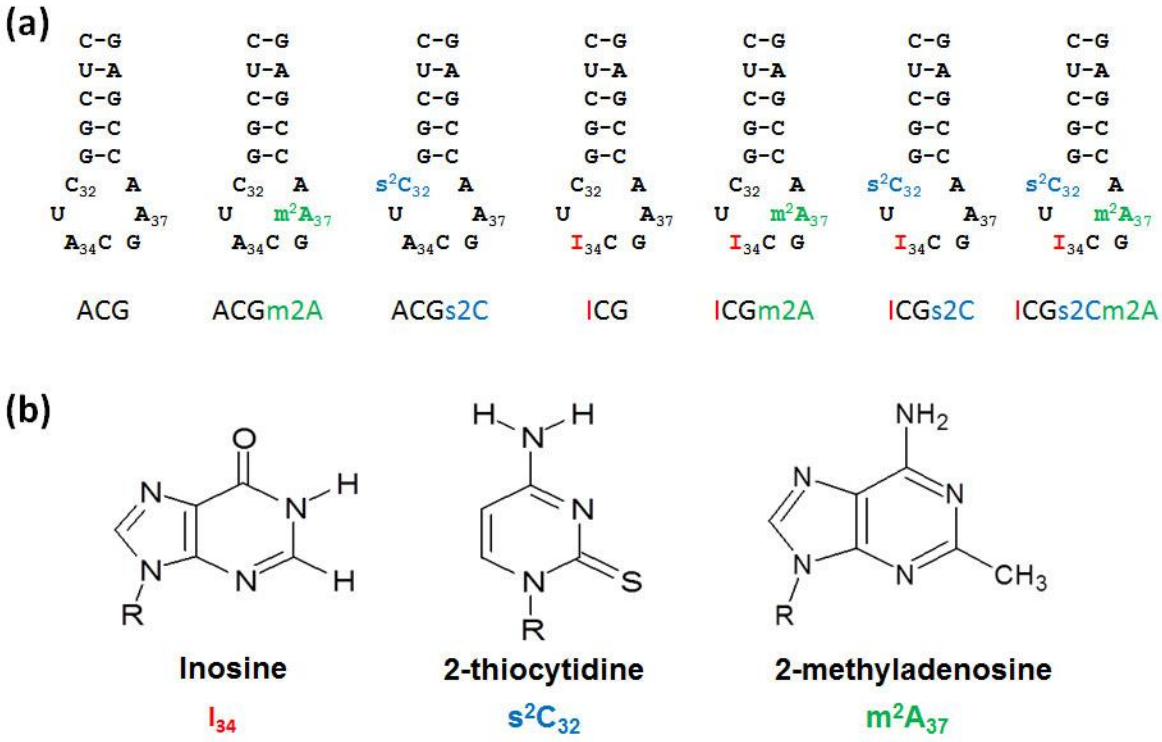

Figure 1. ASL constructs and nucleoside modifications. (a) Seven differentially modified ASL constructs were used for molecular dynamics simulations. The naming conventions below each ASL are used in other figures and text and denote the three anticodon residues followed by position 32 and 37 modifications 2-thiocytidine (s2C) and 2-methyladenosine (m2A). **(b)** The modified nucleosides present in these ASLs are all small modifications. Inosine is a deaminated adenosine in which the position 6 amine is converted to a carbonyl. In 2-thiocytidine, the carbonyl at position 2 is converted to a thionyl. A methyl group is added to the number 2 carbon of adenosine to create 2-methyladenosine.

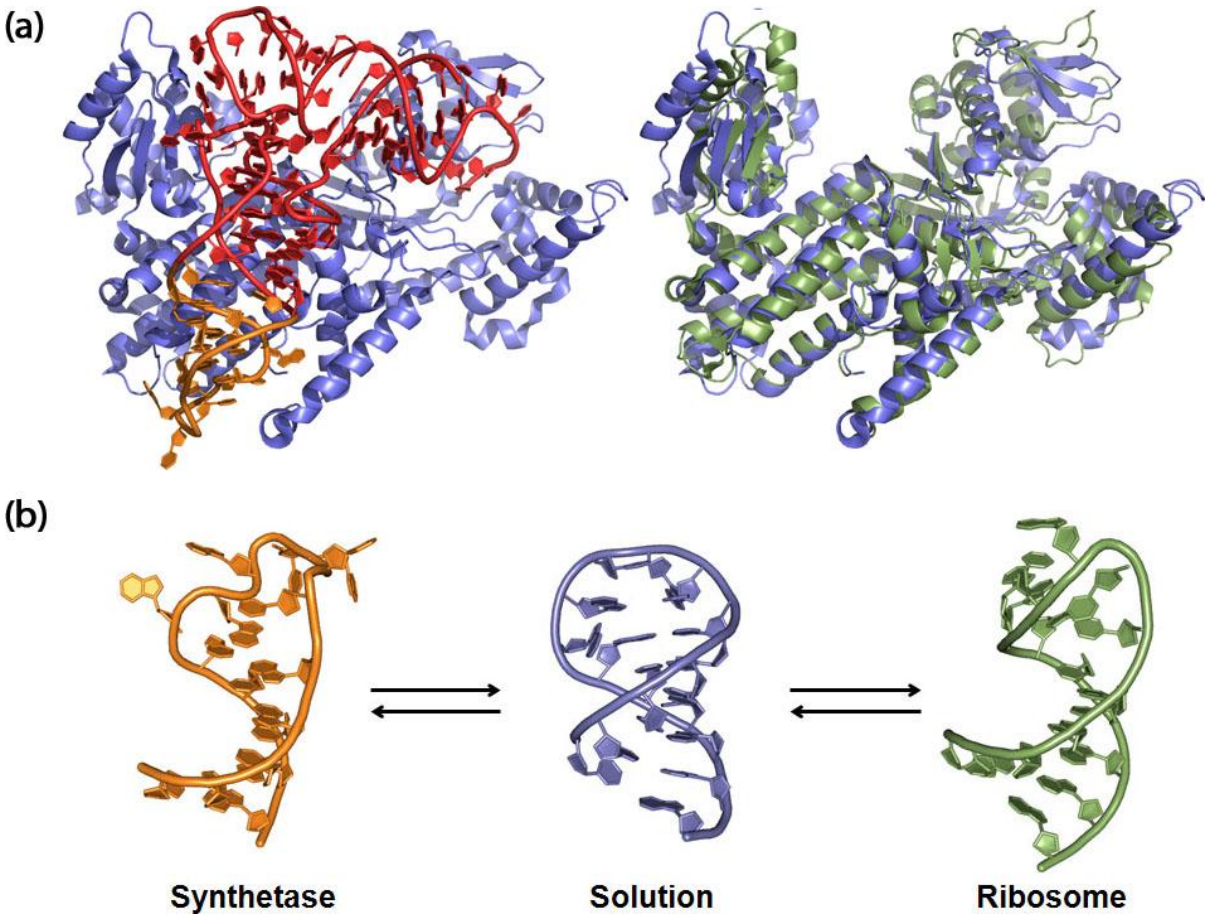

Figure 2. Structural studies have identified three different conformations that must exist within the cell. (a) The crystal structures of the yeast arginyl-tRNA synthetase (blue) with $tRNA^{Arg}_{ICG}$ (red) showing the location of the ASL (orange) on both structures (left) and superimposed with the bacterial synthetase (green) showing homology in the ASL binding region (right). **(b)** An equilibrium must be present between the structure in solution (blue, middle), the canonical U-turn conformation when bound in the ribosomal A-site (green, right) and distorted structure seen when bound to the yeast aminoacyl-tRNA synthetase (orange, left). To achieve proper function, the ASL must be able to switch between the three conformations in a reversible manner.

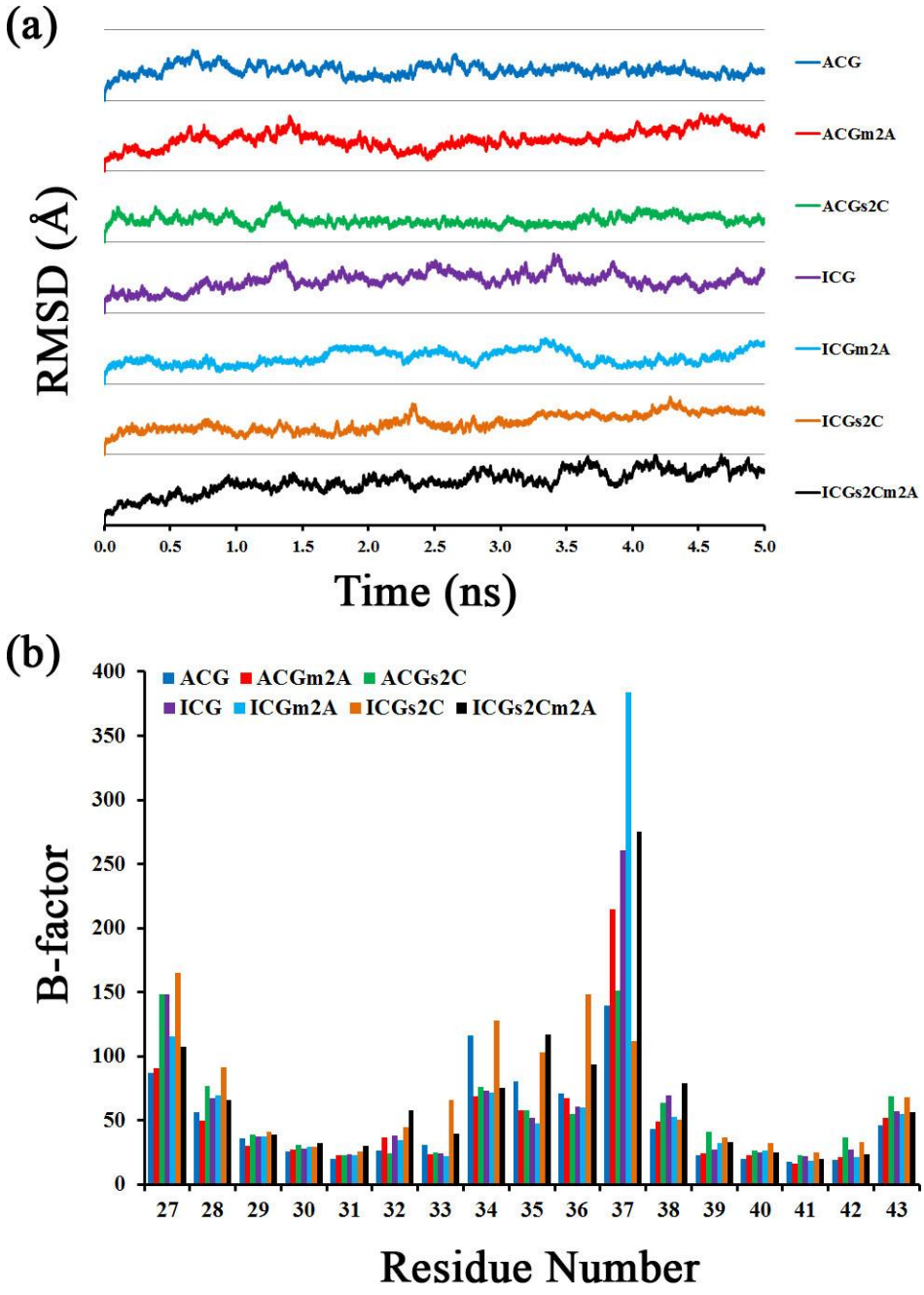

Figure 3. Standard MD simulations of the ASL^{Arg1}_{ICG} solution structure. (a) Representative heavy atom RMSD *vs* time plots over 5ns showed only slight variations. **(b)** Per residue B-factor calculations allowed for a determination of the residue-specific flexibility in the loop. B-factors were calculated based on root mean square fluctuations using the ptraj module from AmberTools1.5. The average of five replicates is plotted.

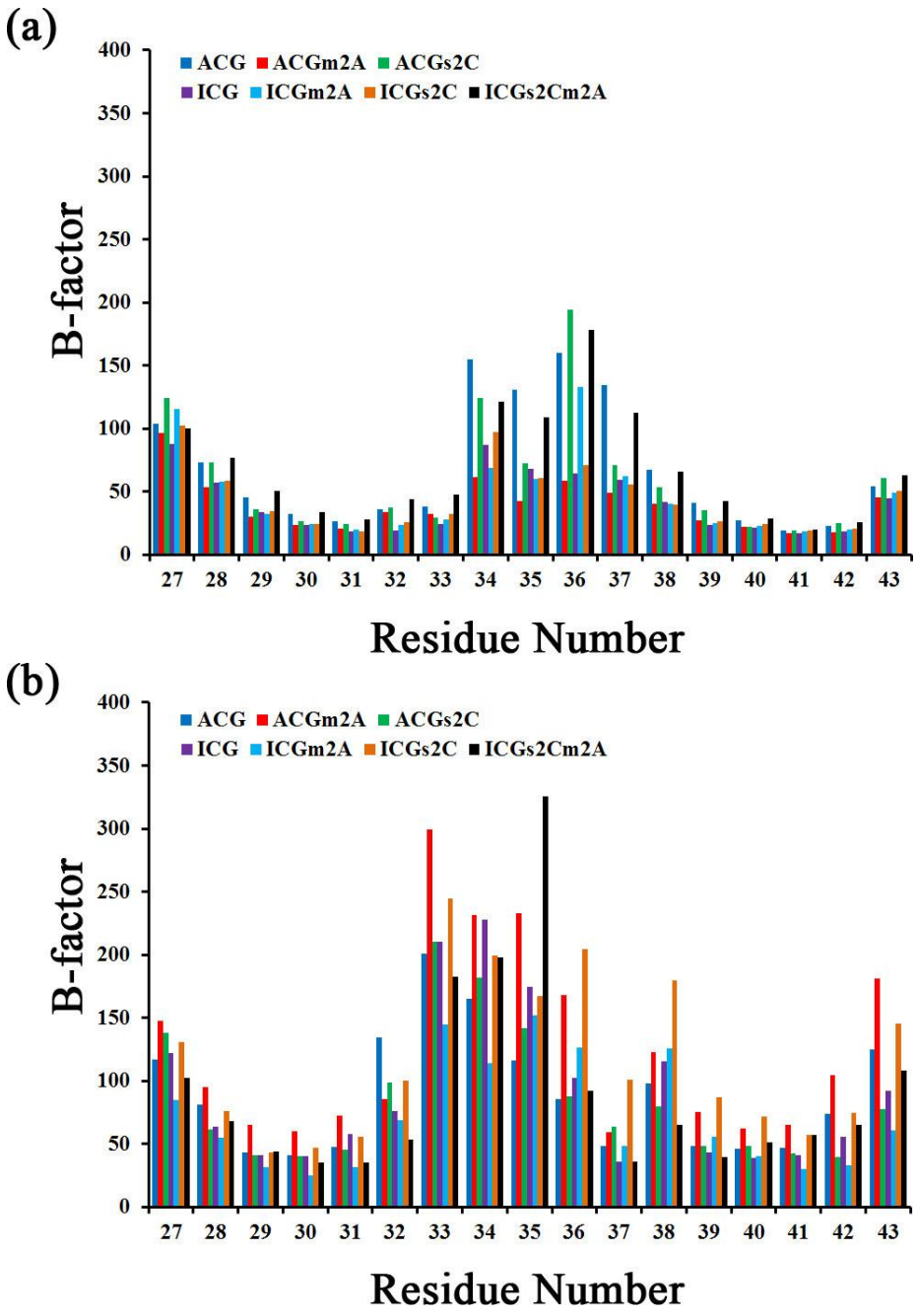

Figure 4. Standard MD simulations of the ASL$^{Arg1}_{ICG}$ ribosome-bound and synthetase-bound structures. The per residue B-factors were calculated from 5ns of standard molecular dynamics simulations using the **(a)** ribosome-bound U-turn and **(b)** distorted, synthetase-bound conformations as starting structures for determination of residue-specific conformational space. B-factors were calculated based on root mean square fluctuations using the ptraj module from AmberTools1.5. The average of five replicates is plotted.

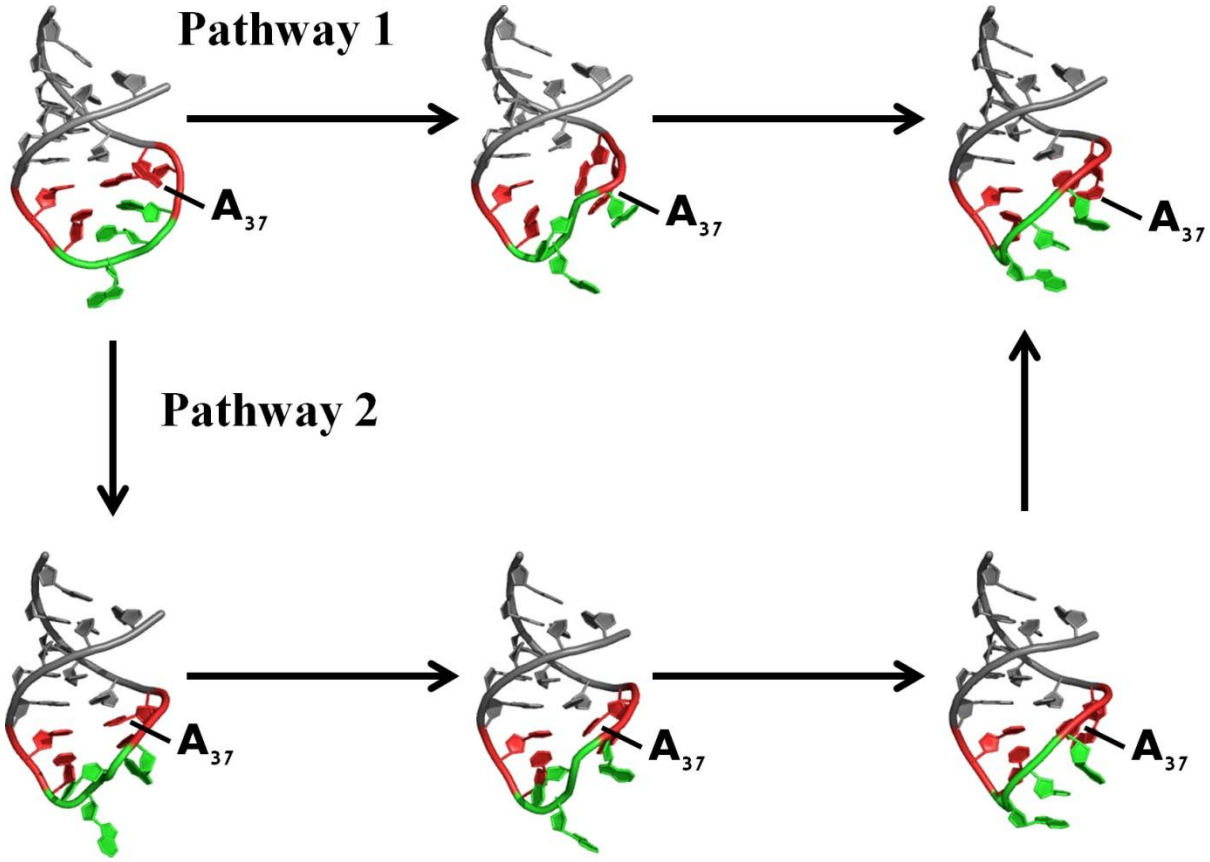

Figure 5. Conformational pathways resulting in a canonical U-turn structure. Pathway 1 is characterized by A_{37} flipping from a solvent-exposed position to stacking between G_{36} and A_{38} as G_{36} flips from the inside to the outside of the loop and into the anti glycosidic conformation. In **Pathway 2**, A_{37} takes the unusual position of stacking between A_{38} and C_{39}, displacing the C_{32}•A_{38} non-canonical base pair, forming a Watson-Crick•Hoogsteen C_{32}•A_{37} base pair. Next, G_{36} has room to flip outside of the loop and into the anti glycosidic rotation, followed by C_{35} flipping outside of the loop and stacking with A_{34} and G_{36}. Finally, the C_{32}•A_{37} base pair is destabilized by the flipping of C_{35} causing A_{37} to stack between G_{36} and A_{38} and resulting in the reformation of the C_{32}•A_{38} base pair.

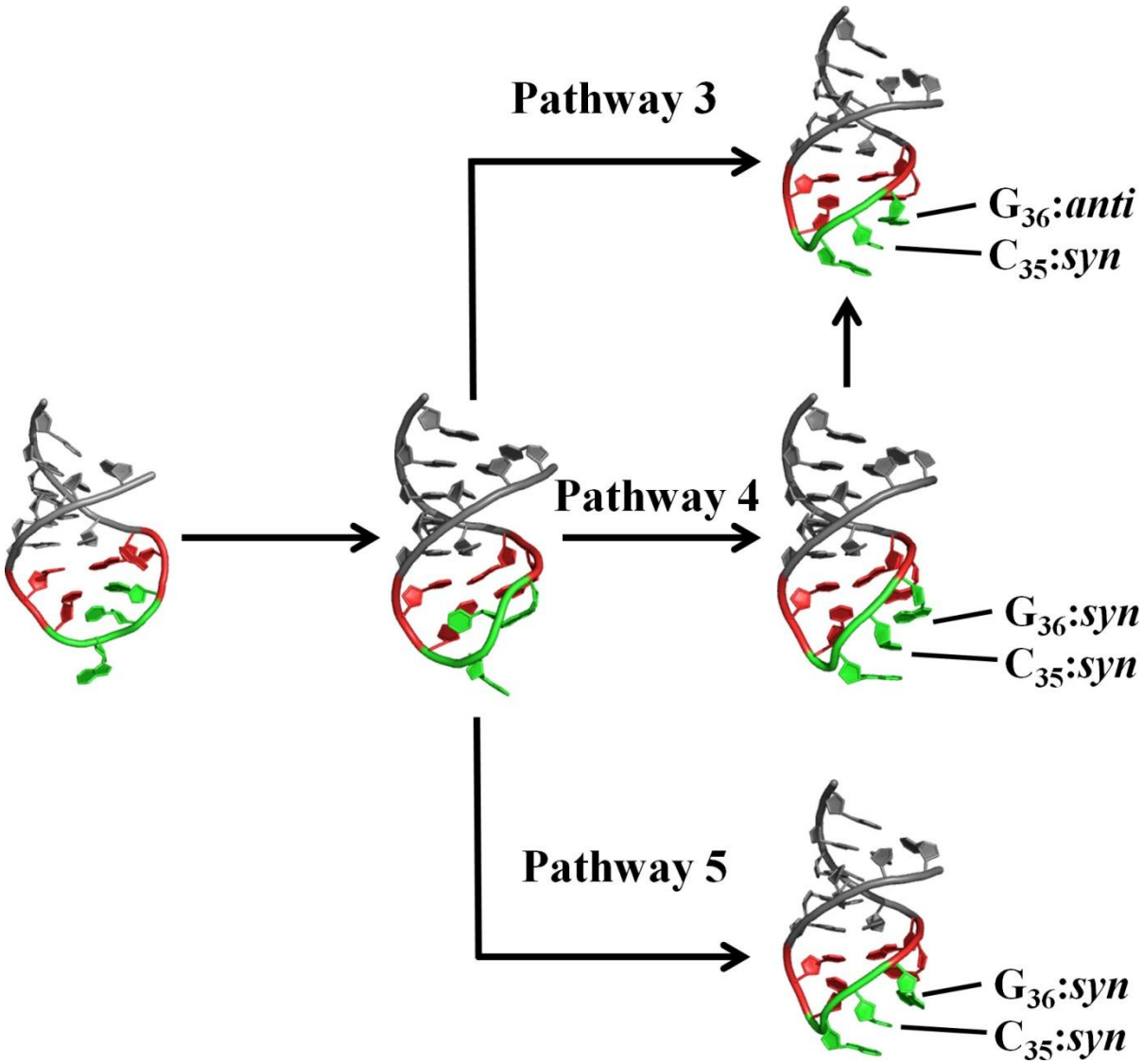

Figure 6. Conformational pathways not resulting in a canonical U-turn structure. In all three pathways, the first step occurs with A_{37} flipping between G_{36} and A_{38} while G_{36} flips from the inside to the outside of the loop and adopts a *syn* glycosidic conformation. In **Pathway 3**, G_{36} flips from *syn* to *anti*-conformation when C_{35} flips outside of the loop and stacks with A/I_{34} and G_{36}. **Pathway 4** is characterized by a more sequential change in which C_{35} flips outside the loop prior to G_{36} rotating from *syn* to *anti*. In **Pathway 5**, C_{35} moves from inside of the loop to outside of the loop as in Pathway 4; however, G_{36} never adopts the *anti*-conformation.

