The Pennsylvania State University

The Graduate School

Chemistry Department

DEVELOPMENT AND UTILIZATION OF SPECTROSCOPIC AND CHEMICAL METHODOLOGIES FOR INVESTIGATING pK_a SHIFTING IN NUCLEIC ACIDS

A Dissertation in

Chemistry

by

Jennifer Wilcox

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2013

The dissertation of Jennifer Wilcox was reviewed and approved* by the following:

Philip C. Bevilacqua
Professor of Chemistry
Dissertation Advisor
Chair of Committee

Scott Showalter
Assistant Professor of Chemistry

Squire Booker
Professor of Chemistry and Biochemistry and Molecular Biology

Andrey Krasilnikov
Associate Professor of Biochemistry and Molecular Biology

Barbara Garrison
Shapiro Professor of Chemistry
Head of the Department of Chemistry

*Signatures are on file in the Graduate School

ABSTRACT

RNA has historically been thought to play an intermediary role in the flow of genetic information between DNA and proteins. With its 2' hydroxyl, RNA is much less stable than DNA, which makes it amenable to a transitional role rather than one of storage. Proteins are equipped with 20 chemically diverse amino acid side chains, surpassing the functional potential of the four heterocycles that make up RNA. However, with the discovery of catalytic RNA (or ribozymes) in 1982, RNA was no longer limited to the transitional role it was once believed to have. The discovery of catalytic RNA shed light on RNA's potential to actively participate in biological processes.

Proteins can act catalytically as a result of an electrostatic contribution with pK_a's of amino acid side chains far from neutrality to form a charged species at neutrality or with side chain pK_a's near neutrality to allow for participation through proton transfer. However, the pK_a's of single stranded nucleobases are far from neutrality, inhibiting participation in catalysis. Because Watson-Crick base pairing requires the formation of hydrogen bonds, the pK_a's of the nucleobases are shifted *even further* from neutrality, significantly decreasing the potential for catalytic activity. The pK_a's of the nucleobases must be shifted towards neutrality in order to act catalytically in a manner analogous to amino acid side chains.

Some non-canonical interactions shift RNA pK_a's *towards* neutrality, enabling the nucleobases to act catalytically. Interactions that facilitate protonation of nucleobases occur in both secondary and tertiary motifs and allow RNA to play a much more diverse role than the intermediary role once thought. Protonated nucleobases participate in many biological processes such as general acid-base catalysis of ribozymes, programmed ribosomal frameshifting, miRNA processing, and RNA editing. We hypothesize that RNA systems can contain protonated motifs

that participate in biological processes but confirmation of protonation in these systems is limited by existing methods.

To determine the driving forces for pK_a shifting, the development of a new method for determining pK_a's was essential. Spectroscopic methods currently exist to quantify pK_a shifting in nucleic acid systems; however, spectroscopic analysis by these methods requires high sample concentration, costly sample preparation and lengthy experimental time. Additionally, some existing spectroscopic methods such as UV-Vis spectroscopy and CD require a large conformational change to produce an observable signal change between protonated and deprotonated states, therefore limiting the size of systems that can be analyzed. I developed a new method using fluorescence spectroscopy to determine pK_a's in nucleic acids. Development of this method utilized fluorescence quenching of 2-aminopurine (2AP), a fluorescent isomer of adenine, when stacked. Positioning the 2AP adjacent to the protonated base allowed for high fluorescence when the base pair wasn't formed and low fluorescence when the base pair was formed and the 2AP fluorescence was quenched due to stacking. The pK_a's obtained with fluorescence spectroscopy agreed with those obtained with a previously developed ^{31}P NMR method. The pK_a of adenine in an A^+•C wobble in a DNA hairpin was measured accurately with fluorescence at one-third of the cost and 10 percent of the required experimental time of ^{31}P NMR. In addition to utilization of the technique on DNA secondary motifs, I also determined the pK_a of a cytidine participating in a base quartet in the beet western yellows virus (BWYV) RNA. Protonation of C8 in the BWYV RNA enables pseudoknot formation and programmed ribosomal frameshifting.

Previous studies have been performed in the Bevilacqua lab to determine the salt, temperature, and context dependence of pK_a shifting in DNA while the sequence remained constant. In nature, however, sequence can also contribute to pK_a shifting. Using ^{31}P NMR, I determined the pK_a of adenine in an A^+•C wobble in an RNA hairpin in various sequence and

structural environments. Strengthening the base pairing of nearest neighbors from two weak A-U nearest neighbors to two strong G-C nearest neighbors enabled an additional pK_a shift of 1.6 units, up to 8.1. A similar dependence on the melting temperature and free energy on the oligonucleotide sequence was found through thermal denaturation experiments as both parameters were more favorable with increased nearest neighbor base pairing strength, indicating that the addition of stability and more favorable interactions contributes to the stability of the construct as well as pK_a shifting. The pK_a's determined were between 6.5 and 8.1, suggesting that sequence could play a role in tuning pK_a's near physiological pH. We hope that the trends presented in this study will serve as a model to aid in the determination of the roles protonated bases play in various dsRNA-mediated processes in biology.

The aforementioned studies use fluorescence and NMR spectroscopies to determine pK_a's in nucleic acids. In those cases, systems were designed with a suspected protonated nucleobase whose identity was known before quantifying pK_a shifting. Although spectroscopic methods are useful in determining the pK_a of nucleobases with elevated pK_a's to confirm protonation and quantify pK_a values, many systems contain protonated nucleobases of unknown identity and spectroscopic methods are unable to readily identify the protonated nucleobase. In an effort to address this problem, I made significant progress towards the development of a chemical methodology to damage and map protonated bases in large RNA systems utilizing elevated reactivity of cationic over neutral nucleobases. Ultimately, the development of a reagent to map protonated bases in large RNA systems should be possible but significant efforts need to be made in the future to determine the optimal reaction and analysis conditions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

2AP, 2-aminopurine

ADAR, adenosine deaminase acting on RNA

BTP, bis-tris-propane buffer

BWYV, beet western yellows virus

BWYV-dm, beet western yellows virus double mutant

dsRNA, double-stranded RNA

Fl, fluorescence

HDV, hepatitis delta virus

miRNA, micro-RNA

PAGE, polyacrylamide gel electrophoresis

RNAi, RNA interference

siRNA, short-interfering RNA

SS and SS', RNAs with two strong G-C nearest neighbor base pairs

SW and SW', RNAs with one strong GC and one weak AU nearest neighbor base pairs

$\mathbf{T_m}$**,** melting temperature

WW and WW', RNAs with two weak AU nearest neighbor base pairs

ACKNOWLEDGEMENTS

This work would not have been possible without the support of family, friends, professors and colleagues. I have had a unique experience with Dr. Scott Showalter, Dr. Squire Booker and Dr. Andrey Krasilnikov as my committee members. Each of them has played a significant role in my research for which I am appreciative. At University of Scranton, I would like to thank Dr. O'Malley and Jeff Hartley for helping me to realize how much I love chemistry and Dr. Baumann for reminding me of my strength when times get tough.

I have been blessed with Dr. Philip Bevilacqua as my mentor. I now understand the importance of cautious optimism; although, I'm pretty sure Phil played Devil's advocate to make sure I was never too cautious or too optimistic. Phil, in addition to teaching me science, you taught me how to be an ethical, well-rounded scientist and person. For that, I will be eternally grateful.

I appreciate all of the help my labmates have provided throughout my graduate career. To current labmates, specifically Chelsea Schmidt, Pallavi Thaplyal, and Dr. Durga Ghosh, thank you for your thoughtfulness, patience, and unending support. To former labmates Dr. Melissa Mullen, Dr. Joshua Sokoloski, and Dr. Rebecca Toroney- Between dance parties, people watching during class change time, and endless baymate discussions on fashion, politics, and Dear Prudence, I have loved every minute of grad school with you three.

