

Marquette University

e-Publications@Marquette

Master's Theses (2009 -)

Dissertations, Theses, and Professional
Projects

Biochemical Analysis of Dual T-Loop RNA Motifs that Perform Shape Recognition of the tRNA Elbow

John Aparicio Aguirre
Marquette University

Follow this and additional works at: https://epublications.marquette.edu/theses_open

 Part of the [Chemistry Commons](#)

Recommended Citation

Aguirre, John Aparicio, "Biochemical Analysis of Dual T-Loop RNA Motifs that Perform Shape Recognition of the tRNA Elbow" (2021). *Master's Theses (2009 -)*. 680.
https://epublications.marquette.edu/theses_open/680

BIOCHEMICAL ANALYSIS OF DUAL T-LOOP RNA MOTIFS THAT PERFORM
SHAPE RECOGNITION OF THE TRNA ELBOW

by

John Aparicio Aguirre, B.Sc
Northland College, Ashland, WI

A Thesis submitted to the Faculty of the Graduate School, Marquette University, in Partial
Fulfillment of the Requirements for the Degree of Master of Science

Milwaukee, Wisconsin

August 2021

ABSTRACT

BIOCHEMICAL ANALYSIS OF DUAL T-LOOP RNA MOTIFS THAT PERFORM SHAPE RECOGNITION OF THE TRNA ELBOW

John Aparicio Aguirre, B.Sc

Marquette University, 2021

T-loops are highly abundant, structurally conserved five nucleotide motifs that facilitate long range intra- and intermolecular interactions and function to stabilize the tertiary fold of RNA. Interlocked or dual T-loop motifs provide additional thermodynamic stability and perform long-range docking of other RNA molecules. Dual T-loop RNA structures are present within the multi-turnover ribonuclease P-based ribozyme, the ribosome, and the bacterial tRNA-sensing riboswitch genetic element. In all three systems, a dual T-loop structure recognizes TYC/D-loop region of tRNA. Interestingly, sequence conservation at specific positions in the T-loops suggest a conserved interlocked RNA architecture. While X-ray crystallography provides insight into the putative hydrophobic stacking interactions, little or no information exists regarding hydrogen bonding or the conformational flexibility of the dual T-loop motifs. Through extensive NMR experiments, we examine the pre-assembly of the dual T-loop motif and define the hydrophobic stacking and hydrogen bonding environment of the secondary structural elements of a minimal construct containing the dual T-loop motif in a bacterial *Bacillus subtilis* *glyQS* T-box riboswitch. To better understand the structural diversity and functional roles of interlocked T-loop RNA motifs, the thermodynamic parameters and recognition properties of the dual-T loop RNA – tRNA interaction will be determined using a combined UV- melt analysis and isothermal titration calorimetry (ITC) approach. Results from these experiments suggest weak binding of the truncated constructs with a dissociation constant of $2.00 \pm 0.47 \mu\text{M}$. In addition, preliminary data on the thermodynamic stability of constructs has been obtained. RNA structural motifs, such as the dual T-loop motif, play an important role in the gene regulation and understanding the recognition patterns could lead to potential application in pharmaceuticals

ACKNOWLEDGEMENTS

John Aparicio Aguirre

During my research period, I have had the pleasure of working with a number of people who assisted in various ways to my research. First, I would like to express my deepest gratitude to my advisor Dr. Nicholas Reiter for the wealth of knowledge, guidance, and encouragement throughout my research.

I would like to thank my committee members Dr. Joseph Clark and Dr. Chae Yifor being a part of my committee and their support throughout the years.

My sincere appreciation goes out to Dr. Sheng Cai, for helping me collect NMR data and learn about various NMR experiments, Dr. Martin St. Maurice for providing the ITC instrument for my experiments, and Dr. Marco Tonelli from NMRFAM in Madison for providing the NMR facility, running, and processing the data collected there.

I would like to thank my group members Dr. Danyun, Ainur, Dulmi, Genesis, Owen, and Julianne and all those who helped me in various ways throughout my research work.

Lastly, I would like to thank my parents who have always encouraged and challenged me throughout my life. I'd also like to thank my sister for all her help and my grandparents their love and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	i
TABLE OF CONTENTS.....	ii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
CHAPTER 1.....	1
INTRODUCTION.....	1
1. General Overview.....	1
1.1. RNA and the Importance.....	1
1.2. RNA Folding and tertiary motifs.....	2
1.2.1.1. Base pairing and secondary structure in RNA.....	3
1.2.1.2. Diversity of RNA tertiary motifs.....	6
1.2.1.3. Universal structural features of transfer RNA.....	8
1.2.2. Fundamental recognition principles of large RNA molecules.....	9
1.3. tRNA recognition and processing by the RNase P enzyme.....	10
1.3.1.1.1. RNase P derivatives and conservation in all three domains of life.....	10
1.3.1.1.2. Classes of RNA and Functionality.....	11
1.3.2. Shape-base substrate recognition and catalysis.....	13
1.3.2.1.1. tRNA recognition and gene regulation by the 5' leader regions of bacterial aminoacyl-tRNA (aa-tRNA) synthetase mRNAs.....	13
1.3.3. Gram-positive bacterial functions.....	13
1.3.4. T-box riboswitches structure and functions.....	14
1.3.5. Base stacking interactions.....	15

1.4. tRNA recognition by the ribosome.....	15
1.4.1.1.1.Anticodon recognition and base pair alignment to the codon message.....	15
1.4.2. Interactions near the 3' end of the tRNA that facilitate peptide bond formation.....	16
1.4.2.1.1.Acceptor stem and TΨC-D loop interactions by the ribosome.....	17
1.5. Common features of large RNAs that Recognize transfer RNA (tRNA).....	17
1.5.1.1.1.Ribonuclease P (RNase P), T-box riboswitch, ribosome.....	17
1.5.1.1.2. T-loop Motif.....	19
1.5.1.1.3.Sequence conservation of single T-loops.....	19
1.5.2. Hydrogen bonding and Dynamics of T-loop formation.....	19
1.5.2.1.1.1.Dual T-loop.....	19
1.1.6.4.2. Interlocked Dual T-loops.....	20
1.7 Introduction to biophysical methods to examine large RNA molecules.....	21
1.7.1. Biochemical analysis.....	21
1.7.2. X-ray crystallography-strengths and limitations.....	21
1.7.3. Cryo-Electron microscopy (Cryo-EM), strengths and limitations.....	22
1.7.4. NMR spectroscopy, strengths and limitations.....	22
1.7.5. Strategy for assignments.....	23
1.8. Scope of Study.....	23
CHAPTER 2.....	25
2.1 Introduction.....	25
2.1.1. Bacillus subtilis glyQS T-box riboswitch Structure and Function.....	25

2.1.2. Evolutionary similarity with in comparison to RNase P and rRNA.....	25
2.1.3. Structure and function of B.sub T-box riboswitch.....	26
2.1.4. Conformational dynamics of B.sub T-box riboswitch dual T-loop.....	27
2.2. Material and Methods.....	28
2.2.1. Materials.....	28
2.2.2. Instruments/Apparatus.....	28
2.2.3. RNA Transcription and purification.....	29
2.2.4. Native gel folding.....	30
2.2.5. Electrophoretic Mobility Shift Assay and Isothermal Titration Calorimetry for binding analysis.....	31
2.2.5.1 Electrophoretic Mobility Shift Assays.....	31
2.2.5.2. Isothermal Titration Calorimetry.....	32
2.2.6. Circular dichroism and UV Melt temperature studies.....	33
2.3 Results and Discussion.....	33
2.4 Conclusion.....	48
CHAPTER 3.....	50
3.1. Introduction.....	50
3.1.1. Nuclear Magnetic Resonance Spectroscopy of RNA.....	50
3.1.2. NMR Structure Determination strategies for Bacillus subtilis glyQS T-box riboswitch Dual T-loop Motifs.....	51
3.1.3. Challenges with structure determination of RNA via NMR.....	53
3.2. Material and Methods.....	54
3.2.1. Materials.....	54

3.2.2. Instruments/Apparatus.....	55
3.2.3. NMR Experiments.....	56
3.2.3.1. Sample Preparation.....	56
3.2.3.2. 1D NMR temperature and Mg ²⁺ dependence experiments.....	56
3.2.3.3. 2D NMR secondary structure elucidation.....	57
3.2.3.3.1. Experimental Design.....	57
3.2.3.4. 3D NMR Experiments.....	59
3.3. Results and Discussion.....	59
3.4. Conclusion.....	72
FUTURE WORK.....	74
REFERENCES.....	77
APPENDIX.....	85

LIST OF TABLES

Table 1. DNA sequences for truncated constructs to examine dual T-loop motif formation across phylogeny. Bolded is the DNA construct used for experimentation.....	35
Table 2. RNA sequences of the B.subWT and mutant constructs with predicted secondary structure depiction.....	36
Table 3. Buffer conditions for optimized folding of RNA constructs for the biochemical analysis experiments.....	38
Table 4. Melting temperature analysis of B.sub T-box riboswitch WT and mutants calculated from UV-melt experiments indicating highest thermostability to be the A30G mutant.....	44

LIST OF FIGURES

Figure 1. Central dogma; the flow of genetic information from DNA transcription to RNA translation to proteins.....	1
Figure 2. The four RNA nucleobases and the canonical Watson-Crick base pairing indicating hydrogen bonding interactions.....	2
Figure 3. Hydrogen bonding faces on the purine and pyrimidine nucleotides (17).....	3
Figure 4. A.) Seven torsion angles of RNA nucleotides including glycosidic torsion angle which determines helps determine hydrogen bonding faces (17). B.) Tertiary representation of the RNA double stranded A-form structure (14).....	4
Figure 5. A.) X-ray crystal structure of tRNA ^{Phe} in the conserved L-shape. B.) Dual T-loop TYC (blue) and D-loops (green) of tRNA discovery. Indicates the intercalation between the T-loops creating stabilizing base stacking interactions (70).....	8
Figure 6. Conserved recognition of tRNA through dual T-loop RNA motifs. A.) Recognition of E-site tRNA by ribosomal RNA T-loops within L1 stalk. (PDB 4V4I). B.) RNase P RNA dual T-loop binding to tRNA via intercalation and base stacking interactions. (PDB 3Q1Q). C.) Binding of T-box riboswitch Stem I region to tRNA ^{Gly} for glycine regulation (PDB 4LCK). (70).	14
Figure 7. Sequence conservation of T-loops within 11 introns and riboswitches and overall all sequence conservation of T-loops, following the UUNRA or UGNRA sequence (6).....	18
Figure 8. A.) Schematic diagram of single T-loop structure with base stacking interactions along the nucleotides 3 to 5, including intercalation of nucleobase in between base 4 and 5 (6). B.) Cartoon diagram indicating the intercalation and base stacking interactions occurring within the B.sub dual T-loops C.) Pymol representation of an intercalated dual T-loop of B.sub T-box riboswitch. Distal (D) T-loop shown in white and apical (A) T-loop pink (PDB: 6POM).....	20
Figure 9. Crystal structure of the B.sub glyQS T-box riboswitch (blue) and B.sub tRNA ^{Gly} (green) complex based on SAXS data. B.sub T-box riboswitch locks tRNA ^{Gly} into place via specific binding interaction (labeled) for gene regulation (36, 73).....	26
Figure 10. Test transcription indicating transcription viability of Gkai glyQS T box riboswitch, B.sub T-box riboswitch WT, TMR RNase P, and Oihe T-box riboswitch,	

respectively, for qualitative analysis of RNA yield. Oihe T-box construct did not transcribe while others indicated fair amount of DNA product.....34

Figure 11. Homogeneous fold of B.sub WT T-box comparison to dimer Gkau and heterogeneous TMR.....37

Figure 12. A.) Dimerization analysis of B.sub T-box construct (Lane 5) indicates monomeric fold compared to unstructured single-stranded RNA (Lane 1) and structured G-quadruplexes (Lanes 2-4), and tRNA (Lane 6)38

Figure 13. Compact folding of B.subWT, C13U, and A30G at 45bp in folding conditions listed in Table 3.39

Figure 14. A.) UV-melting temperature studies indicate A30G (**grey**) has most thermostability with a $T_m = 77^\circ\text{C}$. A11U/C13A (**yellow**), C13U (**blue**), and B.subWT (**orange**) all have similar $T_m = 67, 67, \text{ and } 66^\circ\text{C}$ centered on an absorbance of 260nm. B.) Circular dichroism spectroscopy of B. sub T-box constructs A30G (**grey**), A11U/C13A (**yellow**), C13U (**blue**), and B.subWT (**orange**) and their qualitative formation of double stranded complex at 25°C measuring the absorbance on a range of 190-300nm centered on 260nm. C.) Secondary structure of B.sub wild type highlights the difference between the various point mutants, where each box is color-coded corresponding to the color on the scatter plot. Parameters for CD experiment were 20-90°C ramp @ 5°C increments with Abs. range of 190-300nm.....42

Figure 15. A.) RNA-tRNA complex formation through EMSA studies previously performed by the Reiter group on binding to tRNA (Reiter Proposal)46

Figure 16. ITC binding curves for B.subWT and C13U constructs, respectively. Comparative study was performed which indicated an ~50 fold difference in dissociation constant(K_d) (36).....47

Figure 17. Chemical shift of RNA nucleotides (55). Regions 10-14 ppm indicated base pairing where noncanonical pairs are found between 10-12ppm. Aromatic protons are found in the 6-9ppm range with ribose protons containing the most overlap.....50

Figure 18. Representation of canonical base pairing in RNA. Arrows indicate NOE connectivities stemming from imino proton.....51

Figure 19. General pulse sequence for 1D NMR analysis.....57

Figure 20. A.) 1D temperature dependence studies of folded B.sub T-box riboswitch at a range of 283-298K. Temperature increases from top to bottom at 5 K increments. B.)

Mg²⁺ 1D NMR titration experiment on B.sub T-box riboswitch with increasing concentration of Mg²⁺ from top to bottom (0, 0.2, 0.4, 2, 5, and 10 mM, respectively).....61

Figure 21. Top.) Secondary structure prediction from mfold web servers of B.sub T-box riboswitch construct. Bottom.) Imino base pairing region of ¹H,¹⁵N HSQC. Labeled 1D corresponds to resonances directly underneath in COSY and NOESY experiments. Dashed lines indicate strongly base paired regions. Sequential assignments of the imino helical regions within the construct.....62

Figure 22. Top.) U-A and Bottom.) G-C labeled ¹H-¹⁵N HMQC indicating imino resonances in B.sub WT T-box construct. Both contain 2D NOESY experiment illustrating the relationship between the two experiments.....65

Figure 23. Top.) ¹H-¹³C HSQC of U-A labeled construct displaying the AH2 region 150-155ppm aligned with D2O NOESY sequential walk region to identify starting point. Bottom.) Overlaid G-C (blue) and U-A (red) labeled sample presenting aromatic protons of nucleobases. This indicates high degree of spectral overlap among aromatic protons of RNA.....66

Figure 24. A.) Pymol representation of the two helical regions, helix I and helix II, within the B.sub T-box riboswitch. B.) 2D ¹H-¹H NOESY (black) spectrum overlaid with a ¹H-¹H TOCSY (yellow) recorded in D₂O at 800MHz with a 300ms mixing time.....70

Figure 25. Predicted hydrogen bonding patterns of the base triples within the interdigitated dual T-loops (PDB 6POM).....70

Figure 26. 2D NOESY spectrum of inter- and intraribose walk displaying the simplifying power of selective deuteration. Both regions are the same with the right spectrum being selectively deuterated (3)74

Figure 27. UV melt temperature determination through first derivative analysis of B.subWT T-box construct at 68°C.....85

Figure 28. Enthalpy determination of B.subWT to be -6.60kcal/mol from plot of ln(K_a) versus 1/T(K⁻¹).....86

Figure 29. UV melt temperature determination of B.subA11U/C13A T-box construct at 70°C.....86

Figure 30. Enthalpy determining to be -6.34 kcal/mol from plot of ln(K_a) versus 1/T(K⁻¹) for double point mutant.....86

Figure 31. UV melt temperature determination of B.subC13U T-box construct at 70°C.....	87
Figure 32. Enthalpy determining to be -6.55 kcal/mol from plot of $\ln(K_a)$ versus $1/T(K^{-1})$ for C13U mutant.....	87
Figure 33. UV melt temperature determination of B.subA30G T-box construct at 70°C.....	88
Figure 34 . CD spectroscopy analysis of B.sub mutant constructs indicating ds-character with secondary structure representation indicating point mutant locations (C13U-yellow, A30G-orange, A11U/C13A-grey, and B.subWT-blue at 215nm.....	88
Figure 35. EMSA binding studies to tRNA by B.subWT, helix, A30G, and C13U mutants, respectively. Concentration of tRNA incubated with mutants increase from left to right.....	88
Figure 36. Isothermal Titration calorimetry (ITC) with tRNA (100 μ M) and B.subT-box (1mM) given a K_d of $4.78 \pm 1.8 \mu$ M.....	89
Figure 37. Isothermal Titration calorimetry (ITC) with tRNA (1mM) and B.subWT T-box (88 μ M) indicating no binding.....	90
Figure 38. Isothermal Titration calorimetry (ITC) with tRNA (100 μ M) and B.sub C13U T-box (1mM) given a K_d of $7.35 \pm 342 \mu$ M.....	91
Figure 39. Isothermal Titration calorimetry (ITC) with tRNA (50 μ M) and B.sub C13U T-box (250 μ M) indicating no binding.....	92
Figure 40. 2D 1H - 1H NOESY of B.sub WT T-box construct overlaid with Helix construct for uracil base pairing determination both collected at 600MHz.....	93
Figure 41. 2D 1H - 1H NOESY of B.sub WT T-box construct overlaid with C13U construct for T-loop noncanonical base pairing determination both collected at 600MHz.....	93

CHAPTER 1

Introduction

1. General Overview

1.1. *RNA and the Importance*

Ribonucleic acid (RNA) is one of three essential macromolecules in all forms of life. RNA's role as a central messenger is well documented, as it is transcribed from deoxyribonucleic acid (DNA) and translated into proteins (Figure 1). Initially, RNA was thought to only have a few roles in the translational process after being transcribed from DNA. The three major classes of RNA involved in translation are messenger RNA

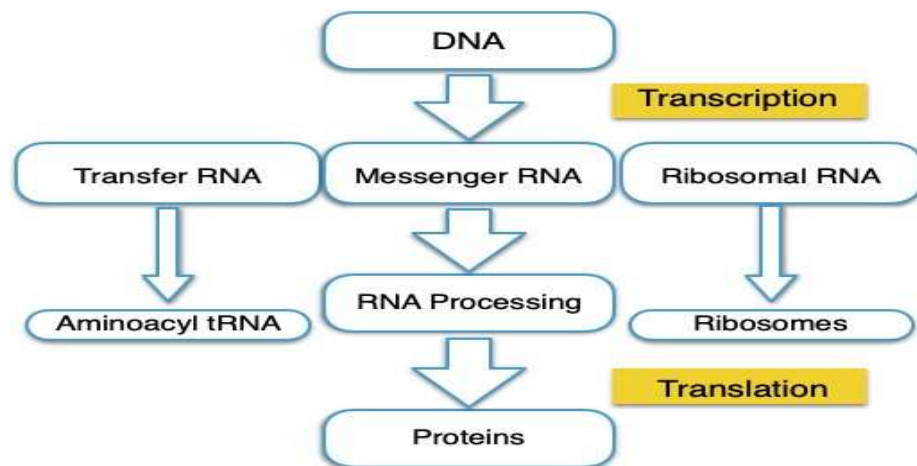


Figure 1. Central dogma; the flow of genetic information from DNA transcription to RNA translation to proteins.

(mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). All three types play an important role in the transfer of genetic information and protein synthesis. However, the vast majority of the human genome is not transcribed into protein coding mRNAs but

rather transcribed into non-coding RNAs molecules (ncRNAs) (41). Characterization of these ncRNAs that do not code for proteins have demonstrated the diverse functions of RNA that are reflected by RNA's ability to fold into a wide range of structures. RNA undergoes extensive processing and modifications from a variety of complex proteins and ribonucleoproteins (RNP) in order to provide for the efficient regulation of gene expression as well as other functions (32, 38). RNA structure and stability is dependent on transient base pairing interactions, small structured tertiary motifs, or even the assembly of a defined, complex three-dimensional fold. The formation of RNA tertiary structure is based on sequence but also involves hydrophobic stacking and base pairing, and is stabilized by cations and divalent metal ions. In the cell, proteins greatly expand the functional complexity of RNAs and these RNP complexes help to mediate fundamental catalytic reactions and cellular transformations.

1.2. RNA Folding and Tertiary Motifs

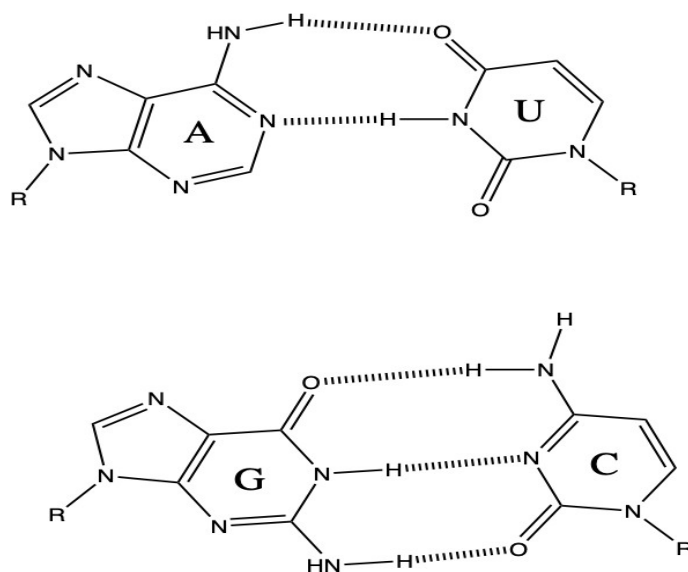


Figure 2. The four RNA nucleobases and the canonical Watson-Crick base pairing indicating hydrogen bonding interactions.

RNA and DNA have very similar primary structure, consisting of a pentose sugar, four different nucleosides and linked via a phosphodiester back- bone forming a long polymer chain.

Nucleotides are the phosphate esters of nucleosides, which are found in both DNA and RNA. The primary difference between the two is based on an additional 2'-hydroxyl found on the furanose ring of RNA (Figure 3). A second difference is the presence of a methyl group at the C5 position of DNA, thymine (T), compared to the presence of a hydrogen at the C5 position of RNA (uracil, U). Similar to DNA, RNA structures stem from the sequence of nucleotides and rely on base pairing interactions to form helical structures. Although DNA and RNA both have seven defined torsion angles along one nucleotide unit, each form very different structures in solution. DNA typically forms a B-form helical structure, where the furanose ring has a C2'-endo conformation to minimize non-bonded interactions (14). RNAs helical pattern conforms to an A-form helix, where the C3'-endo furanose is preferred. RNA has much more structural and functional diversity than DNA due to the presence of the 2' OH functional group.

