RICE UNIVERSITY

The T Box Mechanism and Anticodon Stem-Loops: Molecular and Structural Studies of Glycyl-tRNA Anticodon Stem-Loops and Their Binding to the T Box Specifier Domain.

by

Andrew Thomas Chang

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Edward P. Nikoforowicz, Associate Professor
Biochemistry and Cell Biology

George N. Bennet, E. Dell Butcher Professor
Biochemistry and Cell Biology

Christy Landes, Assistant Professor
Chemistry & Electrical and Computer Engineering

Kathleen S. Matthews, Stewart Memorial Professor
Biochemistry and Cell Biology

Yizhi Jane Tao, Associate Professor
Biochemistry and Cell Biology

HOUSTON, TEXAS
NOVEMBER, 2013
RICE UNIVERSITY

The T Box Mechanism and Anticodon Stem-Loops:
Molecular and Structural Studies of Glycyl-tRNA Anticodon Stem-Loops and Their
Binding to the T Box Specifier Domain.

by

Andrew Thomas Chang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Edward P. Nikonowicz, Associate Professor
Biochemistry and Cell Biology

George N. Bennett, E. Dell Butcher Professor
Biochemistry and Cell Biology

Christy Landes, Assistant Professor
Chemistry & Electrical and Computer Engineering

Kathleen S. Matthews, Stewart Memorial Professor
Biochemistry and Cell Biology

Yizhi Jane Tao, Associate Professor
Biochemistry and Cell Biology

Houston, Texas
November, 2013
Abstract

The T box mechanism is a riboswitch commonly used by Gram-positive bacteria to regulate expression of amino-acid related genes such as aminoacyl-tRNA synthetases (aaRS). The T box riboswitch regulates gene expression by the mechanism of transcription attenuation. The 5'-UTR of the mRNA forms mutually exclusive anti-terminator or terminator structures depending on whether the tRNA is uncharged or charged. This study focuses on the interactions that occur between T box specifier domain (SD) and tRNA anticodon stem-loop (ASL). This intermolecular interaction contributes to the specificity of the T box riboswitch. In bacteria, glycyl-tRNA molecules with anticodon sequences GCC and UCC exhibit multiple extratranslational functions, including transcriptional regulation and cell wall biosynthesis. In this study, the high-resolution structures of three glycyl-tRNA anticodon arms with anticodon GCC and UCC have been determined. Two of the tRNA molecules are proteinogenic, and one is non-proteinogenic and participates in cell-wall biosynthesis. The structures of the three tRNA\textsuperscript{Gly} anticodon arms exhibit small differences, and there is no evidence that they form the canonical U-turn motif. The Specifier domain of the T box riboswitch contains the Specifier sequence that is complementary to the tRNA anticodon and is flanked by a highly conserved purine nucleotide that could result in a fourth base pair involving the invariant U33 of tRNA. We show that the interaction between the T box Specifier domain and tRNA consists of three Watson–Crick base pairs and that U33 confers stability to the complex through intramolecular hydrogen bonding. The NMR data also suggest the ASL may change its structure to form a U-turn when in complex with the T box Specifier domain. Other T box domains also have specific tertiary structure. The NMR data in this study support the fact that the T box apical loop interacts with the AG loop, as seen in the crystal structure. The NMR data suggest that the U70 of the apical loop forms a reversed Hoogsteen A-U base pair with A73 and a U70G mutation has detrimental effects on the structures and interactions between the two loops.
Acknowledgement

This thesis is dedicated to my family, especially my grandmother, Teresa Tsui-Er Hong-Chen, who died from lung cancer in December, 2007, before the Christmas of my first year at Rice. Because I was away from my hometown in Taiwan, I did not meet my grandmother for the last time. For the completion of my graduate study, which is the career path I have chosen, the time of being with family has to be sacrificed. Grandmother Teresa had adored and treasured me the most, yet I could not stay by her side at that time, and I always feel sorry for this. Therefore, for Grandmother Teresa, I complete this thesis, in memory of my dearest Grandmother.

This thesis is completed because of the people mentioned below.

First of all, I would like to thank my advisor Dr. Ed Nikonowicz, who has always been helpful and patient. A good mentor and a good comrade in science, we have co-authored two journal publications which are imbued in this thesis. I would like also to thank the Department of Biochemistry and Cell Biology, especially to my thesis committee: Dr. Yi-Zhi Jane Tao, Dr. Kathy Matthews, Dr. George Bennett, and Dr. Christy Landes, for their scientific and career advices, and to Dr. Janet Braam, Dr. George Bennett, Dr. Mike Stern, and Dr. Bonnie Bartel, for offering me the chance to study at Rice University. I also would like to thank Dr. Shun-I Chu and his family, for their trust and recommendation.

Secondly, I would like to thank my friends and colleagues in the Department of Biochemistry and Cell Biology. My graduate seniors: Dr. Yu-Chang Tsai, Dr. Curtis Chun-Jen Lin, and Dr. Tom Sheng-Yaw Guu, have been my close friends and mentors since the first day I arrived at Houston. I would like to thank them for their companionship and help. I also would like to thank Shirley Shiu-I Liu, Dr. George Zhao Huang, and Dr. Peng Zhai from Silber and Matthews Labs, for great help during my first year. I would like to thank Dr. David Culberson from Pernilla/Olson lab for teaching me the technique of ITC. I would like to thank my colleagues in Nikonowicz Lab: Dr.
Margaret Michnicka for teaching me various techniques of RNA biochemistry, Dr. Jiachen Wang for teaching me techniques in structural calculation, and Dr. Andria Denmon and Dr. Troy Hammerstrom for their encouragement and advice. I would like to give special thanks to my classmate and friend, Dr. Joseph Faust, who not only helped me in my thesis writing but also taught me the ultimate secret technique to prevent leakage during polyacrylamide-gel preparation.

In addition, I would like to thank my landlady, Ms. Jane Su, and the owners of Teacup Cafe, Mr. Chung Huang and Ms. Hsueh-Hua Lee. These people offered comfortable places where I completed the thesis writing. It is because of my landlady that I do not have to worry about finding a place to live. Teacup Cafe provided me not only the best Taiwanese food in Houston but also many happy memories.

I also appreciate the help from the people of Rice Counseling Center, especially Dr. Mark Moeller, Dr. Elaine Frankel, and Dr. Salimah Khan. They help me in managing my obsessive compulsive disorder (OCD), which was triggered after I started my graduate school. Because of their help, I was able to overcome my worries and do experiments.

I also have to give thanks to my instructors from National Taiwan University: Dr. Chun-Nan Lee, Dr. Liang-Yin Lin, Dr. Ya-Chien Yang, and Dr. Wei-Horng Fang, for their recommendation so that I was admitted into Rice. I would like to thank Dr. Wei-Horng Fang and the lab’s seniors Yueh-Fang Hsu and Yu-Lun Lin. They not only taught me the fundamental lab techniques, but also the spirit of not giving up.

Finally, I would like to thank my family, especially my parents, Bao-Taa Chang and Maria Chien-Hui Hong. Without them being Rice alumni and getting married at Rice, I would not have been born. And they raised me and gave the chance to be educated and pursue advanced degrees. It is my luck to be their son, and I really thank them for their caring for so many years.

A Chinese/Taiwanese writer Chih-Fan Chen once wrote: “So many need to be thanked, so just thank God!” Indeed, I thank God for making all these happen.
# Table of Contents

Abstract ............................................................................................................................... ii  
Acknowledgement ............................................................................................................. iii  
Table of Contents .................................................................................................................v  
List of Figures and Tables ................................................................................................. vii  
List of Abbreviations .......................................................................................................... ix  
Chapter 1. Introduction ........................................................................................................1  
   1.1 Goal of the research .......................................................................................... 1  
   1.2 T box mechanism .............................................................................................. 1  
      1.2.1 Identification of the T box sequence .................................................. 2  
      1.2.2 Identification of conserved secondary structures and the T box mechanism .................................. 4  
      1.2.3 Specificity of T-box tRNA-mRNA recognition ......................................... 7  
      1.2.4 Known 3D structures of T box RNA ........................................................ 12  
      1.2.5 Genes regulated by T box ................................................................ 15  
   1.3 tRNA anticodon stem-loop ............................................................................. 18  
      1.3.1 A broader view of tRNA functions .................................................... 18  
      1.3.2 Modified bases in anticodon stem-loop ........................................... 19  
      1.3.3 The U-turn ........................................................................................ 25  
      1.3.4 Proteinogenic and non-proteinogenic glycyl-tRNA .............................. 28  
      1.3.5 Wobble rule and codon usage .......................................................... 29  
Chapter 2. Materials and Methods .....................................................................................32  
   2.1 Preparation of RNA samples ........................................................................... 32  
   2.2 NMR Spectroscopy ......................................................................................... 33  
      2.2.1 Assignments of non-exchangeable protons ...................................... 34  
      2.2.2 Assignments of exchangeable protons ............................................. 38  
      2.2.3 Techniques using difference spectra ................................................ 39  
   2.3 Structure calculation and refinement .............................................................. 41  
      2.3.1 NOE distance constraints ................................................................. 42  
      2.3.2 Dihedral angel and hydrogen bonding constraints .......................... 42  
   2.4 Isothermal titration calorimetry ...................................................................... 45  
   2.5 UV thermal melting ........................................................................................ 46  
   2.6 Denaturing and native analytical polyacrylamide gel electrophoresis .......... 47  
Chapter 3. NMR Structures of tRNA\textsubscript{Gly} anticodon stem-loops .........................48  
   3.1 Design of the hairpin molecules ..................................................................... 50  
   3.2 Thermal Stability ............................................................................................ 52  
   3.3 Resonance Assignments .................................................................................. 52  
   3.4 Effects of the metal ions ................................................................................ 59
List of Figures and Tables

Figure 1.1 The T box mechanism .................................................................2
Figure 1.2 Conserved T box motifs .............................................................4
Figure 1.4 Non-canonical base pairs .........................................................14
Figure 1.3 Watson-Crick base pairs ..........................................................14
Figure 1.5 How the bases pair .................................................................14
Figure 1.6 i6A and its derivative modifications .........................................19
Figure 1.7 Queosine ..............................................................................21
Figure 1.8 Uridine modifications .............................................................21
Table 1.1 The codon translation table .......................................................23
Figure 1.9 Guide for viewing stereo images .............................................25
Figure 1.11 3D structure of tRNA ............................................................26
Figure 1.10 Secondary structure of tRNA ................................................26
Figure 1.12 The U-turn structure .............................................................27
Figure 1.13 The penta-glycine bridge used by Staphylococci ..................28
Table 1.2 Wobble rule .........................................................................30
Table 1.3. tRNAGly isoacceptors in different organisms ..........................31
Figure 2.1 Atom nomenclature of RNA ..................................................34
Figure 2.2 1H-1H NOESY spectrum of ASLGly,GCC .................................35
Figure 2.3 13C-1H HSQC spectrum of ASLGly,GCC ................................36
Figure 2.4 HCN spectrum of ASLGly,GCC .............................................37
Figure 2.5 Two sugar pucker conformations ..........................................44
Figure 2.5 Sigmoidal curve of the ITC isotherm ......................................45
Figure 3.1 Sequences corresponding to the anticodon arms of Staphylococcus aureus (A) tRNAGly,GCC, (B) tRNAGly,UCC, and (C) np-tRNAGly,UCC .......................51
Figure 3.2 UV melting curves of ASLGly,GCC (red), np-ASLGly,UCC (blue), and np-ASLGly,UCC (green). ..............................................................51
Figure 3.3 Comparison of two-dimensional 13C-1H HSQC spectra of the base C6/8 regions of ASLGly,UCC in the (A) absence and (B) presence of Co(NH3)6 3+ and of np-ASLGly,UCC in the (C) absence and (D) presence of Mg2+ ..........................................................54
Figure 3.4 The effect of the addition of Co(NH3)6 3+ on the protonation state of ASLGly,UCC A38 .............................................................................54
Figure 3.5 Sequential assignments of the H8/6-H1’ resonances of the ASLGly s: (a) ASLGly,GCC, (b) ASLGly,UCC, and np-ASLGly,UCC ....................................................55
Figure 3.6 Comparison of the cytidine NH2 regions from 15N-1H HSQC spectra of (A) ASLGly,GCC, (B) ASLGly,UCC with Co(NH3)6 3+, and (C) np-ASLGly,UCC with Mg2+ RNA hairpins. .................................................................56
Table 3.1 Summary of chemical shifts of ASLGly s ....................................57
Table 3.2 Summary of experimental distance and dihedral angle constraints and refinement statistics for ASLGly molecules ..................................................60
Figure 3.7 Superposition of eight converged structures of (A) ASL<sub>Gly,GCC</sub>, (B) ASL<sub>Gly,UCC</sub>, and (C) np-ASL<sub>Gly,UCC</sub> RNA hairpins. .................................................................61
Figure 3.8 Stereoviews of the loop regions of (A) ASL<sub>Gly,GCC</sub>, (B) ASL<sub>Gly,UCC</sub>, and (C) np-ASL<sub>Gly,UCC</sub> RNA hairpins. .................................................................62
Figure 3.9 NOE cross peaks imply ASL<sub>Gly,GCC</sub> form the 6th and 7th base pairs. ....64
Figure 3.10 Comparison of the U-turn motifs ..........................................................68
Figure 4.1 Secondary structures of the Tyr<sub>SD</sub>Gly,GCC and ASL<sub>Gly,GCCac</sub> .............................................76
Figure 4.2 ITC experiments show the binding between SD and ASL. .......................78
Figure 4.3 Disassociation of the complex ..................................................................78
Figure 4.4 Native-PAGE of ASL-SD binding............................................................79
Figure 4.5 Imino 15N-1H HMQC of ASL Gly,GCC ..................................................81
Figure 4.6 Spectral changes of the SD during ASL binding ........................................81
Figure 4.7 NH region of the 1H 1D difference spectra of ASL<sub>Gly</sub>. .........................82
Figure 4.8 Imino proton connectivities between adjacent base pairs in H<sub>2</sub>O 1H-1H NOESY SD-ASL<sub>Gly</sub> complex.................................................................82
Table 4.1 Chemical shift table for SD-ASL binding..................................................83
Figure 4.9 15N-1H HMQC spectrum of U33A ASL<sub>Gly</sub>-bound SD. .........................85
Figure 4.10 Structural modeling of the SD-ASL complex ...........................................87
Figure 4.11 The anticodon-codon pairing in the SD-ASL binding ................................88
Figure 4.12 Structural changes in the loop E motif during SD-ASL binding ...............89
Figure 4.13 The structural changes in ASL ..................................................................91
Figure 5.1 Secondary structures of the T box apical and AG loops ............................95
Figure 5.2 Converged structures of the serS apical loop ............................................96
Figure 5.3 Superimposition of the serS apical loop structures ....................................96
Figure 5.4 Imino 15N-1H HMQC of the serS apical loop and AG loop .....................99
Figure 5.5 The HNCO spectrum of the serS double-loop molecule ..........................100
Figure 5.6 The amino 15N-1H HMQC of the full double-loop molecule ..................100
Figure 5.7 Base pairings between the apical loop and AG loop ...............................104
Figure 5.8 Proposed structure for the serS apical loop in conjunction with the AG loop. .........................................................................................................................105
Figure 5.9 Tertiary interactions between serS apical loop and AG loop .....................105
Figure 5.10 Comparisons of the apical loop with and without the existence of the AG loop .................................................................................................................106
Figure 5.11 Comparison of serS and glyQS double-loop structure ............................107
Figure 5.12 Sequence comparison of the serS and glyQS apical loop and AG box. ....107
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{13}\text{C}$</td>
<td>carbon-13</td>
</tr>
<tr>
<td>$^{15}\text{N}$</td>
<td>nitrogen-15</td>
</tr>
<tr>
<td>1D</td>
<td>one dimension</td>
</tr>
<tr>
<td>1H</td>
<td>proton</td>
</tr>
<tr>
<td>2D</td>
<td>two dimension</td>
</tr>
<tr>
<td>$^{31}\text{P}$</td>
<td>phosphate-31</td>
</tr>
<tr>
<td>3D</td>
<td>three dimension</td>
</tr>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>aaRS</td>
<td>aminoacyl tRNA synthetase</td>
</tr>
<tr>
<td>Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>ASL</td>
<td>anticodon stem-loop</td>
</tr>
<tr>
<td>Asn</td>
<td>asparagine</td>
</tr>
<tr>
<td>Asp</td>
<td>aspartate</td>
</tr>
<tr>
<td>ASV</td>
<td>avian sarcoma virus</td>
</tr>
<tr>
<td><em>B. stearothermophilus</em></td>
<td><em>Bacillus stearothermophilus</em></td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td><em>Bacillus subtilis</em></td>
</tr>
<tr>
<td>BCAA</td>
<td>branched chain amino acid</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>C</td>
<td>cytidine</td>
</tr>
<tr>
<td>cmnm$^5\text{U}$</td>
<td>5-carboxymethylaminomethyluridine</td>
</tr>
<tr>
<td>cmo$^5\text{U}$</td>
<td>5-oxyacetic acid uridine</td>
</tr>
<tr>
<td>Co(NH$_3$)$_6$$^{3+}$</td>
<td>cobalt hexamine</td>
</tr>
<tr>
<td>COSY</td>
<td>correlation spectroscopy</td>
</tr>
<tr>
<td>CT</td>
<td>constant time</td>
</tr>
<tr>
<td>cw</td>
<td>continuous wave</td>
</tr>
<tr>
<td>CX</td>
<td>complex</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>D$_2$O</td>
<td>deuterium oxide</td>
</tr>
<tr>
<td>DMAPP</td>
<td>dimethylallyl pyrophosphate</td>
</tr>
<tr>
<td>DQF</td>
<td>double-quantum filtered</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EF-Tu</td>
<td>elongation factor thermal unstable</td>
</tr>
<tr>
<td>Fe</td>
<td>iron</td>
</tr>
<tr>
<td>G</td>
<td>guanosine</td>
</tr>
<tr>
<td>GARP</td>
<td>globally optimized alternating phase rectangular pulse</td>
</tr>
<tr>
<td>H-bond</td>
<td>hydrogen bond</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>water</td>
</tr>
<tr>
<td>His</td>
<td>histidine</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HMQC</td>
<td>heteronuclear multi-quantum coherence spectroscopy</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence spectroscopy</td>
</tr>
<tr>
<td>I</td>
<td>inosine</td>
</tr>
<tr>
<td>i$^6$A</td>
<td>N6-isopentenyl adenosine</td>
</tr>
<tr>
<td>Ile</td>
<td>isoleucine</td>
</tr>
<tr>
<td>K$_2$C</td>
<td>lysidine</td>
</tr>
<tr>
<td>KCl</td>
<td>potassium chloride</td>
</tr>
<tr>
<td>$K_d$</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>m</td>
<td>medium</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
</tr>
<tr>
<td>M. capricolum</td>
<td><em>Mycoplasma capricolum</em></td>
</tr>
<tr>
<td>M. mycoides</td>
<td><em>Mycoplasma mycoides</em></td>
</tr>
<tr>
<td>Met</td>
<td>methionine</td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>magnesium ion</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>mnm&lt;sup&gt;s2&lt;/sup&gt;U</td>
<td>5-methylaminomethyl-2-thiouridine</td>
</tr>
<tr>
<td>mnm&lt;sup&gt;s&lt;/sup&gt;U</td>
<td>5-methylaminomethyl uridine</td>
</tr>
<tr>
<td>Mo-MuLV</td>
<td>Moloney murine leukemia virus</td>
</tr>
<tr>
<td>ms</td>
<td>millisecond</td>
</tr>
<tr>
<td>ms&lt;sup&gt;2j6&lt;/sup&gt;A</td>
<td>2-methylthio-N6-isopentenyl adenosine</td>
</tr>
<tr>
<td>ms&lt;sup&gt;io6&lt;/sup&gt;A</td>
<td>2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NH</td>
<td>amino</td>
</tr>
<tr>
<td>NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>amino</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleoside triphosphate</td>
</tr>
<tr>
<td>O</td>
<td>oxygen</td>
</tr>
<tr>
<td>OCD</td>
<td>obsessive compulsive disorder</td>
</tr>
<tr>
<td>OH</td>
<td>hydroxy</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>primer binding site</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Phe</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>ppm</td>
<td>part per million</td>
</tr>
<tr>
<td>preQ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>7-aminomethyl-7-deazaguanosine</td>
</tr>
<tr>
<td>Pro</td>
<td>proline</td>
</tr>
<tr>
<td>Q</td>
<td>queuosine</td>
</tr>
<tr>
<td>RMSD</td>
<td>root-mean-square deviation</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>s</td>
<td>strong</td>
</tr>
<tr>
<td>S</td>
<td>sulfur</td>
</tr>
<tr>
<td>S. aureus</td>
<td><em>Staphylococcus aureus</em></td>
</tr>
<tr>
<td>S. epidermidis</td>
<td><em>Staphylococcus epidermidis</em></td>
</tr>
<tr>
<td>S&lt;sup&gt;2&lt;/sup&gt;C</td>
<td>2-thiocytidine</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SD</td>
<td>specifier domain</td>
</tr>
<tr>
<td>sec</td>
<td>second</td>
</tr>
<tr>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>SHAPE</td>
<td>selective 2’-hydroxyl acylation analyzed by primer extension</td>
</tr>
<tr>
<td>t&lt;sup&gt;a&lt;/sup&gt;A</td>
<td>N6-threonylcarbamoyladenosine</td>
</tr>
<tr>
<td>Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>T</td>
<td>melting temperature</td>
</tr>
<tr>
<td>TOSCY</td>
<td>total correlation spectroscopy</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;Arg&lt;/sub&gt;</td>
<td>arginyl tRNA</td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;Asn&lt;/sub&gt;</td>
<td>asparaginyl tRNA</td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;Gly&lt;/sub&gt;</td>
<td>glycyll tRNA</td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;His&lt;/sub&gt;</td>
<td>histidyl tRNA</td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;Ile&lt;/sub&gt;</td>
<td>isoleucyl tRNA</td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;Leu&lt;/sub&gt;</td>
<td>leucyl tRNA</td>
</tr>
<tr>
<td>tRNA(^{\text{Lys}})</td>
<td>lysyl tRNA</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>tRNA(^{\text{Phe}})</td>
<td>phenylalanyl tRNA</td>
</tr>
<tr>
<td>tRNA(^{\text{Ser}})</td>
<td>seryl tRNA</td>
</tr>
<tr>
<td>tRNA(^{\text{Thr}})</td>
<td>threonyl tRNA</td>
</tr>
<tr>
<td>tRNA(^{\text{Tyr}})</td>
<td>tyrosyl tRNA</td>
</tr>
<tr>
<td>tRNA(^{\text{Val}})</td>
<td>valyl tRNA</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>U</td>
<td>uridine</td>
</tr>
<tr>
<td>Val</td>
<td>valine</td>
</tr>
<tr>
<td>vs</td>
<td>very strong</td>
</tr>
<tr>
<td>vw</td>
<td>very weak</td>
</tr>
<tr>
<td>w</td>
<td>weak</td>
</tr>
<tr>
<td>(\Psi)</td>
<td>pseudouridine</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1 Goal of the research

The central dogma is the foundation of current molecular biology and genetics – DNA is transcribed into RNA, and RNA is translated into protein. In the early version of the central dogma, RNA was commonly regarded as an intermediate between DNA and protein, effectively the amplification of the gene recorded as DNA. As time progressed and with the advancement of the science, RNA was determined to be more complicated and more important than merely a messenger. The importance of RNA not only comes from RNA other than mRNA, tRNA, rRNA, but also from the existence of introns, untranslated regions, and single nucleotide polymorphisms. Non-coding RNA is now known, like protein and DNA, to play an important role in gene regulation and expression. The T box, trp operon, and iron responsive element (IRE) are just a few examples.

Like proteins, RNA also contains secondary and tertiary structures. Conserved structural motifs of RNA are closely related to the function of specific RNA molecules. The goal of my thesis is to use NMR spectroscopy and biochemistry to contribute to a better understanding of the structure of two of these important RNA molecules: the tRNA anticodon stem-loop and T box mRNA specifier domain.

1.2 T box mechanism

Attenuation is a mechanism for controlling gene expression at the transcription level: the 5’ leader of the mRNA acts as a riboswitch by forming mutually exclusive alternative structures. glyQS and tyrS are two T box genes for glycyl-tRNA and tyrosyl-tRNA synthetase, respectively,
and their expression depends on the aminoacylation status of the bound tRNA. When an uncharged tRNA binds to the riboswitch, the anti-terminator hairpin stabilizes, and the downstream gene will be transcribed. However, when a charged tRNA binds to the riboswitch, the terminator hairpin forms. The terminator hairpin causes the release of RNA polymerase and thereby terminates transcription (Figure 2.1). This specific type of riboswitch in the mRNA leader is called the T box mechanism, which is widely adopted by Gram-positive bacteria.

1.2.1 Identification of the T box sequence

In 1992, Dr. Henkin and her colleagues first identified the T box sequence. At first, they identified the tyrS gene which codes for tyrosyl-tRNA synthetase in Bacillus subtilis. The gene’s 3’ end is adjacent to the end of the B. subtilis rpsD gene encoding ribosomal protein S4, and the gene

Figure 1.1 The T box mechanism
The T box mechanism works by forming two mutual exclusive structures – the terminator and anti-terminator hairpin loops. The tRNA anticodon stem-loop interacts with the T box specifier domain through anticodon-codon recognition. The terminator is the naturally more favorable structure so that the downstream gene is not transcribed. However, the acceptor stem of an uncharged tRNA can pair with the T box sequence and thus facilitate the formation of the antiterminator hairpin loop, which allows the transcription of the downstream gene.
is homologous to tyrS genes in other Bacillus species (Henkin et al., 1992). The product of this gene was confirmed to be a functional tyrosyl-tRNA synthetase because of the ability of the gene to complement an E. coli tyrosyl-tRNA synthetase mutant (Henkin et al., 1992). By comparing the tyrS leader regions of B. subtilis and Bacillus stearothermophilus, the Henkin group found a possible rho-independent transcriptional terminator in both leaders. It was known that B. stearothermophilus tyrS leader could serve as a transcription terminator in vivo when put into E. coli and in vitro when using E. coli RNA polymerase (Waye et al., 1986). It is most likely the same for other Bacillus tyrS leaders (Henkin et al., 1992). Besides the conserved terminator, they also found a conserved 18-nt long sequence immediately upstream the terminator. This 18-nt long sequence, which they called the T box, is perfectly conserved in B. subtilis and B. stearothermophilus tyrS leaders and therefore makes it extremely interesting and important to be considered as a regulatory signal. However, this T box lacks the sign of other transcription antitermination systems, such as E. coli pheST operon and thrS gene: E. coli pheST contains five tandem codon repeats for phenylalanine (Grunberg-Manago, 1987) and thrS contains a tRNA-like sequence (Moine et al., 1988).

