THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENTS OF CHEMISTRY &
BIOCHEMISTRY AND MOLECULAR BIOLOGY

A COMPUTATIONAL AND EXPERIMENTAL APPROACH TO THE
ANALYSIS OF IONIZED NUCLEOBASES

AMARPREET KAUR AHLUWALIA
SPRING 2013

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree in Biochemistry and Molecular Biology
with interdisciplinary honors in Chemistry and Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Philip C. Bevilacqua
Professor of Chemistry
Thesis Supervisor

C.-P. David Tu
Professor of Biochemistry and Molecular Biology
Honors Adviser

Ken Feldman
Professor of Chemistry
Honors Adviser

* Signatures are on file in the Schreyer Honors College.
ABSTRACT

Catalytic RNAs are responsible for essential physiological functions in organisms. For example, leading models for the origin of life suggest that life began with an RNA-World, where RNA played structural, genetic, and catalytic roles. Perturbation of $pK_a$ values can favor a particular ionization state of a nucleobase at biological pH and control the function of the nucleic acid. While Watson-Crick base pairing typically shifts the $pK_a$ of nucleobases away from neutrality, certain bases may have $pK_a$'s that are shifted towards physiological pH. This thesis details various computational and experimental efforts to identify and characterize the presence of such ionized base pairs. A database of known A\(^*\)C wobble pairings was compiled and analyzed. Small DNA oligonucleotides were synthesized to probe the $pK_a$ shift of guanine in the presence of a cationic base utilizing UV-Visible pH dependent titrations. Raman spectroscopy was employed to determine differences in signal intensity from different electrostatic states of guanine. Surface Enhanced Raman Spectroscopy (SERS) was investigated as an alternative to in-solution Raman analysis and Raman crystallography. This thesis lays groundwork for future experiments in elucidating functional and structural roles of ionized bases in nucleic acids.
# TABLE OF CONTENTS

List of Figures ........................................................................................................................................ v

List of Tables ........................................................................................................................................ vii

List of Abbreviations .......................................................................................................................... viii

Acknowledgements ............................................................................................................................... ix

## Chapter 1: Introduction to Thesis ........................................................................................................ 1

I. Background on Nucleic Acids ............................................................................................................. 1
   A. Nucleic Acid Structure ......................................................................................................................... 1
   B. RNA World Hypothesis ......................................................................................................................... 3
   C. Base Pairing ......................................................................................................................................... 3
   D. Higher Order RNA Structure ............................................................................................................... 5

II. Nucleic Acid Function ......................................................................................................................... 5
   A. Nucleic Acid Catalysis ......................................................................................................................... 6
   B. Biological Catalysis (Proteins vs. Nucleic Acids) ............................................................................... 7
   C. pK_a Shifting of Nucleobases ............................................................................................................. 7

III. Topics Covered ................................................................................................................................... 9

IV. References ......................................................................................................................................... 10

## Chapter 2: Towards Identification of RNAs with Shifted pK_a’s: A Bioinformatics Approach ......... 13

I. Abstract ................................................................................................................................................ 13

II. Introduction ......................................................................................................................................... 14

III. Materials and Methods ...................................................................................................................... 15
   A. MC Annotate Package Analysis ......................................................................................................... 16
   B. Removal of Duplicate Structures ..................................................................................................... 16
   C. *In Silico* Protonation of Bases in Visualization Software ................................................................. 17
   D. False Positive Reduction .................................................................................................................... 17

IV. Results and Discussion ...................................................................................................................... 18
   A. Protonation of Mismatches Near Physiological pH .......................................................................... 23
   B. Tandem A^+C Mismatches Flanked by GC Pairs ............................................................................. 24
   C. Base Triple Interactions ..................................................................................................................... 24

V. Future Directions ................................................................................................................................ 25

VI. Acknowledgements ........................................................................................................................... 25

VII. References ........................................................................................................................................ 26

## Chapter 3: The pK_a of Guanine in Various Contexts: Implications for RNA Catalysis ................. 28

I. Abstract ................................................................................................................................................. 28

II. Introduction ......................................................................................................................................... 28

III. Materials and Methods ...................................................................................................................... 30
2. Atomic Force Microscopy Image Processing and Analysis ........................................69
III. Materials and Methods ..........................................................................................70
   A. Preparation of LNA-containing DNA Tether.......................................................70
   B. HDV Ribozyme Construct and Native Gel Analysis .........................................72
   C. Electrochemical Roughening of Gold Wafer .....................................................73
   D. Atomic Force Microscopy Verification of Electrochemical Roughening ............74
   E. Proteinase K Treatment of Wafers .....................................................................75
   F. Wafer Soaking ....................................................................................................75
IV. Results and Discussion ..........................................................................................77
   A. Rationale Behind Design of DNA-based LNA tether .......................................77
   B. Atomic Force Microscopy Verification of Electrochemical Roughening ............78
   C. Native Gel Results ............................................................................................81
   D. Obtained Spectra and Signal Strength ...............................................................82
V. Future Directions ....................................................................................................87
VI. Acknowledgements ...............................................................................................88
VII. References ............................................................................................................88

Appendix A.1: Tabulation of A•C Wobble Mismatch Structural Data .......................93

Appendix A.2: Representative Base Triples ...............................................................98

Appendix B: PDB Structures from Bioinformatics Survey .......................................99
LIST OF FIGURES

Figure 1-1. The Central Dogma of Molecular Biology .................................................................1
Figure 1-2. Nucleobases and Backbone Structure of DNA and RNA .........................................2
Figure 1-3. Canonical Watson-Crick Base Pairing in Nucleic Acids ...........................................4
Figure 1-4. Reactions of the phosphodiester backbone catalyzed by ribozymes ..........................6
Figure 1-5. Standard pK_a Values in Nucleic Acids ....................................................................8
Figure 2-1. A^•C Wobble Base Pair ..........................................................................................14
Figure 2-2. Representative False Positive Structure ....................................................................17
Figure 2-3. Biologically Relevant Structures Solved Near Physiological pH ...............................22
Figure 3-1. Representative Base Pairings ..................................................................................29
Figure 3-2. Adenine versus N1-Methyl Adenine .........................................................................30
Figure 3-3. Mechanism for Hydroxide-Catalyzed Phosphodiester Hydrolysis ............................34
Figure 3-4. 100mM LiCl pH Probe Calibration ............................................................................37
Figure 3-5. GTP Titration .............................................................................................................38
Figure 3-6. GAAAA, AAGAA, GCCCC, CCGCC Titrations .......................................................39
Figure 3-7. GAAAA, AAGAA, GCCCC, CCGCC Titration Curves Superimposed .....................40
Figure 3-8. AAGAA, AA^+GAA, AAGA^+A, AA^+GA^+A Titrations ........................................41
Figure 3-9. AAGAA, AA^+GAA, AAGA^+A, AA^+GA^+A Titration Curves Superimposed ........42
Figure 3-10. Possible Tautomeric Form of N1-meA .................................................................43
Figure 3-11. Schematic Representation of Stacking Interaction Breakage .................................44
Figure 3-12. A•A trans Watson-Crick/Watson-Crick Pairing ....................................................46
Figure 4-1. Self-Cleavage Mechanisms in HDV Ribozyme .........................................................54
Figure 4-2. Schematic of Raman Microscope Setup ....................................................................56
Figure 4-3. Laser and Microscope Focusing in Raman Analysis ................................................57
LIST OF TABLES

Table 2-1. Tabulation of A⁺•C Wobble Mismatches in PDB Structures ........................................19
Table 3-1. Summary of Experimentally Determined pKₐ Values .....................................................36
Table 4-1. Common Raman Stretches in RNA ..............................................................................51
Table 4-2. GTP Concentrations and Raman Signal Strength .........................................................58
Table 4-3. Key Stretches in GTP Identification .............................................................................61
Table 5-1. Literature Raman Stretches in HDV Ribozyme Crystals ..............................................77
Table 5-2. Key Stretches in Tether Identification ..........................................................................83
Table 5-3. Possible Stretches in HDV/Tether Complex Identification ..........................................84
Table A1.1 A⁺•C Wobble Mismatch Structural Survey Results ....................................................93
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM</td>
<td>Atomic Force Microscopy</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge Coupled Device</td>
</tr>
<tr>
<td>ds</td>
<td>double stranded</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>LNA</td>
<td>locked nucleic acid</td>
</tr>
<tr>
<td>LSRR</td>
<td>Localized Surface Plasma Resonance</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleotide triphosphate (any of ATP, CTP, GTP, or UTP)</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>SERS</td>
<td>Surface Enhanced Raman Spectroscopy</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface Plasma Resonance</td>
</tr>
<tr>
<td>ss</td>
<td>single stranded (DNA or RNA)</td>
</tr>
<tr>
<td>TE buffer</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>Tris</td>
<td>trishydroxymethylaminomethane</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>Ultra Violet-Visible</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude for the support I have received through my years of undergraduate research as well as in the preparation of this document. The completion of this thesis would not have been possible without the support, guidance, and unfailing “cautious-optimism” of Dr. Philip Bevilacqua. Phil, you have been both a professional and personal role model for me, and I am very thankful for all you have done for me these past years. With your help, I have become not only a better researcher, but also a more effective educator and science writer. I look forward to taking all I have learned from you as both a researcher and educator to my own chemistry classroom for the next two years, and I hope to eventually be as meticulous and accomplished a scientist in the future.

In addition, I would like to thank all members of the Bevilacqua lab for their support and unwavering patience in helping me troubleshoot my experiments. In particular, I would like to thank my graduate student mentor, Jenny Wilcox, for guiding me through all initial project development and always being willing to read manuscripts and discuss new project directions. Jenny, I know I don’t say it nearly often enough, but thank you so very much for all your help in the lab these past years. But most of all, thanks for forcing me to relax and let go when experiments just would not seem to work out.

I would like to extend a special thanks to Dr. Ken Feldman and Dr. C.-P. David Tu for their time and helpful suggestions and edits to this thesis. Finally, I would like to thank my professors, friends, and family for their continual support. Mom and Dad, you have always shown a great interest in what I have been involved in and have always challenged me to be better than my best. Thank you both for always believing in me, even when I found it hard to believe in myself.


Chapter 1

Introduction to Thesis

I. Background on Nucleic Acids

A. Nucleic Acid Structure

Deoxyribonucleic Acid (DNA) contains the genetic code for all living organisms and is essential for life. However, genomic DNA does not itself direct protein synthesis. Messenger RNA (mRNA) is produced from DNA in the nucleus during transcription and is then transported into the cytoplasm where the genetic message is translated by the ribosome to synthesize proteins within the cell. This overall flow of genetic material from DNA to RNA to protein via the processes of transcription and translation is known as the central dogma of molecular biology (Figure 1-1). Three RNAs are essential for the success of these processes: mRNA, transfer RNA (tRNA), and ribosomal RNA (rRNA).

Figure 1-1: Schematic depiction of the central dogma of molecular biology. DNA is transcribed to form RNA which is then translated into protein using various cellular machinery.

Nucleic acids are large biopolymers that are comprised of nucleotide monomers: adenine (A), guanine (G), cytosine (C), and either thymine (T) or uridine (U), (Figure 1-2A). Bases are attached to a pentose sugar ring and are covalently linked by a phosphodiester backbone. There are two key structural differences between DNA and RNA: 1) DNA contains the nucleobase thymine while RNA contains uracil, and 2) RNA has a 2’ hydroxyl (OH) group on the pentose sugar while DNA has a 2’H (Figure 1-2B).
Figure 1-2: Nucleobases and backbone structure of DNA and RNA. A) Purines: adenine and guanine; pyrimidines: cytosine, thymine, and uracil; Watson-Crick faces in blue, Hoogsteen faces
indicated by red bracket.\textsuperscript{4} \textbf{B)} Phosphodiester backbone of DNA (left) and RNA (right); 2’-OH (RNA) or H (DNA) noted in blue.

B. RNA World Hypothesis

Despite the intermediary role of RNA suggested in the central dogma, leading models for the origin of life suggest that life began with an RNA-World, where RNA (or a closely related sugar) played structural, genetic, and catalytic roles.\textsuperscript{4,6} Fundamentally, this is because DNA has only genetic roles and proteins have only structural and catalytic roles. RNA is typically single stranded, while DNA is double stranded. This key difference greatly affects possible topologies and functions of RNA and DNA. The double stranded nature of DNA restricts its possible structures to primarily standard B-form helices.\textsuperscript{4,7} On the other hand, single-stranded (ss) RNA can form a wide variety of structures, such as hairpins, helices, and pseudoknots. In addition, deoxyribose is the product of an extremely complex metabolic synthesis, requiring free radical chemistry by ribonucleotide reductase.\textsuperscript{2,6} These characteristics have helped support the widely held ideal that RNA, not DNA, was the first carrier of genetic information.

C. Base Pairing

Nucleobases interact with one another primarily through hydrogen bonding interactions on the Watson-Crick (WC) face (Figure 1-2A). Hydrogen bonds are a type of noncovalent interaction of the form X-H\cdots\cdots Y, where X and Y are electronegative donor and acceptor atoms (N, O, or F).\textsuperscript{2,4} A/T (U) and G/C pairings are commonly known as canonical WC base pairs (Figure 1-3).\textsuperscript{8}

The pattern of hydrogen bond donors and acceptors on the WC face of a nucleobase determines which base it will preferentially pair with. For example, adenine has a hydrogen bond donor at N6 and acceptor at N1.\textsuperscript{2} Uracil and thymine, the bases adenine preferentially pairs with, have a complementary acceptor-donor-acceptor (ADA) pattern (Figure 1-3).
Figure 1-3: Canonical Watson-Crick base pairing in RNA. Adenine preferentially binds with uracil in a two hydrogen bond pair. Cytosine preferentially binds with guanine in a three hydrogen bond pair. Hydrogen bond donor sites (amino and imino protons) are marked in blue, while hydrogen bond acceptor sites (carbonyl oxygens and aromatic nitrogens) are marked in red.⁴
D. Higher Order RNA Structure

The overall structure and fold of the RNA is dictated primarily by hydrogen bond schemes and stacking interactions between bases. Stacking is a form of London dispersion interaction between \( \pi \) molecular orbitals of the heterocyclic rings and is strongest amongst purine-purine interactions.

The secondary structure of a nucleic acid refers to base pairing interactions in the biopolymer. While biological DNA tends to exist mostly as a fully paired double helix, ssRNA can form more complex structures. However, since nucleic acids are a highly polyanionic structure, cations associate with the backbone to compensate for unfavorable repulsive charges between the phosphate groups.

II. Nucleic Acid Function

The fundamental role of nucleic acids in biological systems is to carry and transfer genetic information. In the classical view of the central dogma, RNA was thought to simply act as an intermediary molecule in the transfer of genetic material from DNA to protein. However, a paradigm shift occurred in the 1980’s when Thomas Cech and Sidney Altman discovered the existence of catalytic RNAs (ribozymes). The Cech Lab demonstrated that RNA could remove introns in the absence of protein by a process known as ‘self-splicing’, while the Altman Lab showed that the catalytic activity of bacterial RNase P, a ribonucleoprotein (RNP), was found solely in its RNA component.

Since these groundbreaking discoveries, the known functional roles of RNA have greatly expanded. Today, RNA is known to directly regulate gene expression in a biological process called RNA Interference (RNAi). In addition, gene regulation in prokaryotes and eukaryotes is affected by RNA regulatory elements called riboswitches, which are RNA sensors that can detect and respond to small molecules by affecting gene regulation. Recently, structural studies of the ribosome have shown the key catalytic portion of the machinery to be rRNA, not protein.
components. In landmark studies of the ribosome, Harry Noller (Ehrlich & Darmstaeder Prize, 2007; Gairdner Award, 2007), and later, Tom Steitz (Nobel Prize, 2009) showed that the ribosome, is in fact, a ribozyme.\textsuperscript{15-19}

A. Nucleic Acid Catalysis

Ribozymes typically catalyze reactions that either cleave or ligate the phosphodiester backbone (Figure 1-4). This addition/elimination reaction utilizes nucleophilic attack on the bridging backbone phosphate.\textsuperscript{4,20} The nucleophile tends to vary between large and small ribozymes. While large ribozymes (> ~100 nt) tend to use an external nucleophile such as water, small ribozymes (~40-100 nt), such as the hepatitis delta virus (HDV), hairpin, hammerhead, and \textit{glmS} ribozymes, use an internal 2′OH from its RNA base as the nucleophile that performs the attack.\textsuperscript{20,21}

![Figure 1-4: Reactions of the phosphodiester backbone catalyzed by ribozymes. A) General catalytic mechanism of RNA self-cleavage in large ribozymes. An exogenous nucleophile (Nu:) carries out nucleophilic attack on the bridging phosphate. B) General catalytic mechanism of]
RNA self-cleavage in small ribozymes. Without an exogenous nucleophile, a nearby basic species (B:) deprotonates the 2’ OH; general acid and base roles are often carried out by nucleobases.