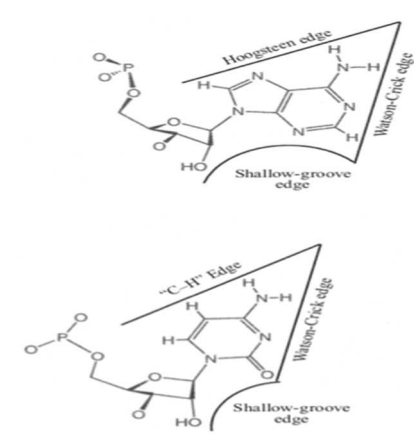


Figure 3. Hydrogen bonding faces on the purine andpyrimidine nucleotides (17).

1.2.1.1. Base pairing and secondary structure in RNA

Primary RNA transcripts upon the formation of base pairs lead to complex secondary structures such as bulges, hairpin loops, and helices through base pairing, stacking, and hydrogen bonding (31). RNAs can form compact structures due to the maximization of hydrogen bonding, base pairing, and base stacking interactions (14, 28). Nucleosides bases in RNA are composed of monocyclic pyrimidines, such as cytosine (C) and uracil (U) and bicyclic purines, adenine (A) and guanine (G). Canonical base pairing in RNA occurs between A:U and C:G (Figure 2A,B). Non-canonical base pairs can occur based on the orientation of nucleosides and the folding of the RNA molecule. Noncanonical base pairing patterns occur on the C-H edge of pyrimidines or the Hoogsteen edge of purines and the sugar edge which involves hydrogen bonding interactions with the ribose 2'-OH (Figure 3). Generally, these are not as favorable as canonical base pairing, but can provide a wealth of stability through alternative base stacking and hydrogen bonding interactions.

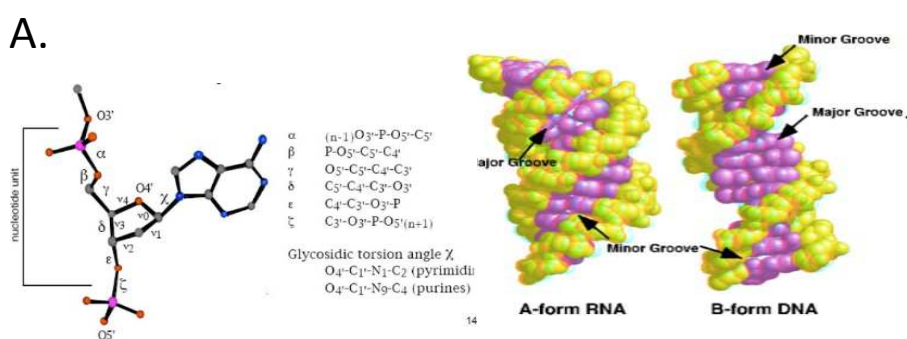


Figure 4. A.) Seven torsion angles of RNA nucleotides including glycosidic torsion angle which determines helps determine hydrogen bonding faces (17). B.) Tertiary representation of the RNA double stranded A-form structure (14)

RNAs have a wide variety of secondary structures based on their folding patterns. A key factor that affects RNA flexibility is the phosphate backbone. Generally, RNA can form an A-form helical structure, dependent on its sequence and solution conditions, and this helix can have a $\sim 6\text{\AA}$ larger diameter than a DNA B-form helix due to the additional 2'-hydroxyl group and avoidance of steric clashes (29, 53). Additionally, an A-form structure creates a narrow and deep major groove with a shallower and wider minor groove compared to other helical conformations (Figure 4B). RNA contains seven torsion angles; six define the conformation of the RNA backbone and one glycosidic torsion angle determines the orientation of the nucleobase with respect to the ribose (Figure 4A). The torsion angles along with the sugar pucker provide the flexibility for the formation of unique structures depending on base pairing, stacking, and hydrogen bonding environment and also help to avoid steric clashes that occur along the RNA biopolymer. The formation of secondary structures is based on constraints that include: bases cannot participate in more than one base pair, a region that must contain at least three unpaired bases, and no cross-strand interactions from one stem to another (29). These constraints distinguish the line between secondary and tertiary RNA structure.

1.1.2.2. Diversity of RNA tertiary motifs

RNA tertiary structure is defined as the arrangement of secondary structural elements into one global conformation. Negatively charged RNA structure is dependent on electrostatic interactions with positive metal ions which are integral to their tertiary fold and function (4). Secondary structure of RNA is more robust and forms faster than tertiary structure due to base pairing (42). The formation of a rigid secondary structure with metal ions provides the building blocks for a wide variety of tertiary motifs. The most

common and prevalent tertiary motif in RNA structure is called coaxial stacking, where two terminal RNAs base stack to form a more stable structure (4). Stacking of the π -electrons provides the release of 0.5-3.0 kcal/mol of free energy depending on the stacking bases. (11, 55). Coaxial stacking occurs at junction motifs, where other motifs meet and fold and help determine the number of conformational states as they are the intersection (4). Junctions typically form in two or three-way intersections but have been analyzed up to 10 (37). A typical, and common, example of a junction can be found in the four-way junction within tRNA. This four-leaf clover junction provides the intersection for the folding into the common L-shape found in tRNA. Bulges are unpaired stretches of nucleotides located within one strand of a nucleic acid duplex or helical structure (26).

Bulge sizes can vary from a single unpaired residue, up to several nucleotides that frequently form flexible unique recognition sites in RNA three-dimensional architectures. By causing a break in the helical structure, bulges widen deep groove through distortion of the RNA backbone and allows for recognition and binding through access to the otherwise inaccessible base pairs (26). A similar type of motif is the kissing loop, which utilizes magnesium ions to form Watson-Crick base pairs between two secondary hairpin structures (11). Formation of base pairs between two previously unpaired bases allows for continual coaxial stacking interactions of two separate helical structures, increasing the overall stability of the structure. Over half of all hairpin structures in RNA are capped by the tetraloop-receptor motif and provide stability for crystallization in addition to participating in tertiary interactions by providing functionality to their substrates (47) The tetraloop can be characterized as providing long range tertiary interactions through sequence specific binding of a four nucleotide common sequence, GNRA or UNCG,

where N is any base and R is purine specific (47). Tetraloop bind to a variety of RNA and protein molecules and serve as a model for another type of tertiary interactions, the T-loop, which perform important binding functions and can be found in ribosomal RNA and transfer RNA.

The discovery of the importance of tertiary folding in the functionality of non-coding RNAs, along with new methodology for structure determination has caused an increase in examination efforts to characterize reoccurring RNA tertiary motifs. RNA tertiary structure is determined by interactions between distant secondary structural elements (11). Basic secondary structure elements, like base pairing and base stacking, lead to tertiary structures, such as co-axial stacking, A-minor motifs, base triples, and T-loops. Co-axial stacking of the nucleobases dominates the stabilization of RNA structure through the pie-electron system. Other tertiary structures, like base triples or T-loops, provide stability through increasing the number of hydrogen bonds, intercalation, or providing binding sites for ligands. RNA in the unfolded state is believed to form the majority of native secondary structural elements but depends on hierarchical formation of tertiary interactions to determine native structure.

1.1.2.3. Universal structural features of transfer RNA

Transfer RNA (tRNA) is an adaptor molecule essential for translation which contains an amino acid and binds to the complementary mRNA containing the specific genetic information needed to begin translation. tRNA is a smaller RNA with approximately 70-100nt folded into a L-shape tertiary structure (Figure 5A). tRNA structural is generally conserved across various bacterial tRNAs despite variations in

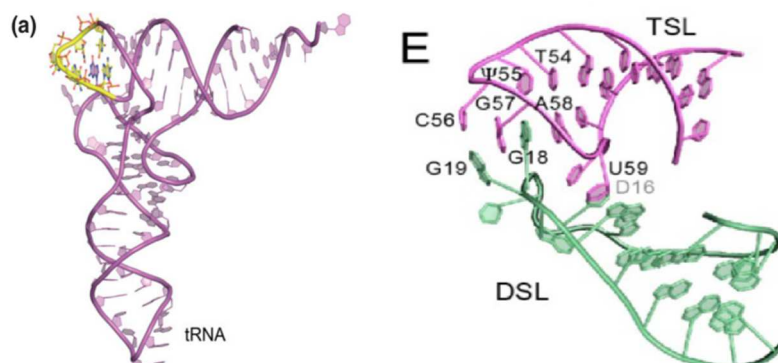


Figure 5. A.) X-ray crystal structure of tRNA^{Phe} in the conserved L-shape. B.) Dual T-loop TYC (blue) and D-loops (green) of tRNA discovery. Indicates the intercalation between the T-loops creating stabilizing base stacking interactions (70).

sequences and base modifications (44). The general structure of consists of a secondary four leaf clover-like structure with a four-way junction allowing for the for the tertiary folding. The universally conserved TYC-D loop elements intercalate and provide the long-range interactions needed to form the conserved L-shape tertiary fold (Figure 5B). This allows for two separate coaxial stacked helical regions to provide a wealth of base pairing and stacking inter- actions leading to a stable and well folded structure.

1.1.2.4. Fundamental recognition principles of large RNA molecules

Recognition of large RNA molecules depends on specific interactions with ribonucleoprotein (RNP) complexes, for example, RNase P or RNA-RNA interactions such as degenerate, non-specific binding of tRNA (24). Specific interactions can be

defined as the preferential association to RNA sequence or structural motifs. Whereas, non-specific interactions associate with RNA without sequence or structural motifs specificity (24). These interactions are the driving force behind the wide variety of functions within RNA and illustrate the importance of tertiary structures for both specific and non-specific interactions, which include base pairing, base stacking, and metal ion mediated folding.

Interactions including base pairing and base stacking mediated by metal ions allows for formation of tertiary structures. RNAs can recognize free bases within the helical region or other secondary motifs to form base pairs in order to stabilize the structure. Hydrogen bonding affects the stability of RNA through the number of bonds in base pairing. Thermodynamically, the traditional Watson-Crick base pair of C:G pair is more stable with three hydrogen bonds compared to the two of a U:A base pair (12). This leads to an approximately 10 kcal/mol difference between the two pairs but is not the major contributor to RNA stability or recognition. Base stacking in RNA structures provides the most stability to the molecule due to π -electron interactions between stacked base pairs and plays an important role in the recognition pathways (11). The typical free energy associated with terminal canonical base pairs G-C and U-A are -3.1 and -0.9 kcal/mol, respectively (61). Thermodynamic stability improves as the number of stacking interactions increases and dependence on the nearest neighbors influences stability as well. Influence of stacking interactions on base pairs are indicated by the comparison of a 5'-UUAA-3' duplex with free energy value of approximately -0.9 kcal/mol, in contrast to a 5'-GUAC-3' duplex with an increase in free energy to -1.8 kcal/mol (61). Stacking interactions thermodynamic properties vary based on

geometrical deformations and how well bases stack with each other in solution. Metal ions help form stable structures through interactions with the polyanionic backbone of RNA. Localized binding pockets in RNA interactions can accumulate negative charge and use abundant cations, such as Na^+ , K^+ , Ca^{2+} , and Mg^{2+} , to stabilize it (11, 52). Mg^{2+} contains both a catalytic and structural role in a variety of structures including the processing of tRNA. These building blocks for secondary structures fold into more intricate tertiary structures through hydrogen bonding, base stacking, and hydrophobic interactions. It is important to study these higher ordered RNA structures to help determine their functions in the cell.

1.3. tRNA recognition and processing by the RNase P enzyme

1.3.1.1. RNase P derivatives and conservation in all three domains of life

Processing and recognition of pre-tRNA by a RNA subunit provides a model system for the accurate study of RNA folding, binding specificity, and catalysis. RNase P is an endonuclease ribozyme which recognizes pre-tRNA in trans and cleaves the 5' end upon binding (32). The RNA portion was discovered to be catalytically proficient for the processing of pre-tRNA in the absence of proteins by the Pace and Altman group (1, 15). This was the first described, RNA based, multiple turnover enzymes discovered, leading to an increased research in the role of RNA catalysis. Sequence analysis of 10 RNase P RNA sequences from all organisms discovered five highly conserved regions (CR) implicating the functional role of the RNA across phylogeny (48).

The catalytic RNA portion of RNase P is universally conserved across all three domains, but the protein subunits are not. RNase P holoenzyme contains drastically different proteins subunits based on each of the different domain. Bacterial RNase P

has shown that the catalytic RNA is active in vitro without the protein subunits but contains one in vivo (25). Archaeal and eukaryotic RNase P RNAs require more protein subunits, 4 or 5 and 10 subunits, respectively, in order to stabilize the RNA subunit and function catalytically, due to the increase in complexity of the organism (7, 9, 15). The bacterial RNase P RNA sequence is not entirely conserved, but does maintain structural organization with a few regions in the sequence remaining highly conserved across all organisms. Through the comparison of RNase mitochondrial RNA processing enzyme (MRP)— found in the nucleolus— in both yeast and mammals, similarities in structure to RNase P, indicating a split into two different enzymes based on location in the cell (9, 15). Given the structure conservation, it is possible the RNase P RNA molecule was primordial (15). The similarities of the RNA component across phylogeny and the three classes of RNA indicate the importance of RNase P RNA in the translational process and recognition of tRNA-like substrates.

1.3.1.2. Classes of RNA and Functionality

Messenger RNA (mRNA), otherwise known as coding RNA, carries genetic information transcribed from DNA in order to produce proteins. Eukaryotic mRNAs have the most variability in function of the three classes of RNA, due to their need to be processed from precursor-mRNA (pre-mRNA) to mature mRNA. Pre-mRNA contain coding and noncoding regions within the same gene and need to undergo splicing before the translational process (67). The removal, or splicing, of introns (noncoding) while linking protein coding regions (exons) is a one step process completed by the spliceosome. Capping of the 5' end with a modified guanine and polyadenylation at the 3' end leads to an increase in stability, which allows for the mature mRNA to be exported

from the nucleus to the ribosome; where the coding regions make proteins and noncoding regions function in gene regulation through recognition of tRNA via highly conserved regions similar to RNase P (2).

Ribosomes are complex macromolecules consisting of multiple protein subunits and mature rRNA. Ribosomal RNA (rRNA) are the building blocks of ribosomes and provide the location for the interaction between aminoacylated tRNA to interact with and mRNA to form proteins (32). rRNA catalyzes the synthesis of peptide bonds by stitching together amino acids. Similar to mRNA, extensive processing is needed in order to stabilize the rRNA. This includes methylation of specific nucleotides to stabilize the rRNA and allow for the formation of the ribosome. Despite being made up of many protein sub- units, rRNA makes up a large portion of the mass and helps stabilize the ribosomal structure through specific structural motifs, similar to RNase P.

Precursor tRNA (pre-tRNA) undergoes multiple processing pathways as it is non-functional in the immature form. Cleavage of by both endo- and exonucleases at the 5' and 3' ends, respectively, along with the addition of CAA sequence at the 3' end and nucleotide modification leads to fully functional tRNA ready for translation. tRNA has clover-like secondary structure that folds into the L-shape tertiary structure. Highly conserved long range interactions through structural motifs help stabilize the 900 turn and help in recognition by mRNA, as well as other substrates.

1.3.2. Shape-base substrate recognition and catalysis

Typically, ribonucleases help catalyze the degradation of RNAs. Instead, ribonuclease P acts as a ribozyme and cleaves pre-tRNA through a hydrolytic mechanism to process tRNA for translation. The RNA component of bacterial RNase P is between

300- 450 nucleotides in length and can be separated into two distinct regions, the catalytic (C) and specificity (S) domain (51). These independent regions form inter-domain interactions creating a concave binding pocket for the correctly positioned tRNA and tRNA-like substrates to bind (58). Binding of substrate tRNA includes three important interactions; base stacking between TYC and D-loops of tRNA with S-domain of RNase P, A-minor interactions with acceptor stem, and base pairing at 3' end of tRNA (41). Interactions with the S-domain assist substrate recognition through base pairing at the CCA-3' end to nucleotides in tRNA as well as the protein component (41). Highly CR II and III within the S-domain form one of the most unique and essential recognition elements, known as a T-loop, for RNase P-tRNA complex formation. CR II and III, otherwise known as J11/12 and J12/11 T-loop regions, utilize base stacking interactions between TYC T-loop of tRNA . These CR II and III are found in other structures including the T-box riboswitch, mRNA, and rRNA, again, showing the evolutionary significance of RNase P RNA.

1.3.2.1.1. tRNA recognition and gene regulation by the 5' leader regions of bacterial aminoacyl-tRNA (aa-tRNA) synthetase mRNAs

1.3.3. Gram-positive bacterial functions

Gram positive bacteria function as gene regulators in the cell. These T-boxes are found upstream of mRNA sequences in order to control the termination of synthetases (69). Specific utilization of the T-box mechanism controls termination and, therefore, expression of aminoacylated tRNA synthase, amino acid biosynthesis, and transporter genes. Specifically, gram positive bacteria do not recognize small molecules and directly bind based on tRNA molecules based on the aminoacylation state.

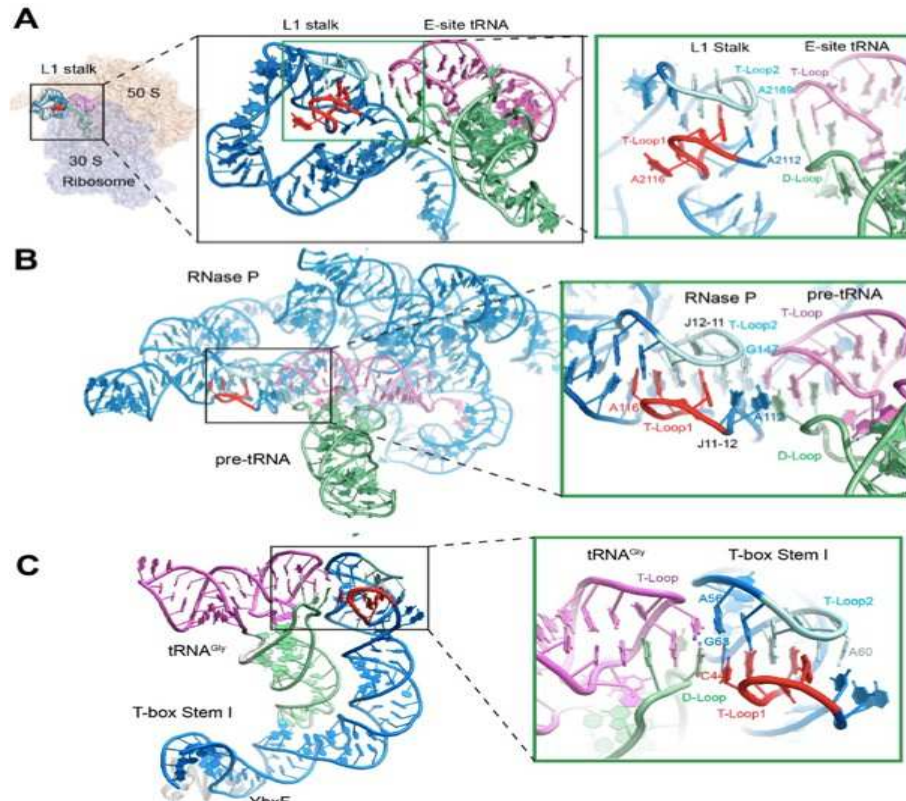


Figure 6. Conserved recognition of tRNA through dual T-loop RNA motifs. A.) Recognition of E-site tRNA by ribosomal RNA T-loops within L1 stalk. (PDB 4V4I). B.) RNase P RNA dual T-loop binding to tRNA via intercalation and base stacking interactions. (PDB 3Q1Q). C.) Binding of T-box riboswitch Stem I region to tRNA^{Gly} for glycine regulation (PDB 4LCK). (70).

1.3.4. T-box riboswitches structure and functions

T-box riboswitches recognize specific RNA elements in response to external signals in order to affect transcription or translation in mRNAs (69). The riboswitch contains two domains, the aptamer and expression platform, which function to bind tRNA through a highly conserved sequence and adopt conformations suitable for binding tRNA, respectively (69, 71). T-boxes consist of the highly conserved Stem I and the expression platform, which are separated by a linker or junction. The expression platform can adopt multiple conformations based on the substrate bound (8). Stem I

contains the conserved regions and that sense or bind tRNA via the conserved structural motif. In general, the T-box riboswitch is ~100nt long and the conserved structural motifs allow for double strand formation which contain a K-turn motif and bulges (8). This compact fold allows for recognition of the cognate tRNA or tRNA-like structure.

1.3.5. Base stacking interactions

Upon recognition of tRNA, X-ray crystal structures show significant base stacking interactions (8, 70). The distal conserved regions of Stem I form a six-nucleotide stack with three coplanar bases (Figure 6C). These base stacking interactions with tRNA to form the stable complex and allow for tRNA processing. This causes global and local conformational changes within tRNA, including the D-loop in tRNA. A conserved C44-G63 base pair in the dual T-loop structure of tRNA is allowed to base stack with the T-box and form the stable stacking pattern (Figure 6C). Intercalation of bases within the highly conserved regions in the T-box riboswitch provide stability and a more favorable binding interface for tRNA (71).

1.1.4. tRNA recognition by the ribosome

1.4.1.1.1. Anticodon recognition and base pair alignment to the codon message

Ribosomes facilitate the processing of tRNA for translation. tRNA is recognized by three distinct parts, aminoacyl, peptidyl, and exit sites (A, P, E) which surround and bind tRNA through both protein and RNA components (70). The A-site recognition occurs at the 30S subunit where the proofreading of the tRNA occurs to help bind cognate tRNA (Figure 6A). The anticodon stem loop (ASL) of tRNA is composed of ~26-44 nucleotides, which can be seen in base stacking interactions with the 16S subunit nucleotides G530, C1054, A1492, and A1493 (56). Interactions between tRNA and

A1492 and A1493 nucleotides are crucial for initial binding through A-minor interactions with the first two base pairs within the codon-anticodon duplex, while G530 interacts with the minor groove of the pairings within the A-site of the ribosome (30). A base stacking interaction occurs between C1054 and a G-C wobble pair within the tRNA ASL. These interactions are crucial for the recognition and proofreading of tRNA^{Phe} to be deacylated (70).

1.4.2. Interactions near the 3' end of the tRNA that facilitate peptide bond formation.

In the P-site, tRNA has been proofread and prepared to form the peptide bond via the peptidyl transferase upon recognition. The P-site functions as a clamp, with tRNA burying 19% of its surface area to either mRNA or the ribosome, in order to form the peptide bonds (70). P-site tRNA is in contact with many ribosomal protein tails providing extensive contacts for stabilization and binding. Base stacking interactions occur between one of the ribosomal tails and tRNA D-loop providing a strong interaction with peptidyl-tRNA (30). The backbone of mRNA codon interacts with nucleotides C1402, C1403, and U1498 in the 16S RNA. One important stabilizing interaction is the C1400 and G966 nucleotides, which base stack with a wobble pair. Mutation at the C1400 location decreases translation efficiency by 12-fold (70). Another important interaction which prevents the movement of tRNA during peptide formation is two A-minor interactions with nucleotides, A1339 and G1338, binding to two G-C pairs characteristic of initiator tRNA. With the tRNA held firmly in place, the 3' end of tRNA, containing the amino acid, allows for peptide bond formation in the peptidyl transferase center (PTC) in the 50S subunit. Stabilization of the 3'-end of tRNA occurs through two G-C base pairs between tRNA and the 50S subunits, along with two A-minor interactions. A

conserved base stacking and 2'-OH hydrogen bonding interaction between tRNA A76 has shown to be important for peptidyl transferase and substrate binding (56). After peptide formation, the A-minor interactions release, similar to a molecular switch, which allow the tRNA to translocate to the E-site (70).