The Henkin group constructed a tyrS-lacZ fusion containing tyrS leader and lacZ reporter gene to determine how the leader affects gene expression. The B. subtilis strain with the fusion gene showed little β-galactosidase expression. The expression of β-galactosidase increased greatly during tyrosine starvation. This result is consistent with the known fact that tyrosyl-tRNA synthetase expression responds to tyrosine starvation (Dale et al., 1971). In addition, this result also indicated that regulation occurs at the transcriptional level. Deletion of both T box and the terminator in the fusion gene led to increased β-galactosidase activity. This result implied a negative-regulatory role of the leader (Henkin et al., 1992). However, deletion of only the T box sequence led to decrease in β-galactosidase activity and loss of response to tyrosine starvation. This result suggested that the T box sequence is essential for antitermination (Henkin et al., 1992). The Henkin group also found that the T box was present in other Bacillus tRNA synthetase genes, such as thrS, thrZ, pheST, and trpS. Sequence analysis indicates these genes share the same regulatory mechanism.
1.2.2 Identification of conserved secondary structures and the T box mechanism

It did not take long to identify other secondary structures in the leaders of the T box containing genes. The Henkin group soon found that there also are a series of three stem-loop structures 5’ to the T box sequence and terminator. They designated these stem-loops I, II, and III. Furthermore, there is a bulge region near the base of stem I containing a triplet sequence specifying the corresponding amino acid (Grundy and Henkin, 1993). In the tyrS leader, the triplet is a tyrosine codon (UAC); in pheS, it is a phenylalanine codon (UUC); in thrS, it is a threonine codon (AAC), and so on. They also found, by altering the triplet in the tyrS-lacZ fusion to codons specifying other amino-acids, that the response of the T box will be sensitive to limitation of a different amino acid. For instance, when changing the tyrosine codon (UAC) in the leader to a phenylalanine codon (UUC), the fusion gene expression now responds to phenylalanine starvation but not tyrosine (Grundy and Henkin, 1993). They also changed the tyrosine codon to a non-sense amber (UAG) or orche (UAA) codon, and then the tyrS-lacZ fusion lost the ability to be induced. However, when using a B. subtilis strain containing the sup-3 allele, a modified lysyl-tRNA acting as amber and orche suppressor, the fusion gene regains partial induction in response to lysine starvation. This result strongly suggested that tRNA is a regulator in this antitermination mechanism (Grundy and Henkin, 1993). According to their result, the Henkin group proposed the model for the T box mechanism. They noticed that in all T box containing genes, the 5’ side of the termina-

![Figure 1.2 Conserved T box motifs](image)

The T box stem I contains several conserved motifs – the apical loop, the AG loop, the F box, the specifier domain, and the kink-turn motif. The T box refers to the conserved nucleotides that can pair with the tRNA acceptor stem. Some T box leaders also contains other structures such as stem II and stem III.
tor stem could base pair with the T box sequence to form an alternative structure containing part of the T box sequence and a bulge on one side, and they called this structure the antiterminator. In the model, uncharged tRNA interacts with the nascent RNA transcript via codon-anticodon recognition and stabilizes the formation of antiterminator. However, because of the steric interference of the amino acid, charged tRNA cannot trigger the formation of the antiterminator structure. Therefore, under normal conditions where tyrosine is abundant, charged tyrosyl-tRNA out-competes the uncharged tRNA, and, as a result, the formation of the terminator hairpin is favorable. Under tyrosine starvation, uncharged tyrosyl-tRNA becomes the major species and interacts with the leader. The interaction of the uncharged tRNA and the leader favors the formation of the antiterminator hairpin and results in continuous transcription of the gene (Grundy and Henkin, 1993).

The Henkin group later confirmed a second interaction site between the tRNA and the tyrS leader. The acceptor arm of the tRNA pairs with the side-bulge of the antiterminator hairpin. They found that the system is sensitive to mutations in the antiterminator bulge. Nucleotide substitutions in the side-bulge could cause 10-fold decrease in the expression of the tyrS-lacZ fusion gene (Grundy et al., 1994). However, a complementary mutation in the tRNA acceptor arm could suppress the effect caused by the substitution in the antiterminator bulge. This result supports the idea that the acceptor arm of the uncharged tRNA interacts with the side bulge of the antiterminator. These data also explain why charged tRNA with the amino acid interferes with the acceptor arm – T box interaction and does not have the same ability to stabilize the formation of the antiterminator (Grundy et al., 1994).

**Sequential and structural analysis of the tyrS leader**

Besides the T box sequence, the Henkin group also determined other conservation elements including the F box, AG box, GNUG box, AGUA-I box, AGUA-II box, and GAAC box. The AG box (bases 52-57) is an internal loop of stem I and is usually adjacent to a helical region, and mutations in the AG box decrease the expression of the gene (Rollins et al., 1997). The GNUG
box (base 68-71) is at the top of stem I. The single base substitution U70G in the GNUG box greatly compromises the expression and the induction elicited by tyrosine starvation. The F box is located between stem II and stem III. It consists of CCGUUA from bases 166-167. The U170C mutation next to the F box resulted in significant loss in basal expression and was uninducible via tyrosine starvation. This result suggested that the U in position 170 was required for antitermination (Rollins et al., 1997).

The GA motif is essential for antitermination

The GA motif is at the base of the stem I of the \textit{tyrS} leader. By analyzing 198 T box leader sequences, the Henkin group found that there was a conserved motif at the base of the stem I in 140 of the leader sequence. Based on the 140 T box leaders, the Henkin group identified a pattern for these conserved sequences and called the motif the “GA motif” because of its highly conserved GA dinucleotides. The GA motif consists of an asymmetric internal loop between two helices. The asymmetric internal loop is of the 5+2 arrangement and contains two opposite oriented GA dinucleotide sequences which are almost completely conserved, and mutations to the GA motif decreases the induction ratio of antitermination \textit{in vivo} (Winkler et al., 2001).

The Henkin group performed site-directed mutagenesis for the \textit{tyrS} leader and found that the GA motif is essential for antitermination. Mutations in the conserved GA dinucleotide sequences yielded the most serious effect in antitermination. Mutations A23G and A109C both led to a decrease in \textit{tyrS-lacZ} expression more than 500 fold compared to wild-type during tyrosine limitation. Mutation A21G resulted in 90-fold decrease of induced \textit{tyrS-lacZ} expression. These results suggested that the GA motif has strong preferences in primary sequence at these positions (Winkler et al., 2001).

The Henkin group suggested that the GA motif may facilitate the binding of tRNA to the T box leader because the GA motif is adjacent to the specifier domain. A similar motif was found in the protein binding site of the yeast L30 pre-mRNA, in which there was a bend of 130 degrees
between the two helices (Mao et al., 1999; Chao et al., 2004). The similarity of the GA motif and the L30 pre-mRNA indicates that the GA motif may have a significant role in the T box leader folding (Winkler et al., 2001).

### 1.2.3 Specificity of T-box tRNA-mRNA recognition

The Henkin group continued their mutagenesis study on the tyrS-lacZ fusion with a broadened spectrum this time. They changed the tyrosine codon on the specifier loop to codons specifying other amino acids besides phenylalanine, and changed the antiterminator to pair with the acceptor arm of sequence when needed. The result was positive. They were able to change the specificity of the tyrS-lacZ fusion from tyrosyl-tRNA to two seryl-tRNAs, threonyl-tRNA, and prolinyl-tRNA. However, the attempts in changing the specificity to tRNAArg, tRNAAsn, and tRNAHis resulted in failures. The structural differences between these tRNAs and tyrosyl-tRNA may be the reason why the antitermination could not occur for these tRNAs (these tRNAs differ the most from the tyrosyl-tRNA) (Grundy et al., 1997).

**In vitro antitermination by interaction of tRNA and the mRNA leader.**

Almost all of the studies about the T box mechanism before 2000 were done in vivo by using tyrosine-starvation induced tyrS-lacZ expression. Several attempts to study the T box antitermination in vitro, including tyrS and thrS of *B. subtilis*, were not successful (Grundy et al., 2002; Luo et al., 1998). It was a breakthrough when the Henkin group successfully used the glyQS leader of *B. subtilis* to demonstrate tRNA directed antitermination in vitro. glyQS is also a T box family gene which encodes glycyl-tRNA synthetase. Like its other T box gene relatives, glyQS contains a conserved T box leader acting as a riboswitch responding to glycine starvation. The T box leader in glyQS contains most of the conserved structures found in tyrS and other T box leaders except stem II and stem IIA/B structures, and the glyQS leader was believed to interact with glycyl-tRNA
The template for their in vitro transcription was a 440-bp PCR product containing from position -135 to 305 of the glyQS gene, and the position of the T box terminator was predicted to be on position 220. The group used the “halted complex assay”, which has different conditions for the initiation and elongation of the RNA transcript. The RNA polymerase they used in the in vitro study was purified from B. subtilis. The glycyl-tRNA used for antitermination was also produced in vitro. The read-through of the glyQS transcript was around 5% without the addition of glycyl-tRNA. Addition of glycyl-tRNA to the assay increased the read-through significantly to around 60%. The glyQS responded specifically to glycyl-tRNA in vitro, and addition of tyrosyl-tRNA resulted in termination. The folding of the nascent glyQS RNA transcript seems to play a crucial role for antitermination. The read-through transcription by adding glycyl-tRNA only worked at high Mg²⁺ concentration and at low NTP concentration. Mg²⁺ facilitates folding of the RNA secondary structure, and low NTP concentration lowers the speed of transcription and may give the transcript more time to fold into the correct structure. The Henkin group also applied the same in vitro antitermination assay on tyrS system, but without success. They also tested the tyrS leader RNA with the specifier changed to a glycine codon and observed transcription read-through. Therefore, they concluded that the codon-anticodon paring was not enough to account for the inability of tyrS to work in vitro (Grundy et al., 2002).

The NusA protein has been found to affect the rates of elongation and sensitivity of RNAP to stop at the termination site (Friedman and Court, 2001). NusA is also involved in several transcription termination systems (Friedman and Court, 2001). To test if NusA also affects the T box mechanism, the Henkin group added the NusA protein to the glyQS in vitro antitermination assay (Grundy et al., 2002). Under conditions without glycyl-tRNA, there was a small increase in termination by adding NusA (Grundy et al., 2002). However, under conditions with glycyl-tRNA, the addition of NusA had no effect on antitermination. This result suggested that T box mechanism could function without NusA, at least under this condition. The group also tested if the source of the RNAP affected the function of antitermination. Using RNAP from E. coli instead of B. subtilis
in the glyQS antitermination assays resulted in the same glycyl-tRNA responsive antitermination, thereby proving that it was the leader of the transcript that led to antitermination but not the RNAP (Grundy et al., 2002).

The Henkin group also tested the antitermination ability of glyQS variants by glycyl-tRNA variants in vitro. When the specifier of glyQS was changed from a glycine codon (GGC) to a cysteine codon (UGC) and the anticodon of the glycyl-tRNA was changed from the glycine anticodon (GCC) to the cysteine anticodon (GCA), the antitermination response was significant, although not as efficient as the wild-type glyQS-tRNA\textsuperscript{Gly}. Another antitermination assay of a variant with the mutation in the glyQS antiterminator bulge and a corresponding mutation in the glycyl-tRNA acceptor arm yielded a positive result. These in vitro antitermination assays successfully identified the determinants for specificity of the T box mechanism and were consistent with previous in vivo studies.

The success of the glyQS in vitro antitermination assay comprises the first and only T box leader shown to be functional in vitro. Under the in vitro conditions used by the Henkin group, there was no other factor required besides the in vitro synthesized glycyl-tRNA for the antitermination to work. Therefore, although not sufficient to prove that there is no other factor involved in the T box mechanism in vivo except the mRNA-tRNA interaction, one can still conclude that the interaction of the glyQS mRNA leader and the tRNA can function independently. Riboswitches sometimes require other factors, such as proteins, for proper function. For instance, the translation attenuation system in E. coli trp operon needs ribosomes. The fact that in vitro antitermination of glyQS requires only RNA indicates that the T box mechanism may be more primitive than other regulation mechanisms (Grundy et al., 2002).

*In vitro study of the interaction of glyQS leader and the glycyl-tRNA*

To verify the interaction of the specifier domain and the tRNA directly, the Henkin group designed in vitro Mg\textsuperscript{2+}-induced cleavage experiments. The rationale of these experiment was using
the nature of RNA cleavage caused by Mg$^{2+}$. Well-structured and paired RNA is more resistant to cleavage induced by Mg$^{2+}$ (Yousef et al., 2005).

When doing the Mg$^{2+}$ cleavage with the glycQS leader alone, the group found efficient cleavage occurred in the proposed loop regions, including the specifier domain and the linker regions between the stems. Adding glycyl-tRNA to the glycQS leader improved cleavage protection in specific areas, including the regions interacting with tRNA in the specifier domain, the antiterminator bulge, the linkers between stems I and III, and the linker between stem III and the antiterminator. Addition of glycyl-tRNA with an extra residue in the acceptor arm showed interesting results. Presumably unable to interact with the antiterminator bulge, the antiterminator bulge was no longer protected from the cleavage induced by Mg$^{2+}$. However, the region of interaction with tRNA in the specifier domain remains protected from Mg$^{2+}$ cleavage, indicating the codon-anticodon interaction is independent of the interaction between antiterminator bulge and the tRNA acceptor arm (Yousef et al., 2005).

In the same study, the Henkin group also used RNaseH mapping to verify the tRNA dependent structural switch of the glycQS terminator and antiterminator. The group designed a DNA oligonucleotide that can anneal with the 3’ side of the terminator. Under the condition of terminator formation, the DNA oligonucleotide would not be able to pair with the RNA transcript because the complementary RNA was in the terminator helix. The DNA oligonucleotide can only pair with the RNA transcript when its complementary RNA is available for pairing, such as under the conditions of antiterminator formation. RNaseH cleaves RNA-DNA hybrids and therefore can be used to probe the hybridization of the DNA oligonucleotide and the RNA transcript. The group found that in the presence of glycyl-tRNA, the glycQS RNA transcript was susceptible to RNaseH cleavage, whereas the transcript was resistant to RNaseH either with just the transcript alone or with the extended tRNA$^{Gly}$. This result confirmed the structural change from terminator to antiterminator in response to uncharged tRNA (Yousef et al., 2005).

Nelson et al. (2006) have successfully monitored the interaction between glycQS specifier domain and glycyl-tRNA anticodon stem-loop by using fluorescence quenching. In the study, they
synthesize two RNA molecules, “the common half” corresponding to the 5’ half of the glyQS stem I, and the other was the “specifier half” corresponding to the 3’ half of the glyQS stem I. In the specifier half, the two nucleotides adjacent to the specifier codon were substituted with 2-aminopurine ribonucleoside, a fluorescence reporter. The two RNA molecules were then annealed. Adding glycyl-tRNA anticodon stem-loop to the complex led to quenching of fluorescence and changes in the fluorescence emission spectrum, and the affinity of the complex for the anticodon stem-loop was calculated to be 20.2 µM with a ΔG˚ of -6.4 kcal/mol. Addition of tRNA^Phe had no effect on the fluorescence emission. The group also changed the specifier codon to a valine codon and tested the binding of anticodon stem-loop to the complex, resulting in quenched fluorescence, indicating the successful change of specificity (Nelson et al., 2006).

In vitro binding of the antiterminator to the tRNA

To test if the binding of antiterminator and tRNA required additional factors, Gerdeman et al. (2002) designed a gel-shift study. The group used ³²P-labelled antiterminator stem-loop and tyrosyl-tRNA. Adding of the tRNA slowed down the movement of the stem-loop on a native gel, but adding E. coli tyrosyl-tRNA (which has a different discriminator base) had no effect. The group then tested the binding of acceptor arm of the tRNA to the antiterminator. They used microhelices (mh-) corresponding to the acceptor arm of tRNA. The mh-UCCA, which is the complementary acceptor stem for the antiterminator stem-loop to bind, whereas using other mis-matched acceptor stems had no effect on stem-loop mobility (Gerdeman et al., 2002).

These in vitro gel-shift assays confirmed the interactions between the antiterminator and the acceptor arm and were consistent with previous in vivo experiments. The binding of the acceptor stem to the antiterminator did not require a protein factor, and the pairing of the variable base on the antiterminator bulge and the discriminator position of the tRNA acceptor arm have decisive roles in the binding specificity (Gerdeman et al., 2002).
1.2.4 Known 3D structures of T box RNA

**Structure of the T box antiterminator**

In 2003, Gerdeman *et al.* published a 3D model for antiterminator stem-loop based on the *tyrS* antiterminator. They solved the structure using NMR, and the structure is the first to be solved for the T box leader. The antiterminator loop motif is a 7-nucleotide bulge. An important feature of a bulge is its flexibility. The bulge structures in the human immunodeficiency type 1 virus (HIV-1) RNA genome are good examples for the flexibility of bulge structures. The TAR and RRE elements of the HIV-1 RNA change shapes after binding to their ligands (Puglisi *et al.*, 1992; Battiste *et al.*, 1994). The 7-nt bulge in the antiterminator interacts with the acceptor arm of tRNA. Therefore, Gerdeman *et al.* (2003) proposed that flexibility of the bulge was important for the tRNA binding feature in the antiterminator.

Gerdeman *et al.* (2003) demonstrated that the backbone of the bulge adopted an omega-like curvature. Except for the terminal bases in the bulge, all other bases in the bulge were flipped outward. The outward extrusion for the bulge bases makes the bases available for hydrogen bonding and as a potential region for molecular interactions. The group also showed there was a 80° kink caused by the 7-nt bulge between the two helices of the antiterminator. Bases in the 5’ portion of the bulge interact with the tRNA acceptor arm. The structure of this part of the bulge was less defined because of the lack of constraints and possible rapid sampling of multiple conformations. The 3’ portion of the bulge appeared to be base stacked, and its structure was better defined and less flexible than 5’ portion of the bulge. C224 of the 70-nt bulge is a well-conserved residue in T box antiterminator. A C to U mutation at this position resulted in decrease in efficiency of antitermination. The group also perform structural studies for a the C224U mutant. The mutation causes disruption of base stacking in the 3’ portion of the bulge. Therefore, the group proposed that the 3’ portion of the bulge partially constrains the conformational flexibility of the bulge. This stacking-induced structure in the 3’ of the bulge may facilitate the induced-fit or tertiary structure capture.
by restricting the number of possible conformations (Gerdeman et al., 2003).

Structure of the T box specifier domain and GA motif

The Nikonowicz lab solved the structure of tyrS specifier domain and GA motif (Wang et al., 2010; Wang and Nikonowicz, 2011). The specifier bulge adopts the structure of loop E motif. The name of the loop E motif comes from the E. coli 5S rRNA loop E, the first fold discovered of this kind that forms the possible binding site for ribosomal protein L25 (Correl et al., 1997). Besides being found in rRNA, loop E motifs are also found in several virus and hairpin ribozymes. The signature of the loop E motif is its non-Watson-Crick base pairs. There is a reverse Hoogsteen A-U, a sheared GA, and an A-A parallel interaction in the loop E motif of tyrS specifier domain (Figure 1.3, 1.4, and 1.5). The unpaired guanine between the two As of the A-A parallel and the reverse Hoogsteen A-U flips outward from the helix and makes a turn in the phosphate backbone. The turn and the flipping out of the guanine makes the A-A parallel possible.

The GA motif beneath the specifier domain was originally proposed to form a kink-turn. The kink-turn is an asymmetric loop between two helices and the asymmetric loop causes a ~120° bend of the two helix axes. Besides the asymmetric loop, there are two non-canonical G-A base pairs in one of the two stems. The kink-turn is seen in the L30 ribosomal protein binding site and in splicesomal U4 RNA.

Although the sequence of the GA motif fulfills the requirement of forming a kink-turn and the two sheared G-A base pairs were observed in NMR spectra, the GA motif does not form a kink-turn. The tri-loop only causes a bend of 25° instead of 120° between the two helices.

Structure of T box top loop along with the AG box

Recently, the GNUG loop was proposed to interact with the AG box (Lehmann et al., 2013). The two domains have sequences resemble of head-to-tail double T-loop structures seen in
Figure 1.3 Watson-Crick base pairs
The Watson-Crick A-U and G-C pairs are the canonical base pairs that makes the A-form RNA helix.

Figure 1.4 Non-canonical base pairs.
In this study, there are terms referring non-canonical base pairs such as parallel A-A, reversed Hoogsteen A-U, sheared G-A, G-U base pair, and A*-C mismatch.

Figure 1.5 How the bases pair.
The reversed Hoogsteen referes to the adenine pair with the uridine via the Hoogsteen edge and the reverse means the two glycosidic bonds are in the trans configuration (when considering the base-pair as a whole molecule). The “parallel” of the parallel A-A refers to that the backbones of the two strands have the same orientation. (Leontis and Westof, 2001)
the L1 stalk of 23S rRNA and in RNase P. The double T-loop structure is known to bind the elbow of tRNA by providing a platform. Another study later confirmed the double T-loop in glyQS Stem I and the data also support the interaction between the double T-loop with the tRNA elbow (Grigg et al., 2013).

The top loop bends toward the AG box and a complex hydrogen bonding network forms between the two domains. The study identified a total of five tertiary interactions between the top loop and the AG box. Three of the five are base triples and two are base-to-sugar hydrogen bonds. The U of the glyQS GNUG loop forms one of the base triples of the double T-loop structure.

**1.2.5 Genes regulated by T box**

*T box regulation in aminoacyl-tRNA synthetase genes*

The T box mechanism has been found to regulate most of the aaRS genes in *Firmicutes*, including *tyrS, thrS, thrZ, leuS, valS, pheS, tyrZ*, and *trpS* in *Bacillus sp.* (Reviewed by Gutiérrez-Preciado et al., 2007). The T box mechanism appears to be the most widely used regulatory mechanism for both class I and class II aaRS enzymes in *Firmicutes*. Gutiérrez-Preciado et al. stated that T box regulated aaRS genes were found generally to be in monocistronic transcriptional units with three major exceptions. The first exception is when the genes encode different polypeptides in heterodimeric enzymes, for instance, *glyQS*. The second exception is when the genes are associated with amino acid biosynthesis, such as *cycS*. The third is when two aaRS enzymes for different amino acids are encoded by the same operon, for example, the *hisS-aspS* operon (Reviewed by Gutiérrez-Preciado et al., 2007).

Generally, there is one T box leader per aaRS gene. However, a few aaRS genes were reported to have multiple T box leaders. *thrZ* is an extreme example in this category, for it has three tandem T box elements. To turn on the gene, three uncharged tRNA\textsubscript{Thr} are required to bind to the regulatory region providing tighter regulation. *thrZ* encodes an isozyme of ThrS, whereas the
major ThrS isozyme is encoded by \textit{thrS}. \textit{thrS} contains only one T box element. Therefore, \textit{thrS} is the gene to be turned on when uncharged tRNA$^{\text{Thr}}$ starts accumulating, and \textit{thrZ} is rarely induced in normal growth condition (Gutiérrez-Preciado et al., 2007).

The number of aaRS genes regulated by the T box mechanism in Gram-positive bacteria is highly variable. The number is high in \textit{Bacillaceae} but low in \textit{Actinomycetes}. This dramatic difference may be the result of evolution or adaptation to the environment (Gutiérrez-Preciado et al., 2007).

\textit{Amino acid biosynthetic genes regulated by T box}

In Gram-negative bacteria, amino acid biosynthetic genes are regulated by regulatory proteins responding to amino acid availability and attenuation of transcription responding to the speed of translation. The \textit{trp} operon in \textit{E. coli} is a good example (Reviewed by Gutiérrez-Preciado et al., 2007). In Gram-positive bacteria, regulatory proteins are also used. However, some bacteria use T box mechanism, for example, in many species in the phylum of \textit{Firmicutes}, the \textit{trp} operon is regulated by the T box mechanism. Several amino acid biosynthetic pathways, including Ser, Pro, Arg, Met, Cys, Ile, Leu, Val, Tyr, Phe, Trp, His, Asn, Asp, and Thr, have been found to be regulated by the T box mechanism. However, Gutiérrez-Preciado et al. (2007) analyzed 559 bacterial genomes and could not find any T box regulated operons for the Lys, Gln, and Gly synthesis pathways.

\textit{Amino acid transporter genes regulated by the T box mechanism}

In addition to aminoacyl-tRNA synthetase genes and amino acid biosynthetic genes, there are also amino acid transporter genes regulated by the T box mechanism (Reviewed by Gutiérrez-Preciado et al., 2007). According to Gutiérrez-Preciado et al., amino acid transporter genes regulated by the T box mechanism are exclusive in \textit{Firmicutes}. They have identified 34 different families of orthologous genes encoding amino acid transporter genes regulated by the T
box mechanism. The group found that T box leaders in these genes could act as an indicator for gene annotations. There are several amino transporter genes that are not well-characterized and therefore are categorized as hypothetical proteins or BCAA permeases. They found that nearly 70% of these genes are regulated by tRNA^{Ile} and 15% by tRNA^{Thr}. Others are also found to be regulated by tRNA^{Leu}, tRNA^{Val}, and tRNA^{Phe}. This result suggested that, although most of genes in this category encode proteins that transport branched-chain amino acids, other genes in this category may encode proteins that transport other amino acids. By using the predictive power of the T box mechanism, the group indicated a few possible errors of gene annotation in GenBank database. For example, by looking the T box leader of a gene in *Lactobacillus plantarum*, the group found that the gene had a His specifier sequence and was likely to encode a histidine transporter but the gene was incorrectly annotated as glutamine transporters (*glnPQH*). The group also found some other possible transporter genes, such as possible transporters for Asp, Asn, Phe, Cys, Met, and Ser, were incorrectly categorized as glutamine transporters.

*Regulatory proteins regulated by the T box mechanism*

The T box mechanism was also found to regulate operons encoding regulatory proteins (reviewed by Gutiérrez-Preciado *et al.*, 2007). Most of the regulatory proteins regulated by the T box mechanism were found to be cotranscribed with aaRS genes or with amino acid biosynthetic genes. The regulation of tryptophan synthesis in *B. subtilis* is related to the T box mechanism. A T box sequence responding to tryptophan regulates the transcription of *rtpA* gene, which encodes the antitermination regulatory protein. The AT protein controls the activity of the Trp-activated regulatory protein (TRAP), which responds to the accumulation of tRNA^{Trp}. Moreover, *rtpA* gene is co-transcribed with the *ycbK* gene. The product of the *ycbK* gene is thought to be a tryptophan transporter, providing an example of T box-regulated regulatory protein genes co-transcribed with amino acid biosynthetic genes (reviewed by Gutiérrez-Preciado *et al.*, 2007).
1.3 tRNA anticodon stem-loop

Transfer RNA, or simply tRNA, is one of the most recognized species of ribonucleic acid in cells along with ribosomal RNA (rRNA) and messenger RNA (mRNA). Just as the name implies, the primary function of tRNA is to transfer amino acids to ribosome for protein translation. The anticodon of tRNA of different amino acid species recognizes and pairs with the complementary codon on mRNA so that the protein will have the correct amino acid sequence.

The genetic code is degenerate since 61 of the 64 codon combinations are not stop codons and there are only 20 amino acids: one codon can only specify one type of amino acids, whereas each type of amino acid uses one or more codons. At least one tRNA species corresponds to each of the 20 amino acids. The codon usage and the number of tRNA species vary between different organisms from 22 in mitochondria and 31 in mammalian cells. Some tRNA species can read more than one codon (wobble), and the wobble position is the first anticodon nucleotide. The wobble rule and codon usage will be discussed in detail in section 1.3.5.