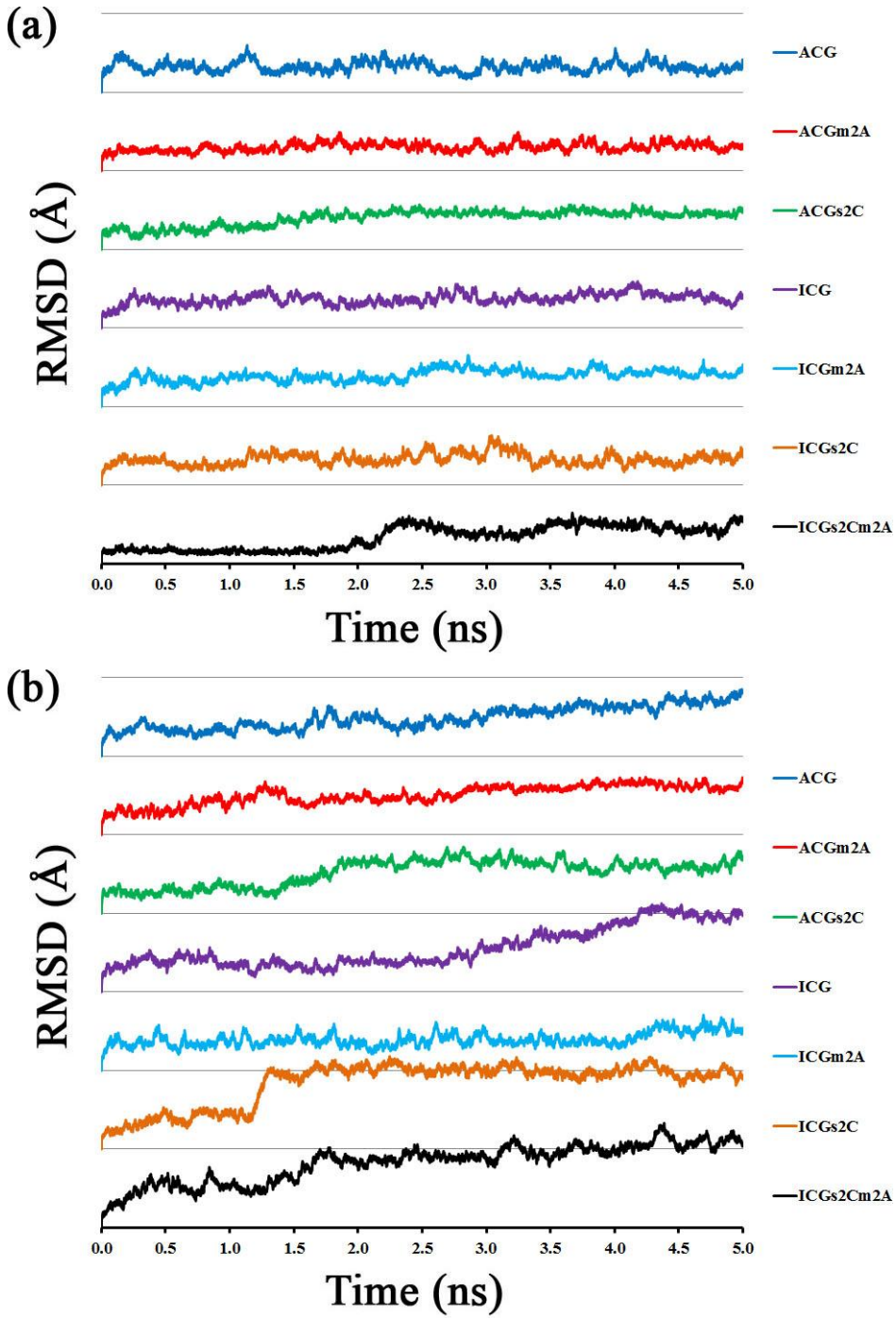

Figure S1. Standard MD simulations of the ASL^{Arg1}_{ICG} ribosome-bound and argRS-bound structures. Representative heavy atom RMSD *vs* time plots over 5ns showed only slight variations in standard MD simulations using starting structures from both **(a)** the ribosome-bound U-turn conformation and **(b)** argRS-bound distorted conformation.

CHAPTER 4. Structure and function effects of $tRNA^{Arg4}_{UCU}$ anticodon domain modifications

To be submitted once completed with the following authorship:

William A. Cantara, Kun Lu , Minhal Makshood, Erick Harr, Rob Kaiser, Paul F. Agris

4.1 Statement of project contribution

I, along with Dr. Agris, acted in a supervisory role for this project. I personally performed the thermal denaturation, circular dichroism and ribosome binding experiments. I also assisted with the NMR experiments and spectral assignments by assigning the ribose resonances. I will also be performing the structure calculations. I had the help of Dr. Kun Lu, who performed all of the NMR experiments and assigned all the aromatic protons of the RNAs. Minhal Makshood and Erick Harr, two undergraduates assisted with the circular dichroism and thermal denaturation experiments, respectively. I acted in an advisory and mentoring role for both undergraduates and coordinated experiments with Dr. Lu.

Abbreviations: ASL, anticodon stem and loop; CD, circular dichroism spectroscopy; COSY, COrrelation SpectroscopY; HSQC, Heteronuclear Single Quantum Correlation; mnm^5U_{34}, 5-methylaminomethyladenosine in position 34; mRNA, messenger RNA; NMR, nuclear magnetic resonance; NOE, Nuclear Overhauser Effect; NOESY, Nuclear Overhauser Enhancement Spectroscopy; ppm, part per million; RMSD, root mean square deviation; s^2C_{32}, 2-thiocytidine in position 32; t^6A_{37}, 2-methyladenosine in position 37; TOCSY, TOtal Correlation Spectroscopy; Ψ_{40}, pseudouridine in position 40.

Keywords: RNA structure, RNA function, 2-thiocytidine, 5-methylaminomethyladenosine, N^6-threonylcarbamoyladenosine

4.2 Abstract

In *Escherichia coli*, $tRNA^{Arg4}_{UCU}$ is responsible for decoding the rare AGA codon. The anticodon stem and loop domain of this isoacceptor contains the three common modifications mnm^5U_{34}, t^6A_{37} and Ψ_{40} as well as the rare s^2C_{32}. Here unmodified ASL^{Arg4}_{UCU}, singly modified ASL^{Arg4}_{UCU}-s^2C_{32} and ASL^{Arg4}_{UCU}-mnm^5U_{34};t^6A_{37};Ψ_{40} were investigated using a combination of functional, biophysical and structural techniques to decipher the structural and functional characteristics that are imparted by these modifications. Codon-specific ribosome binding assays confirm that binding to the cognate AGA codon is dependent on modification. Although s^2C_{32} has an inhibitory effect on wobble decoding in $tRNA^{Arg1,2}_{ICG}$, it is not responsible for negating dual codon recognition of both AGA and AGG in $tRNA^{Arg4}_{UCU}$. Biophysically, these modifications decrease molecular stability and base stacking interactions at optimum growth temperature, but increase the overall melting temperature. Interestingly, s^2C_{32} elicits an extra low temperature melting transition that is characterized by a significant change in secondary structure. To gain a better understanding of the structural properties of these modifications, NMR studies were performed to compare the structure difference between modified and triply-modified $tRNA^{Arg4}$ ASL. Preliminary NMR results suggest that the hypermodified nucleosides introduce more stability to the structure and promote a rigid loop conformation, which is important for codon-anticodon interaction. The modifications also affect the formation of C_{32}•A_{38}^{+} at the base of the anticodon stem by affecting the pKa of A_{38}. Although a spectral characteristics of a canonical U-turn were not observed, several NOEs have been observed between mnm^5U_{34} and t_6A_{37}, which may indicate preorganization of the ASL for codon recognition, reducing the entropic cost of an induced fit in the ribosomal A-site. A full molecular dynamics-restrained NMR solution structure determination of all three ASLs is underway to unambiguously determine the roles of these modifications in the structure:function relationships of this ASL.

4.3 Introduction

Proper and efficient ribosome-mediated protein synthesis relies on the correct 3-dimensional structure and chemistry of tRNA. Studies aimed at detailing the role of tRNA structure and function in translational fidelity have led to the discovery of 93 different naturally occurring modifications present in tRNA [1]. While these modifications, in general, lead to increased order and stability for the molecule as a whole, the great number and diversity of modification chemistries in the anticodon stem and loop (ASL) leads to a variety of translational properties. Indeed, ASL modifications can act as identity determinants for aminoacylation [2], enhance stability [3,4], maintain the translational reading frame [5,6], regulate codon discrimination [7,8], increase synonymous codon affinity [9-11] and prestructure the ASL into a canonical U-turn for proper ribosomal A-site recognition [4].

Translational decoding of rare codons has implications for understanding both epigenetic regulation of gene expression and evolution of codon bias. A compelling trend exists between tRNA expression patters and codon bias in both *E.coli* and yeast [12]. This study revealed that in *E.coli* $tRNA^{Arg3}_{CCU}$ and $tRNA^{Arg4}_{UCU}$, which are responsible for decoding the rarest codons, AGG and AGA, are the most scarcely expressed tRNAs. A regulatory role is evident by the DNA synthesis and repair, protein synthesis, cell cycle regulation and metabolism genes that are enriched with these rare codons [13]. Indeed, genes enriched with these codons show reduced levels of translation and a high frequency of translational frameshifting events [14,15].

The *E.coli* $tRNA^{Arg4}_{UCU}$ ASL contains four endogenous posttranslational modifications, 2-thiocytidine at position 32 (s^2C_{32}), 5-methylaminomethyluridine at position 34 (mnm^5U_{34}), N^6-threonylcarbamoyladenosine at position 37 (t^6A_{37}) and pseudouridine at position 40 (Ψ_{40}) (Figure 1a) [16]. The rare s^2C_{32} modification is found in four of the five arginyl-tRNA isoacceptors and $tRNA^{Ser2}_{GCU}$ in *E.coli*, making it of particular interest [16]. The oxygen to sulfur replacement at position 2 of pyrimidines further stabilizes a C3'-*endo* sugar pucker in addition to an *anti N*-glycosidic conformation [17-20]. Also, the lower electronegativity of the sulfur atom results in better stackability of the base [21]. Interestingly, these properties

should stabilize the cross-loop C_{32}•A_{38} mismatch base pair that commonly closes the loop, resulting in a pseudo five-membered loop that is characteristic of a canonical U-turn conformation [22,23].

In addition to s^2C_{32}, ASL^{Arg4}_{UCU} contains two other loop sequence and modifications that closely match those seen in a hypomodified *E.coli* ASL^{Lys}_{UUU} (Figure 1b). In ASL^{Lys}_{UUU}, the mnm^5U_{34} and t^6A_{37} modifications confer the ability to read both A and G in the third codon position [10]; whereas, in ASL^{Arg4}_{UCU}, it only allows recognition of A [15]. It has been shown that the geometry of the C_{32}•A_{38} mismatch base pair regulates the ability of the ASL to discriminate between cognate and wobble codons [24]. Therefore, the presence of s^2C_{32} may alter the geometry of this pair, rendering the ASL incapable of decoding the other rare AGG codon in addition to AGA. Indeed, s^2C_{32} has been shown to have an inhibitory effect on wobble binding of ASL^{Arg1}_{ICG} to the rare, synonymous CGA codon [7]. ASL^{Lys}_{UUU} also has the uncommon property of forming an extended stem conformation resulting in a tri-nucleotide loop in the absence of modification [25]. In that case, the modifications mnm^5U_{34} and t^6A_{37} functioned to negate intraloop hydrogen bonding and stabilizing a prestructured U-turn-like conformation in solution [4,25,26].

To investigate the roles of posttranscriptional modifications in the structure and function of ASL^{Arg4}_{UCU}, thermodynamic, functional and structural characterizations were performed on synthesized unmodified (ASL^{Arg4}_{UCU}), singly modified (ASL^{Arg4}_{UCU}-s^2C_{32}) and triply modified (ASL^{Arg4}_{UCU}-mnm^5U_{34};t^6A_{37};Ψ_{40})(Figure 1c). Codon binding assays revealed that s^2C_{32} was not responsible for negating the ability of the ASL to wobble decode AGG codons; however, biophysical experiments suggest that it may play a role in modulating the thermal stability of certain secondary structure elements. Despite a lack of characteristic U-turn spectral identifiers, the remaining three modifications appear to stabilize the loop residues into a more ordered conformation, reducing the energetic resources required for adopting a canonical U-turn conformation for A-site ribosome binding.

4.4 Results

4.4.1 Lack of s^2C_{32} does not allow wobble decoding

Two key differences between the loops of ASL^{Arg4}_{UCU} and ASL^{Lys}_{UUU} are the modification of C_{32} to s^2C_{32} in ASL^{Arg4}_{UCU} and difference in the identity of the pyrimidines at position 35 of the anticodon (Figure 1a,b). To determine if, as seen in ASL^{Arg1}_{ICG}, the s^2C_{32} is responsible for the restriction of binding to AGA only, A-site codon specific ribosome binding assays were performed on three synthesized ASL constructs (unmodified ASL^{Arg4}_{UCU}, singly modified ASL^{Arg4}_{UCU}-s^2C_{32} and triply modified ASL^{Arg4}_{UCU}-mnm^5U_{34};t^6A_{37};Ψ_{40}; Figure 1c). Neither the unmodified nor the singly modified ASLs were able to bind to the AGA codon; however, the triply modified ASL bound AGA with a dissociation constant of 188 ± 21 nM (Figure 2a). Interestingly, none of the ASL constructs tested, including the triply modified ASL^{Arg4}_{UCU}-mnm^5U_{34};t^6A_{37};Ψ_{40}, were able to bind to the AGG codon efficiently (Figure 2b), a clear indication that s^2C_{32} is not responsible for negating wobble decoding of AGG. Dissociation constants were able to be determined for some of the other ASL constructs; however, the relatively low amount of binding suggests that these are not physiologically relevant binding events.

4.4.2 Modifications alter biophysical properties of the ASLs

Ribosomal A-site codon binding assays revealed that modifications are required for binding to the cognate AGA codon. To assess the biophysical characteristics caused by modification, circular dichroism and UV-monitored thermal denaturation spectroscopies were performed on all three ASL constructs. One of the hallmark characteristics of the canonical U-turn structure is the presence of a highly stacked and ordered loop; therefore, circular dichroism spectroscopy was used to measure the relative amount of stacking in each construct. All ASLs showed a peak around 270 nm indicating that they all form an A-form RNA stem; however, the triply modified ASL peak was slightly red-shifted suggesting that the modifications may be inhibiting a cross-loop hydrogen bond [27]. The unmodified ASL showed the largest peak (7.53 cm^2/mmol), corresponding to a high level of base stacking,

compared to the triply modified ASL (6.46 cm^2/mmol) at 5°C (Figure 2c). Interestingly, when heated up to 20°C, the ellipticity of both the unmodified and triply modified ASLs decreased by ~5%, whereas the ellipticity of the singly modified ASL dropped by 16.3% (Figure 2c,d). This indicates that, individually, the s^2C_{32} is forming base stacking interactions at low temperature that are easily broken with a small temperature increase.

Both the singly modified and triply modified ASLs showed both a reduction of ~1.5% hyperchromicity (Table 1, Figure 3), a measure of the overall molecular order. Correspondingly, the unmodified ASL showed a more negative enthalpy and free energy change, suggesting that the extra order may result in stabilization of the ASL. Interestingly, the higher melting temperature of triply modified ASL^{Arg4} indicates that the pseudouridine or other modifications are acting to inhibit full melting of the hairpin, independent from the structural order. This is a common property of pseudouridine in A-form RNA helices [4,28]. Although thermodynamic parameters were not calculated for the singly modified ASL, the melting curve reveals a low temperature transition that corresponds with the circular dichroism data (Figure 3). Overall, the biophysical data reveals a modification dependent modulation of the thermodynamic and base stacking properties of the ASL constructs, resulting in a reduction of overall molecular order and reduction in the base stacking of the loop.

4.4.3 NMR resonance assignments

Biophysical characterization suggests that the differences in codon binding seen in the ASLs are related to modification-induced alteration of loop conformation. Specifically, modifications may play a role in reorganizing the loop residues to allow for proper A-site recognition [4,29]. To understand whether structural differences exist and how they may influence codon discrimination, full nuclear magnetic resonance spectroscopy (NMR) characterizations were performed on the unmodified and triply modified ASL constructs. Commonly used protocols for establishment of nucleic acid resonance assignments in the absence of isotope labeling were applied [30-33]. Using a combination of ^1H-^1H homonuclear and natural abundance ^1H-^{13}C and ^1H-^{31}P heteronuclear NMR experiments,

sequence-specific assignment of nucleotide spin systems were mostly completed.

Exchangeable proton spectra were recorded in 90% H_2O/10% D_2O at three different temperatures (283K, 288K and 293K) to observe and identify the imino and amino resonances of the stem residues. The imino signal assignments are consistent with the formation of the predicted five base-pair double-stranded stem region in both unmodified and triply modified ASL^{Arg4} RNAs. Five well-resolved peaks were observed in the low field region (12.00-14.50ppm) of the 1D imino spectra of both unmodified and triply modified ASLs. These were identified as the base-paired stem residue protons $G_{27}H1$, $G_{29}H1$, $G_{39}H1$, $U/\Psi_{40}H3$ and $G_{42}H1$ (Figure 4). No additional imino protons were observed in both samples that would indicate the presence of alternate conformations typically seen in mixtures of hairpin and duplex. In both spectra, one initially broad peak that was significantly broadened by a temperature increase was assigned to $G_{27}H1$ as this is a characteristic feature of the terminal base pair which rapidly exchanges with solvent. Three additional peaks were observed in the triply modified ASL spectrum as a result of t^6A_{37} and Ψ_{40} (Figure 4b). The amide proton resonances at 9.1 and 10.1 ppm were assigned to be N6H and N11H of t^6A_{37} from 2D NOESY and TOCSY connectivities. The N11H resonance of the ureido of the threonyl moiety of t^6A_{37} was observed and has TOCSY connectivity to the Hα and Hβ resonances of the side chain, indicating that N11 acts as a hydrogen bond donor with N1 on the Watson-Crick face. The observation of t^6A_{37} N6H and N11H indicated that the remaining peak at ~10.3 ppm was assigned to N1H of Ψ_{40}. The N1H imino protons have been observed and assigned in a number of tRNAs [34-36], tRNA anticodon stem-loop [37], single-stranded RNA [38] and RNA duplexes [39,40]. To observe the Ψ_{40} N1H, the only structural requirement is that the local structure in the vicinity of Ψ allows for an RNA conformation with two phosphates in position for water coordination. Therefore, the observation of uridine N1H in the triply modified ASL^{Arg4} suggests that the Ψ modification provides additional stabilization to the RNA structure most likely through coordination of a water molecule and the imino proton exchange is sufficiently slowed to allow detection of the Ψ_{40} N1H NMR resonance. The imino and amino protons in the loop residues were not observed due to their rapidly exchanging nature.

The pH effect on the chemical shift of the unmodified ASL^{Arg4} exchangeable protons was characterized at three different pHs of 6.8, 6.0 and 5.5 at 288K (Figure 6a). No dramatic changes were observed for $G_{27}H1$, $G_{29}H1$, $U/\Psi_{40}H3$ and $G_{42}H1$, however, $G_{39}H1$ showed a significant downfield shift corresponding to the pH decrease, indicating the environment of the $G_{39}H1$ is sensitive to the pH change. The A_{38}, 5' base of G_{39}, could form an A^+-C base-pair with C_{32}. To investigate if an A_{38}^+-C_{32} base-pair forms in ASL^{Arg4} and the protonation of A_{38} affects the $G_{39}H1$ chemical shift, we collected the pH-dependent HSQC spectra (Figure 6b). ^1H-^{13}C HMQC spectra confirms formation of an A^+-C base-pair with protonation at N1 of A_{38} as evidenced by an unusual ^{13}C chemical shift of ~ 6-7 ppm for the C2 of A_{38} upon decreasing the pH to 5.5 [28,41]. At pH 6.8, $A_{38}C2$ is involved in exchanging between protonation and deprotonation and is completely broadened out, indicating the pKa of $A_{38}H1$ is around 6.8. The ^{13}C chemical shifts of the C_{32} C5 and C6 carbons remain unchanged upon lowering the pH (data not shown), consistent with protonation of A_{38} rather than C_{32}. Our NMR results confirmed the formation of the A^+-C base-pair and it stabilizes the unmodified ASL^{Arg4} structure. The high pKa of A_{38} affects loop structure even at neutral pH, suggesting that the A^+-C base-pair formation may be biologically relevant. Due to the risk of destabilizing the modified nucleotides under low pH, we haven't decreased the pH of the triply-modified ASL^{Arg4} to study the pH effect on its structure. However at neutral pH, a broadened but detectable resonance with an unusual ^{13}C chemical shift was observed for $A_{38}C2$-H2 in the ^1H-^{13}C HMQC spectra indicating an elevated pKa (data not shown).

A series of NOESY spectra recorded in 100% D_2O at various mixing times from 50 ms to 400 ms were collected for the assignment of non-exchangeable protons in the ribose and base ring. DQF-COSY were also collected to identify intranucleotide connectivities. These homonuclear experiments were used to assign resonances via the "sequential walk" method (Figure 5), starting from the 5'-terminus. A characteristic NOE walk was observed between $G_{27}H1'$-H8 and $C_{32}H1'$-H6 and from $U_{36}H1'$-H6 to $C_{43}H1'$-H6 for both unmodified and triply modified ASL^{Arg4}. The assignments of the H2' resonances were confirmed through observation of the strong cross-peaks to the H1' protons in the short mixing time NOESY experiment. The H3' protons were identified using the ^1H-^{13}C HSQC experiments. A nearly

complete assignment of each nucleotide's proton resonances was accomplished using the sequential connectivity of the H1'-H8/H6, H2'-H8 and H3'-H8 NOE cross peaks to assign the H1', H2' and H3' protons. The remaining anomeric protons (H4' and H5'/H5'') were assigned on the ^{1}H-^{13}C HSQC spectra by their distinctive ^{13}C chemical shift. Due to severe resonance overlaps, only a partial assignment of the H4' and H5'/H5'' protons was achieved.

From the assigned non-exchangeable protons, breaks in the sequence between the connectivity of C_{32} to C_{35} were observed in both ASLs. $A_{38}H2-U_{33}H1'$ was observed in the NOESY spectra of both ASLs indicating the extended hairpin formation in the loop region. The $G_{39}H8-G_{39}H1'$ peak is much broader for the unmodified ASL^{Arg4} when compared to the triply-modified, suggesting the modifications introduced more stabilization to the ASL structure. The characteristic connectivity of $U_{33}H1'-C_{35}H8$ NOE that is routinely associated with the canonical U-turn motif of tRNA anticodon was not unambiguously observed. The canonical conformation for the 33p34 phosphate is one where the α dihedral angle is trans [42], and this can be conveniently monitored by ^{31}P NMR because a downfield shift is seen for the phosphate in a U-turn structure [43,44]. However, no downfield shifted magnetic resonance was observed in the ^{31}P NMR spectra of both unmodified and triply modified ASL^{Arg4} (Figure 7).

A common feature of the U-turn conformation is the presence of a stable A-form stem and loop consisting exclusively of nucleosides exhibiting C3'-*endo* sugar pucker geometry. Therefore, the sugar pucker conformations of each residue were investigated using COSY, DQF-COSY and TOCSY (data not shown). Both the unmodified and triply modified ASLs showed similar characteristic ribose geometry properties. The five base-paired stem residues conformed to a C3'-*endo* conformation ($^{3}J_{H1'-H2'} < 2$) characteristic of a canonical A-form helix. Additionally, the loop residues U_{33}, U_{36}, A_{37} and A_{38} were in the C3'-*endo* geometry; however, U_{33}, U_{34} and C_{35} exhibited C2'-*endo* conformations ($^{3}J_{H1'-H2'} > 5$). To summarize, NMR spectroscopy was used to fully characterize the structural characteristics of both the unmodified and triply modified ASLs, neither of which showed characteristic spectral properties supporting a U-turn conformation in solution.

4.5 Discussion

4.5.1 C_{35} is a positive determinant for A/G-ending codon discrimination

Deciphering the roles of anticodon domain modifications in tRNA function is key to understanding how the genetic code is regulated and evolved. While some modifications are highly conserved for a specific purpose in many different ASLs, it has become evident that modifications are not always modular in the sense that many appear to function in a way that is dependent on the chemical environment created by the other loop residues [7,45]. A high conservation of t^6A_{37} modifications and derivatives thereof in tRNAs responsible for decoding 5'-ANN-3' codons [16] stems from the ability of this residue to form a pseudo-heterotricyclic ring that stacks on top of and stabilizes the low enthalpic U_{36}•A_1 base pair [4]. Contrastingly, however, we show that in the context of ASL^{Arg4}_{UCU}, mnm^5U_{34} allows for the recognition of only the A-ending AGA codon, whereas in ASL^{Lys}_{UUU}, the same modification further increases recognition to both the A-ending AAA and the G-ending AAG codons [10].

The ability of the ASL to modulate codon discrimination at the wobble position is linked both to the geometry the position 34 sugar pucker [20,46,47] and the stem proximal non-canonical interaction between position 32 and 38 [24]. Depending on whether the position 34 modification promotes a C2'-*endo* or C3'-*endo* sugar pucker, the modification tends to either restrict or expand codon recognition, respectively [20,46,47]. In both the unmodified ASL^{Lys}_{UUU} and ASL^{Lys}_{UUU}-t^6A_{37}, U_{34} adopts a C2'-*endo* conformation, rendering the ASL unable to decode AAG. Interestingly, the methylaminomethyl modification at the 5-position of uridine does not generally promote the C3'-*endo* geometry in a pyrimidine rich region [48]. The wild-type ASL^{Lys}_{UUU} contains a 2-thiolation of U_{34} in addition to the mnm^5 modification [16] which is the determinant that promotes the C3'-*endo* conformation of itself and in neighboring residues [19]. Although in ASL^{Lys}_{UUU}, the 5-position modification is sufficient to allow for recognition of both AAA and AAG, the crystal structure of this interaction shows a geometry for the mnm^5U_{34}•G_3 base pair requiring a rotation about the C1'-N1 bond that allows for a bifurcated hydrogen bond to form between N1 and N2 of the codon G_3 and O2 of mnm^5U_{34} [23]. Since this interaction is much less stable than the normal

G•U wobble base pair between that is characterized by two hydrogen bonds, the base pair is characterized by a stacking interaction caused by the oxygen of the 2-position carbonyl group of mnm^5U_{34} being positioned directly over the ring of U_{35} [23].