1.4.2.1.1. Acceptor stem and TΨC-D loop interactions by the ribosome

The E-site facilitates the exit of fully deacylated tRNA after peptide formation. Mobile protein segment, L1 stalk, interacts with the conserved dual T-loop motif within tRNA acceptor end (56). This creates significant base stacking interactions to stabilize the deacylated tRNA (Figure 6A). T-loops from the L1 stalk and the tRNA create 9 base stacking interactions along with A76 2'-OH group A-minor stabilization and intercalation between a G-A pair in 23S RNA (56). Once these interactions occurs, tRNA is pivoted about the junction and ejection from the E-site occurs (56). The tRNA elbow region containing the dual T-loop motif is vital for the recognition of the ribosome as well as many other small non-coding RNAs.

1.5. Common features of large RNAs that Recognize transfer RNA (tRNA)

1.5.1.1.1. Ribonuclease P (RNase P), T-box riboswitch, ribosome

Importance of these tertiary interactions on the function of RNA has been well

Genetic elements: introns and riboswitches (11 T-loops)	A = 9.1% G = 0% C = 9.1% U = 81.8%	A = 0% G = 90.9% C = 0% U = 9.1%	A = 72.7% G = 0% C = 27.3% U = 0%	A = 54.5% G = 45.5% C = 0% U = 0%	A = 81.8% G = 9.1% C = 0% U = 9.1%
Consensus sequence	U	G	A	A/G	A
Overall frequency	A = 6.7% G = 1.9% C = 6.7% U = 84.7%	A = 0% G = 47.6% C = 0% U = 52.4%	A = 35.2% G = 8.6% C = 49.5% U = 6.7%	A = 50.5% G = 49.5% C = 0% U = 0%	A = 92.4% G = 4.8% C = 0% U = 2.8%
Consensus sequence	U	U/G	C	A/G	A

Figure 7. Sequence conservation of T-loops within 11 introns and riboswitches and overall all sequence conservation of T-loops, following the UUNRA or UGNRA sequence (6).

established. Tertiary motifs in solution have not been well characterized. In this study, we examine the T-loop tertiary motif in an attempt to characterize the hydrogen bonding network, as well as examine the binding thermodynamics. The T-loop is found in a wide variety of RNAs, including; the multi-turnover ribonuclease P-based ribozyme, the ribosome, and the bacterial tRNA-sensing riboswitch genetic element all recognize the TΨC/ D loop region of tRNA and tRNA-like molecules (6, 7, 10, 12, 16, 19, 21) (Figure 6).

1.5.1.1.2. T-loop Motif

The T-loop was first identified in the TYC-loop in tRNA where its name originates (31). The T-loop is found in a wide variety of non-coding RNA molecules, from riboswitches to transfer-messenger RNA. T-loops facilitate long range intra-molecular interactions for stabilization of RNA tertiary structure and intermolecular interactions to bind RNA or other ligands (12, 32). They can be found as a single or dual T-loop, where the dual T-loop intercalates with itself to provide stability through base stacking interactions (Figure 8). T-loops are found in both proximal and distal portions of

RNA molecules, cap long and short stems, and add functional diversity to noncoding RNAs with their ability to act as receptors (6).

1.5.1.1.3. Sequence conservation of single T-loops

The T-loop structure is not necessarily sequence based but generally contains either a UUNRA or UGNRA U-turn, where N is a purine and R is any nucleobase (12). Nucleotides at Positions 1 and 5 close the T-loop, typically in a noncanonical U-A base-pair, known as a reverse Hoogsteen base pair, which occurs 84 and 92% of the time respectively (6, 12) (Figure 7). Otherwise, the closing base pair is an A-A as seen in RNase P. A large gap between Positions 4 and 5, purine A or G, allows for intercalation, base stacking interactions, and potential hydrogen bonding to the Position 2 nucleobase, which is either an U or G. Position 2 never contains either an A or C at that location, most likely due to decrease imino hydrogen bonding potential. Position 3 is usually either a C or A, 49.5 and 35.3% of the time, respectively, that is involved in tertiary interactions (6) (Figure 7).

1.5.2. Hydrogen bonding and Dynamics of T-loop formation

1.5.2.1.1.1. Dual T-loop

The dual T-loop formation begins with the intra-molecular intercalation of two T-loops. Closing base pairs from each individual T-loop usually form base stacking interactions with each other. The highly conserved guanosine nucleobase at position 2, in both T-loops, interacts with the 5' closing base pairs phosphate backbone and forms a stabilizing stacking interaction. If intercalation occurs between position 4 and 5, then potential base pairing occurs between intercalated nucleobase and the nucleobase at

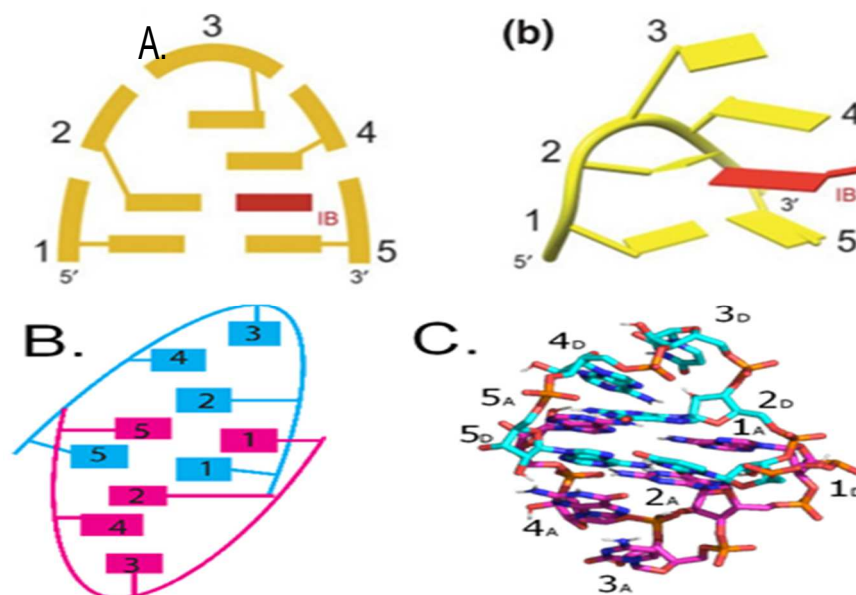


Figure 8. A.) Schematic diagram of single T-loop structure with base stacking interactions along the nucleotides 3 to 5, including intercalation of nucleobase in between base 4 and 5 (1). B.) Cartoon diagram indicating the intercalation and base stacking interactions occurring within the B.sub dual T-loop-sC.) Pymol representation of an intercalated dual T-loop of B.sub T-box riboswitch. Distal (D) T-loop shown in white and apical (A) T-loop pink (PDB: 6POM).

position 2 (Figure 8A). These interactions form a very stable dual T-loop structural motif capable of binding other RNAs or ligands with the free nucleobase at the third position. The multi- turnover ribonuclease P-based ribozyme, the ribosome, and the bacterial tRNA- sensing riboswitch genetic element all recognize the T Ψ C/D loop region of tRNA via this dual T-loop structural motif (6, 7, 10, 12, 16, 19, 21).

1.1.6.4.2. Interlocked Dual T-loops

Interlocked T-loop structure is closed by Hoogsteen U-A pair or noncanonical A-A (N7-amino). At position 2, in bacterial RNA, hydrogen bonds through the 2'-hydroxyl group of the closing base pair (6). Conserved purine at position 4, typically, forms a Hoogsteen/sugar edge hydrogen bond with the 2'-hydroxyl at position 2. In RNase P

RNA, two base triples are formed to create a six nucleotide base stack with the intercalation of both nucleotides at position 2 (6, 36, 50) (Figure 8C). Depending on the sequence conservation, it is generally an A-G-A triple followed by a U-G-A triple. Nucleotide at position 3 are free to interact and bind substrates through base pairing or, preferably, base stacking interactions.

1.7. Introduction to biophysical methods to examine large RNA molecules

Examination of these large RNA molecules is crucial to our understanding of RNA processing within the cell. Misregulation of noncoding RNAs have elucidated the role of them in many diseases including cancer, metabolic disorders, and cardiovascular diseases (68). Formation of these structures is key to helping our understanding. Through extensive biochemical methods, we analyze the biophysical characteristics of B.sub T-box riboswitch.

1.7.1. Biochemical analysis

1.7.2. X-ray crystallography-strengths and limitations

X-ray crystallographic methods provide detailed information on structure in three dimensional crystal lattice. Through careful crystallization techniques, one can determine base stacking, electrostatic, and overall tertiary structure of biomolecules. Previous studies have examined the crystal structure of B.sub T-box riboswitch with pretty high resolution ($\sim 3.8\text{\AA}$) (51). There are limitations to this methodology including the most difficult part which is obtaining a crystal. Usually, RNA constructs would be screened against hundreds of crystallization conditions. Another shortcoming is typically poor resolution which requires molecular refinement against a known structure or crystallization of mutant constructs to analyze specific regions. In order to obtain

detailed, unambiguous base stacking and other structural details, it is necessary have near perfect crystals.

1.7.3. Cryo-Electron microscopy (Cryo-EM), strengths and limitations

Cryogenic electron microscopy (cryo-EM) is a structural analysis method for biomolecules, similar to X-ray crystallography as it provides a 2D image that can be transformed into a 3D structure. Cryo-EM differs from crystallography in that it uses electrons versus light to determine the 2D image gathered by the detector. Cryofixation is used by flash freezing the protein or RNA sample in water to form a vitreous glass which does not disrupt the molecular structure (13). Cryo-EM can produce near atomic resolution images with really no limit to the resolutions obtained. Wavelengths used in these experiments are typically $<0.02\text{\AA}$ indicating the hardware is not the issue with the resolution of the images. Single-particle cryo-EM experiments can be applied to achieve even lower resolution and discriminate between alternative structures. The drawback to cryo-EM is that it is very dependent on the homogeneity of the structure being analyzed and can generate multiple images. Overall, cryo-EM can be a very exciting method for structural analysis especially as modeling technology improves with time.

1.7.4. NMR spectroscopy, strengths and limitations

Nuclear magnetic resonance spectroscopy is the only unambiguous way to assign the exact hydrogen bonding network for biomolecules. Through the use of magnetization, detailed high resolution spectra can be gained to obtain insight into base pairing, conformational equilibria, structure dynamics, and interactions with small ligands. Using coupling constants and dipole-dipole interactions, biomolecules can be assigned with a high degree of certainty. Limitations do arise such as limited chemical-shift dispersion in

RNA, anisotropic with RNA, and limits to size of RNA analyzed. However, these limitations can be overcome with detailed experimentation and isotopic labeling.

1.7.5. Strategy for assignments

RNA is much more difficult to assign than proteins due to only four nucleobases leading to decreased chemical dispersion. Due to this, limits on the size to <50 nucleobases are required for high resolution data. In general, these can be overcome with the use of isotopically labeled experiments and segmental overlapped analysis, but assignments are very difficult. Typical RNA strategy is to start with an assignment of unlabeled imino base pairs with a 2D H₂O NOESY experiment, followed by a 2D D₂O NOESY for assignment of aromatic and sequential walk protons. Conformation of assignments are provided through isotopically labeled HNN-COSY, HSQC, and HMC experiments. Through a combination of these experiments provides a map for the unambiguous assignment of the secondary structure of B.sub WT T-box riboswitch.

1.8. Scope of Study

The processing of pre-tRNA is an essential part of translation that requires recognition of an RNA based enzyme, ribonuclease (RNase) P, to cleave the 5' end. The recognition of the pre-tRNA docking elbow occurs through a specific, phylogenetically conserved RNA motif, known as a T-loop (7). The T-loop motif is an important RNA motif found in a wide variety of processing RNAs, including riboswitches, ribosomal RNA, and some mRNAs (6).

The importance of secondary and tertiary motifs in RNA folding has been well described and provides an important functional role in processing pre-tRNA and gene expression in the cell. The structurally conserved dual T-loop motif across the three

classes of RNA indicates convergent evolutionary strategies for the processing and recognition of the pre-tRNA or tRNA-like molecules. While x-ray crystal and cryo-EM structures give insight into the complex formation and base stacking interactions, little is known about atomic-level structural details of T-loop pre-assembly or thermodynamic parameters associated with tRNA binding.

The main goal of this study is to delineate the secondary structure of *Bacillus subtilis* glyQS T-box riboswitch non-coding mRNA and determine the hydrogen bonding network of base pairs within the dual T-loop motif. T-loop RNA motifs are an essential stabilizing secondary structure, yet not much is known about how they assembled or stabilized or their hydrogen bonding network. The following Master's thesis project defines the secondary structure and seeks to determine the hydrogen bonding network of a dual T-loop motif that exists within the bacterial mRNA *Bacillus subtilis* glyQS T-box riboswitch RNA. Biochemical characterization and multidimensional NMR spectroscopy provide complementary approaches to gain insight into the conformational dynamics of this universal dual T-loop RNA motif. In this study, a 45 nucleotide construct of B.subWT was generated, containing both distal and apical T-loops for examination through extensive NMR spectroscopic methods. In addition, UV-absorbance was used to determine the melting temperature (T_m) and circular dichroism was used to identify potential spectral features the dual T-loop RNA tertiary motif. Finally, isothermal titration calorimetry was performed to examine thermodynamic binding properties and specificity of B.subWT dual T-loops to tRNA

CHAPTER 2

2.1. Introduction

2.1.1. *Bacillus subtilis glyQS T-box riboswitch Structure and Function*

Bacillus subtilis (B.sub) glyQS is a Gram-positive bacteria T-box riboswitch— a type of non-coding RNA located in the 5'-untranslated region of mRNA— that binds to tRNA to regulate glycine expression. Riboswitches are regulatory elements that sense the aminoacylated state of tRNA and can change their conformation based on the charge of tRNA. Riboswitches contain two domains: a aptamer domain and an expression platform. The aptamer region binds to the specified ligand, while the expression platform adopts a favorable conformation based on charged state of tRNA. Similar to RNase P, the highly conserved dual T-loop motif of B.sub T-box riboswitch (T-box) recognizes the TYC and D-loop of tRNA in order to start the processing.

2.1.2. *Evolutionary similarity with in comparison to RNase P and rRNA.*

The T-box riboswitch is believed to coevolve with the mRNAs. All T-box riboswitches contain highly conserved secondary structural elements, which support gene- specific selection of aminoacylated-tRNA (53*). Sequence analysis of T-box riboswitches show highly conserved T-box regions within the across the bacterial domain (63). T-box regions contain the same highly conserved T-loops as RNase P and the L1 stalk in the 50S ribosomal subunit (53). However, the bacterial B.sub T-box riboswitch binds in the opposite orientation compared to RNase P and 50S ribosomal subunit (71). Due to the contrasting conformational binding, it is hard to infer that T-boxes are direct ancestors of RNase P, but their sequences similarities and recognition of aa-tRNA may

indicate that T- boxes independently evolved from RNase P or use a conserved structural scaffold for tRNA recognition (71).

2.1.3. Structure and function of *B.sub* T-box riboswitch

The structure of *B.sub* T-box riboswitch was determined through the cocrystallization of *Geobacillus kaustophilus* glyQ T-box at a resolution of 2.66Å combined with electron cryo-microscopy (cryo-EM) with a resolution of 4.9Å and small-

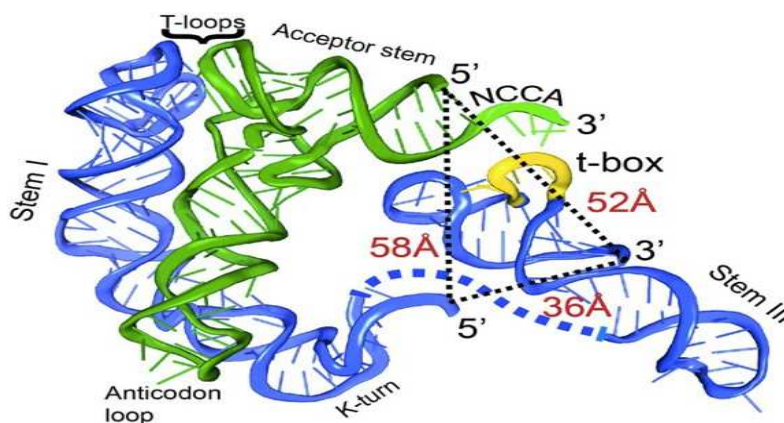


Figure 9. Crystal structure of the *B.sub* glyQS T-box riboswitch (blue) and *B.sub* tRNA^{Gly} (green) complex based on SAXS data. *B.sub* T-box riboswitch locks tRNA^{Gly} into place via specific binding interaction (labeled) for gene regulation (36, 73)

angle X-ray scattering (SAXS) to confirm and provide atomic-level structural information of the T- box riboswitch (72). The examination of the 66 nucleotide length Stem I in the absence of proteins showed intercalation between the interdigitated T-loop structure of the riboswitch and tRNA. The T-box utilizes the same five nucleotide sequences as RNase P for both T-loop motifs. The distal loop is closed by an A-A base pair, but intercalation of the apical G63 T-loop at the second position creates a stabilizing base triple with nucleotides C44 and A56 (71). Intercalation of the distal G at the second position forms an adjacent base triple with the reverse Hoogsteen U-A pair adjacent to

the first base triple. The uracil at Position 3 of the apical T-loop base stacks directly with the CT3 of tRNA TYC T-loop creating a continuous 6 nucleotide base stacking interaction, which stabilizes the complex (6, 36, 72). This is a non-specific interaction for the stabilization of tRNA for processing of the 5' end. While X-ray and cryo-EM structures help determine the complex formation, little is known about the atomic-level structural details of the conformation of the dual T-loop structure of B.sub.

2.1.4. Conformational dynamics of B.sub T-box riboswitch dual T-loop

The structure of riboswitches contain a Stem I— the anti-terminator domain— and a small-to-large Stem III to connect the domains. In the adenosylcobalamin-sensing riboswitch with cofactor B12, T-loops are used to bind to P4 and P6 to add functionality. Crystal structures from three different bacterial riboswitch, each with varying sequences (5'-UGCAA-3', 5'-UGAGA-3', and 5'-UGUAA-3'), all form the same T-loop structure and bind to P4 and P6 regions of B12 enzyme (6). This indicates the conserved structural nature of the T-loop independent of sequence. In riboswitches, the sequence consensus is 5'-UGUNA-3' where N=A or G (6). Upon binding to tRNA the riboswitch forms a cradle-like structure. The interdigitated T-loops form a six layered stack with tRNA inducing pronounced conformational changes to both tRNA and Stem I of B.sub T-box riboswitch. The dynamic nature of this interaction and sequence independence in other riboswitches indicates a stabilizing interaction with the dual T-loop structure. Point mutations of the B.sub T-box riboswitch were made to examine the native fold with mutations in the T-loop region, structural characteristics associated with melting and circular dichroism, and to gain insight into tRNA binding thermodynamics with the various mutants.

2.2. Material and Methods

2.2.1. Materials

Sodium Chloride (NaCl), Potassium chloride (KCl), Monopotassium phosphate (H_2KPO_4), Potassium dihydrogen phosphate (HK_2PO_4), Magnesium chloride (MgCl_2) Tris-base, Sodium hydrogen phosphate (Na_2HPO_4), Glycerol, Diethylaminoethyl (DEAE)-Sphacel, Nucleoside triphosphates (NTPs), Dithiothreitol (DTT), RNasin Ribonuclease Inhibitor (Promega Corporation) Tris-HCl, G25-Sephadex, Spermidine, Bovine serum albumin (BSA), Boric acid, Ethylenediaminetetraacetic acid (EDTA), Ethanol 200 proof, NaOAc, Acrylamide (19:1), Ammonium Persulfate, and tetramethylethylenediamine (TEMED) were purchased from Sigma-Aldrich, VWR, or Fisher Scientific. Single stranded Tie-1 and antisense Tie-1 and guanosine quadriplex (GQ) RNAs for native gel analysis were donated by Ainur Abzhanova and Dulmi Senanayaka, respectively. All reagents were purchased commercially and were of the highest purity available.

Pre-packed 5 mL Ni-NTA and 10 mL Econo-Pac Disposable Chromatography columns were purchased from Bio-Rad. Glass plates, combs, and spacers were purchased along with FischerBiotech Electrophoresis System. All the solutions including buffer solutions were prepared using Nano pure water ($\sim 18.2 \text{ M}\Omega$) from Milli-Q® integral water purification system for ultrapure water and filtered using $0.2 \mu\text{m}$ filters.

2.2.2. Instruments/Apparatus

VWR scientific water bath incubator was used in transformation and transcription of DNA. Beckman coulter bench top and Avanti centrifuge were used for spinning cell lysate, concentrating, and pelleting RNA. FischerBiotech Electrophoresis System FB 600

was used for all gel electrophoresis. A Mineralight multiband UV lamp was used for visualizing RNA bands in gel electrophoresis. Labconco FreeZone 2.5L lyophilizer was used for drying all RNA. Thermo Scientific Nanodrop was used for measuring protein and RNA concentration. Collection of UV-melt and circular dichroism data were collected on Chirascan V100. Data was analyzed using the Chirascan software program. Isothermal titration calorimetry data was collected on a ITC-200 Micro-calorimeter at 25 °C. Binding curves and thermodynamic parameters were determined by nonlinear least-squares fitting analysis using Origin software (OriginLab).

2.2.3. RNA Transcription and purification

DNA Oligonucleotides for truncated constructs of *Bacillus subtilis* (B. sub) glyQS T-box riboswitch were purchased from Integrated DNA Technologies and transcribed using T7 RNA polymerase standard transcription protocols in vitro (65). The secondary structures of the B.sub glyQS T-box riboswitch and mutant sequences were designed using the sequence conservation analysis of riboswitch and RNase P T-loops (6). Constructs were examined using MFOLD (74) to assess whether structural elements were influenced by point mutations. RNAs were purified from the transcription reaction through a 10% denaturing (8M Urea) polyacrylamide gel. RNAs were excised from the gel after UV shadowing analysis identified the band of interest. RNA was eluted with 300 mM sodium acetate (pH 7.0) through overnight rocking at 4 °C. Eluted RNA was precipitated with three volumes of 100 % ice-cold ethanol and stored overnight at -20 °C. Precipitated RNA was centrifuged for 20 minutes at 4300 rpm and dried at room temperature for 1 hr. The RNA concentration was measured by absorbance at 260 nm

with ThermoScientific™ NanoDrop™ UV-Visible spectrometer after resuspension of the pellet in 200 μ L of DEPC water.

Purification was initiated with an anion exchange column (Econo-Pac Disposable Chromatography column, 10mL) with a gradient mixture of low salt (50 mM KPO₄, 100 mM EDTA, and 100-200 mM NaCl) buffer and high salt buffer (50 mM KPO₄, 1.5 M NaCl) with the collection of one mL fractions until the RNA was eluted. After elution, RNA fractions were analyzed on the nanodrop, pooled, and ethanol precipitated with 1.5M NaCl and placed in the -20 °C fridge overnight. RNA was pelleted, resuspended in 200 μ L of DEPC water, and desalted with G-25 Sephadex. RNA was subsequently lyophilized, and kept frozen and stored at -20 °C. An identical procedure was followed for the point mutant constructs with an additional BioRad desalt spin column used for isotopically labeled samples.