B. Biological Catalysis (Proteins vs. Nucleic Acids)

Some reactions in biochemistry are spontaneous, i.e. are thermodynamically favorable ($\Delta G <0$). A classic example of such a thermodynamically spontaneous process would be the complete oxidation of glucose to carbon dioxide and water.$^{2,4-5}$ However, the spontaneous nature of such reactions does not mean that they proceed rapidly. A certain amount of energy (the activation energy) must be put into the system in order for the reaction to proceed, and enzymes catalyze reactions by lowering this activation energy barrier.$^5$

Almost all biological enzymes are proteins. With 20 diverse amino acids and a variety of functionalities in the side chains, proteins can utilize a variety of strategies (covalent, general acid-base, electrostatic, and metal ion catalysis, and catalysis by approximation) to catalyze the reactions of life.$^{1,2,5}$ In contrast, nucleic acids have only four chemically similar side chains (nucleobases) that primarily interact with one another through complementary base pairing. Regardless, RNA is capable of effectively catalyzing reactions in a variety of systems through physical changes (folding) and some chemical changes.$^{22-24}$

C. pH$_a$ Shifting in Nucleobases

Optimal enzyme activity in general acid-base catalysis requires side chains to have pH$_a$ values near the pH of the reaction environment. For an amino acid side chain or nucleobase to optimally function as a general acid or general base, the intrinsic pH$_a$ value must be near the reaction-site pH, allowing the functional protonation state to be regenerated for catalytic turnover.$^{25}$ As such, a key factor in biological catalysis is the tuning of side chain pH$_a$ values to near-neutral physiological pH. Nucleic acids have unshifted pH$_a$ values that are far from
neutrality (Figure 1-5A).\textsuperscript{4,5} In addition, base pair formation favors a particular protonation state of the nucleobase, which typically shifts $pK_a$ values even farther from neutrality (Figure 1-5B).\textsuperscript{4}

It is important to note that while ionization of the bases does not occur in traditional Watson-Crick pairing, it can occur in non-canonical pairings, such as A$^+$•C wobbles. In these cases, the protonation or deprotonation of the base results in the formation of a new structural motif via base pairing, in effect overcoming the energy penalty needed to charge the bases.\textsuperscript{22-24,26}

\textbf{Figure 1-5:} Standard $pK_a$ Values in Nucleic Acids. \textbf{A.} $pK_a$ values of the backbone and sugar are farthest from neutrality, suggesting that the bases themselves are likely sources of potential general acid-bases for catalysis. \textbf{B.} Canonical GC base pair, $pK_a$ shifts due to formation of secondary structure labeled. The extent of $pK_a$ shift depends on reaction site microenvironment.
III. Outline of Topics Covered

This portion of the introduction outlines subsequent chapters of the thesis and includes a small explanation of each study. Chapter 2 deals with cationic bases and is a bioinformatics survey of protonated A•C wobbles base pairs in the Protein Data Base (PDB). Chapter 3 deals with anionic bases and studies pKₐ shifting of guanine in the presence of cationic adenine bases using small synthetic oligonucleotides and UV-visible absorbance detection. Chapters 4 and 5 describe efforts in methodological development of Raman spectroscopy in an effort to probe the ionization state of a single nucleobase within a structure. Chapter 4 details efforts to optimize in-solution analysis of GTP, and Chapter 5 details development of surface enhanced Raman (SERS) analysis of nucleic acids using the HDV ribozyme and an LNA/DNA tether.


This chapter used Monte Carlo (MC) package software and molecular visualization software to analyze base pairing motifs found in RNA structures within the PDB. Details on selection of candidate structures, elimination of false positive candidates, and general methodologies in the development of the database are included.

Chapter 3: The pKₐ of Guanines in Various Contexts: Implications for RNA Catalysis

Synthetic model oligonucleotides were designed in an effort to create an electrostatically beneficial environment for anionic guanine. These oligonucleotides were analyzed by pH dependent UV-visible titrations and pKₐ values of the oligonucleotides were found after correcting for the effect of the presence of various ions on the pH probe.
Chapter 4: *In-Solution Raman Detection of Guanyl Anion Using Model Compound GTP*

In order to gain a virtual “snapshot” of the ionization state of a particular base within a larger structure, Raman spectroscopy was employed. This chapter details methodology in optimization of the use of Raman analysis for in-solution nucleic acid samples.

Chapter 5: *Surface Enhanced Raman Spectroscopy (SERS) Analysis of Nucleic Acids*

This chapter details efforts to increase signal intensity in nucleic acid analysis via Raman spectroscopy. A large RNA structure – the hepatitis delta virus (HDV) ribozyme was selected and annealed to a DNA based tether. This tether contained a 5’ thiol modification allowing for immobilization of the tether/HDV complex on roughened gold for analysis.

IV. **References**


Chapter 2

Towards Identification of RNA Motifs with Shifted pK_a’s: A Bioinformatics Approach


I. Abstract

Charged amino acids make critical contributions to biological catalysis in protein-based enzymes. Similarly, ionized nucleobases play key roles in RNA catalysis as general acid-bases. Advances in spectroscopic techniques have shown pK_a values of adenine and cytosine near neutrality. This bioinformatics study was a structural survey of protonated A▼C wobble mismatches within previously solved RNA structures. Analysis of features such as spatial orientation of bases, bond lengths, and torsion angles within these structures indicates that protonation of bases is not difficult to achieve at physiological pH.
II. Introduction

Ionized bases can play a key role in general acid-base catalysis and can promote conformational switching at near neutral pH.\(^2\)\(^3\) The protonation of an adenine can allow for the electrostatic stabilization of catalytic RNA molecules, while also affording a stable two-hydrogen bond wobble base pair (Figure 2.1). As such, A⁺•C wobble base pairings have a relatively minor effect on A-form helix geometry of dsRNA sections.\(^1\)\(^3\) Recent studies have shown that protonated A⁺•C wobbles can exist in DNA under biologically relevant conditions. Under the most favorable temperature and salt conditions, a \(pK_a\) of 8.0 was observed for the wobble pairing – a shift of 4.5 pH units from the normal \(pK_a\) of single stranded adenine.\(^1\)

The number of RNA structures in structural databases such as the Protein Data Bank (PDB, www.rcsb.org) has risen dramatically over the past decade, allowing for systematic survey and analysis of tertiary structure. While ionized bases exist in nature and have been found in solved structures, databases containing non-canonical pairings are updated infrequently, and as such, are often missing newly solved structures.\(^3\) Therefore, I undertook a systematic probing of structures found within the PDB with regards to experimental conditions and quality of protonated base pair in an effort to gain insight to the role these ionized bases play in natural systems.

The present study first examined RNA and RNA/protein structures within existing collections for A⁺•C wobble pairs: Fox Lab Non-Canonical Base Pair Database (http://prion.bchs.uh.edu/bp_type/) and the Major Lab Base Pair Classification page (http://www-lbit.iro.umontreal.ca/mcannotate/base_pairs/). Next, candidate structures from the PDB were manually searched through for the presence of A⁺•C wobbles. Structures were analyzed with MC

![Figure 2-1: A⁺•C wobble base pair.](image-url)
Annotate and visually inspected in Mac PyMol and DSVisualizer Pro for bona fide A•C wobbles according to the criteria provided below.

II. **Materials and Methods**

Non-canonical Base Pair Database (Fox Lab, University of Houston) and Base Pair Classification page (Major Lab, University of Montreal) were an initial source of information on A•C occurrences. The Fox Lab database is a collection of non-canonical pairings in which protonation states were identified through characteristic NMR signatures and corroborating UV melting curves. The Major Lab classification page was computed using MC Annotate and MC Search only. While protons cannot be directly seen in crystal structures, *in silico* protonation of candidate structures was employed for this study.

PDB files of structures reported in these databases were analyzed using PyMol and DS Visualizer, 3D imaging software packages. MC Annotate, a program that automatically classifies RNA tertiary structure, was also employed. The A•C wobble pairs identified through a PDB file using PyMol are inferred pairings based on measuring two hydrogen bond distances and a bond angle (N3 of cytosine – Hydrogen – N6 of adenine) between corresponding heteroatoms in the two hydrogen bonds. MC-Annotate coding for positively identified A•C wobble pairs was used as a search parameter within the MC-Search software package. This automated search method yielded some false positives (see Figure 2-2), so manual searching of the PDB using keywords like “ribozyme, riboswitch” and other structure names was employed. In addition, to ensure newest structures were found, the search result by structure name was further reduced by year of upload and resolution.

Candidate structures were downloaded from the Protein Data Base zipped as .pdb.gz files and analyzed by the online MC Annotate server to search for the presence of an A•C wobble pair. Manual inspection of the structure for planarity of bases and orientation of the adenine and
cytosine was done using Mac PyMol (Schrodinger, LLC) or DSVisualizer (Accelrys, Inc) and served to confirm MC Annotate output. Torsion angles and bond distances were measured using Mac PyMol and DSVisualizer Pro.

A. MC Annotate Package Analysis:

MC Annotate was chosen for ease in automatic recognition of motifs of interest. The package employs a three step method to calculate likelihood of base protonation by: (i) calculating the probability of hydrogen bonds between a pair of donor and acceptor groups, (ii) accounting for competing donors and acceptors within the pair, and (iii) assigning a type of base pair interaction to the atoms in question. After extrapolating probabilities from all pairs of donor and acceptor groups, a resultant dataset is reduced to a single vector defining distance and angle of the hydrogen bond, and accounts for distortion of the angle caused by a lone pair.

B. Removal of Duplicate Structures

The PDB often contains multiple entries for the same sequence. For example, a search of “30S ribosomal subunit” in the PDB yields 442 structure hits. Many of these structures are crystallization studies of the structure of interest bound to a different antibiotic. As such, to ensure no sequences were counted more than once in this survey, a single representative sequence was selected from a group of structures solved for that sequence at the same experimental conditions (pH and temperature). The representative structure was chosen using a structure that seemed to hold the greatest biological relevance (i.e. close to physiological pH, close to a wild-type sequence of a structure), a method previously employed by Dr. Josh Sokoloski in our lab. A single representative structure was chosen for each experimental pH value a structure was solved at, allowing for comparison of sequence and nearest neighbor interactions between A•C wobbles that formed at near physiological pH versus ones that formed simply because of a fairly acidic environment.
C. *In Silico* Protonation of Bases in Visualization Software

In order to protonate adenine at N1 PyMol, the formal charge on the molecule must be increased prior to manual addition of the hydrogen atom. Using the “atom” selection option in 3-button-viewing mode, the N1 atom of adenine was selected. Upon selection of N1, pk1 (“pick 1” atom) selection was activated. At this point, the formal charge on the molecule was increased by selecting “Build” → “Make pk1 Positive.” Finally, a hydrogen atom was added only to the base in question using the command prompt “h_add sele.”

D. False Positive Reduction:

While the MC Annotate and Search package allowed for fast searching of a vast variety of structures within the PDB, false positive structures were sometimes encountered and were removed upon manual inspection. These false positives were often easily recognizable during initial orientation viewing in PyMol (Figure 2-2).

![Figure 2-2: Representative False Positive Structure. PDB ID: 2CT8. This structure was recognized by MC Search as containing an A•C wobble base pair (indicated in yellow). However, as seen above, the orientation of the adenine is not suitable for the wobble interaction.](image-url)
While most false positives were easily eliminated by visual inspection, distances and angles of possible hydrogen bonds were calculated for all structures, with bond lengths of \( \text{C(N3)•••A(N6)} \) and \( \text{C(O2)-A(N1)} \) measured. If these bond lengths were \( \leq 3.4\text{Å} \), and bond angles were between 140°-180°, the pair was scored as protonated. As seen in Equation 1.1 below, B-factor measures the total vibrational displacement of the atom around the position assigned in a three-dimensional model as a function of total atom displacement, \( \mu_j \). For crystal structures, mean temperature factor \( B_j \) (B-factor) values were calculated for the entire molecule (\( B_{\text{total}} \)) and the bases (\( B_{\text{base}} \)) in question.

\[
B_j = 8\pi^2(\mu_j^2) = 79(\mu_j^2)
\]

(1.1)

\( B_{\text{base}} \) values were then compared to \( B_{\text{total}} \) values. If \( B_{\text{base}} \) values were > 40, the pair was not scored as a protonated base. In addition, \( B_{\text{base}} \) values should be considered in comparison to that of the overall structure. While not seen in analysis of any structures in this survey, if \( B_{\text{base}} \) values are significantly greater than \( B_{\text{total}} \), the pair would not have been scored as a protonated base. Most strong candidates for protonation contained bond angles that fell between 140-180°. However, some candidates had bond angles lower than this (110°-139°), but bond lengths were \( \leq 3.0\text{Å} \) and the pH was relatively acidic. If the pH was below 6, the base was scored as a protonated pair.

IV. Results and Discussion

We found 57 unique occurrences of A••C wobbles in 36 crystal or NMR structures consistent with the above criteria for protonation of the A. Of these, 54 motifs were isolated A••C pairs and three were base triples involving an A••C base pair. For those structure determinations that reported experimental pH, half had a pH in the range of 6.5–7.5, suggesting a \( pK_a \) near biological pH. Representative biologically relevant structures solved near physiological pH are
shown in Figure 2-3. Results of this survey are summarized in Table 1. Full results of the survey, including structure description, sequence context, bond lengths, bond angles, B-factor data, and experimental conditions are listed in Appendix A-1. Reference structures for base triples are listed in Appendix A-2.

Three points of special interest were noted: 1) Protonation of mismatches near physiological pH, 2) Tandem protonated A•C mismatches flanked by GC pairs, and 3) Stabilizing triple interactions as part of or flanking the mismatch. Altogether, this analysis suggests that charging of the bases into a cationic state is not uncommon.

Table 1.1: Tabulation of A•C Wobble Mismatches in PDB Structures

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Structure Description</th>
<th>Number of A•C Mismatches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1BZ2</td>
<td>tRNA Anticodon Loop</td>
<td>1</td>
</tr>
<tr>
<td>1E4P</td>
<td>VSm1 hairpin analog</td>
<td>1</td>
</tr>
<tr>
<td>1EUY</td>
<td>tRNA Synthetase</td>
<td>1</td>
</tr>
<tr>
<td>1FEQ</td>
<td>tRNA Lys3 UUU Anticodon Stem / Loop</td>
<td>1</td>
</tr>
<tr>
<td>1FG0</td>
<td>Large ribosomal subunit + 13 bp minihelix puromycin compound</td>
<td>1</td>
</tr>
<tr>
<td>1FJG</td>
<td>30S Ribosomal Subunit (Complexed with Antibiotics)</td>
<td>1</td>
</tr>
<tr>
<td>1GTR</td>
<td>Anticodon loop recognition by glutaminyl-trna synthetase</td>
<td>1</td>
</tr>
<tr>
<td>1HWQ</td>
<td>VS Ribozyme substrate</td>
<td>1</td>
</tr>
<tr>
<td>1LC6</td>
<td>U6 stem loop RNA</td>
<td>1</td>
</tr>
<tr>
<td>1LDZ</td>
<td>Leadzyme RNA</td>
<td>1</td>
</tr>
<tr>
<td>1NKW</td>
<td>Dienococcus Radiodurans 23S RNA</td>
<td>1</td>
</tr>
<tr>
<td>1O0B</td>
<td>Glutaminyl-tRNA synthetase</td>
<td>1</td>
</tr>
<tr>
<td>PDB Code</td>
<td>Description</td>
<td>Resolution</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>1QRS</td>
<td>Glutaminyl-tRNA synthetase Mutant D235N</td>
<td>1</td>
</tr>
<tr>
<td>1QTQ</td>
<td>Glutaminyl-tRNA synthetase tRNA/Cys complex</td>
<td>1</td>
</tr>
<tr>
<td>1U8D</td>
<td>Guanine riboswitch (hypoxanthine complex)</td>
<td>1</td>
</tr>
<tr>
<td>1U9S</td>
<td>A-type Ribonuclease P</td>
<td>1</td>
</tr>
<tr>
<td>1XM9</td>
<td>t6A37-ASLLysUUU AAA-mRNA Bound to the Decoding Center</td>
<td>1</td>
</tr>
<tr>
<td>1Y26</td>
<td><em>B. subtilis</em> Guanine Riboswitch (Guanine complex)</td>
<td>1</td>
</tr>
<tr>
<td>2BTE</td>
<td>Leucyl-tRNA Synthetase</td>
<td>1</td>
</tr>
<tr>
<td>1EHT</td>
<td>Theophylline-RNA Complex</td>
<td>1</td>
</tr>
<tr>
<td>405D</td>
<td>16bp RNA duplex (pH 7.0)</td>
<td>1</td>
</tr>
<tr>
<td>1L9A</td>
<td>Archae SRP19 + SRP RNA Complex</td>
<td>1</td>
</tr>
<tr>
<td>1FFK*</td>
<td>50S ribosomal subunit (2.4Å Resolution, pH 5.8)</td>
<td>2</td>
</tr>
<tr>
<td>1GID</td>
<td>Group I Ribozyme Domain</td>
<td>2</td>
</tr>
<tr>
<td>1JID</td>
<td>Human SRP19 + SRP RNA Complex</td>
<td>2</td>
</tr>
<tr>
<td>1KFO</td>
<td>RNA Helix zinc-finger protein recognition site</td>
<td>2</td>
</tr>
<tr>
<td>1L1W</td>
<td>SRP19 Splicing Domain</td>
<td>2</td>
</tr>
<tr>
<td>1M9O*</td>
<td>50S ribosomal subunit (pH 6.2)</td>
<td>2</td>
</tr>
<tr>
<td>1S72*</td>
<td>50S ribosomal subunit (pH 7.5)</td>
<td>2</td>
</tr>
<tr>
<td>1Z7F</td>
<td>2’-amine substituted RNA</td>
<td>2</td>
</tr>
<tr>
<td>402D</td>
<td>16bp RNA duplex (pH 5.5)</td>
<td>2</td>
</tr>
<tr>
<td>1D4R</td>
<td>Helix 6; Human SRP</td>
<td>3</td>
</tr>
<tr>
<td>1N32</td>
<td>30S ribosomal subunit</td>
<td>3</td>
</tr>
<tr>
<td>1VQN*</td>
<td>50S ribosomal subunit (no reported pH)</td>
<td>3</td>
</tr>
<tr>
<td>PDB Code</td>
<td>Description</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>1NJI^</td>
<td>50S ribosomal subunit (pH 5.8)</td>
<td>4</td>
</tr>
<tr>
<td>1YI2^</td>
<td>50S ribosomal subunit (pH 5.7)</td>
<td>4</td>
</tr>
</tbody>
</table>