Codon specific A-site ribosome binding assays show a clear distinction between the unmodified and modified ASL in which only the triply modified ASL was able to bind AGA and none of our constructs could recognize AGG (Figure 2a,b). It is clear from previous studies that the fully modified $tRNA^{Arg4}_{UCU}$ can recognize only AGA [15]. Since our construct lacks the endogenous 2-thiolation of C_{32}, we conclude that s^2C_{32} is not responsible for negating the dual codon recognition seen in the similar ASL^{Lys}_{UUU}. The only remaining difference in the identities of the loops between triply modified ASL^{Arg4}_{UCU} and doubly modified ASL^{Lys}_{UUU} is the identity of the position 35 nucleoside. In ASL^{Lys}_{UUU}, the position 35 uridine is responsible for proper positioning of the modified mnm^5U_{34} [23]. Thus the chemical properties and geometry of C_{35} in ASL^{Arg4}_{UCU} may distort the positioning of mnm^5U_{34}, disallowing decoding of AGG. Interestingly, recognition of cognate codon AAA was evident for both unmodified ASL^{Lys}_{UUU} and singly modified ASL^{Lys}_{UUU}-t^6A_{37} [10], whereas unmodified ASL^{Arg4} showed no signs of AGA binding. This was particularly surprising considering that the C_{35}•G_2 base pair is enthalpically preferential to U_{35}•A_2. Therefore, we can conclude that there are either structural or chemical differences in the anticodon residues that cause a preference for cognate codon binding with a U_{35}.

4.5.2 Biophysical properties suggest a conserved function of loop modifications

Although it is clear that, in terms of wobble decoding, the modifications in ASL^{Arg4} function differently than they do in ASL^{Lys}; however, the decoding function appears to be linked to the chemical properties of C_{35} rather than the nature of the modifications. Indeed the thermodynamic characteristics upon modification are quite similar to those of the differentially modified ASL^{Lys}. Indeed, we see a decrease in the values of enthalpic, entropic and Gibbs free energy changes with respect to the unmodified ASL (Table 1) which directly mirrors the differences see in ASL^{Lys} upon modification [10]. Similar to ASL^{Lys}, there is a significant stabilization of the stem in the vicinity of the Ψ modification, which is most likely

caused by the effect of a coordinated water molecule. Also, we see a stabilization of the G_{39} that appears to be the result of a C_{32}•A_{38}^{+} base pair, which has precedent in human ASL^{Lys3} [4,28]. In the unmodified ASL, this base pair may be transient based on the observed pKa of ~6.8. Interestingly, the significant stabilization of G_{39} in the triply modified ASL suggests that the pKa of A_{38} may be shifted higher, allowing for a more stable base pair to form. Also, a slight red-shift seen in the circular dichroism spectra of the triply modified ASL suggests that there may be a difference in the hydrogen bonding properties of this base pair (Figure 3b,c). A slight pKa shift is confirmed by the presence of the unusual upfield shifted A_{38}C2-H2 resonance seen at pH 6.8 in the ^{1}H-^{13}C HSQC spectra of the triply modified ASL. Alternatively, the increased stacking that may occur with the modified, pseudo-heterotricyclic $t^{6}A_{37}$, A_{38} may be positioned in a more stable position for cross-loop base pairing in a pH-dependent manner. Indeed, both ASLs show an A_{38}H2−U_{33}H1' connectivity indicating the formation of an extended stem.

4.5.4 Modifications do not drive the ASL into a canonical U-turn conformation

The sugar conformations of both unmodified and triply modified ASL^{Arg4} were in disagreement with a U-turn conformation. In the canonical U-turn conformation, all stem and loop residues exhibit a C3'-*endo* sugar conformation. Both ASLs were determined to have predominantly C2'-*endo* sugar puckers for U_{33}, U_{34} and C_{35}. This slightly mirrors the trend seen for both unmodified ASL^{Lys}_{UUU} and singly modified ASL^{Lys}_{UUU}-$t^{6}A_{37}$, both of which were determine to exhibit C2'-*endo* conformation for U_{33}, U_{34} and U_{36} [26]. Here we see a difference in which the position 35 and position 36 sugar puckers between ASL^{Lys} and ASL^{Arg4} are reversed. In both cases, modifications are unable to cause a shift in the sugar pucker; however, a crystal structure of ASL^{Lys} bound in the ribosomal A-site reveals that all residues conform to the C3'-*endo* conformation that is characteristic of the U-turn [23]. Therefore, we suggest a similar mechanism in ASL^{Arg4} in which the loop nucleosides are induced into the C3'-*endo* conformation upon A-site binding. However, this would require a certain amount of flexibility for C_{35} to switch sugar conformation. Since proper positioning of this residue is required for stabilizing $mnm^{5}U_{34}$ in a geometry that allows for recognition

of G_3 [23], this flexibility may abrogate this function of the position 35 residue.

Besides the C3'-*endo* sugar puckers that are characteristic of a canonical U-turn, the abrupt backbone turn between position 33 and 34 requires that the α dihedral angle be in the *trans* conformation [42], resulting in a signature downfield resonance for 33p34 in ^{31}P NMR spectra [43,44]. No such resonances were seen for either the unmodified or the triply modified ASL (Figure 7). It is not uncommon for ASLs to have structures that closely resemble a U-turn motif but lack this signature resonance [4]. Similarly, there have been instances where ASLs that do not contain a U-turn motif are able to efficiently bind to codons in the ribosomal A-site [7,10,26].

Interestingly, this appears to be a common characteristic of *E.coli* arginine ASLs [7]. It is unclear how these particular isoacceptors overcome structural limitations during ribosomal binding; however, similarities in the ASLs suggest that there may be a conserved mechanism for overcoming the non-U-turn solution structure. Both $ASL^{Arg1,2}$ [7] and ASL^{Arg4} lack the pre-structuring that is common to many ASLs [4,8,29]; however, both must conform to the canonical U-turn conformation when bound to the ribosome. The fact that both ASLs contain a cytidine at position 35 that is in the C2'-*endo* conformation may allow for a certain amount of deformability in the loops of these ASLs by inhibiting the more stable stacking interactions that occur between nucleosides that are in the C3'-*endo* orientation. Also, the rare s^2C_{32} modification that is common to these two ASLs may provide insights into this mechanism.

4.5.5 Effects of s^2C_{32} on ASL structure

Since s^2C_{32} is a rare modification and common to both $ASL^{Arg1,2}$ and ASL^{Arg4}, it was hypothesized that there may be a connection between this modification and the uncommon properties that characterize these isoacceptors. In $ASL^{Arg1,2}$, the s^2C_{32} modification acts negatively by disallowing recognition of the rare CGA codon [7]. Codon-specific ribosome binding assays revealed that the function of s^2C_{32} is neither to abrogate AGG binding of the fully modified ASL nor to allow binding of the unmodified ASL to the cognate AGA codon

(Figure 2a,b). Therefore, a common functional property was not found. Structurally, s^2C_{32} was able to significantly alter the thermal melting properties of the ASL by introducing a low-temperature transition (Figure 3), suggesting that there may be a cross-strand interaction that is stabilized by this modification. A significant thermal dependence of the circular dichroism spectra correlates this result (Figure 3b,c). A large reduction in the maximum ellipticity between 5°C and 20°C suggests that there is a substantial change in the base stacking properties of the ASL between these temperatures. It is unclear whether this is the result of a destabilized extended stem or another yet to be identified property, but full NMR characterization and structure determination is being performed to decipher the conformational properties elicited by this modification.

4.5.5 Conclusions

Based on a full characterization of both the functional decoding and biophysical properties of all three ASL constructs and preliminary NMR investigation, we conclude a novel function of the anticodon domain modifications in the context of ASL^{Arg4}_{UCU}. Biophysically, mnm^5U_{34} and t^6A_{37} produce similar results of decreasing the molecular order while increasing the thermal stability as seen for ASL^{Lys}_{UUU}-mnm^5U_{34};t^6A_{37} [10]. In ASL^{Lys}_{UUU}, this is the result of negating the extended stem; however, triply modified ASL^{Arg4}_{UCU} still shows strong signs of an extended stem. The extended stem in triply modified ASL^{Arg4}_{UCU} is stabilized by the elevated pKa of A_{38}H1 resulting in a stronger C_{32}•A_{38} base pair. Since we also note a more pronounced stacking interaction between G_{31} and s^2C_{32} and a more stable s^2C_{32}•A_{38} base pair in the singly modified ASL^{Arg4}_{UCU}-s^2C_{32}, we hypothesize that the fully modified ASL will also form this non-canonical base pair which has been shown to be important for codon discrimination [24]. Although ongoing efforts will result in full structure determination of all three ASL constructs, we propose that modification of ASL^{Arg4}_{UCU} results in a highly stabilized non-canonical s^2C_{32}•A_{38} base pair that enhances the codon discrimination that results from the chemical nature of C_{35}. Additionally, as shown previously [4,10], these modifications may also open the stem-distal loop structure allowing for a more U-turn-like conformation and, in the case of t^6A_{37}, most likely stabilizes the U_{36}•A_1

codon•anticodon base pair.

4.6 Materials and methods

4.6.1 Oligonucleotide preparation

Phosphoramidites for s^2C, mnm^5U and t^6A were synthesized as described [7,49], and ASLs and mRNA oligonucleotides were purchased from Dharmacon (Thermo Scientific) and deprotected using standard protocol. RNAs were then dialyzed extensively against 20 mM sodium/potassium phosphate buffer pH 6.8 (10mM Na_2HPO_4 and 10 mM KH_2PO_4) using a 3500 Da cutoff membrane (Pierce). For pH experiments using NMR, adjustment was accomplished using phosphoric acid and sodium hydroxide. Tight-coupled 70S *E.coli* ribosomes for codon binding assays were prepared using sucrose gradient centrifugation as described [50].

4.6.2 Thermal denaturation spectroscopy

Thermal denaturation and renaturation of the ASLs were monitored by UV absorbance at 260 nm with a Cary 100 UV-visible spectrophotometer (Agilent) using Thermal software. All the samples were adjusted to ~0.2 A_{260} and the temperature was ramped from 5 to 95°C at 1°C/min at a rate of 1 data point per minute for 5 cycles per experiment. All experiments were performed simultaneously with a control cell containing buffer only, these data were subtracted as the background before analysis. A total of 30 data sets (15 full cycles) were accumulated over 3 different experiments and the first denaturation was excluded from analysis. Thermodynamic properties were determined using the Meltwin v3.5 software and the error was calculated as the standard error of the mean.

4.6.3 Circular dichroism spectroscopy

Circular dichroism spectra were collected on a Jasco J815 spectropolarimeter in a 1 cm pathlength quartz cuvette in which temperature was controlled to either 5˚C or 20˚C during

data collection. All samples were adjusted to ~0.2 A_{260}/mL (ambient temperature) prior to CD analysis. Spectra were baseline corrected using a control containing buffer only. Using simultaneously collected absorbance data to calculate the concentrations of each sample, the spectra were normalized to molar circular dichroic absorbance ($\Delta\varepsilon = \theta/32980 \bullet C \bullet L \bullet N$ [51]). Spectra shown are the averaged results of two independent experiments consisting of triplicate data collection.

4.6.4 Codon-specific ribosome A-site binding assays

With little variation, ribosome purification and binding assays were performed according to an established protocol [7]. Using UNAFOLD [52], the following 27-mer mRNAs derived from T4 gp32 [53] were investigated for secondary structure:

[i] Control mRNA (GCG): 5' – GAAAAGGAGGUAAAA**AUGCGC**GCACAU – 3'

[ii] A-site AGA: 5' – GAAAAGGAGGUAAAA**AUGAGA**GCACAU – 3'

[iii] A-site AGG: 5' – GAAAAGGAGGUAAAA**AUGAGG**GCACAU – 3'

Heptadecamer ASLs were 5'-^{32}P-radiolabeled and purified using preparative denaturing 15% polyacrylamide gel electrophoresis. mRNA-programmed ribosomes (250 nM ribosome and 2.5 μM mRNA) were P-site saturated with $tRNA^{fMet}$ and incubated for one hour at 37 ˚C with varying concentrations (0, 0.25, 0.5, 1.0, 1.5, 2.5, 5 μM) of unlabeled ASL spiked with up to 2 kCPM radiolabeled ASL in ribosome binding buffer (50 mM HEPES, pH 7; 30 mM KCl; 70 mM NH_4Cl; 1 mM DTT; 100 μM EDTA; 20 mM $MgCl_2$ adjusted to pH 7 with 2 M NaOH). Binding reactions were then filtered through a 0.22 μM nitrocellulose filter (Whatman) using a modified Micro-Sample Filtration Manifold (Whatman Schleicher & Schuell Minifold) 96-well dot blot apparatus [54]. A standard curve was generated by spotting known amounts of radiolabeled ASL to a small strip of nitrocellulose filter. Filters were imaged using a Typhoon phosphor imager (GE Healthcare) and radioactive spot density was calculated using ImageQuant TL software (Amersham). Binding constants were calculated using the one-site specific binding function in Prism v3 (Graphpad). An internal

positive control using unmodified yeast ASL^{Phe} binding to polyuridylic acid and non-specific negative controls using a control mRNA with GCG (mRNA [i]) in the A-site position were performed in tandem with every experiment. All binding curves were performed in triplicate in each of three independent experiments. Determination of ribosome activity and usability of mRNAs was accomplished using filter binding assays consisting of only ribosomes and 5'-^{32}P-radiolabeled mRNAs.

4.6.5 NMR spectroscopy

Complete NMR experiments were performed on unmodified ASL^{Arg4}_{UCU} and triply modified ASL^{Arg4}_{UCU}-mnm$^5U_{34}$;t$^6A_{37}$;Ψ_{40}. All NMR spectra were collected using ionic and thermal conditions (20mM Na/K phosphate buffer at 293K) commonly used for structure determination of RNA stem loops [25,29,55] on a Bruker 700 MHz instrument equipped with a cryoprobe, triple-resonance ^{1}H, ^{13}C, and ^{15}N probe and three-axis pulsed field gradient capabilities and a Bruker 400 MHz instrument equipped with ^{31}P probe. Data were processed using NMRPIPE [56] and the resulting spectra were analyzed using NMRViewJ (One Moon Scientific Inc). For exchangeable protons, ^{1}H-^{1}H NOESY spectra were recorded in 90% H_2O/10% D_2O at a variety of different temperatures (283K, 288K and 293K) and mixing times (50 ms, 100 ms, 200 ms and 400 ms). Non-exchangeable proton spectra were collected at 293K and different mixing times (50 ms, 100 ms, 200 ms and 400 ms). COSY, TOCSY, ^{1}H-^{31}P HETCOR and natural abundance ^{1}H-^{13}C HMQC spectra were recorded at 25°C in 100% in D_2O.

4.7 Acknowledgements

The authors would like to acknowledge the contributions of Dr. C. Stark and Dr. J. Spears for productive conversations and detailed manuscript review. Contributions were also made by Dr. A. Shekhtman and Yu Chen in the form of technical support for NMR instrumentation and computer software, respectively.

4.8 References

1. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FA, Fabris D, Agris PF: The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 2011, **39**(Database issue):D195-201.

2. Tamura K, Himeno H, Asahara H, Hasegawa T, Shimizu M: *In vitro* study of *E.coli* $tRNA^{Arg}$ and $tRNA^{Lys}$ identity elements. *Nucleic Acids Res* 1992, **20**(9):2335-2339.

3. Stuart JW, Koshlap KM, Guenther R, Agris PF: Naturally-occurring modification restricts the anticodon domain conformational space of $tRNA^{Phe}$. *J Mol Biol* 2003, **334**(5):901-918.

4. Vendeix FA, Murphy FV, Cantara WA, Leszczynska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF: Human tRNA(Lys3)(UUU) Is Pre-Structured by Natural Modifications for Cognate and Wobble Codon Binding through Keto-Enol Tautomerism. *J Mol Biol* 2011.

5. Björk GR, Durand JM, Hagervall TG, Leipuviene R, Lundgren HK, Nilsson K, Chen P, Qian Q, Urbonavicius J: Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett* 1999, **452**(1-2):47-51.

6. Urbonavicius J, Qian Q, Durand JM, Hagervall TG, Björk GR: Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J* 2001, **20**(17):4863-4873.

7. Cantara WA, Bilbille Y, Kim J, Kaiser R, Leszczynska G, Malkiewicz A, Agris PF: Modifications Modulate Anticodon Loop Dynamics and Codon Recognition of *E. coli* tRNA(Arg1,2). *J Mol Biol* 2012.

8. Lusic H, Gustilo EM, Vendeix FA, Kaiser R, Delaney MO, Graham WD, Moye VA, Cantara WA, Agris PF, Deiters A: Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial $tRNA^{Met}$. *Nucleic Acids Res* 2008, **36**(20):6548-6557.

9. Ashraf SS, Sochacka E, Cain R, Guenther R, Malkiewicz A, Agris PF: Single atom modification (O-->S) of tRNA confers ribosome binding. *RNA* 1999, **5**(2):188-194.

10. Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, Agris PF: Modified nucleoside dependent Watson-Crick and wobble codon binding by $tRNA^{Lys}_{UUU}$ species. *Biochemistry* 2000, **39**(44):13390-13395.

11. Yarian C, Townsend H, Czestkowski W, Sochacka E, Malkiewicz AJ, Guenther R, Miskiewicz A, Agris PF: Accurate translation of the genetic code depends on tRNA modified nucleosides. *J Biol Chem* 2002, **277**(19):16391-16395.

12. Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985, **2**(1):13-34.

13. Chen GT, Inouye M: Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev* 1994, **8**(21):2641-2652.

14. Gurvich OL, Baranov PV, Gesteland RF, Atkins JF: Expression levels influence ribosomal frameshifting at the tandem rare arginine codons AGG_AGG and AGA_AGA in *Escherichia coli*. *J Bacteriol* 2005, **187**(12):4023-4032.

15. Spanjaard RA, Chen K, Walker JR, van Duin J: Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg). *Nucleic Acids Res* 1990, **18**(17):5031-5036.

16. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J: tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 2009, **37**(Database issue):D159-162.

17. Agris PF, Sierzputowska-Gracz H, Smith W, Malkiewicz A, Sochacka E, Nawrot B: Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J Amer Chem Soc* 1992, **114**(7):2652-2656.

18. Kawai G, Yamamoto Y, Kamimura T, Masegi T, Sekine M, Hata T, Iimori T, Watanabe T, Miyazawa T, Yokoyama S: Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 1992, **31**(4):1040-1046.

19. Smith WS, Sierzputowska-Gracz H, Sochacka E, Malkiewicz A, Agris PF: Chemistry and structure of modified uridine dinucleosides are determined by thiolation. *J Amer Chem Soc* 1992, **114**(21):7989-7997.

20. Yokoyama S, Watanabe T, Murao K, Ishikura H, Yamaizumi Z, Nishimura S, Miyazawa T: Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc Natl Acad Sci USA* 1985, **82**(15):4905-4909.

21. Bilbille Y, Vendeix FAP, Guenther R, Malkiewicz A, Ariza X, Vilarrasa J, Agris PF: The structure of the human $tRNA^{Lys3}$ anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs. *Nucleic Acids Res* 2009, **37**(10):3342-3353.

22. Murphy FV, Ramakrishnan V: Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 2004, **11**(12):1251-1252.

23. Murphy FV, Ramakrishnan V, Malkiewicz A, Agris PF: The role of modifications in codon discrimination by $tRNA^{Lys}_{UUU}$. *Nature Struct Biol* 2004, **11**(12):1186-1191.

24. Olejniczak M, Uhlenbeck OC: tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 2006, **88**(8):943-950.

25. Durant PC, Bajji AC, Sundaram M, Kumar RK, Davis DR: Structural effects of hypermodified nucleosides in the *Escherichia coli* and human $tRNA^{Lys}$ anticodon loop: The effect of nucleosides s^2U, mcm^5U, mcm^5s^2U, mnm^5s^2U, t^6A, and ms^2t^6A *Biochemistry* 2005, **44**(22):8078-8089.

26. Stuart JW, Gdaniec Z, Guenther R, Marszalek M, Sochacka E, Malkiewicz A, Agris PF: Functional anticodon architecture of human $tRNA^{Lys3}$ includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t^6A. *Biochemistry* 2000, **39**(44):13396-13404.

27. Samejima T, Hashizume H, Imahori K, Fujii I, Miura K: Optical rotatory dispersion and circular dichroism of rice dwarf virus ribonucleic acid. *J Mol Biol* 1968, **34**(1):39-48.

28. Durant PC, Davis DR: Stabilization of the anticodon stem-loop of $tRNA^{Lys,3}$ by an A^+-C base-pair and by pseudouridine. *J Mol Biol* 1999, **285**(1):115-131.

29. Vendeix FAP, Dziergowska A, Gustilo EM, Graham WD, Sproat B, Malkiewicz A, Agris PF: Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry* 2008, **47**(23):6117-6129.

30. Flinders J, Dieckmann T: NMR spectroscopy of ribonucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* 2006, **48**:137-159.

31. Furtig B, Richter C, Wohnert J, Schwalbe H: NMR spectroscopy of RNA. *Chembiochem* 2003, **4**(10):936-962.

32. Varani G, Adoul-ela F, Allain FHT: NMR investigation of RNA structure. *Progr Nuclear Magnetic Resonance Spectroscopy* 1996, **29**(1-2):51-127.

33. Wijmenga SS, van Buuren B: The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* 1998, **32**:287-387.

34. Davis DR, Griffey RH, Yamaizumi Z, Nishimura S, Poulter CD: 15N-labeled tRNA. Identification of dihydrouridine in *Escherichia coli* $tRNA^{fMet}$, $tRNA^{Lys}$, and $tRNA^{Phe}$ by ^1H-^{15}N two-dimensional NMR. *The Journal of biological chemistry* 1986, **261**(8):3584-3587.

35. Griffey RH, Davis D, Yamaizumi Z, Nishimura S, Bax A, Hawkins B, Poulter CD: ^{15}N-labeled *Escherichia coli* tRNA^{fMet}, tRNA^{Glu}, tRNA^{Tyr}, and tRNA^{Phe}. Double resonance and two-dimensional NMR of N1-labeled pseudouridine. *The Journal of biological chemistry* 1985, **260**(17):9734-9741.

36. Roy S, Papastavros MZ, Sanchez V, Redfield AG: Nitrogen-15-labeled yeast tRNA^{Phe}: double and two-dimensional heteronuclear NMR of guanosine and uracil ring NH groups. *Biochemistry* 1984, **23**(19):4395-4400.

37. Clore GM, Gronenborn AM, Piper EA, McLaughlin LW, Graeser E, van Boom JH: The solution structure of a RNA pentadecamer comprising the anticodon loop and stem of yeast tRNA^{Phe}. A 500 MHz ^{1}H-N.M.R. study. *The Biochemical journal* 1984, **221**(3):737-751.

38. Davis DR: Stabilization of RNA stacking by pseudouridine. *Nucleic acids research* 1995, **23**(24):5020-5026.

39. Davis DR, Veltri CA, Nielsen L: An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNA^{Lys}, tRNA^{His} and tRNA^{Tyr}. *Journal of biomolecular structure & dynamics* 1998, **15**(6):1121-1132.

40. Hall KB, McLaughlin LW: Properties of a U1/mRNA 5' splice site duplex containing pseudouridine as measured by thermodynamic and NMR methods. *Biochemistry* 1991, **30**(7):1795-1801.

41. Legault P, Pardi A: In situ probing of adenine protonation in RNA by ^{13}C NMR. *J Am Chem Soc* 1994, **116**(18):8390-8391.

42. Saenger W: Principles of Nucleic Acid Structure. Springer-Verlag, New York; 1983.

43. Jucker FM, Pardi A: GNRA tetraloops make a U-turn. *RNA* 1995, **1**(2):219-222.

44. Schweisguth DC, Moore PB: On the conformation of the anticodon loops of initiator and elongator methionine tRNAs. *J Mol Biol* 1997, **267**(3):505-519.

45. Ashraf SS, Guenther RH, Ansari G, Malkiewicz A, Sochacka E, Agris PF: Role of modified nucleosides of yeast tRNA(Phe) in ribosomal binding. *Cell Biochem Biophys* 2000, **33**(3):241-252.

46. Agris PF: Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* 1991, **73**(11):1345-1349.

47. Agris PF: The importance of being modified: roles of modified nucleosides and Mg^{2+} in RNA structure and function. *Prog Nucleic Acid Res Mol Biol* 1996, **53**:79-129.

48. Sundaram M, Durant PC, Davis DR: Hypermodified nucleosides in the anticodon of $tRNA^{Lys}$ stabilize a canonical U-turn structure. *Biochemistry* 2000, **39**(41):12575-12584.