2.2.4. Native gel folding

Native folding of RNAs took place in cold room (4°C) and all buffer solutions were kept at 4°C prior to the start of the experiment. Native gels were made with final concentration of 6% and contained 1X THE, 50mM NaCl, and 10mM MgCl₂. Gels were pre-ran at 50V and wells were washed with 1X THE buffer before addition of samples. Samples were folded using PCR protocol 1 (Series: 92°C for 3 minutes, 40°C for 2min, 50°C for 15min, and 37°C for 30min) where addition of MgCl₂ was added to samples after two minutes at 4°C for optimal fold. Samples were folded in various folding buffers, as previously described, at a concentration of 5 μ M with 1 μ L of glycerol added to the sample just before loading (8, 20, 32, 72). Once samples were loaded, 1X THE dye was added to ensure proper distance traveled by the samples. Gel ran at 50V for

approximately two hours in the 40°C room and visualized with toluidine blue staining dye for 10 minutes then destained with distilled water for 30 min. A picture was taken upon removal of the gel from destaining solution for further analysis.

2.2.5. Electrophoretic Mobility Shift Assay and Isothermal Titration Calorimetry for binding analysis

Labeling of the 5' end with ^{32}P of tRNA

2.5 μ l of tRNA (40 μ M) was added in a single RNase-free microcentrifuge tube along with 3.75 μ l dephosphorylation buffer (2 M Tris, pH 8.0), 5 μ l calf intestinal phosphatase (CIP; 0.1 U/ μ l), and filled to a final volume of 150 μ l with nuclease-free water. Sample was incubated for one hour at 37°C. Phenol chloroform extraction was performed by the addition of 100 μ l of 50 mM KOAc/200 mM KCl pH 7 and 500 μ l phenol chloroform. Phenol extracted by spinning the tube at 4 °C in a centrifuge at 4300rpm for 5 minutes. Supernatant was transferred to a clean tube, ethanol precipitated, and washed the pellet once with ice cold 200 proof ethanol and dried. Kinase reaction was performed as described (16), by mixing 3 μ l tRNA (1.4 μ mol tRNA) and 2 μ l [γ - ^{32}P] ATP (7000 Ci/ mmol) in 2 μ l 10X Kinase Buffer (500 mM Tris, pH 7.5, 100 mM MgCl₂, 50 mM DTT). The reaction was incubated at 37 °C for one hour to optimize phosphorylation. Spin columns removed excess γ - ^{32}P] ATP (7000 Ci/mmol) and was washed with water.

2.2.5.1. Electrophoretic Mobility Shift Assays

B.sub glyQS T-box riboswitch RNA and tRNA(Gly) were folded separately in folding buffer as previously described (21, 73). B.sub T-box RNA and point mutants were added at varying concentrations to 10 μ M tRNA(Gly) for a final volume of 10 μ l and were incubated at room temperature for 30 minutes. All experiments were run on a 6%

acryl- amide native gel (1X THE, 50mM KCl, 15mM MgCl₂) at 120 volts at room temperature. Gels were visualized using with UV light and/or stained with toluidine blue.

2.2.5.2. Isothermal Titration Calorimetry

Thermodynamic parameters associated with tRNA(Gly) binding to B.sub glyQS T-box riboswitch RNA were determined using an ITC-200 Micro-calorimeter at 25 °C. Before the experiment, two interacting RNAs (in 10 mM KPO₄, 50 mM KCl, 10mM MgCl₂) were folded as previously described (64). For preliminary binding analysis of the dual T-loop B.sub RNA with tRNA, the sample cell was filled with 300 µl of tRNA at a concentration of 50 µM and B.sub was titrated into the cell at 600 µM. For each ITC experiment, the titration was carried out by a stepwise (2 µl) injection of B.sub from a stirred syringe (1500 rev/min) into the sample cell. Injections were spaced by 150 seconds and values for the change in enthalpy, association constant, and stoichiometry were determined by nonlinear least-squares fitting analysis using Origin software (OriginLab).

2.2.6. Circular dichroism and UV Melt temperature studies

RNA constructs (in 10 mM KPO₄ and 100 mM (NH₄)₂SO₄) were folded at a concentration of 45 µg/mL or 1.0 absorbance units as previously described and compact fold was tested on native gel as described above (64). Experiments were performed on a Chirascan V100, collecting on a range of 20-90 °C measuring at 5 °C increments at an absorbance range of 190-280 nm centered on 260 nm. Data was analyzed using the Chirascan software program. Melting temperature curves were determined using first derivative plot on the secondary axis. Thermodynamic parameters were calculated by plotting the ln(K_a) vs. 1/T(K-1) and formulating trend-lines, which were used to solve for enthalpy, entropy, and Gibbs free energy of dissolution.

2.2.6.1. Calculations of thermodynamic parameters for unfolding of B.sub T-box Riboswitch following changes in ellipticity as a function of temperature

Initial values for ΔH , T_M , and $[\theta]_F$ and $[\theta]_U$ are estimated through a nonlinear least squares fitting routine. Values for best fit to raw data are found in fitting procedures previously described based on assumption of a homogeneous monomer (18). Molecules undergo a transition from folded (F) to unfolded (U) and determination of the association constant of folding can be represented by the equation (1):

$$K = [F] / [U] \quad (1)$$

which can be used to calculate fraction folded at any temperature through the equation:

$$\alpha = [F] / ([F] + [U]) \text{ or } \alpha = (\theta_t - \theta_U) / (\theta_F - \theta_U) \quad (2)$$

where θ_t is the ellipticity at any temperature, θ_F is the ellipticity of fully folded RNA, and θ_U is the ellipticity of fully unfolded RNA (59). Fraction folded is constrained to $0.15 < \alpha < 0.85$. This determines the association constant, K_a , which is used to determine enthalpy and entropy. To fit the ellipticity changes as a function of temperature (K) one uses the Gibbs-Helmholtz equation 3.

$$\ln(K_a) = \Delta H/R (1/T(K) + \Delta S/R) \quad (3)$$

This will give a figure shown in the appendix where two trendlines can be fit to determine thermodynamic parameters (Appendix Figures 28, 30, 32, 34). Melting temperatures were determined from first derivative of UV-melt data.

2.3. Results and Discussion

The binding of the dual T-loop motif to tRNA is an essential step for all T-box riboswitches. This interaction is found in a variety of other RNA

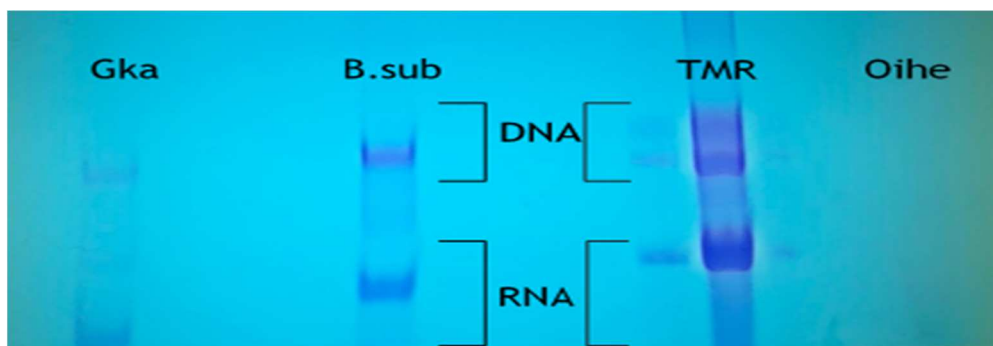


Figure 10. Test transcription indicating transcription viability of Gkai glyQS T box riboswitch, B.sub T-box riboswitch WT, TMR RNase P, and Oihe T-box riboswitch, respectively, for qualitative analysis of RNA yield. Oihe T-box construct did not transcribe while others indicated fair amount of DNA product

molecules that bind to pre-tRNA and tRNA-like substrates. Elucidation of the secondary structure and thermostability of the truncated T-box riboswitch construct is important to gain insight into the formation of the dual T-loop and the properties of binding to tRNA. RNA molecules have a variety of complex folding patterns and are dependent on base stacking, divalent metal cations, noncanonical and canonical base pairing, and phosphate backbone structure. Our goal in this study was to determine optimal folding conditions for formation of dual T-loop structure and to investigate the importance of sequence conservation within the T-loop. ITC, CD, native gel shift analysis, and UV-melting studies provide insight into the global structure and thermodynamic parameters associated with tRNA binding to dual T-loop containing RNAs.

2.3.1. Transcription of homogeneous RNA constructs

Initially, four truncated constructs were created based on previous crystal structures of B.sub T-box riboswitch WT, *Oceanobacillus iheyensis* (Oihe) T-box riboswitch, *Geobacillus kaustophilus* (Gkau) glyQS T box riboswitch, and *Thermotoga maritima* (TMR) RNase P RNA. These sequences from various types of RNA were







chosen based on their ability to bind aa-tRNA in the cell with the highly conserved dual T-loop structural motif (6, 36, 50, 71). DNA sequences were truncated from Stem I regions to 45, 39, 42, and 37 nucleotides (30-75) from the full length structures, respectively, in order to contain two helical regions, Stem Ia and Stem Ib, and the distal and apical T-loop regions (74) (Table 1). Constructs were made to contain a minimal dual T-loop motif based on sequence conservation across various classes of RNA (Figure 7). Secondary structure prediction software, MFOLD and ViennaRNAfold, were used to analyze thermodynamic stability and conformational homogeneity associated with the predicted truncated fold (17). Truncated T-loop structures were made if the difference in predicted free energy of the ensemble differed by at least at 0.3 kcal/mol between the nearest alternative conformation. These programs use nearest neighbor energy parameters to find the minimized energy secondary structures of single stranded RNA (72).

Test transcriptions were performed for qualitative analysis of transcription yield, which indicated Oihe was a poor transcriber shown by the purity gel (Figure 10). Other transcripts were indicated the presence of RNA but in low quantities compared to DNA transcripts and NTPs. Optimization of the transcription reaction with an increase of

Table 1. DNA sequences for truncated constructs to examine dual T-loop motif formation across phylogeny. Bolded is the DNA construct used for experimentation.

Gene	RNA type	DNA Sequence
<i>Oceanobacillus ihayensis</i>	T-box Riboswitch	5'-TAA TAC GAC TCACTATAG GGT GAT TAG CGA CTC TAG GAT AGT GAA AGC TAG AGG ACC-3'
<i>Bacillus subtilis</i>	T-box Riboswitch	5'- TAA TAC GAC TCA CTA TAG GGC TCA TGA AAG CGA CCT TAG GGC GGT GTA AGC TAA GGA TGA GCC-3'
<i>Geobacillus kaistophilus</i>	T-box Riboswitch	5'-TAA TAC GAC TCA CTA TAG GGT TCC GAA GCG AGC CGG GGA GGG TGG CGG CCC GGT GAA C C-3'
<i>Thermotoga maritima</i>	RNase P	5'-TAA TAC GAC TCACTATAG GGC CAT AGA GAAG AT GGG TAACCG GGT GGAACG GCC C-3'

Table 2. RNA sequences of the B.subWT and mutant constructs with predicted secondary structure depiction.

RNA Construct	Sequence (5'-3')	Vienna Fold (Entropic structure)
tRNA (Gly)	GGGAAGUAGUUCAGUGGUAGAACACCACCU- UGCCAAGGUGGGGGUCGCGGUUCGAAUCC- CGUCUCCCC	
B.sub WT	GGCUC AUG AAAGCG ACCUAGGGCG- GUGUAAGCUAAGGAUGAGCC	
B.sub C13U	GGCUC AUG AAAGUG ACCUAGGGCG- GUGUAAGCUAAGGAUGAGCC	
B.sub A30G	GGCUC AUG AAAGCG ACCUAGGGCGGUGU- GAGCUAAGGAUGAGCC	
B.sub A11U/C13A	GGCUC AUG AAUGAG ACCUAGGGCG- GUGUAAGCUAAGGAUGAGCC	
B.sub Helix (U7G/ U19C)	GGCUCAGGAAAGCGACCUCAGGGCG- GUGUAAGCUGAGGCUAGCC	

guano- sine triphosphate concentration to 7 mM and the addition of a second equivalence of T7 RNA polymerase after one hour significantly increased the qualitative yield for most RNAs except for the RNA from *Oceanobacillus iheyensis* (Oihe). Large scale transcription reactions were performed on the B.sub, Gkau, and TMR constructs to obtain a quantitative yield and perform conformational analysis. Initial transcriptions yielded low quantities (approximately 0.1-1 miligram) of RNA in 5 mL transcriptions, but through the use of new 5X transcription buffer, (200mM Tris-HCl pH 7.5, 50 mM Spermidine, 25mM BSA, 1 mM MgCl₂, and 10 mM DTT) transcription yield increased to approximately 14 milligrams for Bsub and Gkau, where the TMR transcription yield



Figure 11. Homogeneous fold of B.sub WT T-box comparison to dimer Gkai and heterogeneous TMR.

was approximately 2.3-5.9 milligrams. The increase in yield is due to the stability of the DNA transcripts, freshly prepared buffers, and the structures of the RNA product.

2.3.2. Determination of conformational homogeneity and ion dependence.

Initial folding conditions were tested based on previous folding of full-length constructs (8, 20, 74). Folding in HEPES buffer (50 mM HEPES pH 7, 100 mM KCl, 10 mM MgCl₂) yielded homogeneous B.sub fold but the Gkai fold was not as stable and required higher concentrations of Mg²⁺ (20, 21). Various folding temperatures were tested on the Gkai, B.sub, and TMR constructs, using both, standard and stepwise. Folding conditions had little effect on the Gkai and B.sub constructs but stepwise folding provided a more stable fold for TMR. Previous studies examined the effects of Mg²⁺ on folding of RNA, which showed a preference for higher, 3mM-17mM, concentrations leading to a more compact fold (46). Examination into Mg²⁺ dependence at 2mM and

Table 3. Buffer conditions for optimized folding of RNA constructs for the biochemical analysis experiments.

Experiment	Buffer	pH
ITC	50mM HEPES, 100mM KCl	7.0
	10mM KPO ₄ , 100mM KCl	7.2
NMR	10mM KPO ₄ , 50mM KCl	7.2
CD	10mM KPO ₄ , 100mM NH ₄ (SO ₄)	7.0

10mM concentrations indicated some differences for the Gkai under native gel conditions as the 42 nt construct migrates higher than the B.sub construct at 45 nt (Figure 11). Folding of the TMR RNase P construct showed poor homogeneity in the folding analysis (Figure 11). Conformation variability in the Gkai and TMR construct excluded these RNAs from further examination due to heterogeneity unacceptable for NMR experiments. Since TMR had low transcriptional yield and alternate conformations under the conditions tested, it was also excluded from further biochemical characterization studies.

Confirmation of the structured B.sub T-box construct was further examined and compared to known structured and unstructured RNAs. Previous studies of longer T-box riboswitch constructs (182mer) indicated refolding into multiple conformations after

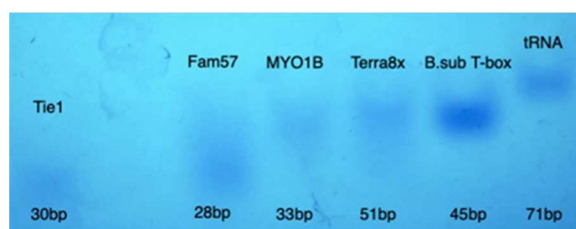


Figure 12. A.) Dimerization analysis of B.sub T-box construct (Lane 5) indicates monomeric fold compared to unstructured single-stranded RNA (Lane 1) and structured G-quadruplexes (Lanes 2-4), and tRNA (Lane 6)

purification through denaturing methods (8). Native folding tests were performed to ensure no dimerization was occurring within the B.sub T-box structure. Unstructured single stranded RNAs (termed Tie-1 and anti-sense Tie-1) allow us to compare folded versus unfolded RNA, with lengths of 25 and 29 nt, respectively. This indicated that B.sub T-box folded as a monomer in solution and was not degraded into individual nucleotides. Structured G-quadruplexes (GQ); Fam57, MYO1B, and Terra8x, as well as tRNAGly maintain robust folded structures at lengths of 28, 33, 51, and 71 nt, respectively (Figure 12). Although B.subWT T-box construct looks close to MYO1B in length, formation of GQ structure causes a slower movement through the gel due to a bulky conformation. 71mer tRNA was used to ensure no dimerization, as dimerized T-box would be approximately 90 nt in length. Although, the separation between B.sub T-box and tRNAGly is not indicative of a 26 nt difference, this may be due to the difference

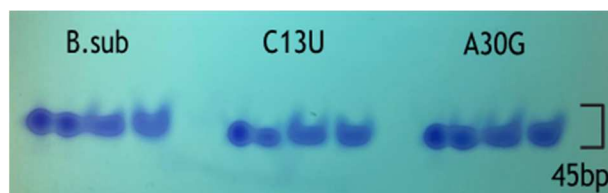


Figure 13. Compact folding of B.subWT, C13U, and A30G at 45bp in folding conditions listed in Table 3.

in compaction. Comparison of the structured and un- structured RNAs, along with B.-sub WT construct, indicates no dimerization during folding and shows proper folding length of B.subWT, as it is in between TERRA 8x (51nt), FAM57(28nt), and MYO1B (33nt). Optimization of folding conditions was completed with no alternative conformations which allows for NMR analysis to be performed.

In addition to the B.sub T-box WT construct, we transcribed three other variants of the monomeric B.sub T-box RNA. A mutant C13U T-box is a known mutation of the

distal T-loop that is vital for base stacking with the D-loop of tRNAGly (36, 71). A uracil at third position within the T-loop of a riboswitch is typically not observed but secondary structure prediction indicates that it may still form a dual T-loop fold (Figure 7). We analyzed the formation of the mutant T-loop and recognition of tRNAGly through a minimal T-loop construct. Our 45mer T-box C13U construct was analyzed via purity gel indicating proper length and homogeneous transcription (Figure 13). We expect to observe a similar difference in dissociation binding constants compared with previous studies, where this C13U mutation shows a ~40 decrease in binding affinity to tRNAGly compared to the full-length WT T-box riboswitch (69). Elucidation of the folding and binding properties of a mutation in the apical T-loop have not been studied yet. Mutation to the apical loop through an A30G was performed based on sequence analysis indicated preference for purine at the fourth position in the T-loop (Figure 7) (6). The predicted secondary structure of the A30G mutant indicates the formation of a tri-loop instead of a T-loop with the mutated G forming a slightly unstable G-C pair closing the tri-loop (Table 2). This interaction will provide insight into the importance of the apical T-loop in stability of the dual T-loop motif and examine binding affinity of the distal T-loop to tRNAGly. Lastly, a double-point mutant was made in the distal T-loop to a more universally conserved T-loop sequence (5'-UGAGA-3'). A future goal of this study is to compare binding affinities and thermodynamic stabilities of various dual T-loop sequences. Mutation of each position within the dual T-loop region will allow us to examine whether sequence conservation within the apical and distal regions of the dual T-loop is consistent with apparent binding specificity to tRNAGly.

2.3.3. Circular dichroism and UV-melt analysis of B.sub T-box riboswitch RNA

Circular dichroism (CD) is a quick, sensitive method that measures the absorption of left- and right-handed circularly polarized light by chiral molecules. CD is a relatively useful method for characterization of nucleic acid molecules at low concentrations and can determine conformational changes through titration experiments or isomerizations between two-state conformers (34). CD of RNA can be useful to differentiate between single-stranded and double stranded RNA, in addition to the type of helical conformation that the RNA folds into. CD spectroscopy is a very sensitive method and is dependent on temperature, pH, and salt concentrations and the entire range of the spectra should be analyzed in order to make inferences on structure (34). Folding conditions of constructs were modified to eliminate the use of chlorides due to decreases in spectral resolution at wavelengths below 210nm, which significantly improved the quality of the spectra (35). Our results were used to gain qualitative understanding of the nature of our B.subWT T-box construct structure in comparison to point mutants.

The CD spectrum for known A-form helical structure shows a dominant positive peak at 260 nm, a smaller negative peak 235 nm, and a negative peak around 210 nm. Conversely, a CD spectra with a broad 278 maxima, a minor peak at 237 nm (-), and 220 (+/-) features are diagnostic of a ssRNA. Analysis of CD at 25 0C over at a range of 195-300nm, indicates A-form helical structure can be seen for all constructs of the T-box, including mutants (Figure 14). B.sub T-box WT shows the typical positive and negative peaks at 260nm and 210nm, respectively, indicating a double-stranded A-form helical structure. All point- and double-point mutants obtain the same general CD profile confirming mfold predictions and native gel shift analysis of formation of compact tertiary structure folding. Subtle differences between point mutants and WT suggest the

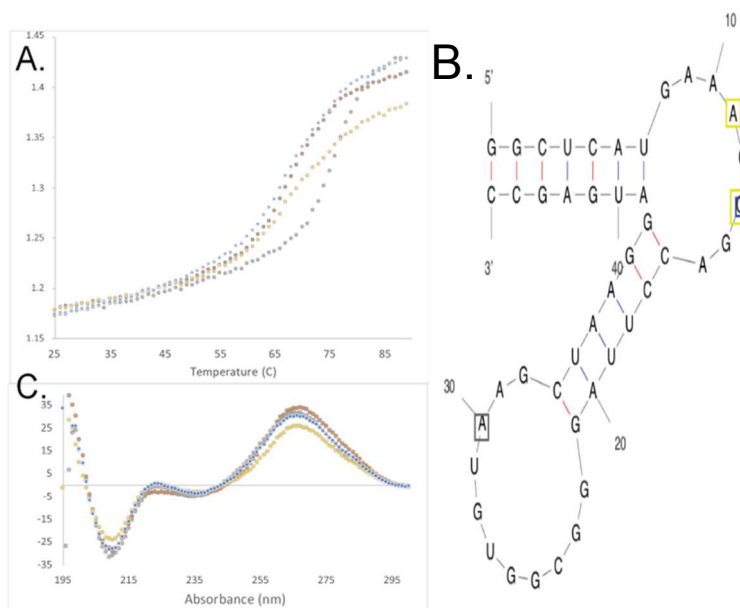


Figure 14. A.) UV-melting temperature studies indicate A30G (grey) has most thermostability with a $T_m = 77^\circ\text{C}$. A11U/C13A (yellow), C13U (blue), and B.subWT (orange) all have similar $T_m = 67, 67,$ and 66°C centered on an absorbance of 260nm. B.) Circular dichroism spectroscopy of B. sub T-box constructs A30G (grey), A11U/C13A (yellow), C13U (blue), and B.subWT (orange) and their qualitative formation of double stranded complex at 25°C measuring the absorbance on a range of 190-300nm centered on 260nm. C.) Secondary structure of B.sub wild type highlights the difference between the various point mutants, where each box is color-coded corresponding to the color on the scatter plot. Parameters for CD experiment were 20-90°C ramp @ 5°C increments with Abs. range of 190-300nm.

formation and stability of T-loops and additional analysis will be required to potentially differentiate CD signatures of various B. sub dual T-loop RNA constructs. Nonetheless, point mutants, C13U and A30G, maintain similar ellipticity pattern in defined A-form helical region. Interestingly, the double-point mutant has a noticeably smaller ellipticity in both ranges, possibly indicating a less stable T-loop structure or formation of a shorter helical structure. Further analysis would be needed to determine this observed difference.