^These structures are the most recently solved within a larger subset of PDB files.
Figure 2-3: Representative biologically relevant structures solved near physiological pH; A•C base pair shown in red.  
A. U6 snRNA stem loop; PDB ID: 1LC6.  
B. SRP19 + SRP RNA complex; PDB ID: 1L9A.  
C. Guanine riboswitch; PDB ID: 1U8D.  
D. Glutaminy-l-tRNA Synthetase; PDB ID: 1O0B.
A. Protonation of Mismatches Near Physiological pH

Of the 29 structures surveyed that reported an experimental pH, 16 had a pH range between 6.5-7.5. The concept that nucleobases can ionize at physiological pH lends support to their existence in biology and participation as general acids or bases in nucleic acid catalysis. The structures solved at physiological pH were varied and included tRNA anticodon loops, U6 stem loop RNA, and various riboswitches.

Of the 16 structures reported as solved within the pH range of 6.5-7.5, 9 contained a GC pair immediately above or below the protonated A•C. As GC pairs have three contiguous hydrogen bonds, they are fairly strong, so these GC pairs likely anchor the helix containing the protonated mismatch. Surprisingly, three structures (1BZ2, 1Z7F, 405D) contained only flanking AU pairings to the protonated A•C.

Structure 1BZ2 was an NMR structure of the anticodon loop of tRNA<sup>Lys</sup>. Researchers found a drastic increase in melting temperature (T<sub>m</sub>) of the tRNA when the pH was lowered to 5 due to the presence of an A•C base pair. The formation of this protonated base pair in turn induces the formation of a dynamic AU pair, which shortens the normally seven-nucleotide anticodon loop to a three-nucleotide loop.<sup>9</sup>

Structure 1Z7F is a crystal structure of a 16 base pair RNA duplex. In this study, Weeks and coworkers found that 2'-amino cytidine (C<sup>N</sup>) nucleotides are easily accommodated in the helix as either a canonical C<sup>N</sup>G pairing or non-canonical A•C<sup>N</sup> pairing.<sup>10</sup> In both aforementioned cases, more easily broken AU nearest neighbors would allow for a more dynamic structure upon ionization of the adenine base.

Structure 405D was the crystal structure of a 16-mer RNA duplex solved at pH 7.<sup>9a</sup> Researchers found angles λ<sub>1</sub> [N1(C)-C1'(C)-C1'(A)] and λ<sub>2</sub> [N9(A)-C1'(A)-C1'(C)] in the A•C mismatches similar to those of GU wobbles in different environments and suggested the importance of the AU pairing in this. The 5' stacking of the AU pairing has a low twist angle...
(27˚) and high rise (3.3Å), while the 3' AU nearest neighbor has a higher twist angle (36˚) and lower rise (2.9Å). It is thought that this relation in twist and rise due to the AU pairs helps give the duplex more freedom to bend.\textsuperscript{11}

B. **Tandem A⁺•C Mismatches Flanked by GC Pairs**

Tandem A⁺•C mismatches flanked by the GC pairings were found in the Signal Recognition Particle (SRP) + SRP RNA complex (Structure 1L9A) as well as an RNA loop at pH 5.5 (Structure 402D). The SRP is a ribonucleoprotein (RNP) that mediates protein targeting to cellular membranes, and SRP19 is one of two proteins that organizes its core.\textsuperscript{12,13}

While the tandem protonated A⁺•C wobbles are not directly involved in binding of SRP RNA to SRP19, it is thought that the presence of non-Watson-Crick pairings like the tandem A⁺•C wobbles and a G•U wobble together help widen the major groove of SRP RNA. This widened major groove is one of the key features recognized by SRP19 as it prepares to bind SRP RNA.\textsuperscript{13} These tandem A⁺•C wobbles are flanked by GC Watson-Crick pairings, which are likely needed to better anchor the widened helix, and prevent greater distortion due to repulsion of two adjacent positive charges. As such, the presence of GC Watson-Crick flanking pairs is unsurprising, when considering that RNA bends as large as 21˚ have been observed at tandem protonation sites.\textsuperscript{14}

C. **Base Triple Interactions**

Base triples were found involving a protonated A⁺•C base pair or flanking the protonated base pair. Stabilizing interactions such as base stacking or a more extensive hydrogen bonding network can help compensate for a possible energetic penalty in ionization of the nucleobase itself. This aforementioned increase in overall stability of larger structures due to the formation of an ionized wobble base pair may explain why protonated A⁺•C wobble pairs can be found in larger structures.
V. **Future Directions**

This method of categorization and annotation of specific bases including MC Annotate package analysis, manual searching through the PDB, and structure visualization and analysis, can be applied to identify other non-canonical pairings, like GU wobbles, in biological systems. Then, to better understand the role the specific bases play in functional RNAs, biochemical assays that specifically target these bases can be designed and employed. Another possible online server is the RNA Characterization of Secondary Structure Motifs (RNA CoSSMos), which allows search for mismatches and loops in the PDB by nearest neighbors.\(^\text{15}\)

However, a key problem in utilization of automated analysis of PDB structures is the number of false positives found by the software. As such, development of a similar methodology for the systematic probing and compilation of structural data is essential to successfully reduce the large size of the initial numbers of sample structures.

Structural data from these databases will continue to play a large role in the development of better scripts and coding parameters for the possibility of designing a self-training and correcting software package. Ultimately, this method of characterization would lead to better recognition of motifs of interest and concurrent reduction of the number of false positives or negatives within the data set. This application to software like MC Search can then be used to automatically update existing databases of non-canonical pairings.

VI. **Acknowledgements**

I would like to thank Marc-Frédéric Blanchet (Major Lab) for all of his assistance in troubleshooting MC Annotate server issues and help in designing preliminary coding parameters for MC Search based on annotated results. In addition, I would like to thank James Watney (Schrodinger, LLC) for all of his help and patience in designing the *in silico* base protonation protocol for large nucleic acid structures.
VII. References


Chapter 3

The $pK_a$ of Guanines in Monomeric and Oligomeric Contexts:

Implications for RNA Catalysis

I. Abstract

As discussed in Chapter 1, Watson-Crick base pairing of nucleobases typically shifts $pK_a$ values farther from neutrality. However, ionized mismatches can occasionally form at neutral pH and have the potential to aid in general acid-base catalysis. While the presence of cationic base pairs like the A$^+$$\cdot$C wobble is commonly accepted, as presented in Chapter 2 of this thesis, the presence of an anionic base at physiological pH is hotly contested. Guanine is properly positioned in the hammerhead, glmS, hairpin and VS ribozymes to deprotonate at N1 and act as a general base, yet positioning alone does not prove ionization. Of these ribozymes, the hairpin has the potential for simultaneous protonation of a nearby adenine to stabilize the anionic guanine. My study in this chapter uses small model sequences to probe for the $pK_a$ shifting of guanine in the presence of N1-meA.

II. Introduction

Unlike amino acids, nucleobases typically exist in an uncharged form at biological pH.\textsuperscript{1} However, protonated bases can occasionally form and play an important role in biologically relevant structures and have $pK_a$ values shifted towards neutrality. This $pK_a$ shift towards neutrality can lead to cationic base formation of adenine and cytosine and anionic base formation of guanine and uracil.\textsuperscript{2} A better understanding of driving forces behind $pK_a$ shifting could
provide a means to predict possible roles the bases may play in nucleic acid catalysis and roles of bases in uncharacterized systems.

Although ionization of bases in Watson-Crick base pairing is typically disfavored, non-canonical base pairs containing ionized bases have been found. Among protonated bases, A•C wobble mismatches are found to form at physiological pH with relatively minor impact on helix geometry (Chapter 2). While the presence of such protonated cationic bases is now commonly accepted, the possibility for anionic bases is contested. This is because adding negative charge to a highly negatively charged structure seems improbable due to charge-charge repulsion. However, there could be factors that favor deprotonation of a nucleobase. For example, the electrostatic stabilization of an anionic base by a nearby cationic base, or binding of a metal ion to the base or phosphate, could help stabilize these negative charges. Cadmium ion, as well as other transition metals such as Mn$^{2+}$, has been shown to bind N7 of guanine, which may help lower the guanine pK$_a$ by stabilizing the anionic form of the base. In addition non-Watson-Crick base pairing could stabilize the anionic base. Figure 3-1B highlights a unique, hypothetical Watson-Crick bonding scheme containing a deprotonated, anionic guanine and compares it to a GC Watson-Crick pair.

**Figure 3-1:** Representative base pair schemes. **A.** Structure of stable GC base pair with three, contiguous hydrogen bonds. **B.** Schematic of G•G$^-$ base pair. In the deprotonated state, this pairing also has three contiguous hydrogen bonds although it requires one of the Gs to be in the syn conformation.
An understanding of the electrostatic state of a nucleobase allows for better elucidation of possible roles that base may play in catalysis. Of particular interest is general acid-base catalysis within self-cleaving ribozymes.\textsuperscript{7,8} The hammerhead, glmS, VS, and hairpin ribozymes all have a guanine positioned to act as a general base via deprotonation at the N1 position.\textsuperscript{8-11} Of these ribozymes, the hairpin also has the potential for protonation of a nearby adenine that could in principle electrostatically stabilize the formation of anionic guanine.\textsuperscript{10} To determine the pK\textsubscript{a} of a nucleobase near a protonated base, we designed 5-mer synthetic oligonucleotides using adenine, guanine, and N1-methyl adenine (N1-meA) bases (Figure 3-2). A methylated adenine was chosen in an effort to ensure a positive charge remained with the base over the pH range studied, although we later came to realize that this base likely ionizes over this region as well based on comments from Dr. Feldman during the oral defense (see below). pH dependent UV-visible titrations were then employed to determine the pK\textsubscript{a} values for the synthetic oligonucleotides.

![Figure 3-2: Protonated adenine versus N1-methyl adenine (N1-meA).](image)

III. Materials and Methods

pH dependent titrations from approximately pH 8.0-12.25 were employed using ultraviolet visible (UV-Vis) spectroscopy to determine the pK\textsubscript{a} of the oligonucleotides. DNA
oligonucleotides were used for these experiments since RNA degrades at high pH, which would have made the titrations less reliable. UV-Vis data were collected from 220-320 nm. The greatest change in absorbance with pH was seen at 260 nm. As such, all representative curves collected in this chapter use absorption data from 260 nm.

A. Preparation of DNA oligonucleotides

In an effort to probe the effects of the nearby cationic base (N1-methyl adenine) on the ability of guanine to deprotonate, the following synthetic DNA oligonucleotides were synthesized at the Penn State DNA sequencing facility: 5’-AAGAA-3’, A(A+)GAA, AAG(A+)A, and A(A+)G(A+)A (where ‘A+’ represents N1-meA). Since the \( pK_a \) of guanine is greater than 9.0, all sequences were designed with one guanine and no thymine or other guanines to ensure only a single \( pK_a \) change was being tracked through the titration. To probe sequence context effects and stacking interactions of charge neutral adenine, the following synthetic DNA oligonucleotides were manufactured by Integrated DNA Technologies (IDT®; Coralville, Iowa): GAAAA, AAGAA, GCCCCC, and CCGCC. IDT® provided mass spectrometry data for these oligonucleotides; as such, these nucleotides were not prepared for mass spectrometric analysis, although efforts are ongoing to characterize the others (see below).

To favor presence of a single cationic counterion, oligonucleotides were subjected to salt exchange and desalting by the following procedure: 1) Dialysis using a 1000 Da MWCO membrane against 100 mM of KCl, 2) Dialysis against sterile water. Dialysis was performed for 6 hours against KCl in an eight-well microdialysis system attached to a pump set at a flow rate of 2mL/min.

B. Ionic Strength Effects on pH Probe Measurements

A stainless steel, IFSET NMR Probe (Hach Chemical) containing a KCl core and pH meter (IQ150 meter, IQ Scientific Instruments) were used to determine the pH of samples during
titration. In the presence of weak acid or base, ionic strength will affect the proton and hydroxide ion concentration in solution and alter the detectable pH of the solution.

To determine the actual pH of the sample solution, the pH probe was calibrated using a series of standards generated from dilute solutions of strong acids (HCl) and bases (KOH, LiOH) in 100 mM KCl or LiCl was prepared and meter response to the solutions was measured. The equation of the resulting linear calibration curve was used to adjust the experimentally determined pH of samples from the titration. K⁺ ions were observed to have no significant impact on pH measurement, which agreed with previous observations by Dr. Nathan Siegfried.² LiCl, however, did impact experimental pH measurements slightly at pH >11.5 and needed to be corrected. Incorrectly low meter readings at pH >11.5 led to unexplained increases in absorbance at pH >12.5. Application of the correction from probe calibration allowed for the titration to be stopped at the proper pH and prevent this increase beyond upper baseline for all subsequent titrations.

C. Probe Cleaning and Shakedown Procedures

When working with unbuffered solutions in alkaline pH for long periods of time, the IFSET pH probe must be regularly cleaned and gently shaken to resettle the KCl core. If the probe is not regularly cleaned and shaken down, it may begin to respond slowly, drift during pH readings, and/or calibrations may fail. General cleaning followed by probe shakedown should be performed every three titrations. Also, before and after all titrations, the probe must be rinsed well with distilled water.

1. General Cleaning Procedure

The probe was washed in ~50 mL of warm purified water containing a small amount of Alconox soap. The probe was then rinsed for approximately 30 seconds with distilled water and soaked in fresh pH 7 buffer for ~30 minutes. No calibrations were performed during this time period.
2. Probe Shakedown

After the 30-minute soak in fresh pH 7 buffer, the probe was removed from the buffer and gently dried with a KimWipe, then capped. Holding the probe horizontally, the probe was shaken vigorously up and down for about 5 seconds, and then held vertically so the gel core could resettle.

D. Titrations (GTP and 5-mer oligonucleotides)

All titrations were performed in a 3 mL quartz cuvette on a Beckman Coulter DU 650 Spectrophotometer. No buffer was present during dialysis or trials to allow easier changing of pH. Guanosine triphosphate (GTP, 10 µM) in 100 mM LiCl was used as a control sample. (Lithium salts needed to be used for GTP titrations to disfavor the formation of G-quadruplexes during the titration.) Synthetic oligonucleotides (10 µM) were titrated in 100 mM KCl since K⁺ was shown to have no significant impact on pH measurement, which thus allowed us to avoid a meter correction.

Small amounts (0.25 µL-6.50 µL) of 0.01 M-1.0 M base (LiOH or KOH) were added to the cuvette and the sample was mixed using a 1000 µL pipette. The total volume of base added was less than 1% of the total sample volume, so no volumetric corrections were needed following data analysis.² The pH was taken before and after data collection, then averaged to ensure accurate meter reading of sample pH. These values were generally within 0.05 pH units. For GTP, the averaged pH value was then adjusted using the LiCl calibration to determine the true pH of the sample. For all remaining titrations, the average meter reading served as sample pH without adjustment since KCl did not show significant impact on pH measurement.