I have been blessed with the company of fantastic people throughout my graduate career who have forever touched my life. Dr. Beth EH, thank you for always going out of your way to care for me. Jay Tarolli, thank you for bringing a little piece of Pine Bush into the middle of

nowhere PA and for giving me a beautiful roof over my head with constant laughter. I'll go ahead and say it for you…. you're welcome. Travis Cordes, thank you for caring for me throughout the most difficult part of graduate school. Your support has made every day bearable and I will be forever grateful. Margot Zaccardi, thank you for keeping my sanity from completely dwindling. I shutter to think about what the last 7 months of graduate school would have been like had we not been office-mates. No more fakin' it- we made it.

Above all, I would like to thank my family for their unending love and support. Allison, thank you for sharing your love for Penn State with me. Whether it was to teach me the words to 'Hail to the Lion' the whole drive to PSU or where to study in the library, you helped make this place home. Eric, I can't thank you enough for taking care of me like I was your own sister. Listing all of the things that you've done for me makes it seem like you just cleaned my apartment but you lifted a tremendous burden for which I'm forever grateful. I could not have made it through long night thesis writing sessions without FaceTiming with Daniel. He was the pick-me-up I needed to get me through the remaining hours of work. Mom and Dad, thank you for being so proud of me. I have been so lucky to have you both by my side and to know that you were there to help me through difficult times or to share in my triumphs. This work would not have been possible without your support and love.

Above all, don't fear difficult moments. The best comes from them.

– Rita Levi-Montalcini,

Recipient of the 1986 Nobel Prize in medicine and

Former longest living Nobel laureate at 103 years old.

Chapter 1

Introduction

[Portions of this chapter have been previously published: "Charged Nucleobases and Their Potential for RNA Catalysis" by Jennifer L. Wilcox, Amarpreet Ahluwalia, and Philip C. Bevilacqua in *Accounts of Chemical Research* 2011 **44** 1270-1279.]

Throughout history, RNA has been believed to act only as an intermediate between DNA, which stores the genetic code, and proteins, which catalyze biological reactions. It was not until recently that the diverse functional potential that RNA possesses was discovered. RNA can use folding to regulate gene expression,[1] as a result of metabolite binding to riboswitches,[2-4] and through RNAi, in which an miRNA or siRNA binds to mRNA to affect translation.[5-7] The discovery that RNA can catalyze reactions as an RNA enzyme (ribozyme)[8, 9] further supported the concept that RNA has additional roles in biological processes. One of the ways that RNA can have diverse structural and functional properties is due to ionization of the nucleobases through shifting of pK_a's towards neutrality. This chapter will focus on RNA structure and how pK_a shifting in nucleic acids can affect structure and function.

1.1 Nucleic Acid Structure

Nucleic acids are comprised of four nucleobases: Adenine (A), Cytosine (C), Guanine (G) and Thymine (T) in DNA or Uracil (U) in RNA (Figure 1-1A), which are planar, pseudo-aromatic ring systems and are further categorized by the make-up of their ring systems. Pyrimidines (C, T or U) are 6-membered heterocycles while purines (A and G) are fused-ring systems, made up of a 6-membered pyrimidine ring fused to a imidazole 5-membered ring. An N-

glycosidic bond at the N1 of pyrimidines and N9 of purines connects the nucleobase to a pentose sugar ring, which is bonded to a phosphate backbone to make a nucleotide (Figure 1-1B). Multiple nucleotides are bonded together to make oligonucleotides, which comprise a strand of DNA or RNA. Two strands of DNA or RNA base pair to form a helical structure (Figure 1-1C), which will be discussed in more detail in Section 1-1.[10]

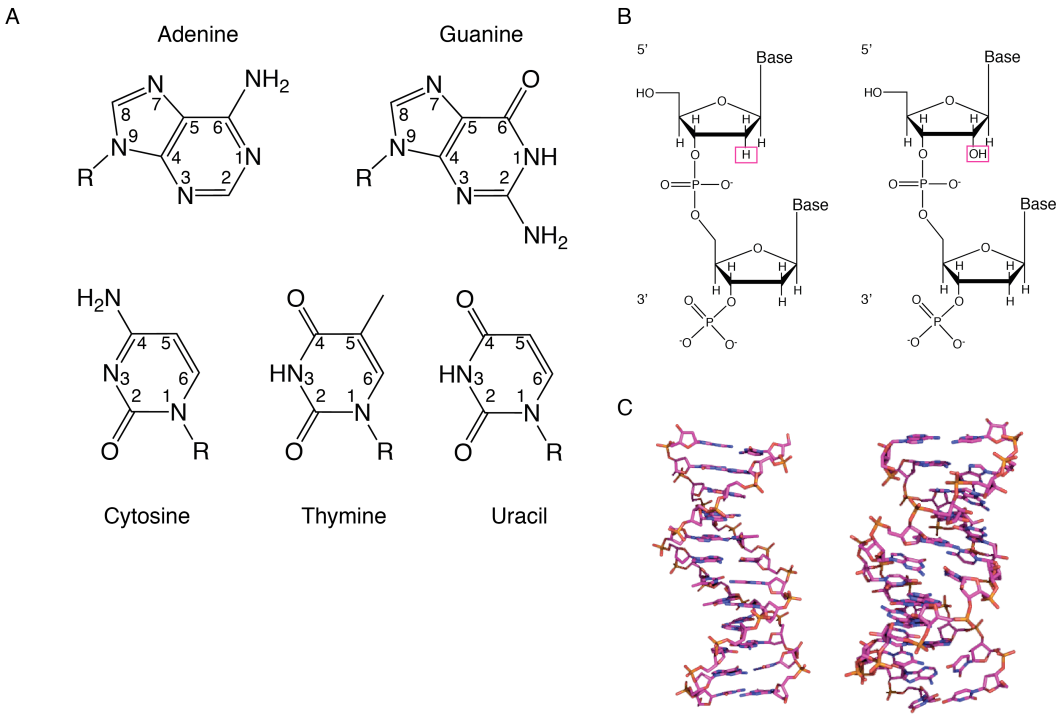


Figure **1-1**: Molecular make up of nucleic acids. (A) DNA and RNA nucleobases. Adenine and guanine are purines with fused ring systems. Cytosine, Thymine in DNA and Uracil in RNA are 6-membered pyrimidine rings. The nucleobases are attached to the pentose ring through N-glycosidic bonds at the N9 of purines and N1 of pyrimidines. (B) Both DNA (left) and RNA (right) are made up of nucleotides comprised of a nucleobase, pentose sugar ring and phosphate. DNA has a hydrogen and RNA a hydroxyl at the 2' position of the sugar ring, boxed in pink. (C) DNA and RNA form helices. DNA (left) typically forms a B-form helix while RNA (right) typically forms an A-form helix. DNA adapted from PDB 1ELH and RNA from PDB 1BNA.

The main molecular difference between DNA and RNA is the substitution of a hydrogen at the 2' position in DNA for the hydroxyl group in RNA (Figure 1-1B). This seemingly inconsequential substitution gives rise to a plethora of differences between the nucleic acids'

structure and function. Both DNA and RNA can form a duplex, in which two strands base pair to form a helical structure but the conformation upon helix formation is very different. In nature, DNA is typically double stranded while RNA is more commonly single stranded, adopting more complex folding patterns. DNA is typically found in a B-form helical structure, as shown in Figure 1-1C, which has a wide major groove, a narrow minor groove and, overall, a more elongated structure with 3.4 Å rise per base pair. Double stranded RNA, however, prefers an A-form helix, which has a narrow, deep major groove, a wide, shallow minor groove, and a more compact structure with just 2.9 Å rise per base pair. Molecular variance not only leads to structural variability but also to function. RNA's 2' hydroxyl makes it susceptible to alkaline degradation. Because of this, the genetic code is stored in DNA and transcribed to RNA.[10]