A hypothesis could be made that A30G mutant forms stronger helical regions due to increased ellipticity in 210 nm region, which would be congruent with the predicted secondary structure of A30G believed to form a tri-loop more similar to a hairpin than a T-loop (Table 2). The main difference between point mutants and WT is seen in 220-235nm range, where the mutants all have a higher ellipticity compared to the B.subWT construct (Figure 14B). This region has been shown to be associated with hairpin turns but could be indicative of T-loop structure since no hairpins are seen within predicted secondary structures (35). Further examination on this region would need to be studied but could lead to a potential novel methodology for indication of the dual T-loop structural motif.

UV-melting temperature studies were performed on each of the mutants described above to examine melting profiles and determine thermodynamic properties associated with melting. UV-melting studies examines the transition between native structure and disordered nucleotides, where an increase in absorbance at 260nm is called hyperchromicity, leading to denaturing of the folded RNA (35). Our study analyzed the difference in thermodynamic properties between the various T-box WT and point mutants. Melting temperatures were measured at 260nm at a temperature range of 25-90 0C with increasing increments of 1 0C per min. First derivative analysis provided the T_m for T- boxWT, C13U, A30G, and A11U-C13A mutants at 68, 70, 77, and 67 C, respectively (Table 4, Appendix Figures 27, 29, 31, 33). C13U mutation was not believed to cause much change in melting temperature due to the mutation of the free base pair at the top of the T-loop which is utilized for long distance stacking interactions rather than T-loop stability. The double-point mutant, on the other hand, contains a more favorable

Table 4. Melting temperature analysis of B.sub T-box riboswitch WT and mutants calculated from UV-melt experiments indicating highest thermostability to be the A30G mutant.

Constructs	T_m (C)
<i>B.subWT</i>	68
<i>B.subA30G</i>	77
<i>B.subA11U/C13A</i>	67
<i>B.subC13U</i>	70

closing U-A base pair that theoretically should increase the stability of the T-loop but a possible decrease in base stacking might be the cause of only the slight melting temperature increase. This data indicates the consistency of stability within the distal T-loop, where even a double point mutant only changes the T_m by 20C. Mutation of A30G in the apical T-loop significantly increases melting temperature by 110C, indicating the importance of the apical T-loop in thermostability. The fourth position, A in the WT apical T- box, is primarily involved in base stacking interactions but mutation to a G could potentially base pair with a cytosine allowing for the extension of the helical region (6). More likely, as seen in the predicted secondary structure, is the formation of the tri-loop with increase base pairing in providing more stability. We do not know if there is any intercalation with the distal T-loop but the increase in thermostability indicates that a stable structure is formed.

Initial thermodynamic parameters were calculated, as previously described, using absorbance profiles (57). Preliminary data do not necessarily match the melting profiles described above. B.sub T-box WT has a calculated free energy of -14.24 kcal/mol which is the most stable out of the constructs measured. Double-point mutant, A11U-C13A, has the shallowest melting profiles and CD absorbance shows, qualitatively, the weakest he-

lix formation but has a free energy of -13.40 kcal/mol (Figure 13, Table 4). This could be due to additional hydrogen bonding between the closing U-A base pair compared to the A-A pair seen in the WT structure, leading to more base stacking interactions within the dual T-loop. C13U mutant had a free energy of -12.86kcal/mol, much lower than anticipated. Formation of the C13U mutant could decrease free energy due to a decrease in base stacking interactions and loss of base pairing to G32 adjacent to the T-loop structure. Lastly, A30G has the lowest free energy at -10.74kcal/mol potentially leading to a more ordered structure, with an entropy contribution lower than all other constructs, due to a potential G-C base pair and formation of the tri-loop. This may explain the most pronounced signatures observed in the CD spectrum at 210nm that may indicate increased helical character (Figure 14C). Preliminary results provide insight into the thermodynamic parameters associated with tertiary structure formation. Further validation of the thermodynamic parameters of each RNA construct will be needed to assess how mutational changes influence the global RNA structure.

2.3.4. Electrophoretic mobility shift assays and isothermal titration calorimetry of B.-sub T-box riboswitch binding to tRNA^{Gly}

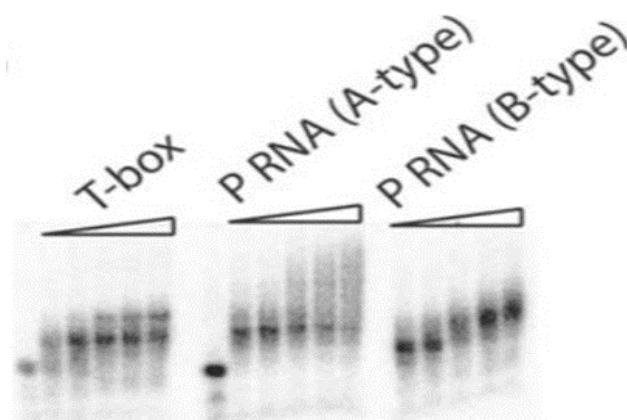


Figure 15. A.) RNA-tRNA complex formation through EMSA studies previously performed by the Reiter group on binding to tRNA (Reiter Proposal).

B. sub T-box riboswitch is known to process tRNA by cleaving the 5' end through a hydrolytic mechanism. tRNA^{Gly} binds bacterial mRNA via hydrophobic base stacking with the dual T-loop motif and its own TYC/D-loops (46). Examination of full length B. sub T-box binding to tRNA^{Gly} was found to have a dissociation constant of $15 \pm 3 \text{ nM}$ (46). Our study examines binding of tRNA^{Gly} exclusively via the minimal construct containing the dual T-loop motif of *Bacillus subtilis* glyQS T-box riboswitch.

Qualitative analysis of the binding interaction between truncated constructs and tRNA was performed via electrophoretic mobility shift assays (EMSA). Previous EMSA studies by on binding of *T. maritima* and the *B. subtilis* RNase P to tRNA^{Gly} indicated formation of tRNA^{Gly}-T-loop complex with increasing amounts of T-box being added to the mixture (Figure 15). Complex was shown as the top band since combination of T-loops and tRNA^{Gly} would create a slower moving band (Figure 15). Through careful experimentation, we were able to see slight complex formation of product even though it was predicted to be a weak binding partner with just T-loops available to form the complex (Appendix Figure 36). This EMSA binding assay gives an initial, qualitative perspective that a minimal dual T-loop RNA from various functional RNAs can interact with tRNA.

Isothermal titration calorimetry is a powerful and quantitative method to measure to the physical basis of molecular interactions. Following the methodology performed by Mondragon group, we titrated our truncated constructs into tRNAGly at millimolar concentrations. Previous studies determined the dissociation to be relatively small, nanomolar concentrations, for binding of the RNA motif to tRNAGly (46). A contrasting study found that Stem I T-loops could not bind tRNAGly without the additional anti-terminator region of mRNA or had very weak interactions (72). Through exhaustive optimization methods, we obtained initial formation of T-boxWT-tRNA complex with a binding enthalpy of 189.5 kcal/mol or a dissociation constant of $2.00 \pm 0.47 \mu\text{M}$ (Figure 16A). Previous studies showed a 50-fold reduction in tRNAGly binding with a mutated C13U in the distal T-loop (46, 69, 72). We created the C13U mutant to analyze this same interaction and determine binding parameters of a truncated, minimal version of the B.sub T-box RNA. Analysis showed a binding enthalpy of 327.3 kcal/mol with a dissociation constant of $1.53 \pm 0.14 \mu\text{M}$, due to the disruption of the G-C pair that directly

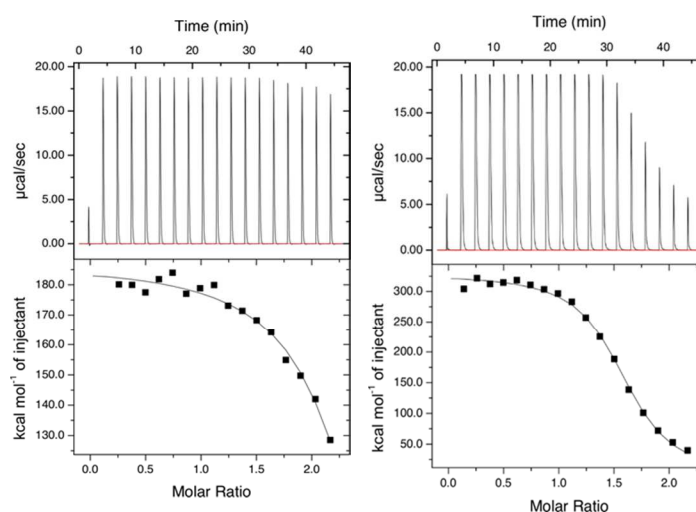


Figure 16. ITC binding curves for B.subWT and C13U constructs, respectively. Comparative study was performed which indicated an ~50 fold difference in dissociation constant (K_d) (36).

base stacks with a tRNA^{Gly} T \square C/D-loops (Figure 16B). This was an unexpected result as it was hypothesized that the C13U mutant would cause a decrease in binding. There is a slight preference for a G-U base stack (-10.6kcal/mol) over a G-C base stack (-9.3kcal/mol), the mutated U could adopt a non-planar orientation, caused by lack of base pairing to the G (28). Subsequent shifting of the U would lead to a decrease in base stacking ability and cause the overall decrease in binding enthalpy. Further studies are needed to examine this binding interaction. Point mutant construct, A30G, had no binding observed to tRNA with a dissociation constant of $149.9 \pm 332 \mu\text{M}$ (Appendix Figure 38). The decrease was likely due to disruption of the distal T-loop and a decrease in hydrogen bonding in the base triple formed between the distal and apical T-loops which normally form a Hoogsteen: Sugar: Watson-crick A:G:A triple (6, 8, 46, 69). This preliminary thermodynamic data needs further study to examine effects concentration of samples, as well as, ensuring a homogeneous fold in the correct conformation of tRNA, which tends to dimerize in non-physiological conditions (46).

2.4. Conclusion

RNA characterization has become increasingly important in understanding natural processing of genes and other functions within the cell. The primary goal of this study was to find suitable folding conditions for B.sub T-box riboswitch and to analyze the structural and thermodynamic properties of the dual T-loop regions within the truncated construct and point mutants. Investigation found suitable folding conditions leading to a compact, structured fold suitable for biochemical analyses. Previous circular dichroism analysis have been able to identify specific helical structures within nucleic acids but it is possible that a novel feature indicative of a T-loop motif in RNA appears within the 220-

235 nm region of the CD spectrum. Examination of other stable dual T-loop structures in other RNAs will need to be tested to confirm this finding. In addition, isothermal titration calorimetry indicates weak binding to tRNA^{Gly} via truncated dual T-loop motif of B.sub T-box riboswitch and a C13U mutant. This work is helpful for us to characterize a dual-T-loop RNA macromolecule in solution. The tRNA binding parameters of the dual T-loop may also be inferred by defining the hydrogen bonding network via NMR titration experiments.

CHAPTER 3

3.1. Introduction

3.1.1. Nuclear Magnetic Resonance Spectroscopy of RNA

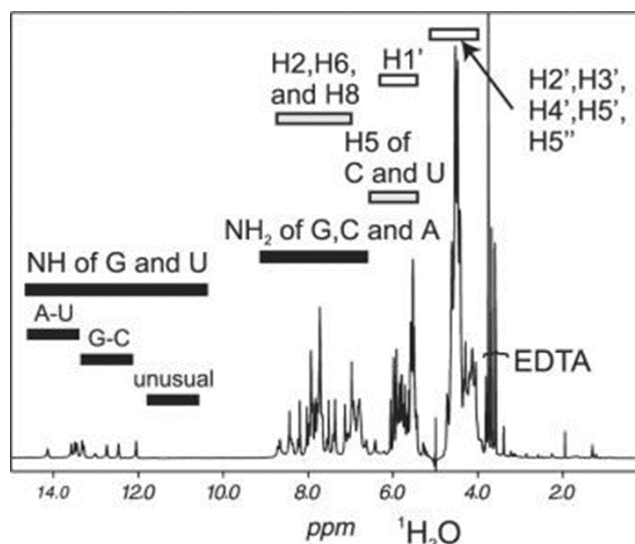


Figure 17. Chemical shift of RNA nucleotides (55). Regions 10-14 ppm indicated base pairing where noncanonical pairs are found between 10-12ppm. Aromatic protons are found in the 6-9ppm range with ribose protons containing the most overlap.

NMR is the only method to unambiguously detect and delineate complex hydrogen bonding networks of large biomolecules. Elucidation of RNA structure can provide key insight into their diverse functional roles within the cell. The human body encodes for many more RNA than proteins but RNA only account for approximately 1% of the total PDB deposits. This is due to the difficulties in sample preparation, cost of labelling experiments, and heterogeneity in the RNA tertiary fold. NOE-based assignment strategies require known assumptions about chemical shift patterns and are inherently less reliable due to chemical shift degeneracy (62). Due to these limitations, application of heteronuclear two- and three-dimensional resonance spectroscopy experiments to isotopically labelled RNA is vital. NMR spectroscopy of RNA presents

many challenges but is crucial to understand RNA's structural versatility, its inherent conformational dynamics, and its hydrogen bonding networks.

3.1.2. NMR Structure Determination strategies for *Bacillus subtilis* glyQS T-box riboswitch Dual T-loop Motifs

NMR spectroscopy is a powerful tool for the analysis of RNA molecules in

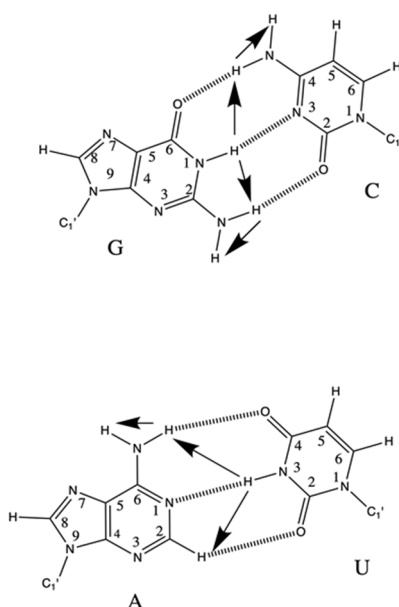


Figure 18. Representation of canonical base pairing in RNA. Arrows indicate NOE connectivities stemming from imino proton.

solution. Through the application of a magnetic field, nuclei with spin properties can be used to detect distinct atoms and determine their spatial location within a structure (33). NMR experiments for RNA can elucidate base pairing patterns, in both canonical and noncanonical regions, as well as conformational equilibria, delineate secondary structural motifs, local structure and dynamics, and define atomic interactions with ligands or other small molecules (54). Based on the targeted nuclei's spin properties, specific pulse sequences can be optimized to gain structural information based on the atom's coupling constants, dipole-dipole interactions, and overall chemical environment. In contrast to

small molecules, where structures can sometimes be solved using 1D or simple 2D experiments, RNA needs extensive NMR experimentation to elucidate structural features. This is due to the size of the molecule, as well as the reduction in chemical-shift dispersion with the biopolymer containing four different nucleotides (54). High resolution structures for RNA can be determined up to 50 nucleotides and intermediate resolution for about 100 nucleotides via NMR.

Structure determination for RNA can be achieved through detailed experimentation of unlabeled and labeled 2D and 3D experiments. Secondary structure can be determined through the use of 2D ^1H - ^1H H_2O and D_2O NOESY experiments, which provide information on base pairing and sequential patterns. The nuclear overhauser effect (NOE) uses spin magnetization transfer from one spin active nuclei to another that have close spatial proximity, within 5\AA , via cross-relaxation. The dipole-dipole coupling contributes to the spin-lattice relaxation, where excited spins relax to equilibrium by the release of thermal energy to the lattice, and provide structural information and correlate to peak intensity (33). Intensities of the cross-peaks are directly correlated with the distance between the interacting nuclei.

Imino-imino NOE cross-peaks occur sequentially and can be used to determine the base pairing pattern within RNA. Cross-peaks between exchangeable imino peaks and aromatic protons in the 2D H_2O NOESY are important for structure determination and can help find key starting points for the sequential walk spectrum, such as the Adenine H2 which has a strong correlation with the imino proton in a U-A base pair. Experiments in D_2O are used to determine the sequential walk of aromatic base protons to H1' of the ribose NOE patterns in the 5' to 3' direction in helical regions (Figure 16). Spectral

overlap between aromatic and ribose protons poses challenges for assignment of this region with the aromatic protons H2, H6, and H8 chemical shift all being between 6.5-8ppm (Figure 17). Ribose H1' and pyrimidine H5 atoms have similar chemical shifts ranging from 5.5- 6.5ppm (Figure 17). Overlay with a 1H-1H D2O TOCSY can identify the H5 and H6 of pyrimidines through the use of scalar couplings and help with sequential walk analysis. Severe overlap frequently occurs between aromatic and ribose protons, especially with RNA oligomers greater than 30 nucleotides (Figure 17).

Overlap of chemical shifts between uracil and guanosine cause ambiguity in the assignment of 2D 1H-1H H2O NOESY spectrum. The chemical shifts of uracil imino resonances are between 13-15 ppm where guanine iminos have a chemical shift of 12-13.5ppm (Figure 17). The decreased chemical shift dispersion can be overcome through the use isotopically ¹³C,¹⁵N labelled RNA in 2D and 3D HNN-COSY, HSQC, and HMQC experiments. The uracil and guanosine imino resonances in the nitrogen dimension are separated by ~10ppm which makes it easier to distinguish base pairing overlap. HNN-COSY experiment elucidates base pairing through the use of large scalar coupling between N-H—N type hydrogen bonding patterns. Through the use of ¹³C and ¹⁵N HSQC and HMQC experiments, confirmation of imino assignments can be made in the 2D 1H-1H H2O NOESY along with determination of aromatic protons for 2D 1H-1H D2O NOESY sequential walk region.

3.1.2.1. Challenges with structure determination of RNA via NMR

Although NMR is a very useful tool for structure determination in RNA, it has a few drawbacks which have limited RNA structural analysis. Chemical shift dispersion, between the four nucleotides, is significantly decreased when compared to proteins. Since

the dominating structure is the A-form helix in double-stranded RNA, many of nucleotides have the same chemical environment leading to spectral overlap, especially in the imino canonical base paired regions. Imino-imino NOE cross-peaks occur sequentially and therefore cannot always be differentiated between intra- and intercatenar contacts (54). Another factor leading to difficulties in RNA assignments is the size of the RNA. RNAs greater than 40 nucleotides in length lead to very challenging assignments due to anisotropic effects. In relaxation experiments, dipolar interactions are examined with respect to molecular tumbling but this also probes tumbling by chemical shift anisotropy (CSA) tensor. Fluctuations in the effective field stemming from molecular tumbling causes CSA and dipolar interactions to cancel each other out. As the size of the molecules increase, so does the spectral complexity. There is a need for higher spectrometer fields to increase resolution in the proton dimension of a spectrum, yet it is a substantial challenge to unambiguously assign a large (>40 nt) RNA at high magnetic field strength (5). Asymmetry within the RNA molecule leads to fast spin-spin relaxation times (short T₂) causing line broadening and poor resolution spectra. Other issues include water exchange with helical regions and heterogeneous conformation due to RNA degradation. To combat these limitations, the use of a variety of 1D, 2D, and 3D NMR experiments, both labeled and unlabeled, provide detail and elucidate the secondary structure conformation of *Bacillus subtilis* glyQS T-box riboswitch.

3.2. Material and Methods

3.2.1. Materials

Sodium Chloride (NaCl), Potassium chloride (KCl), Monopotassium phosphate (H₂KPO₄), Potassium dihydrogen phosphate (HK₂PO₄), Magnesium chloride (MgCl₂)

Tris-base, Sodium hydrogen phosphate (Na_2HPO_4), Glycerol, Diethylaminoethyl (DEAE)-Sphacel, Nucleoside triphosphates (NTPs), Dithiothreitol (DTT), RNasin Ribonuclease Inhibitor (Promega Corporation) Tris-HCl, G25-Sephadex, Spermidine, Bovine serum albumin (BSA), Boric acid, Ethylenediaminetetraacetic acid (EDTA), Ethanol 200 proof, NaOAc, Acrylamide (19:1), Ammonium Persulfate, and tetramethylethylenediamine (TEMED) were purchased from Sigma-Aldrich, VWR, or Fisher Scientific. All reagents were purchased commercially and were of the highest purity available. Pre-packed 5 mL Ni-NTA and 10 mL Econo-Pac Disposable Chromatography columns were purchased from Bio-Rad. Glass plates, combs, and spacers were purchased along with Fischer Biotech Electrophoresis System. All the solutions including buffer solutions were prepared using Nano pure water ($\sim 18.2 \text{ M}\Omega$) from Milli-Q® integral water purification system for ultrapure water and filtered using $0.2 \mu\text{m}$ filters.

3.2.2. *Instruments/Apparatus*

VWR scientific water bath incubator was used in transformation and transcription of DNA. Beckman coulter bench top and Avanti centrifuge were used for spinning cell lysate, concentrating, and pelleting RNA. FischerBiotech Electrophoresis System FB 600 was used for all gel electrophoresis. A Mineralight multiband UV lamp was used for visualizing RNA bands in gel electrophoresis. Labconco FreeZone 2.5L lyophilizer was used for drying all RNA. Thermo Scientific Nanodrop was used for measuring protein and RNA concentration. Collection of 1D and initial 2D NMR spectra was recorded on a Varian 600 MHz equipped with a cryogenic probe at Marquette University. 2D and 3D

NMR spectra were collected on a Bruker 600 and 800 MHz equipped with a cryogenic probe at NMRFAM in Madison, WI.

3.2.3. NMR Experiments

3.2.3.1. Sample Preparation

RNAs were made as described in Chapter 2 with isotopically U-A and G-C labeled ^{13}C , ^{15}N RNA. Isotopically labeled NTPs were kept in the dark to avoid degradation through light exposure. RNAs were folded at a concentration of 1 mM with 10 mM potassium phosphate buffer, pH 7.5, 50 mM potassium chloride, and 10 mM magnesium chloride with 5 % D_2O . Folding protocol was performed as previously described (64). For D_2O experiments, NMR sample and buffer were flash frozen in liquid nitrogen prior to overnight lyophilization, resuspended in 100 % D_2O and repeated lyophilization to remove all residual water, and resuspended in 100% D_2O NMR buffer at pH 7.0. The sample volume for all samples was 350 μL in a 5mm Shigemi tube.

3.2.3.2. 1D NMR temperature and Mg^{2+} dependence experiments

The dual t-loop region within B.sub glyQS T-box riboswitch was folded in 10mM KPO_4 and 50mM KCl using PCR folding protocol 1 (920 C for 3 minutes, 40C for 2 minutes, 500 C for 15minutes, and 370C for 30 minutes). After collection of 1D without Mg^{2+} , first addition of Mg^{2+} was added to the sample and mixed for 2 minutes each time. A range of Mg^{2+} was added starting from 0.1 μM with an incremental increase of 0.1 μM until 0.5 μM then titrated to concentrations of 1, 2, 5, and 10mM Mg^{2+} . NMR titration experiment was performed at 293 K with sine shift for imino region optimization. For temperature dependence study, 1D ^1H spectra were collected on B.subWT with 10mM Mg^{2+} added to the buffer on a Varian 600 MHz spectrometer

equipped with cryoprobe at Marquette University. Temperature range collected was from 283-298 K, increasing incrementally by 5 K. Spectra were processed and analyzed using the Topspin software program.