49. Agris PF, Malkiewicz A, Kraszewski A, Everett K, Nawrot B, Sochacka E, Jankowska J, Guenther R: Site-selected introduction of modified purine and pyrimidine ribonucleosides into RNA by automated phosphoramidite chemistry. *Biochimie* 1995, **77**(1-2):125-134.

50. Powers T, Noller HF: Dominant lethal mutations in a conserved loop in 16S rRNA. *Proc Natl Acad Sci USA* 1990, **87**(3):1042-1046.

51. Sosnick TR: Characterization of tertiary folding of RNA by circular dichroism and urea. *Curr Protoc Nucleic Acid Chem* 2001, **Chapter 11**:Unit 11 15.

52. Markham NR, Zuker M: DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* 2005, **33**(Web Server issue):W577-581.

53. Fahlman RP, Dale T, Uhlenbeck OC: Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol Cell* 2004, **16**(5):799-805.

54. Wong I, Lohman TM: A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci USA* 1993, **90**(12):5428-5432.

55. Denmon AP, Wang J, Nikonowicz EP: Conformation effects of base modification on the anticodon stem-loop of Bacillus subtilis tRNA(Tyr). *J Mol Biol* 2011, **412**(2):285-303.

56. Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A: NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995, 6(3):277-293.

4.9 Figures and tables

Table 1. Thermodynamic properties of ASL^{Arg4}_{UCU} constructs derived from thermal melting.

	ΔH (kcal/mol)	ΔS (cal/K*mol)	ΔG (kcal/mol,37°C)	T_m (°C)	%HCT
Unmod	-50.09 ± 0.86	-148.61 ± 2.55	-4.00 ± 0.07	63.93 ± 0.13	11.9 ± 0.1
3x Mod	-41.72 ± 0.93	-123.16 ± 2.77	-3.53 ± 0.08	65.65 ± 0.22	10.2 ± 0.1

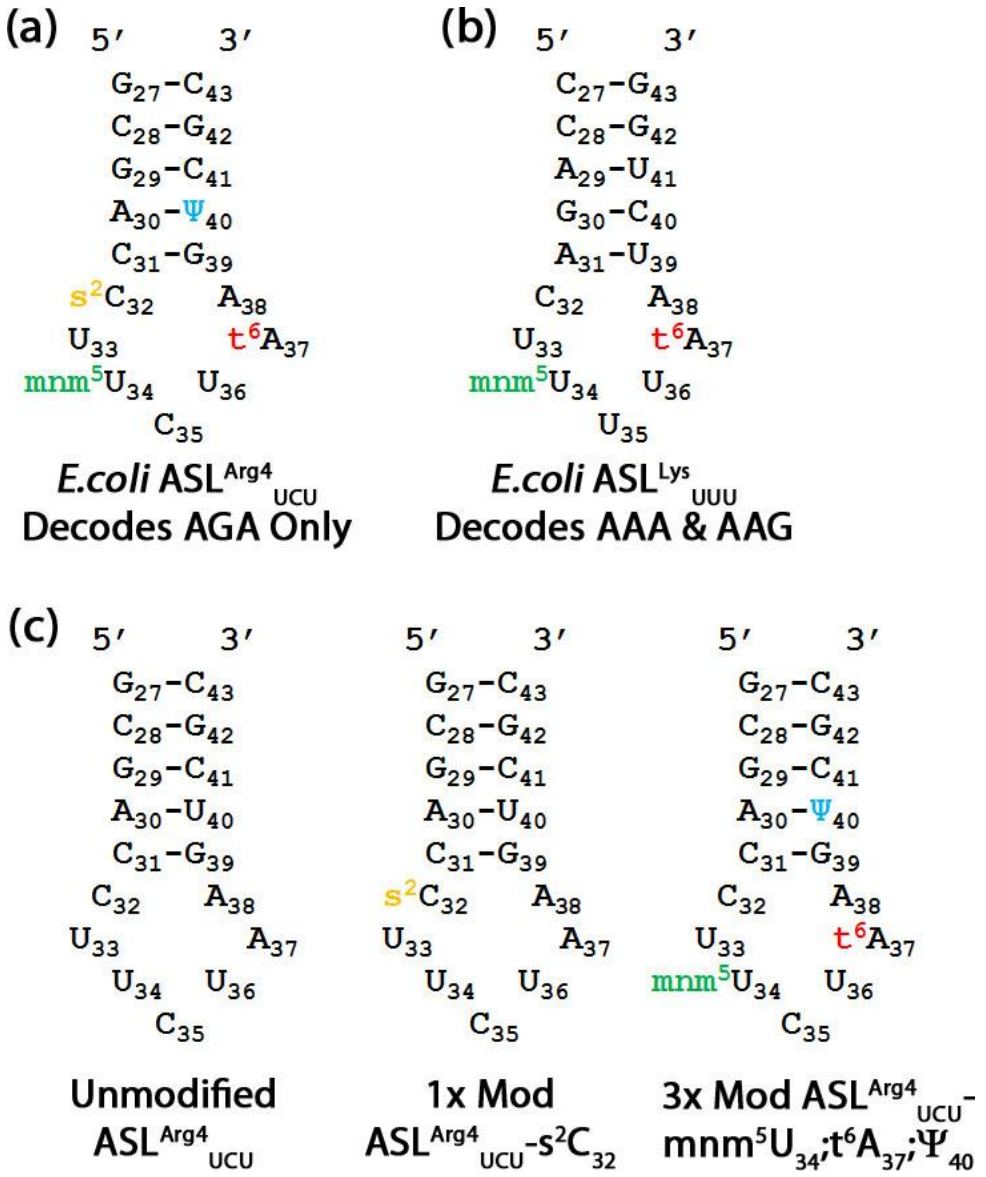

Figure 1. Primary structures of ASL$^{Arg4}_{UCU}$ constructs. (a) The primary sequence of the anticodon domain of wild-type *E.coli* ASL$^{Arg4}_{UCU}$ has four modifications: s^2C_{32} (gold), mnm^5U_{34} (green), t^6A_{37} (red) and Ψ_{40} (blue). **(b)** This sequence can be compared to the hypomodified *E.coli* ASL$^{Lys}_{UUU}$-mnm^5U_{34};t^6A_{37}, which has a very similar loop sequence but has a very different decoding phenotype [10]. **(c)** Three ASL constructs were synthesized to determine the function of s^2C_{32}, both biophysically and functionally.

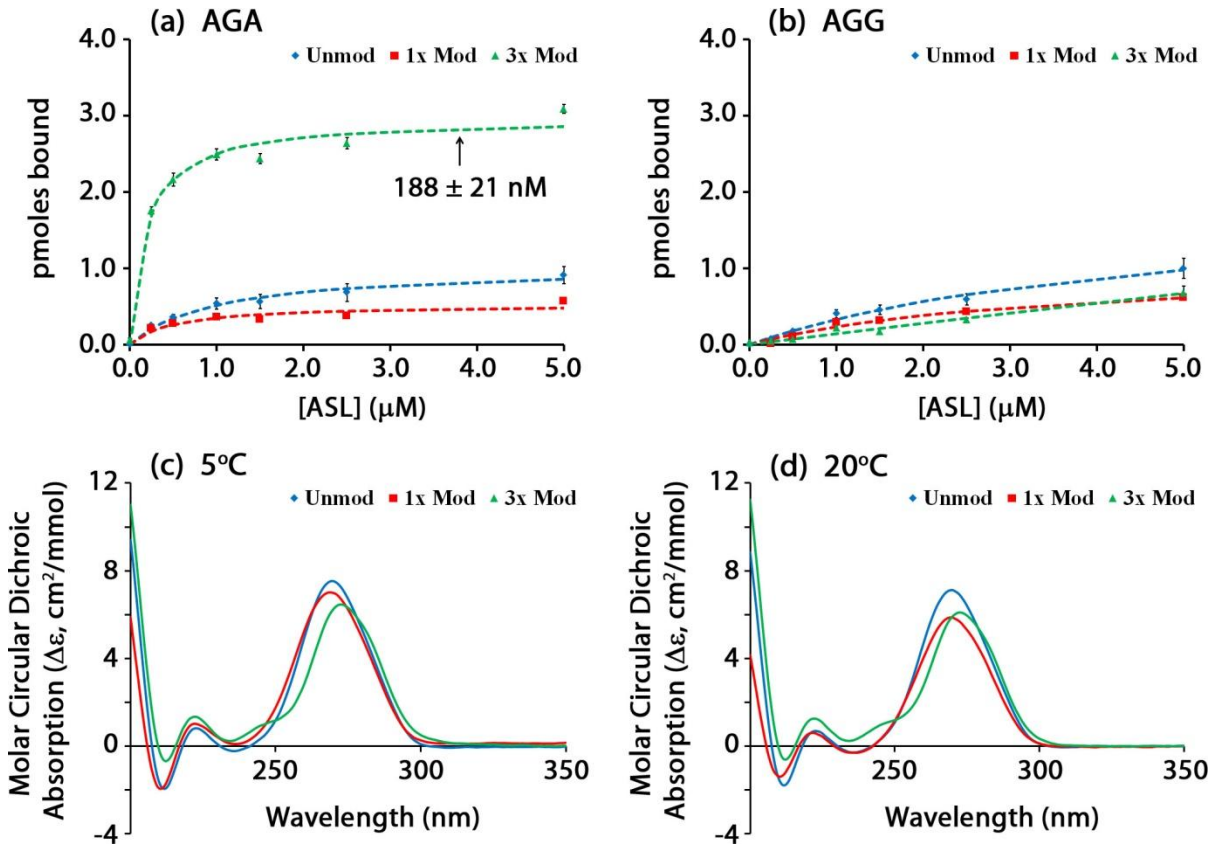

Figure 2. Codon binding assays and circular dichroism spectra. Ribosomal codon binding assays were performed with both **(a)** AGA and **(b)** AGG in the A-site, showing that only the triply modified ASL bound efficiently to the AGA codon. Circular dichroism spectra were recorded at **(c)** 5°C and **(d)** 20°C to determine both the base stacking properties of each ASL and the thermal stability of the stacking interactions.

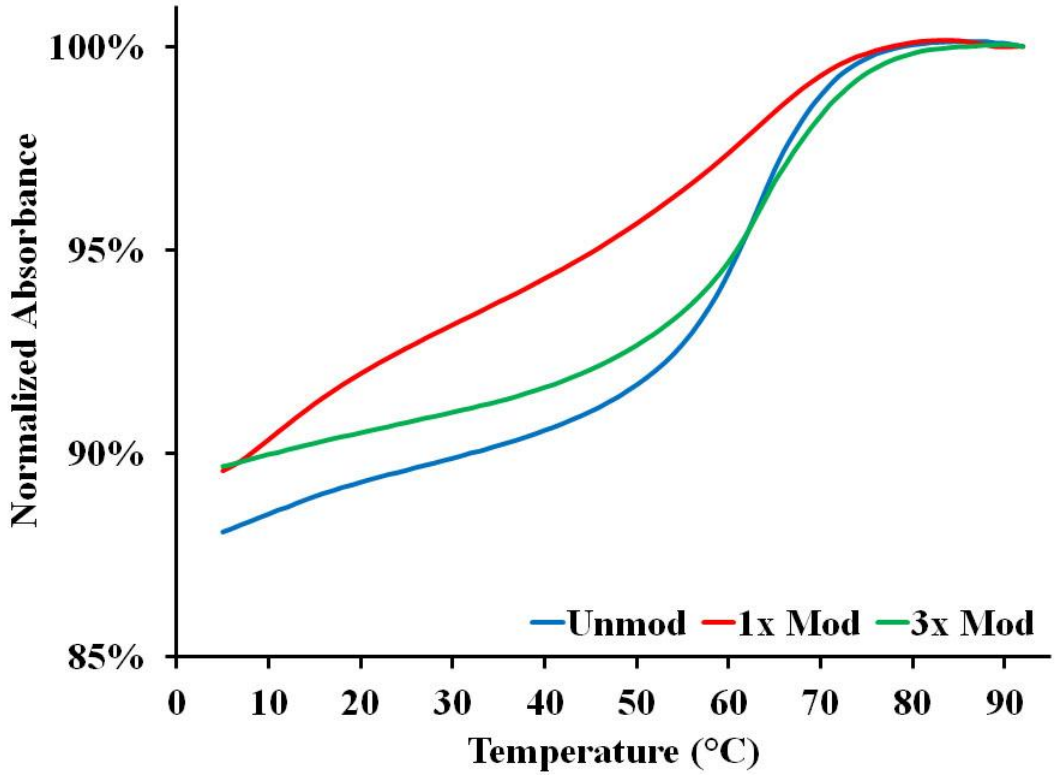

Figure 3. Thermal denaturation and renaturation spectra. Thermal denaturation cycles were monitored by UV abosorbance for the three differentially modified ASLs: ASL^{Arg4}_{UCU} (blue), ASL^{Arg4}_{UCU}-s^2C_{32} (red) and ASL^{Arg4}_{UCU}-mnm^5U_{34};t^6A_{37};Ψ_{40} (green). Curves shown are each the average of 15 full cycles normalized to 100% at the fully denatured state. The arrow indicates a possible intermediate conformation for ASL^{Arg4}_{UCU}-s^2C_{32} suggested by a an additional change in concavity that did allow for thermodynamic analysis using a two state model as used for the other two curves (Table 1).

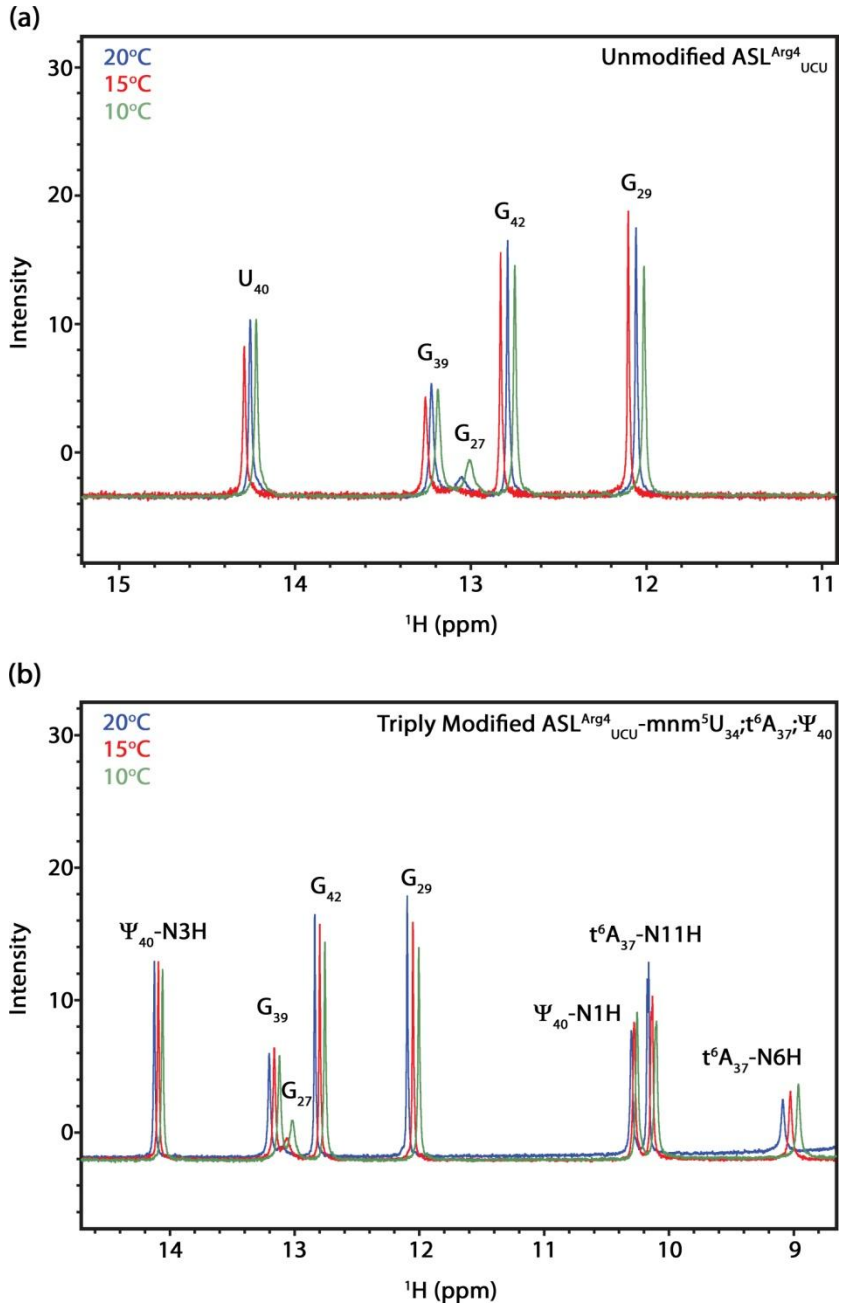

Figure 4. Temperature-dependent 1D imino spectra confirms presence of modifications. **(a)** Unmodified and (b) triply modified ASLs both showed small variations in the imino region of the NOESY spectra analogous to temperature differences. Also, triply modified ASL^{Arg4} showed additional peaks corresponding to exchangeable protons present in t^6A_{37} and Ψ_{40}. Both spectra were recorded at pH 6.8.

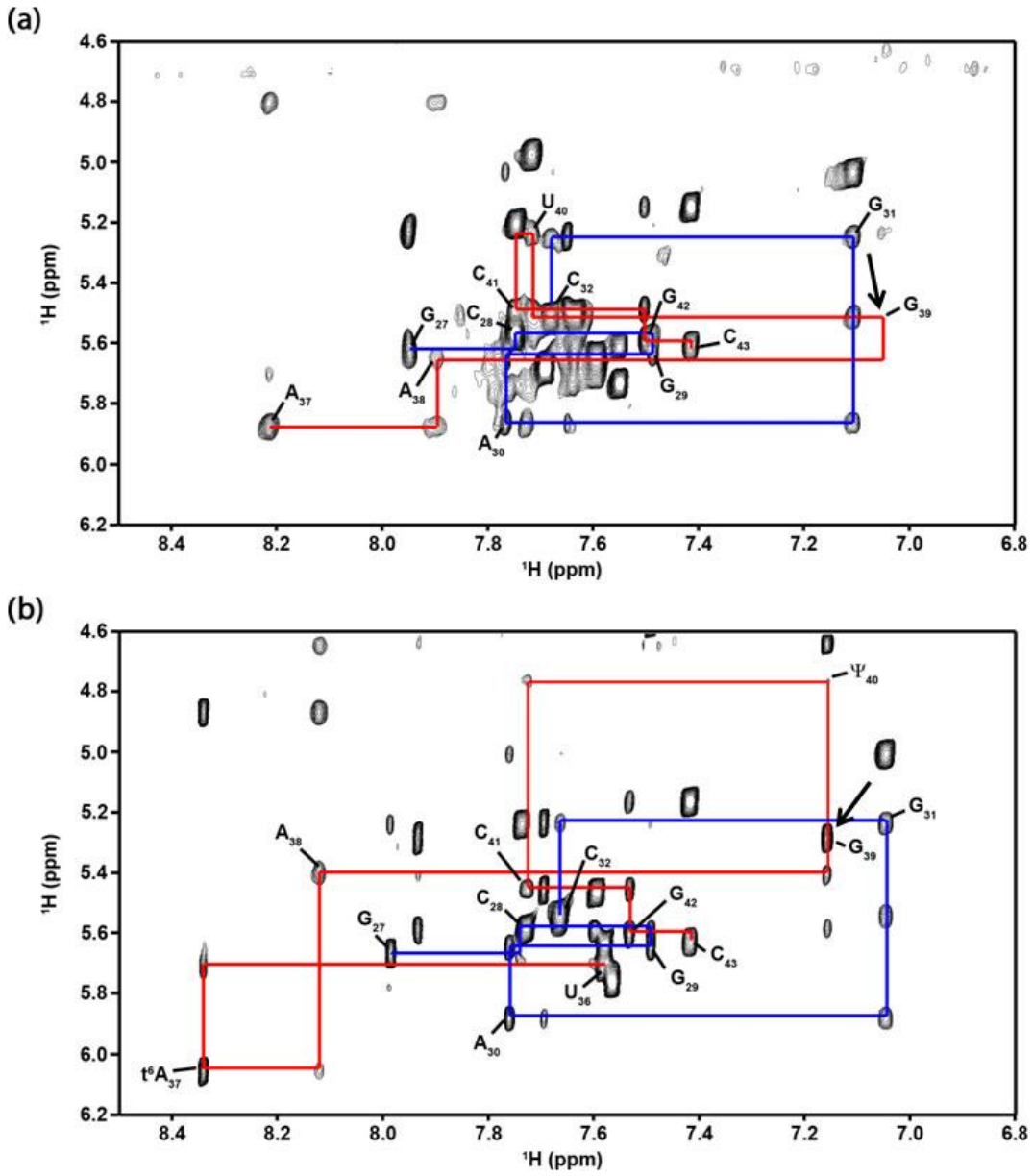

Figure 5. Sequential H1'-aromatic NOE "walk" spectra. A characteristic NOE walk was observed between G_{27}H1'-H8 and C_{32}H1'-H6 (blue) and from A_{37}H1'-H8 to C_{43}H1'-H6 (red) for unmodified ASL^{Arg4} **(a)** and triply modified ASL^{Arg4} **(b)**. Arrows indicate the position of resonances G_{39}H8 resonances. Both spectra were recorded at pH 6.8.

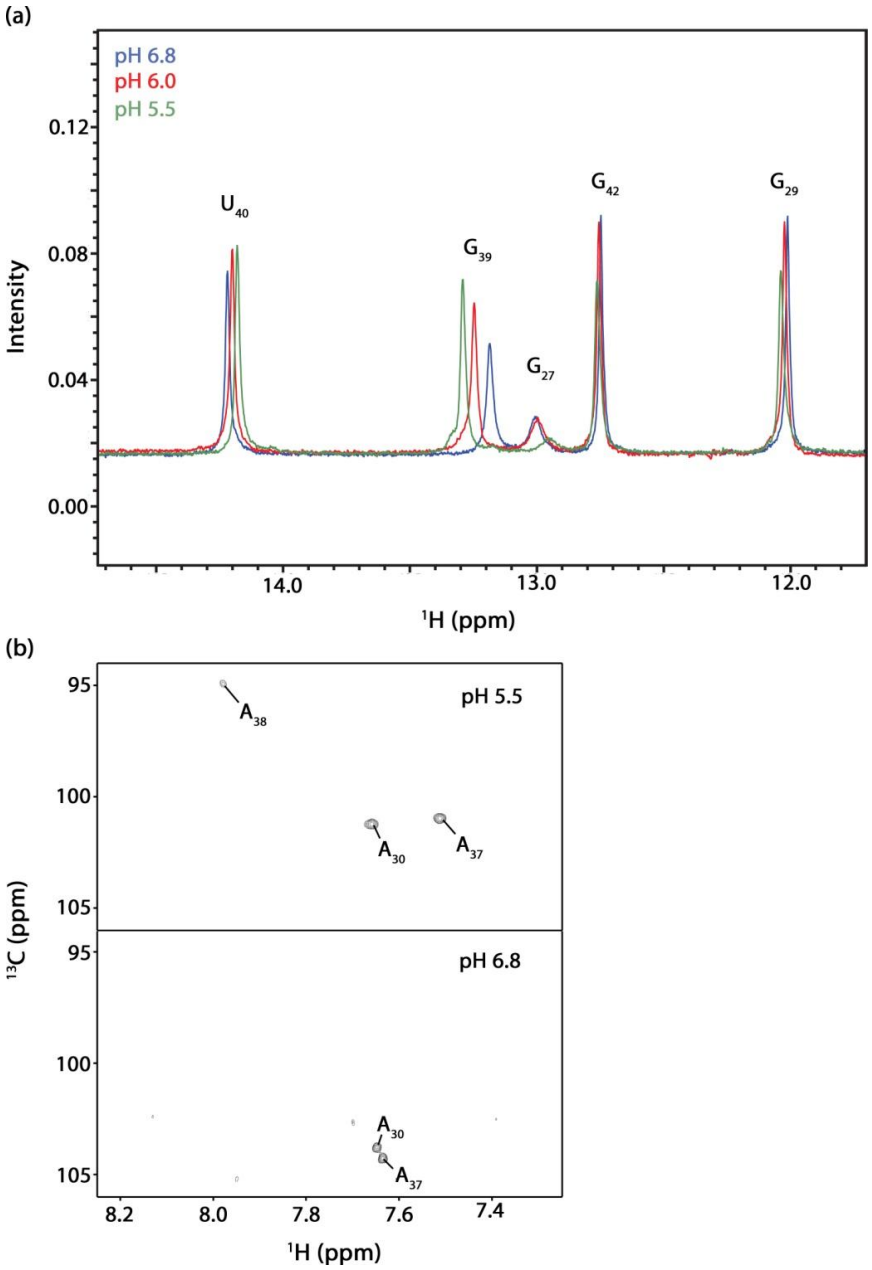

Figure 6. pH-dependent protonation state of unmodified ASL^{Arg4} A_{38}. (a) The imino region of the NOESY spectrum for unmodified ASLArg4 shows a significant peak shift for G_{39}H1 as the pH is decreased from 6.8 (blue) to 6.0 (red) and 5.5 (green). **(b)** Correspondingly, an unusual shifted A_{38}C2 in the ^{1}H-^{13}C HSQC spectrum collected at pH 5.5 indicated the protonation of A_{38}, which could affect the G_{39}H1 resonance.