The nucleobases interact with one another through hydrogen bonding, a noncovalent interaction between an electronegative atom (in nucleic acids N or O) and a proton bonded to another electronegative atom. The functionalities of the nucleobases can be broken down into two groups: hydrogen bond donors and acceptors. A donor functional group can hydrogen bond with an acceptor if facilitated by geometry. G and C base pair to form three hydrogen bonds and A and U base pair to form two hydrogen bonds (Figure 1-2A). G-C and A-U base pairs are also referred to as Watson-Crick base pairs and are the strongest base pairs formed in RNA, forming hydrogen bonds without distortion of the phosphate backbone structure. Non-canonical base pairs can form, some with minor disruption to the A-form phosphate backbone (Examples in Figure 1-2B).[10]

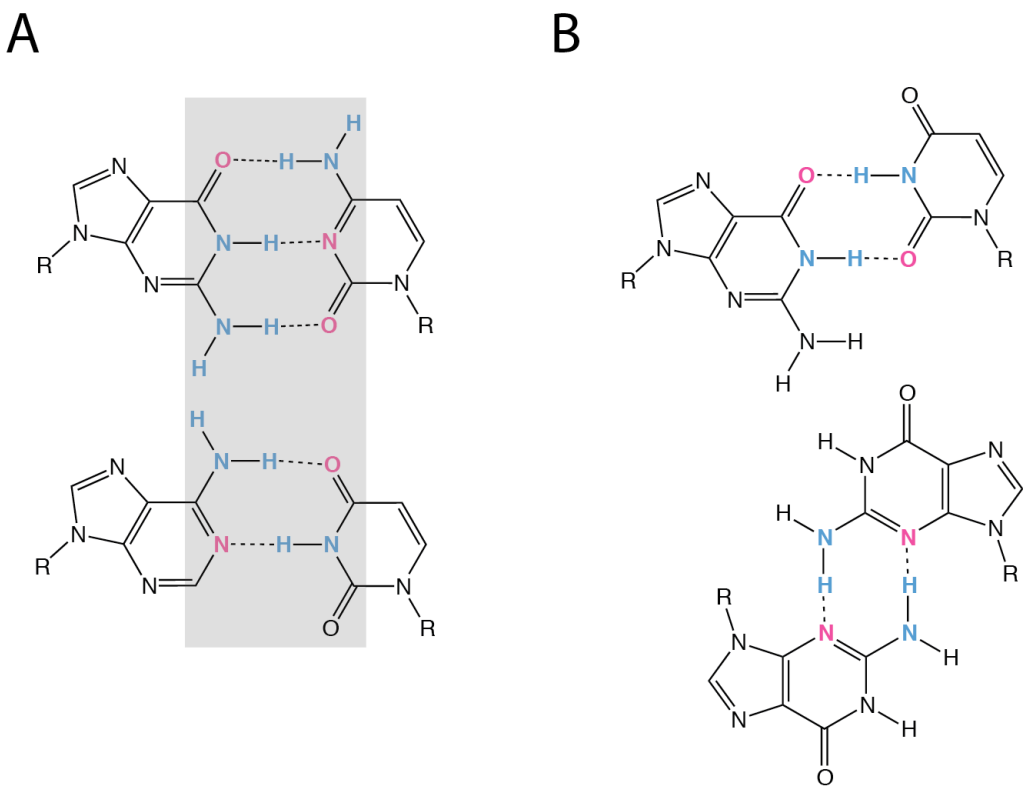


Figure **1-2**: Base pairing in RNA is driven by hydrogen bonding. Hydrogen bond donors (labeled in blue) form hydrogen bonds with acceptors (labeled in pink). (A) Watson-Crick base pairs. G and C base pair to form three hydrogen bonds and A and U with two. Hydrogen bonding occurs on the Watson-Crick face, shaded in gray. (B) Examples of non-canonical base pairs. In a G-U wobble, two base pairs are formed with minor distortion to the phosphate backbone. In a G-G symmetrical base pair, interactions between the N3 and amino group of C2 are observed, but one of the G's must be in the syn conformation, or the strands run parallel to each other.

In addition to hydrogen bond formation, bases interact in a vertical manner through stacking of the π molecular orbitals of the ring systems Because of the greatest potential for stacking between the larger ring system, stacking is strongest among purine-purine rings.

Nucleic acids typically fold in a hierarchical manner, forming primary, secondary, then tertiary structures (Figure 1-3).[11, 12] A commonly occurring secondary motif is the hairpin, also known as a stem-loop, which consists of a base paired region (stem) and a loop of unpaired nucleotides. A hairpin can be fully base paired, or have a mismatch or bulge, or any combination

of these elements (Figure 1-3B). As mentioned previously, RNA is most commonly single stranded, permitting the formation of more complex combinations of secondary structure motif through tertiary interactions (Figure 1-3C). The diverse structures that RNA adopts leads to its functional diversity.[10]

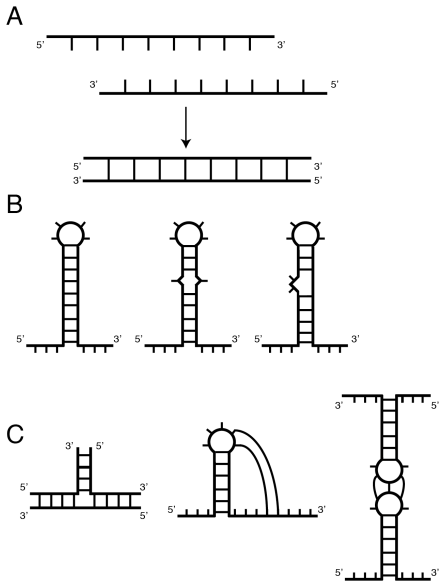


Figure **1-3**: Hierarchical folding of nucleic acids. (A) Two strands of oligonucleotides base pair to form a duplex. It is important to note that nucleic acids base pair in an anti-parallel manner with one strand from 5' to 3' and the other from 3' to 5'. (B) Common secondary structure motifs. A hairpin, also referred to as a stem-loop is made up of a base paired stem and a single stranded end and typically unpaired nucleotides at the base of the stem. The stem can be completely base paired (left), contain a mismatch (center) or a bulge (right) or any combination of these structural motifs. (C) Common tertiary structure motifs. Multiple secondary structures can combine to form a tertiary interaction such as a three way helical junction (left), a pseudoknot (center) or a kissing loop (right), among others.

1.2 Nucleic acid Function

As described in the Central Dogma of Molecular Biology, nucleic acids' primary function is to transfer the genetic code from the form in which it is stored (DNA) to its catalytically active form (protein), consisting of transcription of DNA to RNA and translation of

RNA to protein. Transcription begins with DNA, the more stable nucleic acid, which binds to a polymerase. Each DNA base templates which RNA base is to be added opposite it through Watson-Crick base pairing. The sequence of the DNA is read by the polymerase and the complementary RNA strand is synthesized. Transcription produces many different RNAs that can be further categorized as either a messenger RNA (mRNA), which codes for proteins, or non-coding RNAs, which have been shown to have a multitude of functions. A messenger RNA (mRNA) is then translated into a protein in the ribosome. The combination of three RNA bases (codon) specifies which amino acid should be added to the protein during translation, beginning with a start codon and ending with a stop codon. The transfer of genetic information from DNA to RNA to proteins is imperative for an organism to live.[13]

Since its discovery in the 1950's, there have been many additions to the processes of transcription and translation. It has since been discovered that RNA can be used to synthesize DNA in a process called reverse transcription.[14, 15] DNA and RNA can both be used to make more of themselves through DNA and RNA replication, respectively.[16, 17] And lastly, in the presence of antibiotics, DNA can make proteins through direct translation, bypassing the RNA intermediate.[18]

RNA has its own important functions, however, and can directly regulate translation through a silencing mechanism known as RNA interference (RNAi).[5-7] In RNAi, a pri-miRNA is transcribed, cleaved by Drosha-DGCR8 and Dicer-- two large protein complexes-- to form a mature miRNA. That miRNA then base pairs with mRNA, inhibiting translation and triggering degradation of the mRNA by the RISC.[19] Secondly, in ribosomal frameshifting, RNA forms a frameshift element through tertiary structure, creating a kinetic barrier that causes the ribosome to pause during translation, slip, and misread the mRNA its decoding.[20] This process is commonly employed by viruses with small genomes in order to code for multiple proteins by a single mRNA. Third, upon binding of a certain metabolite, RNA can inhibit or activate protein

synthesis, acting as a riboswitch.[2-4] These are just a few examples of the diverse roles RNA can play beyond the intermediary role it was once assumed to have.