3.2.3.3. 2D NMR secondary structure elucidation

Spectra of the truncated B.sub glyQS T-box riboswitch RNA containing dual-T-loop regions were collected at 293 K on a Bruker 800 MHz spectrometer equipped with a cryogenic probe at NMRFAM in Madison, WI. Spectra was processed with NMRPipe and analyzed using Sparky (65).

3.2.3.3.1. Experimental Design

Exchangeable imino resonances from 1H-1H-NOESY experiments were recorded

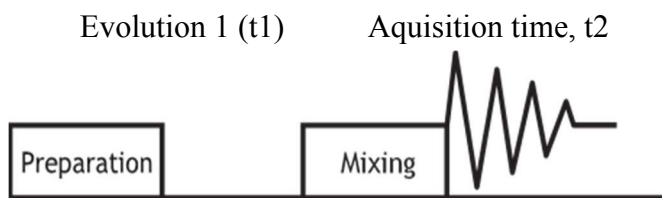


Figure 19. General pulse sequence for 1D NMR analysis.

in water with a 200 ms mixing time, with sweep widths of 16.6 and 10.0 kHz in F1 and F2, respectively. In a general 2D NMR experiment, magnetization is created through a radio frequency (RF) pulse corresponding the specific frequency needed to excite the atom of interest after which the magnetization evolves and provides information in the indirect dimension (y-axis) (Figure 19). Then the magnetization is mixed with magnetization of interest and acquired through direct observation (Figure 17). Nuclear overhauser effect (NOE) allows for the transfer of nuclear spin polarization from one spin active nuclei through space to another via cross-relaxation. A RF pulse is applied to a

specific nuclei of interest, which gets excited to different energy level. Dipole-dipole interactions through space with excited nuclei induce spin transitions of nearby magnetic nuclei. The relaxation of the non-irradiated nuclei follows a Boltzman distribution, resulting in chemical shift information and the intensity of an observed signal (49*). Pulse sequences can be optimized based on the nuclei on interest, given a nuclei's magnetogyric ratio and the abundance of ^1H , ^{13}C , and ^{15}N in the biomolecule. Sequential resonance assignments required collection of a 2D ^1H - ^1H -NOESY (D_2O) with a longer (200-300ms mixing time) to analyze the aromatic protons attached to the nucleobases and their corresponding interactions with ribose protons. In addition, ^1H - ^1H -TOCSY (D_2O) experiment provides 2- and 3- bond scalar couplings of an RNA with a 45 ms mixing time.

Quantitative HNN-COSY experiments were collected at 293 K as previously described (71). Through the use of INEPT (insensitive nuclei enhanced by polarization transfer) period, utilization of the $1J_{\text{HN}}$ coupling provides transfer of magnetization to donor ^{15}N . Transfer of magnetization through the $h2J_{\text{NN}}$ coupling ($^{15}\text{N}_{\text{acceptor}}$ - $^{15}\text{N}_{\text{donor}}$) via a COSY delay period followed by a t_1 period, which provides information on chemical shifts in the indirect dimension. Magnetization is transferred back to donor ^{15}N and transverse relaxation-optimized spectroscopy (TROSY) converts relaxing component (^1H)- $^{15}\text{N}_{\text{donor}}$ into ($^{15}\text{N}_{\text{donor}}$)- ^1H doublet detected in the final evolution period, t_2 (22).

^1H - ^{15}N -NOESY-HMQC spectra were collected on isotopically labeled GC and UA samples to distinguish individual bases pairings. Further identification of aromatic

and ribose regions were analyzed with collection of ^1H - ^{15}N - HMQC, ^1H - ^{13}C - HMQC, ^1H - ^{15}N NOESY- HSQC, and ^1H - ^{13}C -NOESY-HMQC experiments.

3.2.3.4. 3D NMR Experiments

A 3D ^1H - ^1H - ^{13}C NOESY HSQC spectrum was collected at 298 K on isotopically labeled GC and UA samples to distinguish aromatic NOE contacts and to correlate chemical shifts of ^{13}C to non-exchangeable protons. To distinguish severe spectral overlap of ribose protons, future 3D HCCH-COSY and HCCH TOCSY experiments can be obtained, utilizing scalar coupling between proton and isotopically labeled ^{13}C .

3.3. RESULTS

3.3.1. 1D NMR experiments in preparation for 2D structural analysis.

It is important to find a stable temperature for RNA when performing NMR spectroscopy in order to maintain native structure. Imino protons of nucleobases in RNA oligonucleotides are protected from chemical exchange with the bulk water as long as they are hydrogen-bonded (53). In the RNA double-helix, nucleobases can still undergo base-pair opening even with a stabilized structure. Temperature studies were performed to elucidate optimal temperatures for imino characterization within our B.sub T-box RNA construct at a range of 283-298K. Analysis of the 1D spectrum (Figure 17A) shows sharper peaks at higher temperatures, especially in the canonically base paired regions. At higher temperatures, the formation of peaks, around 11.2 and 11.8 ppm, indicate the formation of noncanonical base pairs in that region. Concurrently, peaks at 10.2 and 11 ppm disappear give evidence of a base pair that is in exchange with water molecules and is not observable within the timescale of the experiment (54). It appears that these types

of base pairs may exist in the dual T-loop RNA. Our NMR data shows an optimal temperature of 298 K for NMR data collection.

1D ^1H NMR titration experiments were performed on the B.subWT construct with increasing concentrations of Mg^{2+} . Divalent magnesium plays an important role in the folding of the RNA constructs. Stabilization of the phosphate backbone through interactions with the anionic oxygen allows for the formation of unique structures, like the T-loop (28). Mg^{2+} can form stabilizing interactions with N7 positions of purines but requires hydration (28). Figure 17B depicts the differences in the imino region at higher concentrations of Mg^{2+} . The first three titrations, at micromolar concentrations, show relatively no change in base pairing compared to the first 1D with no Mg^{2+} . At millimolar concentrations, a dramatic change in the 13 ppm region occurs as base pairing becomes more defined and is stabilized by the metal ion. A comparison between 2mM and 10mM Mg^{2+} shows the importance of Mg^{2+} to base pairing with three defined peaks in the 13.6-14 ppm range at 10mM Mg^{2+} indicating formation of UA base pairs. Mg^{2+} has been shown to be involved in the folding of the dual T-loop structure, with Mg^{2+} interactions near the distal T-loop (50). Increasing Mg^{2+} concentrations backup this finding with more imino resonances distinguishing themselves in the noncanonical region. For structural studies of RNA via NMR spectroscopy, the formation of a compact structure in solution is essential to determining the correct conformation.

3.3.2. 2D NMR experiments to elucidate secondary structure of Bacillus subtilis glyQS T-box riboswitch

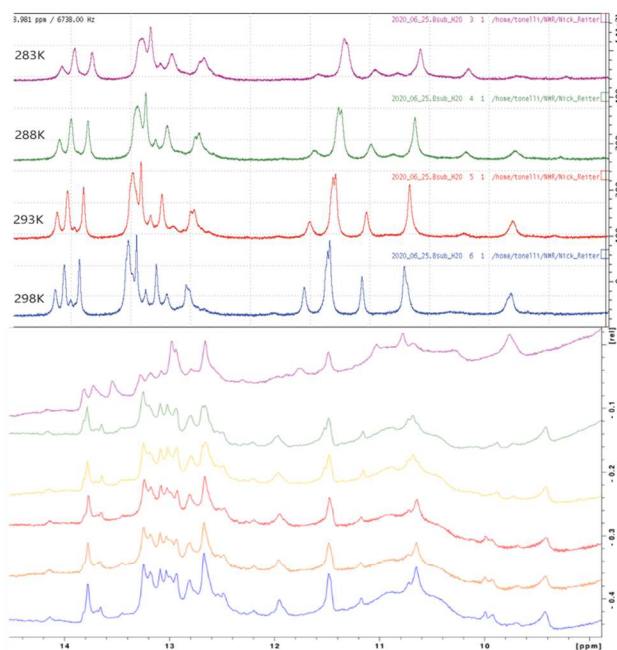


Figure 20. Top.) 1D temperature dependence studies of folded B.sub T-box riboswitch at a range of 283-298K. Temperature increases from top to bottom at 5 K increments. Bottom.) Mg²⁺ 1D NMR titration experiment on B.sub T-box riboswitch with increasing concentration of Mg²⁺ from top to bottom (0, 0.2, 0.4, 2, 5, and 10 mM, respectively).

Using 2D ¹H-¹H H₂O NOESY NMR experiments, the helical regions of unlabeled B.subWT could be assigned (Figure 21). NOESY spectra allows us to correlate protons to each other through space with the use of magnetization transfer from one spin active nuclei to another. Imino proton resonances of the guanines (G) and uracils (U) provide valuable information about base pairing but can only be seen when hydrogen bonded due to fast exchange with water molecules (54). The imino region is found between 10-15 ppm and by counting the imino proton resonances it is possible to determine the number of base pairs in a well folded RNA. Our assignment strategy was to

Figure 21. Top.) Secondary structure prediction from mfold web servers of B.sub T-box riboswitch construct. Bottom.) Imino base pairing region of 1H,15N HSQC. Labeled 1D corresponds to resonances directly underneath in COSY and NOESY experiments. Dashed lines indicate strongly base paired regions. Sequential assignments of the imino helical regions within the construct.

15ppm range, with U- A and G-C pairs having overlapping chemical shifts of 13-15ppm and 12-13.5ppm, respectively (54). We can define two distinct uracils, U40 and U18, through overlaid HNN- COSY with the 2D 1H-1H NOESY experiment, which confirmed assignment of strongly base paired nucleotide (Figure 21B and C). HNN-COSY of U-A labelled nucleotides, confirmed base pairing at 13.99 and 13.86ppm (Figure 18 B and C). Imino-imino NOE cross-peaks occur sequentially and cross strand, allowing for the assignment of helical regions. With unambiguous assignments of U40 and U18, we can follow the sequential NOE cross-signals to determine the helical patterns in both helix I and II (Figure 18A). As shown in figure 18B, we are able to define five strongly base paired U-A. Since there are only eight U's within the truncated construct, it is an obvious starting point for determining the sequential assignment of imino resonances.

In the crystal structure and cryo-EM structure (3Q1Q/6POM), seven of the eight U's are base paired, with two being in a noncanonical pairing, and one free U29 at the top of the apical loop. The HNN-COSY indicated five strongly base paired U-A's, which we can use to unambiguously assign based on NOE connectivities. We defined the first imino resonance as U40 at 13.99 ppm, which was used to assign helix I, and has two NOE cross peaks in the NOESY spectrum (Figure 21C). We determined the first connection to be a G-C pair at 13.45ppm, while the second connection is hypothesized to be a unique, non-canonically paired U7 bound to A through C4 carbonyl:amino N6 and H6:N1 (Figure 21B and C). This would be a very weak noncanonical base pair explaining the absence of a base pair in HNN-COSY experiment, but the imino resonance that appears in the non- canonical region would be bound to a Mg²⁺ metal shown to be involved in folding at that region (50). The noncanonical U7 has no other NOE cross

peaks due to loss of base pairing as the structure opens up to form the distal T-loop, but NOE connectivities can be followed through the rest of helical region from G41 to G2. The imino resonance for G1 could not be found, most likely caused by water exchange.

Assignment of helix II has presented more complications due to high resonance overlap. Assignment of U18 at 13.86ppm was distinguished through the connection to both a U and G but more specifically, the imino resonances in the noncanonical region stemming from U34 which also has imino resonance NOE cross peaks connected to a U and G. U18 is nearby two very stable G-C base pairs indicated in the HNN-COSY, G37 and G38 with chemical shifts of 12.84 and 12.68ppm, respectively (Figure 21B and C). U19 has overlapping chemical shifts with a G and U, where the U is believed to stem from heterogeneous conformation, and is distinguished from G through 3D-imino NOESY-HSQC connection to U18 (Appendix Figure 41). U34 is near the top of helix II and is predicted to be more dynamic since it is near the unstructured region before the T-loop. In the predicted entropic secondary structure (Table 2), base pairs near the top of helix II have a higher propensity for disorder than the bottom of helix II, including U34, due to consecutive, weaker U-A base pairs. We assigned U34 at 13.17ppm because of the NOE cross peaks to G21 in the more disordered, noncanonical region (Figure 18C) and connection was confirmed in the 3D-imino NOESY-HSQC experiment (Appendix Figure 41). A helix mutational study was performed to try to elucidate the overlapped regions from 13.0-13.4 ppm. The mutant contained two base pair mutations; the first being U7G and A39C and the other was U18C and A35G with the purpose of combating spectral overlap between three adjacent U's. In the spectrum, changes to both U40 and U18 connectivities but especially in the second helical region directly connected to U18

(Appendix Figure 39). Helix II shows regions of high flexibility with noncanonical base pairs not seen in the B.subWT construct. Mutational analysis along with HNN-COSY helped with initial assignments but confirmation of those assignments was needed with further experimentation.

To combat chemical shift overlap of exchangeable protons, 2D ^1H - ^{15}N HMQC and 3D ^1H - ^1H - ^{15}N imino NOESY-HSQC experiments were obtained. These experiments help to elucidate the secondary structure of the B.sub T-box construct and complement the 2D H_2O NOESY. HSQC and HMQC NMR experiments utilize the scalar coupling, similar to the HNN COSY, between ^{15}N donor and the attached proton

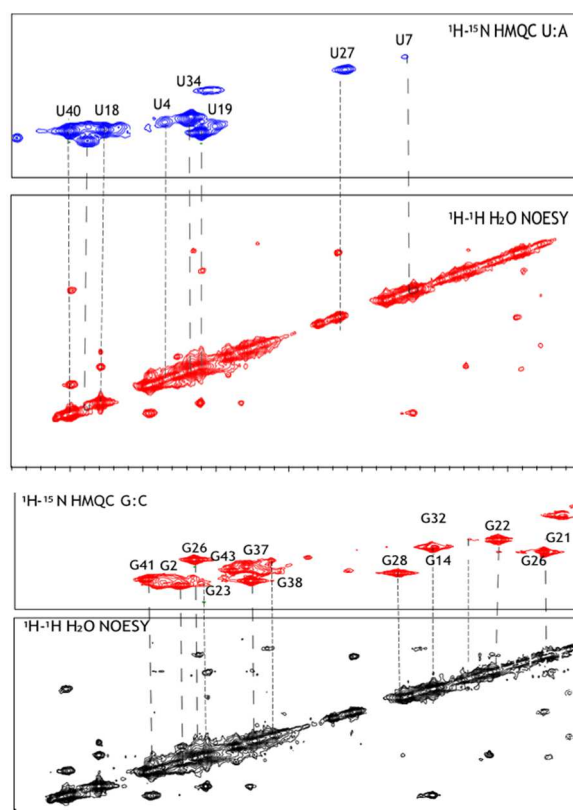


Figure 22. Top.) U-A and Bottom.) G-C labeled ^1H - ^{15}N HMQC indicating imino resonances in B.sub WT T-box construct. Both contain 2D NOESY experiment illustrating the relationship between the two experiments

(2JHN). These experiments are very useful because they provide another way to distinguish U-A and G-C pairs due to the separation in the nitrogen dimension by ~ 10 ppm (54, 65). In our 2D ^1H - ^1H NOESY, it was difficult to identify imino resonances in the 13-13.5 ppm range due to spectral overlap. The U-A imino HMQC spectrum shows very similar resonances to the HNN-COSY and contains all five base paired peaks but includes resonances in the 12-13 ppm range that are not seen in the COSY (Figure 22B).

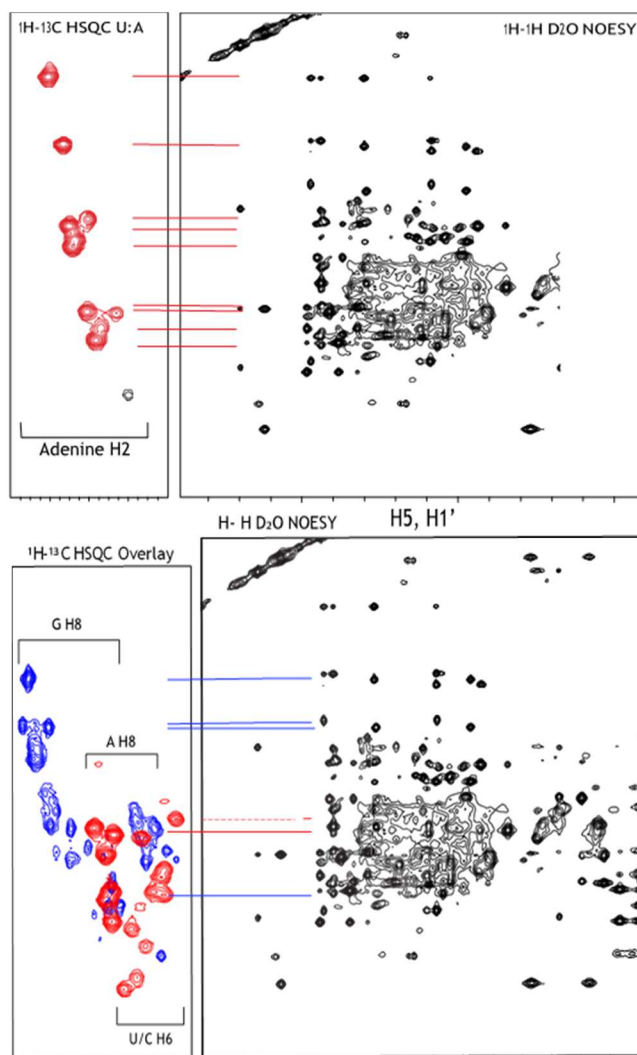


Figure 23. Top.) ^1H - ^{13}C HSQC of U-A labeled construct displaying the AH2 region 150-155 ppm aligned with D₂O NOESY sequential walk region to identify starting point. Bottom.) Overlaid G-C (blue) and U-A (red) labeled sample presenting aromatic protons of nucleobases. This indicates high degree of spectral overlap among aromatic protons of RNA.

In the HMQC U-A spectrum, there are two peaks that are unlabeled, both of which do not have any NOE cross peaks in the H₂O NOESY spectrum. Interestingly, there is only one unassigned U7 within the helical region, where U7 is predicted to be metal bound or forming a noncanonical C- H:Watson-Crick U-A pair. Evidence of this pairing could be seen through a weak peak at 11.72 ppm but more confirmation would be needed to confidently assign peak (Figure 22B). Examination into the 3D 1H-1H-15N imino NOESY-HSQC indicated no NOE cross peaks when examining the potential U at 11.72ppm or 158.2 in the 15N dimension (Figure 19B). The G-C HMQC identified all of the guanines within the helical region and matched resonances of the G-C HNN-COSY. Some peaks are not labeled within the G-C HMQC due to conformational heterogeneity not identified in the earlier HNN-COSY (Figure 22A). These resonances may be due to some RNA degradation that may be occurring within the NMR sample. Of note, several guanine H1 resonances extend into the noncanonical region but contain decreased NOE contacts that are likely due to fast exchange with solvent. Although the region before the T-loop is unstructured, and therefore prone to water exchange, imino connections in the noncanonical region are seen from G21 to G22 (Figure 21 C and Figure 22 A). G22 is positioned and predicted to be stabilized by the phosphate backbone oxygen of the G32, as well as, a stacking interaction with G23. The G23 H1 imino resonance is stabilized through hydrogen bond interactions with the phosphate backbone of A31. The G23 H1 has a resonance assignment of 13.09 ppm in the canonical base pairing region but is also thought to be close to a metal site according to the crystal structure. Helical regions with a high degree of confidence but unambiguous assignments cannot be made at this time. Assignment in the T-loop region have been a challenge, which may suggest that the dual

T-loop exhibits a more flexible and dynamic conformation in its unbound state. In contrast, the helical region is well resolved with the combination of the NOESY, HNN COSY, HMQC, and 3D 1H-1H-15N imino NOESY- HSQC experiments. Nonetheless, sequential walk assignment are still needed within parts of the lower and upper helical stem regions.

3.3.3. 2D NMR D2O experiments to correlate through-space interactions between the aromatic to ribose sugar protons

The interresidual sequential walk assignment of RNA helical regions is performed in D2O in order to minimize the H2O signal (~ 4.7 ppm) that significantly obscures the chemical shift resonances of the ribose region. This allows for the all aromatic protons to show NOE cross peaks to their own H1' sugar protons and to the adjacent nucleotide in the 5' direction (54). The distance between protons visualized can be extended up to 6-7 Å for this experiment due to spin diffusion through the H2' sugar proton (54). Difficulties for this experiment lie in finding a suitable starting point for the sequential walk. Severe spectral overlap occurs in this experiment, shown in Figure 20, but can be overcome with the 2D 1H-13C HSQC experiments, which use scalar couplings between aromatic protons and their respective carbons (2JHC). The C2H2 of adenine has a distinct chemical shift, 150-155 ppm, from other aromatic proton and can be used to delineate all H2s within the sequential walk region (Figure 22A). This serves as one suitable starting point to confirm and validate the sequential NOE pattern. Once AH2s are labeled, a spectral overlay of the D2O NOESY with a 1H-1H TOCSY can be used to identify the H5-H6 of pyrimidines with another scalar coupling experiment (Figure 21). Simple experiments like these can provide much needed clarity to very complicated spectra and

helped tremendously in our sequential walk assignment of the B.subWT T-box riboswitch construct.