1.3.1 A broader view of tRNA functions

tRNA is generally known for its cellular role in translation, where the tRNA carries amino acids to the ribosomes. However, in some organisms, tRNAs are not only used in translation but also in other biological processes. The previously mentioned T box mechanism is an example of tRNA being used in gene regulation. Some other biological processes using tRNA include viral replication, antibiotic synthesis, suppression of pre-mRNA splicing, and bacterial peptidoglycan synthesis.

As reviewed by Mak and Kleiman (1997), tRNA is used as primer for reverse transcription in several retrovirus species. Notable examples are human immunodeficiency virus (HIV), avian sarcoma virus (ASV), and Moloney murine leukemia virus (Mo-MuLV). Different viruses may use different tRNA species as primer: HIV uses tRNA^{Lys}, ASV uses tRNA^{Trp}, and MuLV uses
tRNA\textsuperscript{Pro}. To start the viral reverse transcription, tRNA binds to the primer binding site (PBS) of the viral genomic RNA. Moreover, specific primer tRNA species are selectively packaged into the virus particle. (Kleiman, 2002; Mak and Kleiman, 1997)

### 1.3.2 Modified bases in anticodon stem-loop

It has been known for half a century that there are base modifications in RNA molecules. There are more than 100 different base modifications in RNA, 75 of them are found in tRNA. Many of the modifications occur in the anticodon stem-loop. The two most modified nucleoside positions in tRNA are position 34 and 37. Position 34 is the first anticodon nucleotide, and position 37 is 3’ adjacent to the anticodons (reviewed by Gustilo et al., 2008; reviewed by Agris et al., 2007; reviewed by Nishimura and Watanabe, 2006).

The anticodon stem-loop interacts with the T box specifier domain, and tyrosyl-tRNA anticodon stem-loop interacts with tyrS specifier domain. Within the tyrosyl-tRNA anticodon stem-loop, there are four base modifications. These four base modifications are queuosine on position 34 (Q34), 2-methylthio-N6-isopentenyl adenosine (ms\textsuperscript{2}i6A37) on position 37, and pseudouridine on position 39 (Ψ\textsubscript{39}) of the tRNA molecule.

The ms\textsuperscript{2}i6A actually contains two distinct modifications - methylthiolation and isopentenylation (Figure 1.6). The adenosine is isopentenylated to become i6A by the enzyme MiaA, and

![Image of chemical structures showing the modification process]

**Figure 1.6 i6A and its derivative modifications**
The MiaA enzyme modifies adenine to i6A (Cabello-villegas et al., 2002), and MiaB subsequently modifies the residue to ms\textsuperscript{2}i6A (Hernández et al., 2007). In some bacterial species such as Salmonella, the residue can be modified by MiaE to ms\textsuperscript{2}i6o6A (Mathevon et al., 2007).
MiaB modifies it to become ms\textsuperscript{2}i\textsuperscript{6}A by adding a methylthio group to the isopentenylated adenosine (reviewed by Booker et al., 2007). The ms\textsuperscript{2}i\textsuperscript{6}A modification in position 37 is in all \textit{E. coli} tRNA reading codons starting with U except for serine (reviewed by Nishimura and Watanabe, 2006). In \textit{Salmonella typhimurium}, the base is further modified to the hydroxylated derivative ms\textsuperscript{2}io\textsuperscript{6}A\textsubscript{37} by MiaE (Mathevon et al., 2007).

MiaA recognizes the anticodon stem-loop, and uses the A\textsubscript{37} and dimethylallyl pyrophosphate (DMAPP) as precursors for the reaction. Our lab has successfully performed \textit{in vitro} isopentenylation reaction by using MiaA to make base-modified tRNA\textsuperscript{Phe} anticodon stem-loops (Cabello-Villegas et al., 2004). MiaB also recognizes the anticodon stem-loop; however, it needs the adenosine to be isopentenylated first. Pierrel \textit{et al.} (2004) have successfully purified an MiaB enzyme from \textit{Thermotoga maritima} and used the protein to generate (in very small quantities) anticodon stem-loop containing ms\textsuperscript{2}i\textsuperscript{6}A\textsubscript{37} successfully.

MiaB is a metalloprotein containing two 4Fe-4S clusters, and uses S-adenosylmethionine (SAM) and the isoprenylated adenosine as precursors for methylthiolation (Figure 5; Hernandez et al., 2007). Not being a typical enzyme, MiaB is not only a catalyst but also a precursor itself, acting as the sulfur donor for the methylthiolation reaction. This “self-sacrifice” propensity makes MiaB a single turnover enzyme \textit{in vitro} (reviewed by Booker et al., 2007).

The queuosine, or 7-[[4.5-cis-dihydroxy-2-cyclopenten-1-yl]-amino]methyl]-7-deaza-guanosine, is found exclusively in position 34 in tRNA for Tyr, His, Asn, and Asp. The Q34 modification is incorporated by the enzyme Tgt – tRNA-guanine transglycosylase (reviewed by Garcia and Kittendorf, 2005). Tgt facilitates the exchange of guanine with preQ\textsubscript{1} on the anticodon stem-loop, and has specificity (Nakanish \textit{et al.}, 1994). Eukaryotic Tgt incorporates queuine directly whereas prokaryotic Tgt enzyme usually does not (reviewed by Garcia and Kittendorf, 2005). It is suggested that the difference in binding pocket size between prokaryotic and eukaryotic Tgt enzymes permits the difference in substrate specificities (Romier \textit{et al.}, 1997). In most prokaryotes, preQ\textsubscript{1} is then modified by other enzymes to queuosine on site (Nakanish \textit{et al.}, 1994). Therefore, the Q34 modification normally involves steps (Figure 1.7).
Structural effects in anticodon stem-loop caused by base modifications.

Our lab had previously solved and compared the structure of unmodified *E. coli* tRNA
Phe anticodon stem-loop and i6A37 modified *E. coli* tRNA
Phe anticodon stem-loop. The unmodified anticodon stem-loop of *E. coli* tRNA
Phe is well structured, although without any modifications (Cabello-Villegas et al., 2002). The stem is an A-form helix; however, the loop region is not the canonical U-turn motif as seen in the crystal structure of fully modified yeast tRNA
Phe (Cabello-Villegas et al., 2002). The structure of i6A modified tRNA
Phe anticodon stem-loop remains largely unchanged. However, the i6A modification does increase the mobility of the nucleotides near the loop and disrupts the base pairing of U33 and A37, thus increasing the dynamics of the loop.
In addition, the loop also begins to favor a canonical U-turn motif, although it remains dynamic. The UV melting experiment also suggested that the i^6A modification decreases the base stacking in the anticodon stem-loop (Cabello-Villegas et al., 2002).

Besides the i^6A37 modified E. coli tRNA_{phe} anticodon stem-loop, our lab also solved the structure of the Ψ32 modified E. coli tRNA_{phe} anticodon stem-loop (Figure 1.8). The Ψ32 modified E. coli tRNA_{phe} anticodon stem-loop has a similar structure as the unmodified anticodon stem-loop (Cabello-Villegas et al., 2005). However, Ψ32 modified anticodon stem-loop possesses a less compact tri-loop and has increased overall stability and local base stacking. Moreover, the chemical shift differences between the two molecules indicate the structures of the loop regions are different (Cabello-Villegas et al., 2005).

Additional evidence indicates that base modifications in anticodon positions are important for the structure of the anticodon stem-loop. Durant et al. (2005) have solved the structure of tRNA^{lys} anticodon stem-loop. Their result indicated that the base modifications in the tRNA^{lys} anticodon stem-loop favored the forming of the canonical U-turn motif seen in fully modified tRNA anticodon stem-loops.

The effect of base modifications on codon selection and recognition

Base modifications facilitate codon selection and recognition, and base modifications are involved in codon bias and rare codon usage. tRNAs recognizing codons starting with U almost all contain i^6A37 or its derivatives and tRNAs recognizing codons starting with A often contain t^6A or its derivatives (Nishimura and Watanabe, 2006). According to these results, Nishimura and Watanabe suggested that the modifications at this position are important for codon recognition.

In addition, base modifications may be connected with rare codon usage. There are five tRNA isoacceptors to read the six arginine codons in E. coli. The tRNA^{Arg4}(UCU) isoacceptor is the tRNA that decodes the AGA and AGG arginine codons (Table 1.1). AGA and AGG are rare codons used by E. coli (reviewed by Gustilo et al., 2008). The tRNA^{Arg4}(UCU) isoacceptor
has a 2-thiocytidine at position 32 (s²C32), 5-methylaminomethyl uridine (Figure 1.8) at position 34 (mnm⁵U34), and N6-threonylcarbamoyladenosine at position 37 (t⁶A37). It is known that the s²C32 affects the decoding of the AGG codon (Jager et al., 2004). Agris et al. (2007) have suggested that these modifications may be crucial for decoding the rare codons. Compared to the isoacceptor for rare arginine codons, the isoacceptors decoding the two common arginine codons CGU and CGC have an inosine (I34) at position 34. Therefore, Agris et al. (2007) also suggested that there was a connection between base modifications and rare codon usage.

Another example is the isoacceptor for isoleucine. The AUA codon for isoleucine is the fifth rarest codon in *E. coli*. One of the isoleucine isoacceptors contains the lysidine modified C34 (K²C34). The anticodon is CAU in this tRNA. Without this modification, the isoacceptor will recognize methionine codon AUG instead of isoleucine codon AUA (reviewed by Gustilo et al., 2008).

The mnm⁵s²U34 (Figure 1.8) and t⁶A37 modifications are present in the tRNA⁰lys species

<table>
<thead>
<tr>
<th>1st nucleotide</th>
<th>2nd nucleotide</th>
<th>3rd nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>Phe Ser Tyr Cys</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>Phe Ser Tyr Cys</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Leu Ser Term Term</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Leu Ser Term Trp</td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>Leu Pro His Arg</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>Leu Pro His Arg</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Leu Pro Gln Arg</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Leu Pro Gln Arg</td>
<td>G</td>
</tr>
<tr>
<td>A</td>
<td>Ile Thr Asn Ser</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>Ile Thr Asn Ser</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Ile Thr Lys Arg</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Met Thr Lys Arg</td>
<td>G</td>
</tr>
<tr>
<td>G</td>
<td>Val Ala Asp Gly</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>Val Ala Asp Gly</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Val Ala Glu Gly</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Val Ala Glu Gly</td>
<td>G</td>
</tr>
</tbody>
</table>

**Table 1.1 The codon translation table**
The standard codon translation table which shows the 64 codon combinations and their corresponding amino acid.
of all organisms. For the recognition of the AAA and AAG codons by tRNA$_{\text{Lys}}$, these two modifications are required. These two modifications increase the recognition of lysine codons by the tRNA$_{\text{Lys}}$ (reviewed by Gustilo et al., 2008).

The 5-oxyacetic acid uridine at position 34 (cmo$^5$U34) is a conserved modification in tRNA for alanine, leucine, proline, serine, threonine, and valine (Figure 1.8). This modification is required for decoding of the Valine GUG codon in _E. coli_ because the isoacceptor containing UAC codon without this modification can only decode the UAC anticodon. Moreover, unlike the unmodified tRNA, valine isoacceptors with this modification are able to decode all four valine codon in vivo. An _in vivo_ knock-out study showed that _Salmonella enterica_ was able to survive with all proline isoacceptors knocked out except the one containing cmo$^5$U34 (Nasvall et al., 2004). Furthermore, deletion of an enzyme involved in cmo$^5$U34 modification resulted in significant reduction of growth (reviewed by Gustilo et al., 2008).

*Modified bases in tRNA reduce frameshifting errors during translation*

The Björk group proposed a model for how a hypomodified tRNA causes frameshifting during translation (Björk et al., 1999; Urbonavičius et al., 2001). Depending on which site in the translation complex the hypomodified tRNA exerts its effect, two different mechanisms – the A-site effect and the P-site effect – are proposed in the model. For the A-site effect, the hypomodified tRNA has decreased ability to enter the A-site and is out-competed by other near-cognate tRNAs so that a near-cognate tRNA enters the A-site. After a three nucleotide translocation, the near-cognate tRNA does not fit optimally in the P-site. This lack of fit increases the frequency of slippage and frameshifting. Another situation that yields the A-site effect is when a hypomodified cognate tRNA moves into the A-site slowly and causes a pause in translation that offers time and a chance for the slippage to occur. For the P-site effect, the hypomodified cognate tRNA may go through the A-site selection; after the translocation, the hypomodified tRNA becomes a hypomodified cognate peptidyl-tRNA. The hypomodification causes a less optimal codon-anticodon pairing similar to a
near-cognate tRNA and thus increases the chance of slippage and frameshifting. A hypomodified tRNA can cause frameshifting by both the A-site and P-site effect (Urbonavicius et al., 2001).

The lack of Q34 modification leads to frameshifting by the A-site effect in codon species UAU (Tyr) and CAU (His) but not UAC (Tyr), CAC (His), or AAU (Asn). This result suggests that the Q modification is more important for efficient interaction with codons ending in U than with codons ending in C. Furthermore, there was no P-site effect for Q deficiency to cause frameshifting (Urbonavicius et al., 2001).

The effect of ms2i6oA37 deficiency on frameshifting by the P-site effect has also been tested. In miaA and miaB mutants, a 9-fold increase in frameshifting was observed when decoding the tyrosine codon UAU and 6- to 7-fold increase when decoding the tyrosine codon UAC (Urbonavicius et al., 2001). Similar results were found in miaA and miaB mutants for decoding phenylalanine codons. In a miaE mutant, the frameshifting was increased by 20%. The conclusion is that the ms2 modification is the main contributor of reading frame maintenance for decoding tyrosine and phenylalanine codons, whereas the i6 and hydroxy groups play only minor roles (Urbonavicius et al., 2001).

1.3.3 The U-turn

U-turn is an important RNA tertiary structural motif. First discovered in the anticodon

![RNA](RNA.png)

**Figure 1.9 Guide for viewing stereo images**
There are plenty of stereo images in this thesis. Stereo images have better demonstration of 3D molecular structure because of the depth of field. Viewing stereo image may be challenge for some people, so here is a tutorial by using the word “RNA” as an easy practice before running into any other stereo figures. One should adjust his/her focus of the eyes trying to overlaying the two “RNA” into a single one in the middle. If successfully putting eyes out of the normal focus to the correct focal point, one should see that “N” is closer than “R” and “A” is closer than “N”.
Figure 1.10 Secondary structure of tRNA
The tRNA consists of four major motifs – the acceptor stem (green), the D arm (purple), the anticodon stem-loop (red), and the $\text{T}\Psi\text{C}$ arm (blue). Extra nucleotides (white) may occur between the domains, especially the extra arm between anticodon stem-loop and $\text{T}\Psi\text{C}$ arm, which is variable in size among different tRNA species.

Figure 1.11 3D structure of tRNA.
Above is a stereo image of yeast tRNA$^{\text{Phe}}$ (PDB 1TRA; Westof and Sundaralingam, 1986). The backbone is color-coded according to the secondary structure in Figure 1.9. The anticodon stem-loop (red) forms a U-turn. There are tertiary interactions between the domains especially the ones between the D arm (purple) and the $\text{T}\Psi\text{C}$ arm (blue). The tertiary interactions cause the 3D structure of the tRNA to be more like an L-shape rather than the clover shape of the secondary structure.
stem-loop and TΨC loop of yeast tRNA\textsuperscript{Phe} crystal structure (Figure 1.9, 1.10, 1.11), U-turn is the first terminal-loop motif to be characterized (Robertus \textit{et al.}, 1974). The iconic feature of the U-turn is an abrupt 180° turn in the phosphate backbone. In tRNA anticodon stem-loops adopting U-turn, the anticodon nucleotides follow the 180° turn (Figure 1.12).

There are two sequence motifs that can form the U-turn: UNR and GNRA (N=any nucleotides; R=purines), in which UNR is similar to that found in tRNA\textsuperscript{Phe} and GNRA tetra-loop has a modified form of a U-turn (Jucker and Pardi, 1995). A U-turn is stabilized by at least two hydrogen bonds (Figure 1.12). The UNR type of U-turn found in tRNA anticodon stem-loops, the two hydrogen bonds are U33 2’-OH with R35 N7 or C35 N4H2 and U33 N3H with the phosphate oxygen (O\textsuperscript{p1}, O\textsuperscript{p2}) after residue 35. In the GNRA tetra-loop, the same two hydrogen bonds seen in UNR also exist but are from the G instead and an additional sheared G-A hydrogen bond is present (Jucker and Pardi, 1995).

Among available tRNA structures of tRNA\textsuperscript{Phe}, tRNA\textsuperscript{Lys} and tRNA\textsuperscript{Cys}, the anticodon stem-loops adopt a U-turn. U-turns also occur in other RNA molecules. Examples of non-tRNA U-turn include the hammerhead ribozyme (Pley \textit{et al.}, 1994), HIV LTR (Puglisi and Puglisi, 1998), U2

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig12.png}
\caption{The U-turn structure}
\label{fig12}
\end{figure}

This stereo image shows a U-turn motif and the iconic hydrogen bonds. The anticodon nucleotides are colored in green and the conserved U33 is colored in pink. The two hydrogen bonds stabilize this U-turn are U33 H3 to C35 OP1 and C35 H41 to U33 O2’.
snRNA loop IIa (Stallings and Moore, 1997), and several motifs in 23S rRNA (Huang et al., 1996; Conn et al, 1999; Culver et al., 1999). With GNRA tetra-loops having potential to form U-turns, the U-turn is now deemed as a more universal motif (Jucker and Pardi, 1995), and is predicted to occur in more places in the 16S and 23S rRNA (Gutell et al., 2000).

The U-turn allows the nucleotides following to be exposed to solvent. Therefore, the U-turn is thought to provide an interface for tertiary interactions. The U-turns seen in the tRNA anticodon stem-loop, tRNA TΨC loop, hammerhead ribozyme, and GNRA tetra loop all involve in inter- or intra-molecular tertiary RNA-RNA interactions.

### 1.3.4 Proteinogenic and non-proteinogenic glycyl-tRNA

Peptidoglycans are used by Gram-positive bacteria as a major component of the cell wall, whereas in Gram-negative bacteria the peptidoglycan layer is much thinner and lipopolysaccharide is the majority. The interpeptide bridge that crosslinks the peptidoglycan chains varies in length between bacterial species ranged from 0 to 5 amino acid residues (Vollmer et al., 2008). In *Staphylococci*, the peptidoglycan contains an unique pentaglycine bridge between peptidoglycan chains. The pentaglycine bridge makes the cell wall of *S. aureus* 80% cross-linked and is believed to be more flexible. Proteins contributing to virulence can anchor on the highly cross-linked cell wall. It has been shown that the cross-linkage is necessary for methicillin resistance in *S. aureus*.

![Figure 1.13 The penta-glycine bridge used by Staphylococci](image)
(MRSA). Disrupting the pentaglycine lowered the resistance of methicillin resistance. Unlike other amino acids in peptidoglycan synthesis, the glycines of the pentaglycines bridge are added by the FemABX complex from glycy1-tRNA (Figure 1.13).

In the 1970s, two novel glycy1-tRNA isoacceptors (named tRNA\textsuperscript{Gly1A} and tRNA\textsuperscript{Gly1B}) were discovered and isolated and do not participate in protein synthesis but in cell wall synthesis (Stewart \textit{et al.}, 1971; Roberts \textit{et al.}, 1974). According to previously published sequence analysis, these two glycy1-tRNAs use the anticodon UCC (Roberts \textit{et al.}, 1974). However, the first U in the anticodon is not modified to cmnm\textsuperscript{5}U as is the case of proteinogenic UCC glycy1-tRNA (Roberts \textit{et al.}, 1974). Recently, Giannouli \textit{et al.} also found a similar glycy1-tRNA in \textit{S. aureus}. The non-proteinogenic glycy1-tRNA found in \textit{S. epidermidis} has low binding affinity for elongation factor thermal unstable (EF-Tu), explaining its non-proteinogenic characteristic. Interestingly, Giannouli \textit{et al.} (2009) also found the newly identified glycy1-tRNA isoacceptor in \textit{S. aureus} has low binding affinity toward \textit{S. aureus} EF-Tu. Therefore, they proposed the newly found UCC glycy1-tRNA in \textit{S. aureus} shares the same role as the non-proteinogenic UCC glycy1-tRNA in \textit{S. epidermidis}. Although not stated in their paper, the U34 in the new glycy1-tRNA is probably also not modified because its anticodon stem-loop shares the exact same sequence as the \textit{S. epidermidis} tRNA.

1.3.5 Wobble rule and codon usage

Francis Crick proposed the wobble rule in 1966 because there are 4\textsuperscript{3}=64 codon combinations but effectively there are less than 45 tRNAs. In the original version of the wobble rule, G and U can pair with each other and inosine (I) can pair with A, C, and U (Table 1.2). In the revised version of the wobble rule, the effect of base modification is included (Table 1.2). It is worth noting that in the new version of the wobble rule, U can recognize all four nucleotides on the wobble position (Table 1.2), and introducing base modification on U, for example, an S\textsuperscript{2}U modification, strips the ability of U to pair with U and C (Lim and Curran, 2001).

The frequency of a specific codon and the proportion of a specific tRNA isoacceptor differ
from organism to organism. It is also commonly known that because codon usage varies between species, poor protein expression during molecular cloning and genetic engineering can result from rare codon usage. The problem may be solved through codon optimization. The use of rare codons in some genes may facilitate protein folding. Translation is stalled or slowed when a rare codon is encountered, so the protein has more time to fold into the correct structure or a chaperone can be recruited to help (reviewed by Marin, 2008; Zhang et al., 2009).

In Mycoplasma, only glycyll-tRNA isoacceptor with anticodon UCC exists (Table 1.3). Although the usage of the GGC codon ranges from none to 20 per thousand. It is possible that some mycoplasma only use one glycyll-tRNA isoacceptor to recognize all four glycine codons according to the revised wobble rule.

Watson-Crick base pairs contribute to the formation of standard A-form helix. However, with the diversity of the RNA tertiary structures, there are significant numbers of non-canonical base pairs, including wobble G-U, sheared G-A, and A-A parallel. Almost every combination of base pairs exists in RNA structures. These non-canonical base pairs are often responsible for the tertiary RNA-RNA or RNA-protein interactions.

Table 1.2 Wobble rule
The above table summarizes the original wobble rule proposed by Crick in 1966 and the modified version proposed by Lim and Curran in 2001. In the modified version the unmodified U is proposed to pair with all nucleotides (Lim and Curran, 2001).
Table 1.3. tRNA\textsuperscript{Gly} isoacceptors in different organisms

The table above summarizes the tRNA\textsuperscript{Gly} isoacceptors in different organisms. Not all organisms have tRNA\textsuperscript{Gly} corresponded to each glycyl codon. \textit{Mycoplasma} has only one tRNA\textsuperscript{Gly} isoacceptor for all glycyl codons whereas \textit{Caenorhabditis elegan} and other higher organisms have one for each glycyl codon.

<table>
<thead>
<tr>
<th>tRNA\textsuperscript{Gly}</th>
<th>C. elegan</th>
<th>E. coli</th>
<th>S. cerevisiae</th>
<th>B. subtilis</th>
<th>M. capricolum</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GCC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CCC</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ACC</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Chapter 2. Materials and Methods

2.1 Preparation of RNA samples

The RNA molecules used in this study were prepared mostly by in vitro transcription using T7 polymerase (Milligan et al., 1987). Commercial chemically synthesized RNA molecules were also used but rarely because of high cost. The sequence of the planned RNA molecule usually needs optimization. The transcription starts best with guanines, so three consecutive Gs usually serve as the starting nucleotides. The secondary structures of RNA molecules are predicted using mFold to screen for alternative folds. Ideally only one structure is predicted to form. Swapping the positions of two paired nucleotides may eliminate alternative foldings. Rarely, an A-U base pair needs to be changed to a G-C base pair or vice versa. Non-paired nucleotides are always the same as the original sequence. Chemically synthesized DNA oligonucleotides were used as templates for transcription. A T7 promoter sequence is always added before the desired sequence template. The top strand and bottom strand of the template were ordered separately and annealed later to form the double-stranded template stock at 20 µM with 10 mM Tris-HCl pH 8.0 and 10 mM EDTA. The annealing of template was done by heating to 95°C for 60 seconds and then cooling at room temperature.

The transcription buffer was typically 10 mM Tris-HCl pH 8.0, 25 mM MgCl₂, 5 mM DTT, 1 mM Spermidine, 0.01% Triton-X 100, 4 mM NTP mix, 10 mM GMP, 1 µM DNA template, and ~1000 U/mL T7 polymerase. Alternatively, 2 mM of ¹⁵N/¹³C labelled NTP mix was used to prepare ¹⁵N/¹³C RNA molecules (Nikonowicz et al., 1992; Michnicka et al., 1993). The transcription reaction was incubated at 37°C for at least 3 hours. To stop the reaction, an equimolar amount of EDTA to the Mg²⁺ was added. Normally, a small-scale reaction of 40 µL was performed first to test the efficiency of transcription, and then the reaction was scaled up to a volume of 1 mL to 10 mL, depending on the amount of RNA needed. The RNA was then precipitated with ethanol and
resuspended in 50% formamide and 25 mM EDTA with bromophenol blue and xylene cyanol as loading dye. The sample was heated at 95°C for 3 minutes and then loaded into 5 mm urea PAGE (10%-20%, depending on size of the molecule) for separation (500 V for ~16 hours). After the dye running to the desired position, the RNA band was visualized using UV shadowing and the band was then cut out. RNA was then electro-eluted from the gel slices using EluteTrap apparatus (Schleicher & Schuell). The eluted RNA was ethanol precipitated again to remove the Tris buffer and then resuspended in high-salt phosphate NMR buffer (20mM potassium phosphates at pH 6.8, 10 mM KCl, 1M NaCl, 2 mM EDTA). The sample was dialyzed using Amicon Centricon or Ultra (3000 M.W. cutoff) two times against high-salt NMR buffer and three times against low-salt NMR buffer (10 mM potassium phosphate at pH 6.8, 10 mM KCl, 0.2 mM EDTA). The final volume of the RNA sample was adjusted to 330 µL using low-salt NMR buffer in 10% D₂O or 99.6% D₂O depending on the type of NMR experiments.

The NMR samples for the ASL-SD complex were prepared in a specific way. The mixing has to be performed gently. Simply mixing two concentrated ASL and SD samples together could not form the complex properly, possibly due to aggregation. Desired amounts of SD and ASL RNA were annealed and were concentrated to volumes of 200 µL and 100 µL, respectively. Desired amounts ultra-pure Mg²⁺ were added to the buffer to a final concentration of 2 mM during the concentration process. To mix the sample, 10 µL of the ASL was added each time to the SD with intervals of two minutes between each mix.