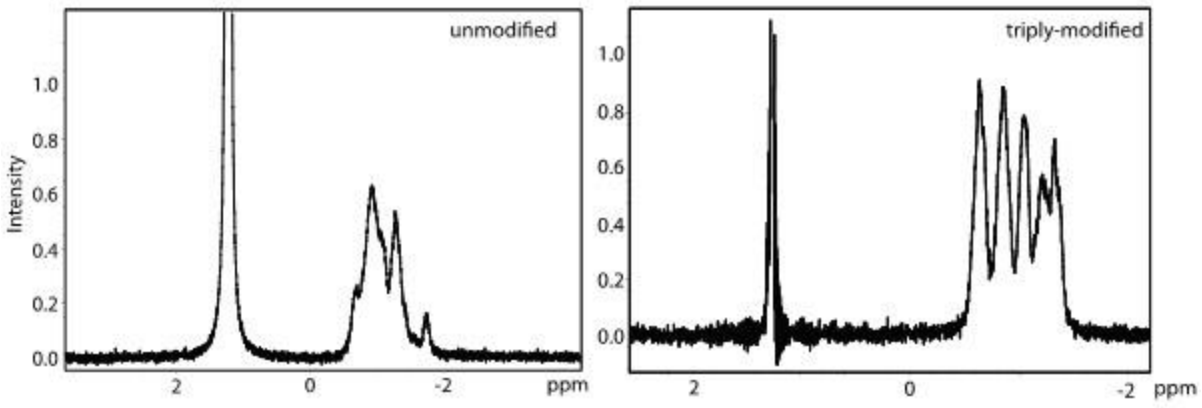

Figure 7. ^{31}P NMR spectra of unmodified and triply-modified ASL^{Arg4}. No downfield shifted resonance, characteristic of a typical U-turn, were observed for both ASLs.

CHAPTER 5. Concluding remarks

5.1 Overcoming problems associated with decoding six-fold degenerate codons

The universal genetic code (see Chapter 1, Figure 1) consists of three amino acids that are coded by six different codons, arginine, leucine and serine. For each set, a single aminoacyl-tRNA synthetase must recognize all isoacceptors. Not only must multiple isoacceptors adopt a uniform conformation for ribosome binding, they must also have the specific structure and chemistry required for specific recognition by the cognate aminoacyl-tRNA synthetase. This becomes a problem with split codon boxes not only because of the ambiguity of the chemistry at the wobble position, but also to allow for different nucelosides at positions 35 and/or 36. For *E.coli* arginyl-tRNAs, the common C_{35} is used as a strong identity determinant for aminoacylation [1,2]. Based on the crystal structure of the yeast $tRNA^{Arg}_{ICG}$ in complex with its cognate aminoacyl-tRNA synthetase, C_{35} must be in a position away from other nucleosides and solvent exposed in order to make specific contacts with the protein backbone and sidechains [3]. This requirement for aminoacylation has interesting implications for decoding. First, the anticodon stem and loop domain (ASL) must have significantly more deformability than that of ASLs from other isoacceptors. Second, a significant conformational change must occur between binding to the aminoacyl-tRNA synthetase and adopting the canonical U-turn seen while decoding synonymous codons in the ribosomal A-site [4]. Finally, an induced-fit model for these structural changes requires a specific conformational intermediate to allow for a conformational equilibrium in which the tRNA can be recycled after decoding to be aminoacylated again.

In the studies described herein, we have identified the likely conformational intermediate for ASL^{Arg1}_{ICG} [5]. Despite the unconventional and singular structure of this ASL in different modification states, our data shows a variety modification-dependent of thermodynamic and functional properties. We show, for instance that with only the I_{34} modification, all three synonymous codons are able to be decoded; however, the addition of either s^2C_{32} or m^2A_{37}

completely negates binding to the rare CGA codon. Although the biophysical data indicates that modifications modulate the dynamic properties of the ASL, these properties do not empirically point to a specific mechanism for the codon discrimination. Computational studies presented here describe the environment-dependent, residue-specific flexibility alterations that are imparted in different states of modification. It is clear that the conformational transitions described will require specific residues to have a large conformational space. Interestingly, modifications have typically been seen to induce rigidity into the loop [6,7], which would not be beneficial to a molecule that requires such drastic conformational changes. Therefore, despite having only preliminary computational evidence, we postulate a model in which the state of ASL modification is dynamic and that modifications alter the landscape between the different conformations to both modulate structural transitions and allow for regulation of decoding ability.

Although ASL^{Arg1}_{ICG} has the ability to decode three codons, the remaining three isoacceptors recognize only one codon each. Of these three, ASL^{Arg4}_{UCU} stands out because of its decoding difference with a slightly hypomodified form of ASL^{Lys}_{UUU} despite their nearly identical loop sequence and modifications (See Chapter 4, Figure 1a,b). The fact that the particular geometry of the non-canonical 32•38 base pair participates in regulation of codon discrimination suggested that the s^2C_{32} modification may be responsible for negating wobble decoding [8]. Indeed, s^2C_{32} was shown to inhibit the ability of ASL^{Arg1}_{ICG} to decode the CGA wobble codon [5]. However, in the present study, we have definitively shown that the s^2C_{32} modification is not responsible for negating wobble decoding. Interestingly, analysis of thermodynamic and base stacking properties indicates a similar effect of modifications on the biophysical characteristics of the loop as seen for ASL^{Lys}_{UUU} [9]. A structural analysis shows that modifications participate in allowing cognate decoding through stabilization of the non-canonical 32•38 base pair through modulation of the pKa of A_{38} and enhancement of base stacking interactions between G_{31} and s^2C_{32}.

5.2 Anticodon domain modifications as regulators of cellular function

The fact that the modification state of the $tRNA^{Arg1}_{ICG}$ ASL can regulate whether the isoacceptor decodes all three codons or only the two common codons (CGU and CGC) indicates a role for modifications in regulating the function of the tRNA. This would serve three advantageous functions. First, this would counteract the need for two different isoacceptors to decode the common and rare codons. Also, there is not a need for another loop sequence to adopt the same conformations for recognition by both the ribosome and the aminoacyl-tRNA synthetase. Finally, the function of the tRNA, and thus translation of CGA-containing genes, could be regulated by a single modification enzyme rather than a full set of tRNA transcription and maturation factors. The role of modification enzymes as regulators of cellular function has been reported previously in yeast where tRNA methylation is regulated during instances of cellular stress [10]. Here, the expression patterns and activity of the modifications enzymes for s^2C_{32} (TtcA,[11]) and m^2A_{37} (TrmG,[12]) are potential regulators of CGA-containing gene expression at the translational level. Although a thorough examination of genes that include CGA codons has not been determined, genes that are moderately to lowly expressed are enriched with these codons [13]. Taken together, these are very clear indications that the modification state of $tRNA^{Arg1}_{ICG}$ may be dynamically regulated in order to further control expression of CGA-containing mRNAs. A further study using analysis of tRNA modification during different environmental conditions or growth states in addition to analysis of expression patters and codon bias would successfully test the predictions of these studies.

5.3 Conclusions

In conclusion, our analysis of $tRNA^{Arg}$ ASL modifications has shown an interesting dual role for the rare s^2C_{32} modification. In one instance, it has a clear inhibitory effect on wobble decoding; however, in the context of ASL^{Arg4}_{UCU}, our results indicate that it may enhance cognate codon binding. Indeed, we have also identified a dual role for mnm^5U_{34} in

$tRNA^{Arg4}_{UCU}$ and $tRNA^{Lys}_{UUU}$ to either inhibit or allow wobble decoding, respectively. This dual role of ASL modifications underscores the subtlety of modifications and how a single modification can be made to perform different functions depending on the sequence and chemical context of the ASL. In this way, one of the key problems with degeneracy of the genetic code, namely the simultaneous need for structural uniformity and sequence variety in the ASL, can be alleviated through a smaller subset of modification chemistries. Therefore, we propose a model in which the modification state of the ASL is modulated in response to environmental conditions and results in functions that are dependent on the context of the other loop nucleosides and modifications.

5.4 References

1. Sissler M, Giege R, Florentz C: Arginine aminoacylation identity is context-dependent and ensured by alternate recognition sets in the anticodon loop of accepting tRNA transcripts. *EMBO J* 1996, **15**(18):5069-5076.

2. Tamura K, Himeno H, Asahara H, Hasegawa T, Shimizu M: *In vitro* study of *E.coli* $tRNA^{Arg}$ and $tRNA^{Lys}$ identity elements. *Nucleic Acids Res* 1992, **20**(9):2335-2339.

3. Delagoutte B, Moras D, Cavarelli J: tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J* 2000, **19**(21):5599-5610.

4. Murphy FV, Ramakrishnan V: Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 2004, **11**(12):1251-1252.

5. Cantara WA, Bilbille Y, Kim J, Kaiser R, Leszczynska G, Malkiewicz A, Agris PF: Modifications Modulate Anticodon Loop Dynamics and Codon Recognition of E. coli tRNA(Arg1,2). *J Mol Biol* 2012.

6. Kawai G, Yamamoto Y, Kamimura T, Masegi T, Sekine M, Hata T, Iimori T, Watanabe T, Miyazawa T, Yokoyama S: Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 1992, **31**(4):1040-1046.

7. Vendeix FAP, Dziergowska A, Gustilo EM, Graham WD, Sproat B, Malkiewicz A, Agris PF: Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry* 2008, **47**(23):6117-6129.

8. Olejniczak M, Uhlenbeck OC: tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 2006, **88**(8):943-950.

9. Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, Agris PF: Modified nucleoside dependent Watson-Crick and wobble codon binding by $tRNA^{Lys}_{UUU}$ species. *Biochemistry* 2000, **39**(44):13390-13395.

10. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ: A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet* 2010, **6**(12):e1001247.

11. Jager G, Leipuviene R, Pollard MG, Qian Q, Bjork GR: The conserved Cys-X1-X2-Cys motif present in the TtcA protein is required for the thiolation of cytidine in position 32 of tRNA from Salmonella enterica serovar Typhimurium. *J Bacteriol* 2004, **186**(3):750-757.

12. Qian Q, Curran JF, Bjork GR: The methyl group of the N6-methyl-N6-threonylcarbamoyladenosine in tRNA of Escherichia coli modestly improves the efficiency of the tRNA. *J Bacteriol* 1998, **180**(7):1808-1813.

13. Sharp PM, Li WH: Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. *Nucleic Acids Res* 1986, **14**(19):7737-7749.

APPENDICES

APPENDIX A. 70S ribosome preparation

A.1 Protocol Notes

This protocol is adapted from ribosome preparation protocols performed in the Wollenzien laboratory at North Carolina State University and the Joseph laboratory at the University of California, San Diego. Once flash frozen, these ribosomes are usable indefinitely as long as they are kept at -80°C. The typical aliquot consists of 20uL for use in codon-specific ribosomal filter binding assays; however, this amount can be safely adjusted for use in other capacities. It is absolutely necessary to use an *Escherichia coli* variant such as MRE600 that has a reduced amount of RNA degradation. Be sure plan ahead; this protocol takes at least two days and, in some cases, can take three days to increase purity or activity.

A.2 Buffer Preparation

Buffers A and B should be prepared fresh for each ribosome purification. After buffer preparation, all buffers should be kept at 4°C or on ice for the duration of the protocol.

100mL Buffer A: 20mM TrisHCl, 20mM $MgCl_2$, 200mM NH_4Cl, 4mM 2-mercaptoethanol

1 mL	2 M Tris HCl (pH 7.5)
2 mL	1 M $MgCl_2$
8 mL	2.5 M NH_4Cl
15.6 µL	2-mercaptoethanol

50mL Buffer B: 20mM TrisHCl, 10mM $MgCl_2$, 500mM NH_4Cl, 1mM 2-mercaptoethanol

0.5 mL	2 M Tris HCl (pH 7.5)
0.5 mL	1 M MgCl2
10 mL	2.5 M NH4Cl
3.9 µL	2-mercaptoethanol

A.3 Cell Growth and Lysis

1. Weight out ~8g of frozen *E.coli* MRE600 cells.
2. Resuspend cells in 20mL Buffer A.
    a. **IMPORTANT:** Keep everything on ice!
3. Lyse the cell suspension by sonication.
    a. Clean the sonicator probe with 95% ethanol and rinse with MQH_2O.
    b. Sonicate cell suspension for 30 seconds and mix by pipetting up and down.
    c. Repeat step #4 five more times.
    d. Clean the sonicator with 95% ethanol and rinse with MQH_2O.

A.4 Sucrose Gradient Purification

1. Aliquot the lysed cell mixture into four ultracentrifuge tubes (balance them carefully!).
2. Centrifuge tubes in a Beckman TI 50.2 rotor.
    a. Turn on diffusion.
    b. Turn on vacuum (wait for pressure to go below 500 microns).
    c. Set the time (20 minutes).
    d. Set the speed (16k RPM; 30854 x g)
    e. Set the temperature for 4°C.
    f. Turn off the brake
    g. Once the vacuum light is on, push the start button
3. Pour the supernatant into four fresh centrifuge tubes (balance them carefully!).
4. Spin again in ultracentrifuge (follow the pattern in 2a-g) at 30k RPM (108472 x g) for 3 hours at 4°C. The pellet will contain crude ribosomes.
5. Discard the supernatant and invert tubes for 5 minutes to dry at 4°C.
6. Wash each pellet (crude ribosomes) in 1mL Buffer B.

7. Resuspend each pellet in 7mL Buffer B using small (micro) stir bars. Do this slowly and on ice. It should take at least 1 hour.
8. Prepare 50mL of 1.1M sucrose cushion using Buffer B.
9. Load 4mL ribosome suspension onto an 8mL 1.1M sucrose cushion (two samples) in ultracentrifuge tubes. Carefully balance the tubes!
10. Spin again in ultracentrifuge (follow the pattern in 2a-g) at 30k RPM (108472 x g) for 18 hours at 4°C.
11. Wash each pellet with 5mL ice cold Buffer B.
12. Resuspend each pellet in 1mL Buffer B. Do this carefully using the micro stir bars as in step 7 and be sure this step is performed entirely on ice.
13. Make 20µL aliquots and immediately flash freeze them in liquid nitrogen. Store in the -80˚C freezer.
14. Once thawed, the ribosome concentration of each aliquot can be determined as follows: 1 A_{260} = 23-24 pmoles tight coupled 70S ribosomes (I typically use 23).
15. To improve purity, after step 11, steps 7-11 can be repeated. This is not typically necessary.

APPENDIX B. ^{32}P 5'-end labeling and urea PAGE purification of RNA

B.1 Protocol Notes

This protocol is set up for samples that are 100μM in concentration. The protocol will label 100pmoles of RNA using T4 Polynucleotide Kinase (T4 PNK) from New England Biolabs. You can label more or less by adjusting the labeling recipe accordingly. You should set aside enough time in your schedule for a 1 hour incubation followed by a lengthy polyacrylamide gel electrophoresis (generally around 2.5 hours). Samples may be frozen prior to gel purification; however, yields may be slightly lower. Take note that the labeling reactions can be performed while the gel is being pre-run.

B.2 Labeling Reaction

1. For single labeling reactions, prepare using the following recipe; however, for multiple reactions, a mastermix can be prepared using 10x PNK Buffer, water, and ^{32}PγATP and then subsequently adding specific RNAs and T4 PNK last.

Volume (μL)	Ingredient
2.5	10x PNK Buffer
15.5	MilliQ H_2O
5.0	^{32}PγATP
1.0	PNK Enzyme
1.0	100μM RNA

2. Incubate the reactions for 1 hour at 37˚C. Longer incubations are possible; however, RNA degradation can occur for extensively long incubations.
3. After labeling, add 15μL 2x Urea Loading Buffer (0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol FF, 8M urea, 0.5M EDTA) to each sample.

4. Heat to 90˚C for 2 minutes.
5. Place samples in ice until ready to be loaded into the gel.

B.3 Urea Polyacrylamide Gel Electrophoresis (PAGE)

6. Prepare 70mL of 15% polyacrylamide:bis-acrylamide (29:1) in 1x TBE (45mM Tris-borate, 1mM EDTA). NOTE: These solutions are typically prepared in large quantities before performing this protocol. Also, keep this solution cold to ensure that the gel does not polymerize while pouring.
7. Add 70µL TEMED (N,N,N′,N′-Tetramethylethylenediamine) and 280µL 10% APS (ammonium persulfate).
8. Pour gel
    a. Thoroughly clean the glass plates, spacers and comb with ethanol and water.
    b. Assemble the plates with two spacers along the long sides and secure with large binder clips.
    c. Place the assembled plates horizontally on top of four P200 pipet tip boxes and elevate the top end (with the notch) to ~15˚.
    d. Using a 25mL glass pipet, add polyacrylamide mixture while tapping the glass to ensure that no bubbles are formed and that the solution traverses the length of the glass plates.
    e. Remove whatever was used to elevate the glass plates and insert the comb.
    f. Allow the gel to solidify at room temperature for 30-45 minutes or until fully polymerized.
    g. Gently remove the comb and assemble the gel apparatus.
    h. Fill the top reservoir with 1x TBE to ~1cm above the wells and the bottom reservoir to ~1cm above the level of the bottom of the glass plates. Use a syringe filled with running buffer (1x TBE) to ensure that there are no bubbles at the bottom of the glass plates. Also, use this syringe to flush out the wells.
    i. Pre-run the empty gel at 62 watts for ~1 hour to heat the gel.
9. Load the samples into the desired wells of the gel.

10. Run the gel at 62 watts for ~2.5 hours or until the first dye front reaches ~2cm above the bottom of the gel plates (this may need to be adjusted depending on what size of RNA is being prepared; this preparation assumes a 17-mer).

B.4 Gel Imaging

11. Remove one of the glass plates while keeping the gel on the other plate.
12. Place small (~2mm x 2mm) pieces of nitrocellulose filter around the gel for alignment.
13. Spot radioactive dye (1x Urea Loading Buffer in 50% ethanol spiked with 1uL ^{32}PγATP; will need to be spiked every so often to account for the radioactive decay) onto the pieces of filter paper for alignment and allow them to dry.
14. Cover the gel and glass plate with saran wrap.
15. Place the gel into the phosphoimaging cassette and close.
16. Let the gel expose for ~5 minutes.
17. Quickly remove the gel from the cassette and close it back up.
18. Scan the cassette with the Typhoon Phosphor Imager in the Molecular Core.
19. Save the image as a *.tif file.
20. Print the image in the original scale (Photoshop allows this).

B.5 Gel Extraction & Precipitation

21. Align the gel on top of the image with the alignment spots.
22. Remove the front portion of the saran wrap so that the gel is accessible.
23. Cut out the gel pieces that correspond to the positions of the ASLs on the printed image.
24. Place the gel pieces into individual 1.5mL microfuge tubes.
25. Pipet 100uL of RNA Elution Buffer (0.5M NH_4OAc, 0.1% SDS, 0.1mM EDTA) into each microfuge tube.
26. Place the tip of a P-1000 pipet tip into a Bunsen burner flame to seal it off.

27. Crush the gel into small pieces using the sealed-off end of the pipet tip.
28. Wash the gel pieces from the pipet tip with 400uL of RNA Elution Buffer into the microfuge tube.
29. Place the microfuge tubes into the rotator and rotate for 2 hours at room temperature.
30. Place the rotator into the cold room and rotate for another 18 hours.
31. Remove microfuge tubes from the rotator and spin them for 5 minutes at max speed on the microcentrifuge.
32. Pipet supernatant into a 2mL Costar filter tube and centrifuge for 2 minutes at max speed. NOTE: If gel pieces still contain more than 50% of the radioactivity (check using Geiger counter), steps 28-29 can be repeated using 100µL RNA Elution Buffer.
33. Remove the filter and add 1mL 100% ethanol and 50uL 3M ammonium acetate.
34. Place samples in the freezer for 1 hour.
35. Remove samples from the freezer and centrifuge in the cold room for 15 minutes and aspirate the supernatant.
36. Allow the tubes to sit out with the lid open until the tubes are dry.
37. Store samples in the plexiglass box in the -20˚C freezer.

APPENDIX C. Codon-specific ribosomal A-site filter binding assays and analysis

C.1 Protocol Notes

Prior to performing this protocol, both the control ASL^{Phe} and the ASL of interest (ASL^{Xxx}) should be 5' ^{32}P-end labeled. Each labeled ASL should be at an activity of 2,000-72,000cpm/μL and 2,000-168,000cpm/μL for ASL^{Phe} and ASL^{Xxx}, respectively. The assay is set up to be performed in a 96-well PCR plate. This assay will determine the ability of a single ASL to bind to either of two different codons. It contains both an internal positive control (ASL^{Phe}) and a negative control mRNA to be run in parallel within each assay. Also, the data analysis will be performed using a standard curve to obtain the exact number of picomoles bound per data point. This assay will take a significant portion of a day (at least 4 hours) and should be planned and prepared for ahead of time. It is usually good practice to start the incubators at least one hour prior to beginning the assay. Also, prior to performing this assay, the mRNAs and ribosome activity should be tested by performing this assay with labeled mRNAs and not using ASL or $tRNA^{fMet}$.

C.2 Preparation of Mastermixes

Mastermixes should be prepared according to the spreadsheet that has been set up to calculate amounts of each ingredient and each sheet should be labeled with the correct date and purpose for easy input into a lab notebook and to keep a correct digital copy (Figure 1a). In general, all ASLs, mRNAs and $tRNA^{fMet}$ should be prepared to a concentration of 100uM; however, these values can be altered in the setup area of the protocol (Figure 1b). Enter the correct values for the radioactivity into the CPM column of the setup table (Figure 1b). Also, ensure that the labels are automatically changed accordingly (Figure 1c).

1. Prepare 5x RB buffer (ideally, this should be prepared in advance; Figure 2).
2. Turn on the water bath to 42˚C and the heat block to 80˚C.
3. Remove one ribosome aliquot from the -80˚C freezer and measure the absorbance at A_{260} of 1μL in 1000μL total aqueous volume.

4. Multiply the $OD_{260}/\mu L$ by 23 pmoles/OD_{260} to obtain to obtain the ribosome concentration in μM. Enter this value into the setup table (Figure 1b).
5. Prepare mastermixes on ice.
    a. For the ASL mastermixes, add all ingredients except for 5x buffer, incubate at 80˚C for 2 minutes, allow to cool to room temperature and place on ice prior to addition of 5x buffer.
    b. All other mastermixes should be prepared by adding water and 5x buffer prior to the remaining ingredient.
    c. Label PolyU, control mRNA, mRNA #1 and mRNA #2 with "A", "B", "C" and "D", respectively. These will serve as the basis for the working mixes.

C.3 Assay Protocol

6. Incubate ribosome, $tRNA^{fMet}$ and ASL mastermixes at 42˚C for 10 minutes.
7. Slow cool ribosome, $tRNA^{fMet}$ and ASL mastermixes to 37˚C (usually takes 25-30 minutes).
    a. While these mastermixes are cooling, add the noted amounts of 1x RB buffer to the noted wells of a PCR plate (Figure 3). From here on, keep the PCR plate on ice at all times unless otherwise specified.
    b. Place the PCR plate on ice when finished.
    c. Cut a piece of nitrocellulose filter paper to fit onto a filter binding apparatus. Place into a pipet tip lid and soak with 1x RB buffer.
    d. Place the filter paper, P200 tip box and filter binding apparatus into the refrigerator to equilibrate to 4˚C.
8. Incubate ribosome, $tRNA^{fMet}$ and ASL mastermixes at 37˚C for 10 minutes.
9. Prepare the initial working mixes.
    a. Add 115μL ribosome mastermix to "Tube A" (PolyU)
    b. Add 115μL ribosome mastermix to "Tube B" (Control mRNA)
    c. Add 115μL ribosome mastermix to "Tube C" (mRNA #1)
    d. Add 115μL ribosome mastermix to "Tube D" (mRNA #2)

e. Incubate these mixes for 15 minutes at 37˚C.

10. While working mixes are incubating, prepare ASL mastermix dilutions and add them to the PCR plate.
    a. Add 13µL of ASL^{Phe} mastermix to 117µL 1x RB buffer.
    b. Add 18µL of ASL^{Xxx} mastermix to 162µL 1x RB buffer.
    c. Add the ASL mastermixes to the PCR plate (Figure 4).
11. Add 115µL tRNAfMet mastermix to working mixes B-D and 115µL of 1x RB buffer to working mix A. Incubate these working mixes at 37˚C for 15 minutes.
12. Add 15µL of the noted working mix to the noted PCR plate wells (Figure 5). Pipet up and down to mix.
13. Incubate the PCR plate at 37˚C for 60 minutes.
14. Incubate the PCR plate on ice for 30 minutes.
15. Place the equilibrated filter on the filter binding apparatus and attach a slow vacuum.
16. Wash the filter with 150µL ice cold 1x RB buffer in each well twice.
17. Add 100µL ice cold 1x RB buffer to each well of the PCR plate and immediately pipet the contents onto the filter binding apparatus.
18. Wash 2x with 150µL ice cold 1x RB buffer.
19. Remove the filter from the apparatus (careful not to touch any of the wells) and let sit on KimWipes to dry.
20. Spot 0.5, 1.0, 1.5, 2.0 and 2.5µL of ASL mastermix dilutions in triplicate onto a separate piece of filter paper (standard curve).
21. When dry, wrap both filters in Saran wrap and expose to a phosphor screen for 45 minutes.
22. Remove the filters from the screen and scan using the Typhoon Phosphor Imager (GE Healthcare) using the setting for highest resolution.