1.3 pK_a Shifting in Nucleic Acids

A major difference between RNA and protein enzymes is the lack of diversity of the RNA side chains. Proteins are composed of 20 chemically diverse amino acids, whereas RNA has just four similar heterocycles.[21] Amino acid side chains participate in catalysis in one of several ways: they can have a pK_a shifted far from neutrality to form a charged species, enabling participation in an electrostatic manner or they have an ionizable proton with a pK_a near neutrality.[22] It is well established that the charged amino acids play key roles in protein enzyme catalysis: lysine and arginine act as oxyanion holes to stabilize charge development in the transition state, whereas histidine cycles between neutral and positive when it transfers protons. Thus, one might inquire whether the nucleobases play similar roles.

Unlike amino acids, the pK_a's of single stranded nucleobases are far from neutrality (Figure 1-4). A and C have intrinsic pK_a's of 3.5 and 4.2, respectively, while G and T or U have pK_a's of 9.2.[21] Watson-Crick base pairing shifts pK_a's even *farther* from neutrality[23], decreasing the potential for participation in general acid-base catalysis (Figure 1-4B). In a G-C base pair, because protonation at N1 facilitates hydrogen bonding and base pair formation, the pK_a is shifted 2-3 units above the pK_a of single-stranded G. Similarly, protonation at the N3 of C would inhibit hydrogen bonding and base pair formation with G, causing the pK_a to shift 2-3 units below the pK_a of single-stranded C. The pK_a's in an A-U Watson-Crick base pair are shifted further away from neutrality as well, shifting the pK_a of N1 of A 2-3 units below and the pK_a of N3 of U 2-3 units above their single stranded counterparts.

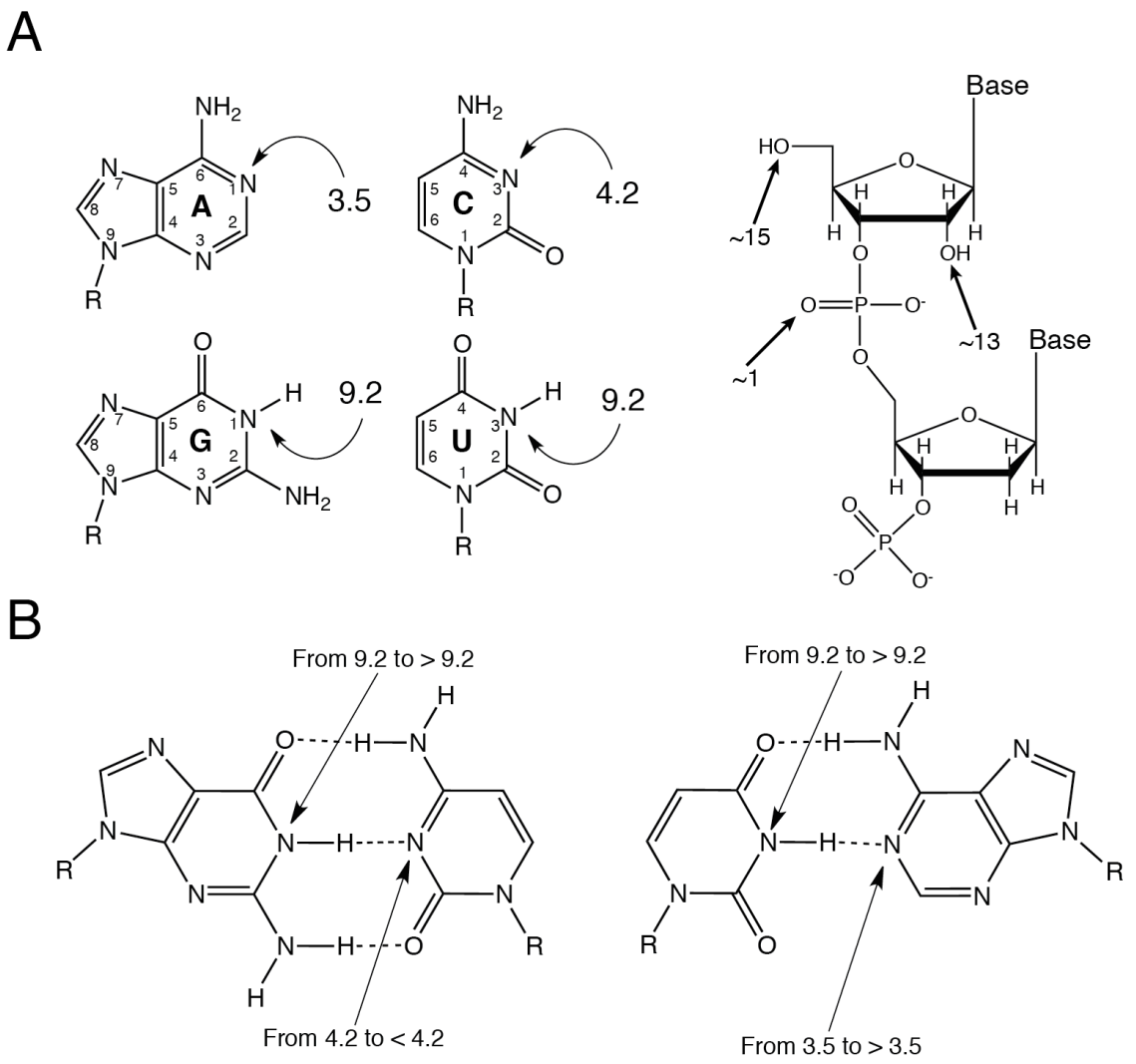


Figure **1-4**: pK_a's of nucleic acids. (A) pK_a's of nucleobases. A and C have pK_a's below neutrality, of 3.5 and 4.2, respectively, giving rise to their deprotonated neutral states at physiological pH. G and U have pK_a's of 9.2 and are both neutral and protonated at physiological pH. The pK_a's of the sugar and phosphate are also removed from neutrality. (B) Watson-Crick base pairing shifts pK_a's even further from neutrality since protons are sequestered in hydrogen bonding.

With Watson-Crick base pairing causing pK_a's of the nucleobases to shift further away from neutrality, it is of interest to understand how nucleic acids can act enzymatically. Interestingly, examples are known where A and C shift toward neutrality. Such pK_a shifts are

divided into “Class I” and “Class II”, depending on whether the loaded proton is sequestered in base pairing (Figure 1-5) and both classes have potential to participate in catalysis.[24] Class I primarily serves as an oxyanion hole since the loaded proton participates in base pairing. In Class II, the proton is not sequestered in base pairing, therefore is available to participate in proton transfer. These roles are not exclusive, however, owing to acidification of external amino protons upon protonation of an imino proton.[25] pK_a values for Class I motifs from 7 to 8 have been reported in model systems, indicating potential for populating the protonated state at neutrality, and lysine/arginine-like behavior.[24] Also, pK_a values for Class II motifs near 7 have been reported for A and C, indicating potential for histidine-like behavior.[24]

Figure **1-5**: Protonated bases pairs in RNA.[24] (A) Examples of Class I pK_a shifting are shown. In Class I, the ionized proton (pink) is sequestered in base pairing, enabling these base pairs to play important structural and electrostatic roles. (B) Examples of Class II pK_a shifting are shown. In Class II pK_a shifts, the ionized proton (shown in pink) is not sequestered in base pairing, allowing for direct participation catalytically. (C) Proposed mechanism for general acid-base catalysis of the HDV ribozyme.[26]