Assignment of the sequential walk began with finding a suitable AH2 from the HSQC and the imino NOESY experiment. Our construct contains 12 total adenines with 10 involved in pairing and all 10 of the base paired adenines were found in the HSQC. Comparison with the 1H-1H NOESY defined base pairs with adenines H2 labeled in the aromatic region and were used for the starting point of the sequential walk within the A-form helical regions of helix I and II (Figure 23). A20H2 at 6.782 ppm contained well resolved NOE cross peaks to other protons and was therefore used as one starting point. A20H2 had a confirmed connection to the cross strand A35H2 at 6.393ppm but no clearly defined connection to the U19 directly below in the 5' direction. From A35H2, the entire sequential walk on the 5'-3' side of helix II can be followed, even showing a connection to helix I A39H8 at 7.676 ppm. Confirmation of the A39H8 peak assignment was validated in the 1H-13C HSQC with a resonance frequency of 139.7 ppm which is congruent with the average chemical shift at 139ppm (Figure 22B). Helix I was assigned with the starting point of following A39H8 down to its own H2 at 8.002 ppm. From that point, we were able to follow the 5'-3' side ranging from nucleotides A39-C44 with pretty high confidence. There were a few points of contention when assigning the sequential walk spectrum. Although, the TOCSY helped identify unambiguous pyrimidines, the assignments were difficult due to lack of or poorly resolved NOE cross peaks (Figure 23). In 3'-5' regions of both helix I and II, there is consecutive stacking of

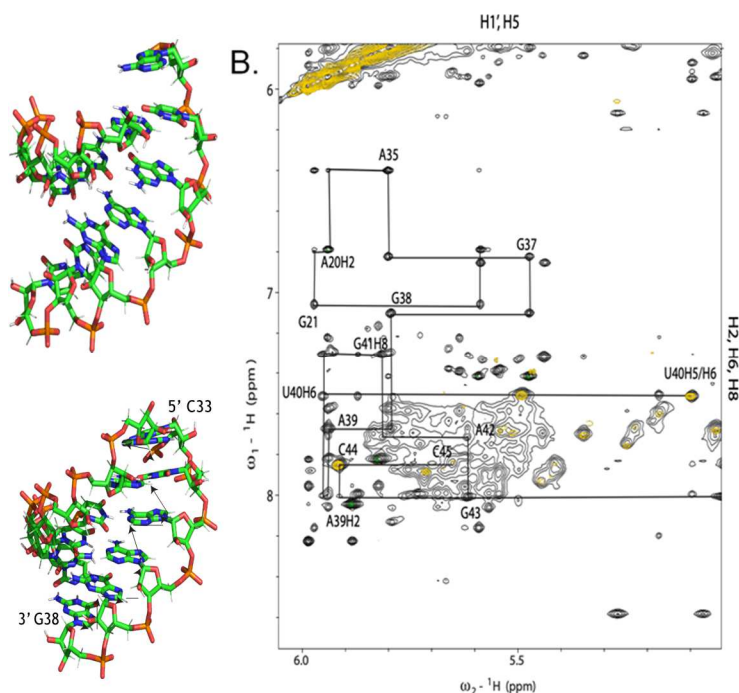


Figure 24. A.) Pymol representation of the two helical regions, helix I and helix II, within the B.sub T-box riboswitch. B.) 2D ${}^1\text{H}$ - ${}^1\text{H}$ NOESY (black) spectrum overlaid with a ${}^1\text{H}$ - ${}^1\text{H}$ TOCSY (yellow) recorded in D_2O at 800MHz with a 300ms mixing time.

pyrimidines, which are approximately 0.5-3.5 kcal/mol less stable in energy than pyrimidine-purine or purine-purine stack (28). This is best illustrated near the 3'-end of the construct, where the connection between G43, C44, and C45 only is found through a H8 to H5/H6 to H5/H6 NOE cross peaks, respectively, without any H1' interactions (Figure 24B). The sheared formation of the base pairs would allow for the connection to

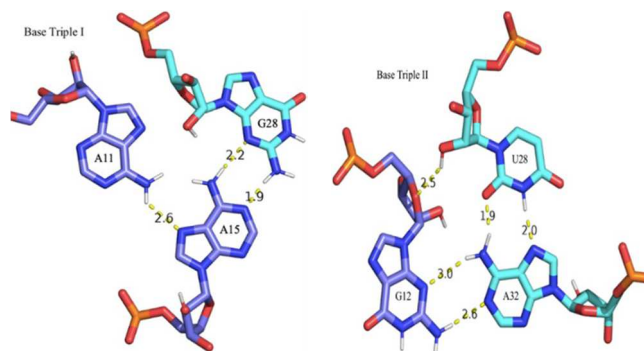


Figure 25. Predicted hydrogen bonding patterns of the base triples within the interdigitated dual T-loops (PDB 6POM).

the H5/H6 while increasing the distance of the H1' on the ribose diminishing any cross peaks that would normally be seen. Assignment of pyrimidines were confirmed through a ^1H - ^{13}C HSQC spectra of the ribose, which used scalar coupling of C1'-H1' and C5-H5 with distinguished resonance frequencies in the carbon dimension. G-C and U-A base pairs have similar chemical shifts for C1' (87-95ppm) but differ in the C5 range at 96-99ppm and 101-106ppm, respectively (Appendix Figure 42 A and B). Another issue with these assignments is the difficulty in confirmation through ^1H - ^{13}C HSQC spectra. Figure 20B illustrates the spectral overlap that occurs between GC8 (133-139ppm), AC8 (138-140ppm), and pyrimidine C6s (138-142ppm). This spectral overlap between aromatic protons was difficult to overcome but with extensive homo- and heteronuclear NMR experimentation, we were able to determine the sequential walk and confirm the helical secondary structural elements.

3.3.4. Noncanonical base pairs within the T-loop region of B.sub T-box riboswitch

Proton resonances in the noncanonical region of an RNA can be more difficult to assign than the canonical A-form regions of an RNA. These noncanonical base pairs often exhibit weaker hydrogen bonding interactions and contain imino protons that undergo fast exchange with the water solvent. In the B.sub T-box riboswitch construct, extensive base stacking interactions have been observed in crystal and cryo-EM structures but the hydrogen bonding network of the dual T-loop region has yet to be elucidated. The formation of two base triples occurs through intercalation of position 2 of one T-loop into positions 1 and 5 of the other T-loop creating AGA and UGA triples (Figure 25). An identifiable peak within the noncanonical region of the ^1H - ^1H H₂O NOESY is the Hoogsteen U27-A32 base pair at 12.17ppm and confirmed in the ^1H - ^{15}N UA labeled

HMQC (Figure 21B and 22A). U27 imino has a strong NOE cross peak to the preceding G26 at a distance of $\sim 3.3\text{\AA}$, which is base paired to G14 O6 from the distal T-loop (Figure 21C). A weak NOE cross peak is seen between the G26 imino resonances and G32 imino at a distance of $\sim 5.2\text{\AA}$ (Figure 21C). There are only two more imino resonances within the dual T-loop structure, nucleotides G12 in the distal T-loop and U29 in the apical T-loop (Figure 8C). Neither of these nucleotides form base pairs but G12's H1 imino resonance is hypothesized to be stabilized by the 2' OH functional group of the A15 phosphate back-bone in the crystal structure. G12 could also be stabilized by the proposed Mg^{2+} binding site near the distal T-loop (50). This would cause an unusual chemical shift that would most likely be seen in the noncanonical region as well. Spectral overlap is still the major obstacle in the noncanonical region and will require additional experimentation.

3.4. Conclusion

NMR allows for the solution structure determination of relatively large RNAs and a major goal of the current study was to elucidate secondary structure of a dual T-loop containing B.subWT T-box RNA. Helical base pairing regions were determined through the unambiguous assignment method of HNN-COSY paired with 2D 1H-1H NOESY and 2D 1H-15N HMQC experiments (Figure 21 B and C and Figure 22). Base pairing assignments allowed for assignment of many non-exchangeable protons within the 2D NOESY experiment as well as additional 3D NMR experiments. The previously determined crystal structures help us to understand the tertiary fold and local regions that give rise to the dual T-loop structure but NMR provide solution information on the hydrogen bonding network and flexibility of the RNA (8, 69, 70, 72). Assignments near

the more dynamic regions of the helix near the T-loop need further examination to ensure proper assignments were made. Although, base pairing interactions are shown to occur in the helical regions, the spectral overlap of NOE cross peaks in the sequential walk region make assignments especially challenging for the 3'-5' ends of the RNA. Additional experiments are needed to confirm the helical regions, as well as, define the hydrogen bonding interactions within the dual T-loop region.

CHAPTER 4

FUTURE WORK

NMR experiments for complete assignment and tertiary structure determination of B.subWT T-box riboswitch

The assignment of the T-loop region and tertiary structure for this construct is very difficult due to the complex spectra associated with a 45mer RNA construct. This

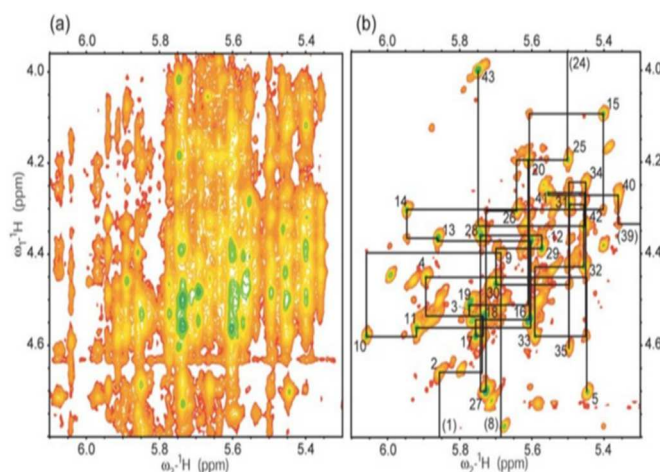


Figure 26. 2D NOESY spectrum of inter- and intraribose walk displaying the simplifying power of selective deuteration. Both regions are the same with the right spectrum being selectively deuterated (3).

leads to complications with conformational heterogeneity, anisotropic tumbling, and spectral overlap. In order to overcome these challenges additional 3D HCCH-COSY and 3D HCCH-TOCSY experiments will be required. This will help to clarify the ribose proton connectivities of the RNA. Variations of these experiments can be performed including HCCH- TOCSY/¹H-¹³C HSQC in D₂O to obtain H2-H8 correlations for confirmation of structural analysis. The conformation of the phosphodiester backbone and the ribose sugar pucker determine the overall structure of the RNA molecule. Experiments to help elucidate the conformation and sugar pucker are, therefore,

important in structure determination. Another method to overcome the overlap in the ribose region is to selectively deuterate the pentose ring. Extreme overlap within the sugar pucker region can be seen in Figure 22, and without the use of selectively labeled resonances it would be nearly impossible to assign. The drawback to this method is loss of backbone proton assignments of the deuterated protons, although not many NOE cross peaks stem from the H4', H5', and H5'' protons so not much structural information is lost. The more significant drawback is the expense of the selective deuteration (about \$1500-2000 per 2mg of NTP) in an experiment that requires millimolar concentrations. With this data determination of the solution structure could be elucidated and would clearly define the hydrogen bonding network of the dual T-loop structure.

RNA is a dynamic molecule that form a variety of structures on different timescales (49). The dynamics of the dual T-loop motif is important to analyze to gain a better understanding of the conformational flexibility within the structure. There are both 1D and 2D NMR experiments that can be applied to analyze dynamics in RNA. Analysis of line shape within 1D experiments can provide a qualitative measure of dynamics based on line shape and intensity that correlates to timescale at three different speeds: slow, intermediate, and fast. In our future experiments, the use of the intermediate ^{13}C relaxation dispersion experiments allows for the analysis of dynamics on the Larmor timescales, which correspond to MHz frequencies, and influences bond lengths and intermolecular distance causing changes in dipolar coupling and CSA. This will allow us to analyze the short term conformations at the micro- to millisecond timescale and provide chemical shifts of the excited state (49). Calculations can be made to determine

quantitative changes within our structure and provide insight into the dynamics motions of nucleotides within the T-loop region.

Biochemical characterization of dual T-loop structure via circular dichroism

Structure analysis of the dual T-loop structure has implicated the possibility of a novel method for RNA structure determination through circular dichroism methods. We revealed data that showed unique structural differences between mutant constructs of the B.sub T-box dual T-loop structure. The 220-235 nm region provides a specific signature of interest and may be indicative of T-loop formation. Slight differences can be seen when comparing double stranded DNA CD profile to our B.subWt construct. Both contain the highly positive peak around 260nm but DNA has a much more negative peak in that region (34). Also, hairpin formation is found in the CD profile near 275nm indicated by a very slight bump which can be seen in the B.subWT construct indicating the possible signature of hairpin-like, T-loop. Introducing more point mutants in the apical T-loop, which had the greater impact on UV-melting temperature studies, could distinguish different thermodynamic contributions of the locations within that T-loop, as well as, structural information. Future experiments will focus on salt dependence studies for T-loop formation and engineering point mutants constructs for further analysis of the T-loop region. Our study looks to define the solution structure of the minimal dual T-loop containing B.sub T-box riboswitch that enables binding to tRNA and elucidate the thermodynamic properties associated with it.

BIBLIOGRAPHY

1. Altman, S., et. al. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*. **1983**, 35, 849-857.
2. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **2003**, 72, 291-336.
3. Butcher, S.E., et. al. RNA Helical Packing in Solution: NMR Structure of a 30 kDa GAAA Tetraloop-Receptor Complex. *J. Mol. Biol.* **2005**, 351, 371-382.
4. Butcher, S.E. and Pyle, A.M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of Chemical Research*. **2011**, 44, 1302-1311.
5. Case, D.A. The use of chemical shift and their anisotropies in biomolecular structure determination. *Current Opinion in Structural Biology*. **1998**, 8, 624-630.
6. Chan, C.W., Chetnani, B., Mondragón, A. Structure and function of the T-loop Structural motif in noncoding RNAs. *Wire RNAs*. **2013**, 4, 507-522
7. Chen, J.L., Pace, N.R. Identification of the universally conserved core of the ribonuclease P RNA. *RNA*. **1997**, 3(6), 557-560.
8. Chetnani, B. and Mondragon, A. Molecular envelope and atomic model of an anti-terminated glyQS T-box regulator in complex with tRNA^{Gly}. *Nucleic Acids Research*. **2017**, 45, 8079-8090.
9. Craft, J., et. al. The RNA Processing Enzyme RNase MRP is Identical to the Th RNP and Related to RNase P. *Science*. **1989**, 245, 137
10. Donghi D., Sigel R.K.O. Metal Ion–RNA Interactions Studied via Multinuclear NMR. In: Hartig J. (eds) Ribozymes. *Methods in Molecular Biology (Methods and Protocols)*. **2012**, 848, 1-29.
11. Doudna, J.A., et. al. Tertiary motifs in RNA structure and folding. *Angew. Chem.Int. Ed.* **1999**, 38, 2326-2443.
12. Eddy, S. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet.* **2001**, 2, 919–929.
13. Egelman, E.H. The current revolution of Cryo-EM. *Biophysical Journal*. **2016**, 110,

1008-1012.

14. Egli, M. DNA and RNA structure. *Nucleic Acids in Chemistry and Biology*. **2006**.
15. Engelke, D. and Fierke, C.A. The evolution of RNase P. *RNA*. **2015**, 21, 517-518.
16. Erat, M.C., Sigel, R.K.O. Determination of the intrinsic affinities of multiple site specific Mg^{2+} ions coordinated to domain 6 of a group II intron ribozyme. *Inorg Chem*. **2007**, 46, 11224-11234.
17. Fiala, R.; Sklenář, V. NMR of Nucleic Acids; **2002**.
18. Greenfield, N.J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protocols*. **2006**, 1, 2527-2535.
19. Griggs, J.C., et al. tRNA recognition by the T box RNA. *National Academy of Sciences*. **2013**, 110(18), 7240-7245
20. Griggs, J.C., et. al. T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *PNAS*. **2013**, 110, 7240–7245.
21. Grundy FJ, Winkler WC, Henkin TM. tRNA-mediated transcription antitermination in vitro: codon- anticodon pairing independent of the ribosome. *PNAS*. **2002**, 99, 11121-11126.
22. Grzesiek, S. Direct detection of N-H N hydrogen bonds in biomolecules by NMR spectroscopy. *Nature Protocols*. **2008**, 3, 242-248.
23. Guerrier-Takada C, Altman S. Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science*. **1984**, 223(4633), 285-286.
24. Harris, M.E. and Jankowsky, E. Specificity and non-specificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol*. **2015**, 16, 533-544.
25. Hartmann, E. and Hartmann, R.K. The enigma of ribonuclease P evolution. *Trends Genet*. **2003**, 9, 561-569.
26. Hermann, T. and Patel, D.J. RNA bulges as architectural and recognition motifs. *Structure*. **2000**, 8, 47–54.
27. Herschlag, D., et. al. Quantitative test of a reconstitution model for RNA folding thermodynamics and kinetics. *PNAS*. **2017**, 7688-7696.

28. Hobza, P., et. al. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Nucleic Acid Sci.* **2002**, 61, 3-31.
29. Hofacker, I.L. and Stadler, P.F. RNA Secondary Structures. *In Reviews in Cell Biology and Molecular Medicine.* **2006**.
30. Jenner, L. et. al. Structural rearrangements of the ribosome at the tRNA proofreading step. *Nature Structural and Molecular Biology.* **2010**, 17, 1072-1079.
31. Kazantsev, A.V. and Pace, N.R. Bacterial RNase P: a new view of an ancient enzyme. *Nature Reviews.* **2006**, 4, 729-740.
32. Krasilnikov, A.S. and Mondragón, A. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA.* **2003**, 9, 640-643.
33. Kriz, G.S., et. al. Introduction to Spectroscopy. *Saunders College Publishing.* **1979**.
34. Kypr, J., et. al. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids research.* **2009**, 37, 1713-1725.
35. Kypr, J., et. al. Circular dichroism spectroscopy of nucleic acids. *Nucleic Acids Research.* **2011**, 1, 573-584.
36. Lan, P., et al. Structural insight into precursor tRNA processing by yeast ribonuclease P. *Science.* **2018**, 362(6415).
37. Laing, C. and Schlick, T. Analysis of four-way junctions in RNA structures. *J. Mol. Biol.* **2009**, 390, 547-59.
38. Lai, D. et. al. On the importance of cotranscriptional RNA structure formation. *RNA.* **2013**, 19, 1461-1473.
39. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics.* **2015**, 31(8), 1325-7.
40. Leontis, N.B. and Westhof, E. Conserved geometrical base-pairing patterns in RNA. *Quarterly Reviews of Biophysics.* **1998**, 31, 399-455.
41. Li S, Zhang J., et. al. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat Struct Mol Biol.* **2019**, 26(12), 1094-1105.
42. Mathews, D.H. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Research.* **2006**, 34, 4912-4924.

43. Milligan, J.F., et al. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates, *Nucleic Acids Research*. **1987**, 15(21), 8783–8798.
44. Mondragon, A., et. al. Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tR-NAs. *RNA*. **2019**, 1-34.
45. Nagaswamy, U. and Fox, G.E. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA*. **2002**, 8, 1112–1119.
46. Nikonowicz, E.P., et. al. Capture and release of tRNA by the T-loop receptor in the function of the T-box riboswitch. *Biochemistry*. **2017**, 56, 3549-3558.
47. Nikonowicz, E.P., et. al. Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA binding proteins. *Wiley Interdiscip. Rev. RNA*. **2014**, 5, 1-28.
48. Pace, N. and Chen, J.L. Identification of the universally conserved core of ribonuclease P RNA. *RNA Society*. **1997**, 3, 557-560.
49. Petzold, K., et. al. RNA dynamics by NMR spectroscopy. *ChemBioChem*. **2019**, 20, 2685-2710.
50. Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., Mondragón, A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*. **2010**, 468 (7325), 784-791.
51. Reiter, N.J. RNA lab methods: a laboratory manual. **2010**, entry
52. Roland, K.O., et. al. Metal ion-RNA interactions studied via multinuclear NMR. *Methods in Molecular Biology*. **2012**, 848, 253-274.
53. Sampath, R., et. al. RNAMotif, and RNA secondary structure definition and search algorithm. *Nucleic acids research*. **2001**, 29, 4724-4735.
54. Schwalbe, H., et. al. NMR spectroscopy of RNA. *ChemBioChem*. **2003**, 4, 936-962.
55. Schwalbe, H., et.al. Direct observation of the temperature-induced melting process of the *Salmonella* fourU RNA thermometer at base-pair resolution. *Nucleic Acids Research*. **2010**, 38, 3834–3847.
56. Scott, L.G. and Hennig, M. RNA Structure Determination by NMR. *Bioinformatics: Methods in Molecular Biology*. **2008**.

57. Selmer, M. Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science*. **2006**, 313, 1935-1942.
58. Tincoco Jr., I. and Puglisi, J.D. Absorbance melting curves of RNA. *Methods in Enzymology*. **1989**, 180, 304-325.
59. Torres-Larios, A. Swinger, K.K., Pan, T., and Mondragón, A. Crystal structure of the RNA component of bacterial ribonuclease P. *Nature*. **2005**, 437, 584-587.
60. Torres-Larios, A. Swinger, K.K., Pan, T., and Mondragón, A. Structure of ribonuclease P—a universal ribozyme. *Curr Opin Struct Biol*. **2006**, 16, 327-335.
61. Tu, Z., et al. An improved system for competent cell preparation and high efficiency plasmid transformation using different Escherichia coli strains. *Electronic Journal of Biotechnology*. **2005**, 8, 717-3458.
62. Turner, D.H., et.al. Improved free-energy parameters for prediction of RNA duplex stability. *Pro. Natl. Acad. Sci*. **1986**, 83, 9373-937
63. Varani, G. et. al. NMR investigation of RNA structure. *Prog. Nucl. magnetic Reson. Spectrosc*. **1996**, 29, 51-127.
64. Vitreschak, A.G., et. al. Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA*. **2008**, 14, 717-735.
65. Wan, F., et. al. Cryo-electron microscopy structure of an archaeal ribonuclease P holoenzyme. *Nature Communications*. **2019**; 10(1): 2617.
66. Wöhnert, J., et al. Direct identification of NH...N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Research*. **1999**, 27, 153104–3110.
67. Wu, J., et. al. Cryo-EM structure of the Human Ribonuclease P Holoenzyme. *Cell*. **2018**, 175, 393-404.
68. Yao, R.W. et. al. Cellular functions of long noncoding RNAs. *Nature Cell Biology*. **2019**, 21, 542-551.
69. Yao, R. W.; Wang, Y.; Chen, L. L. Cellular Functions of Long Noncoding RNAs. *Nature Cell Biology*. **2019**.
70. Zhang, J., Chetnani, B., et. al. Specific structural elements of the T-box riboswitch drive the two-step binding of the tRNA ligand. *eLife*. **2018**, 7, 1-22.

71. Zhang, J. and Ferrè-D'Amare. The tRNA Elbow in structure, recognition, and evolution. *Life*. **2016**, 6, 1-11.
72. Zhang, J. and Ferrè-D'Amare. Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*. **2013**, 500, 363-366.
73. Zhang, J., et. al. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat. Struc. Mol. Bio.* **2019**, 26, 1094-1105.
74. Zhunag, Z., Jaeger, L., and Shea, J.E. Probing the structural hierarchy and energy landscape of an RNA T-loop hairpin. *Nuc. Acids Research*. **2007**, 35, 6995-7002.
75. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, 31, 3406-3415.

APPENDIX

Chapter 2 Appendix

Materials

Luria broth (LB broth), LB Agar, Ampicillin, Phenylmethylsulfonyl fluoride (PMSF), Imidazole, and Isopropyl B-D-1-thiogalactopyranoside (IPTG) were purchased from Sigma-Aldrich, VWR, or Fisher Scientific. All reagents were purchased commercially and were of the highest purity available. Bacterial cells, cell lysates, and proteins were handled on ice, in 4°C cold room, or in 4°C temperature-controlled refrigerator. BL21 gold competent cells and P266L T7 polymerase plasmid were donated by Ainur Abzhanova and made following the Tu protocol (13).

Instruments and Apparatus

Thermo scientific precision incubator was used for incubating LB agar plates. For growing and inducing cell cultures, New Brunswick Scientific Innova 4330 refrigerated incubator shaker was used. Lysing of cells was performed using Branson Sonifier sonicator. Beckman coulter bench top and Avanti centrifuge were used for spinning cell lysate, concentrating, and pelleting RNA. Thermo Scientific Nanodrop was used for measuring protein and RNA concentration.