Plots of absorbance (A) at 260 nm versus pH were fit to equation 1 to obtain the pKₐ. A₀ and Aₚ represent the absorbance values of the unprotonated and protonated states, respectively and n_Hill represents the Hill coefficient. The Hill coefficient serves as an upper limit to the number of proton binding sites, and equation 3-1 assumes an all-or-none model if multiple sites
are present. Hill coefficients were found to be between 0.987-0.997 for titrations. All representative titration curves below have \( n_{\text{Hill}} \) values fixed to 1.00 and experimentally determined \( pK_a \) values are listed in Table 3-1.

\[
A = A_P + (A_{U} - A_P) \frac{1}{1 + 10^{n_{\text{Hill}}(pK_a - pH)}}
\]  

(3-1)

IV. Results and Discussion

A. Rationale Behind Design of Oligonucleotide Sequences

Small nucleic acid systems often have similar interactions to large nucleic acid systems. As such, thermodynamic parameters can be determined from small systems and the same constraints can often later be applied to larger ones. Since the \( pK_a \) of guanine is greater than 9.0, and titrations were run from approximately pH 8-12, DNA oligonucleotides were used instead of RNA to protect against denaturing of the oligonucleotide during titrations (Figure 3-3).¹

Figure 3-3: Mechanism for hydroxide-catalyzed phosphodiester hydrolysis. The presence of a 2′-OH in RNA allows for deprotonation by hydroxide, ultimately producing a cyclic 2′,3′-phosphate on one base and a free 5′-OH on the other.¹
Additionally, base stacking reflects “vertical” interactions between bases along a single strand of a nucleic acid. Since adenine is a strong stacker, these interactions should help add stability to the system. Finally, while the pKₐ of unpaired guanine is usually 9.5, the pKₐ of all other bases in the oligonucleotide sequences are ~4.0. As such, all changes in absorbance should be due to changes in the protonation state of guanine.

B. Determination of pKₐ by pH-Dependent UV-Visible Titrations

One problem seen early in this study was the denaturing of the sample towards high pH, as incorrectly low meter readings at pH >11.5 led to unexplained increases in absorbance beyond the upper baseline. Application of the correction from probe calibration allowed for the titration to be stopped at the proper pH (Figure 3-4). A UV-Vis monitored pKₐ determination for GTP in the presence of 100mM LiCl yielded an average pKₐ of 10.09, slightly elevated from the pKₐ of unpaired guanine due to the presence of the anionic triphosphate (Figure 3-5).

Remaining titrations were performed in 100mM KCl. Performing in this medium had two major advantages: 1) there was no discernable effect of the salt on the pH probe so the pH did not need to be corrected, and 2) 100mM KCl is close to physiological concentration. Effects of stabilizing stacking interactions by adenine were investigated by pKₐ determination of oligonucleotide sequences GAAAA, GCCCC, CCGCC, and AAGAA (Figures 3-6, 3-7).

The effect of a nearby cationic base on the pKₐ shift of guanine was investigated through the titrations of oligonucleotide sequences AAGAA, AA⁺GAA, AAGA⁺A, and AA⁺GA⁺A. Upon the addition of methylated adenines, the pKₐ of guanine in the sequence shifted even farther from neutrality than in GTP or the non-methylated oligonucleotides (Figures 3-8, 3-9). All titrations were repeated, and average pKₐ values plus deviations are listed in Table 3-1.
Table 3-1: Experimentally Determined $pK_a$ Values from Titrations

<table>
<thead>
<tr>
<th>System Description</th>
<th>Salt Conditions</th>
<th>$pK_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 µM GTP</td>
<td>100 mM LiCl</td>
<td>10.09 ± 0.04</td>
</tr>
<tr>
<td>10 µM GAAAA</td>
<td>100 mM KCl</td>
<td>10.61 ± 0.01</td>
</tr>
<tr>
<td>10 µM GCCCC</td>
<td>100 mM KCl</td>
<td>10.45 ± 0.04</td>
</tr>
<tr>
<td>10 µM CCGCC</td>
<td>100 mM KCl</td>
<td>10.37 ± 0.03</td>
</tr>
<tr>
<td>10 µM AAGAA</td>
<td>100 mM KCl</td>
<td>10.54 ± 0.03</td>
</tr>
<tr>
<td>10 µM AA^GAA</td>
<td>100 mM KCl</td>
<td>11.12 ± 0.06</td>
</tr>
<tr>
<td>10 µM AAGA^A</td>
<td>100 mM KCl</td>
<td>11.07 ± 0.05</td>
</tr>
<tr>
<td>10 µM AA^GA^A</td>
<td>100 mM KCl</td>
<td>11.64 ± 0.07</td>
</tr>
</tbody>
</table>

The $pK_a$ values of oligonucleotide sequences were elevated ~0.5 units from the $pK_a$ of GTP, possibly due to the presence of a greater number of phosphates in the anionic backbone of these nucleotides compared to GTP or due to stacking interactions stabilizing charge-neutral guanine ($G^0$). The $pK_a$ values between the oligonucleotides differed by only ~0.2 units, and the $pK_a$ of oligonucleotides containing adenines was slightly higher than that of sequences that contained cytosines (a poor base stacker).

Addition of a single methylated base increased the $pK_a$ to >11.0. The experimentally determined $pK_a$ values of AA$^+$GAA and AAGA$^+$A were 11.12 and 11.07, respectively, showing similar behavior regardless of the position of N1-meA with respect to the guanine. Addition of two methylated bases in the oligonucleotide AA$^+$GA$^+$A yielded an even more elevated $pK_a$ of 11.64 (Figures 3-8, 3-9).
Figure 3-4: 100mM LiCl pH Probe Calibration. Overall, little deviation (0.1-1.36%) was seen in pH meter readings.
Figure 3-5: Determination of the pKₐ of GTP in the presence of 100mM LiCl. The extrapolated pKₐ value was 10.09.
Figure 3-6: Titration curves of A. GAAAA, B. AAGAA, C. GCCCC, and D. CCGCC in the presence of 100mM LiCl.
Figure 3-7: Superimposed titration curves of GAAAA, AAGAA, GCCCC, and CCGCC.
Figure 3-8: Titration curves of A. AAGAA, B. A(A\textsuperscript{+})GAA, C. AAG(A\textsuperscript{+})A, and D. A(A\textsuperscript{+})G(A\textsuperscript{+})A in the presence of 100mM LiCl.
C. Tautomerization of N1-meA

One key problem in the utilization of N1-meA, which was brought out by Dr. Feldman during the oral defense of the thesis, is the formation of tautomeric forms of the nucleobase (Figure 3-10). Compounds like N1-meA and N1,N9-dimethyl adenine (N1,N9-dimeA) have been shown to lose a proton from the N6 position. This deprotonation could then lead to the

Figure 3-9: Superimposed titration curves of AAGAA, A(A⁺)GAA, AAG(A⁺)A, and A(A⁺)G(A⁺)A.
formation of a charge-neutral heterocyclic structure that would affect conclusions drawn from this titration.

**Figure 3-10:** Possible tautomeric form of N1-meA. A. Possible resonance form of N1-meA where the positive charge is localized to N6. B. Possible imino form of N1-meA, where the positive charge is alleviated through deprotonation at N6.

For N1-meA, pKₐ values given in literature have varied from 8.25 to ~9.3.¹⁵,¹⁷ As such, when titrating through these pH ranges, deprotonation at N6 to form the imino tautomer is probable. Results of pH dependent UV-Vis titrations of N1-meA in the presence of 0.100 M NaNO₃ at 25°C have allowed for quantification of the tautomeric equilibrium, K_T, between the imino and amino forms of the substituted base in the presence of 0.100 M NaNO₃ (Equation 3-2).¹⁴,¹⁵
The $K_T$ of N1-meA was found to be $10^{5.08}$, suggesting that the imino tautomer of N1-meA is strongly favored.\textsuperscript{14,15}

An inspection of the UV absorbance curves for GTP vs. N1-meA-containing oligonucleotide sequences reveals differences in shape of the absorbance curve at 220nm at pH $> 10$. This change in absorbance may indicate the presence of the imino tautomer of N1-meA. As such, the attempt to design oligonucleotide systems that could aid in deprotonation of a base via addition of cationic adenines likely did not yield a proper electrostatic environment. Nonetheless, incorporation of the N1-meA bases shifted the $pK_a$ of guanine farther from neutrality than in GTP.

To understand this, we start by considering AAGAA, which has no methylated A's. In this case, the $pK_a$ is elevated from 10.1 in GTP to 10.54. This increase in $pK_a$ is likely due to in part to the presence of more phosphate groups in the oligonucleotide sequences, leading to a higher effective negative charge density in the synthetic oligonucleotides compared to GTP. Additionally, stacking occurs in oligonucleotide sequences, but not in GTP. The stacking seems to favor $G^0$ and it is possible that anionic guanine does not stack well. Another point to note is the shape of the absorbance curves when comparing GTP titrations to the titrations of the oligonucleotides. While the GTP curve shows decreasing absorbance, the oligonucleotide titrations all show increasing absorbance. Absorbance increase is generally indicative of the breaking of stacking interactions (Figure 3-11).

Next, we consider the N1-meA-containing oligonucleotides. Assuming the charge neutral imino tautomer of the base predominates, there is now no cationic stabilization for the deprotonation of the base.
However, the presence of hydrophobic methyl groups on N1-meA adjacent to the guanine of interest may increase the stacking strength of N1-meA compared to adenine. In this regard, an increase in the experimentally determined $pK_a$ of guanine within these sequences is chemically reasonable. While it is difficult to deprotonate a base in such a highly negatively charged local environment, it becomes even more difficult to do so in the presence of stabilizing stacking interactions and absence of electrostatic stabilization of the negative charge. If nothing else, my data suggest that any downward shift of the $pK_a$ of the guanine by the cationic bases is not substantial enough to shift it below the $pK_a$ of the protonated base (i.e. below ~9).

D. Problematic Data Fits in $pK_a$ Extraction

Direct measurement of the $pK_a$ of active site residues of various ribozymes have been attempted using a variety of techniques including fluorescence, NMR, Raman crystallography, and computational analysis.\textsuperscript{2,10,13,18} The Fedor lab has probed the $pK_a$ shift for G8, a guanine base in the HDV ribozyme that has been implicated in acid-base catalysis. Their data showed similar difficulties in achieving an upper baseline to pH titrations.\textsuperscript{2} As such, the $pK_a$ shift of guanine may be farther from neutrality than once realized.

V. Future Directions

A. Analysis of N1-meA Containing Oligonucleotides by Mass Spectrometry

The oligonucleotide sequences synthesized at Penn State are pending mass spectrometric analysis to verify correct incorporation of N1-meA bases. Efforts have been made to replace the cationic potassium counter-ion with ammonium via precipitation and dialysis in order to facilitate electrospray mass spectrometry (ES-MS). As the sequences are only five nt long, they are very difficult to precipitate out of solution, and efforts have yielded samples too small for mass spectrometric analysis. Dialysis of these oligonucleotides against 2.0 M ammonium acetate ($\text{NH}_4\text{OAc}$) was also attempted. However, a leak was detected in the membrane and some sample
was lost. As such, small amounts of remaining samples are still pending mass spectrometric analysis.

B. Computational Analysis of G•G Motif and Possibility for Isosteric Base Pairs

To determine if the G•G base pair motif is biologically relevant, computational analysis was employed using a similar methodology as presented in Chapter 1. Attempts were made to search for the G•G motif in existing databases in the hopes of finding a candidate structure that could be used for a similar study as presented in Chapter 1 using an MC Search/MC Annotate pipeline. However, no structures could be found that displayed ideal base positioning for this interaction.

As such, future computational analysis will search for isosteric base motifs of the G•G pair – namely an A•A \textit{trans} Watson-Crick pair (Figure 3-12). In addition, future analysis will use donor/acceptor patterns in addition to isosteric bases as a search parameter. However, MC Search is not yet capable of searching solely by such donor/acceptor patterns. The analysis of such base motifs is a possible project for Yin Tang, a bioinformatics and genomics graduate student, who is looking for a project involved in pattern recognition.

\begin{center}
\includegraphics[width=0.5\textwidth]{figure3-12.png}
\end{center}

\textbf{Figure 3-12:} A•A \textit{trans} WC/WC base pair.

C. \textit{8}-Fluoro-Guanine Base Titrations

In addition to difficulties in modeling cationic bases, a key problem in these titrations is achieving a definitive upper base line through the titration. Despite efforts to create an electrostatically beneficial environment for the deprotonation of guanine at near-neutral pH,
upper baseline values were still not achievable. Instead of attempting to create an
electrostatically favorable environment for the base to deprotonate, another possibility is to
directly depress the pKₐ of the nucleobase through the incorporation of 8-substituted guanine.
Fedor Lab (The Scripps Research Institute) used 8-azaG, which has a pKₐ of 8.05 versus 9.2 for
guanine, and tracked pKₐ shifts by fluorescent titrations. Future directions to further depress the
pKₐ and achieve an upper base line may be possible through the substitution of a strong electron
withdrawing group like fluorine.

VI. Acknowledgements

I would like to thank Dr. Josh Sokoloski and Jesse Biel for all of their help in attempting
to prepare these oligonucleotides for mass spectrometry analysis. In addition, I would like to
thank the support staff at Hach Chemical for their assistance in the design of multiple probe
cleaning procedures for the IFSET probe.

VII. References


found in known RNA structures. Nucleic Acids Res. 28: 375-376.

shifting in an A•C wobble: effects of helix position, temperature, and ionic strength.
Biochemistry 49: 3225-3236.

nucleobase pKₐ values by indirect labeling and demonstration of a pKₐ of neutrality in


Chapter 4

In-Solution Raman Detection of Guanyl Anion Using Model Compound GTP

I. Abstract

This study investigates the potential for in-solution Raman analysis to detect electrostatic differences in nucleobases as an alternative to Raman crystallography. As discussed in Chapter 3, nucleobases can participate in general acid-base catalysis. Of particular interest is the potential for anionic guanine to act as a general base in the Hammerhead, GlmS, hairpin, and VS ribozymes. This study reveals the potential to use Raman microscopy to detect anions as they populate within a system using guanosine triphosphate (GTP) as a model compound. Results show adequate spectral acquisition and differentiation of the anionic and charge-neutral form of guanine at fairly high GTP concentrations (≥ 100 mM).

II. Introduction

Since its discovery in 1928, Raman spectroscopy, a light scattering technique, has provided a unique fingerprint for molecular structures.\(^1\) When a photon interacts with a molecule, elastic (Rayleigh) scattering or inelastic (Raman) scattering may result. Elastic scattering occurs when light of the same wavelength and energy of the incident photon is reflected, while inelastic (Raman) scattering occurs when the reflected photon has a different wavelength and energy than the incident photon.

Raman spectroscopy measures the energy change of inelastically scattered monochromatic light, and Raman spectra plot intensity of this energy change before and after the scattering of the incident photon. As such, since electrons that are promoted into the virtual
excited state almost immediately relax to the ground state, Raman spectroscopy can provide an instantaneous picture of the molecule of interest.\textsuperscript{1-3} In addition, Raman has several key advantages for biological researchers: in addition to being a non-invasive technique, water has a very weak Raman absorption (as explained below), so the Raman spectrum of a sample in water should be dominated by the sample itself.\textsuperscript{2-3}

For a molecular vibration to generate a Raman signal, there must also be a distortion of the electron density near the vibrating nuclei. As such, electron rich groups (ex: C=O, C=N, C=C, C-S, S-S, S-H) generally tend to produce more intense Raman signals than electron poor groups.\textsuperscript{1,4} In nucleic acid analysis, vibrations within the nucleobases or the backbone phosphates typically dominate the Raman spectrum (Table 4-1).\textsuperscript{4,5} In contrast, in light absorption spectroscopy, the vibrations of many kinds of unsymmetrically bonded atoms can contribute significantly to the overall spectrum, leading to very complicated data with many overlapping bands. It is worth noting that while Raman has many useful characteristics, it is relatively insensitive. For example, only about 1 photon in $10^7$ photons is inelastically scattered. Ways to overcome this insensitivity include Raman crystallography,\textsuperscript{5} and Surface Enhanced Raman Spectroscopy (Chapter 5).