C.4 Data Analysis

23. Use ImageQuant TL to quantify each spot and export the data into an MS Excel sheet.
24. Use the standard curve to calculate pmoles bound and subtract the negative control.

25. Use Graphpad Prism to calculate the binding constants using the One-Site Specific binding equation.

C.5 Figures

Date 2/19/2012

Purpose This will be the first of three assays to test the ability of Arg4 UCU triply modified to bind to the rare Arg codons AGA & AGG. **a**

Component	µM	CPM
Ribosome	16.7	---
fMet	100	---
PolyU	10	µg/µL
GCG	100	---
AGA	100	---
AGG	100	---
Phe	100	50000
Arg	100	50000

b

Ribosome Mix		
5pmol/rxn	**5µL/rxn**	
	1	100
Ribosome	0.3	29.9
5x Buffer	1.0	100.0
diH_2O	3.7	370.1
Total	5.0	500.0

tRNA fMet Mix		
25pmol/rxn	**5µL/rxn**	
	1	74
fMet	0.3	18.5
5x Buffer	1.0	74.0
diH_2O	3.8	277.5
Total	5.0	370.0

PolyU Mix		
5ug/rxn	**5µL/rxn**	
	1	23
PolyU	0.5	11.5
5x Buffer	1.0	23.0
diH_2O	3.5	80.5
Total	5.0	115.0

GCG Mastermix		
50pmol/rxn	**5µL/rxn**	
	1	23
100µM GCG	0.5	11.5
5x Buffer	1.0	23.0
diH_2O	3.5	80.5
Total	5.0	115.0

AGA Mastermix		
50pmol/rxn	**5µL/rxn**	
	1	23
100µM AGA	0.5	11.5
5x Buffer	1.0	23.0
diH_2O	3.5	80.5
Total	5.0	115.0

AGG Mastermix		
50pmol/rxn	**5µL/rxn**	
	1	23
100µM AGG	0.5	11.5
5x Buffer	1.0	23.0
diH_2O	3.5	80.5
Total	5.0	115.0

ASL Phe Mastermix		
5pmol/rxn	**5µL/rxn**	
Cold Phe	1200	12.0
Hot Phe	72	1.4
5x Buffer	---	12.0
diH_2O	---	34.6
Total	---	60.0

ASL Arg Mastermix		
5pmol/rxn	**5µL/rxn**	
Cold Arg	2800	28.0
Hot Arg	168	3.4
5x Buffer	---	28.0
diH_2O	---	80.6
Total	---	140.0

c

Figure C1. Example spreadsheet for ASL^{Arg4}_{UCU} to test the ability to decode codons AGA and AGG using a control mRNA with a GCG codon. (a) The date and purpose area allows for proper organization and exact record keeping. **(b)** The setup table allows for editing of concentrations and radioactivity of the different components of the assay. **(c)** The recipes for the mastermixes are laid out in table format. The recipe values are calculated based on a standard setup in which each mastermix is prepared in excess to allow for pipet error.

RB Buffer + 20mM $MgCl_2$		
	1x	**5x**
HEPES (pH 7.0)	50mM	250mM
KCl	30mM	150mM
NH_4Cl	70mM	350mM
DTT	1mM	5mM
EDTA	0.1mM	0.5mM
$MgCl_2$	20mM	100mM

Figure C2. Ribosome binding buffer should be prepared ahead of time. A 5x buffer is typically prepared and diluted to obtain 1x buffer. Both will be needed during the course of the assay. Also, the value of $MgCl_2$ can be adjusted if the initial assays of ribosome activity show low activity.

	1	2	3	4	5	6	7	8	9	10	11	12
A	5.00	5.00	5.00	2.50	2.50	2.50	0.00	0.00	0.00	4.00	4.00	4.00
B	3.50	3.50	3.50	2.50	2.50	2.50	0.00	0.00	0.00			
C	5.00	5.00	5.00	2.50	2.50	2.50	0.00	0.00	0.00	4.00	4.00	4.00
D	3.50	3.50	3.50	2.50	2.50	2.50	0.00	0.00	0.00			
E	5.00	5.00	5.00	2.50	2.50	2.50	0.00	0.00	0.00	4.00	4.00	4.00
F	3.50	3.50	3.50	2.50	2.50	2.50	0.00	0.00	0.00			
G	5.00	5.00	5.00	2.50	2.50	2.50	0.00	0.00	0.00	4.00	4.00	4.00
H	3.50	3.50	3.50	2.50	2.50	2.50	0.00	0.00	0.00			

Figure C3. Addition of 1x RB buffer. For step 6a, the noted amounts of 1x RB buffer (in μL) should be added to the noted wells.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.00	0.00	0.00	2.50	2.50	2.50	5.00	5.00	5.00	1.00	1.00	1.00
B	1.50	1.50	1.50	2.50	2.50	2.50	5.00	5.00	5.00			
C	0.00	0.00	0.00	2.50	2.50	2.50	5.00	5.00	5.00	1.00	1.00	1.00
D	1.50	1.50	1.50	2.50	2.50	2.50	5.00	5.00	5.00			
E	0.00	0.00	0.00	2.50	2.50	2.50	5.00	5.00	5.00	1.00	1.00	1.00
F	1.50	1.50	1.50	2.50	2.50	2.50	5.00	5.00	5.00			
G	0.00	0.00	0.00	2.50	2.50	2.50	5.00	5.00	5.00	1.00	1.00	1.00
H	1.50	1.50	1.50	2.50	2.50	2.50	5.00	5.00	5.00			

Figure C4. Addition of ASL mastermix. For step 10c, the noted amounts (in µL) of ASL mastermix should be added to the noted wells. Rows A-B should be ASL^{Phe} and rows C-H should be ASL^{Xxx}. Also, the 10x dilution mastermixes should go into the wells in the red boxes and the original mastermix should go into the other wells.

	1	2	3	4	5	6	7	8	9	10	11	12
A	A	A	A	A	A	A	A	A	A	A	A	A
B	A	A	A	A	A	A	A	A	A			
C	B	B	B	B	B	B	B	B	B	B	B	B
D	B	B	B	B	B	B	B	B	B			
E	C	C	C	C	C	C	C	C	C	C	C	C
F	C	C	C	C	C	C	C	C	C			
G	D	D	D	D	D	D	D	D	D	D	D	D
H	D	D	D	D	D	D	D	D	D			

Figure C5. Addition of working mixes. For step 12, 15µL of the noted working mixes should be added to the noted wells.

APPENDIX D. Standard molecular dynamics simulations of ASLs using AMBER11

D.1 Protocol Notes

This protocol is designed for the explicit solvent simulation of a heptadecamer (17mer) ASL neutralized with Na^+ ions, minimized, equilibrated for 120ps and simulated under production conditions for a total of 6ns. The protocol can be adjusted to allow for any number of different types of simulations. The main protocol described uses only the SANDER module in AMBER11 using AmberTools1.5 for setup; however, I have appended a section at the end that makes use of GPGPU acceleration using PMEMD. Prior to attempting this protocol, it is advised that the user familiarize themselves with the use of Linux/Unix operating systems. Also, the user should meet with members of the Research IT department at the University at Albany or The RNA Institute to discuss usage of the HPC cluster or GPGPU workstations/cluster, respectively. The protocols herein can be extrapolated to perform other types of simulation either with AMBER11, future implementations of AMBER, GROMACS or NAMD. All text for input files, commands and scripts are denoted by the use of **`Courier New`** (monotype) bold font and are boxed. Commands will be denoted by the use of either a $ for command line or > for *leap* commands (do not actually include the $ or > in the command). All of the simulations presented here can be manipulated by changing variables in the input files. Please consult the AmberTools1.5 or AMBER11 manual prior to making adjustments to your simulations.

D.2 Preparation of Parameter Files

Prior to generation of input files, two things will be necessary: (1) a pdb file containing only your molecule (this can be edited to introduce mutations or modifications) and (2) the proper force field parameter files (only necessary for modified residues). The input files will be created using either the *tleap* or *xleap* modules from AmberTools1.5 for either command line or graphical interfaces, respectively. This protocol will make use of *xleap* to allow for investigation of the structure for anomalies.

1. Plan out your simulations and create organized folders accordingly.
2. Navigate to the path of your pdb file.
3. Execute *xleap* (this command assumes that $AMBERHOME is defined and $AMBERHOME/bin is added to your PATH).

```
$xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB
```

4. Inside of the GUI, use the following commands.
    a. Load any modification parameters that are necessary (an example for s^2C is shown).

```
>loadamberprep 2SC.prepin
>loadamberparams 2SC.frcmod
```

    b. Load the pdb file (ensure that the correct atoms are added, if not correct, you may need to go back and edit the pdb file to be more leap-friendly).

```
>model = loadpdb "XXX.pdb"
```

    c. Check the structure for problems including improper or missing bonds, unbounded atoms or distorted structure.

```
>edit model
```

    d. Neutralize the ASL with Na^+ ions.

```
>addions model Na+ 0
```

    e. Solvate the molecule with explicit TIP3P water molecules in a truncated octahedral water box (the example is for a 10.0Å cutoff).

```
>solvateoct model TIP3PBOX 10.0
```

    f. Save the prmtop and inpcrd files.

```
>saveamberparm model XXX.prmtop XXX.inpcrd
```

D.3 First Minimization

During the first minimization, we are mostly interested in allowing the solvent molecules to minimize. We do this mostly because the solvent molecules were arbitrarily placed by *xleap* and are not in their lowest energy state. To accomplish this, the solute is held fixed under a very high restraint force during minimization.

5. Create the following input file for the first minimization step and give it a name (example: ACG_min1.in).
    a. Input file (example: ACG_min1.in)

```
ACG heptadecamer: initial minimisation with solvent + Na ions
&cntrl
   imin   = 1,
   maxcyc = 1000,
   ncyc   = 500,
   ntb    = 1,
   ntr    = 1,
   cut    = 10
/
Hold the RNA fixed
500.0
RES 1 17
END
END
```

    b. Parameter definitions

```
imin=1                # 1 = minimisation turned on.
maxcyc=1000           # Total number of steps. 1000 as default.
ncyc=500              # Steps of steepest descent. 500 as default.
ntb=1                 # 1 = constant volume periodic boundaries.
ntr=1                 # 1 = Use position restraints based on GROUP.
cut=10                # Angstrom cutoff.
Held RNA fixed        # Type of molecule. RNA as default.
Force Const. 500.0    # Force constant. 500 kcal/mol/A^2 as default.
Fixed Res 1-17        # Total number of residues to restrain.
```

6. Execute the first minimization using the following command.
    a. For serial minimization:

```
$sander -O -i ACG_min1.in -o ACG_min1.out -p ACG.prmtop -c
ACG.inpcrd -r ACG_min1.rst -ref ACG.inpcrd
```

b. For parallel minimization on the cluster (the –N flag denotes the number of nodes, -n 8 denotes the number of processors, -J is for the name of the job on the cluster, –p to denote that a batch job is being used and --wrap is for the execution command) **NOTE:** The –N and –n flags can be adjusted. For instance minimizations do not require 8 processors to be efficient, 2 or 4 should be plenty.

```
$sbatch –N 1 –n 8 –J ACG –p batch --wrap="mpirun –np 8 sander.MPI -O -i ACG_min1.in -o ACG_min1.out -p ACG.prmtop -c ACG.inpcrd -r ACG_min1.rst -ref ACG.inpcrd"
```

c. Parallel simulations can be monitored using the command: `$squeue`

D.4 Second Minimization

Following minimization of the solvent, it is important to allow for minimization of the entire system. To accomplish this, the solute is now released from any restraints and the entire system is allowed to minimize. During this step, a very large number of cycles are used (15,000) to ensure full minimization of the system. This is probably overkill; however, not fully minimizing the system can result in simulation artifacts later. Here, where very little computational resources are required per cycle, is the best place to use excessive cycles.]

7. Once the first minimization is finished, create the input file for the second minimization step and give it a name (example: ACG_min2.in).
   a. Input file (example ACG_min2.in)

```
ACG heptadecamer: initial minimisation, whole system
&cntrl
   imin   = 1,
   maxcyc = 15000,
   ncyc   = 1000,
   ntb    = 1,
   ntr    = 0,
   cut    = 10
/
```

   b. Parameter definitions

```
imin=1          # 1 = minimisation turned on.
maxcyc=15000    # Total number of steps. 15000 as default.
ncyc=1000       # Steps of steepest descent. 1000 as default.
ntb=1           # 1 = Constant volume periodic boundaries.
ntr=0           # 0 = Do not use position restraints.
cut=10          # Use the same cutoff
```

8. Execute the second minimization using the following command.
    a. For serial minimization:

```
$sander -O -i ACG_min2.in -o ACG_min2.out -p ACG.prmtop -c
ACG_min1.rst -r ACG_min2.rst
```

    b. For parallel minimization on the cluster (the –N flag denotes the number of nodes, -n 8 denotes the number of processors, -J is for the name of the job on the cluster, –p to denote that a batch job is being used and --wrap is for the execution command) **NOTE:** The –N and –n flags can be adjusted. For instance minimizations do not require 8 processors to be efficient, 2 or 4 should be plenty.

```
$sbatch -N 1 -n 8 -J ACG -p batch --wrap="mpirun -np 8 sander.MPI -
O -i ACG_min2.in -o ACG_min2.out -p ACG.prmtop -c
ACG_min1.rst.inpcrd -r ACG_min2.rst"
```

    c. Parallel simulations can be monitored using the command: `$squeue`

D.5 First Equilibration

During the first equilibration, it is necessary to allow the system to slowly heat up from 0-300k. During this time, it is possible for the system to undergo very drastic, unwanted structural transitions. Therefore, it is very important to ensure that the solute (ASL) is held fixed with a small restraint force.

9. Once minimizations are complete, create the input file for the first equilibration step and give it a name (example: ACG_equ1.in).

a. Input file (example ACG_equ1.in)

```
ACG heptadecamer: 20.000 picoseconds MD with restraint on RNA
&cntrl
   imin     = 0,
   irest    = 0,
   ntx      = 1,
   ntb      = 1,
   ntr      = 1,
   cut      = 10,
   ntc      = 2,
   ntf      = 2,
   tempi    = 0.0,
   temp0    = 300.0,
   ntt      = 3,
   ig       = -1,
   gamma_ln = 1.0,
   nstlim   = 10000,
   dt       = 0.002,
   ntpr     = 100,
   ntwx     = 100,
   ntwr     = 1000,
/
Hold the RNA fixed with weak restraints
10.0
RES 1 17
END
END
```

b. Parameter definitions

```
imin=0                # 0 = minimisation turned off.
irest=0               # 0 = restart MD using no initial vel info.
ntx=1                 # 1 = no initial velocity info. Use inpcrd.
ntb=1                 # 1 = use constant volume periodic boundaries.
cut=10                # Use the same cutoff.
ntr=1                 # 1 = use position restraints based on GROUP.
ntc=2                 # 2 = use SHAKE for all H-containing bonds.
ntf=2                 # 2 = use SHAKE for all H-containing bonds.
tempi=0.0             # Initial temperature. 0.0k as default.
temp0=300.0           # Final temperature. 300.0k as default.
ntt=3                 # Use Langevin dynamics to control temp.
ig=-1                 # -1 = use a random seed for initial traj.
gamma_ln=1.0          # Use 1.0ps^-1 collision frequency.
nstlim=10000          # Total number of md steps. 10000 as default.
dt=0.002              # Time step. 2fs as default (0.002ps).
ntpr=100              # Step interval to write to output file.
ntwx=100              # Step interval to write to traj file.
ntwr=1000             # Step interval to write to restart file.
type=$m1_type         # Use the same molecule type.
fconst=10.0           # Force constant. 10 kcal/mol/A^2 as default.
residues=1-17         # Use the same number of residues.
```

10. Execute the first equilibration using the following command.
    a. For serial equilibration:

```
$ sander -O -i ACG_equ1.in -o ACG_equ1.out -p ACG.prmtop -c
ACG_min2.rst -r ACG_equ1.rst -x ACG_equ1.mdcrd -ref ACG_min2.rst
```

    b. For parallel equilibration on the cluster (the –N flag denotes the number of nodes, -n 8 denotes the number of processors, -J is for the name of the job on the cluster, –p to denote that a batch job is being used and --wrap is for the execution command) **NOTE:** The –N and –n flags can be adjusted. For instance the first equilibration does not require 8 processors to be efficient, 4 should be plenty.

```
$sbatch –N 1 –n 8 –J ACG –p batch --wrap="mpirun –np 8 sander.MPI -
O -i ACG_equ1.in -o ACG_equ1.out -p ACG.prmtop -c ACG_min2.rst -r
ACG_equ1.rst -x ACG_equ1.mdcrd -ref ACG_min2.rst"
```

    c. Parallel simulations can be monitored using the command: `$squeue`

D.6 Second Equilibration

The second equilibration is performed under production conditions in order to allow for the system to have time to fully equilibrate at 300k.

11. Once the first equilibration is complete, create the input file for the second equilibration step and give it a name (example: ACG_equ2.in).
    a. Input file (example ACG_equ2.in)

```
ACG heptadecamer: 100.000 picoseconds MD with no restraints, whole
system
&cntrl
   imin     = 0,
   irest    = 1,
   ntx      = 5,
   ntb      = 2,
```

```
   pres0    = 1.0,
   ntp      = 1,
   taup     = 2.0
   ntr      = 0,
   cut      = 10,
   ntf      = 2,
   ntc      = 2,
   tempi    = 300.0,
   temp0    = 300.0,
   ntt      = 3,
   gamma_ln = 1.0,
   nstlim   = 50000,
   dt       = 0.002,
   ntpr     = 100,
   ntwx     = 100,
   ntwr     = 1000
/
```

b. Parameter definitions

```
imin=0          # 0 = minimisation turned off.
irest=1         # 1 = restart MD using previous vel info.
ntx=5           # 5 = use initial vel and box info from rst file.
ntb=2           # 2 = use constant pressure periodic boundaries.
pres0=1.0       # Constant pressure value in atm. 1.0 as default.
ntp=1           # 1 = isotropic pos. scaling to maintain pressure.
taup=2.0        # Relaxation time in ps. 2.0 as default.
cut=10          # Use the same cutoff.
ntr=0           # 0 = do not use position restraints.
ntc=2           # 2 = use SHAKE for all H-containing bonds.
ntf=2           # 2 = use SHAKE for all H-containing bonds.
tempi=300.0     # Initial temperature. 300.0k as default.
temp0=300.0     # Final temperature. 300.0k as default.
ntt=3           # Use Langevin dynamics to control temp.
gamma_ln=1.0    # Use 1.0ps^-1 collision frequency.
nstlim=50000    # Total number of md steps. 50000 as default.
dt=0.002        # Time step. 2fs as default (0.002ps).
ntpr=100        # Step interval to write to output file.
ntwx=100        # Step interval to write to traj file.
ntwr=1000       # Step interval to write to restart file.
```

12. Execute the second equilibration using the following command.

c. For serial equilibration:

```
$ sander -O -i ACG_equ2.in -o ACG_equ2.out -p ACG.prmtop -c
ACG_equ1.rst -r ACG_equ2.rst -x ACG_equ2.mdcrd
```

d. For parallel equilibration on the cluster (the –N flag denotes the number of nodes, -n 8 denotes the number of processors, -J is for the name of the job on the cluster, –p to denote that a batch job is being used and --wrap is for the execution command) **NOTE:** The –N and –n flags can be adjusted. For instance the second equilibration will require 4-8 processors to be efficient.

```
$sbatch –N 1 –n 8 –J ACG –p batch --wrap="mpirun –np 8 sander.MPI -
O -i ACG_equ2.in -o ACG_equ2.out -p ACG.prmtop -c ACG_equ1.rst -r
ACG_equ2.rst -x ACG_equ2.mdcrd"
```

e. Parallel simulations can be monitored using the command: `$squeue`

D.7 CPU-based Production Simulations using SANDER.MPI

Production simulations should be run using the same parameters as those used for the equilibration; however, they are typically on a larger scale. It is typically advisable to use the first 1ns of production simulation as an extended equilibration. ASLs have typically been simulated for ~5ns; however, as computational resources become more available, this number should grow. Please plan your simulations according to accepted practices described in the literature.

13. Once both minimizations and equilibrations are complete, create the input file for the production step and give it a name (example: ACG_pro1.in) NOTE: The example input file created below is for a 1ns simulation. You can change the length of the simulation by adjusting the nstlim parameter.

    f. Input file (example ACG_pro1.in)

```
ACG heptadecamer: 1 nanoseconds production MD with no restraints,
whole system
&cntrl
   imin     = 0,
   ig       = -1,
   irest    = 1,
   ntx      = 5,
   ntb      = 2,
   pres0    = 1.0,
```

```
   ntp      = 1,
   taup     = 2.0
   ntr      = 0,
   cut      = 10,
   ntf      = 2,
   ntc      = 2,
   tempi    = 300.0,
   temp0    = 300.0,
   ntt      = 3,
   gamma_ln = 1.0,
   nstlim   = 500000,
   dt       = 0.002,
   ntpr     = 100,
   ntwx     = 100,
   ntwr     = 1000
/
```

g. Parameter definitions

```
imin=0          # 0 = minimisation turned off.
ig=-1           # -1 = use a random seed for initial traj.
irest=1         # 1 = restart MD using previous vel info.
ntx=5           # 5 = use initial vel and box info from rst file.
ntb=2           # 2 = use constant pressure periodic boundaries.
pres0=1.0       # Constant pressure value in atm. 1.0 as default.
ntp=1           # 1 = isotropic pos. scaling to maintain pressure.
taup=2.0        # Relaxation time in ps. 2.0 as default.
cut=10          # Use the same cutoff.
ntr=0           # 0 = do not use position restraints.
ntc=2           # 2 = use SHAKE for all H-containing bonds.
ntf=2           # 2 = use SHAKE for all H-containing bonds.
tempi=300.0     # Initial temperature. 300.0k as default.
temp0=300.0     # Final temperature. 300.0k as default.
ntt=3           # Use Langevin dynamics to control temp.
gamma_ln=1.0    # Use 1.0ps^-1 collision frequency.
nstlim=50000    # Total number of md steps. 50000 as default.
dt=0.002        # Time step. 2fs as default (0.002ps).
ntpr=100        # Step interval to write to output file.
ntwx=100        # Step interval to write to traj file.
ntwr=1000       # Step interval to write to restart file.
```

14. Execute production simulation using the following command.

h. For serial production simulation:

```
$ sander -O -i ACG_equ2.in -o ACG_equ2.out -p ACG.prmtop -c
ACG_equ1.rst -r ACG_equ2.rst -x ACG_equ2.mdcrd
```

i. For parallel production simulation on the cluster (the –N flag denotes the number of nodes, -n 8 denotes the number of processors, -J is for the name of the job on the cluster, –p to denote that a batch job is being used and --wrap is for the execution command) **NOTE:** The –N and –n flags can be adjusted. For instance the production simulations will require at least 8 processors to be efficient.

```
$sbatch –N 1 –n 8 –J ACG –p batch --wrap="mpirun –np 8 sander.MPI -O -i ACG_equ2.in -o ACG_equ2.out -p ACG.prmtop -c ACG_equ1.rst -r ACG_equ2.rst -x ACG_equ2.mdcrd"
```

j. Parallel simulations can be monitored using the command: `$squeue`

D.7 GPGPU-based Production Simulations using PMEMD.CUDA

GPGPU-based production simulations using PMEMD should be run using the same parameters as those used for the CPU-based simulation; however, GPGPU simulations are designed and will scale better with larger systems and longer time scales. It is typically advisable to use the first 1ns of production simulation as an extended equilibration even when using PMEMD. You should be able to improve your simulation speed for anything >12k atoms (in explicit solvent).