2.2 NMR Spectroscopy

The NMR experiments of this study were performed using Varian Inova 500 MHz (¹H-[¹³C,¹⁵N,³¹P] probe) and 600 and 800 MHz (¹H-[¹³C,¹⁵N] cryoprobe). All data was analyzed and processed using Felix 2007 software (Felix NMR Inc., San Diego, CA). Essential NMR experiments performed for structural elucidation include ¹³C-¹H HSQC, ¹H-¹H NOESY, HCCH-TOCSY, HCN, and ¹³C-¹H NOESY-HSQC.
2.2.1 Assignments of non-exchangeable protons

Resonance assignments start with the sequential walk of inter- and intra-residue cross peaks of ribose H1’ and base H8/H6 using 1H-1H NOESY (Figure 2.2). Cross peaks in NOESY represent the magnetization transfer of spin to spin through space, a phenomenon known as the nuclear Overhauser effect (NOE), from which the experiment is named. The intensity of the peak is inversely related to the distance of the two spins, and therefore provides important spatial in-
formation of the spin system. However, without additional information from the $^{13}$C-$^1$H HSQC (Figure 2.3) and HCN spectra (Figure 2.4), figuring out the NOESY sequential walk would be very difficult. In an $^{13}$C-$^1$H HSQC spectrum, cross peaks represent the correlation between proton and carbon such as the base carbon C8/C6 (Figure 2.1) and the attached hydrogen H8/H6. In an HSQC spectrum, protons with the same chemical shift can be deduced by the different carbon chemical shifts of their attached carbon (Nikonowicz et al., 1993) (Figure 2.3). Moreover, the carbon chemical shift can also define the identity of the carbon and its attached proton. All purine H8s, pyrimidine H6s, and adenine H2s are in the same chemical shift range of 7-8.5 ppm. The chemical shifts of purine C8s and pyrimidine C6s are around 140 ppm, but adenine C2s are around 150 ppm. Chemical shifts of pyrimidine H5s are in the same range as of the H1’s between 5 and 6 ppm. The C1’ chemical shifts are around 90 ppm, whereas cytidine C5s are around 97 ppm and uridine C5s are round 104 ppm. Ribose protons other than H1’s are also in the same chemical shift range of 4-5 ppm. However, chemical shifts of C2’s and C3’s are at the seventies, C4’s at the eighties, and
C5's and C5''s at the sixties. Besides carbon chemical shift information, couplings of the carbon resonances also provide important information for assignments of different base species. Although purine C8/H8 and pyrimidine C6/H6 cross peaks are all in the same region, purine C8/H8 cross peaks are singlets, but pyrimidine C6/H6 cross peaks are doublets in the carbon dimension.

The HCN experiment utilizes the magnetization transfer from base H8/H6 and ribose H1’ through the C8/C6 and C1’ respectively and then to the base N9/N1 (Figure 2.4; Sklenár et al., 1993). This magnetization transfer provides the correlations of the base H8/H6 and ribose H1’ to the same N9/N1 and thus the correlation of the H8/H6 and H1’ of the same nucleotide (Figure 2.4). Therefore, the HCN experiment is a very powerful tool in deducing the NOESY sequential walk since the H8/H6 and H1’ pairs in the HCN spectrum should appear as intra-residue cross peaks.
Another feature of the HCN experiment is the ability to differentiate uridine and cytidine H6s. In the HCN spectrum, uridine N1s show up the most upfield at 145-150 ppm and cytidine N1s show up slightly downfield at 150-155 ppm. Adenine and guanine N9s are the most downfield at around 170 ppm. Therefore, the HCN experiment also acts as an alternative of $^{13}$C coupling in HSQC in differentiating purine H8s and pyrimidine H6s. The 3D $^{13}$C-edited NOESY-HSQC experiment is used in confirming the correctness of the NOESY sequential walk by isolating NOE peaks by carbon chemical shifts (Nikonowicz and Pardi, 1992). Moreover, the 3D NOESY-HSQC experiment can also identify parts where the sequential walk is overlapped itself or buried within the strong pyrimidine H5-H6 cross peaks.

After the completion of the NOESY sequential walk, an 3D HCCH-TOCSY experiment and sometimes an additional HCCH-COSY experiments are used in conjunction with the 3D

**Figure 2.4 HCN spectrum of ASL$_{Gly,GCC}^{Gly,GCC}$**

The HCN experiment gives correlation of the base H8/H6 and H1’ of a same residue. H8/H6s and H1’s are showed in pairs. The 15N chemical shift also gives information about the kind of the base. in the NOESY spectrum. Another feature of the HCN experiment is the ability to differentiate uridine and cytidine H6s. In the HCN spectrum, uridine N1s show up the most upfield at 145-150 ppm and cytidine N1s show up slightly downfield at 150-155 ppm. Adenine and guanine N9s are the most downfield at around 170 ppm. Therefore, the HCN experiment also acts as an alternative of $^{13}$C coupling in HSQC in differentiating purine H8s and pyrimidine H6s. The 3D $^{13}$C-edited NOESY-HSQC experiment is used in confirming the correctness of the NOESY sequential walk by isolating NOE peaks by carbon chemical shifts (Nikonowicz and Pardi, 1992). Moreover, the 3D NOESY-HSQC experiment can also identify parts where the sequential walk is overlapped itself or buried within the strong pyrimidine H5-H6 cross peaks.
$^{13}$C-$^1$H NOESY-HSQC to assign the ribose carbon and proton resonances. Unlike NOESY, the magnetization transfer is through covalent bonds in 3D HCCH-TOCSY and HCCH-COSY. The HCCH-COSY shows the correlations between directly coupled carbons (Pardi and Nikonowicz, 1992) whereas, the HCCH-TOCSY shows the correlations not only between directly coupled carbons but also between indirectly coupling protons. Therefore, an HCCH-COSY shows correlations between adjacent protons whereas an HCCH-TOCSY shows the correlations of the whole spin system, in this case, the protons of the entire ribose ring. $^{13}$C-Edited 3D HCCH-TOCSY provides extra separations based on the $^{13}$C chemical shifts (Pardi, 1995). By using the 3D HCCH-TOCSY, the chemical shifts of C2'/H2', C3'/H3', C4'/H4', and C5'/H5'/H5'' of a correlated C1'/H1' can be assigned. Sometimes, it is difficult to distinguish between C2'/H2' and C3'/H3' in a 3D HCCH-TOCSY experiment. Since only H1' and H2' correlations are visible in the HCCH-COSY, HCCH-COSY can be used to specifically assigned the H2’s without the interference of H3’s. An 3D HCCH-COSY RELAY experiment shows not only the H1’ and H2’ correlations but also the extended correlations of H1’ and H3’. Therefore, by using an HCCH-COSY in conjunction with an 3D HCCH-COSY RELAY experiment, the H2’ and H3’ chemical shifts can be differentiated and assigned to their corresponding residue.

### 2.2.2 Assignments of exchangeable protons

Base imino (NH) and amino (NH$_2$) protons exchange with hydrogen or deuterium ions in the solvent and are therefore called exchangeable protons. Protons exchanging with solvent can only be detected using 90% H$_2$O (10% D$_2$O) as solvent during NMR experiments. Assignments of the solvent exchangeable protons are important because their chemical shifts provide information about hydrogen bonding and base pairing. NMR experiments performed in H$_2$O include $^1$H-$^1$H NOESY and $^{15}$N-$^1$H HSQC/HMQC.

The $^{15}$N-$^1$H HSQC/HMQC experiment shows correlations between protons and their attached nitrogen (Nikonowicz et al., 1992). Proton and nitrogen chemical shifts retrieved from
the spectrum not only can distinguish resonances between different bases but also indicate hydrogen bonds. The imino proton resonances disperse between 10-16 ppm, whereas amino proton resonances locate in a region between 6-10 ppm depending on hydrogen bonding status. Guanine imino nitrogen resonances are usually upfield (145-150 ppm), whereas uridine imino nitrogen resonances are usually downfield (160-165 ppm). The guanine H1 chemical shifts are 12-13.5 ppm for Watson-Crick base pairs and 10-12 ppm for non-Watson-Crick base pairs such as G-U and G-A and for non-H-bonded guanines if observable. The uridine H3 chemical shifts are 13-15 ppm in Watson-Crick base pairs, 11-12 ppm in G-U base pairs, and 10-12 ppm when unpaired. Usually, the two cytidine amino protons are in slow exchange (rotation of the carbon-nitrogen bond), so two distinct amino proton resonances for each cytidine can be observed. In G-C paired cytidines, one amino proton of the pair locates around 8 ppm, where H-bonded cytidine amino protons should be, and the other one at 6-7 ppm. For unpaired cytidines, the two amino protons are both in the 6-7 ppm range, a chemical shift range for non-H-bonded cytidine amino protons. The chemical shift of cytidine amino N4 also gives information about hydrogen bonding. N4 chemical shifts from non-H-bonded cytidine aminos are upfield between 92-95 ppm, whereas H-bonded ones are downfield between 95-100 ppm. However, amino protons other than cytidines are typically undetectable using $^{15}$N-$^1$H HSQC because of the two amino protons exchange at an intermediate rate. (Varani et al., 1996)

An H$_2$O NOESY shows cross peaks from exchangeable protons in addition to cross peaks seen in a D$_2$O NOESY. Cross-strand NOEs can be observed in an H$_2$O NOESY from nucleotides involved in base pairing (Dieckmann and Feigon, 1997). Watson-Crick A-U base pairs can be detected by NOEs between adenine H2s and uridine imino H3s, and Watson-Crick G-C base pairs can be characterized by NOEs between guanine imino H1 to cytidine amino protons (H41 and H42). The non-Watson-Crick G-U base pairs are characterized by strong NOEs between guanine imino H1s and uridine imino H3s.

2.2.3 Techniques using difference spectra
Constant time (CT)-HSQC experiments with $^{31}$P coupled or decoupled can be used to measure the coupling by $^{31}$P to the carbons – the coupling constant $J_{CP}$. The coupling constant is determined by the intensity difference of cross peaks between the $^{31}$P coupled and decoupled CT-HSQC spectra, given that the intensity of the peak is proportional to $\cos(\pi J_{PC} t)$, where $t$ is the evolution time of the CT-HSQC experiment (Legault et al., 1995). Therefore, the $J_{PC}$ can be determined based on the following equation: $J_{PC} = \frac{\cos^{-1}(1-\Delta I)}{\pi t}$, where $\Delta I$ is the ratio of the peak intensities between $^{31}$P coupled and $^{31}$P decoupled CT-HSQC spectra. $J_{PC}$ of carbons coupled to more than one $^{31}$P can only be determined to a certain range since the $J_{PC}$ calculated from the above mentioned equation is the sum from the two phosphates. The method and the calculation are adapted from the detailed work done by Legault et al. (1995). This technique is used in the calculation of coupling constant $J_{PC4'}$ and $J_{PC2'}$. The $J_{PC4'}$ and $J_{PC2'}$ provide information about $\beta$ and $\epsilon$ torsional angels respectively.

Another difference technique is used for the 1D imino $^1$H spectrum of samples mixed with both labelled and unlabelled RNA molecules, in this case – labelled anticodon stem-loop and unlabelled specifier domain. The peaks coming from the unlabelled molecule are the same whether $^{15}$N is decoupled or not, whereas the peaks coming from the labelled molecule split when $^{15}$N is coupled and degenerate when decoupled. Subtraction of the $^{15}$N decoupled spectrum by the coupled spectrum yields a 1D imino $^1$H spectrum with only peaks coming from the labelled molecule (the anticodon stem-loop), thereby removing the interference from the unlabelled molecule (the specifier domain). The purpose of a 1D imino $^1$H spectrum is to detect protons that are in an intermediate exchange rate. Peaks from these protons do not show up in the HMQC because of the longer time frame of the entire pulse sequence. Moreover, by setting different $^{15}$N chemical shift as the center of the decoupling, the imino $^{15}$N chemical shifts of the labelled molecule (anticodon stem-loop) can be approximated. The imino $^1$H resonance shows incomplete cancellation when the $^{15}$N decoupling center is at or near the corresponding imino $^{15}$N resonance. When $^{15}$N decoupling is centered at 144 ppm, only peaks from guanine imino H1 show up in the difference 1D spectrum,
whereas only peaks from uridine H3 show up when the decoupling is centered at 160 ppm.

### 2.3 Structure calculation and refinement

Structural calculations and refinements in this study were performed using XPLOR-NIH software input by using its Python interpreter (Schwieters et al., 2003; Schwieters et al., 2006). There are two Python scripts – annealing and refinement scripts. The scripts were adapted from the template scripts on the XPLOR-NIH website and modified by Jiachen Wang and later Andrew Chang for different structural studies. XPLOR-NIH utilizes the simulated annealing algorithm.

A simulated annealing performs molecular dynamics by starting at high temperature and then decreasing the temperature gradually. At high temperature atoms can move more freely, to search for a wide range of possible energy states for a region of global minimum. As the temperature decreases, the atoms become more restricted, thus narrowing the search region, and the energy state approximates the local minimum. The annealing script has a high starting temperature to search for proper structural candidates. The refinement script is applied to the selected candidates and with low starting temperature to fine-tune and optimize the structure.

A linear RNA structure with the exact same sequence is generated using Insight II or 3DNA as the starting coordinate (Lu and Olson, 2003). After several runs of the annealing script, output structures with more correct geometry (for instance, correct base pairs) may substitute the original linear molecular as the new starting coordinate for better computational efficiency. Constraints are additional “rules” added to the molecular dynamics. Violations of these “rules” also increase the calculated energy. Therefore, the system tends to follow the constraints for minimum energy. Constraints are what shape the calculated structures and make the structures converge. Constraints are set based on NMR experimental observations and analysis of the assumed secondary structure. Three basic types of constraints are used in NMR structural calculation — NOE distance, dihedral angle, and hydrogen bonding.
2.3.1 NOE distance constraints

NOE distance constraints are derived from the intensities of cross peaks in the NOESY type spectra. The distance between two protons is inversely related to the intensity of the NOE cross peaks. Intensities of the NOE cross peaks are divided into five levels: very strong (vs), strong (s), medium (m), weak (w), and very weak (vw). Each distance constraint is set with a upper and lower limit thus the distance is restricted within a range. All cross peaks are set with a lower limit of 1.8 Å. Upper limit is set as 3.0 Å for very strong, 4.0 Å for strong, 5.0 Å for medium, 6.0 Å for weak, and 7.0 Å for very weak peaks. The H5-H6 intra-residue cross peaks can be used as a reference to be the strongest peaks (vs) since H5-H6 distances are rigid.

The base to 1’ region contains inter- and intra-residual cross peaks used in the sequential walk. In the same region, other peaks valuable for distance constraints include those inter-residual pyrimidine H5 to H6/H8 and adenine H2 to H1’/pyrimidine H5. The left-top and right-bottom corners of the NOESY are base to ribose region. Peaks in this region are intra- and inter-residual H8/ H6/adenine H2 to ribose protons other than H1’ NOEs that are also used in distance constraints. Base H8/H6s usually give cross peaks to ribose protons of themselves and the ribose protons of their 5’ neighbor, although cross peaks to the 3’ neighboring ribose may also be seen in few special cases. The middle top and bottom regions are the 1’ to ribose region. An H1’ usually gives strong cross peaks to its own and 5’ neighboring H2’ and H3’ and cross peaks to the H4’ and H5’/H5” are usually weaker. Cross peaks to the 3’ neighboring ribose are also rarely seen. The diagonal region of the NOESY contains some medium to very weak inter-residual cross peaks. In the center diagonal region are H1’ to H1’, H1’ to pyrimidine H5, and pyrimidine H5 to pyrimidine H5 cross peaks. In the downfield diagonal region are H8/H6 to H8/H6, H8/H6 to adenine H2, and adenine H2 to adenine H2 cross peaks.

2.3.2 Dihedral angle and hydrogen bonding constraints
Dihedral angle constraints can improve the convergence of the calculated structure greatly. Some dihedral angles are just loosely constrained to prevent structurally undesirable results. Most of the dihedral angle constraints involve the phosphate backbone, which compensate the lack of distance information for the phosphate backbone. Each dihedral angle is defined by four different atoms (Varani et al., 1996).

The $\alpha$ and $\zeta$ angles are defined by $O3'-P-O5'-C5'$ and $C3'-O3'-P-O5'$ respectively. $\alpha$ and $\zeta$ are usually loosely constrained. In a normal A-form helix, the $\alpha$ and $\zeta$ are constrained as $65^\circ\pm40^\circ$ and $70^\circ\pm40^\circ$ respectively. In other unknown regions, $\alpha$ and $\zeta$ are constrained as $0^\circ\pm120^\circ$ just to avoid trans conformations or left unconstrained.

The $\beta$ and $\varepsilon$ angles are defined by $P-O5'-C5'-C4'$ and $C4'-C3'-O3'-P$ respectively. For standard A-form helix and most cases, the $\beta$ and $\varepsilon$ dihedral angles are expected to be trans and restrained to $180^\circ\pm30^\circ$ and $210^\circ\pm40^\circ$ respectively. Coupling constants measured by the CT-HSQC can be used to confirm the conformation of these two angles. A $J_{PC4'}$ larger than 5 Hz indicates a trans $\beta$ angle. Gauche$^+$ and gauche$^-$ conformations of $\beta$, on the other side, correspond with $J_{PC4'}$ smaller than 5 Hz. A trans $\varepsilon$ angle is bundled with large $J_{PC4'}$ (> 5 Hz) and small $J_{PC2'}$ (< 5 Hz), whereas a gauche$^-$ $\varepsilon$ is on the opposite with small $J_{PC4'}$ and large $J_{PC2'}$. A gauche$^-$ $\varepsilon$ is also one of the signs of a C2'-endo sugar pucker (Figure 2.5).

The $\gamma$ dihedral angle is defined by $O5'-C5'-C4'-C3'$ and is gauche$^+$ in both A-form and B-form helices. Therefore, $\gamma$ angles are restrained as $60^\circ\pm40^\circ$ in assumed helical regions and are left unconstrained in other regions.

The $\delta$ angle is defined by $C5'-C4'-C3'-O3'$ and is related to the conformation of sugar pucker. The $\delta$ is constrained to $85^\circ\pm30^\circ$ and $160^\circ\pm30^\circ$ for C3’-endo and C2’-endo sugars respectively.

The glycosidic angle, $\chi$, usually does not need to be constrained except in cases where special structural motif is present. The $\chi$ dihedral angle is usually anti. A strong H8/H6 to H1’ NOE is indicative of an unusual syn conformation.

The sugar pucker is defined by three $\nu$ angles: $\nu_1$ (O4’-C1’-C2’-C3’), $\nu_2$ (C1’-C2’-C3’-
C4’), and v3 (C2’-C3’-C4’-O4’). For C3’-endo sugars, the three angles – v1, v2, and v3 are constrained as -26.2°, 36.5°, and -33.0°, all with a range of ±15°. For C2’-endo sugars, the three angles are constrained as 24.9°, -34.9°, and 33.3°.

Hydrogen bonding constraints consist of both distance and dihedral angle constraints. The distance constraints are set as 2.0 Å between the donor and acceptor. For base pairs, at least two 180° dihedral angle constraints are introduced per base pair to make the two paired bases co-

Figure 2.5 Two sugar pucker conformations.
In normal A-form RNA helix, the sugar ring adopted the normal C3’-endo configuration (b). C2’-endo configuration (a) may be adopted in B-form helix and some special motifs such as loop E motif. The different sugar configurations cause the adjacent phosphate backbones to turn differently.
 Isothermal titration calorimetry experiments of this study were performed using the VP-ITC calorimeter by MicroCal Inc. By titrating ligand to its receptor, ITC measures the heat released or absorbed for each injection, and the thermodynamic parameters and the binding constant can then be calculated. The plot of the molar ratio versus the heat changes of each injection form a titration curve that is called the isotherm. During the course of the titration, heat released/absorbed starts at a plateau where titrator is much more in excess and then decreases as the titrator is getting more and more saturated to another plateau near zero where the titrant is in excess. As a result, the shape of an isotherm is sigmoidal and is very much like a pH curve in an acid-base titration. The sigmoidicity of the isotherm is determined by the c value, which is defined by the product of the number of binding site (n), binding constant (K_B), and the concentration of the titrant (M_tot): 

\[ c = nK_B M_{tot} \]

The optimal c value for an ITC experiment is between 10 and 100 (Salim and Feig, 2009) (Figure 2.5). For simplicity, one can view the process of finding the K_B is to look for a c value that best fits the experimental isotherm since the c value affects the sigmoidicity. The fitting and data processing can be done by using the Origin 7.0 software provided by the instrument manufacturer. The software is optimized with macros specifically for ITC data analysis.

Figure 2.5 Sigmoidal curve of the ITC isotherm. The figure above shows different c values and their differences in the sigmoidal shape. A larger c value leads to a more steep inflection of the curve (Wiseman et al., 1989).
Derivation of the original ITC equations can be found in the original ITC paper (Wiseman et al., 1989).

For ITC experiments of this study, tRNA anticodon stem-loop at ~0.35 mM was used as titrate in the syringe, and T box specifier domain about 0.03 mM was used as titrant in the cell. The experimental temperature was 10°C, and a total of 30 injections with 5 minutes between each injection were performed for each ITC experiment. Buffer matching between the titrate and titrant dramatically affects the outcome of an ITC experiment and is something that is easily neglected because it is not involved in the experimental calculations or parameters. The two RNA samples are annealed (heated to 95°C and then snap cooled to 4°C) and dialyzed against the same ITC buffer (usually low salt NMR buffer with 2-20 mM Mg²⁺) at least twice by using Amicon Centricon or Ultra.

### 2.5 UV thermal melting

UV thermal melting curves of different RNA molecules are different because of different levels of base stacking and thermal stability. The increase in UV absorption when DNA or RNA “melts” from double-stranded to single-stranded is called the hyperchromic effect, which can be explained using the exitonic coupling theory (D’Abramo et al., 2013). UV melting studies were performed using about 0.5 OD_{260} unit RNA sample dissolved in buffer low salt NMR buffer. The samples were heated to 90°C for 60 s and snap-cooled on ice before each melting experiment. A_{260} absorbance spectra from 10 to 92 °C were recorded (1.0 °C/min) on a Jasco J-815 circular dichroism spectrometer equipped with a Peltier heating apparatus. The melting curves were acquired in triplicate and averaged. The melting temperature (T_m) of the RNA molecules were the temperature interpolated at 50% of the maximum absorbance. The hyperchromicity is determined by the increased percentage of absorption from the minimum to the maximum. The hyperchromicity of double-stranded DNA/RNA denaturation may be up to 30%.
2.6 Denaturing and native analytical polyacrylamide gel electrophoresis

Two types of analytical PAGE – denaturing and native – are generally used. Denaturing PAGE is prepared with 6 M urea and is used to check the RNA sample integrity, size, and amount. On a denaturing PAGE, the running speed of an RNA sample is solely dependent on the size of the molecule. Native PAGE is prepared without urea but with other reagents required for proper RNA folding. In the case of T box, magnesium is needed. The denaturing PAGE usually runs with 800 V at room temperature, and native PAGE runs with 50 to 100 V at 4°C. Native PAGE is used in this study to observe binding between the T box specifier domain and tRNA anticodon stem-loop. Band shifts and intensity changes relative to the free control are expected to happen when the complex is formed. The magnesium concentration in the gels varies from 2 mM to 20 mM. The concentration of RNA in a loading of 10 μL 1:1 complex is approximately 0.05 mM for each molecule. Visualization of the RNA bands on the gel can be done by UV shadowing or Stains-All staining, which has higher sensitivity.
Chapter 3. NMR Structures of tRNA$^{\text{Gly}}$ anticodon stem-loops

Glycyl-tRNA (tRNA$^{\text{Gly}}$) is one of the tRNA families with multiple extra-translational roles. In many Gram positive bacteria, expression of the gene encoding glycyl-tRNA synthetase is regulated by the T box riboswitch mechanism and tRNA$^{\text{Gly}}$ with the anticodon GCC (tRNA$^{\text{Gly},\text{GCC}}$) serving as a sensor molecule. This transcription attenuation mechanism is sensitive to the ratio of charged tRNA to uncharged tRNA in the cell. The 5’ untranslated region of the mRNA sequence-specifically binds tRNA molecules and forms one of two alternative hairpin secondary structures, the terminator hairpin and the antiterminator hairpin, depending upon the charge state of the bound tRNA. T box riboswitch selection of the appropriate tRNA molecule for regulation of the downstream gene occurs through pairing of the Specifier codon nucleotides with the nucleotides of the tRNA anticodon. Additional base pairing involving the universally conserved U33 of tRNA also has been proposed (Yousef et al., 2005). In many bacteria, including Bacilli and Staphylococci, the glycyl T box riboswitch is specified by the codon 5’-GGC-3’ which is complementary to the tRNA-$^{\text{Gly}}$ isoacceptor tRNA$^{\text{Gly},\text{GCC}}$, but the Specifier codon 5’-GGA-3’ also is represented and would be predicted to bind the isoacceptor tRNA$^{\text{Gly},\text{UCC}}$ (Vitreschak et al., 2008). Binding of the non-cognate glycyl-tRNA isoacceptor tRNA$^{\text{Gly},\text{UCC}}$ to the 5’-GGC-3’ Specifier sequence cannot be excluded and may contribute to regulation of the glyQS operon.

Peptidoglycan cell wall biosynthesis in bacteria involves a non-ribosomal peptidyltransferase mechanism that utilizes aminoacylated tRNA molecules as substrates for the peptide polymerization reaction. The peptidyltransferase enzymes FemABX catalyze the formation of short homopolymers that cross-link the glycan moieties and increase rigidity of the cell wall (Benson et al., 2002; Berger-Bächi, et al., 1998). A glycyl-tRNA, first identified and sequenced in Staphylococci, participates in cell wall synthesis but is not involved in ribosome-catalyzed protein synthesis and is designated a non-proteinogenic glycyl-tRNA (np-tRNA$^{\text{Gly}}$) (Bumsted et al., 1968; Roberts et al.,...
1974; Giannouli et al., 2009). These np-tRNA\textsuperscript{Gly} molecules bear the anticodon sequence 5’-UCC-3’ and are charged by glycyl-tRNA synthetase. The np-tRNA\textsuperscript{Gly} has a cytidine at position 37 rather than the purine found in proteinogenic glycyl-tRNA molecules, and the U34 base is not modified as it is in the proteinogenic tRNA\textsuperscript{Gly,UCC} of many bacteria (Sprinzl et al., 1998). The \textit{Staphylococcal} np-tRNA\textsuperscript{Gly} molecules contain A49-U65 and A51-U63 base pairs at the base of the T-arm rather than the G49-U65 and G51-C63 base pairs found in most proteinogenic tRNA molecules. In \textit{Thermus thermophilus}, substitution of the A51-U63 base pair for G51-C63 results in the loss of a direct contact between E390 of elongation factor thermal unstable (EF-Tu) and the amino group of G51, substantially weakening the affinity for EF-Tu (Nissen et al., 1995; Nissen et al., 1999; Sanderson and Uhlenbeck, 2007). This weakened affinity for EF-Tu limits participation of np-tRNA\textsuperscript{Gly} in ribosomal protein synthesis and presumably ensures a stable pool of proteinogenic tRNA\textsuperscript{Gly} for translation during cell wall biosynthesis (Giannouli et al., 2009; Sanderson and Uhlenbeck, 2007).