Table 4-1: Common Raman Stretches in RNA\textsuperscript{4,6}

<table>
<thead>
<tr>
<th>Wavenumber (cm$^{-1}$)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>315</td>
<td>Adenine</td>
</tr>
<tr>
<td>362</td>
<td>O-P-O</td>
</tr>
<tr>
<td>404</td>
<td>C5’-O-P</td>
</tr>
<tr>
<td>432</td>
<td>C3’-O-P</td>
</tr>
<tr>
<td>531</td>
<td>Backbone</td>
</tr>
<tr>
<td>576</td>
<td>C=O bend (Cytosine, Guanine)</td>
</tr>
<tr>
<td>598</td>
<td>Cytosine, Guanine</td>
</tr>
<tr>
<td>629</td>
<td>Cytosine</td>
</tr>
<tr>
<td>645</td>
<td>Cytosine, Adenine</td>
</tr>
<tr>
<td>668</td>
<td>Guanine (ring breathing)</td>
</tr>
<tr>
<td>724</td>
<td>Adenine (ring breathing)</td>
</tr>
<tr>
<td>785</td>
<td>Cytosine, Uracil (ring breathing)</td>
</tr>
<tr>
<td>813</td>
<td>O-P-O stretch</td>
</tr>
<tr>
<td>848</td>
<td>Ribose or O-P-O stretch</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>863, 918, 1045</td>
<td>Furanose C-O, C-C stretch</td>
</tr>
<tr>
<td>1101</td>
<td>PO² symmetric stretch</td>
</tr>
<tr>
<td>1157</td>
<td>Cytosine</td>
</tr>
<tr>
<td>1250</td>
<td>Ring Mode (C, U, A)</td>
</tr>
<tr>
<td>1297</td>
<td>Ring Mode (C, A)</td>
</tr>
<tr>
<td>1318</td>
<td>Ring Mode (G)</td>
</tr>
<tr>
<td>1337</td>
<td>Ring Mode (A)</td>
</tr>
<tr>
<td>1377, 1418</td>
<td>Ring Mode (A, G, C)</td>
</tr>
<tr>
<td>1686</td>
<td>C=O Stretch (Uracil)</td>
</tr>
<tr>
<td>1720</td>
<td>C=O Stretch (Guanine)</td>
</tr>
</tbody>
</table>

Bands below 200cm⁻¹ contain delocalized helix modes and lattice vibrations and can be useful in structural study. Bands in the 600-800cm⁻¹ region have been assigned as ring-breathing motions. Bands in the 1100-1700cm⁻¹ region arise from conjugated single and double bond formation within the bases, including exocyclic C=O and C-NH₂ groups. It is important to note that the spectral profiles in the 1100-1700cm⁻¹ range of dsDNA and dsRNA can greatly vary due to the presence of thymine in DNA and uracil in RNA.

A. Applications of Raman Spectroscopy in Biomolecular Analysis

Raman spectroscopy has been intensively utilized in analysis of biomolecular compounds. In nucleic acid analysis, Alex Rich utilized Raman crystallography in his landmark studies of the conversion of B-DNA to Z-DNA by examining the Raman spectra of crystals. More recently, Raman microscopy has emerged as a key tool for biomedical researchers. Experimental approaches are being developed and updated to allow for the study of complex biological processes and interactions such as protein folding, membrane assembly, and even complete cell analysis.

Raman crystallography is the application of Raman spectroscopy to crystal structures. This technique has been successfully utilized for proteins, and has recently been successfully applied to catalytic RNA molecules, also yielding a wealth of data. Crystallographic study allows for increased concentration of material (typically mM) and has a fairly low luminescent
background. However, crystal growth is both a time and labor intensive process, so applications of Raman spectroscopy that can yield high-resolution data in a much shorter amount of time are desirable.

A key limitation to the application of Raman crystallography is the high concentration of molecules (mM) needed to obtain a good spectrum. In-solution analysis is a possible solution to this concentration issue since it is more akin to the natural state of a cell, where nucleic acid concentrations range from nM to pM levels. However, a fairly low signal to noise ratio exists due to lower concentrations of the nucleic acid in-solution (typically µM). In this regard, optimization of in-solution Raman analysis of nucleic acids could provide great insight to structural questions.

B. Guanines in Ribozyme Self-Cleavage

Guanines are critical in the self-cleavage reactions of many ribozymes, but their precise function in catalysis is unresolved. Considering a general acid-base model, guanine is thought to act as a general base, deprotonated at the N1 position, by accepting a proton from the 2’-hydroxyl. The protonated form then stabilizes a 2’-oxyanion leaving group by hydrogen bonding from the N1 position (Figure 4-1A). Another possibility is another molecule in solution acts as the general base and deprotonates the 2’-hydroxyl. Here, the guanine remains in the charge neutral, protonated state throughout the reaction and serves to stabilize the oxyanion intermediate (Figure 4-1B).
Figure 4-1: Possible self-cleavage mechanisms in the hairpin ribozyme and the role G8 may play. **A.** Deprotonated guanine acts as general base and accepts a proton from the 2’-hydroxyl. **B.** Another molecule in solution acts as the general base and accepts a proton from the 2’-hydroxyl; guanine stabilizes the oxyanion intermediate.

$pK_a$ perturbation can favor a particular ionization state of a nucleobase and control the function of the nucleic acid. Preliminary data from pH-dependent UV-visible titrations (Chapter 3) indicates the $pK_a$ of guanine shifts farther from neutrality, suggesting that the neutral, protonated form of the base predominates at physiological pH. A variety of analytical and spectroscopic methods can be employed to determine shifts in $pK_a$. Understanding the nature of $pK_a$ shifting in guanine allows for better understanding of possible roles it may have in larger systems (i.e. catalysis, stabilization, etc).

Possible approaches to characterization of the native electronic state of guanine in ribozymes include analytical and spectroscopic methods. Most successful has been the use of fluorescence titrations targeting fluorescence changes of the base of interest (guanine) as a function of pH.\textsuperscript{13} Whereas these pH-dependent titrations can probe for the $pK_a$ of an entire RNA,
these techniques cannot accurately predict the exact base that has been ionized at near-neutral pH. To gain a virtual snapshot of the molecule of interest, a technique like Raman spectroscopy can prove to be very useful.

III. Materials and Methods

A. Preparation of Hanging Drop Samples

Various concentrations of GTP (sodium salt, Sigma), ranging from 50mM-100mM, were prepared in an effort to find a minimum concentration needed to still yield a readable Raman spectrum (Table 4-2). Each sample was adjusted to two different pH levels (~7.0-7.5 and 10) using 0.5M LiOH to prevent formation of G-quadruplexes. At neutral pH, the uncharged, protonated form of GTP should dominate; while at pH 10 the anionic deprotonated form of GTP should dominate. No purification steps were performed for in-solution analysis.

4µL drops (various concentrations of GTP) were pipetted to the center of a siliconized cover slip, and suspended by covering a well in a PCR sample tray. Prior to suspension of the drop, 50µL of water was placed in the well to help prevent evaporation of the drop during data collection. The siliconized cover slips were inverted and the area around the drop sealed with grease. Data was collected using a Renishaw inVia microRaman instrument (Figure 4-2).
Figure 4-2: Schematic of Raman Microscope Setup. (Adapted from Gong et al. “Raman Crystallography of RNA.” Methods: 2009, 49: 101-111). Red laser light from a 633 nm HeNe ion laser was used to excite RNA in-solution. After the light travels through a series of mirrors and lenses in the microscope (red lines), backscattered photons are collected through a 20x objective lens, and are then directed to a CCD detector which records the data. The illuminating lamp (black lines) allows for visualization of the hanging drop through a video CCD output on the computer screen. This visualization allows for more accurate positioning of the sample in relation to the laser beam.
B. Sample Positioning and Laser Calibration

After a 10 minute warm up for the 633nm HeNe laser, the sample tray was placed on the stage. The sample tray was first manually positioned by hand, and then more precisely by small adjustments of the stage so the sample was directly under an objective lens. At this point, all external lights were turned off, and the laser was autofocused on an internal silicon reference. Finally, a laser auto-align was performed prior to analysis.

C. Laser Focus to Hanging Drop

After focusing the 10x and 20x objective lenses, the laser was used to focus on the drop beyond the plane of the coverslip with minimal power and intensity. To do this, the laser was first focused to the coverslip where the beam scattering was minimized. Then, the beam was forced slightly beyond the plane of the coverslip until the light surrounded and collected on the center of the drop (Figure 4-3).

D. Data Acquisition

To avoid stray incident lighting reaching the detector, all external lights in the room were turned off prior to data acquisition. A total acquisition time of 300s was recorded with ten 30s accumulations for a range 200cm\(^{-1}\) – 1800cm\(^{-1}\). Due to the long acquisition time and intensity of the laser, individual drops were used up to three times before a short cooling period.
IV. Results and Discussion

A. Rationale Behind Choice of GTP

Since a large amount of material would be needed for in-solution analysis, GTP was chosen as it is an inexpensive, commercially available substance. In addition, the Carey Lab (Case Western Reserve University) provided spectra obtained for GTP [unpublished data] for comparison.

B. Obtained Spectra and Signal Strength

A minimum concentration of 100mM was needed to yield a readable spectrum and provide any information about the electrostatic state of the GTP. Major peak assignments used for identification of GTP (pH 7) and deprotonated GTP (pH 10) are shown in Table 4-3.

### Table 4-2: GTP Concentrations and Raman Signal Strength

<table>
<thead>
<tr>
<th>GTP Concentration</th>
<th>Signal Strength</th>
<th>Raman Intensity of 1487 cm(^{-1}) peak (counts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM</td>
<td>No signal</td>
<td>N/A</td>
</tr>
<tr>
<td>75mM</td>
<td>Minimal signal, high background</td>
<td>N/A</td>
</tr>
<tr>
<td>85mM</td>
<td>Fair signal, some background</td>
<td>~8000</td>
</tr>
<tr>
<td>100mM</td>
<td>Strong signal, some background</td>
<td>~18000</td>
</tr>
</tbody>
</table>

For concentrations below 85mM GTP, no signal is received. Figure 4-4 shows a spectrum taken using 50mM GTP, 20x objective lens, and 300s acquisition time. The increase in signal corresponds to reading of the siliconized cover slip plus possible background luminescence from impurities.

At GTP concentrations at and above 100mM, a strong signal is received and adequate Raman spectra are observed using a detection time of 300s. Figures 4-5 and 4-6 show the spectra of GTP and anionic GTP, respectively. Irregularities in background are noted at Raman shifts below 200 cm\(^{-1}\) and above 1730 cm\(^{-1}\) possibly due to the presence of an impurity in the sample or
slight slipping of the stage during long acquisition times causing improper focus. Despite these irregularities, a difference spectrum of GTP (pH 10-pH 7) can be obtained, which allows for differentiation between the protonated and deprotonated form of GTP (Figure 4-7). Table 4-3 summarizes key stretches used to identify these forms.

**Figure 4-4:** Raman spectrum of 50mM GTP, with high background noise from siliconized cover slip and buffer or impurities. Regardless of 300s acquisition time, no signal from the sample was obtained.
Figure 4-5: Representative Raman spectrum of GTP at pH 7.0. Obtained using 100 mM GTP, 20x objective lens, 300s accumulation time. No difference spectra taken.

Figure 4-6: Representative Raman spectrum of anionic GTP at pH 10. Obtained using 100 mM GTP, 1.0 M LiOH, 20x objective lens, 300s accumulation time. No difference spectra taken.


<table>
<thead>
<tr>
<th>Assignment</th>
<th>Wavenumber (cm⁻¹)</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1189, 1345, 1592</td>
<td>G⁻</td>
</tr>
<tr>
<td>B</td>
<td>1685</td>
<td>C=O (G)</td>
</tr>
<tr>
<td>C</td>
<td>1323</td>
<td>G, ring mode</td>
</tr>
<tr>
<td>D</td>
<td>667</td>
<td>G, ring breathing</td>
</tr>
<tr>
<td>E</td>
<td>589, 1486, 1177</td>
<td>G</td>
</tr>
<tr>
<td>F</td>
<td>338</td>
<td>O-P-O</td>
</tr>
</tbody>
</table>

**Figure 4-7:** Difference Raman spectrum (GTP, pH 10 – GTP, pH 7). Individual spectra for both species are seen above in Figures 4-5 and 4-6. Stretches A-F, labeled above, are identified in Table 4-3.
Overall, Raman spectroscopy can be used to isolate spectra of GTP at fairly high concentrations. It was found that the 20x objective and longer acquisition times (Figure 4-7) yielded spectra of similar quality to those obtained by the Carey Lab [unpublished data]. In the case of anionic GTP, three equal height medium-broad peaks at 1590cm\(^{-1}\), 1474cm\(^{-1}\), and 1345cm\(^{-1}\) are easily discernable from the sharp, tall peak at 1486cm\(^{-1}\) in the charge neutral form. In the difference spectrum, the 1487cm\(^{-1}\) stretch was most easily discerned between the two forms of guanine. The square “peak” seen from 1750cm\(^{-1}\) – 1800cm\(^{-1}\) is likely due to the presence of an impurity in the sample or could be due to slight movement of the stage during long acquisition times.

Ultimately, Raman spectroscopy should be utilized for the characterization of oligonucleotides and large nucleic acid structures rather than analysis of nucleotide triphosphates (NTPs). Identification of the protonation state of a single nucleobase has been achieved using Raman crystallography of the HDV ribozyme.\(^5\) Whereas it is possible to detect changes in the protonation state of a single base using in-solution analysis, the high concentration of sample needed renders it an inefficient process for analysis of larger, more complex structures.\(^5\)

V. Future Directions

Overall, this study shows the possibility for in-solution differentiation between charge neutral guanine and the guanyl anion. In larger RNA complexes, substitution of heavy atoms within nucleobases can be used to separate the guanine of interest from other guanines within the structure. Incorporation of heavy atoms in structures will shift the wavenumber at which

There are certain challenges that must be addressed to conduct Raman experiments in-solution instead of in-crystals – foremost is the low signal-to-noise ratio. While increasing the concentration of RNA in solution helps, high concentrations of RNA in solution can form dimers and larger aggregates, which are problematic.\(^4,5\) In addition, RNA in solution has a fairly high
luminescent background, and focusing of the microscope objective and laser on a single hanging drop can be challenging.\textsuperscript{5} As such, future in-solution work will focus on prior purification of samples to minimize background noise, or use of longer wavelength lasers.

A possible method to decrease the needed concentration of sample for Raman analysis would include borrowing techniques from traditional surface enhancement and applying them to nucleic acid analysis. Surface Enhanced Raman Spectroscopy (SERS) has been shown to successfully detect picomolar amounts of nucleic acids as biomarker tags. A protocol designed by El Amri and co-workers successfully utilized silver colloids allowing for Raman spectroscopy to identify subpicomolar concentrations of nucleic acid.\textsuperscript{15}

The El Amri protocol has previously only been used to identify the presence of nucleic acid in a complex mixture of particles, allowing for the isolation of nucleic acids from clays and silts. However, application of SERS techniques may amplify signal enough to determine native electrostatic states of nucleobases at given pH. As such, future work test SERS characterization of nucleobases and application to in-solution Raman spectroscopy, as shown in Chapter 5.

VI. \textit{Acknowledgements}

I would like to thank Nichole Sullivan and members of the Badding lab for all of their help and instruction in use of the department Raman microscope.

VII. \textit{References}


Surface Enhanced Raman Spectroscopy (SERS) Analysis of Nucleic Acids

I. Abstract

This study investigates the potential for application of SERS analysis to larger nucleic acid systems as an alternative to two other Raman methods used for RNA: Raman crystallography and in-solution Raman analysis. As discussed in Chapter 4, Raman is a highly insensitive method, and SERS typically amplifies Raman signal by a factor of $10^6$. As such, lower concentrations of nucleic acid can be used, which eliminates potential aggregation associated with high RNA concentrations. A thiolated linker sequence was designed to base pair with the HDV ribozyme, and thiol-metal dative bonds were used to immobilize the nucleic acid to the surface of roughened gold for signal amplification. Preliminary results show tether binding with excellent signal to noise and ribozyme base pairing of lesser signal intensity.