Because Class II protons are not sequestered in base pairing, they can directly participate in general acid-base catalysis in an RNA enzyme (ribozyme).[24] Small ribozymes such as the hepatitis delta virus (HDV),[27] hairpin,[28] hammerhead,[29] Varkund satellite (VS)[30] and *glmS*[31] ribozymes all employ a mechanism similar to that shown in Figure 1-5C, in which the 2’ OH acts as a nucleophile and attacks an adjacent phosphate. Cleavage of the phosphate backbone and donation of a nearby proton occur, leaving the product bases with a 2’,3’-cyclic phosphate and a terminal 5’ OH.[32] In contrast, large ribozymes such as RNase P,[33] the Group I,[34] and Group II introns[35] utilize an exogenous nucleophile that attacks a nearby phosphate, causing cleavage of the backbone and produce a 2’,3’-*cis*-diol and 5’ phosphate termini typically using mutations rather than side chains.[32]

1.4 pK_a Determination

To fully understand the degree to which pK_a's are shifted towards neutrality, spectroscopic methods such as ^{1}H-, ^{13}C-, and ^{15}N-detected NMR,[36-38] UV-vis,[39] CD,[40] and Raman crystallography[41, 42] have been developed to determine pK_a's in nucleic acids. Although these methods have been successfully applied to determine pK_a's of nucleic acids, each has significant limitations. NMR techniques were the first to be employed to determine pK_a's in nucleic acids and are still widely used, however NMR requires high sample concentration and isotopic labeling of ^{13}C or ^{15}N, which is difficult and costly to obtain. Both UV-Vis and CD require a large conformational change to produce an observable change in signal from protonated to deprotonated states, which can make these methods problematic when attempting to analyze large RNA systems where any conformation change may be small on a percentage basis. Since in-solution Raman spectroscopy is relatively insensitive, Raman crystallography was used by our lab to determine the pK_a of C75 in the hepatitis delta virus (HDV) ribozyme using HDV crystals, increasing the Raman signal,[41] however crystals are often difficult to grow, adding complications to sample preparation and the pK_a is not for the molecules free in solution.

Our lab also developed a method using ^{31}P NMR to determine the salt, temperature and context dependence of pK_a shifting in DNA.[43, 44] This method requires the substitution of a phosphorothioate into the phosphate backbone of the nucleic acid, which shifts the peaks of interest downfield and separate from the phosphate backbone or phosphate envelope. The change in chemical shift is directly related to the pH, allowing for determination of the pK_a of the ionized nucleobase. The main limitation of this method is the requirement of large sample concentration, typically millimolar concentrations. Phosphorothioate modification also requires the chemical synthesis or purchase of modified nucleic acids, which adds to sample cost. The amount of time required for pK_a determination is also a limitation of this method. An NMR titration entails

approximately 20 hours of active experimenter time, depending on the strength of the magnet being used. For example, I found a titration using a 360 MHz magnet to take 38 hours and 14 hours on a 600 MHz magnet. This lengthy experiment time poses as a danger to sample integrity. Subjecting RNA to high temperature or high magnesium for the duration of the experiment could be detrimental to the sample, causing degradation.

Recently, the fluorophore 8-aza-G has been used as a reporter for a pK_a in the hairpin and *glmS* ribozymes.[45-48] On the one hand, this base is relatively non-perturbing. However, it has several limitations. One of the main limitations of this method is the preparation of 8-aza-G which requires eight enzyme transformations. Moreover, only five of the eight enzymes are available commercially with purification of the unavailable enzymes (rbsK, prsA, and glyD) found in literature.[49, 50] Using this method, a pK_a of >10.6 was extrapolated for G8 in a complex analogous to domain A in the hairpin ribozyme, although the nature of the experiment required termination of the titration at approximately pH 10, creating difficulties in determining an accurate pK_a. Because of the pK_a elevation of 8-aza-G versus unmodified G, the unmodified pK_a is not always observed, causing difficulties.

As discussed throughout this section, understanding the driving forces for pK_a shifting in nucleic acids would benefit the scientific community, providing insight into mechanism for general acid-base catalysis and how protonation can affect biological processes through structure formation. However, the nucleic acid field lacks a method for determination of pK_a's quickly, inexpensively and accurately. The development of a new method for determining pK_a's in nucleic acids is vital to the advancement of this field.

In addition, pK_a determination has been limited to select RNAs of interest and few studies have explored the driving forces for pK_a shifting in nucleic acids. Recently, the salt, temperature and context dependence was performed on model DNA oligonucleotides by our lab.[44] Although that study contributed greatly to the trends of pK_a shifting in DNA, the focus of that work was

how external environmental factors affect pK_a shifting without consideration of how sequence affects pK_a shifting. Knowledge of how sequence affects pK_a's could be used as a preliminary means of predicting which base pairs are protonated in biological systems without directly determining the pK_a of the nucleobase in question.

1.5 Thesis objectives

As discussed in Section 1.3, the existing methods for determining pK_a's in nucleic acids have significant limitations. **Chapter 2** describes the development of a method using 2-aminopurine fluorescence to determine the pK_a in DNA oligonucleotides. When inserting a reporter into a system, it is problematic if the reporter perturbs the system on which it is reporting. Comparison of the pK_a values obtained using fluorescence to pK_a's using ^{31}P NMR were in good agreement, indicating that insertion of the reporter 2-aminopurine did not significantly perturb the pK_a at least in these instances. The fluorescence method was then used to determine the pK_a of a protonated C in the tertiary structure of beet western yellows virus RNA. Protonation of C8 in this system enables formation of a pseudoknot, which causes ribosomal frameshifting during translation. Since fluorescence is a highly sensitive method, the amount of sample required for these assays is significantly decreased compared to NMR, which also decreases the sample cost. Experimental time is also significantly improved, as a fluorescence titration can be done in as little as an hour. Efforts towards high throughput application of this method are discussed.

Chapter 3 examines the sequence and context dependence of pK_a shifting in RNA. In this study, the nearest neighbors of an A^+•C base pair were varied from a two weak Watson-Crick base pairs, to one weak and one strong Watson-Crick base pair, to two strong Watson-Crick base pairs. A difference of 1.5 pK_a units was seen for these three constructs. As shown in Section 1-1

RNA tends to adopt more complex structures so the effect of bulge addition on the pK_a was examined. Addition of a bulge two base pairs away from the protonated base perturbed the pK_a by just 0.5 pK_a units but one base pair away disrupted helix formation throughout the stem. These data give insights into the magnitude of pK_a shifts under physiological conditions. Since sequence and context varies in biologically relevant RNA's, these data provide trends for predicting if protonated bases pairs will form in physiological environments.

As mentioned, several spectroscopic methods exist to report on the pK_a of a nucleobase, but current methods lack the ability to detect which base is ionizing. Since identification of the ionized nucleobase would be beneficial to gain mechanistic insight, a chemical method to determine which base is protonated in a large RNA system was pursued**. C**hapter 4 comprises efforts made towards the development of a method for the identification of a protonated nucleobase in a large RNA system. Various reagents were used in an attempt to chemically damage the protonated base. Two different approaches for analysis were used to detect damages: reverse transcription and mass spectrometry. In the case of reverse transcription, the RNA samples were treated with a chosen reagent, such as bisulfite, which damages the Watson-Crick face, and then analyzed using denaturing polyacrylamide gel eletrophoresis, anticipating a stop in reverse transcription one position before the damaged base. For mass spectrometric detection, a reagent was also added and the samples were analyzed by mass spectrometry to determine if the reagent covalently bonded to the sample, which would have been evident by the shift in mass. Progress was made towards the development of a chemical method for mapping protonated bases in large RNA systems using both reverse transcription and mass spectrometry as analysis techniques.

1.6 References

1. Lee, F., and Yanofsky, C. (1977) Transcription termination at the trp operon attenuators of Escherichia coli and Salmonella typhimurium: RNA secondary structure and regulation of termination, *Proc Natl Acad Sci U S A* **74**, 4365-4369.

2. Winkler, W. C., and Breaker, R. R. (2003) Genetic control by metabolite-binding riboswitches, *Chembiochem* **4**, 1024-1032.

3. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., and Breaker, R. R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria, *Cell* **113**, 577-586.

4. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine, *Nat Struct Biol* **10**, 701-707.

5. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs, *Science* **294**, 853-858.

6. Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans, *Science* **294**, 858-862.

7. Lee, R. C., and Ambros, V. (2001) An extensive class of small RNAs in Caenorhabditis elegans, *Science* **294**, 862-864.

8. Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981) In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence, *Cell* **27**, 487-496.

9. Baer, M., and Altman, S. (1985) A catalytic RNA and its gene from Salmonella typhimurium, *Science* **228**, 999-1002.

10. Bloomfield, V. C., DM. Tinoco, I (2000) Nucleic Acids- Structures, Properties, and Functions, University Science Books, Sausalito.

11. Tinoco, I., Jr., and Bustamante, C. (1999) How RNA folds, *J Mol Biol* **293**, 271-281.

12. Brion, P., and Westhof, E. (1997) Hierarchy and dynamics of RNA folding, *Annu Rev Biophys Biomol Struct* **26**, 113-137.

13. Crick, F. H. (1958) On protein synthesis, *Symp Soc Exp Biol* **12**, 138-163.

14. Temin, H. M., and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus, *Nature* **226**, 1211-1213.

15. Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature* **226**, 1209-1211.

16. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* **239**, 487-491.

17. Commoner, B., Lippincott, J. A., and Symington, J. (1959) Replication of tobacco mosaic virus. Kinetics of virus protein and ribonucleic acid biosynthesis, *Nature* **184**, 1992-1998.

18. McCarthy, B. J., and Holland, J. J. (1965) Denatured DNA as a direct template for in vitro protein synthesis, *Proc Natl Acad Sci U S A* **54**, 880-886.

19. He, L., and Hannon, G. J. (2004) MicroRNAs: small RNAs with a big role in gene regulation, *Nat Rev Genet* **5**, 522-531.

20. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression, *Nature* **331**, 280-283.

21. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer-Verlag, New York.

22. Fersht, A. R. (1985) Enzyme Structure and Mechanism, 2nd ed., Freeman, New York.

23. Legault, P., and Pardi, A. (1997) Unusual Dynamics and pKa Shift at the Active Site of a Lead-Dependent Ribozyme, *J. Am. Chem. Soc.* **119**, 6621-6628.

24. Bevilacqua, P. C., Brown, T. S., Nakano, S., and Yajima, R. (2004) Catalytic roles for proton transfer and protonation in ribozymes, *Biopolymers* **73**, 90-109.

25. Gueron, M., and Leroy, J. L. (1995) Studies of base pair kinetics by NMR measurement of proton exchange, *Methods Enzymol* **261**, 383-413.

26. Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme, *Science* **287**, 1493-1497.

27. Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. (2010) A 1.9 A crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage, *Biochemistry* **49**, 6508-6518.

28. Wilson, T. J., Nahas, M., Araki, L., Harusawa, S., Ha, T., and Lilley, D. M. (2007) RNA folding and the origins of catalytic activity in the hairpin ribozyme, *Blood Cells Mol Dis* **38**, 8-14.

29. Blount, K. F., and Uhlenbeck, O. C. (2005) The structure-function dilemma of the hammerhead ribozyme, *Annu Rev Biophys Biomol Struct* **34**, 415-440.

30. Lilley, D. M. (2004) The Varkud satellite ribozyme, *RNA* **10**, 151-158.

31. Ferre-D'Amare, A. R. (2010) The glmS ribozyme: use of a small molecule coenzyme by a gene-regulatory RNA, *Q Rev Biophys* **43**, 423-447.

32. Bevilacqua, P. C., and Yajima, R. (2006) Nucleobase catalysis in ribozyme mechanism, *Curr Opin Chem Biol* **10**, 455-464.

33. Esakova, O., and Krasilnikov, A. S. (2010) Of proteins and RNA: the RNase P/MRP family, *RNA* **16**, 1725-1747.

34. Vicens, Q., and Cech, T. R. (2006) Atomic level architecture of group I introns revealed, *Trends Biochem Sci* **31**, 41-51.

35. Michel, F., Costa, M., and Westhof, E. (2009) The ribozyme core of group II introns: a structure in want of partners, *Trends Biochem Sci* **34**, 189-199.

36. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs, *Biochemistry* **30**, 8242-8251.

37. Legault, P., and Pardi, A. (1994) In situ Probing of Adenine Protonation in RNA by 13C NMR, *J. Am. Chem. Soc.* **116**, 8390-8391.

38. Wang, C., Gao, H., Gaffney, B. L., and Jones, R. A. (1991) Nitrogen-15-labeled oligodeoxynucleotides. 3. Protonation of the adenine N1 in the A.cntdot.C and A.cntdot.G mispairs of the duplexes {d[CG(15N1)AGAATTCCCG]}2 and {d[CGGGAATTC(15N1)ACG]}2, *J. Am. Chem. Soc.* **113**, 5486-5488.

39. Moody, E. M., Lecomte, J. T., and Bevilacqua, P. C. (2005) Linkage between proton binding and folding in RNA: a thermodynamic framework and its experimental application for investigating pKa shifting, *RNA* **11**, 157-172.

40. Abeysirigunawardena, S. C., and Chow, C. S. (2008) pH-dependent structural changes of helix 69 from Escherichia coli 23S ribosomal RNA, *RNA* **14**, 782-792.

41. Gong, B., Chen, J. H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. (2007) Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography, *J. Am. Chem. Soc.* **129**, 13335-13342.

42. Guo, M., Spitale, R. C., Volpini, R., Krucinska, J., Cristalli, G., Carey, P. R., and Wedekind, J. E. (2009) Direct Raman measurement of an elevated base pKa in the active site of a small ribozyme in a precatalytic conformation, *J. Am. Chem. Soc.* **131**, 12908-12909.

43. Moody, E. M., Brown, T. S., and Bevilacqua, P. C. (2004) Simple method for determining nucleobase pK(a) values by indirect labeling and demonstration of a pK(a) of neutrality in dsDNA, *J. Am. Chem. Soc.* **126**, 10200-10201.

44. Siegfried, N. A., O'Hare, B., and Bevilacqua, P. C. (2010) Driving forces for nucleic acid pK(a) shifting in an A(+).C wobble: effects of helix position, temperature, and ionic strength, *Biochemistry* **49**, 3225-3236.

45. Da Costa, C. P., Fedor, M. J., and Scott, L. G. (2007) 8-Azaguanine reporter of purine ionization states in structured RNAs, *J. Am. Chem. Soc.* **129**, 3426-3432.

46. Liu, L., Cottrell, J. W., Scott, L. G., and Fedor, M. J. (2009) Direct measurement of the ionization state of an essential guanine in the hairpin ribozyme, *Nat Chem Biol* **5**, 351-357.

47. Cottrell, J. W., Scott, L. G., and Fedor, M. J. (2011) The pH dependence of hairpin ribozyme catalysis reflects ionization of an active site adenine, *J Biol Chem* **286**, 17658-17664.

48. Viladoms, J. l., Scott, L. G., and Fedor, M. J. (2011) An Active-Site Guanine Participates in glmS Ribozyme Catalysis in Its Protonated State, *J. Am. Chem. Soc.* **133**, 18388-18396.

49. Tolbert, T. J., and Williamson, J. R. (1996) Preparation of Specifically Deuterated RNA for NMR Studies Using a Combination of Chemical and Enzymatic Synthesis,Ä†, *J. Am. Chem. Soc.* **118**, 7929-7940.

50. Scott, L. G., Tolbert, T. J., and Williamson, J. R. (2000) Preparation of specifically 2H- and 13C-labeled ribonucleotides, *Methods Enzymol* **317**, 18-38.

Chapter 2

A Simple Fluorescence Method for pK_a Determination in RNA and DNA Reveals Highly Shifted pK_a's

[Published as a paper entitled "A Simple Fluorescence Method for pK_a Determination in RNA and DNA Reveals Highly Shifted pK_a's" by Jennifer L. Wilcox and Philip C. Bevilacqua in *Journal of the American Chemical Society* **2013**, *135* 7390-7393.]