Glycine is a member of a four-codon box family, a set of four codons that designate the same amino acid and whose first two nucleotides are the same. In bacteria, these boxes are read by up to three different tRNA isoacceptors. Species of \textit{Bacillus} and \textit{Staphylococcus} use two tRNA\textsuperscript{Gly} isoacceptors, with the anticodon sequences 5’-U*CC-3’ and 5’-GCC-3’ where U* is a modified uridine. Modifications on U34 can lead to opposite functional effects, enhancement of the ability of U to wobble or restriction of wobbling and enhancement of discrimination (Agris, 2008). In the case of lysine, which occupies a mixed-codon box, U34 is modified to 5-methylaminomethyl-2-thiouridine (mmn^s^2U) and pairing is restricted to A and G. The U34 modification uridine 5-oxo-acetic acid (cmo\textsuperscript{5}U) allows a single tRNA isoacceptor to decode at least three valine codons in bacteria (Sprintz et al., 1998; Agris, 2004). However, modification of U34 is not always needed for enhanced wobbling. In \textit{Mycoplasma mycoides} and in mitochondria and chloroplasts, one tRNA isoacceptor with anticodon sequence 5’-UCC-3’ reads all four glycine codons with equal efficiency (Samuelsson et al., 1983; Claesson et al., 1990; Claesson et al., 1995). Notably, 5’-GGA-3’ and 5’-GGU-3’ combined represent ~95% of glycine codons used in \textit{M. mycoides} whereas these codons are used ~75% of the time in \textit{S. aureus} or \textit{B. subtilis}. 
tRNA molecules interact with a variety of proteins and other RNA molecules in the cell to fulfill a multitude of functional roles. The glycyl-tRNA isoacceptors selected for this study participate in translation, transcription regulation, and cell wall biosynthesis. These selected glycyl-tRNA isoacceptors also allow examination of the possible structural influences of base type at positions 34 and 37. Structures of the anticodon arms of the three glycyl-tRNA molecules, tRNA\text{Gly,GCC}, tRNA\text{Gly,UCC}, and np-tRNA\text{Gly,UCC}, were determined in this study. Although the participation of the np-tRNA\text{Gly,UCC} in translation is likely to be restricted by low affinity for EF-Tu, the pyrimidine-37 residue of np-tRNA\text{Gly,UCC} and other np-tRNAs may limit contributions of these molecules to other anticodon-dependent processes. The structures of the tRNA\text{Gly} anticodon arms differ from one another, but none of them form the classical U-turn motif seen in some tRNA anticodon arms. All of these RNA molecules form stems with at least five base pairs. The anticodon loop of tRNA\text{Gly,GCC} becomes more dynamic and disordered in the presence of multivalent cations, whereas the anticodon loops of tRNA\text{Gly,UCC} and np-tRNA\text{Gly,UCC} become more structurally ordered by these ions. Although the U-turn is integral to ribosomal codon-anticodon pairing, it is not known if this motif is required for T box regulation or cell wall biosynthesis. A more dynamic loop structure may better accommodate the different functional roles of tRNA\text{Gly}.

3.1 Design of the hairpin molecules

Three tRNA anticodon stem-loops (ASL), namely ASL\text{Gly,GCC}, np-ASL\text{Gly,UCC}, and ASL\text{Gly,UCC} have been designed and used in this study, and the complete sequences used are as shown in Figure 3.1. For effective \textit{in vitro} RNA synthesis, some changes have been made on the sequences while trying to match the sequences to the original corresponding tRNAs as much as possible.

The sequence of ASL\text{Gly,GCC} is based on residues C27-G43 of the tRNA\text{Gly,GCC} of \textit{Bacillus subtilis} from tRNAdb (Jühling \textit{et al.}, 2009). The first two G-C base-pairs were flipped, and the A-U base-pair was changed to a G-U base-pair, making the ASL\text{Gly,GCC} sequence starting with three repetitive guanosines.
Figure 3.1  Sequences corresponding to the anticodon arms of *Staphylococcus aureus* (A) tRNA\textsubscript{Gly,GCC}, (B) tRNA\textsubscript{Gly,UCC}, and (C) np-tRNA\textsubscript{Gly,UCC}.
Nucleotide numbering corresponds to the full-length tRNA molecule. Residues 27–30 and 40–43 were changed to permit transcription using T7 RNA polymerase and to facilitate comparison of structural and thermodynamic effects of the loop sequences. Also shown are the chemical structures of modifications cmnm5U, 5-carboxymethylaminomethyl uridine, cmo5U, uridine 5-oxyacetic acid, and m6A, (N6-methylallyl) adenine.

Figure 3.2  UV melting curves of ASL\textsuperscript{Gly,GCC} (red), np-ASL\textsuperscript{Gly,UCC} (blue), and np-ASL\textsuperscript{Gly,UCC} (green).
The ASL\textsuperscript{Gly} molecules exhibit a single melting transition that occurs above 55 °C, suggesting minimal stacking of the unpaired loop nucleotide bases. The melting transitions generally agree with the predicted secondary structures of the molecules, with ASL\textsuperscript{Gly,UCC} displaying the lowest melting transition and having the fewest Watson–Crick base pairs. The addition of Mg\textsuperscript{2+} to ASL\textsuperscript{Gly,UCC} (light blue) increases the T\textsubscript{m} by ~10 °C but causes an ~1.5 °C decrease when added to np-ASL\textsuperscript{Gly,UCC} (light green). In the absence of Mg\textsuperscript{2+}, the apparent T\textsubscript{m} values of ASL\textsuperscript{Gly,GCC}, ASL\textsuperscript{Gly,UCC}, and np-ASL\textsuperscript{Gly,UCC} are 72.0, 60.4, and 73.1 °C, respectively.
The np-ASL<sub>Gly,UCC</sub> was based on residues C27-G43 of the tRNA<sub>Gly</sub>IB from *Staphylococcus epidermidis* (Roberts *et al.*., 1974). The first G-C base-pair was flipped, the second base-pair was changed to a G-U base-pair, and the third A-U base-pair was changed to a G-C base-pair, making the sequence start with three repetitive Gs. The ASL<sub>Gly,UCC</sub> was based on residues U27-A43 of the tRNA<sub>Gly</sub>(UCC) of *B. subtilis* from tRNAdb 2009 (Jühling *et al.*., 2009), and the stem region was modified to match with the stem of np-ASL<sub>Gly,UCC</sub>.

### 3.2 Thermal Stability

The thermal stability of ASL<sub>Gly,GCC</sub>, ASL<sub>Gly,UCC</sub>, and np-ASL<sub>Gly,UCC</sub> (Figure 3.2) were investigated using UV melting experiments to determine overall molecular stability (Tm). The normalized UV thermal denaturation curves indicate that the ASL<sub>Gly</sub> molecules melt in one step (Figure 3.2). Two-stage melting transitions of other anticodon arms (tyrosine and phenylalanine) have been observed with the lower temperature (<50 °C) transition corresponding to the destacking of the loop nucleotides (Cabello-Villegas *et al.*., 2004; Denmon *et al.*, 2011). In these two systems, the anticodon contains an adenine nucleotide which may facilitate stacking of the unpaired loop nucleotides. Two of the glycine molecules (ASL<sub>Gly,GCC</sub> and np-ASL<sub>Gly,UCC</sub>) display similar T<sub>m</sub> values around 70 °C, and the T<sub>m</sub> of ASL<sub>Gly,UCC</sub> is 10 °C lower. Both ASL<sub>Gly,GCC</sub> and np-ASL<sub>Gly,UCC</sub> form 6-7 Watson-Crick base pair stems, whereas ASL<sub>Gly,UCC</sub> forms five regular base pairs and a protonated C-A<sup>+</sup> base pair at pH 6.3 (see below). Additionally, the degree of base stacking differs among the molecules. The hyperchromicity associated with melting is least for np-ASL<sub>Gly,UCC</sub> (9%) and greatest for ASL<sub>Gly,GCC</sub> and ASL<sub>Gly,UCC</sub> (16% and 14%, respectively). Mg<sup>2+</sup> raises the T<sub>m</sub> of ASL<sub>Gly,UCC</sub> to the level of np-ASL<sub>Gly,UCC</sub> and has only a slight (~1 °C) destabilizing effect on np-ASL<sub>Gly,UCC</sub> (Figure 3.2).

### 3.3 Resonance Assignments
Due to self-complementarity, the RNA sequences used in this study can adopt either hairpin (monomer) form or a duplex form with internal loops of different sizes. Therefore, the oligomeric states of the RNA molecules were assessed using the NH spectra of each of the molecules. The hairpin forms have moderate line widths (9-14 Hz $^1$H) and NH peak patterns that are independent of RNA concentration. ASL$_{\text{Gly,UCC}}$ and np-ASL$_{\text{Gly,UCC}}$ yield a single set of NH resonances consistent with the hairpin helix and give rise to single bands on native PAGE gels. The NH spectrum of ASL$_{\text{Gly,GCC}}$ also yields a single set of peaks immediately after annealing. However, additional peaks appear in the spectrum after several hours. The hairpin-duplex equilibrium was confirmed using a NOE-based $^{15}$N-filtered spectrum. At low RNA concentration (<0.4 mM) and <5 mM NaCl, the duplex constitutes <5% of the RNA population. Native PAGE analysis also shows that ASL$_{\text{Gly,GCC}}$ forms a mixture of monomer and dimer species.

Under conditions of low salt and RNA concentration <0.4 mM, ASL$_{\text{Gly,GCC}}$ yields good quality spectra with no evidence of duplex. However, the base $^1$H and $^{13}$C resonances of ASL$_{\text{Gly,UCC}}$ and np-ASL$_{\text{Gly,UCC}}$ exhibit limited dispersion with a few resonances broadened by chemical exchange (Figure 3.3). To improve spectral quality, Mg$^{2+}$ and Co(NH$_3$)$_6$$^{3+}$ were tested for their ability to bind the RNA molecules and promote conformational homogeneity. For np-ASL$_{\text{Gly,UCC}}$, 3.0 mM Mg$^{2+}$ yields quality spectra with improved resolution. For ASL$_{\text{Gly,UCC}}$, 2.0 mM Co(NH$_3$)$_6$$^{3+}$ was found to improve spectral dispersion and reduce exchange broadening, whereas Mg$^{2+}$ was less effective and resulted in general resonance broadening. Therefore, the solution NMR studies for ASL$_{\text{Gly,UCC}}$ and np-ASL$_{\text{Gly,UCC}}$ were performed in the presence of Co(NH$_3$)$_6$$^{3+}$ and Mg$^{2+}$, respectively.

The non-exchangeable $^1$H and $^{13}$C resonances of the three ASL$_{\text{Gly}}$ molecules (Figure 3.1) were assigned using standard heteronuclear techniques (Wang and Nikonowicz, 2011; Denmon et al., 2011). Most of the base and ribose $^1$H and $^{13}$C correlations are resolved for each of the molecules (Figure 3.3). In the absence of metal ion, the A38 C2 resonance of ASL$_{\text{Gly,UCC}}$ is slightly broadened and has a chemical shift of 148.6 ppm indicative of N1 protonation (Figure 3.4). With the addition of Co(NH$_3$)$_6$$^{3+}$, the A38 base deprotonates and the C2 resonance shifts downfield to
Figure 3.3 Comparison of two-dimensional $^{13}$C−$^1$H HSQC spectra of the base C6/8 regions of ASL$^{\text{Gly,UCC}}$ in the (A) absence and (B) presence of Co(NH$_3$)$_6^{3+}$ and of np-ASL$^{\text{Gly,UCC}}$ in the (C) absence and (D) presence of Mg$^{2+}$.

Addition of Co(NH$_3$)$_6^{3+}$ leads to substantial sharpening and upfield movement of the A38 cross-peak of ASL$^{\text{Gly,UCC}}$.$^{3+}$

Figure 3.4 The effect of the addition of Co(NH$_3$)$_6^{3+}$ on the protonation state of ASL$^{\text{Gly,UCC}}$ A38.

The adenine C2 resonances in the absence (gray) and presence (black) of Co(NH$_3$)$_6^{3+}$. Addition of Co(NH$_3$)$_6^{3+}$ causes the A38 C2 resonance to shift downfield.
Figure 3.5 Sequential assignments of the H8/6-H1' resonances of the ASL$^{Gly}$s: (a) ASL$^{Gly,GCC}$, (b) ASL$^{Gly,UCC}$, and np-ASL$^{Gly,UCC}$. The sequential assignments are continuous in all ASL$^{Gly}$s with ASL$^{Gly,GCC}$ having the most resolved spectrum among the three. Connections are showed in arrows and H1’-H8/H6 intra-residual cross-peaks are marked blue.
153.2 ppm. The ribose spin systems, except for the incompletely labeled 5’-terminal nucleotides, were identified using 2D or 3D HCCH-COSY and HCCH-TOCSY experiments. For ASL<sub>Gly,GCC</sub> and np-ASL<sub>Gly,UC</sub>, spectral overlap even in the 3D spectrum limited unambiguous assignment of some 4’ and 5’ resonances. Intra-residue base-to-sugar correlations were identified using 2D H(C)N experiments to obtain correlations H6-N1, H8-N9, and H1’-N1/N9. All pyrimidine correlations and all purine (ASL<sub>Gly,GCC</sub> and np-ASL<sub>Gly,UC</sub>) or four of six purine (ASL<sub>Gly,UC</sub>) correlations were identified in these spectra. The G28 and A38 correlations of ASL<sub>Gly,UCC</sub> are not observed due to chemical exchange.

Sequential assignments for the non-exchangeable resonances were made using 2D NOESY and 3D <sup>13</sup>C-edited NOESY experiments to identify sequential H6/8-H1’ NOE connectivities. The sequential base-1’ NOE connectivities are continuous through all 17 nucleotides at long mixing times (≥400 ms). The connectivity is continuous in the 180 ms NOESY spectrum of ASL<sub>Gly,GCC</sub> (Figure 3.5), but at this mixing time the inter-residue NOE is weak between nucleotides U33 and U34 and broken between nucleotides U34 and C35 in spectra of ASL<sub>Gly,UC</sub> and np-ASL<sub>Gly,UC</sub>.
Table 3.1 Summary of chemical shifts of ASLGly.

Chemical shift information of (a) ASL Gly,GGC, (b) ASL Gly,UCC, and (c) np-ASL Gly,UCC are listed. N/A=not available; N/O=not observed.
In addition, while peak overlap near the diagonal prevented detection of many potential H6-H6 inter-residue NOEs, sequential H5-H6 crosspeaks were identified that support stacking of several pyrimidine bases in the loop regions of ASL
^{Gly, UCC} and np-ASL
^{Gly, UCC}.

The NH and NH$_2$ resonances were assigned using $^1$H-$^1$H NOESY and HNCCH experiments. For all molecules, the NH resonances of the first five neighboring base pairs yield NOE connectivities between each other. For ASL
^{Gly, GCC} and np-ASL
^{Gly, UCC}, the NH resonance connectivities extend to a sixth base pair. The NH spectrum of ASL
^{Gly, GCC} contains a broad guanine NH resonance at 10.56 ppm that was assigned to G34. The U33 NH resonance at 12.95 ppm of ASL
^{Gly, GCC} is broad and very weak as well. The U33 and U34 NH resonances of ASL
^{Gly, UCC} and np-ASL
^{Gly, UCC} also are broad and only observed in the 1D spectra between 10.5 and 11.5 ppm. The cytidine and adenine NH$_2$ resonances were assigned via scalar correlations using HSQC and HNCCH experiments. The upfield chemical shifts (7.20 and 6.69 ppm) of the C32 H4 resonances of ASL
^{Gly, UCC} are consistent with the lack of intra-molecular hydrogen bonding (Figure 3.6). In all molecules, the C35 and C36 NH$_2$ $^1$H resonance pairs and the NH$_2$$^{15}$N resonances are shifted upfield by ~1 ppm ($^1$H) and ~2 ppm ($^{15}$N) relative to those of the base paired cytidine residues (Figure 3.6). The C37 NH$_2$ resonances of np-ASL
^{Gly, UCC} exhibit similar upfield shifts.

The inter-nucleotide phosphate $^{31}$P resonances are clustered between -3.54 and -4.60 ppm for all molecules and partial assignments were obtained using HCP or $^{31}$P-$^1$H hetero-TOCSY-NOESY spectra (O’Neil-Cabello et al., 2004). The P–H3’ correlations and several P–H4’ and P–H5’/H5’’ correlations are present in $^{31}$P-$^1$H HetCor spectra and provide independent confirmation of the $^{31}$P assignments. Notably, the $^{31}$P resonances of ASL
^{Gly, UCC} and np-ASL
^{Gly, UCC} remain tightly grouped indicating that the metal ions have little effect on the phosphate backbone conformation (Figure 3) and point to weak metal ion coordination to the phosphate backbone. A complete list of resonance assignments, including non-protonated positions, is given in Table 3.1.
3.4 Effects of the metal ions

In high-resolution structure studies of tRNA, Mg\(^{2+}\) and Co(NH\(_3\))\(_6\)\(^{3+}\) ions have been observed proximal to the loop-helix junction in the anticodon stem-loop of tRNA. As noted above, the multivalent ions Co(NH\(_3\))\(_6\)\(^{3+}\) and Mg\(^{2+}\) were needed to obtain high-quality spectra for ASL\(^{\text{Gly,UCC}}\) and np-ASL\(^{\text{Gly,UCC}}\), respectively. In the absence of multivalent ions, the A38 C2 resonance at 148.6 ppm marks the presence of a C32-A\(^{38}\) base pair in ASL\(^{\text{Gly,UCC}}\) (Figure 3.3). Both Co(NH\(_3\))\(_6\)\(^{3+}\) and Mg\(^{2+}\) lead to loss of this base pair and deprotonation of A38. Mg\(^{2+}\) reduces the conformational variability of the anticodon nucleotides at 10 mM concentration, but causes broadening of resonances throughout the spectrum and substantial spectral overlap in the NOESY. Co(NH\(_3\))\(_6\)\(^{3+}\) (2 mM) also reduces the conformational variability of the loop nucleotides, but does not lead to excessive line broadening. The most significant chemical shift change caused by Co(NH\(_3\))\(_6\)\(^{3+}\) is the N1 base resonance of G28 (~0.3 ppm). The remaining chemical shift changes involve loop nucleotide resonances and are minor (<0.1 ppm). The resonances of np-ASL\(^{\text{Gly,UCC}}\) exhibit a similar pattern of chemical shift changes with 3 mM Mg\(^{2+}\). The addition of Co(NH\(_3\))\(_6\)\(^{3+}\) to np-ASL\(^{\text{Gly,UCC}}\) does not produce additional resonance changes, but was used to localize ion binding. The 90% \(^1\text{H}_2\text{O}\) NOESY spectrum supports non-specific association of Co(NH\(_3\))\(_6\)\(^{3+}\) in the loop region. In the 90% \(^1\text{H}_2\text{O}\) NOESY spectra, NOEs between the guanine and uridine NH and cytidine NH\(_2\) protons of the stem with the Co(NH\(_3\))\(_6\)\(^{3+}\) protons confirm the expected coordination of Co(NH\(_3\))\(_6\)\(^{3+}\) at the G28-U42 base pair (Cabello-Villegas et al., 2002). Although the NOE spectral data defining the stem location of Co(NH\(_3\))\(_6\)\(^{3+}\) are very good, the cobalt hexamine complex can only be restricted to the major groove side of the loop. Very weak NOE cross peaks are observed between Co(NH\(_3\))\(_6\)\(^{3+}\) and the base protons of residues (Wang et al., 2010; Wang and Nikonowicz, 2011; Schwieters et al., 2003; Cabello-Villegas et al., 2002). However, there is evidence for coordination involving the U33 O4 carbonyl atom of the base. The chemical shifts of the U33 C2 and C4 nuclei (152.2 and 168.2 ppm, respectively) are consistent with participation of the O4 atom in a stable hydrogen bond or in metal ion coordination. The addition of Co(NH\(_3\))\(_6\)\(^{3+}\) supports metal ion coordination...
Table 3.2 Summary of experimental distance and dihedral angle constraints and refinement statistics for ASL<sub>Gly</sub> molecules

<table>
<thead>
<tr>
<th>Constraint</th>
<th>ASL&lt;sub&gt;Gly,GCC&lt;/sub&gt;</th>
<th>ASL&lt;sub&gt;Gly,UCC&lt;/sub&gt;</th>
<th>np-ASL&lt;sub&gt;Gly,UCC&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NOE distance constraints</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intraresidue&lt;sup&gt;a&lt;/sup&gt;</td>
<td>116</td>
<td>106</td>
<td>110</td>
</tr>
<tr>
<td>interresidue</td>
<td>94</td>
<td>81</td>
<td>95</td>
</tr>
<tr>
<td>mean number per residue</td>
<td>12.4</td>
<td>11</td>
<td>12.1</td>
</tr>
<tr>
<td><strong>NOE constraints by category</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>strong (1.8 - 3.0 Å)</td>
<td>32</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>medium (1.8 - 4.5 Å)</td>
<td>108</td>
<td>80</td>
<td>89</td>
</tr>
<tr>
<td>weak (1.8 - 6.0 Å)</td>
<td>56</td>
<td>82</td>
<td>84</td>
</tr>
<tr>
<td>very weak (1.8 - 7.0 Å)</td>
<td>14</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td><strong>base pair constraints</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>32</td>
<td>26</td>
<td>32</td>
</tr>
<tr>
<td><strong>dihedral angle constraints</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ribose ring&lt;sup&gt;b&lt;/sup&gt;</td>
<td>48</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>backbone</td>
<td>71</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>H bonds</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>mean number per residue</td>
<td>7.7</td>
<td>9.1</td>
<td>9.2</td>
</tr>
<tr>
<td><strong>Violations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>average distance constraints &gt; 0.6 Å&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>average dihedral constraints &gt; 0.5°&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>RMSD from ideal geometry&lt;sup&gt;e&lt;/sup&gt;</strong></td>
<td>0.39 ± 0.17</td>
<td>0.83 ± 0.27</td>
<td>0.75 ± 0.9</td>
</tr>
</tbody>
</table>

<sup>a</sup> Only conformationally restrictive constraints are included.
<sup>b</sup> Three torsion angles within each ribose ring were used to constrain the ring to either the C2’-endo or C3’-endo conformation.
<sup>c</sup> A distance violation of 0.6 Å corresponds to 5.0 kcal energy penalty.
<sup>d</sup> A dihedral angle violation of 0.5° corresponds to 0.05 kcal energy penalty.
<sup>e</sup> Calculated against the minimized average structure.
at the U33 O4 by causing additional (~0.2 ppm) upfield and downfield shifts of the C2 and C4 resonances. Coordination of the Mg$^{2+}$ and Co(NH$_3$)$_6$$_{3+}$ ions may also involve bridging between phosphate groups across the major groove; however, the $^{31}$P spectra of both molecules are minimally altered by Co(NH$_3$)$_6$$_{3+}$ relative to Mg$^{2+}$.

For the ASL$^{Gly,GCC}$ RNA, intermediate concentrations of Mg$^{2+}$ and Co(NH$_3$)$_6$$_{3+}$ ions cause modest exchange broadening of the anticodon nucleotide base resonances. The U33 NH resonance is exchange broadened and not observed, and the position of the U39 resonance is shifted upfield ~0.3 ppm. No other resonances in this region exhibit substantial chemical shift changes. The A38 H2 is shifted upfield 0.8 ppm and the A37, U33, and C36 base 6 and 8 resonances are broadened by chemical exchange. Also, no intermolecular NOEs between Co(NH$_3$)$_6$$_{3+}$ protons and loop nucleotide protons could be identified. Thus, although Mg$^{2+}$ and Co(NH$_3$)$_6$$_{3+}$ associate with the loop region of ASL$^{Gly,GCC}$, the interaction appears weaker than the interaction of these ions with either ASL$^{Gly,UCC}$ or np-ASL$^{Gly,UCC}$.

![Figure 3.7 Superposition of eight converged structures of (A) ASL$^{Gly,GCC}$, (B) ASL$^{Gly,UCC}$, and (C) np-ASL$^{Gly,UCC}$ RNA hairpins.](image)

Views are into the major grooves of the anticodon loops. The rmsds between the individual structures and the average structure are listed in Table 3.2. The greatest variability occurs among the anticodon bases and reflects the smaller number of constraints for these residues.
Figure 3.8 Stereoviews of the loop regions of (A) ASL$^{\text{Gly,GCC}}$, (B) ASL$^{\text{Gly,UCC}}$, and (C) np-ASL$^{\text{Gly,UCC}}$ RNA hairpins.

The color scheme is as follows: red for anticodon bases N34, C35, and C36, blue for N31 and N39, green for N32 and N38, pink for U33, and pink/brown for N37.
3.5 NMR structures of tRNA\textsuperscript{Gly} anticodon stem-loops

The structures of the ASL\textsuperscript{Gly} molecules were calculated using a restrained molecular dynamics routine starting from 50 sets of coordinates with randomized backbone dihedral angles. The calculations used a total of 188-211 conformationally restrictive distance constraints and 131-156 dihedral angle constraints (Table 3.2) to produce 8 converged structures for each molecule (Figure 3.7). Structures were classified as converged if they were consistent with the NMR data and maintained correct stereochemistry. All converged structures have no constraints violated by more than 0.1 Å. When the structures are arranged in order of increasing overall energy, the converged structures form a plateau with similarly low overall energies and constraint violation energies. The root mean square deviations (RMSDs) of the heavy atoms between the individual structures and the minimized mean structures are 0.39 Å, 0.83 Å, and 0.75 Å for, ASL\textsuperscript{Gly,GCC}, ASL\textsuperscript{Gly,UCC}, and np-ASL\textsuperscript{Gly,UCC}, respectively. The global fold of ASL\textsuperscript{Gly,GCC} is a hairpin composed of a seven base pair stem and a three nucleotide loop (Figure 3.7). The overall fold of np-ASL\textsuperscript{Gly,UCC} is a six base pair stem with a five nucleotide loop (Figure 3.8). The overall fold of ASL\textsuperscript{Gly,UCC} (a five base pair stem with a five to seven nucleotide loop) is similar to np-ASL\textsuperscript{Gly,UCC}, but the loss of some inter-residue NOEs among loop nucleotides and the loss of secondary structure proximal to the loop caused by the absence of the 32-38 base pair results in a somewhat less precisely defined loop conformation (Figure 3.7). The minimized average structures are shown in Figure 3.8.

For all of the molecules, the helical base stack along the 5’ side of the loop is continuous though residue 34 and is conserved among the converged sets of structures. Base stacking along the 3’ side of the loop varies among the structures. For ASL\textsuperscript{Gly,GCC}, the 3’ strand stacking begins with A37. For ASL\textsuperscript{Gly,UCC} and np-ASL\textsuperscript{Gly,UCC}, the 3’ strand stacking generally begins with C35, but individual structures show moderate deviations (Figures 3.7). The C35 and C36 nucleotides are common to each of the three families of structures, and they distribute along the major groove or minor groove sides of the loops depending upon the family. The general positions of these residues within individual structures of each family, though, are uniform and exhibit few excursions to the
opposite side.