II. Introduction

A. Surface Enhanced Raman Spectroscopy

1. SERS Signal Amplification

Since its landmark discovery in 1928 by C.V. Raman (Nobel Prize, 1930) Raman spectroscopy has been used to determine molecular structure and identify the presence of molecules in systems.\(^1\) A key obstacle in Raman spectroscopy is separating the weak inelastically scattered light of interest from the intense elastically scattered light (known as Rayleigh scattering) and fluorescent light impurities.\(^2,3\) Surface Enhanced Raman Spectroscopy (SERS) is
one kind of Raman spectroscopic technique that has been used in other systems to enhance signal intensity of the spectrum. For example, SERS has been applied to a variety of analytical systems including anthrax detection,\textsuperscript{4-6} chemical-warfare agent detection,\textsuperscript{4,7} and even \textit{in vivo} glucose sensing.\textsuperscript{4,8} SERS is a highly sensitive, molecule-specific, and high-resolution technique in which the intensity of inelastically scattered photons is typically enhanced by a factor of $10^6$ compared to normal Raman spectra.\textsuperscript{4,9}

Adsorption of an analyte to a (roughened) metal surface allows for the increase of the Raman scattering cross-section. Oscillations in the metal electron density (called surface plasmons) create an electromagnetic field. When surface plasmons are excited and collectively resonate with an incident photon of light, Surface Plasma Resonance (SPR) occurs.\textsuperscript{4,9,10} When a metal surface is struck by incident light of wavelength that is much greater than the size of the particles on the metal surface, inelastic scattering and excitation events will take place. However, all molecular excitation except SPR can be ignored and Localized Surface Plasma Resonance (LSPR) occurs.\textsuperscript{9,10}

LSPR sites act as antennae of sorts and enhance the Raman signal, allowing the surface to become SERS active. In order for this enhancement to occur, features must be created on the metal surface with dimension around the same order of magnitude of the excitation wavelength.\textsuperscript{10} These small features are created through roughening of the surface by electrochemical means. On a roughened metal surface, the resultant electric field of plasmons can propagate parallel and perpendicular to the metal surface. On a smooth surface, however, plasmon waves are bound parallel to the surface and can thus propagate only parallel to the surface. The perpendicular propagation of electric field is essential for high intensity SERS detection, and this is achieved by so called “roughening” of the surface.\textsuperscript{9,11}

SERS substrates are typically gold or silver since they are capable of producing plasmons along the metal/air interface, can be handled easily in a variety of environments, and can be
electrochemically roughened to create LSPR active sites.\textsuperscript{12} Whereas silver roughening provides the largest signal enhancement, the nanostructures can undergo oxidation. Gold roughening produces large signal enhancement and is relatively chemically inert.\textsuperscript{12,13}

2. \textit{Self-Assembled Monolayers and Surface Adsorption}

Self Assembled Monolayers (SAMs) of organic substances form during the adsorption of the substance to the SERS surface.\textsuperscript{4} While some molecules may naturally interact with the SERS substrate, organic molecules are generally immobilized on a metal SERS surface via SAM formation.\textsuperscript{14,15} Molecules are synthesized containing a functional head group (commonly, thiols and phosphonates) that acts as an anchor between the molecule and surface (Figure 5-1).\textsuperscript{14}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{Figure5-1.png}
\caption{Schematic representation of Self-Assembled Monolayers formed on a metal surface. The head group interacts with and adheres the molecule of interest to the metal surface. A spacer arm is used to provide a well-defined thickness and act as a physical barrier between the functional group of interest and metal-head group interface region.\textsuperscript{14,15}}
\end{figure}

The unshared pair of electrons from a sulfur atom allows thiol-containing molecules to form fairly strong dative bonds with metal surfaces. The gold-thiol bond is known to be fairly stable (bond dissociation energy \(\sim\)100-150 kJ/mol; heat of adsorption tens of kcal/mol), and thiol groups can be introduced within a biological macromolecule through chemical modification.\textsuperscript{16} Hegner first described successful deposition of DNA oligonucleotides with a 5’ thiol modification to a flat gold surface, allowing for enhanced resolution by atomic force microscopy (AFM).\textsuperscript{13}
This procedure has been utilized successfully in other microscopic analyses, and thiolation of organic molecules remains a commonly used technique in SERS analysis today.

3. Distance Dependence of SERS Intensity

One other consideration in SAM design and deposition is the distance dependence of SERS signal amplification.\textsuperscript{3,4} SERS does not require that the material of interest be in direct contact with the surface, but rather within a certain sensing area a few nanometers from the surface. However, the Raman intensity decays significantly as the distance between the sample and metal surface is increased. The distance dependence of SERS has been theoretically approximated as:

\[ I = \left(1 + \frac{r}{a}\right)^{-10} \]  

(5-1)

where \(I\) is the Raman signal intensity, \(a\) is the average size of the field enhancing features on the SERS surface, and \(r\) is the distance from the surface to the adsorbate.\textsuperscript{4,17}

B. Locked Nucleic Acids

Locked Nucleic Acids (LNAs) are high-affinity, conformationally restricted RNA mimics that can be synthetically incorporated into oligonucleotide sequences (Figure 5-2).\textsuperscript{18} The ribose ring is “locked” in an ideal conformation for Watson-Crick binding via a methylene bridge between O2’ and C4’. This forces the sugar into the C3’-endo conformation compatible with A-form helix formation, and helps increase hybridization affinity by reducing the conformational entropy loss upon hybridization. Since these nucleotides are locked in a conformation that supports Watson-
Crick binding, strands that include LNAs demonstrate exceptional thermal stability when hybridized to DNA or RNA oligonucleotides.\textsuperscript{18,19}

It is estimated that the incorporation of a single LNA monomer raises the melting temperature ($T_m$) of a duplex by 2-8°C.\textsuperscript{19} In addition, LNA oligonucleotides have been shown to have remarkable specificity to discriminate single-base mutations in a target system. These key characteristics have proven LNA useful in a variety of applications, but especially in the development of biosensors.\textsuperscript{20,21} In addition, LNA/DNA chimeric oligonucleotides are often used with high success, and reduce the total cost of synthesis, since LNA can be quite expensive to make.

C. Atomic Force Microscopy

1. Introduction to Atomic Force Microscopy

Atomic Force Microscopy (AFM) is a high-resolution type of scanning probe microscopy able to resolve images on the nanometer scale. Unlike Scanning Tunneling Microscopy (STM), which can only image conducting or semiconducting surfaces, AFM can image almost any kind of surfaces including biological samples.\textsuperscript{3,22} Modern AFM typically uses laser beam deflection, in which a laser is reflected from the back of a reflective cantilever onto a position-sensitive detector.\textsuperscript{22,23}

2. Atomic Force Microscopy Image Processing and Analysis

Unlike traditional microscope images, AFM images have three-dimensional topography content (i.e. have height information).\textsuperscript{3,23} AFM images are stored in a computer as a three dimensional array of numbers. As such, the number array can be processed and analyzed using AFM image processing software (WSxM, Nanotec Electronica). AFM images will always have a background slope or curvature that needs to be removed from the image in a “processing” step. A histogram of the image is then created by making a plot of the number of pixels in the image versus the colors of the pixels themselves. Line profiles are the most common form of AFM
image analysis and are a two-dimensional profile, or cross section, that is extracted from the AFM image. The line profile may be taken in any direction along the image (i.e. horizontally, vertically, at an angle, etc.). The line profile plots heights against the distance along the profile line, effectively allowing for a two-dimensional rendering of part of the three-dimensional topologies.23

III. Materials and Methods

A. Preparation of LNA-containing DNA Tether

Surface attachments of ribozymes have been previously used in a variety of single molecule analyses, notably studies of the group I intron by Herschlag Lab and the HDV ribozyme by Andrea Szakal, former graduate student in the Bevilacqua Lab, in single-molecule experiments.24,25 To immobilize the HDV ribozyme to the roughened surface for SERS analysis, a 14-nucleotide LNA-containing DNA tether including a 5’ thiol modification was prepared by Exiqon (Woburn, Massachusetts) for these experiments. The tether is complementary to the 85-99 position of the 1-99 form of the HDV ribozyme (Figure 5-3). Exiqon performed HPLC purification of the oligonucleotide tether and mass spectrometric analysis.

In general, LNA-containing residues will bind very tightly to other LNA-containing residues. To help minimize the formation of stable secondary structures in LNA-containing oligonucleotide sequences, GC content should be kept between 30-60%. In addition, stretches of more than 3 consecutive LNA bases and 3 consecutive natural Gs or Cs should be avoided (Exiqon). While the incorporation of an LNA into a nucleic acid structure is fairly expensive, each LNA present typically raises the $T_m$ of the sequence by 2-8°C. As such, two LNA bases were incorporated into the tether design, which raised the $T_m$ of the DNA oligonucleotide to 54°C.
LNAs are now commonly used in biosensor design, and it is suggested that LNA bases are incorporated at regular intervals along the tether sequence.\textsuperscript{26,27} The location of the LNA bases was chosen based on analysis using the Exiqon LNA Oligo Optimizer tool (http://www.exiqon.com/ls/Pages/ExiqonOligoOptimizerTool.aspx) and scoring system. This tool helped to avoid significant self-hybridization and secondary structure formation based on placement of the LNA bases in the oligonucleotide sequence. Sequences are inputted to the Oligo Optimizer tool, and raw scores are computed. Scores below “60” can be viewed as a very rough estimate of the melting temperature (in °C) of the secondary structures. As such, the secondary structure of an LNA-containing oligonucleotide with a score below 20 is unlikely to be stable at room temperature. While scores below “40” (40°C melting temperature) are considered good sequences, we endeavored to keep scores below “20.” The final tether design had a score of “10” due to the presence of two consecutive G/C bases.
Figure 5-3: LNA-containing DNA tether and HDV 1-99 construct. A. DNA-based LNA tether used in SERS experiments. Red color represents LNA bases. B. HDV ribozyme 1-99 construct with the tether shown in pink. (Adapted from Chadalavada et al. (2000) A Role for Upstream RNA Structure in Facilitating the Catalytic Fold of the Genomic Hepatitis D Virus Ribozyme. J. Mol. Biol 49: 101-111. C. Schematic of deposition to gold surface. Colors of portions of the HDV ribozyme match those in Figure 5-3B.

B. HDV Ribozyme Construct and Native Gel Analysis

The 1-99 HDV ribozyme construct (~20 μL, 35.9 μM) used for SERS analysis was previously prepared and characterized by Jennifer Wilcox, graduate student in the Bevilacqua Lab. Native gels were used to check the effects of 10 mM MgCl₂ and the annealing process on the structure of the HDV ribozyme. A 0.75mm thick gel was cast (12% acrylamide – 37.5:1
crosslinking ratio, 0.25X TBE). Sample concentrations were checked using a NanoVue spectrophotometer and adjusted when brought up in 5% glycerol such that the final absorbance of each sample was ~1.00. The gel was run at 8°C at 200V constant voltage with 2.5X TBE running buffer for ~5 hours. Finally, the gel was treated with SYBR Gold Stain for ~15 minutes, then exposed under UV light.28

A 110 nt marker, pri-miR-16-1 was prepared and characterized by Kaycee Quarles, graduate student in the Showalter Lab.

Sequence: 5’ - GGG UGA UAG CAA UGU CAG CAG UGC CUU AGC AGC ACG UAA AUA UUG GCG UUA AGA UUC UAA AAU UAU CUC CAG UAU UAA CUG UGC UGC UGA AGUAAG GUU GAC CAU ACU CUA C - 3’

This sequence is a partial hairpin and will fold somewhat on a native gel. However, it is fairly representative of a 110 nt species and was used as a general marker on the gel.

C. Electrochemical Roughening of Gold Wafer

The metal films were deposited onto silicon substrates (~1.5 nm of native oxide). Prior to deposition, silicon substrates were cleaned with piranha solution (75% concentrated sulfuric acid, H₂SO₄, 25% hydrogen peroxide, H₂O₂), rinsed copiously with water and ethanol, and then blown dry with nitrogen. Approximately 10nm of chromium was first deposited from a heated rod coated with chromium. This chromium coating acts as an adhesion layer and was followed by deposition of ~100 nm of gold deposited from a resistive boat (background pressure < 3x10⁻⁸ Torr during all deposition processes).

All solutions were prepared using deionized Millipore-grade water and degassed for ~20 minutes at room temperature (23°C) directly before use. All electrochemical experiments were performed at room temperature in a homemade Teflon corrosion cell, controlled by a potentiostat (VoltaLab PGZ402). Electrochemical roughening of a 2.0 in diameter gold wafer was accomplished with repeated oxidation and reduction cycles (ORC) based on Weaver’s procedure
using 0.1 M NaCl, a platinum wire auxiliary electrode, and an Ag/AgCl reference electrode.\textsuperscript{29,30} A cyclic potential voltimetric scan consisted of ramping potential at 0.3 V/s from -0.256 V to 1.445 V and holding for 1.3 seconds, then decreasing potential to -0.256 V and holding for 10 seconds. This wave was repeated for 30 cycles (Figure 5-4).\textsuperscript{29}

\textbf{Figure 5-4:} Cyclic voltammogram at 300mV/s of the 30 cycle roughening process in 0.100 M NaCl. This graph tracks the oxidation-reduction cycle treatment of gold wafers and was charted using PeriStat software.

\textbf{D. Atomic Force Microscopy (AFM) Verification of Electrochemical Roughening}

At this point, Atomic Force Microscopy (AFM) was performed to examine the roughening of the surface. Tapping mode AFM images were taken using a Dimension Series 3000 scanning probe microscope (Digital Instruments, Santa Barbara, CA). Commercial silicon cantilever probes approximately 125 µm long with typical resonant frequencies between 294 and 375 kHz were used. Height, amplitude, and phase images were recorded at room temperature.

RMS Amplitude and Vertical Deflection were observed to be around 0V prior to focusing on the probe tip. At this point, we focused directly on the wafer surface, using mars and blemishes on the surface to help ensure the microscope was properly focused. Cantilever auto tuning in tapping mode was performed, and then the cantilever was engaged for the scan in tapping mode.
E. Proteinase K Treatment of Wafers

To ensure no nucleases or proteins remained on the wafers following handling during the electrochemical roughening procedure and AFM verification, wafers were treated with Proteinase K. First, wafers were broken into small pieces using a scribe, and pieces were placed in a sterile petri dish. For example, a circular wafer 5.08 cm in diameter was broken into 12 pieces of ~0.7 cm by ~0.3 cm. The wafers were incubated with Proteinase K (100 µg/mL) in 0.01 M Tris (pH 7.5), 5 mM EDTA, and 0.5% SDS for 40 minutes at 37˚C.

Following this treatment, Proteinase K was inactivated by incubating the plates at 70˚C for ~15 minutes. Wafers were removed from the oven and allowed to cool to room temperature, then rinsed ~15 times with sterile, deionized water and gentle agitation. The wafers were individually rinsed with 100% ethanol, blown dry using nitrogen, and placed in new, dry petri dishes.

F. Wafer Soaking

Small 3mL glass conical vials were used for soaking of samples. Caps were washed using Alconox soap dissolved in purified water, while vials were soaked in piranha solution for 10 minutes, then rinsed repeatedly with purified water and allowed to dry overnight. Caps and cleaned vials were then autoclaved to remove any nucleases.

The DNA/LNA tether and HDV ribozyme mixture was heat denatured at 60˚C for 1 minute and allowed to cool on ice for 5 minutes prior to transfer to the conical vial and soaking of the wafer. Wafers were soaked in a 2 mL total volume solution containing 10 mM MgCl₂, 10 mM Tris (pH 7.5), and nucleic acid (5 µg/mL).

Four vials were set up: 1) buffer only – containing no nucleic acid; 2) HDV only – HDV 1-99 construct (5 µg/mL); 3) Tether only – DNA/LNA tether (5 µg/mL); and 4) HDV+Tether – HDV 1-99 construct (3.75 µg/mL) and tether (1.25 µg/mL). Clean, dry wafer pieces were placed in conical vials and all wafers were soaked for 36 hours at 4˚C, with Raman scans performed at
24 and 36-hour time points. Following this data collection, waters were placed back in their soaking solutions, and were left at room temperature for 18 hours to see if the increase in temperature allowed for better sample deposition and increased signal intensity.

SERS was recorded on a Renishaw inVia microRaman instrument in the Badding Lab. The instrument consists of a 35 mW HeNe laser (632.8 nm) as the light source, a motorized microscope stage sample holder and a CCD detector (Figure 4-2). A 50x objective lens was used during data acquisition. While the wafers were not blown dry for this series of experiments, the laser and microscope were focused directly to the surface, not on any liquid present on the surface (Figure 5-5). In addition, while physical marring is not indicative of the electrochemical roughening process, the microscope was first focused to these easily visible striations on the surface when searching for SERS active sites.

Figure 5-5: Gold wafer as viewed through 50x objective lens. As noted above, scratches and mars are not indicative of a SERS active site, but were used as a reference point during microscope focus to the surface.
IV. Results and Discussion

A. Rationale Behind HDV Ribozyme/Tether Complex Design

Raman crystallography has been previously used to characterize the HDV ribozyme.\textsuperscript{31-34} As such, we attempted a SERS analysis of this structure to see if we could obtain similar quality spectra. Key stretches used in the identification of HDV are shown in Table 5-1, and are based upon the extensive literature of such stretches.\textsuperscript{31-34}

To help increase the binding affinity of the HDV ribozyme to a relatively short (16 nt) thiolated tether, LNAs were incorporated. It was desirable to use as short of a tether as possible in order to maximize SERS given its strong distance dependence. While the tether only hydrogen bonded to a 15 nt stretch of the HDV ribozyme, an addition adenine was added to the 3’ end of the tether to add stabilizing stacking interactions to the structure.