2.1 Abstract

Charged nucleobases exist in RNA and DNA at neutral pH owing to pK_a shifting. These bases can affect polymerase fidelity and participate in ribozyme general acid-base catalysis. In addition, protonated RNA bases influence miRNA processing and viral frameshifting. It is therefore important to have a simple and rapid method for determining the pK_a of nucleobases in RNA and DNA. Here we describe the development of a method in which we use 2-aminopurine (2AP), a fluorescent isomer of adenine, to report on the pK_a of a nearby ionizing base in both in DNA secondary structure and RNA tertiary structure. We observe large, up to 5-fold, quenching in fluorescence upon protonation of a nearby base. Using this method, we identify highly shifted pK_a's of 7.6 for adenine in a DNA oligonucleotide and 8.15 for cytidine in a tertiary structure element from beet western yellows virus (BWYV) RNA. These pK_a values, which were corroborated by ^{31}P NMR measurements and comparison to literature, are shifted over 4 units from their unpaired standard values. This versatile fluorescence method can be used to determine pK_a's for ionization of both A and C and suggests that shifted pK_a's are prevalent in DNA and RNA secondary and tertiary structures.

2.2 Introduction

The unpaired nucleobases in RNA and DNA are typically uncharged, with their pK_a's far removed from neutrality. At biological pH, both A and C are usually deprotonated at their imino nitrogen, with unperturbed pK_a's in their unpaired states of 3.5 and 4.2, respectively, while G and U/T typically have protonated imino nitrogens with pK_a's of ~9.3.[1, 2] Upon Watson-Crick base pair formation, the pK_a's shift even further away from neutrality, by two or more pK_a units, because of hydrogen bonding, which makes charged nucleobases even less probable at neutral pH.[3]

In certain specialized cases, however, non-Watson-Crick base pairing or a highly negative potential can shift pK_a's *towards* neutrality, enabling the formation of structures with positively charged A and C bases and potentially negatively charged G and U bases.[4] These cases are of interest because they allow nucleobases to influence the fidelity of replication, miRNA processing, and viral frameshifting; also, they can facilitate the participation of the bases in general acid-base catalysis.[5-8] In particular, protonated A^+•C base pairs are found in both DNA and RNA, where they are important for structure and function in the leadzyme, spliceosome and viral RNA.[4]

Beet western yellows virus (BWYV) RNA belongs to a group of viral RNAs that utilize -1 ribosomal frameshifting to regulate translation and requires formation of a pseudoknot to initiate a change in the reading frame. Protonation of C8 in BWYV allows the pseudoknot structure to form at neutral pH. The C8 residue participates in a base quartet with G12, C26 and A25.[9] An analogous base quadruple is found in the C41 quartet in the genomic hepatitis delta virus (HDV) ribozyme,[10] and in other viral frameshifting RNAs.[11] Mutational studies indicate that this quartet is critical for viral frameshifting. The pK_a of the protonated base has been estimated

from optical monitored thermal melting using a thermodynamic linkage model, but has not been measured directly.[12]

The nucleobase 2-aminopurine (2AP) is fluorescent and has been widely used to study the dynamics and folding of RNA and DNA.[13-16] 2AP is structurally similar to adenine and it readily forms a two-hydrogen bond base pair with thymine in DNA (Figure 1A). The fluorescence nucleobase analogue's fluorescence is quenched by stacking,[17] making it an ideal reporter of base pair formation and binding.

In this study, we used 2AP as a reporter of base pair formation and began by looking at pK_a shifting in DNA hairpins (Figure 2-1B). Since the fluorescence of 2AP is known to be quenched by stacking, we placed it adjacent to an ionizing A^+•C wobble. We reasoned that at pH lower than the pK_a, the ionizing base would protonate and form its base pair with C. This interaction would restore native structure to the hairpin, which would lead to stacking and quenching of 2AP. Likewise, at pH higher than the pK_a, the ionizing base would deprotonate so its base pair with C would not form, which would lead to higher fluorescence.

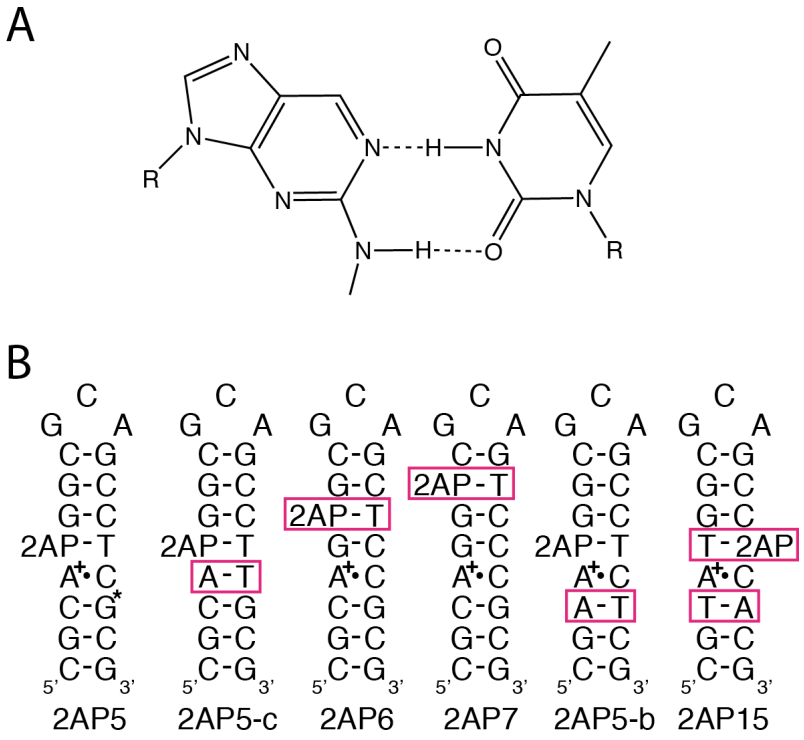


Figure **2-1**: Fluorescent base pair and representative model DNA oligonucleotides. (A) 2AP base pairs with thymine in a two hydrogen bond, non-wobble geometry. (B) DNA constructs used in fluorescence experiments. The naming corresponds to the position of 2AP in the helix. Position of phosphorothioate substitution, used in ^{31}P NMR constructs only, is denoted with an asterisk. '2AP5-c' is a control in which the A^{+}•C base pair is changed to A-T. 2AP6 and 2AP7 move the position of 2AP to the sixth and seventh position in the helix, respectively. In '2AP5-b' the base pair below the A^{+}•C base pair is changed from a C-G to A-T, while in '2AP15', the 2AP-T base pair is flipped. The number after 2AP refers to the position of 2AP in the helix from the 5' end.

2.3 Materials and Methods

2.3.1 Preparation of DNA

DNA oligonucleotides were synthesized, deblocked, and desalted by the manufacturer (Integrated DNA Technologies, Coralville, Iowa). Oligonucleotides underwent a two-step dialysis procedure to removed undesired cationic counterions. They were dialyzed into 100 mM KCl for 6

h to exchange the cation and distilled, deionized water for 16 h to remove excess K^+ using an eight-well microdialysis system (Gibco-BFL Life Technologies, Rockville MD) attached to a peristaltic pump set at a flow rate of ~2mL/min. The DNA sequences were as follows, where '2' represents 2-aminopurine:

```
2AP5:      5'CGC A2G GCG CAG CCT CGC G
2AP6:      5'CGC AG2 GCG CAG CTC CGC G
2AP7:      5'CGC AGG 2CG CAG TCC CGC G
2AP5-b:    5'CGA A2G GCG CAG CCT CTC G
2AP5-c:    5'CGC A2G GCG CAG CCT TGC G
2AP15:     5'CGT ATG GCG CAG CC2 CAC G
```

Mass spectrometry was performed by the manufacturer to confirm synthesis and purity. Purity of DNA oligonucleotides was confirmed via TLC.