### 3.5.1 Structure of the loop regions of ASL<sub>Gly</sub> molecules

The loop of ASL<sub>Gly,GCC</sub> is composed of nucleotides G34–C36 and is closed by a distorted U33–A37 base pair (Figure 3.8). The position of the A37 base is restricted by NOE cross peaks (A37 H2-G34 H1’ and A37 H2-A38 H1’) that are characteristic of adenine nucleotides in helices. The hydrogen bond functional groups between U33 and A37 are aligned, but the bases exhibit moderate buckling (25° average among converged structures). The intensity of the U33 NH resonance is weak and could result from solvent exchange. However, the U33 C2 and C4 chemical

![Figure 3.9 NOE cross peaks imply ASL<sub>Gly,GCC</sub> form the 6th and 7th base pairs.](image)

Above is the alignment of two spectra of ASL<sub>Gly,GCC</sub> with H<sub>2</sub>O <sup>1</sup>H-<sup>1</sup>H NOESY on the top and the imino <sup>15</sup>N-<sup>1</sup>H HMQC on the bottom. The U32 and U33 imino H3s have chemical shifts of 13.68 and 13.94, respectively, and the chemical shifts are within the range of the Watson-Crick base-paired uridine H3s. Their respective NOE cross peaks to the H2s of A37 (7.28 ppm) and A38 (7.54 ppm) in the H<sub>2</sub>O <sup>1</sup>H-<sup>1</sup>H NOESY also confirm the formation of the hydrogen bonds. The upfield shifted U41 H3 (12.27 ppm) is consistent with the formation of a G-U base pair.
shifts (152.8 and 168.6 ppm, respectively) and A37 N1 chemical shift (222.1 ppm) are indicative of weak hydrogen bonding and are consistent with the distorted U33–A37 pair. The chemical shifts of U33 and U34 imino H3 protons are in the range of the Watson-Crick base pairs, and the NOE cross peaks of A37 H2–U33 H3 and A38 H2–U34 H3 in the H2O 1H-1H NOESY also support the U33–A37 and U34–A38 base pairs (Figure 3.9). The G34 base extends across the helix axis with the U33 and C35 bases stacking above and beneath, respectively, the major groove edge of the G34 base. The G34-C35 stacking is supported by H8-H6 and H8-H5 NOE cross peaks. The C35 base and its NH2 group point away from the helix axis toward the solvent. The C36 base also resides on the major groove side of the loop with its NH2 group extending toward the solvent, but is not coplanar with the C35 base. The 15N and 1H chemical shifts of the C35 and C36 NH2 groups are shifted upfield (Figure 3.6) and are consistent with the absence of intra-molecular interactions. The C36 base is vertically displaced from the A37 base and does stack.

The loop of ASL Gly,UCC is composed of seven nucleotides, residues C32-A38. Although cobalt hexamine leads to deprotonation of the A38 base, the C32 and A38 bases maintain an alignment characteristic of the C32-A+38 base pair. The U33 and A37 bases are coplanar and stack beneath C32 and A38, respectively. The A37 base is rotated toward the minor groove causing the N6 amino group to align for hydrogen bonding with the U33 O2 atom. However, the distance between the heavy atoms is long (4.0 Å average) and the U33 C2 and A37 N6 chemical shifts, 153.8 ppm and 80.1 ppm respectively, do not reflect hydrogen bonding involving the O2 and H6 atoms. The base of U34 is on the major groove side of the loop and stacks beneath the U33 base. The bases of C35 and C36 lay on the minor groove side of the helix and stack on each other and beneath A37 with their Watson-Crick edges pointing away from the helix axis toward the solvent. The 1H and 15N resonances of the C35 and C36 NH2 groups are shifted upfield (5.7-6.8 ppm 1H and 93 ppm 15N) and are characteristic of solvent exposed cytidine amino groups (Figure 3.6).

The loop of np-ASL Gly,UCC is composed of residues U33-C37. The bases of U33 and C37 stack beneath C32 and G38, respectively, and neither base is laterally displaced towards the major or minor grooves. However, C37 is rotated so that the Watson-Crick edge of the base points into
the minor groove rather than toward the helix axis. The base of C36 is parallel and stacked with C37 in ~80% of the structures, but points down and away from the C37 base in the remaining structures. Both orientations are in agreement with the observed NOE cross peak pattern. Like C37, the Watson-Crick edge of C36 points out toward the minor groove side of the loop in all structures. The base of U34 stacks beneath U33 and is rotated out towards the major groove. C35 is positioned on the minor groove side of the loop at the apex of the phosphate backbone turn. The base of C35 is laterally displaced from and slightly below the base plane of C36. With the exception of U33, the functional groups along the Watson-Crick edges of the loop nucleotide bases are solvent exposed and none form intra-molecular interactions. The C35-C37 NH$_2$ chemical shifts (Table 3.1) reflect the solvent exposure of these bases.

The sugar-phosphate backbone conformations of the loop nucleotides are surprisingly uniform (Figure 3.8). The majority of the nucleotides maintain the C3’-endo ring pucker, but the C35 and C36 ribose groups have observable H1’–H2’ couplings and ribose $^{13}$C chemical shifts that indicate a mixture of C2’- or C3’-endo ring pucker conformations. In addition, the uniformly small (<5 Hz) P–C2’ coupling constants for the loop residues place the ε torsion angles in the trans conformation characteristic of RNA. The phosphate backbone torsion angles β, γ, and ε of residues 34-36 were loosely constrained for the calculations, but most have values within the range common to RNA helices, and the majority of deviations from standard values involve residue C34. The $^{31}$P chemical shifts for all inter-residue phosphorus atoms are tightly clustered between -3.5 and -4.6 ppm, indicating that the α and ζ torsion angles throughout the loop and helix adopt a gauche conformation (Pardi, 1995).

3.5.2 Structure of the ASL$^{\text{Gly}}$ stems

The conformations of the stems of the three ASL$^{\text{Gly}}$ molecules are very similar. The geometry of the hairpins from the terminal G27-C43 base pair to 32-38 interaction is primarily A-form (Figure 3.7). The C32-A*38 of ASL$^{\text{Gly,UCC}}$ is deprotonated with addition of cobalt hexamine, but the
relative arrangement of the two bases is largely unchanged as evidenced by intra- and inter-strand NOEs involving C32 and A38. The base-base and base-ribose NOE cross peaks among the stem nucleotides are continuous and their intensities are consistent with the A-form helical geometry. The torsion angles of the sugar-phosphate backbones involving the stem residues of all molecules also are within the limits of A-form geometry and are supported by chemical shift and J-coupling data.

3.6 Discussion

Although the bulk of tRNA in the cell is destined for aminoacylation and delivery to the ribosome for protein synthesis, a variety of alternative functional roles exist for tRNA (Dieckmann and Feigon, 1997). Besides their central role in translation, tRNA molecules participate in the addition of amino acids to membrane lipids and the N-termini of peptides (Peschel et al., 2001; Roy and Ibba, 2008), the biosynthesis of antibiotics and of the crosslinks in peptidoglycan cell walls (Bumsted et al., 1968; Roberts, 1974; Garg et al., 2006; Gondry et al., 2009), the regulation of transcription and translation (Henkin and Yanofsky, 2002), and viral replication (Barat et al., 1989; Colicelli and Goff, 1986). Not all tRNA molecules are multifunctional, but in some bacteria the glycyl-tRNA family has extra-ribosomal roles in transcriptional regulation and cell wall biosynthesis. Here we have examined the structural features of three glycyl-tRNA molecules that serve these three functions.

3.6.1 Comparison of the ASL\textsuperscript{Gly} structures

The structures of the tRNA\textsuperscript{Gly} anticodon arm stems (residues 27-32 and 38-43) are nearly the same, even though ASL\textsuperscript{Gly,UCC} has a C-A mismatch rather than the U-A and G-C pairs present in ASL\textsuperscript{Gly,GCC} and np-ASL\textsuperscript{Gly,UCC}, respectively. The A38 of ASL\textsuperscript{Gly,UCC} is deprotonated in the presence of Co(NH\textsubscript{3})\textsubscript{6}\textsuperscript{3+}, but the basic structural features of the C32-A38\textsuperscript{+} base pair – coplanarity of the two
Figure 3.10 Comparison of the U-turn motifs
Displayed are the structures from (A) the solution NMR structure of fully modified human tRNA\textsubscript{Lys,3} (PDB entry 1FL8), (B) the X-ray crystal structure of ribosome-bound E. coli tRNA\textsubscript{Lys,3} (PDB entry 1XMO), (C) the X-ray crystal structure of modified E. coli aa-tRNA\textsubscript{Cys} in complex with EF-Tu/GDPNP (PDB entry 1B23), and (D) the X-ray crystal structure of the hammerhead ribozyme (PDB entry 299D). The U-turn folds present in panels A–C are characterized by stacking of the anticyodon bases (purple) along the minor groove side of the helix and the NH–P hydrogen bond from U33 (green). In panel D, the GUAA sequence shows the U-turn fold characteristic of the GNRA tetraloops, where the G (green) forms the NH–P hydrogen bond. The first (n\textsubscript{i}) and third (n\textsubscript{i+2}) nucleotides of the turns are labeled. The 2’-OH–base hydrogen bond between n\textsubscript{i} and n\textsubscript{i+2} involves base N7, when n\textsubscript{i+2} is a purine and the exocyclic amino N\textsuperscript{4}H\textsubscript{2} when n\textsubscript{i+2} is cytidine. The n\textsubscript{i}–n\textsubscript{i+2} interaction has not been observed in ASLs when uridine occupies the n\textsubscript{i+2} position (A and B), although O4 in principle could perform this function. Modifications of residues 34 and 37 (A and B) provide additional hydrogen bonds in the turn.
bases and minimal lateral displacement from the helix axis – are conserved. In all of the molecules, the base of residue 34 stacks with U33, and is moderately displaced toward the major groove edge of U33. Also common to the three structures is the minor groove displacement of the base of residue 37. The most notable structural differences among the molecules are the orientations of nucleotides C35 and C36. In ASL^{Gly,GCC}, these nucleotides are positioned on the major groove side of the loop whereas in ASL^{Gly,UCC} and np-ASL^{Gly,UCC}, these residues are located on the minor groove side of the loop (Figure 3.8). The anticodon nucleotides of the three sequences also exhibit similar sugar pucker behavior. The riboses of nucleotides C35 and C36 appear to oscillate between C2’-endo and C3’-endo conformations as they give rise to cross peaks with modest (≈5 Hz) coupling in the DQF-COSY spectrum and intermediate 13C chemical shifts. To determine if the base positions are equally accommodated by the canonical ribose conformations, we performed a series of simulations where the ribose pucker conformations of C35 and C36 were fixed to C2’-endo or C3’-endo. In about 50% of the ASL^{Gly,GCC} structures, the C2’-endo pucker leads to a small downward rotation of the C36 base toward the helix axis and away from the major groove side of the loop. For ASL^{Gly,UCC} and np-ASL^{Gly,UCC}, the enforcement of C2’-endo pucker widens the curvature of the phosphate backbone through the loop with minimal effect on the positions of the bases in the loop. Neither the number of NOE violations nor the overall energies for the three molecules are changed significantly by applying the C2’-endo constraints. Thus, the average conformational states of the ribose puckers are consistent with the calculated sets of structures introduce a conformational malleability that may be needed to optimize anticodon loop nucleotide interactions in different contexts.

3.6.2 Comparison with structures of other anticodon stem-loop sequences

The crystal structure of fully modified yeast tRNA^{Phe} revealed what is now known to be a common RNA structural motif designated the U-turn (Kellogg and Schweitzer, 1993; Cabello-Villegas et al., 2004). This functionally important mo-
tif has since been observed in the structures of a host of other RNA molecules including the signal recognition particle (Shi and Moore, 2000) and RNA tetraloops (Kieft and Tinoco, 1997; Gorenstein, 1984) (Figure 3.10). In the context of the tRNA anticodon arm, the U-turn has been observed in crystal forms of unmodified \textit{E. coli} tRNA$^{\text{Phe}}$ (anticodon 5’-GAA-3’) (Byrne \textit{et al.}, 2010), an \textit{E. coli} tRNA$^{\text{Cys}}$–EF-Tu complex (anticodon 5’-GCA-3’) (Nissen \textit{et al.}, 1999), the ribosome-bound \textit{E. coli} tRNA$^{\text{Lys,3}}$ (anticodon 5’-UUU-3’) (Murphy \textit{et al.}, 2004), and the solution forms of partially modified \textit{E. coli} tRNA$^{\text{Phe}}$ and fully modified tRNA$^{\text{Lys,3}}$ (Cabello-Villegas \textit{et al.}, 2004; Durant \textit{et al.}, 2005) (Figure 3.10). The U-turn motif of the tRNA anticodon arm is characterized by a 120° turn of the phosphate backbone between residues 33 and 34 and stacking of the three anticodon nucleotide bases on the 3’ side of the loop. The U-turn of tRNA has been observed to contain hydrogen bonds between the U33 2’-OH and the N7 or N4H$_2$ of purine or cytidine residues at position 35 and between the U33 N$_3$H and the non-bridging phosphoryl oxygen 5’ to residue 36.

The U-turn structural motif that is ubiquitous among anticodon loops of tRNA molecules in crystal form is not adopted by any of the ASL$^{\text{Gly}}$ molecules in this study. In a classic U-turn, the non-sequential U33 H1’ to N35 H6/8 distance is ≈3.8 Å and should give rise to a moderately intense NOE cross peak. Although spectral overlap would occlude the non-sequential U33-C35 1’-base NOE cross peak in np-ASL$^{\text{Gly,UCC}}$, this peak is not present in the NOE spectra of ASL$^{\text{Gly,GCC}}$ or ASL$^{\text{Gly,UCC}}$. In addition, NOE cross peaks in the spectra of the ASL$^{\text{Gly,GCC}}$ and ASL$^{\text{Gly,UCC}}$ (including U33 H1’ to A38 H2 and G34 or U34 H1’ to A37 H2) are not compatible with the U-turn motif. The reversal of the phosphate backbone occurs smoothly between residues 34 and 36 (Figure 3.8) and does not turn abruptly between U33 and G34/U34 as observed for the classic U-turn (Schwieters \textit{et al.}, 2003; Kellogg and Schweitzer, 1993; Cabello-Villegas \textit{et al.}, 2004; Banerjee \textit{et al.}, 2010). In the U-turn, the trans conformation of the G34/U34 backbone angle $\alpha$ facilitates the sharp turn and is expected to cause the corresponding $^{31}$P resonance to shift downfield 2-3 ppm (Pardi, 1995, Schmitz \textit{et al.}, 1999). None of the phosphorus spectra of the ASL$^{\text{Gly}}$ molecules display this unusual feature and are consistent with the gauche angles adopted by the converged structures.
The absence of the U-turn fold from the tRNA anticodon loop in solution is not unique to the ASL\textsuperscript{Gly} molecules. The anticodon arms of unmodified and N\textsuperscript{6}-dimethylallyl-A37 (i\textsuperscript{6}A37)-modified tRNA\textsuperscript{Tyr}, Ψ39-modified and N\textsuperscript{6}-threonylcarbamoyl-A37 (t\textsuperscript{6}A37)-modified human tRNA\textsuperscript{Lys,3}, unmodified \textit{E. coli} tRNA\textsuperscript{Phe}, and unmodified \textit{E. coli} tRNA\textsuperscript{Val} also do not have the U-turn structure (Jucker an Pardi, 1995; Pley \textit{et al.}, 1994; Byrne \textit{et al.}, 2010). Interestingly, the addition of Mg\textsuperscript{2+} to unmodified \textit{E. coli} tRNA\textsuperscript{Val}, i\textsuperscript{6}A37-modified \textit{E. coli} ASL\textsuperscript{Phe}, and [5-methylcarboxymethyl, 2-thiouridine-34 (mcm\textsuperscript{5}s\textsuperscript{2}U34), t\textsuperscript{6}A37]-modified \textit{E. coli} ASL\textsuperscript{Lys} molecules leads to formation of the U-turn (Varani \textit{et al.}, 1996; Kim \textit{et al.}, 1974; Byrne \textit{et al.}, 2010). These data highlight crucial roles for base modification and Mg\textsuperscript{2+} in the formation of the U-turn fold in some sequences. In the glycyl-tRNAs, though, modification of the anticodon loop is sparse or non-existent. There are no modifications in the anticodon arm of bacterial tRNA\textsuperscript{Gly,GCC}. The presence of Mg\textsuperscript{2+} disrupts the U33-A37 base pair and increases the mobility of the loop nucleotides, but does not cause reorganization of the loop nucleotides into the U-turn motif. In some bacterial species, the anticodon arm of tRNA\textsuperscript{Gly,UCC} contains the 5-carboxymethylaminomethyl (cmnm\textsuperscript{5}) functionality on U34 (Figure 3.1). This modification is chemically similar to the mcm\textsuperscript{5}s\textsuperscript{2}U34 of \textit{E. coli} tRNA\textsuperscript{Lys,3}. Like tRNA\textsuperscript{Gly,UCC}, the anticodon sequence of tRNA\textsuperscript{Lys,3} is composed only of pyrimidine bases (5'-UUU-3') with a uridine nucleotide at position 34. In its modified form, a salt bridge exists between the U33pU34 phosphoryl oxygen and the mcm\textsuperscript{5} modification. However, the adoption of the U-turn in tRNA\textsuperscript{Lys,3} also requires thiolation at the base C2 position which strongly promotes the C3'-\textit{endo} ribose pucker and the t\textsuperscript{6}A37 modification (Kim \textit{et al.}, 1974). The cmnm\textsuperscript{5} modification also promotes the C3'-\textit{endo} ribose pucker, but much more weakly than C2 thiolation (Kim \textit{et al.}, 1974; Schmitz \textit{et al.}, 1999). Thus, it is possible that introduction of the cmnm\textsuperscript{5} modification into ASL\textsuperscript{Gly,UCC} could predispose the loop to adopt a U-turn fold; however, this modification alone is unlikely to substantially change the structure of the pyrimidine rich ASL\textsuperscript{Gly,UCC} loop (Kim \textit{et al.}, 1974).
3.6.3 Functions of the tRNA\textsuperscript{Gly} molecules

The physical properties of the glycyl anticodon stem-loops will have an impact on their various functions. In bacteria that utilize glycine for synthesis of the peptidoglycan cell wall, it appears important that the pool of charged np-tRNA\textsuperscript{Gly,UCC} not be diverted into the translation machinery. Exclusion of the np-tRNA\textsuperscript{Gly,UCC} from protein synthesis can be accomplished by two mechanisms. First, the absence of guanine nucleotides at the base of the TΨC stem of np-tRNA\textsuperscript{Gly,UCC} is predicted to weaken the complex with EF-Tu:GTP (Giannouli \textit{et al.}, 2009; Sanderson and Uhlenbeck, 2007), thereby limiting ribosome binding of aminoacyl np-tRNA\textsuperscript{Gly,UCC}. A second feature of np-ASL\textsuperscript{Gly,UCC} that may protect against participation in ribosome-catalyzed translation is the cytidine nucleotide at position 37. In proteinogenic tRNA molecules, residue 37 is a purine nucleotide that is modified in most cases. The purine-37 residue confers additional stacking energy for the codon-anticodon helix. tRNA\textsuperscript{Phe} molecules containing unmodified G or A nucleotides at position 37 bind the ribosome A-site more rapidly and are released more slowly than tRNA\textsuperscript{Phe} molecules containing C or U (Murphy \textit{et al.}, 2004). Thus, the C37 of np-tRNA\textsuperscript{Gly,UCC} may act to further limit its contribution to protein biosynthesis, a possible stop-gap since the sole glycyl-tRNA synthetase enzyme relies in part on a common fold of the tRNA\textsuperscript{Gly,UCC} and np-tRNA\textsuperscript{Gly,UCC} anticodon loops for activity. Interestingly, although the structures of the proteinogenic and non-proteinogenic tRNA\textsuperscript{Gly} anticodon arms are remarkably similar and exhibit similar responses to metal ions, the T\textsubscript{m} of the np-ASL\textsuperscript{Gly,UCC} stem is 10 °C higher than the T\textsubscript{m} of the ASL\textsuperscript{Gly,UCC} stem which can be attributed in part to the different 32-38 pairings in the two molecules (Figure 3.2). Also, if np-tRNA\textsuperscript{Gly} lacks the ability to interact with the codon GGC to any significant degree, then regulation of the glyQS riboswitch would be responsive only to the pool of uncharged proteinogenic tRNA\textsuperscript{Gly}.

Glycine is a member of the four-codon box family in which the codon nucleotide sequences differ only at the third, or wobble position, and the four possible codon combinations designate a single amino acid. In \textit{E. coli}, there are three glycine isoacceptor tRNAs. tRNA\textsuperscript{Gly,CCC} is the most discriminating and translates only the codon GGG to any significant degree (Samuelsson \textit{et al.}, 2004).
E. coli tRNA^{Gly,UCC}, in which the U34 modification is not identified but could be mmn^5 or cmmn5, translates all four glycine codons but translates codons GGU and GGC with ~25% the efficiency of tRNA^{Gly,GCC} (Samuelsson et al., 1983). In B. subtilis, which contains only two isoacceptors, tRNA^{Gly,GCC} and tRNA^{Gly,UCC}, U34 of tRNA^{Gly,UCC} has the modification cmmn5. The cmmn5 modification can facilitate stacking of residues 34 and 35 and increase codon discrimination. tRNA^{Gly,UCC} is the sole glycyl-tRNA present in Mycoplasma mycoides, and the only anticodon arm modification is N6-methyl A37 (Figure 3.1). Like the E. coli tRNA^{Gly,UCC}, the M. mycoides glycyl-tRNA translates the four glycine codons without discrimination (Samuelsson et al., 1983; Durant et al., 2005), but the translational efficiency of codons GGU and GGC is ~50% the efficiency of E. coli tRNA^{Gly,GCC} (Samuelsson et al., 1983). A key to the ability of the M. mycoides tRNA^{Gly,UCC} to read codons efficiently without discrimination was determined to be the C32-A38 base pair (Durant et al., 2005). Although E. coli tRNA^{Gly,UCC} also contains the C32-A38 base pair, the modification of U34 and/or lack of modification of A37 may curtail the ability of tRNA^{Gly,UCC} to read the four glycine codons with equal efficiency.

The ability of tRNAs to discriminate codon triplets based on the third codon nucleotide is a balance of contributions from modification of tRNA residues 34 and 37, the identity of residues 32 and 38, and the purine/pyrimidine composition of residues 35 and 36 (Agris, 2008; Durant et al., 2005). The property of codon discrimination exhibits modest correlation with the propensity of the U34 ribose pucker to adopt the C3'-endo conformation, which leads to a more conformationally ordered anticodon loop. The C2 thiolation present in some split codon boxes strongly reinforces the C3'-endo conformation (Schimtz et al., 1999). The mcm^5 and cmmn^5 also tend to favor the C3'-endo conformation, as does the 5-oxoacetic acid (cmo^5) modification, but to an even lesser extent (Schimtz et al., 1999; Sundaram et al., 2000). The ribose conformations of residue 34 in ASL^{Gly,GCC} and ASL^{Gly,UCC} are a mixture of C3'- and C2'-endo conformations. While cmmn^5 modification of U34 could change this equilibrium in B. subtilis (or E. coli) tRNA^{Gly,UCC}, the tRNA^{Gly,GCC} is unmodified and would be unchanged in vivo.

In six of the eight four-codon boxes, the cognate tRNA molecules with U34 contain the
modification cmo⁵ that suppresses wobble base discrimination (Sundaram et al., 2000) (Figure 3.1) and expands codon reading. One question that arises is why U34 of glycyl and arginyl tRNAs carry the mnm⁵ or cmnm⁵ modifications rather than cmo⁵ when restricted codon reading appears unnecessary. In tRNA{Gly} and tRNA{Arg}, U34 is followed by a cytidine whereas in the six other four-codon boxes, U34 is followed by a purine nucleotide. It is possible that A or G at position 35 facilitates a stacking of the loop bases, as evidenced by Mg²⁺-induced U-turn formation in unmodified tRNA{Val} (Byrne et al., 2010), that is less easily accomplished by the cytidine base and modification of U34 with cmo⁵ indirectly compensates for this positive contribution to loop ordering.
Chapter 4. Binding of tRNA anticodon stem-loop to T box specifier domain

The Specifier Domain (SD) is a structural element in the 5’ region of the mRNA leader that is variable in size and contains the Specifier sequence, three nucleotides that are complementary to the anticodon nucleotides of the cognate tRNA. The specificity of the riboswitch for tRNA is primarily achieved through pairing of the Specifier sequence nucleotides with the anticodon of tRNA (Grundy and Henkin, 1993; Grundy et al., 1997), and changes in this sequence can switch the specificity of the T box riboswitch to allow recognition of other tRNA species (Grundy and Henkin, 1993; Grundy et al., 1997; Grundy et al., 2002). In addition to the Specifier sequence, the SD contains a loop E structural motif that is necessary for proper regulatory function and structure maintenance (Rollins et al., 1997; Wang et al., 2010; Wang and Nikonowicz, 2011). The SD also contains a highly conserved purine residue immediately 3’ to the Specifier sequence that is positioned to pair with the invariant U33 of tRNA (Grundy et al., 2002, Vitreschak et al., 2008), creating the potential for a fourth base pair between the anticodon loop and the Specifier loop. This residue is protected from Mg$^{2+}$ cleavage in the tRNA-mRNA leader complex (Yousef et al., 2005), supporting the possibility that the tRNA-SD interaction involves four base pairs (Nelson et al., 2006). Recent SHAPE analysis of tRNA in complex with stem I of the Geobacillus kaustophilus glyQS T box riboswitch (which includes the Specifier domain) indicated protection of U33 but not of the conserved adenine (Griggs et al., 2013). However, in the recently reported co-crystal structure of the Oceanobacillus iheyensis glyQ riboswitch Stem I with tRNA Gly, the Specifier-tRNA interaction involves three base pairs, and the U33 of tRNA loops out (Zhang and Ferré-D’Amaré, 2013).

We have used NMR spectroscopy and isothermal calorimetry (ITC) to examine the interaction between the anticodon arm of tRNA Gly,GCC (ASL Gly) and the SD of the Bacillus subtilis tyrS mRNA leader containing the glycyl Specifier sequence, GGC. The SD-ASL Gly complex is formed...
by three stacked intermolecular Watson-Crick G-C base pairs. The conformation of the ASL Gly loop transitions from dynamic and disordered to a moderately stable U-turn structural motif in the complex. A U33A mutant of ASL Gly that is unable to form the canonical U-turn motif retains the ability to bind SD, but with reduced affinity. Our data are consistent with a configuration where the conserved purine 3’ to the Specifier sequence and the conserved purine 3’ to the anticodon (residue 37 of tRNA) stack against the ends of the intermolecular helix and may confer additional stability to the complex. These results are consistent with the co-crystal structure of the glyQ T box-tRNA complex and observations of solution SHAPE experiments (Griggs et al, 2013; Zhang and Ferré-D’Amaré, 2013).

4.1 Design of the SD and ASL molecules

The nucleotide sequences corresponding to the Specifier domain of the tyrS T box riboswitch with glycyl Specifier sequence GGC TyrSSD,GGC and the anticodon arm of tRNA Gly,GCCac. The Specifier (red) and anticodon (green) nucleotides confer specificity to the interaction. The highly conserved A (magenta) in the Specifier domain and the invariant U (orange) of the anticodon loop also have been proposed to pair. Nucleotides of the SD that form the loop E motif are shown in blue.