Spectral profiles in the 1100-1700cm\textsuperscript{-1} range of DNA and RNA can greatly vary due to the presence of thymine in DNA and uracil in RNA.\textsuperscript{35,36} As such, a DNA based tether was chosen to help prevent obscuring interesting features in the HDV ribozyme. The 16-nucleotide sequence was designed to be complementary to the 85-99 3’ overhang arm of the HDV ribozyme, which is not part of the ribozyme and thus should not interfere with its fold.

Table 5-1: Key Raman Stretches in the HDV Ribozyme\textsuperscript{31-34}

<table>
<thead>
<tr>
<th>Wavenumber (cm\textsuperscript{-1})</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>633, 670</td>
<td>G</td>
</tr>
<tr>
<td>725</td>
<td>A</td>
</tr>
<tr>
<td>784</td>
<td>C, U</td>
</tr>
<tr>
<td>812</td>
<td>Backbone</td>
</tr>
<tr>
<td>917, 999, 1045</td>
<td>Ribose</td>
</tr>
<tr>
<td>1101</td>
<td>PO\textsubscript{2}^-</td>
</tr>
<tr>
<td>1181</td>
<td>G</td>
</tr>
<tr>
<td>1232</td>
<td>U</td>
</tr>
<tr>
<td>1252</td>
<td>C</td>
</tr>
<tr>
<td>1322</td>
<td>G</td>
</tr>
<tr>
<td>1337</td>
<td>A</td>
</tr>
<tr>
<td>1376, 1421</td>
<td>A, G</td>
</tr>
</tbody>
</table>
### B. Atomic Force Microscopy Verification of Electrochemical Roughening

Atomic Force Microscopy (AFM) was used to verify success of the electrochemical roughening process. 100nm sections were isolated for smooth, unroughed wafers (Figure 5-6) and compared to electrochemically roughened gold wafers (Figure 5-7). Roughened wafers showed isolated white patches at the sites of roughening. In addition, surface profiles of roughened samples showed more noise near peaks, characteristic of successful surface roughening.
Figure 5-6: AFM characterization of smooth, unroughened wafers. A. AFM scan of unroughened wafer with histogram analysis of height distributions on the smooth wafer surface. B. Profile data and selection. The line profile is a two dimensional profile extracted from the AFM image that shows heights found along that selection.
Figure 5-7: AFM characterization of electrochemically roughened wafers.  

A. AFM scan of roughened wafer with histogram analysis of height distributions on the smooth wafer surface. Example roughened surface patches (white in color) are circled in red.  

B. Profile data and selection. The line profile is a two dimensional profile extracted from the AFM image that shows heights found along that selection.
C. Native Gel Results

Native gel results show multiple bands for the HDV ribozyme sample (Figure 5-6). The presence of two bands for the monomeric form of the 1-99 wild-type HDV ribozyme has been previously shown under similar salt and renaturation conditions. It has been shown that changing salt concentrations does not affect the electrophoretic mobility of the HDV ribozyme sample, but increasing concentrations of the HDV ribozyme can increase population of a dimeric form, not seen on this gel.

Figure 5-8: Native gel results. Samples were renatured by incubating at 90°C for 1 minute, followed by cooling on ice for ~10 minutes. The gel was run at 8°C at 200V constant voltage
with 2.5X TBE running buffer for ~5 hours. Finally, the gel was exposed to SYBR Gold Stain (Invitrogen, Molecular Probes) for ~15 minutes, then exposed under UV light.

Lanes 4-7 of the native gel show similarity in structure regardless of ratio of ribozyme to tether during the incubation process. Lanes 4-5 contain a mixture of HDV ribozyme plus the DNA/LNA tether, mixed at a 3:1 ratio. Lane 4 shows the native structure of this complex, and Lane 5 was heat denatured immediately prior to loading the gel. As seen in Figure 5-6 above, a band can be seen that has migrated the same distance as the tether-only sample. Similarly, Lanes 6-7 show a mixture of HDV ribozyme plus the DNA/LNA tether mixed at a 2:1 ratio. Lane 7 was heat denatured immediately prior to loading the gel, and a band can be seen that has migrated the same distance as the tether-only sample.

D. Obtained Spectra and Signal Strength

Minimal background signal (~1000 counts) was obtained for buffer only and HDV ribozyme-only samples. This suggests that there is a minimum of gold-reactive species in the buffer and ribozyme and minimal adsorption of non-thiolated RNA to the surface. Tether-only samples yielded high intensity (~140,000 counts) spectra with clear peaks (Figure 5-8). Assignment for key peaks in the tether identification is given in Table 5-2.\textsuperscript{38,39} In particular, the C-S thiol stretch is easily discernable at 2936 cm\textsuperscript{-1} and the nucleic acid stretches are found within the 200-1600 cm\textsuperscript{-1} range, as expected. The tether/HDV complex did show a high intensity (~60,000 counts), but few stretches were discernable within the spectrum (Figure 5-9). This spectrum strongly suggests the tether is still present. Possible peak assignments are listed in Table 5-3.\textsuperscript{38,39}
Table 5-2: Key Stretches in Tether Identification

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Wavenumber (cm⁻¹)</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2936</td>
<td>Thiol</td>
</tr>
<tr>
<td>B</td>
<td>2137</td>
<td>Strong polarized bond(^a) (possible degradation product)</td>
</tr>
<tr>
<td>C</td>
<td>1579</td>
<td>Deoxy-G</td>
</tr>
<tr>
<td>D</td>
<td>1533</td>
<td>T</td>
</tr>
<tr>
<td>E</td>
<td>1396</td>
<td>Deoxy-A</td>
</tr>
<tr>
<td>F</td>
<td>1179</td>
<td>G (LNA)</td>
</tr>
<tr>
<td>G</td>
<td>1023</td>
<td>Deoxy-C</td>
</tr>
<tr>
<td>H</td>
<td>743</td>
<td>-(CH(_2))(_n) linker</td>
</tr>
<tr>
<td>I</td>
<td>639</td>
<td>T</td>
</tr>
<tr>
<td>J</td>
<td>525</td>
<td>-(CH(_2))(_n) linker</td>
</tr>
</tbody>
</table>

\(^a\)The region from 2000-2300 cm⁻¹ is fairly “silent” in Raman spectroscopy and very few compounds absorb in this region. One compound that is known to absorb well in this region is carbon monoxide. However, there is a minimal likelihood of carbon contamination from surface preparation practices, so it is more likely that this is a photodegradation product of some sort.

Figure 5-9: Representative SERS spectrum of DNA/LNA tether (5 μg/mL). The 50x objective lens was used and data was acquired for 30s. Peak assignments given in Table 5-2 above.
Table 5-3: Possible Stretches in HDV/Tether Complex Identification

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Wavenumber (cm⁻¹)</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2938</td>
<td>Thiol</td>
</tr>
<tr>
<td>B</td>
<td>2139</td>
<td>Strong polarized bond (possible degradation product)(^a)</td>
</tr>
<tr>
<td>C</td>
<td>1525</td>
<td>Deoxy-G</td>
</tr>
<tr>
<td>D</td>
<td>1154</td>
<td>Deoxy-G</td>
</tr>
<tr>
<td>E</td>
<td>524</td>
<td>-(CH₂)₆ linker</td>
</tr>
</tbody>
</table>

\(^a\)The region from 2000-2300 cm⁻¹ is fairly “silent” in Raman spectroscopy and very few compounds absorb in this region. One compound that is known to absorb well in this region is carbon monoxide. However, there is a minimal likelihood of carbon contamination from surface preparation practices, so it is more likely that this is a photodegradation product of some sort.

![Figure 5-10](image.png)

**Figure 5-10:** Representative SERS spectrum of HDV ribozyme plus DNA/LNA tether (18 hour soak, room temperature). Obtained using 5 µg/mL total nucleic acid concentration and 3:1 ratio of HDV:tether. The 50x objective lens was used and data was acquired for 30s.
One possible problem due to annealing the HDV ribozyme to the tether prior to deposition is that the tether/HDV complex is improperly orientated to the metal surface, leading to lack of increase in signal intensity. Samples that were originally soaked for 36 hours at 4°C were re-soaked at room temperature for 18 hours to see if the signal intensity would increase. Results are shown in Figure 5-8, and were similar to those before the additional room temperature soak. In addition, a wafer that was previously soaked in solution containing only the tether was then soaked in a solution that only contained the HDV ribozyme. It was hoped that the addition of HDV ribozyme to tether that had already successfully bound the metal wafer would allow for increased signal intensity. However, this trial yielded spectra similar to that shown in Figure 5-8. In addition, a wafer fragment containing the tether/HDV complex was incubated for 1 minute in 8M urea, then washed well in 10 mM MgCl₂, 10 mM Tris-HCl buffer. Upon washing with urea and then washing with buffer, the spectrum of the DNA/LNA tether was seen.

As such, there may be some level of success in these results. For example, the RNA typically appears in the 600cm⁻¹ – 1800cm⁻¹ range and signal is being detected that seems to align with previously obtained spectra from HDV ribozyme crystals. In particular, features corresponding to nucleobase ring modes and backbone stretches from the HDV ribozyme may be present. As such, subsequent scans will focus on reducing noise within the 600cm⁻¹ – 1800cm⁻¹ range.

While wafers were treated with Proteinase K and washed thoroughly, it is possible that some of the protein remained on the wafer. However, key stretches typically seen in proteins are not discernable on the tether-only and the HDV complex spectra. For example, tyrosine and tryptophan give a clear shoulder peak at 1618 cm⁻¹, not seen in either spectrum. In addition, one of the most discernable stretches in Raman analysis of proteins is an intense peak at 1668 cm⁻¹ caused by the amide linkage. This 1668 cm⁻¹ peak is also missing in the obtained spectra.
Overall, SERS analysis may turn out to be a valuable technique for Raman signal amplification of ribozymes, as preliminary data suggest that large complexes can be somewhat characterized. As is seen in Figures 5-8 and 5-9, the HDV/tether complex seems to be annealing well to the tether. If the complex and tether were not annealing during the deposition process, acquired spectra of the HDV/tether complex should appear identical to the tether only spectrum. Also, the tether in the complex spectrum is now double-stranded, which can affect its signal-to-noise ratio.

Finally, it is possible that the length of the linker and tether is detracting from the surface enhancement capabilities via LSPR. Calculation of the mean square end-to-end distance of the tether shows the single stranded tether extends 9.36 nm from the metal surface, assuming the inter-base separation to be the fully extended chain length and no stabilizing interactions present in the tether strand. The tether, once annealed to ribozyme would extend ~4.35 nm (.29 nm/bp × 15 bp) from the surface, which was hoped to be close enough to the surface for adequate spectral acquisition. While the annealing of the tether to the ribozyme brings the ribozyme much closer to the metal surface, this distance may be too large to have significant signal increase for overall complex, as noted in the introduction.

The strong stretch at ~2134 cm⁻¹ is most troubling. As suggested by Professor Lasse Jensen, carbon monoxide and carbonyl photodegradation products are known to absorb at this wavenumber. Consultation with Nichole Sullivan, a graduate student in Allara Lab, revealed minimal likelihood of carbon contamination during the deposition of gold to silicon wafers and subsequent roughening of gold surface. In addition, spectra taken of the wafers with only buffer present do not show this particular stretch. It is possible that there is a photodegradation product produced during the deposition and scanning of nucleic acid samples. The possibility of photodegradation will be verified through subsequent testing of stable and well-characterized systems like benzene thiol. As such, electrochemically roughed SERS amplification using the
aforementioned system design remains an inefficient process for structural study of large nucleic acids, although other tether designs may improve this, as may colloidal SERS (see below).

V. Future Directions

Various soaking techniques will be employed to determine effects of buffer and salt concentration on ribozyme structure that may play a role in the adsorption process. Smaller wafers will be used to minimize the chances of missing a SERS active site during Raman data acquisition of the surface. In addition, distance dependence of SERS amplification must be considered, so a shorter tether should be used. One possibility is designing an extension off the top of the P1 portion of the ribozyme and then annealing that to the enzyme portion of HDV. Additionally, the *auf bau* method can be applied to this project: building up a duplex, then a small hairpin, and finally moving to larger structures like the HDV ribozyme.

Another possibility is microstructured optical fibers (MOFs), which increase the pathlength of the sample for SERS analysis and have been successfully used in biomolecular analysis. In addition, the surface preparation process could be varied, and nanostructured surfaces can be designed for use in SERS analysis of nucleic acids. Electrochemical surface roughening, while successful, does not create regular SERS active sites on a metal surface, and is very hard to reproduce. It is possible that signal is being amplified significantly on a particular area of the wafer and was simply missed during the scanning process. Scanning was performed manually, by scanning mars to the gold surface and moving slowly from one edge of the wafer piece to another. Typically, 10-12 spots were scanned per wafer piece, because there is no uniform distribution of SERS active sites on the wafers.

Metal colloids have recently become the premier SERS nanostructure, effectively replacing surface roughening in many types of analyses. Colloidal metal SERS has grown in popularity as fairly inexpensive and versatile technique, and allows for the creation of more reproducible surface environments at the atomic level. Colloidal SERS utilizes “hot spots” –
regions between two or more colloidal nanoparticles with unusually high Raman scattering enhancing capabilities.\textsuperscript{10,43} The presence of such hot spots allows for single molecule detection and may help increase sensitivity to structures of interest. Colloidal SERS has previously been successfully utilized by El Amri and co-workers in the detection of subpicomolar concentrations of nucleic acid.\textsuperscript{44}

VI. \textit{Acknowledgements}

I would like to express my sincere gratitude to Nichole Sullivan (Allara Lab, Penn State) for all of her help in electrochemical roughening, AFM verification, control experiment design, and her time in editing this chapter. I would also like to thank Professor Lasse Jensen for his time and assistance in Raman stretch identification and Brian Glassner and Remi Adenika (Exiqon) for their assistance in LNA-containing tether design. Finally, I would like to thank Kaycee Quarles (Showalter Lab, Penn State) for teaching me how to cast native gels.