15. Once both minimizations and equilibrations are complete, create the input file for the production step and give it a name (example: ACG_pro1.in) NOTE: The example input file created below is for a 50ns simulation. You can change the length of the simulation by adjusting the nstlim parameter.

    k. Input file (example ACG_pro1.in)

```
ACG heptadecamer: 1 nanoseconds production MD with no restraints, whole system
&cntrl
   imin     = 0,
   ig       = -1,
   irest    = 1,
   ntx      = 5,
```

```
   ntb      = 2,
   pres0    = 1.0,
   ntp      = 1,
   taup     = 2.0
   ntr      = 0,
   cut      = 10,
   ntf      = 2,
   ntc      = 2,
   tempi    = 300.0,
   temp0    = 300.0,
   ntt      = 3,
   gamma_ln = 1.0,
   nstlim   = 25000000,
   dt       = 0.002,
   ntpr     = 100,
   ntwx     = 100,
   ntwr     = 1000
/
```

l. Parameter definitions

```
imin=0          # 0 = minimisation turned off.
ig=-1           # -1 = use a random seed for initial traj.
irest=1         # 1 = restart MD using previous vel info.
ntx=5           # 5 = use initial vel and box info from rst file.
ntb=2           # 2 = use constant pressure periodic boundaries.
pres0=1.0       # Constant pressure value in atm. 1.0 as default.
ntp=1           # 1 = isotropic pos. scaling to maintain pressure.
taup=2.0        # Relaxation time in ps. 2.0 as default.
cut=10          # Use the same cutoff.
ntr=0           # 0 = do not use position restraints.
ntc=2           # 2 = use SHAKE for all H-containing bonds.
ntf=2           # 2 = use SHAKE for all H-containing bonds.
tempi=300.0     # Initial temperature. 300.0k as default.
temp0=300.0     # Final temperature. 300.0k as default.
ntt=3           # Use Langevin dynamics to control temp.
gamma_ln=1.0    # Use 1.0ps^-1 collision frequency.
nstlim=25000000 # Total number of md steps. 25000000 as default.
dt=0.002        # Time step. 2fs as default (0.002ps).
ntpr=100        # Step interval to write to output file.
ntwx=100        # Step interval to write to traj file.
ntwr=1000       # Step interval to write to restart file.
```

16. Currently, there is only the capability of using a single GPGPU (Tesla C2075) at a time on a single workstation (tesla1 or tesla2). Execute production simulation using the following command.

    m. For serial production simulation:

```
$ pmemd.cuda -O -i ACG_equ2.in -o ACG_equ2.out -p ACG.prmtop -c
ACG_equ1.rst -r ACG_equ2.rst -x ACG_equ2.mdcrd
```

D.8 Data Analysis

Analysis scripts will vary depending on what you are trying to see in your simulations. In this section, I have included an example analysis script that calculates the backbone, mass weighted, non-weighted and heavy atom RMSD vs time. Also, the overall molecular and per residue RMSF and B-factors are calculated. Finally, the script has one example each of how to calculate a dihedral angle, sugar pucker and atomic distance vs time. The script then creates an input file for the creation of a pdb file that contains an ensemble of snapshot structures taken over the course of 5ns at intervals of 10ps. The script then invokes the *ptraj* module from AmberTools1.5 to execute the input files. Last, the process_mdout.perl script creates detailed tables of the contents of the sander output files. Of course, more detailed analyses can be performed by simply adding more parameters to the input files or by determining specific properties to examine by investigating the resulting pdb file.

```
#!/bin/bash

if [[ -d MD2 ]] ; then
 rm -rf MD2
fi

mkdir MD2
cd MD2

if [[ -d RMSD ]] ; then
 rm -rf RMSD
fi

mkdir RMSD

##### Analysis of Trajectory File #####

cat > Soln_rms-First.in << EOF
trajin ACG_So_P2.mdcrd
center :1-17 mass origin
image origin center
rms first out RMSD/Soln_BB.rms @P,O3',O5',C3',C4',C5' time 0.2
rms first out RMSD/Soln_NW.rms :1-17 time 0.2
```

```
rms first mass out RMSD/Soln_MW.rms :1-17 time 0.2
rms first out RMSD/Soln_HA.rms :1-17&!@/H time 0.2
atomicfluct out RMSD/RMSF_All.rms :1-17&!@/H bymask
atomicfluct out RMSD/RMSF_Res.rms :1-17&!@/H byres
atomicfluct out RMSD/Bfact_All.rms :1-17&!@/H bymask bfactor
atomicfluct out RMSD/Bfact_Res.rms :1-17&!@/H byres bfactor
dihedral chi32 :6@O4' :6@C1' :6@C2 :6@N1 out RMSD/Soln_xC32.dih
dihedral chi36 :10@O4' :10@C1' :10@C4 :10@N9 out RMSD/Soln_xG36.dih
pucker   puckA34   :8@C1'   :8@C2'   :8@C3'   :8@C4'   :8@O4'   out
RMSD/Soln_A34.pck
distance C32A38 :6@N3 :12@N1 out RMSD/Soln_C32A38.dst
EOF

cat > Soln_pdb.in << EOF
trajin ACG_So_P2.mdcrd 50 2500000 50
center :1-17 mass origin
image origin center
strip :WAT
strip :Na+
trajout traj.pdb pdb nobox append
EOF

ptraj  /home/will/Arg12/Solution/leap_files/ACG.prmtop  <  Soln_rms-
First.in
ptraj /home/will/Arg12/Solution/leap_files/ACG.prmtop < Soln_pdb.in


##### Creation of Summary data from .out files #####

if [[ -d Analysis ]] ; then
 rm -rf Analysis
fi

mkdir Analysis

cp ACG_So_P2.out Analysis/
cp /home/will/Downloads/process_mdout.perl Analysis/
cd Analysis/

process_mdout.perl ACG_So_P2.out

done
```

APPENDIX E. Targeted molecular dynamics simulations of ASLs using AMBER11

E.1 Protocol Notes

Targeted molecular dynamics is a method of biasing a system toward a particular conformation. In doing this, it is possible to determine hypothetical conformational pathways between both small and large conformational changes. The software does this by introducing an energy penalty that is based on the RMSD between the current conformation and a reference structure (the one that you are biasing the system to adopt). There are a few ways to accomplish this task, one is by simply setting the target RMSD to zero and optimizing the force factor. While this works very well for small conformational changes, there is a drawback that the force on your molecule will not be constant throughout the simulation. Also, if you are attempting to compare different systems, you may find that some systems have different optimum force constants, rendering your comparison dependent on an outside bias. Another way to accomplish this is to use a slightly larger force constant, but with enough leeway to allow for some sampling occur at each time point. While doing this, you can linearly reduce the target RMSD over time. This will keep the force more constant, will alleviate the possibility of getting stuck in energy wells and you can use the same settings for multiple systems and adequately compare them. The force constant should still be optimized to allow for proper conformational sampling during the simulation.

E.2 Input File Preparation

A key necessity of this type of simulation is that the total number of atoms in the starting structure and the reference structure (including solvent atoms) must be exactly the same. They do not all need to be included in the mask, but they should be the same nonetheless. Also, both structures should be fully minimized prior to performing targeted molecular dynamics. Typically, I will perform minimization, equilibration and an initial 1ns of production simulation and use the resulting structure as the starting point for targeted molecular dynamics (See Appendix D2-D7). Since the starting structure is the base structure,

I typically use the number of atoms in this structure and create the input files for the reference structure to conform to this exact number of atoms.

1. Plan out your simulations and create organized folders accordingly.
2. Create input files, minimize, equilibrate and run the first 1ns of production simulation based on the protocol for standard molecular dynamics simulation (Appendix D2-D7).
3. Navigate to the path of your reference pdb file (use a pdb file that is generated from the explicit solvent minimized min2.rst file and strip off all water and ion molecules).
4. Execute *xleap* (this command assumes that $AMBERHOME is defined and $AMBERHOME/bin is added to your PATH).

```
$xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB
```

5. Inside of the GUI, use the following commands.
    a. Load any modification parameters that are necessary (an example for s^2C is shown).

```
>loadamberprep 2SC.prepin
>loadamberparams 2SC.frcmod
```

    b. Load the pdb file (ensure that the correct atoms are added, if not correct, you may need to go back and edit the pdb file to be more leap-friendly).

```
>model = loadpdb "XXX.pdb"
```

    c. Check the structure for problems including improper or missing bonds, unbounded atoms or distorted structure.

```
>edit model
```

    d. Neutralize the ASL with Na^+ ions.

```
>addions model Na+ 0
```

e. Solvate the molecule with explicit TIP3P water molecules in a truncated octahedral water box (the example is for a 10.0Å cutoff).

```
>solvateoct model TIP3PBOX 10.0
```

i. Here, the value of 10.0 will most likely not give you the exact number of water molecules as that of the starting structure.

ii. First optimize this by trying different values for the box cutoff and determining which value gives you the least amount of EXCESS water molecules.

iii. Now, after you have added the water box, delete the excess water molecules using the following command to remove residue number 3654: `remove model model.3654`

f. Save the prmtop and inpcrd files.

```
>saveamberparm model XXX.prmtop XXX.inpcrd
```

E.3 Targeted Molecular Dynamics Simulation

For the targeted molecular dynamics simulations, you will need to optimize the force parameter and determine what length of time will be good for your simulations. For the example here, I will be using a weighting that will linearly decrease the target RMSD from 3.5Å (the initial is ~3.6Å) to 0Å over a time of 4ns. The simulation will then be left to run for an additional 1ns with the target RMSD set at 0Å.

6. Create the input file for the targeted molecular dynamics and give it a name (example: ACG-Ri_tgtmd.in).

a. Input file (example: ACG-Ri_tgtmd.in)

```
ASL Arg12 ACG Heptadecamer: 5ns Targeted MD from Solution to
Ribosome-bound
 &cntrl
        imin = 0, irest = 1, ntx = 5,
        ntb = 2, pres0 = 1.0, ntp = 1,
        taup = 2.0, ig = -1,
        cut = 10, ntr = 0,
        ntc = 2, ntf = 2,
```

```
        tempi = 300.0, temp0 = 300.0,
        ntt = 3, gamma_ln = 1.0,
        nstlim = 2500000, dt = 0.002,
        ntpr = 100, ntwx = 100, ntwr = 1000,
        irest=1, nmropt=1,
        itgtmd=1, tgtrmsd=3.5, tgtmdfrc=1.00,
        tgtfitmask=":1-17",
        tgtrmsmask=":1-17 & !@H=",
 /
 &wt
        TYPE='TGTRMSD', istep1 =1, istep2 = 2000000,
        value1 = 3.5, value2 = 0.0,
/
&wt
        TYPE='TGTRMSD', istep1 =2000000, istep2 = 2500000,
        value1 = 0.0, value2 = 0.0,
/
 &wt
   type="END",
 /
```

b. Parameter description

```
imin=0          # 0 = minimisation turned off.
ig=-1           # -1 = use a random seed for initial traj.
irest=1         # 1 = restart MD using previous vel info.
ntx=5           # 5 = use initial vel and box info from rst file.
ntb=2           # 2 = use constant pressure periodic boundaries.
pres0=1.0       # Constant pressure value in atm. 1.0 as default.
ntp=1           # 1 = isotropic pos. scaling to maintain pressure.
taup=2.0        # Relaxation time in ps. 2.0 as default.
cut=10          # Use the same cutoff.
ntr=0           # 0 = do not use position restraints.
ntc=2           # 2 = use SHAKE for all H-containing bonds.
ntf=2           # 2 = use SHAKE for all H-containing bonds.
tempi=300.0     # Initial temperature. 300.0k as default.
temp0=300.0     # Final temperature. 300.0k as default.
ntt=3           # Use Langevin dynamics to control temp.
gamma_ln=1.0    # Use 1.0ps^-1 collision frequency.
nstlim=2500000  # Total number of md steps. 2500000 as default.
dt=0.002        # Time step. 2fs as default (0.002ps).
ntpr=100        # Step interval to write to output file.
ntwx=100        # Step interval to write to traj file.
ntwr=1000       # Step interval to write to restart file.
nmropt=1        # 1 = Allow restraints or wt changes to be read.
itgtmd=1        # 1 = Set targeted MD on
tgtrmsd=3.5     # Set the initial target RMSD. 3.5 as default.
tgtmdfrc=1.00   # Set the target MD force to 1.00
tgtfitmask=":1-17"   # Defines the target atoms for rms fitting.
tgtrmsmask=":1-17 & !@H="  # Defines the atoms for calculation of
RMSD.
```

17. Execute the targeted MD simulation using the following command.

    n. For serial targeted MD simulation:

```
$ sander -O -i ACG-Ri_tgtmd.in -o ACG-Ri_TMD.out -p ACG.prmtop -c
ACG_So_P1.rst   -r   ACG-Ri_TMD.rst   -x   ACG-Ri_TMD.mdcrd   -ref
Rib_ACG.inpcrd
```

    o. For parallel targeted MD simulation on the cluster (the –N flag denotes the number of nodes, -n 8 denotes the number of processors, -J is for the name of the job on the cluster, –p to denote that a batch job is being used and --wrap is for the execution command) **NOTE:** The –N and –n flags can be adjusted. For instance the targeted MD simulations will require at least 8 processors to be efficient.

```
$sbatch –N 1 –n 8 –J ACG –p batch --wrap="mpirun –np 8 sander.MPI -
O   -i   ACG-Ri_tgtmd.in   -o   ACG-Ri_TMD.out   -p   ACG.prmtop   -c
ACG_So_P1.rst   -r   ACG-Ri_TMD.rst   -x   ACG-Ri_TMD.mdcrd   -ref
Rib_ACG.inpcrd"
```

    p. Parallel simulations can be monitored using the command: `$squeue`

7. Targeted MD simulations can be analyzed using ptraj as described for standard MD simulations (Appendix D8).

APPENDIX F – The RNA modification database, RNAMDB: 2011 update

William A. Cantara, Pamela F. Crain, Jef Rozenski, James A. McCloskey, Kimberly A. Harris, Xiaonong Zhang, Franck A. P. Vendeix, Daniele Fabris, Paul F. Agris

Published in Nucleic Acids Research, 2011, Vol. 39 (Database issue), p. D195-201.

F.1 Statement of Project Contribution

This project was solely the responsibility of me. While other authors were instrumental in teaching me the history of the database, how to understand the flow of the database programming architecture and for migration of the database onto University at Albany servers, I wrote the paper, designed the tables and figures and assisted with the development of the database as a highly customizable template for future upgrading into a web portal. With the help of Leslie Zuker and Andrew Wright, my contribution to this project continues with the design and implementation of the most up-to-date version of the database (http://mods.rna.albany.edu/) that includes new functionality: updated references with direct links to primary literature, updated text and acknowledgements, improved search and content management functionality and enhanced graphics.

F.2 Abstract

Since its inception in 1994, *The RNA Modification Database* (RNAMDB, http://rna-mdb.cas.albany.edu/RNAmods/) has served as a focal point for information pertaining to naturally occurring RNA modifications. In its current state, the database employs an easy-to-use, searchable interface for obtaining detailed data on the 109 currently known RNA

modifications. Each entry provides the chemical structure, common name and symbol, elemental composition and mass, CA registry numbers and index name, phylogenetic source, type of RNA species in which it is found, and references to the first reported structure determination and synthesis. Though newly transferred in its entirety to The RNA Institute, the RNAMDB continues to grow with two notable additions, agmatidine and 8-methyladenosine, appended in the past year. *The RNA Modification Database* is staying up-to-date with significant improvements being prepared for inclusion within the next year and the following year. The expanded future role of The RNA Modification Database will be to serve as a primary information portal for researchers across the entire spectrum of RNA-related research.

F.3 Introduction

The chemical composition of an RNA molecule allows for its inherent ability to play many roles within biological systems. This ability is further enhanced through the site selected addition of the 109 currently known post-transcriptional modifications catalyzed by specific RNA modification enzymes [1]. These naturally-occurring modifications are found in all three major RNA species (tRNA, mRNA and rRNA) in all three primary phylogenetic domains (archaea, bacteria and eukarya) as well as in a handful of other RNA species such as snRNA [2-5] (Table 1). The modifications are one of the most evolutionarily conserved properties of RNAs. Due in large part to comprehensive investigations into the structural and functional roles of modified nucleotides in tRNA, significant advancements have been achieved in our understanding of the various roles played by these modifications [2,6-9]. The need to provide a comprehensive, searchable database to house this wealth of knowledge led to the first iteration of *The RNA Modification Database* (RNAMDB) in 1994 [10]. The current version of the database, now housed at The RNA Institute at the University at Albany SUNY, contains all naturally-occurring, RNA-derived modified ribonucleosides for which the chemical structures are known. The RNAMDB provides a user-friendly, searchable interface that directs the user to a detailed information page for each database entry (Fig. 1).

Users are invited to submit comments regarding existing entries, including errors and omissions, as well as suggestions for improvements to the following email address: rnamdb@albany.edu.

F.4 Database Contents

A comprehensive introduction page provides a detailed overview of the current state of the RNAMDB, an in-depth description of the entries and a page specifically dedicated to modifications found in ribosomal RNA. The overview section of the introduction page provides a focus on commonly used symbols representing RNA modifications, reasons for exclusion of certain modified nucleosides and an extensive listing of reviews for further reference.

Presently, the information page for each RNAMDB entry contains the following details (Fig. 2a-g).

a) Chemical structure of the modified nucleoside: this structure is simplistic and is not meant to imply the preferred orientation of dihedral angles or sugar pucker

b) Common name and symbol

c) The searchable Chemical Abstracts registry numbers and index name, which may provide stereochemical information not shown in the chemical structure

d) Elemental composition and molecular weight

e) Phylogenetic source of the RNA (ie. archaea, bacteria or eukarya) and the type of RNA species in which it is found (ie. tRNA, rRNA, etc.)

f) Literature references for the first reported structure and synthesis. Reports of more recent synthesis or refinements can be found using the Chemical Abstracts registry numbers noted above

g) Comments pertaining to any of the information provided on the entry information page provided by database curators

An example of a single entry information page for 2-methyladenosine has been provided (Fig. 2). The RNAMDB contains 109 such entries that consist of 93 found in tRNAs, 31 in rRNA, 13 in mRNA and 14 in other RNA species such as snRNA, snoRNA and miRNA [3,4,6] (Table 1).

In addition to the Introduction and Search pages, the website contains a compilation of links to other helpful databases and search tools, and provides access to a suite of programs designed to assist in the identification and characterization of nucleic acids by mass spectrometry (Masspec Toolbox). These dedicated programs are capable of calculating the mass of ribo- and deoxyribo-oligonucleotides from their sequence, as well as predicting the masses of products obtained by submitting a certain species to tandem mass spectrometry (MS/MS), or to digestion by endo- and exonucleases. Other tools can calculate nucleotide and elemental composition from a certain mass and provide the corresponding isotopic distribution. All of these tools allow for users to account for the presence of chemically modified nucleotide residues.

F.5 Recent Additions

Since the publication of the last database overview [3], fourteeen new entries have been added to the RNAMDB, including agmatidine and 8-methyladenosine, which have both been added within the past year [11-25] (Table 2). Agmatidine (C^+) was recently found in archaeal AUA-decoding $tRNA^{Ile2}$, where it performs a function similar to lysidine, a modified cytidine residue containing lysine in place of the C2-oxo group, in bacteria [23]. In *E. coli*, the presence of lysidine at the first position of the anticodon allows for AUA-decoding $tRNA^{Ile2}$ to recognize A and not G in the third position of the codon [26]. Agmatidine has a structure very similar to lysidine in which agmatine (decarboxy-arginine) replaces the C2-oxo group of cytidine (Fig. 3a) and, at the first position of the anticodon, performs the same function of

differentiating between A and G at the third position of the codon in *H. marismortui* [23]. The discrimination between A and G that results from the presence of lysidine or agmatidine is hypothesized to play an important role in the stabilization of potential positively charged tautomeric structures at neutral pH, which are able to selectively base pair with A [23].

Similar to 2-methyladenosine (m^2A) [27], 8-methyladenosine (m^8A, Fig. 3b) is present in the large subunit of bacterial ribosomes, specifically in the peptidyl transferase center at position A2503 of 23S rRNA as confirmed by comparison to the chemically synthesized mononucleoside [24]. Whereas the m^2A chemistry shows only a small amount of antibiotic resistance [28], the m^8A moiety confers extensive resistance to five major classes of peptidyl transferase-targeting antibiotics [29]. Interestingly, the m^8A mononucleoside was also synthesized in 1993 and shown to be a potent inhibitor of vaccinia virus [25]. This modification is a clear example of why it is important to understand the purpose of RNA modifications in living systems and how they can be utilized for therapeutic purposes. As an interesting side note, the presence of dimethylated A2503 (m^2m^8A) was also reported [24]; however, this modification has yet to be fully characterized or structurally confirmed.

F.6 The RNA Institute Takes Over as Curator

As of January 2009, upkeep and supervision of the RNAMDB has been turned over to The RNA Institute. Already, this transfer has resulted in a major layout makeover in May 2009 to match the theme of the newly established institute. As a part of its mission, The RNA Institute will be expected to become a proposed hub of cutting edge RNA research and a leading provider of structural and functional information in the field of RNA modifications. For this reason, The RNA Institute will provide the logical base for new enhancements aimed at expanding the focus of the current infrastructure to cover other important aspects of modern RNA research. These enhancements will be embodied by the proposed creation of a portal that will constitute a primary resource for researchers across the entire spectrum of RNA-related disciplines. The portal will not only include the current RNAMDB and

associated tools, but will also contain a collection of pertinent experimental protocols, fundamental information about RNA structure and function and additional informatics tools. From this new infrastructure, users will be able to access other informatics resources for RNA science, including different repositories of experimental protocols, force field parameters for molecular dynamics simulations [30], and sister databases dedicated to non-natural RNA modifications. The infrastructure for all of the above-mentioned improvements will be implemented on a rolling basis as they are completed during 2011. A full-featured RNA portal is a goal to be reached in 2012.

The experimental protocols made available by the RNAMDB portal will comprise established procedures that are common praxis in a typical RNA lab, but will also include new cutting-edge methods developed by the broader RNA community. The RNA Institute will be expected to play an important role in populating this section, due to the research interests of its members and the nature of their collaborations, which involve the development of enabling technologies based on a wide range of experimental approaches, including nuclear magnetic resonance, X-ray crystallography, single-molecule spectroscopy and, with the recent opening of The RNA Mass Spectrometry Center, high-resolution mass spectrometry. However, any investigator engaged in RNA research will be encouraged to share his/her favorite techniques with the broader RNA community through the portal. Any new protocol will require proof, either through publication in a peer-reviewed journal or empirical representative data that demonstrate its usefulness, effectiveness, and reproducibility. Links to other repositories of information on specific methods, such as peer-reviewed scientific journals, will allow users to conveniently access the most up-to-date and well-characterized methods in the literature. While it is important to note that this enhancement will take a great deal of time to populate in the next two years with useful protocols from the various disciplines within the RNA community, the initial steps for incorporation into the RNAMDB are currently in progress.

The proposed portal will be hosting a number of tools for predicting RNA secondary structure folding, the energy of those structures, and their predicted biophysical properties,

such as melting temperature. For instance, the popular RNA folding programs Mfold [31] and UNAfold [32] developed by Dr. Michael Zuker will soon be accessible through the current RNAMDB infrastructure and will be subsequently incorporated in the portal. Full integration will be achieved by making the information stored in the RNAMDB directly accessible by the folding algorithms in such a way as to enable users to calculate the possible effects of RNA modifications on energetics and structure stability. Since a drastic restructuring of the source code will be required, full integration of Dr. Zuker's folding programs will likely be implemented during the 2012-13 academic year; however, the programs are currently included, in their previous state, as a part of the RNAMDB. In a similar fashion, we plan on modifying existing software to enable the productive utilization of the information in the RNAMDB. For example, full integration with the programs in the existing Masspec Toolbox suite will greatly benefit the MS identification and characterization of natural RNAs extracted from living organisms. The new tools developed for the portal will be designed to take advantage of the information stored in the site and, thus, will be capable of accessing the different types of data and seamlessly communicating with one another across the board.

The proposed portal will include a constantly updated compilation of hyperlinks to external sources that may provide additional information for the identification and the structure-function investigation of RNA. In addition to repositories of experimental protocols, these links will direct users to pertinent peer-reviewed journals, *the RNA Society*, *the Transfer RNA Database* [33], *the Collaboratory for MS3D* [34], and many others. Links will also seamlessly point users to the external primary sources of information contained in the RNAMDB website. For instance, users will no longer be required to copy and paste the Chemical Abstracts registry numbers from the RNAMDB to the CAS Registry website. Instead, they will simply need to click on the hyperlinked CAS Registry number. Similarly, all references to both internal and external sources will be hyperlinked to provide quick access to all primary information. Additionally, it has become apparent that the information contained within the RNAMDB can be supplemented by cross-referencing with other relevant information residing in other databases such as *Modomics* [1] and *tRNAdb* [33].

Other planned enhancements to the RNAMDB functionalities include incorporation and cross-referencing with new sister databases containing the growing amount of information on both natural and non-natural RNA modifications. A sister database of force field parameters for molecular dynamics simulations will also be included as a supplement to the newly incorporated table of modified nucleoside base pairing free energy (ΔG) minima [35]. Although excluded from the current RNAMDB, non-natural modifications are gaining increasing interest in both academic and corporate settings. These types of modifications include those used as research tools because of their ability to fluoresce [36], to restrict certain geometries, including backbone alterations [37], to enable unusual base pairing schemes [38] and to act as tools for probing structure [39,40]. Furthermore, other non-natural nucleosides that have been discovered as the products of interactions with pharmaceuticals and as a result of environmental stimuli will be also included.

Finally, it is becoming apparent that a more modern user interface will be needed to accommodate these new enhancements. With recent technological improvements, such as deep transcriptome analysis [41] and high-throughput mass spectrometry [42,43], which have led to advancements in our current understanding of RNA in biological systems, it has become necessary to create a more automated method for researchers to submit their findings to the RNAMDB for consideration. The new user interface is being developed for implementation during 2011 and will seek to create a more user-friendly experience while adding improved functionality. The new interface will allow the curators to dedicate more of their time and resources towards validating entries and developing new features. In this direction, while past iterations allowed for unpublished modifications to be included (Table 2), the new implementation and curating structure will require publication as a criterion for entry into the RNAMDB. This will create a benchmark for the inclusion of modifications that have been diligently characterized and will reduce the likelihood of inaccurate entries.