Figure 4.1 Secondary structures of the TyrSSD,GGC and ASL Gly,GCCac.
The glyQS T box is regulated by the tRNA\textsuperscript{Gly,GCC}, as demonstrated by the in vitro antitermination experiments. The anticodon stem-loop from this tRNA, ASL\textsuperscript{Gly,GCC}, has a stem of seven base pairs as shown in the previous chapter. We originally planned using ASL\textsuperscript{Gly,GCC} in the study of binding between T box specifier domain and tRNA anticodon stem-loop, since the glyQS system is the only successful in vitro T box system so far and ASL\textsuperscript{Gly,GCC} does not contain base modifications. However, ASL\textsuperscript{Gly,GCC}, not a full tRNA molecule, has a critical setback: it forms a homodimer when concentrated. Duplex ASL\textsuperscript{Gly,GCC} cannot bind to T box specifier domain and thus cannot be used in the study. A negative ITC result from Zhang and Ferré-D’amaré suggested the same conclusion (Zhang and Ferré D’Amaré, 2013). To overcome the problem, we linked the glyQS SD and ASL\textsuperscript{Gly,GCC} together in a single molecule. However, the approach proved to be futile because we could not observe any sign of SD-ASL binding using NMR. Finally, we incorporated the C32-A38 mismatch into the ASL\textsuperscript{Gly,GCC} as a new anticodon stem-loop (Figure 4.1). The new anticodon stem-loop (ASL\textsuperscript{Gly,GCCac}) is the molecule used in our later binding studies.

The native glyQS specifier domain was originally planned to be used in the binding study. To make better use of the NMR and structural data from the previous tyrS SD study, a hybrid SD is used. The new SD (TyrS\textsubscript{SDggc}) contains the glycyl codon 5’-GGC-3’ but otherwise wild-type tyrS SD sequence (Figure 4.1).

4.2 Successful binding between SD and ASL using ITC and native PAGE

Several ITC experiments have been performed to test the binding between T box SD and tRNA ASL. Our ITC experiment using the combination of TyrS\textsubscript{SDggc} and ASL\textsuperscript{Gly,GCCac} shows significant affinity between the two molecules with a dissociation constant (K\textsubscript{d}) of 0.37 µM. To test the effect of the GA motif on the binding, a molecule of TyrS\textsubscript{SDggc} with kink-turn (TyrS\textsubscript{SDKTggc}) has also been tested for its affinity with the ASL\textsuperscript{Gly,GCCac}. The existence of the kink-turn does slightly
Figure 4.2 ITC experiments show the binding between SD and ASL. The $K_d$ of the binding of $\text{TyrSSD}_{SGD,GGC}$ to the $\text{ASL}_{Gly,GCCac}$ is 0.37 $\mu$M, determined by ITC (b). Adding the kink-turn (a) to the specifier domain ($\text{TyrS}_{SDKT,GGC}$) causes about a three-fold increase in the affinity ($K_d = 0.10 \mu$M). An A33U mutation on the ASL decreases the affinity by three fold to a $K_d$ of 1.1 $\mu$M (c).

Figure 4.3 Disassociation of the complex
To confirm the shifted band is the complex, the band was cut out and eluted from the gel of the native-PAGE. The eluted RNA was then run on an urea PAGE. Staining of the urea gel showed two bands corresponded to the dissociates SD and ASL.
Figure 4.4 Native-PAGE of ASL-SD binding.
Constant amounts of SD are titrated with increasing amounts of ASL. The TyrS<sub>SD,GGC</sub> binding with ASL<sub>Gly,GCCac</sub> is shown in (a), ASL<sub>Gly,GGC</sub> in (b), and A33U ASL<sub>Gly,GCC</sub> in (c). The concentration of Mg<sup>2+</sup> is 2 mM except in (d), where no Mg<sup>2+</sup> is present and thus no band shifts.
increase the affinity between SD and ASL. The ITC combination of TyrS\textsubscript{SDKTgge} with kink-turn and ASL\textsubscript{Gly,GCCac} yielded a K\textsubscript{d} of 0.10 µM, which is a 3.5-fold stronger binding than without the kink-turn. Compared with the wild-type gly\textsubscript{QS} specifier domain, the engineered TyrS\textsubscript{SDgge} has a higher affinity for the ASL\textsubscript{Gly,GCCac} ASL. ITC results published previously from our lab showed a K\textsubscript{d} between gly\textsubscript{QS} SD and ASL\textsubscript{Gly,GCCac} to be 1.47 µM (Wang and Nikonowicz, 2011).

Native-PAGE was also used to monitor the binding between the SD and ASL. A constant amount of TyrS\textsubscript{SDgge} was mixed with variable titers of ASL\textsubscript{Gly,GCCac} and vice versa from 1/4 to 8X. The samples were then run on a 10% PAGE without denaturant. When the amount of ASL\textsubscript{Gly,GCCac} was constant, the low molecular weight band of free ASL diminished according to the increasing titer of SD. While the amount of TyrS\textsubscript{SDgge} was constant, a new high molecular weight band corresponding to the complex appeared, and the SD band diminished as the titer of ASL increased.

To confirm the shifted SD band is actually the SD-ASL complex, a gel slice of the shifted band corresponding to the complex was cut out. The RNA in the slice was eluted and then loaded into an urea gel, in which the RNA would be denatured and the mobility was based solely on molecular weight. The result of the urea gel showed two bands, one in high molecular weight and one in low molecular weight regions. The RNA eluted from the gel slice contains both ASL and SD (Figure 4.3). Therefore, the “complex” seen in the native PAGE is not only shifted from the free SD but also confirmed to have ASL.

To have better signal-noise in the NMR experiment, one approach is to reduce the Mg\textsuperscript{2+} concentration. Previously, the ITC experiments and native gel were all performed with at least 10 mM Mg\textsuperscript{2+} and 5 mM Mg\textsuperscript{2+} for NMR. To determine if lower Mg\textsuperscript{2+} concentration works, two native-PAGEs were performed such that one had 2 mM Mg\textsuperscript{2+} and one had no Mg\textsuperscript{2+} (Figure 4.4). The gel without Mg\textsuperscript{2+} showed no band shift, indicating no binding between the ASL and SD. The gel with 2 mM Mg\textsuperscript{2+} showed band shifts. Thus, the binding can occur in a Mg\textsuperscript{2+} concentration as low as 2 mM, which would improve NMR spectral quality.

### 4.3 NMR evidence of SD-ASL complex formation
**Figure 4.5 Imino 15N-1H HMQC of ASL Gly,GCCac**
The free ASL is shown in purple and the SD in pink. The ASL Gly,GCCac shows slight chemical shift changes when bound to the SD. The G39 and G40 H1 resonances shift downfield for about 0.1 ppm. The G40 peak of SD-bound may actually an overlap peak of the U33, so the relative intensities of the peaks change from G39>G40 in free ASL to G40>G39 in bound ASL.

**Figure 4.6 Spectral changes of the SD during ASL binding.**
Mg$^{2+}$-bound SD (purple) imino $^1$H-$^15$N cross-peaks are overlaid with the cross-peaks from ASL bound SD (pink). When SD bound with ASL, the N1 resonance of G26, the G of the sheared G-A of loop E motif, shifts upfield so does the G11. Two new peaks from G29 and G30 representing the anticodon-codon pairing shows up (asterisk). Mg$^{2+}$ is necessary for SD-ASL binding. The Mg$^{2+}$-free SD is shown in green. Addition of Mg$^{2+}$ caused some spectral changes, especially the upfield shift of G11.
Figure 4.7 NH region of the $^1$H 1D difference spectra of ASL$^{Gly}$. The $^1$H 1D difference spectrum of free ASL is shown on the left and SD-bound on the right. The $^{15}$N carrier positions of 145 and 161 ppm are centered on the guanine and uridine regions, respectively, for low power CW $^{15}$N decoupling. The top spectra were acquired using broadband (GARP) $^{15}$N decoupling centered at 153 ppm. In the complex, the U33 resonance intensifies and collapses to a single peak with a frequency that corresponds to a U in a U-turn motif.

Figure 4.8 Imino proton connectivities between adjacent base pairs in H$_2$O $^1$H-1H NOESY SD-ASL$^{Gly}$ complex. The dashed lines trace the connectivities of the ASL$^{Gly}$ molecule in the complex. For clarity, the (*) was added to indicate ASL$^{Gly}$ nucleotides. The ASL$^{Gly}$ is in 50% molar excess and the dotted lines trace the connectivities among the NH resonances of excess ASLGly molecules. The labels identify cross peaks between NH protons of neighboring base pairs. Connectivity through the Specifier codon triplet, G29–G30–C31, is highlighted in blue (G29 and G30 of SD and *G34 of the ASL). The inter-base pair connectivity is broken between G6–C33 and C31–*G34 and between G29–*C36 and the reversed Hoogsteen U12–A27. The absence of a G6–*G34 cross peak indicates that the lower stem and the Specifier– anticodon helix are not coaxially stacked and that A32 may provide a platform to stabilize the C31–*G34 base pair.
The imino (NH) resonances of the SD and ASL<sub>Gly</sub> molecules were monitored using 15N-1H HMQC spectra and assigned using NOESY-based experiments. Preliminary studies to identify and optimize conditions that stabilize the SD-ASL<sub>Gly</sub> complex demonstrated that Mg<sup>2+</sup> was necessary. Addition of Mg<sup>2+</sup> to ASL<sub>Gly</sub> disrupts the C32-A<sup>38</sup> base pair of the hairpin, but has little effect on the NH spectrum. Addition of Mg<sup>2+</sup> to SD reinforces the loop E motif as evidenced by the appearance of NH resonances for nucleotides U12 and G26 that participate in reversed Hoogsteen U-A and sheared A-G base pairs, respectively (Figure 4.6). The NH resonance of the bulged G11 nucleotide shifts 1.0 ppm upfield in response to Mg<sup>2+</sup> binding. However, at the base of the Specifier loop, the NH resonances of U35, U5, and G6 are doubled. The splitting between the resonances is most pronounced for residues proximal to the loop-helix junction. Exchange cross peaks between the split resonances in the NOESY spectrum (not shown) are indicative of two conformations at the base of the Specifier loop.

The SD-ASL<sub>Gly</sub> complexes were prepared such that one of the RNA molecules was 13C/15N-

<table>
<thead>
<tr>
<th>Table 4.1 Chemical shift table for SD-ASL binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above is a table summarizing the imino and amino proton and nitrogen chemical shifts observed in the study of SD-ASL binding.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ASL&lt;sup&gt;Gly&lt;/sup&gt; free</th>
<th>ASL&lt;sup&gt;Gly&lt;/sup&gt; complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH, N</td>
<td>NH, N</td>
<td>NH, N</td>
</tr>
<tr>
<td>G2</td>
<td>12.43</td>
<td>146.61</td>
</tr>
<tr>
<td>U7</td>
<td>10.69</td>
<td>160.5&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>G8</td>
<td>10.37</td>
<td>146.5&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>G13</td>
<td>12.62</td>
<td>147.62</td>
</tr>
<tr>
<td>G14</td>
<td>13.23</td>
<td>147.71</td>
</tr>
<tr>
<td>U15</td>
<td>14.37</td>
<td>162.15</td>
</tr>
<tr>
<td>SD free</td>
<td>SD complex</td>
<td></td>
</tr>
<tr>
<td>NH, N</td>
<td>NH, N</td>
<td>NH, N</td>
</tr>
<tr>
<td>G2</td>
<td>12.72</td>
<td>146.73</td>
</tr>
<tr>
<td>G3</td>
<td>12.28</td>
<td>146.30</td>
</tr>
<tr>
<td>G5</td>
<td>13.35</td>
<td>161.27</td>
</tr>
<tr>
<td>U5</td>
<td>11.84</td>
<td>145.53</td>
</tr>
<tr>
<td>G6</td>
<td>11.80</td>
<td>145.37</td>
</tr>
<tr>
<td>G11</td>
<td>9.44</td>
<td>143.62</td>
</tr>
<tr>
<td>U12</td>
<td>12.73</td>
<td>160.72</td>
</tr>
<tr>
<td>U14</td>
<td>12.48</td>
<td>147.28</td>
</tr>
<tr>
<td>U15</td>
<td>12.11</td>
<td>146.35</td>
</tr>
<tr>
<td>U18</td>
<td>11.69</td>
<td>159.19</td>
</tr>
<tr>
<td>U19</td>
<td>n.o.</td>
<td>n.o.</td>
</tr>
<tr>
<td>G21</td>
<td>9.78</td>
<td>142.42</td>
</tr>
<tr>
<td>G22</td>
<td>13.40</td>
<td>147.44</td>
</tr>
<tr>
<td>G23</td>
<td>14.33</td>
<td>162.34</td>
</tr>
<tr>
<td>G26</td>
<td>11.60</td>
<td>149.35</td>
</tr>
<tr>
<td>G29</td>
<td>n.o.</td>
<td>11.85</td>
</tr>
<tr>
<td>G30</td>
<td>n.o.</td>
<td>13.14</td>
</tr>
<tr>
<td>U35</td>
<td>13.88</td>
<td>161.45</td>
</tr>
</tbody>
</table>

*Uncertainty in chemical shift is +/− 1.5 ppm. Chemical shift determined from NOESY spectrum of complex.*
labeled and the other was unlabeled. In the HMQC spectrum of SD, complex formation leads to the appearance of two additional guanine NH resonances at 11.85 ppm (G29) and 13.14 ppm (G30) (Figure 4.6). These resonances give rise to a cross peak in the NOESY spectrum (Figure 4.8). The resonance at 13.14 ppm also has a NOE cross peak with a NH resonance at 12.78 ppm. A $^{15}$N-edited NOESY spectrum indicated this latter NOE cross peak is inter-molecular, establishing the identities of the NH resonances at 13.14 and 12.78 ppm as G30 of SD and *G34 of ASL$^{Gly}$, respectively. Also in the complex, split SD resonances of U35, U5, and G6 collapse into single peaks with chemical shifts unique to the complex (Figure 4.6). These results indicate that binding of ASL$^{Gly}$ imposes conformational ordering to the base of the Specifier loop. Changes in the U12 and G26 NH resonances also point to additional ordering within the loop E motif either to accommodate or in response to ASL$^{Gly}$ binding. The U12 NH resonance intensifies and the G26 $^{15}$N NH resonance shifts 4.0 ppm upfield. The shift of the G26 resonance is consistent with formation of the extensive hydrogen bond network involving the base O6, N1, and N2 atoms of this residue in the context of a loop E motif (Cabello-Villegas et al., 2004).

4.3.1 ASL$^{Gly}$ may form a U-turn when in complex with T box SD

The free form of ASL$^{Gly}$ has a five base pair stem, but the seven nucleotide loop does not adopt the archetypal U-turn motif as detailed in Chapter 3. The $^{15}$N-$^1$H HMQC spectrum contains four major peaks corresponding to the stem nucleotides and a fifth weak peak that corresponds to the terminal G27 residue (Figure 4.5). There are three additional peaks in the NH $^1$H spectrum but the resonances are exchange broadened and do not give rise to peaks in the HMQC spectrum. Because there is no evidence for multiple conformations of the stem nucleotides in other NH (NOESY) or base (HMQC and NOESY) spectra, these resonances were tentatively determined to correspond to the unpaired U33 and G34 residues. The exchange-broadened peaks were characterized using $^1$H 1D difference experiments (Figure 4.7). The $^{15}$N carrier was positioned at different chemical shifts in the NH nitrogen region and pairs of $^1$H spectra were recorded with and without
low power CW $^{15}$N decoupling. The NH $^1$H resonances from labeled ASL$_{Gly}$ display incomplete cancellation when the $^{15}$N carrier is positioned at or near the frequencies of the corresponding NH nitrogen resonances. $^{15}$N decoupling centered at 160 ppm (the region of the $^{15}$N spectrum characteristic of uridine bases) leads to the appearance of two peaks, 10.69 ppm and 11.45 ppm (Figure 4.7). Decoupling centered at 144 ppm (the guanine region of the $^{15}$N spectrum) leads to the appearance of a peak at 10.37 ppm. Thus, these peaks were assigned to U33 and G34, respectively. Titration of unlabeled SD with labeled ASL$_{Gly}$ causes ~0.1 ppm downfield shifts of the loop-proximal G39 and G40 NH $^1$H resonances in the $^{15}$N-$^1$H HMQC spectrum. Although no new cross peaks become apparent, the G34 (previously assigned using the NOESY spectrum) and G39 resonances partially overlap and give rise to a single intense peak centered at 12.72 ppm (Figure 4.5). To locate the U33 NH resonance in the complex, 1D difference spectra were again acquired. Imino $^1$H resonances from the unlabeled SD molecule are not affected by the $^{15}$N decoupling field and thus cancel when the spectral pairs are subtracted. In the complex, the two U33 NH resonances become a single peak at 11.64 ppm when $^{15}$N decoupling was centered at 160 ppm (Figure 4.7). $^{15}$N decoupling centered at 144 ppm confirms the shift of the G34 NH resonance away from 10.37 ppm in the complex. Although the U33 NH $^1$H resonance in the complex is moderately broad, the chemical shift of the resonance is consistent with the spectroscopic signature of a uridine nucleotide that participates in a U-turn motif (Sundaram et al., 2000; Cabello-Villegas et al., 2004). Hydrogen bonding involving a 2’-OH or bound water molecule also can lead to NH $^1$H chemical shifts over this range, but

Figure 4.9 $^{15}$N-$^1$H HMQC spectrum of U33A ASL$_{Gly}$-bound SD.

Above is the $^{15}$N-$^1$H HMQC spectrum of the NH region of SD in complex (black) with two mole equivalents of the U33A mutant of ASL-Gly. Peaks highlighted in red correspond to resonances of the Mg$^{2+}$-bound SD molecule. The G26, G29, and G30 resonances (blue boxes) appear to be in intermediate exchange and are not present in the spectrum. With the addition of one equivalent of native ASL$_{Gly}$, G26, G29, and G30 appear and peaks from the free SD disappear.
a moderately stable U-turn that limits U33 dynamics is consistent with SHAPE experiments that show protection of this residue after complex formation (Grigg et al., 2013).

To further explore the importance of the U33 residue for SD binding, a U33A mutant of ASL Gly was examined (Figure 4.1). The invariant U33 of tRNA molecules is the basis of the structural motif known as the U-turn (Ogle et al, 2002). The U-turn motif reverses the direction of the phosphate backbone and orients the anticodon bases for presentation to the mRNA codon at the ribosome. The motif is stabilized by cross-strand hydrogen bonds involving the U33 NH and 2’-OH groups and by stacking of the anticodon bases. The A33 mutant is not isosteric with U33 and cannot form the characteristic hydrogen bond network, but A33 also is not expected to sequester the anticodon bases and prevent intermolecular pairing. Indeed, the NMR spectrum of the U33A mutant of ASL Gly indicates that the loop is disordered (not shown). However, if U33 of tRNA pairs with A32 of SD, the A33 substitution is expected to severely impair complex formation by introducing an A-A mismatch. Native PAGE reveals an incomplete band shift of U33A ASL Gly by SD, and NMR spectra indicate U33A substitution weakens the affinity of ASL Gly for SD. The HMQC spectrum of SD in the presence of two-fold excess U33A ASL Gly (Figure 4.9) displays features of free SD (resonances U5, U35, and G6) and bound SD (resonances U5, U35, G6, and U12) which indicate a lifetime for the complex > 0.01s. Not unexpectedly, NH resonances G29 and G30 continue to exchange with solvent and are absent from the spectrum. Interestingly, the G26 resonance is exchange-broadened beyond detection, indicating a lifetime for the complex of ~0.003s. The different lifetimes indicated by the peaks suggest that formation of the complex involves structural changes in the loop E motif (possibly involving G11 or G26) that are not experienced at the base of the Specifier loop. The addition of one equivalent of ASL Gly to the sample yields the SD spectrum corresponding to the complex, consistent with the stronger affinity of the cognate ASL. The NMR spectral results are supported by ITC measurements that reveal the complex between SD and A33-ASL Gly is about three times weaker than the complex between SD and U33-ASL Gly (Figure 4.2) (0.37 µM vs 1.1 µM).

Two compensatory mutants of SD, A32U and A7U/A32U, were tested for their ability to
Figure 4.10 Structural modeling of the SD-ASL complex.
Above are two stereo images showing the two possible SD-ASL binding orientations. The ASL can either bind to the minor groove (a) or major groove (b) of the SD. The minor groove model has a less distorted SD structure.
associate with U33A-ASL Gly and restore complex stability. However, complex formation could not be detected by NMR or by native PAGE analysis. In the SD molecule, U32 within the context of A7 extends the lower helix into the Specifier loop by one Watson-Crick base pair, reducing the number of unpaired nucleotides from seven to five, and limiting the conformational freedom of the Specifier sequence. The A7U/A32U double mutant does not appear to form an additional base pair, but the absence of A32 in both mutants also could result in loss of important stacking interactions as observed in the co-crystal structure (Zhang and Ferré-D’Amaré, 2013).

4.4 Structural modeling of the ASL-SD complex

Modeling of the SD-ASL complex was done by modeling the two molecules separately and then modeling again with the merged coordinates and the intermolecular hydrogen bonding constraints. The NMR spectra of TyrS SDgc showed similar patterns and features as seen in the wild-type tyrS SD, especially the resonances of the loop E motif. The signature of the loop E motif is the

Figure 4.11 The anticodon-codon pairing in the SD-ASL binding. Model of the interaction between the Specifier Sequence (green) and the ASL Gly loop (red). The loop E motif nucleotides are highlighted in blue. The ASL possibly forms a U-turn and three base pairs are involved in the SD-ASL interaction. U33 although does not pair with A32 of the SD, it still help stabilize the U-turn structure.
S-turn, which contains three non-canonical base pairs: an A-A parallel, a reversed Hoogsteen A-U, and a sheared G-A. Between the A-A parallel and the reversed Hoogsteen A-U, there is a guanine looped out. Therefore, modeling of the TyrS SD used the dihedral angels and the NOE distances of the tyrS SD structure by our lab (Wang et al., 2010). The modeling of ASL GlyGCCac used the dihedral angles and NOE distances from the ASL GlyUCC molecule. When modeling the complex, hydrogen bonding constraints between the anticodon-codon pairing were added and also for the U-turn. The dihedral angel constraints of the codon nucleotides had to be loosened to avoid violations (Figure 12).

The ASL could bind either through either the minor groove or the major groove of the SD (Figure 4.10). Without further modification to the constraints, the ASL binds to the SD via minor groove according to simulation result. To force the ASL to bind to the SD via major groove, a

Figure 4.12. Structural changes in the loop E motif during SD-ASL binding.
The original loop E structure in the tyrS SD (Wang et al., 2010) has the Watson-Crick edges of the second and third codon nucleotides facing the major groove side and the first to the minor groove side (blue). To pair with the anticodon nucleotides, the condon nucleotides have to be all on either the minor (green) or the major (yellow) groove side. The phosphate backbone, therefore, has to change to accommodate the anticodon pairing. The three codon nucleotides and the A32 are shown in thick color sticks and nucleotides in the loop E motif are shown in thin color sticks.
lot more efforts have been put to modify the constraints, and the simulation showed that a major
groove binding could not be achieve without unwinding part of the helix. Therefore, binding of
ASL to SD through the minor groove should be naturally more favorable according to our simula-
tion. Indeed, in the crystal structure by Zhang and Ferré-D’Amaré, the binding occurs through the
minor groove of SD. Pairing via the minor groove face of SD positions the 3’ flanking unpaired
adenine bases, A37 of ASLGly and A32 of SD, to stack against each end of the anticodon-Specifier
sequence minihelix and enhance stability of the interaction (Figure 4.11). These features are ex-
hibited in the crystal structure. The crystal structure also shows the loop E motif in the Specifier
domain collapses to a more compact state when the tRNA is binds, consistent with cross peak
changes in the NH spectrum of SD (Figure 4.6).

Forming a fourth base-pair also seemed very unlikely after the simulation was performed.
A few attempts were made to introduce the hydrogen bonding constraints of the fourth base pair.
However, the simulation gave unfavorable results which marked the constraints violated. Indeed,
to form a fourth base pair the U33 of ASL has to be on the same groove as the other three anticodon
nucleotides, and this also applies to A32 and the codon nucleotides of the SD. To be on the same
groove as the anticodon nucleotide means the turning of the phosphate backbone has to be moved
forward before U33, this would change a lot to the overall structure of the ASL.

4.5 Discussion

In the ASL, the base of U33 is not positioned for intermolecular pairing in either the solu-
tion model or crystal structure. However, increased protection of the U33 NH proton from solvent
exchange and reduced dynamics of the U33 ribose in the complex, as evidenced by SHAPE results
(Grigg et al., 2013), suggest U33 adopts a more intrahelical orientation in solution than the looped
out conformation observed in the co-crystal (Figure 4.13). In addition, the extrahelical confor-
mation of residue 33 would be expected to accommodate the U33A substitution and not result in the
impaired binding exhibited by the mutant. The U33A substitution could lower complex stability
through steric clash in an intrahelical orientation or through loss of intramolecular hydrogen bonds characteristic of the U33 base that stabilize loop architecture. Although the U-turn motif is not a major conformer of the ASL$^{Gly}$ in solution (Denmon et al., 2011), the apparent absence of this structural motif in the free molecule is not unique to the ASL$^{Gly}$ loop (Correl et al., 1997; Quigley and Rich, 1976; Murphy et al., 2004). Indeed, the tRNA anticodon loops found to adopt the U-turn motif free in solution also contain extensive natural nucleotide base modifications (Quigley and Rich, 1976).

In agreement with biochemical studies, the solution model and crystal structure of the complex indicate tRNA is selected through complementary pairing of Specifier and anticodon
bases (Grundy and Henkin, 1993; Grundy et al., 1997; Grundy et al., 2002), and the rejection of non-cognate or near-cognate tRNA appears dependent on intermolecular helix stability. On the ribosome, invariant nucleotides G530, A1492, and A1493 of 16S rRNA form a hydrogen bond network involving 2’-OH groups that interrogate the minor groove of the codon-anticodon helix (Ogle et al., 2001; Ogle et al., 2002). This network is critical to selection of cognate tRNAs and rejection of near-cognate tRNAs (Ogle et al., 2002). Although presumably less critical than in translation, a similar “proof-reading” mechanism to ensure cognate tRNA selection is not apparent for the Specifier-anticodon interaction. Instead, cognate tRNA selection may rely on the geometry of the Specifier-anticodon helix to promote favorable interactions, such as ordering of the SD loop E motif upon ASL binding and optimized stacking of the flanking purines, to slow tRNA dissociation. Slower tRNA dissociation could then allow docking of the tRNA elbow and apical loop of stem I to further stabilize the complex (Grigg et al., 2013; Zhang and Ferré-D’Amaré, 2013).

4.5.1 tRNA^{Gly,GCC} may be the only isoacceptor used in T box regulation.

So far the only function of the non-proteinogenic tRNA in Staphylococci is for peptidoglycan synthesis (Roberts et al., 1974; Giannouli et al., 2009). Staphylococci also utilize the T box mechanism (Gutiérrez-Preciado et al., 2009). Therefore, it is possible that non-proteinogenic tRNAs also participate in the T box regulation through wobble and thus provide an additional cellular function. To test if the np-ASL^{Gly,UCC} could recognize the GCC codon on the SD using wobble, an ITC experiment was performed between the np-ASL^{Gly,UCC} and the TyrSSDgcc. The result showed no binding occurs between the two molecules. Native gel analysis was also performed to test the binding between TyrS_{SDgcc} and ASLs other than ASL^{Gly,GCC}. The three ASL molecules used in Chapter 3 – ASL^{Gly,GCC}, ASL^{Gly,UCC}, and np-ASL^{Gly,UCC} – were tested for their affinity for TyrSSDgcc using native gel. Of the three ASLs, only ASL^{Gly,GCC} demonstrated a robust band shift. The low molecular weight bands for ASL^{Gly,UCC} and np-ASL^{Gly,UCC} stayed visible, indicating no binding between them and the TyrS_{SDgcc}. Although ASL^{Gly,GCC} was able to completely shift the band of SD on
a native gel, we could not observe the ASL spectral change we observed in the ASL\textsuperscript{Gly,GCCac} using NMR.