VI. \textit{References}


### Appendix A1: Tabulation of A*-C Wobble Mismatch Structural Data in PDB Structures

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Location of Pair</th>
<th>Sequence Context</th>
<th>Nearest Neighbor (Above)</th>
<th>Nearest Neighbor (Below)</th>
<th>Structure Description</th>
<th>Structure Data Source Type</th>
<th>Experimental pH</th>
<th>C(N3)***A(N6) (Å)</th>
<th>C(O2)***A(N1) (Å)</th>
<th>O-H-N Bond Angle (°)</th>
<th>B(total) from Wilson Plot</th>
<th>Mean B(total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1BZ2</td>
<td>A38-C32</td>
<td>SUA3/SUA3</td>
<td>U*A</td>
<td>A*U</td>
<td>tRNA Anticodon Loop</td>
<td>NMR</td>
<td>7.0</td>
<td>3.00</td>
<td>2.83</td>
<td>168.29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1E4P</td>
<td>A7-C18</td>
<td>SGA3/SAC5</td>
<td>A*G</td>
<td>G*C</td>
<td>VSm1 hairpin analog</td>
<td>NMR</td>
<td>5.0</td>
<td>3.27</td>
<td>2.92</td>
<td>171.23</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1EU5</td>
<td>A926-C944</td>
<td>FAAC7/FAAC7</td>
<td>C*G, A</td>
<td>C*G</td>
<td>tRNA Synthetase</td>
<td>X-Ray Diffraction</td>
<td>7.0</td>
<td>3.32</td>
<td>3.41</td>
<td>147.44</td>
<td>56.00</td>
<td>Null</td>
</tr>
<tr>
<td>1FEQ</td>
<td>A38-C32</td>
<td>SUA3/SUC5</td>
<td>A*U</td>
<td>U<em>6A</em></td>
<td>tRNA Lys3 UUU Anticodon Stem / Loop</td>
<td>NMR</td>
<td>5.6</td>
<td>3.00</td>
<td>2.78</td>
<td>154.66</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1FG0</td>
<td>A2485-C2104-2536</td>
<td>SUA3/SUC3</td>
<td>A-A-G</td>
<td>C, G</td>
<td>Large ribosomal subunit + 13 bp minihelix puromycin compound</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>3.02</td>
<td>2.89</td>
<td>159.18</td>
<td>52.00</td>
<td>Null</td>
</tr>
<tr>
<td>1FG7</td>
<td>A1500-C1402</td>
<td>SUA3/SUC5</td>
<td>C*G</td>
<td>A, C(U)</td>
<td>30S Ribosomal Subunit (Complexed with Antibiotics)</td>
<td>X-Ray Diffraction</td>
<td>6.5</td>
<td>3.08</td>
<td>2.47</td>
<td>176.26</td>
<td>75.30</td>
<td>71.30</td>
</tr>
<tr>
<td>1GTR</td>
<td>A26-C44</td>
<td>SUA3/SUC5</td>
<td>C*G, A</td>
<td>C*G</td>
<td>Anticodon loop recognition by glutaminyl-tRNA synthetase</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>2.84</td>
<td>3.39</td>
<td>136.26</td>
<td>Null</td>
<td>Null</td>
</tr>
<tr>
<td>1HWQ</td>
<td>A8-C23</td>
<td>SUA3/SUC3</td>
<td>G*C</td>
<td>A*G</td>
<td>VS Ribozyme substrate</td>
<td>NMR</td>
<td>6.6</td>
<td>2.61</td>
<td>2.92</td>
<td>109.30</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1LC6</td>
<td>A18-C6</td>
<td>SUA3/SUC5</td>
<td>U*</td>
<td>G*C</td>
<td>U6 stem loop RNA</td>
<td>NMR</td>
<td>7</td>
<td>2.58</td>
<td>3.24</td>
<td>135.46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gene</td>
<td>Accession</td>
<td>Type</td>
<td>Mutant</td>
<td>Nucleotide Change</td>
<td>Description</td>
<td>Technique</td>
<td>Kd</td>
<td>Ki</td>
<td>IC50</td>
<td>EC50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>------</td>
<td>--------</td>
<td>------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>----</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1LDZ</td>
<td>A25-C6</td>
<td>G&gt;C</td>
<td>G&gt;C</td>
<td>Leucyl tRNA synthetase (full length)</td>
<td>NMR</td>
<td>5.5</td>
<td>3.00</td>
<td>2.83</td>
<td>156.52</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1NKW</td>
<td>A292-C369</td>
<td>G&gt;U</td>
<td>U&gt;A</td>
<td>Dicytosine Adenylates 23S RNA</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>2.96</td>
<td>3.66</td>
<td>157.50</td>
<td>Null</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>1N32</td>
<td>A1500-C1402</td>
<td>C&gt;G</td>
<td>A&gt;C</td>
<td>3OS Subunit</td>
<td>X-Ray Diffraction</td>
<td>6.5</td>
<td>3.00</td>
<td>2.52</td>
<td>161.30</td>
<td>75.17</td>
<td>70.51</td>
<td></td>
</tr>
<tr>
<td>106B</td>
<td>A926-C944</td>
<td>C&gt;G</td>
<td>C&gt;G, A</td>
<td>Glutamine aminosyl tRNA synthetase (I-Glu AmpCpp bound)</td>
<td>X-Ray Diffraction</td>
<td>7</td>
<td>2.85</td>
<td>3.18</td>
<td>154.60</td>
<td>43.70</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>1QRS</td>
<td>A26-C44</td>
<td>G&gt;C</td>
<td>A&gt;C</td>
<td>Glutamyl-tRNA synthetase Mutant</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>3.18</td>
<td>3.45</td>
<td>143.09</td>
<td>Null</td>
<td>31.25</td>
<td></td>
</tr>
<tr>
<td>1QTQ</td>
<td>A926-C944</td>
<td>G&gt;C</td>
<td>A&gt;C</td>
<td>Glutamyl-tRNA synthetase Mutant</td>
<td>X-Ray Diffraction</td>
<td>7</td>
<td>3.13</td>
<td>3.27</td>
<td>171.92</td>
<td>34.50</td>
<td>49.30</td>
<td></td>
</tr>
<tr>
<td>1U9S</td>
<td>A192-C178</td>
<td>G&gt;G</td>
<td>G&gt;G</td>
<td>Specificity domain of Ribonuclease P of the A-type</td>
<td>X-Ray Diffraction</td>
<td>6.5</td>
<td>3.12</td>
<td>2.72</td>
<td>152.64</td>
<td>64.90</td>
<td>87.00</td>
<td></td>
</tr>
<tr>
<td>1XMQ</td>
<td>A1500-C1402</td>
<td>G&gt;C</td>
<td>A&gt;C</td>
<td>AUC-triple &quot;like&quot;</td>
<td>X-Ray Diffraction</td>
<td>6.5</td>
<td>3.07</td>
<td>2.65</td>
<td>168.96</td>
<td>Null</td>
<td>77.07</td>
<td></td>
</tr>
<tr>
<td>1EHT</td>
<td>A10-C21</td>
<td>C&gt;C</td>
<td>C&gt;C</td>
<td>Thymine-RNA Complex</td>
<td>NMR</td>
<td>6.8</td>
<td>2.74</td>
<td>2.76</td>
<td>151.60</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2BTE</td>
<td>A38-C32</td>
<td>NONE</td>
<td>A&gt;U</td>
<td>Leucyl-tRNA synthetase + tRNA-Leu</td>
<td>X-Ray Diffraction</td>
<td>6.5</td>
<td>2.96</td>
<td>3.00</td>
<td>152.82</td>
<td>83.60</td>
<td>82.20</td>
<td></td>
</tr>
<tr>
<td>405D</td>
<td>Aa11-Cb22</td>
<td>A&gt;U</td>
<td>U&gt;A</td>
<td>GCAGACUUAUAUCUGC duplex</td>
<td>X-Ray Diffraction</td>
<td>7</td>
<td>2.80</td>
<td>2.52</td>
<td>166.32</td>
<td>Null</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>1L9A</td>
<td>A139-C158</td>
<td>G&gt;C</td>
<td>A&gt;C</td>
<td>SRP19 + SRP RNA Complex</td>
<td>X-Ray Diffraction</td>
<td>7.4</td>
<td>3.17</td>
<td>3.79</td>
<td>148.94</td>
<td>11.60</td>
<td>98.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1FK*</td>
<td>A1005-C963</td>
<td>C-G</td>
<td>A, C</td>
<td>50S ribosomal subunit (2.4 Å Resolution, pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>2.94</td>
<td>2.58</td>
<td>159.79</td>
<td>47.70</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A2485-C2104</td>
<td>U, C-G</td>
<td>A-G, G, A</td>
<td>50S ribosomal subunit (2.4 Å Resolution, pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>3.21</td>
<td>2.70</td>
<td>149.10</td>
<td>47.70</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>1GID</td>
<td>A104-C216</td>
<td>G-C</td>
<td>A-G</td>
<td>Group I Ribozyme Domain</td>
<td>X-Ray Diffraction</td>
<td>6</td>
<td>2.85</td>
<td>3.13</td>
<td>147.69</td>
<td>Null</td>
<td>45.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ab104-Cb216</td>
<td>G-C</td>
<td>A-G</td>
<td>Group I Ribozyme Domain</td>
<td>X-Ray Diffraction</td>
<td>6</td>
<td>2.80</td>
<td>2.92</td>
<td>147.45</td>
<td>Null</td>
<td>45.00</td>
<td></td>
</tr>
<tr>
<td>1JID</td>
<td>A139-C158</td>
<td>G-C</td>
<td>C-A+</td>
<td>SRP19 complexed with SRP RNA</td>
<td>X-Ray Diffraction</td>
<td>5.6</td>
<td>2.86</td>
<td>2.69</td>
<td>175.28</td>
<td>19.50</td>
<td>27.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A157-C140</td>
<td>C-A+</td>
<td>G-C</td>
<td>SRP19 complexed with SRP RNA</td>
<td>X-Ray Diffraction</td>
<td>5.6</td>
<td>3.02</td>
<td>2.68</td>
<td>167.22</td>
<td>19.50</td>
<td>27.63</td>
<td></td>
</tr>
<tr>
<td>1KFO</td>
<td>A6-CG,4</td>
<td>-</td>
<td>-</td>
<td>RNA Helix recognized by zinc-finger protein</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Null</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A6-CG,4</td>
<td>-</td>
<td>-</td>
<td>RNA Helix recognized by zinc-finger protein</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Null</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>1L1W</td>
<td>A5-C24</td>
<td>G-C</td>
<td>A-C</td>
<td>SRP 19 Splicing Domain</td>
<td>NMR</td>
<td>6.5</td>
<td>3.09</td>
<td>3.05</td>
<td>159.79</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A23-C6</td>
<td>G-C</td>
<td>A-C</td>
<td>SRP 19 Splicing Domain</td>
<td>NMR</td>
<td>6.5</td>
<td>2.94</td>
<td>3.02</td>
<td>170.59</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1M9O*</td>
<td>A1005-C963-C959</td>
<td>C-G</td>
<td>A, C</td>
<td>50S ribosomal subunit (pH 6.2)</td>
<td>X-Ray Diffraction</td>
<td>6.2</td>
<td>2.86</td>
<td>2.58</td>
<td>173.44</td>
<td>49.70</td>
<td>49.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A2485-C2104-C2536</td>
<td>A, A</td>
<td>C-G, U</td>
<td>50S ribosomal subunit (pH 6.2)</td>
<td>X-Ray Diffraction</td>
<td>6.2</td>
<td>2.92</td>
<td>2.51</td>
<td>148.20</td>
<td>49.70</td>
<td>49.60</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Name</td>
<td>Mixture</td>
<td>Mutation</td>
<td>Conditions</td>
<td>Technology</td>
<td>Energy</td>
<td>Tm</td>
<td>Kd</td>
<td>Kpm</td>
<td>Null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>---------</td>
<td>----------</td>
<td>-------------------------------------</td>
<td>------------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1S72*</td>
<td>A1005-C963-C959</td>
<td>A, C</td>
<td>C-G</td>
<td>50S ribosomal subunit (pH 7.5)</td>
<td>X-Ray Diffraction</td>
<td>7.5</td>
<td>2.94</td>
<td>2.57</td>
<td>166.41</td>
<td>47.70</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>A2485-C2104</td>
<td>A, A(A)</td>
<td>U, C-G</td>
<td>50S ribosomal subunit (pH 7.5)</td>
<td>X-Ray Diffraction</td>
<td>7.5</td>
<td>3.03</td>
<td>2.60</td>
<td>151.14</td>
<td>47.70</td>
<td>Null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1Z7F</td>
<td>Aa11-C22</td>
<td>U+G</td>
<td>A+U</td>
<td>16 base pair RNA duplex containing a C-Amismatch</td>
<td>X-Ray Diffraction</td>
<td>7</td>
<td>3.06</td>
<td>2.57</td>
<td>168.40</td>
<td>26.50</td>
<td>33.40</td>
<td></td>
</tr>
<tr>
<td>Aa27-C6</td>
<td>A+U</td>
<td>U+A</td>
<td>16 base pair RNA duplex containing a C-Amismatch</td>
<td>X-Ray Diffraction</td>
<td>7</td>
<td>3.12</td>
<td>2.74</td>
<td>165.80</td>
<td>26.50</td>
<td>33.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>402D</td>
<td>Aa104-Cb112</td>
<td>G+C</td>
<td>A+C</td>
<td>RNA Loop with tandem A+C mismatches</td>
<td>X-Ray Diffraction</td>
<td>5.5</td>
<td>3.54</td>
<td>3.11</td>
<td>118.37</td>
<td>Null</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>Aa113-Cb104</td>
<td>C+A+</td>
<td>G-C</td>
<td>RNA Loop with tandem A+C mismatches</td>
<td>X-Ray Diffraction</td>
<td>5.5</td>
<td>3.09</td>
<td>2.82</td>
<td>118.37</td>
<td>Null</td>
<td>Null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D8R</td>
<td>Ab5-C24</td>
<td>G+C</td>
<td>C+A</td>
<td>Helix 6; Human SRP</td>
<td>X-Ray Diffraction</td>
<td>5.5</td>
<td>3.02</td>
<td>2.62</td>
<td>145.93</td>
<td>Null</td>
<td>59.40</td>
<td></td>
</tr>
<tr>
<td>Aa5-C24</td>
<td>G+C</td>
<td>C+A+</td>
<td>Helix 6; Human SRP</td>
<td>X-Ray Diffraction</td>
<td>5.5</td>
<td>2.95</td>
<td>2.51</td>
<td>153.84</td>
<td>Null</td>
<td>59.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aa23-Cb6</td>
<td>G+C</td>
<td>C+A+</td>
<td>Helix 6; Human SRP</td>
<td>X-Ray Diffraction</td>
<td>5.5</td>
<td>2.93</td>
<td>2.51</td>
<td>154.72</td>
<td>Null</td>
<td>59.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample ID</td>
<td>Amino Acid Change</td>
<td>50S ribosomal subunit (no reported pH)</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>3.13</td>
<td>3.21</td>
<td>156.89</td>
<td>Null</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>---------------------------------------</td>
<td>------------------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1005-C963</td>
<td>C&gt;G A&gt;C</td>
<td>50S ribosomal subunit (no reported pH)</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>2.76</td>
<td>2.52</td>
<td>171.33</td>
<td>Null</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2485-C2104</td>
<td>C&gt;U A&gt;A</td>
<td>50S ribosomal subunit (no reported pH)</td>
<td>X-Ray Diffraction</td>
<td>Null</td>
<td>3.05</td>
<td>2.64</td>
<td>153.44</td>
<td>Null</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1005-C963</td>
<td>C&gt;G A&gt;C</td>
<td>50S ribosomal subunit (pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>3.11</td>
<td>2.51</td>
<td>158.20</td>
<td>21.34</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2485-C2104</td>
<td>C&gt;U C&gt;G</td>
<td>50S ribosomal subunit (pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>2.86</td>
<td>2.61</td>
<td>156.44</td>
<td>21.34</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2488-C2534</td>
<td>C&gt;U C&gt;G</td>
<td>50S ribosomal subunit (pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>2.91</td>
<td>2.57</td>
<td>166.41</td>
<td>21.34</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2485-C2104</td>
<td>C&gt;U A&gt;A</td>
<td>50S ribosomal subunit (pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>3.17</td>
<td>2.55</td>
<td>149.52</td>
<td>21.34</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1005-C963</td>
<td>C&gt;G A&gt;C</td>
<td>50S ribosomal subunit (pH 5.7)</td>
<td>X-Ray Diffraction</td>
<td>5.7</td>
<td>3.17</td>
<td>3.10</td>
<td>162.70</td>
<td>56.60</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2488-C2534</td>
<td>C&gt;U C&gt;G</td>
<td>50S ribosomal subunit (pH 5.8)</td>
<td>X-Ray Diffraction</td>
<td>5.8</td>
<td>2.91</td>
<td>2.57</td>
<td>166.41</td>
<td>56.60</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2485-C2104</td>
<td>C&gt;U C&gt;G</td>
<td>50S ribosomal subunit (pH 5.7)</td>
<td>X-Ray Diffraction</td>
<td>5.7</td>
<td>2.85</td>
<td>2.56</td>
<td>145.67</td>
<td>56.60</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reference Structures (Base Triples of Interest)

Figure A1 – CA+C N3-amino; amino-N3; amino-carbonyl

Figure A2 – ACA⁺ amino-carbonyl
Appendix B: Protonated Bases of Interest Within PDB Files

Figure B1 – Structures Containing a Single Protonated A^+C Wobble
Figure B2. Structures Containing Two Protonated A•C Wobble Base Pairs
A•C wobble pairs shown in red in grey structures below.

IFFK

1GID
Figure B3. Structures Containing Three Protonated A⁺·C Wobble Base Pairs
Figure B4. Structures Containing Four Protonated A•C Wobble Base Pairs
ACADEMIC VITA

Amarpreet Kaur Ahluwalia

73 Simmons Hall, University Park, PA 16802 • a.ahluwalia5080@gmail.com

Education

M.Ed. Candidate in PreK-12 Administration and Education
University of Missouri, College of Education, St. Louis, MO

B.S., Biochemistry and Molecular Biology, 2013, Penn State University, University Park, PA
Honors in Chemistry and Biochemistry and Molecular Biology
Minors: Chemistry, English

Honors and Awards

• Schreyer Honors College Academic Excellence Scholarship
• Conferences & Institutes Study Abroad Award: “Irish Landscape, Literature, and Visual Arts” (Summer 2010)
• University Libraries Award for Information Literacy, First Place (Spring 2012)
• 3M Chemistry Fellowship (Summer 2012)
• Dean’s List

Research and Professional Experience

A Computational and Experimental Approach to the Analysis of Ionized Nucleobases in Nature
December 2009 – May 2013
PI: Philip C. Bevilacqua, Ph.D.
Department of Chemistry
The Pennsylvania State University, University Park, PA

Light – A Particularly Wavelike Phenomenon
June – August 2008
The Pennsylvania Governor’s School for the Sciences
Carnegie Mellon University, Pittsburgh, PA

Chemistry Department, The Pennsylvania State University
Teaching Assistant/Grader – Chemistry 110H
Professor: Philip C. Bevilacqua, Ph.D.
**Campus Activities and Community Service**

- Teach for America 2013 Corps Member – St. Louis
- Schreyer Honors College Scholar Assistant – “The GLOBE”
  PSU Engineering – Academic Summer Enhancement Program Chemistry Tutor
- Big Brothers, Big Sisters – Site-based Big Sister
- Mortar Board National Honor Society, National Society of Collegiate Scholars

**Presentations**


**Refereed Publication**