F.7 Availability

The RNA Modification Database is freely available to researchers via the web address, http://rna-mdb.cas.albany.edu/RNAmods/. Additionally, the programs contained in the Masspec Toolbox are also free for public use through the link on the database homepage and via the web address, http://rna-mdb.cas.albany.edu/RNAmods/rnamass.htm. Users of information contained within this database are requested to cite this article as their source of information.

F.8 Acknowledgements

The authors are grateful for the contributions of Phil Durant for the 3D graphic of yeast $tRNA^{Phe}$ on the homepage and Patrick Limbach for his extensive work on the initial assembly and publication of the database. Additionally, we would like to thank Richard Walker, Steven Pomerantz and Byron Bossenbroek who contributed to the initial compilation of the database. The electronic version of the database was initially designed by Di He, Joe Zhang and Nancy Lombardo (University of Utah). Maintenance of the RNAMDB is supported by The RNA Institute at the University at Albany.

F.9 References

1. Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K: MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res* 2009, **37**(Database issue):D118-121.

2. Curran JF: Modification and Editing of RNA. In. Edited by Grosjean H, Benne R. Washington D.C.: ASM Press; 1998: 493-516.

3. Rozenski J, Crain PF, McCloskey JA: The RNA Modification Database: 1999 update. *Nucleic Acids Res* 1999, **27**(1):196-197.

4. McCloskey JA, Rozenski J: The Small Subunit rRNA Modification Database. *Nucleic Acids Res* 2005, **33**(Database issue):D135-138.

5. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X: Methylation as a crucial step in plant microRNA biogenesis. *Science* 2005, **307**(5711):932-935.

6. Yokoyama S, Nishimura S: tRNA Structure, Biosynthesis and Function. In. Edited by Söll D, RajBhandary UL. Washington D.C.: ASM Press; 1995: 207-223.

7. Agris PF, Vendeix FAP, Graham WD: tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* 2007, **366**(1):1-13.

8. Agris PF: Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* 2008, **9**(7):629-635.

9. Gustilo EM, Vendeix FA, Agris PF: tRNA's modifications bring order to gene expression. *Curr Opinion in Microbiol* 2008, **11**(2):134-140.

10. Limbach PA, Crain PF, McCloskey JA: Summary: the modified nucleosides of RNA. *Nucleic Acids Res* 1994, **22**(12):2183-2196.

11. McCloskey JA, Liu XH, Crain PF, Bruenger E, Guymon R, Hashizume T, Stetter KO: Posttranscriptional modification of transfer RNA in the submarine hyperthermophile Pyrolobus fumarii. *Nucleic Acids Symp Ser* 2000(44):267-268.

12. Pettit G, Yamauchi K, Einck J: Selective methylation of nuclosides1. *Synth Comm* 1980, **10**:25-35.

13. Raviprakash K, Cherayil JD: 2'-O-methyl-1-methyl adenosine: a new modified nucleoside in ragi (Eleusine coracana) tRNA. *Biochem Biophys Res Commun* 1984, **121**:243-248.

14. Robins MJ, MacCoss M, Lee AS: Nucleic acid related compounds. 20. Sugar, base doubly modified nucleosides at the 5'-terminal "cap" of mRNAs and in nuclear RNA. *Biochem Biophys Res Commun* 1976, **70**(2):356-363.

15. Suzuki T, Wada T, Saigo K, Watanabe K: Novel taurine-containing uridine derivatives and mitochondrial human diseases. *Nucleic Acids Res Suppl* 2001(1):257-258.

16. Wada T, Shimazaki T, Nakagawa S, Otuki T, Kurata S, Suzuki T, Watanabe K, Saigo K: Chemical synthesis of novel taurine-containing uridine derivatives. *Nucleic Acids Res Suppl* 2002(2):11-12.

17. Zhou S, Sitaramaiah D, Noon KR, Guymon R, Hashizume T, McCloskey JA: Structures of two new "minimalist" modified nucleosides from archaeal tRNA. *Bioorg Chem* 2004, **32**(2):82-91.

18. Kasai H, Goto M, Ikeda K, Zama M, Mizuno Y, Takemura S, Matsuura S, Sugimoto T, Goto T: Structure of wye (Yt base) and wyosine (Yt) from Torulopsis utilis phenylalanine transfer ribonucleic acid. *Biochemistry* 1976, **15**(4):898-904.

19. Sauerwald A, Sitaramaiah D, McCloskey JA, Soll D, Crain PF: N6-Acetyladenosine: a new modified nucleoside from Methanopyrus kandleri tRNA. *FEBS Lett* 2005, **579**(13):2807-2810.

20. Aritomo K, Wada T, Sekine M: Alkylation of 6-N-acylated adenosine derivatives by the use of phase transfer catalysis. *J Chem Soc, Perkin Trans* 1995, **1**:1837.

21. McCloskey J, Van Wagoner R, Hashizume T, Nomura M, Sako Y: *20th International tRNA workshop Banz, Germany* 2003:41.

22. Nyilas A, Chattopadhyaya J: Synthesis of O2'-methyluridine, O2'-methylcytidine, N4,O2'-dimethylcytidine and N4,N4,O2'-trimethylcytidine from a common intermediate. *Acta Chem Scand B* 1986, **40**(10):826-830.

23. Mandal D, Kohrer C, Su D, Russell SP, Krivos K, Castleberry CM, Blum P, Limbach PA, Soll D, RajBhandary UL: Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proc Natl Acad Sci U S A* 2010, **107**(7):2872-2877.

24. Giessing AM, Jensen SS, Rasmussen A, Hansen LH, Gondela A, Long K, Vester B, Kirpekar F: Identification of 8-methyladenosine as the modification catalyzed by the radical SAM methyltransferase Cfr that confers antibiotic resistance in bacteria. *RNA* 2009, **15**(2):327-336.

25. Van Aerschot AA, Mamos P, Weyns NJ, Ikeda S, De Clercq E, Herdewijn PA: Antiviral activity of C-alkylated purine nucleosides obtained by cross-coupling with tetraalkyltin reagents. *J Med Chem* 1993, **36**(20):2938-2942.

26. Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S: Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* 1988, **336**(6195):179-181.

27. Kowalak JA, Bruenger E, McCloskey JA: Posttranscriptional modification of the central loop of domain V in Escherichia coli 23 S ribosomal RNA. *J Biol Chem* 1995, **270**(30):17758-17764.

28. Toh SM, Xiong L, Bae T, Mankin AS: The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA. *RNA* 2008, **14**(1):98-106.

29. Long KS, Poehlsgaard J, Kehrenberg C, Schwarz S, Vester B: The Cfr rRNA methyltransferase confers resistance to Phenicols, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A antibiotics. *Antimicrob Agents Chemother* 2006, **50**(7):2500-2505.

30. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel H, SantaLucia J: AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Chem Theory and Computation* 2007, **3**:1464-1475.

31. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003, **31**(13):3406-3415.

32. Markham NR, Zuker M: UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* 2008, **453**:3-31.

33. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J: tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 2009, **37**(Database issue):D159-162.

34. Yu ET, Hawkins A, Kuntz ID, Rahn LA, Rothfuss A, Sale K, Young MM, Yang CL, Pancerella CM, Fabris D: The collaboratory for MS3D: a new cyberinfrastructure for the structural elucidation of biological macromolecules and their assemblies using mass spectrometry-based approaches. *J Proteome Res* 2008, **7**(11):4848-4857.

35. Vendeix FAP, Munoz AM, Agris PF: Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA* 2009, **15**(12):2278-2287.

36. Nutiu R, Li Y: Aptamers with fluorescence-signaling properties. *Methods* 2005, **37**(1):16-25.

37. Zhou C, Chattopadhyaya J: The synthesis of therapeutic locked nucleos(t)ides. *Curr Opin Drug Discov Devel* 2009, **12**(6):876-898.

38. Bergstrom DE: Unnatural nucleosides with unusual base pairing properties. *Curr Protoc Nucleic Acid Chem* 2009, **Chapter 1**:Unit 1 4.

39. Fabris D, Yu ET: Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D. *J Mass Spectrom* 2010, **45**(8):841-860.

40. Yu ET, Zhang Q, Fabris D: Untying the FIV frameshifting pseudoknot structure by MS3D. *J Mol Biol* 2005, **345**(1):69-80.

41. Iida K, Jin H, Zhu JK: Bioinformatics analysis suggests base modifications of tRNAs and miRNAs in Arabidopsis thaliana. *BMC Genomics* 2009, **10**:155.

42. Douthwaite S, Kirpekar F: Identifying modifications in RNA by MALDI mass spectrometry. *Methods Enzymol* 2007, **425**:1-20.

43. Kellersberger KA, Yu ET, Merenbloom SI, Fabris D: Atmospheric pressure MALDI-FTMS of normal and chemically modified RNA. *J Am Soc Mass Spectrom* 2005, **16**(2):199-207.

F.10 Tables and figures

Table F1. The RNA modifications in the database are present in all phylogenetic groups and in many of the known RNA species.

RNA	Phylogenetic Source		
	Archaea	**Bacteria**	**Eukarya**
tRNA	43	45	51
rRNA		1*	4*
SSU	11	8	18
LSU	8	15	12
5S	3		1
5.8S			5
mRNA			13
tmRNA		2	
snRNA			11
Chromosomal RNA			2
Other Small RNA			1

*Denotes rRNA modifications in which the subunit of origin is either unknown or reported as a mixture of subunits.

Table F2. Fourteen modifications have been added to the database since the last update was published in 1999.

Entry	Symbol	Common Name	Phylogenetic Source	RNA Type	Refs
96	$m^{1}Gm$	1,2'-*O*-dimethylguanosine	Archaea	tRNA	(11,12)
97	$m^{1}Am$	1,2'-*O*-dimethyladenosine	Eukarya	tRNA	(13,14)
98	$tm^{5}U$	5-taurinomethyluridine	Eukarya	tRNA	(15,16)
99	$tm^{5}s^{2}U$	5-taurinomethyl-2-thiouridine	Eukarya	tRNA	(15,16)
100	imG-14	4-demethylwyosine	Archaea	tRNA	(17,18)
101	imG2	isowyosine	Archaea	tRNA	(17)
102	$ac^{6}A$	N^{6}-acetyladenosine	Archaea	tRNA	(19,20)
103	$inm^{5}U$	5-(isopentenylaminomethyl)uridine	Bacteria	tRNA	(21)
104	$inm^{5}s^{2}U$	5-(isopentenylaminomethyl)- 2-thiouridine	Bacteria	tRNA	(21)
105	$inm^{5}Um$	5-(isopentenylaminomethyl)-2'-*O* methyluridine	Bacteria	tRNA	(21)
106	$m^{2,7}Gm$	N^{2},7,2'-*O*-trimethylguanosine	Archaea	tRNA	(21)
107	$m^{4}{}_{2}Cm$	N^{4},N^{4},2'-*O*-trimethylcytidine	Archaea	tRNA	(22)
108	C^{+}	agmatidine	Archaea	tRNA	(23)
109	$m^{8}A$	8-methyladenosine	Bacteria	rRNA	(24,25)

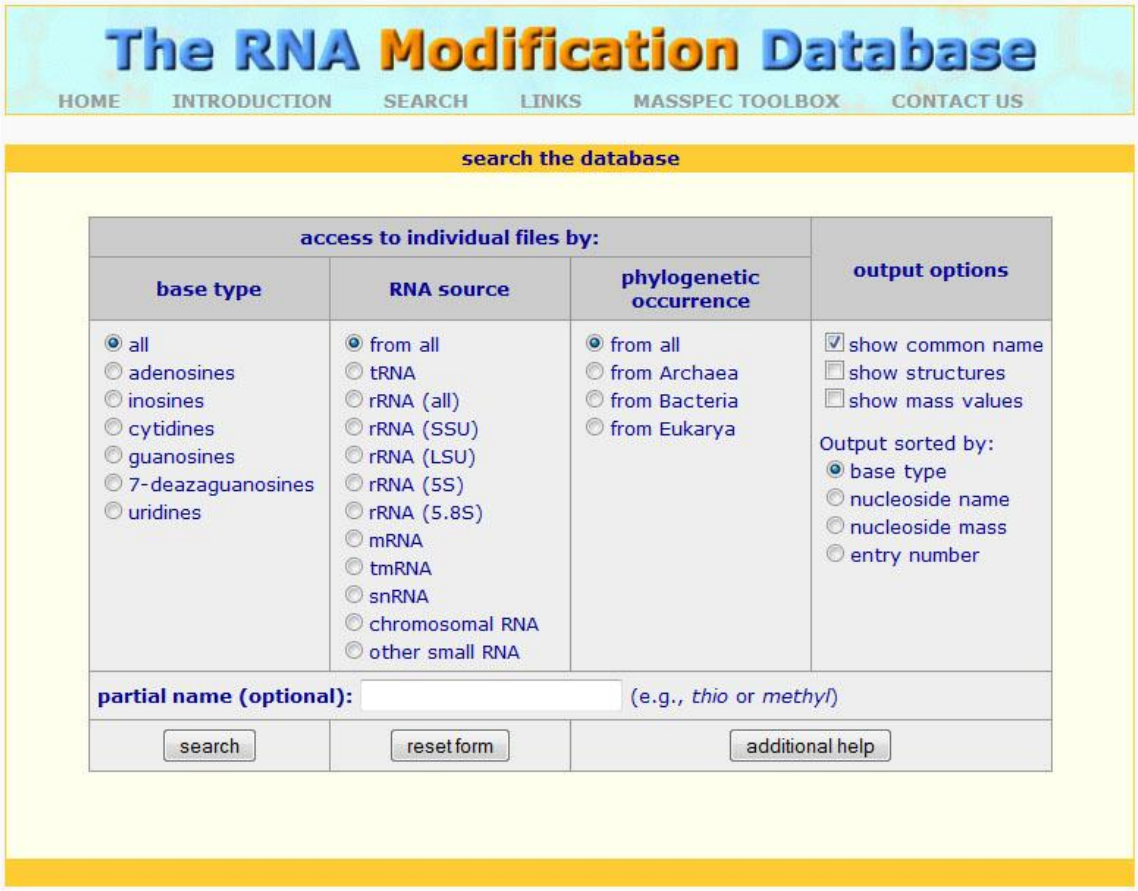

Figure F1. The user-friendly search page makes it easy for users to find RNA modifications in the database based on a variety of different criteria.

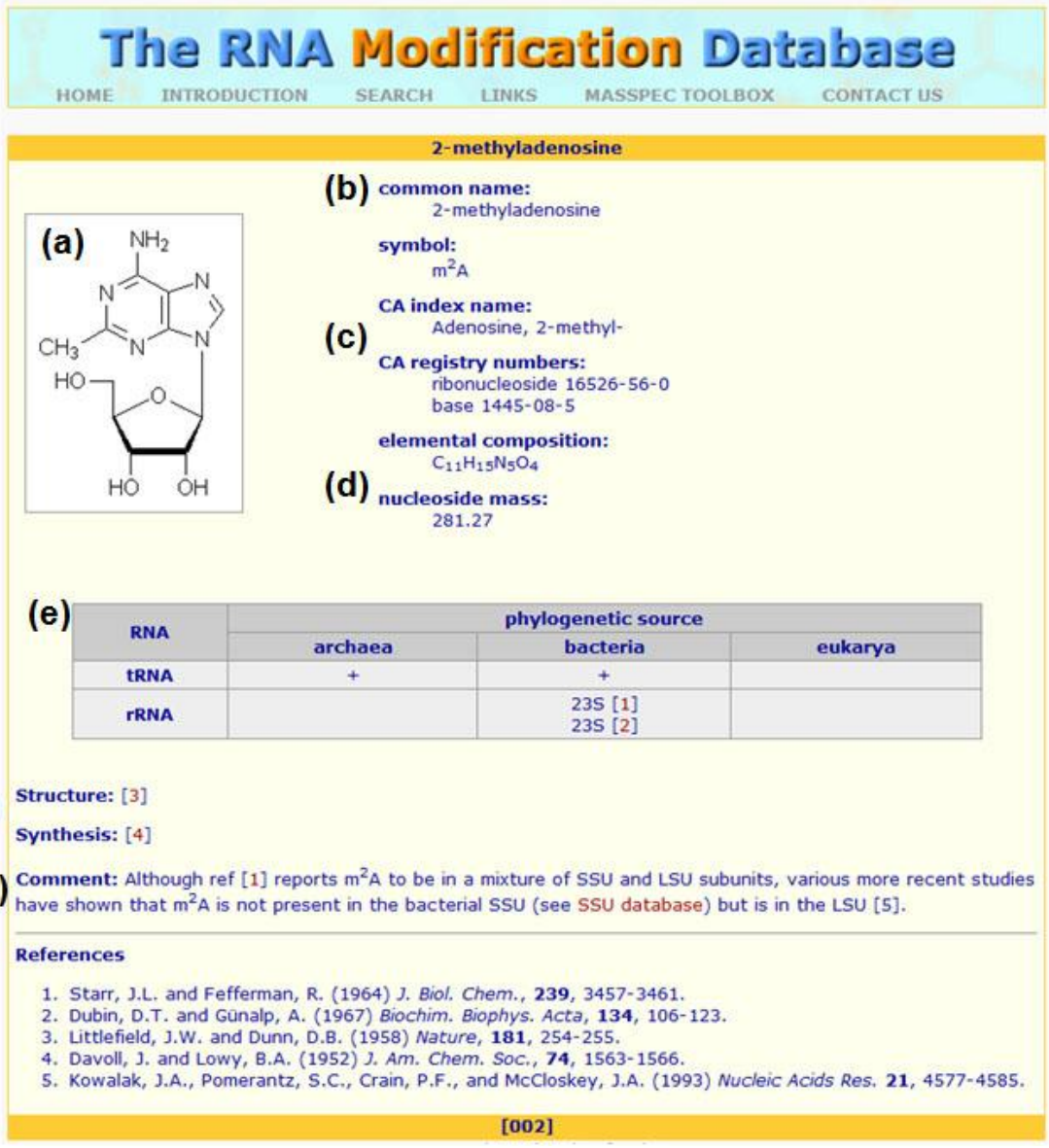

Figure F2. The entry information page for 2-methyladenosine shows a typical representation of the statistics that is contained within the database.

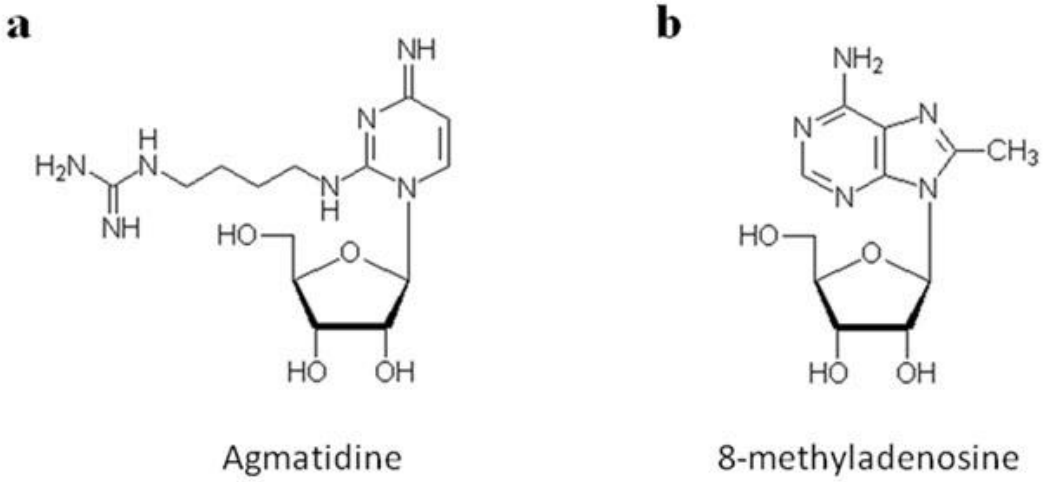

Figure F3. The two most recent additions to The RNA Modification Database, **(a)** agmatidine and **(b)** 8-methyladenosine, have been discovered and structurally characterized within the past year.

APPENDIX G. Curriculum vitae

G.1 Education

North Carolina State University, Raleigh, North Carolina, 2007-Present

- Ph.D. in Molecular and Structural Biochemistry (Spring 2012, expected)
- PI: Paul F. Agris
- **Research Topic:** Decipher the structural and functional roles of naturally-occurring posttranscriptional modifications in the anticodon stem and loop domain of the six-fold degenerate $tRNA^{Arg}$ isoacceptors in *E.coli* using structural and biophysical characterization and functional codon binding assays.

Juniata College, Huntingdon, Pennsylvania, 2001-2005

- B.S. in Biochemistry and Physics, (2005)
- PI: Jill B. Keeney
- **Research Topic:** Examination of the mechanism of high temperature transposition of the endogenous Ty1 retroelement in *S.cerevisiae* using genetic approaches; analysis of the production of virus-like particle proteins using SELDI-TOF mass spectrometry.

G.2. Professional Experience

Conductive Technologies, Inc., York, Pennsylvania, 2006-2007

- Biosensor Scientist in the Research and Development Division
- **Research Topic:** Development of novel glucose biosensor formulation for increased sensitivity and performance, reduction of hematocrit bias and decrease in result time.

Croda Inc., Mill Hall, Pennsylvania, 2005-2006

- Quality Control Chemist in the QC/QA Department

Exygen Research Inc., State College, Pennsylvania, 2003

- Research Technician in Liquid Chromatography and Mass Spectrometry
- **Research Topic:** Analytical determination of synthetic chemical compounds present in biological matrices.

G.3. Teaching Experience

North Carolina State University

- Chemistry/Physics Tutor; Initiative for Maximizing Student Diversity (2009-2010)
- Teaching Assistant; Advanced Biochemistry Laboratory (2007)

North Carolina State University & *University at Albany, SUNY*

- Mentored one high school student and three undergraduate students (2007-Present)

G.4. Awards and Honors

- 2011 – American Society for Biochemistry and Molecular Biology Travel Award
- 2010 – The RNA Institute Travel Award
- 2010 – Department of Biological Sciences Travel Award
- 2009 – American Society for Biochemistry and Molecular Biology Travel Award
- 2008 – Samuel B. Tove Award for Outstanding Teaching
- 2004 – William J. von Liebig Summer Research Fellowship
- 2001 – 2005: Calvert Ellis Scholarship
- 2001 – 2005: Homer C. and Ethel F. Will Scholarship

G.5. Research Techniques

Handling and purification of RNA, DNA and protein:

- Preparative PAGE RNA purification (unlabeled and radiolabeled)
- Purification of DNA and molecular cloning
- Preparative HPLC of small RNA and DNA oligomers
- Recombinant protein expression and purification
- *In vitro* transcription
- Site-directed mutagenesis

Binding assays:

- Radioligand filter binding assays

- Electrophoretic mobility shift assays
- Fluorescence quenching assays
- Isothermal titration calorimetry

Biophysical and structural characterization of RNA:

- Thermal denaturation and circular dichroism spectroscopy
- Differential scanning calorimetry
- NMR: currently being trained to solve the structure of three 17-mer hairpin RNAs
- X-ray crystallography: currently being trained to solve structures of the bacterial ribosome with a 17-mer hairpin RNA bound in the A-site

Computational characterization of RNA:

- Implicit and explicit solvent, targeted molecular dynamics and umbrella sampling
- GPU accelerated molecular dynamics
    - Assisting with the setup of a GPU-based computing center at The RNA Institute
- Docking simulations of RNA with large chemical libraries of ligands
- Software: Linux OS, AMBER, Autodock, Vina, PyMol, VMD, Chimera, Grace, Phenix

Robotics for high throughput screening:

- Use of a Perkin Elmer Janus system for setup of high throughput assay development

G.6. Publications

- **Cantara WA**, Lu K, Makshood M, Harr E, Kaiser R, Malkiewicz A, Agris PF: Structure and function effects of $tRNA^{Arg4}_{UCU}$ anticodon domain modifications. 2012. **Manuscript in preparation**.

- **Cantara WA**, Harr E, Agris PF: Molecular dynamics simulations reveal modification induced alteration in *E. coli* $tRNA^{Arg1}$ conformational transitions. 2012. **Manuscript in preparation**.

- **Cantara WA**, Makshood M, Agris PF: Efficient native batch purification of RNA with homogenous 5' and 3' ends. 2012. **Manuscript in preparation**.

- **Cantara WA**, Bilbille Y, Kim J, Kaiser R, Leszczyńska G, Malkiewicz A, Agris PF: Modifications modulate anticodon loop dynamics and codon recognition of *E. coli* $tRNA^{Arg1,2}$. J Mol Biol 2012, 416(4):579-597.

- Vendeix FAP, Murphy FV, **Cantara WA**, Leszczynska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF: Human $tRNA^{Lys3}_{UUU}$ is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. J Mol Biol 2012, 416(4):467-485. **Featured Article**

- **Cantara WA**, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF: The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res 2011, 39 (Database issue):D195-201.

- Lusic H, Gustilo EM, Vendeix FA, Kaiser R, Delaney MO, Graham WD, Moye VA, **Cantara WA**, Agris PF, Deiters A: Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial $tRNA^{Met}$. Nucleic Acids Res 2008, 36(20):6548-6557.