It has been shown that all glycy1 T box leaders use the 5'-GGC-3' codon (Gutiérrez-Preciado \textit{et al.}, 2009). According to the modified rule of wobbling proposed by Lim and Curran, the unmodified U at the wobble position of anticodon can recognize all four nucleotides (Lim and Curran, 2001). This perspective make sense because glycine is a four-codon box family, and in \textit{Mycoplasma mycoide}, mitochondria, and chloroplast, the tRNA\textsuperscript{Gly,UCC} is the solely isoacceptor that reads all four different codons with equal efficiency (Samuelsson \textit{et al.}, 1983). The ASL\textsuperscript{Gly,UCC} was expected to bind the Tyr\textsubscript{SDggc} using wobble. However, according to our results, this may not be the case. The discrepancy between T box and ribosomal wobble may due to the fewer components in the T box recognition. Unlike in the ribosome where additional interactions are present between the ribosome and the tRNA (Selmer \textit{et al.}, 2006), the T box mRNA leader and the tRNA have been shown to be the only necessary components besides the RNA polymerase for \textit{in vitro} anti-termination (Grundy \textit{et al.}, 2002). Therefore, without the help such as that provided by ribosomal proteins to stabilize the anticodon-codon mini-helix between the SD and ASL, the stabilization of the anticodon-codon mini-helix is solely dependent on the three base pairs and the stacking of the mini helix. A mismatch and two base pairs are just not enough to hold the complex together. The specificity of the T box recognition, in turn, is strict and has no room for wobble.

4.5.2 Unmodified ASL\textsuperscript{Tyr} failed to stabilize a complex with wild-type \textit{tyrS} SD

An ITC experiment was performed for the wild-type \textit{tyrS} SD and the unmodified tRNA\textsuperscript{Tyr} anticodon stem-loop. ASL\textsuperscript{Tyr} bears the anticodon GUA and the G is modified to a queosine (Q) \textit{in vivo}. Besides the Q, the A next to the third anticodon nucleotides is also modified to ms2i6A \textit{in vivo}. The \textit{tyrS} SD uses the 5'-UAC-3' specifier codon. The experiment was to determine if the \textit{tyrS} SD could bind the unmodified ASL\textsuperscript{Tyr} by forming two A-U and one G-C inter-molecular base-pairs as the three G-C pairs between Tyr\textsubscript{SDggc} and ASL\textsuperscript{Gly,GCCac}. The ITC result, however, shows no bind-
Both ASL<sup>Gly,GCC</sup> and unmodified ASL<sup>Tyr</sup> do not adopt the U-turn conformation when free in solution (Denmon et al., 2011). Even though our NMR data for the binding between ASL<sup>Gly,GCCac</sup> and Tyr<sub>SD</sub> suggests that the ASL may change conformation and form a U-turn when in complex with the SD, it is very likely that, without base modifications, ASL<sup>Tyr</sup> cannot stabilize a U-turn and form a complex with tyr<sub>SD</sub>. The condition needed to form a tyr<sub>SD</sub> and ASL<sup>Tyr</sup> complex is more strict than the glycyl system. Compared with three G-C base pairs, two A-U and one G-C base pairs may not be enough to stabilize the complex and a U-turn; additional bonds may be necessary. Given the fact that base modifications help stabilizing the U-turn in tRNA ASLs (reference), base modifications may also, in turn, facilitate complex formation between T box SD and tRNA ASL. Although being the first identified T box, the in vitro antitermination for tyr<sub>S</sub> system has been unsuccessful and the only successful in vitro system is gly<sub>QS</sub>. It has been postulated that the in vitro antitermination of tyr<sub>S</sub> does not work because of either improper folding or the lack of base modifications. The structure of tyr<sub>S</sub> SD is solved and the engineered GCC variant is also proved to bind the ASL, so improper folding is unlikely. In conclusion, the lack of base modifications may be the primary reason for the unsuccessful tyr<sub>S</sub> in vitro termination and ITC binding assays.
Chapter 5. Interactions between the T box AG loop and apical loop

There are several conserved motifs in a T box riboswitches and one of them is the apical loop of the stem I. Beneath the apical loop is the AG loop known as the AG motif. The apical loop has a conserved GNUG sequence from base 68 to 71. The GNUG nucleotides have been shown to be functionally important. Mutation of the conserved U70 of the GNUG completely abolished T box gene expression and its ability to be induced for antitermination in vivo (Rollins et al., 1997). Furthermore, a Mg$^{2+}$ cleavage study also suggested the region is structured (Yousef et al., 2005).

Figure 5.1 Secondary structures of the T box apical and AG loops. The top of the stem I are broken into two molecules – the apical loop (a) and the AG loop (b). A full double-loop molecule (c) is used to study the interactions between the two loops. Dashed lines indicate the proposed tertiary interactions.
Figure 5.2 Converged structure of the *serS* apical loop.
The structure of *serS* apical loop shows base stacking in the loop region. The backbone makes a turn between G68 and G69. There is no evidence of forming G-U base pairs.

Figure 5.3 Superimposition of the *serS* apical loop structures
Above is a figure showing the superimposition of the eight structures of *serS* apical loop that have the lowest energies. The best eight structures align with an RMSD of 1.572 Å. The loop region shows some variance but also the very similar turn of the phosphate backbone and the stacking of the loop bases.
In 2013, Lehmann and colleagues reported the sequence similarity between T box stem I, ribosome L1 stalk (Selmer et al., 2006), and RNAse P (Krasilnikov et al., 2003). For the first time, a model was proposed and constructed based on prol T box stem I, showing a head-to-tail double T-loop structure that interacts with the tRNA elbow (Lehmann et al., 2013). Recent crystal structures of glyQS T box stem I by Grigg et al. and by Zhang and Ferré-D’Amaré further confirmed the model and the tertiary interactions between the apical and AG loops (Grigg et al., 2013; Zhang and Ferré-D’Amaré, 2013). The two crystal structures also showed that the double T-loop structure provides a binding platform for the elbow of tRNA (parts of T-loop and D-loop). In this study, we used solution NMR to solve the high-resolution structure of the serS T box apical loop per se, and then further analyzed the effect caused on the apical loop by the AG loop and the U70G mutation. The serS loop displays good base stacking when on its own, and the unusual chemical shifts we observed after the attachment of AG loop suggest complicated structural changes caused by the loop-loop interaction whereas the U70G mutation lacks some of these traits.

### 5.1 Structure of the serS apical loop

We utilized NMR spectroscopy and then constructed a structural model of the serS apical loop. The NMR spectra of serS apical loop are well-resolved, indicating that the serS apical loop is well-structured. Unlike the apical loops of glyQS and tyrS, the serS apical loop is not a GNRA tetra-loop candidate. The sequence of bases 69-72 in serS is GUGG, whereas the sequence is GUGA in glyQS and tyrS. We also examined the tyrS apical loop using NMR but the spectra indicate multiple conformations. Secondary structure predictions of the serS apical loop indicate the potential of forming one to two additional G-U base pairs, which could further stabilize the structure (Figure 5.1). However, no NMR evidence of G-U base pairs was observed. The $^{15}$N-$^1$H HMQC for the imino groups shows only two uridine H3 at 14.00 and 14.08 ppm, which correspond respectively to U61 and U78 that form a Watson-Crick base pair in the stem. Uridine H3s in G-U base pairs are expected to show up slightly upfield at 12-13 ppm. However, this does not
exclude the possibility that additional G-U pairs are in intermediate exchange regimes and thus could not be detected by the $^{15}$N-$^1$H HSQC.

Our structure calculation indicates that the bases in the loop region are well-stacked (Figure 5.2). The stacking starts from G64 through G68, and then the phosphate backbone makes a turn between G68 and G69 before continuing the stacking through A19. The stacking is consistent with inter-base NOEs. Although weak, H8/H6 to H8/H6 cross peaks are observable between the loop bases except between G68 and G69. Other evidence supporting the stacking includes the inter-residual H2 NOEs between the three adenines from A73 to A75 and the inter-residual H5 cross peaks of U66 and U67. Inter-residual H8 and H5 cross peaks are also observed between G65 and U66, U67 and G68, G69 and U70, and U70 and G71.

### 5.2 Effect of the AG loop

Recently, two crystal structures of glyQS T-box demonstrated the loop-loop interactions between the side and apical loops (Grigg et al., 2013; Zhang and Ferré-D’Amaré, 2013). The two loops form a structure similar to that in RNAse P (Krasilnikov et al., 2003) and the L1 stalk of the ribosome E-site (Selmer et al., 2006). We observed NMR evidence of interactions between the serS side and apical loops. Attaching the AG-rich AG loop to the apical loop construct (Figure 5.1) causes significant changes to the NMR spectral pattern of the imino and amino groups (Figure 5.4). We observed several new and resonances that are not in the chemical shift range of standard Watson-Crick base pairs and could be from the complex hydrogen bonding networks between the two loops as seen in the crystal structures. As mentioned in the previous section, we only observed one uridine N3 imino in the $^{15}$N-$^1$H HSQC spectrum of the serS apical loop (Figure 5.4). When the AG loop is attached, several new U imino resonances appear in the $^{15}$N-$^1$H HSQC. Besides the two downfield Watson-Crick adenine paired and one upfield guanine paired uridine N3H resonances from the two stem regions, a new peak that we assigned to the U70 N3H resonance shows up at 12.97 ppm (Figure 5.4). This imino proton resonance has cross peaks to one of the adenine H6s
Figure 5.4 Imino 15N-1H HMQC of the serS apical loop and AG loop.
The imino spectrum of the serS apical loop (a) only shows two uridine resonances from the A-U base pairs. In the imino spectrum of the AG loop (b), a few upfield guanine resonances are not present. The imino spectrum of the full double-loop molecules (c) show upfield guanines and the U70 that participate in the interactions between the two loops. The U70G mutation (d) disrupt the interactions between the two loops so that the upfield guanines are missing from the spectrum.
Figure 5.5 The amino $^{15}$N-$^1$H HMQC of the full double-loop molecule. Assignments of the adenine H$_6$s that participate in the interactions between the apical and AG loops. The A74 H$_6$1 and G69 H21 resonances are unusually downfield for adenine and guanine amino protons.

Figure 5.6 The HNCO spectrum of the serS double-loop molecule. The HNCO experiment shows the correlation between uridine H3 and C2/C4 and guanine H1 and C2/C6. The uridine C2 resonates more upfield in a Watson-Crick A-U base pair than a reversed Hoogsteen, whereas the uridine C4 resonates more downfield in a Watson-Crick A-U base pair than in a reversed Hoogsteen. U70 C2 and C4 chemical shifts indicate the formation of the reversed Hoogsteen base pair. U67 may also form a reversed Hoogsteen A-U based on another observed downfield-shifted C2.
that we assigned to A73 in the H2O 1H-1H NOESY. The corresponding bases in the glyQS stem I of U70, A73, and G54 form a base triple by a reverse Hoogsteen A-U base pair and a parallel G-A base pair to the same adenine; thus, both amino H6s are paired. The chemical shifts of the U70 H3 and the fact that both H6s of the two A73 N6 amino protons are protected from solvent exchange and show up in the 15N-1H HSQC match the characteristics of forming a base triple (Figure 5.5). Furthermore, the HNCO experiment confirms the formation of the U7-A73 reverse Hoogsteen base pair. As seen in other U in reversed Hoogsteen A-U base pairs (Fürtig et al., 2003), the C2 resonance of U70 shifts downfield and the C4 resonances shifts upfield comparing to the C2 and C4 resonances in Watson-Crick A-U base pairs (Figure 5.6).

In the crystal structure, G54, G56, and G69 form a triple where both H2s in the G69 are paired with one to the O6 of G54 base and the other to the O2' of G54. Examination of the coordinates indicates that G56 and G69 have the ability to form a bifurcated hydrogen bond, with G69 H1 and H21 both acting as hydrogen donors to the O6 of G56. Upfield shifted guanine H3 resonance was observed in the G-G bifurcated hydrogen bond in the loop E motif of the 5S rRNA (Dallas and Moore, 1997). The H1 chemical shift of G69 is 10.18 ppm, which is the chemical shift being observed (10.20 ppm) in the loop E motif of 5S rRNA. The upfield shift of the a H1 resonance in the bifurcated hydrogen bond was explained by its inability to pair with a nitrogen (Dallas and Moore, 1997). One strong guanine amino N2H2 cross peak showed up at 8.63 ppm in the 15N-1H HSQC and its much weaker partner at 7.34 ppm can be detected by going down several contour levels (Figure 5.5). A cross peak of this H2 to the H1 of G69 is observed in the H2O NOESY.

5.2.1 Spectral feature of the AG loop

An AG loop construct was designed and consisted of the AG loop and the two flanking stems. The apical loop was substituted by a UUCG tetra loop. This AG loop molecule serves as a control for the interaction between the apical loop and the AG loop. As expected, the 1H-15N HMQC of the AG loop *per se* did not show peaks from the apical loops nor the peaks of G58 and
G60. The guanine and uridine imino protons of the two stems are still present, as are resonances in the downfield region. The most upfield guanine imino H1 at 10.26 ppm is the guanine of the UUCG tetra-loop. The G62 and U80 iminos of the closing G-U pair of the upper stem still resonate at about the same position as in the spectrum of the double-loop molecule.

Based on these results, the AG loop is sure to form a different structure without the apical loop. As the iminos from the stems and the G-U base pairs stay the same, the two stems should still wind as a continuous helical structure. The AG loop residues are not protected from solvent and are more dynamical, supported by the fact that the imino resonances of G58 and G60 are not present.

5.3 Effect of the U70G mutation

The U70G mutation, which changes the conserved GNUG to GNGG, has been shown to decrease the basal level of T box gene expression and completely shut down the induction by uncharged tRNA in vivo (Rollins et al., 1997). Therefore, the mutation is causing a disastrous effect on proper T box function. We have examined the serS U70G mutant to identify the possible structural effect on the apical loop.

The U70G mutant of the serS apical loop does not show spectral features other than what we see in the wild type. Therefore, the mutation appears to affect the structure of apical loop minimally. We also examined the H-15N HSQC of the U70G mutant of serS double-loop molecule and noticed significant spectral changes. A few NH cross peaks previously observed in the wild type spectrum are missing from the mutant spectrum. Besides the mutated U70, the NH resonance of U80 from the G-U base pair is no longer observable, and neither are the U59 and U76 from the A-U base pairs in the apical stem. The NH resonance of the U83 from the distal stem remains unchanged. The guanine imino resonances are significantly broader than what is seen in the wild-type spectrum. Several guanine imino resonances are either disappear or broaden. The position that was occupied by G53 and G58 turns out to be a single large peak. There is also a strong resonance at 11.5 ppm that could be the broadened G60 and G71. Although the disappearance of adenine paired
uridine imino resonances in the apical stem may be a sign of disruptive folding in the apical region, most of the upper stem resonances are still present. However, it is clear that the U70G mutation at least affects the lifetime of the apical loop-AG loop interactions, if it does not completely disrupt the folding of the apical stem and its tertiary interactions with the AG loop.

5.4 Discussion

In solution, the molecules are tumbling and moving around, and the folds of biological macromolecules are dynamic to a certain degree. Furthermore, inter- and intra-molecular interactions are actually in a dynamic equilibrium. In this chapter, we determined the structure of a T box apical loop and showed its capability to form tertiary interactions with the AG loop in solution. The serS apical loop forms one structure in isolation and then changes to another structure when interacting with the AG loop. In solution, one can imagine that a fraction of the population of the serS apical loop interchanges very rapidly between the two structures (Figure 5.10). Moreover, since the apical loop is the first completed domain during the transcription of the T box leader, there may be a transient state in which the apical loop folds without the AG loop before transcription continues to complete the other half of the AG loop.

5.4.1 Modeling of the serS apical and AG loop full structure

The glyQS apical loop and AG loop forms a double T-loop structure in the crystal. We suggest that, supported by our NMR data, the serS apical loop and AG loop should do the same. The double T-loop structure is also found in RNase P and ribosomal L1 stalk. However, no NMR structure nor chemical shift data were available for the double T-loop structure. The NMR data we present here for the serS apical loop and AG loop are the first for this class. The double T-loop consists of five and six pairing layers in the two crystal structures respectively. Based on the crystal structure of Grigg et al. (2013), we predicted the hydrogen bond network between the serS apical
and AG loops: serS could at least form 4 out of 5 of the pairing layers (Figure 5.9). In the crystal structure by Grigg et al. (2013), the first pairing layer is a triple with one Watson-Crick G-C pair plus another guanine (Figure 5.11). The AG loop of serS lacks the C and the closing G in the apical loop to form the Watson-Crick G-C pair (Figure 5.12). The two corresponding nucleotides in serS are both adenines. We originally predicted that the A55, A74, and G68 could form a base triple of anti-parallel A-A and parallel G-A as the fourth pairing layer. However, the modeling results show several violations, and the structure is not energetically favorable under these hydrogen bond constraints. By analyzing the 3D model, we found that it is possible for A55 and G68 to form a parallel G-A and U67 and A74 to form a reverse Hoogsteen A-U. By doing so, A74 stacks beneath

Figure 5.7 Base pairings between the apical loop and AG loop.
The color coding in the figure is in consistent with Figure 5.1. The serS apical loop and AG loop could form five pairing layers with two non-canonical base pairs and two base triples as shown in the picture.
**Figure 5.8 Proposed structure for the serS apical loop in conjunction with the AG loop.**
This stereo image shows a simulated structure of the serS double-loop molecule. The modeling result is based on the H-bond constraints shown in Figure 5.7.

**Figure 5.9 Tertiary interactions between serS apical loop and AG loop.**
Above is the stereo close-up of the pairing layers formed between the apical and AG loops. The H-bonds and the labels are according to the colors used in Figure 5.7.
G68 and gives more room for the turning of the phosphate backbone. Indeed, the simulation shows this conformation to be more energetically favorable and has no violation (Figure 5.8).

In the crystal structure of Grigg et al. (2013), the two amino H2s of G69 are inferred to be hydrogen bonded, with one to the O6 of G56 and the other to the O2’ of G54. This configuration also yielded violations in our simulation. According to the geometry in the crystal structure, the two amino H2s of G69 may form a bifurcated hydrogen bond to the O6 of G56, if one of them does not pair with the O2’ of G11. Moreover, as mentioned earlier, the bifurcated H-bond of a G-G mismatch seen in 5S rRNA loop E motif has the same chemical shift. After changing the original constraints to the bifurcated configuration, the simulation indicated no violation and a better result in energy.

The subsequent base interactions are G54-A73-U70 and A53-A57-G71, respectively. The
Figure 5.11 Comparison of \textit{serS} and \textit{glyQS} double-loop structure.

The crystal structure from \textit{glyQS} (blue) (PDB 4JRC; Grigg et al., 2013) and the \textit{serS} structure (green) show some differences. The crystal structure adopts a sharper turn for the apical loop.

\textit{serS} \hspace{3cm} \textit{glyQS}

\begin{align*}
G &\quad G \quad U \\
U &\quad G \\
U &\quad G \\
65G &\quad A \\
G &\quad A \\
A &\quad U \\
A &\quad U \\
G &\quad C \\
G &\quad C \\
G &\quad C \\
A &\quad A \\
A &\quad G \\
A &\quad G \\
A &\quad G \\
U &\quad U 80 \\
\end{align*}

\begin{align*}
G &\quad G \\
68G &\quad G \\
65G &\quad A \\
G &\quad G \\
G &\quad G \\
G &\quad G \\
G &\quad C \\
G &\quad C \\
G &\quad C \\
\end{align*}

Figure 5.12 Sequence comparison of the \textit{serS} and \textit{glyQS} apical loop and AG box.

Several sequence differences are present between the \textit{serS} and \textit{glyQS}. Four nucleotides precede the GNUG (68–71) sequence in \textit{serS} whereas in \textit{glyQS} there is only three. However, both apical loops contain 11 nucleotides because \textit{glyQS} has an additional nucleotide after the GNUG. The \textit{serS} also lacks the G-C closing base pair in the \textit{glyQS} apical loop. One less nucleotide before the GNUG may cause \textit{glyQS} apical loop to adopt a sharper turn as overlayed in Figure 5.11. The A55 in \textit{serS} cannot form the interactions of C55 in \textit{glyQS}. 
fifth pairing layer in the *glyQS* crystal structure consists of an inter-loop G-A pair. In the *serS*, an uridine is at the position of the corresponding adenine. The G71 has the potential of forming a G-U pair with the U52 and becoming the fifth pairing layer. However, the imino resonance of U52 is difficult to detect using NMR because the pair is at the solvent interface of the intra-loop helix.

In G-A pairs where the guanine imino H1 is not hydrogen bonded, such as a sheared G-A or a parallel G-A, the guanine H1s resonate upfield in the region of unpaired guanines (Li *et al.*, 1991) and are protected from solvent exchange thus detectable using $^{15}$N-$^1$H HSQC. The tentatively assigned G54 and G68 fulfills the previous description. The G68 imino H1 of the apical loop is not protected from solvent exchange and thus does not show up in the spectrum. The appearance of the G68 imino H1 resonance in the $^{15}$N-$^1$H HSQC when the AG loop is added implies the inter-loop interactions.

The GNUG sequence in the apical loop is highly conserved among T box leaders and the U70G mutation is known to be inhibiting the basal expression of the regulated gene and its ability to be induced by amino acid starvation *in vivo*. In our comparative results of the wild-type and U70G mutant *serS* apical and AG loops, the mutation, if does not disrupts the whole structural folding, affects the lifetime of the double T-loop structure dramatically so that a few key uridine resonances have been missing.
Chapter 6. Conclusion and future perspective

6.1 Summary of the research

Free ASL\textsubscript{Gly} shows no NMR evidence of the U-turn. In this thesis, three high-resolution ASL\textsubscript{Gly} structures have been solved using NMR spectroscopy, and the result was published in 2012 (Chang and Niknowicz, 2012). The data suggest that different ASL\textsubscript{Gly}s can form five to seven base pairs depending on their sequence. However, all of them do not possess the NMR spectral features seen in the U-turn structure, which is adopted by tRNA in the ribosome and some fully modified ASLs (Robertus \textit{et al.}, 1974; Sundaram \textit{et al.}, 2000; Cabello-villegas \textit{et al.}, 2002).

Base modifications may be required for proper recognition in some T box mRNA leaders. So far, the only successful in vitro T box antitermination system is \textit{glyQS}. All other biochemical and structural studies showing ASL-SD complex formation utilized \textit{glyQS} system. In this study, a hybrid of \textit{tyrS} SD with glycyl codon was shown to form a complex with ASL\textsubscript{Gly}, which ruled out the improper folding of SD and supported that the base modifications on ASL may be the reason for the unsuccessful \textit{tyrS} in vitro termination. Indeed, native \textit{tyrS} SD did not bind to unmodified ASL\textsubscript{Ty} based on the ITC result.

The spectral evidence of the anticodon-codon pairing of the SD-ASL binding was demonstrated and published in 2013 (Chang and Nikonowicz, 2013). The ASL adopts a U-turn configuration when in complex with the SD, supported by the chemical shift of U33. The SD-bound ASL\textsubscript{Gly} shows a U33 chemical shift previously identified in a U-turn structure, but the free ASL\textsubscript{Gly} did not. Therefore, although perhaps weak, the U-turn appears to be the preferred structure for tRNA ASL in T box SD recognition, similar to what found in the ribosome. Base modifications on the tRNA ASLs help to stabilize the U-turn (Sundaram \textit{et al.}, 2000; Cabello-Villegas \textit{et al.}, 2002). Therefore, base modifications may be required for proper anticodon-codon recognition in the T box mechanism, if a U-turn has to be present.
The U70 of the T box leader sequence is important because it forms a reversed Hoogsteen A-U base pair with A73. The T box apical loop contains a conserved sequence of GGUG. The U70G mutation decreases basal gene expression and strips the gene induction by amino acid starvation in vivo (Rollins et al., 1997). Crystal structures have shown that the apical loop interacts with AG loop by a head-to-tail double T-loop structure, which contains multiple base triples (Grigg et al., 2013; Zhang and Ferré-D’Amaré, 2013). Our NMR data in this study are consistent with the crystal structure. The serS apical loop per se did not show spectral features of the full molecule containing both apical loop and AG loop. In the molecule containing both apical loop and AG loop, the U70 showed spectral features of an uridine in reversed Hoogsteen A73-U70 base pair. The U70G mutation not only causes the disappearance of the reversed Hoogsteen spectral feature but also other resonances representing the interaction, suggesting a detrimental structural disruption. Although the double T-loop is held by more than five base-pair layer, lose of this single reverse Hoogsteen A74-U70 base pair can cause the entire structure to fall apart.

### 6.2 Future perspective

An ASL\(^{5’}\)-UCC-3’-SD\(^{5’}\)-GGA-3’ model

As mentioned in Chapter 4, the anticodon-codon interactions in T box are possibly very strict, and the binding of the two domains is solely held by the three base pairs and the stacking of the mini-helix. From our studies, the ASL\(^{\text{Gly,GCC}}\) can pair with TyrS\(_{\text{SD,GGC}}\), but the unmodified ASL\(^{\text{Tyr,GUA}}\) cannot pair with the native TyrS\(_{\text{SD,UAC}}\). To test if two GC and one AU base pairs generates a complex, binding studies can be performed using ASL\(^{\text{Gly,UCC}}\) and tyrS SD hybrid with a GGA codon.

The serS SD is also a potential target for further binding studies. The codon used is the serS is UCC, which should pair with the tRNA\(^{\text{Ser,GGA}}\). The ASL of tRNA\(^{\text{Ser,GGA}}\) contains no base modifications, and therefore can be used directly in the study.
High resolution structure of serS side and apical loop.

In Chapter 5, a simulated model of the interacting serS apical and AG loops is presented. A high resolution structure may be solved in the future. The sequential assignment of the entire double T-loop may be challenging. However, by selectively labelled one or two types of the NTPs, resolution can be improved for overlapped resonances.

Structures of other T box motif

The functions of the stem II and stem III in the T box leader are still unknown. Solution of their structure and comparison with similar RNA motifs may help understanding their roles in the T box mechanism. Stem IIA/B was proposed to form a pseudo-knot structure. Using NMR spectroscopy, one can verify the predicted base pairs and secondary structures.

Structure of T box-acceptor stem complex

The pairing of the T box and discriminating bases on the tRNA acceptor stem bears sequence similarity to the pairing of TyrS_{SDggs} and ASL_{Gly,GCCac}. Therefore, by using the same NMR techniques, the pairing should be able to be verified. Just like for the SD-ASL binding in Chapter 4, whether three or four base pairs are involved in the binding will be the key component to investigate.
Bibliography