Preparation of T7 RNA polymerase protein

Preparation of LB-agar plates and LB growing media

Mixed 12 g of premixed LB/agar powder with 300 mL of distilled water in a 500 mL flask and placed in the autoclave in liquid. After LB broth was cooled to approximately 60°C, 300 uL of 100ug/mL Ampicillin stock was added, swirled, and plated in

sterile petri dishes near flame. Plates wrapped in tin foil and stored at 4°C.

Transformation of plasmid

Removed BL21 gold competent cells and P266L T7 polymerase plasmid out of -80°C and thawed on ice. Agar plates were warmed up to room temperature. Mixed 1 µl of P266L T7 polymerase plasmid into 40 µL of competent cells in an eppendorf tube and incubated the competent cell/plasmid mixture on ice for 20 minutes. Heat shocked the transformation tube in 42°C water bath for 45 seconds and placed the tube back on ice for 5 minutes. 500 µl LB solution was added to the bacteria and grow in 37°C shaking incubator for 1 hour. Plated all of the transformation solution onto LB agar plate containing ampicillin and incubated plates at 37°C overnight. A single colony from the LB agar plate was scooped using a pipet tip and added to 5mL of LB broth with ampicillin (200ug/mL) in a falcon tube and kept in a 37°C shaking incubator for 16 hours.

Inoculation and Elution

Inoculated 6 liters of auto-induction media with starter, grew at 37°C for 3 hours and checked OD600 every hour until it no longer increased. Then 1.4 mL of IPTG was added to LB solution and grown overnight at 25°C until saturation. Spun down cultures, resuspended pellets to 50 mL in ice-cold lysis buffer (50mM sodium phosphate pH 8.0, 500mM NaCl, 4 mM imidazole, 0.1% v/v Triton X-100). Added 125uL PMSF to 0.5 mM and small spatula tip with lysozyme to falcon tube. Homogenized lysate with sonicator on ice (2.5 minutes on, 1minute off, 38 minute total processing) and spun down pellet lysate at 27,000rpm for 25 minute at 4°C. Added lysate over nickel sepharose column equilibrated in lysis buffer (50mM sodium phosphate pH 8.0, 500mM NaCl, 4 mM

imidazole, 5% glycerol) for batch purification and placed on rocker for 1 hour. Elution gradient wash buffer was used with an increasing amount of imidazole up to 40% and placed on the rocker for 5 min. Sample was spun down and supernatant was collected, with each elution after 20% imidazole a mixture of 500uL DTT and EDTA (10mM and 5mM, respectively) were added and filled to 50mL with 50% glycerol and stored at 4°C. 20uL of each elution was saved and stored before addition of glycerol for SDS-PAGE gel on all fractions. Samples were concentrated with a 10K ultracentrifuge filter and aliquoted 1mL (~5mg/mL) in T7 RNAP box in -20°C.

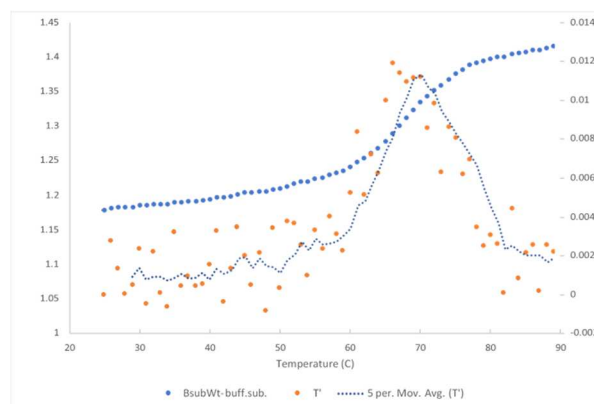


Figure 27. UV melt temperature determination of B.subWT T-box construct at 68°C

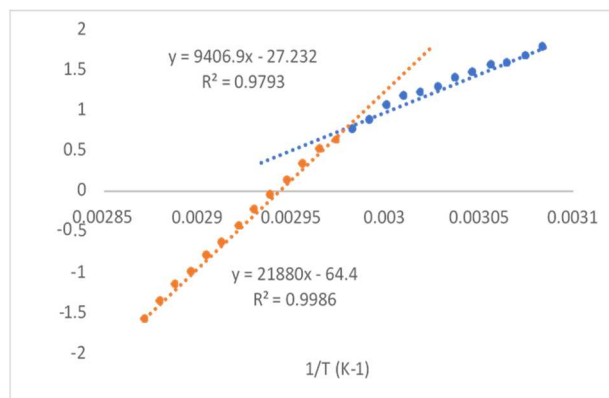


Figure 28. Enthalpy determination of B.subWT to be -6.60kcal/mol from plot of $\ln(Ka)$ versus $1/T(K^{-1})$.

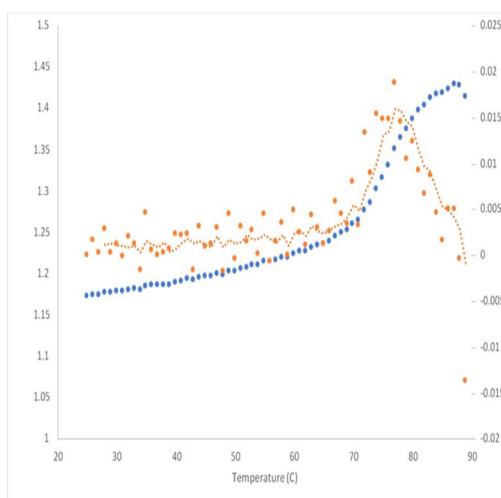


Figure 29. UV melt temperature determination of B.subA30G T-box construct at 77°C

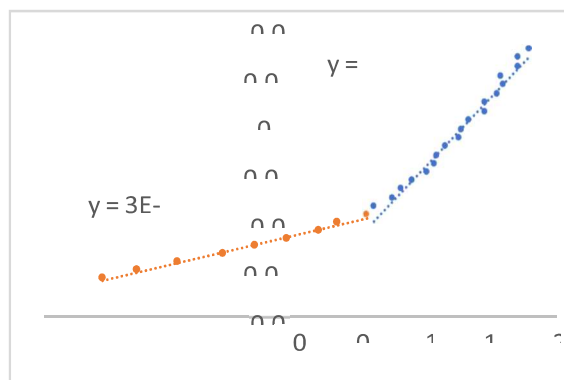


Figure 30. Enthalpy determining to be -5.01kcal/mol from plot of $\ln(Ka)$ versus $1/T(K^{-1})$ for A30G.

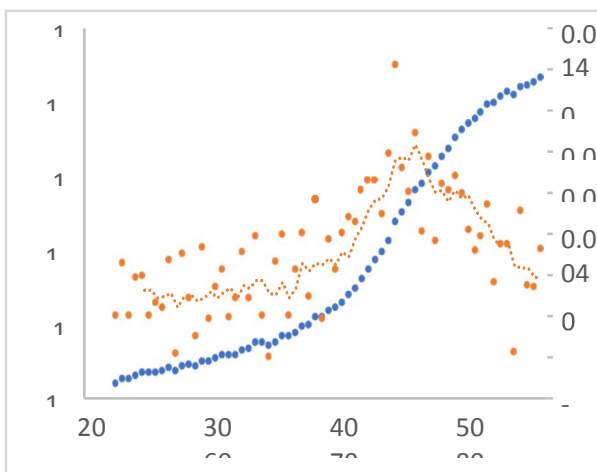


Figure 31. UV melt temperature determination of B.subA11U/C13A T-box construct at 70°C.

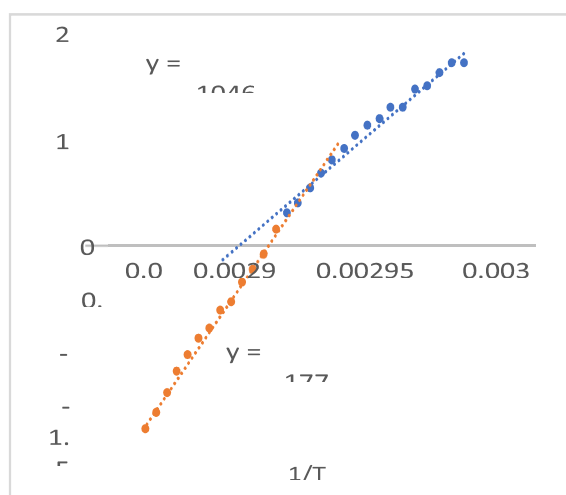


Figure 32. Enthalpy determining to be -6.34 kcal/mol from plot of $\ln(K_a)$ versus $1/T(K^{-1})$ for double point mutant.

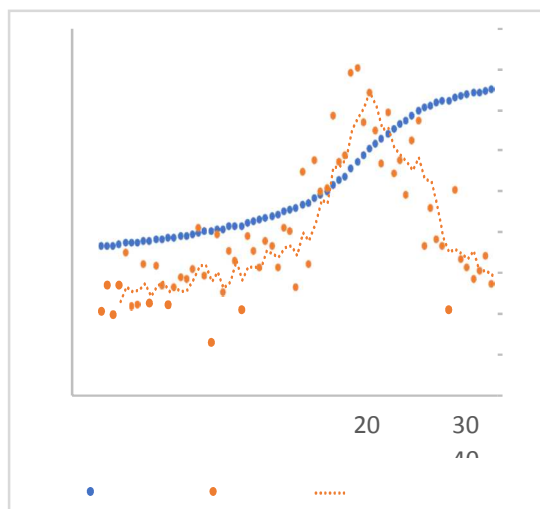


Figure 33. UV melt temperature determination of B.subC13U T-box construct at 70°C

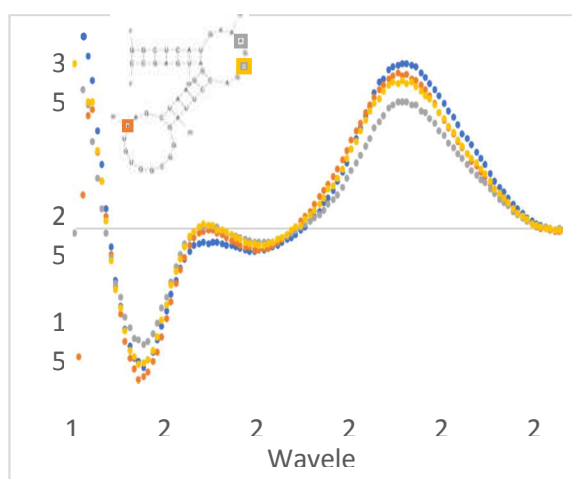


Figure 34. CD spectroscopy analysis of B.sub mutant constructs indicating ds-character with secondary structure representation indicating point mutant locations (C13U- yellow, A30G- orange, A11U/C13A- grey, and B.subWT- blue).

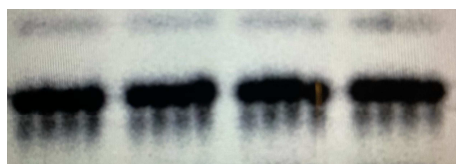


Figure 35. Our EMSA binding studies to tRNA by B.subWT, helix, A30G, and C13U mutants, respectively. Concentration of tRNA incubated with mutants increase from left to right.

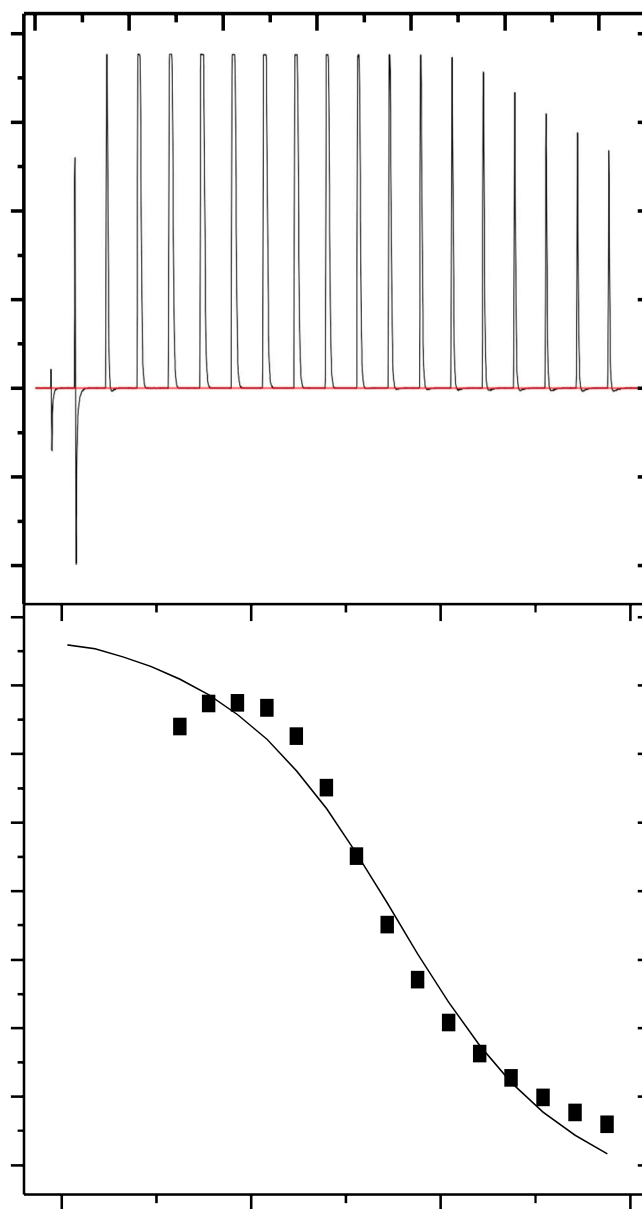


Figure 36. Isothermal Titration calorimetry (ITC) with tRNA (100 μ M) and B.sub WTT-box (1mM) given a K_d of $4.78 \pm 1.8 \mu$ M.

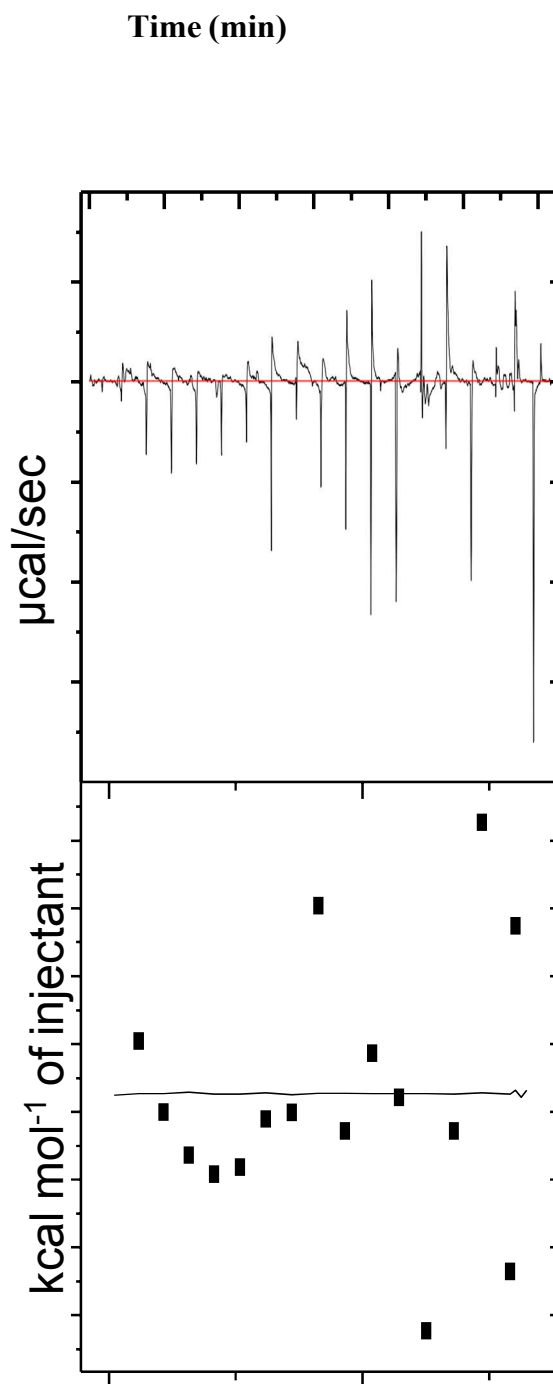


Figure 37. Isothermal Titration calorimetry (ITC) with tRNA (1mM) and B.subWT T-box (88 μM) indicating no binding.

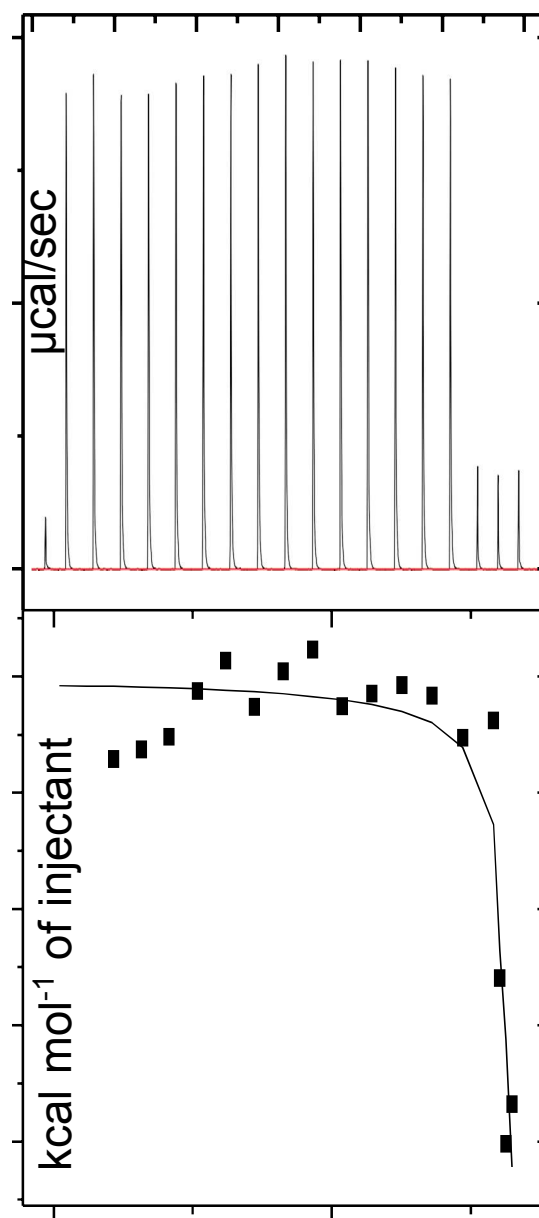


Figure 38. Isothermal Titration calorimetry (ITC) with tRNA ($50\mu\text{M}$) and B.sub C13UT-box ($250\mu\text{M}$) given a K_d of $7.35\pm 342\mu\text{M}$.

Time (min)

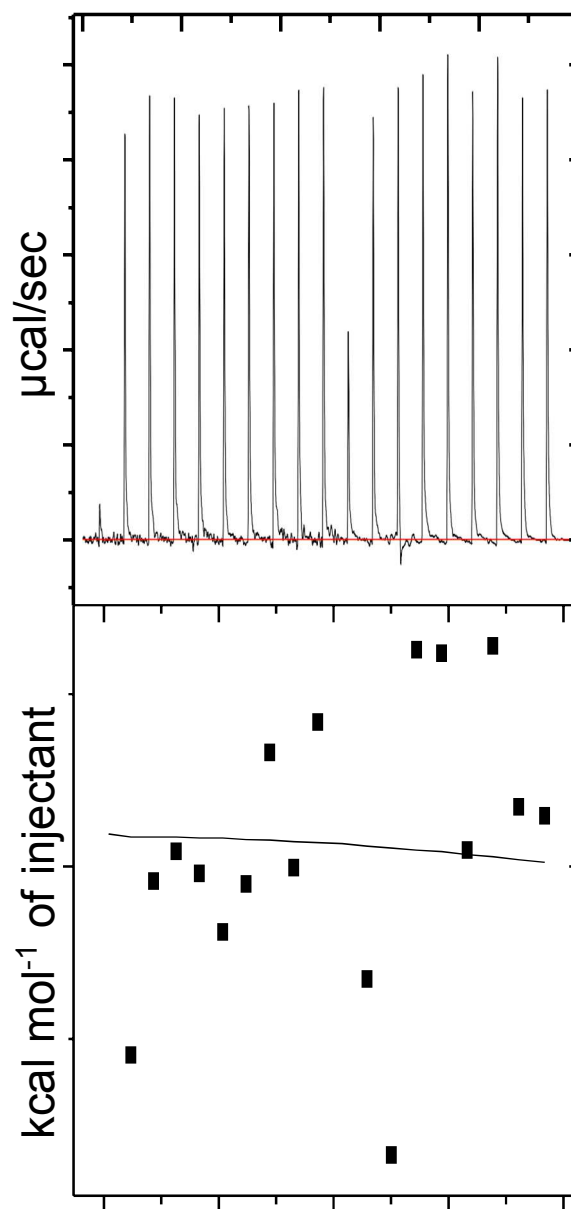


Figure 39. Isothermal Titration calorimetry (ITC) with tRNA ($50\mu\text{M}$) and B.sub C13UT-box ($250\mu\text{M}$) given a indicating no binding

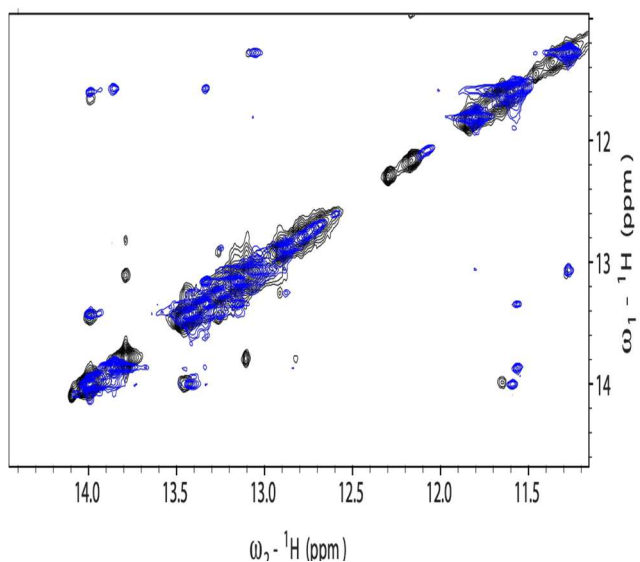


Figure 40. 2D 1H-1H NOESY of B.sub WT T-box construct overlaid with Helix construct for uracil base pairing determination both collected at 600MHz.

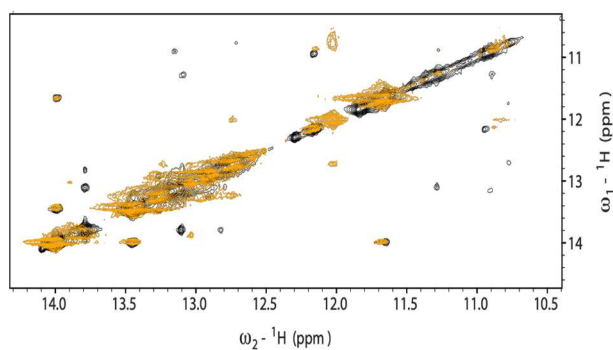


Figure 41. 2D 1H-1H NOESY of B.sub WT T-box construct overlaid with C13U construct for T-loop noncanonical base pairing determination both collected at 600MHz.