For experiments utilizing a fluorescence plate reader, samples were renatured prior to each experiment by heating to 90 °C for 2 min and cooling to room temperature for 15 min in a master mix containing 10 μM DNA and 100 mM KCl. The renatured master-mix was added to individual wells of a 96-or 384- well plate (Grenier- full- and half- area plates with black sides and clear bottoms) with 10 mM bis-tris-propane (VWR) buffer, over a pH range of 5.5 to 9.5.

2.3.2 Preparation of RNA

RNA oligonucleotides were synthesized by manufacturer (Dharmacon, Lafayette, Colorado) and deblocked according to manufacturer's protocol. Samples were dialyzed as previously described. The RNA sequences used were as follows, where '2' represents 2-aminopurine:

```
BWYV:      5'GGC GCG GC2 CCG UCC GCG GAA CAA ACG G
BWYV-dm:   5'GGC GCG GC2 CCA UCC GCG GAA CAA AUG G
```

Lack of degradation of both RNA samples after completion of the entire titration was confirmed by denaturing gel electrophoresis with appropriate size markers and staining with SYBR Gold.

2.3.3 Fluorescence of 2AP DNA and 2AP RNA

Nucleic acid concentrations for fluorescence-detected pH titrations were 10 μM. Titrations were conducted in the absence of buffer in a total volume of 3 mL in a 1 cm pathlength quartz cuvette. Omission of buffer allowed pH to be readily varied and removed a potential fluorescence artifact. We used 10 μM oligonucleotide, even though signal allowed lower concentrations, to allow the DNA/RNA to buffer itself to large pH changes during the titration. DNA titrations were performed in the background of 100 mM KCl, while RNA titrations were in 500 mM KCl to allow for literature comparison.[12]

Oligonucleotides were renatured prior to each experiment by heating to 90 °C for 2 min and cooling on the bench for 15 min. Before renaturation, a small amount of KOH (~1 μL) was added to bring the pH up to the starting point of the experiment. Titrations involved the addition of small volumes of HCl (totaling ~5 μL HCl for the titration) of various concentrations . Upon each addition, the sample was mixed by capping the cuvette with parafilm and inverting, the pH was taken, a spectrum was collected, and the pH was retaken. The pH value used in the pK_a determination was the average of the values measured before and after the spectrum was collected, and were very similar to each other. The pH was monitored with a Hach IQ150 or IQ160 meter, using Hach NMR tube micro probes. Identical pK_a's were obtained on a representative sample with the two different meters. Before the start of each titration, the pH probe was calibrated with colorless pH 7 and 10 standards (BDH General- VWR), which were replaced periodically. (Non-colored pH standards were used to prevent any potential leaching of fluorescent contaminants into the probe and thus the sample.) During the high-pH portion of the titration, the pH was checked periodically with the pH 7 standard, which confirmed absence of

drift. Around pH 7, the probe was recalibrated with colorless pH 4 and 7 standards and the titration continued.

Fluorescence spectra were collected on a Horiba Jobin Yvon Fluoromax-4 and analyzed using FluorEssence and KaleidaGraph software. Samples were excited at 304 nm and emission spectra collected from 325 to 400 nm. To maximize signal-to-noise for different samples, slit widths were adjusted so that the maximum fluorescence intensity was below the 3e+6 limit of the instrument. For wild-type and double mutant BWYV RNA oligonucleotides, however, the same slit widths were used to allow for direct comparison of these two samples. All fluorescence experiments were conducted at room temperature (~22 °C). Fluorescence intensity at 371 nm was plotted versus pH to obtain a titration curve. All DNA and RNA curves were fit to a Henderson-Hasselbalch equation (1), and the pK_a and Hill coefficients were determined.

$$S = S_A + \frac{S_{HA} - S_A}{1 + 10^{n(pH - pK_a)}} \qquad (2\text{-}1)$$

For DNA experiments using a fluorescence plate reader for high-throughput pK_a determination, Synergy4 and SpectraMax2 plate readers were used in the Ades lab and Cameron lab, respectively. Like fluorometer experiments, samples were excited at 304 nm and data were collected from 325 to 400 nm. In attempts to ratiometrically correct for inconsistencies with 2AP fluorescence, Alexa fluor 647 hydrazide (Life Technologies, Grand Island, NY) was used. Bis-tris propane buffer and a master mix of 2AP5 DNA, Alexa fluor 647 hydrazide, KCl were added to each well of a 96-well plate to a final concentration of 10 mM bis-tris-propane, 10 µM 2AP5 DNA, 10 µM Alexa fluor 647 hydrazide, and 100 mM KCl. For 2AP fluorescence, samples were excited at 304 nm and emission spectra were collected from 325 to 400 nm. For Alexa fluor 647 hydrazide fluorescence, samples were excited at 647 nm and emission spectra were collected from 655 to 725 nm. Alexa fluor 647 hydrazide fluorescence intensities at 650 nm were used to correct for changes in fluorescence due to instrumental discrepancies.

2.3.4 NMR of DNA

NMR spectra were collected on a Bruker AV-3-600 spectrophotometer at 25 ^{o}C using previously described parameters.[18] Samples were in 100 mM KCl and 10% D_2O. DNA concentrations ranged from 1.8 to 5.2 mM in a total volume of 500 µL. An internal coaxial tube containing 1% trimethyl phosphate (TMP) in 5% D_2O was used as a reference and set to 0 ppm. The pH was determined as in the fluorescence experiments, except more concentrated acid was needed to adjust the pH owing to the larger concentration of DNA. While the sample was in the NMR, the probe soaked in the pH 7 standard. If the pH read during soaking was off by more than 0.04 pH units, the probe was recalibrated. Also, when the titration reached pH 7, the probe was recalibrated with pH 4 and 7 standards, as described above. The chemical shift of the R_p and S_p phosphorothioate resonances were monitored as a function of pH and fit to equation (1) as in the fluorescence experiments.

2.4 Results and Discussion

2.4.1 Overview

The first four sections of the Results and Discussion describe the development of a simple and efficient fluorescence method for determining pK_a's of nucleobases. This portion includes determining the optimal positioning of the fluorophore 2AP in a DNA helix, control experiment to confirm that the change in fluorescence was due to ionization by using a non-ionizing base pair, and comparison to pK_a's determined by ^{31}P NMR. The fifth and sixth sections include the utilization of the method to determine the pK_a of a protonated base in the pseudoknot of BWYV RNA, and confirmation that the change in fluorescence was due to ionization by using

a double mutant BWYV RNA in which the changed bases result in maintenance of structure but removes the potential for ionization.

2.4.2 Fluorescence-Detected pH Titration of 2AP5

A fluorescence-detected pH titration of the 2AP5 construct (Figure 2-1B), in which the 2AP was positioned adjacent to the protonated base, is provided in Figure 2-2. Fluorescence emission intensity increased between pH 5.5 and 9.5. At low pH, the A^+•C wobble was formed due to protonation of adenine and stacked with 2AP, therefore giving rise to the quenched fluorescence seen in the emission spectra. Similarly, at high pH, adenine was not protonated and the A^+•C base pair was not formed, therefore not stacking with 2AP and not quenching fluorescence. Plotting the fluorescence intensity at the maximum intensity wavelength versus the pH gave a titration curve, which allowed for the determination of a pK_a of 7.58 when fit to the Henderson-Hasselbalch equation. Fitting of the curve also gave a Hill coefficient of 0.87, indicating a two-state system. To confirm that 2AP was the only substantial contribution of fluorescence, the background fluorescence of KCl was subtracted for one titration (Figure A2). Because the background substitution had no effect on the titration of 2AP5, non-background subtracted fluorescence intensities were used for further experiments. Remarkably, upon going from low to high pH, the fluorescence intensity increased almost 5-fold, indicating that 2AP's fluorescence is significantly quenched by formation of the protonated base pair and is therefore an excellent reporter of base pair formation as a result of protonation in DNA.

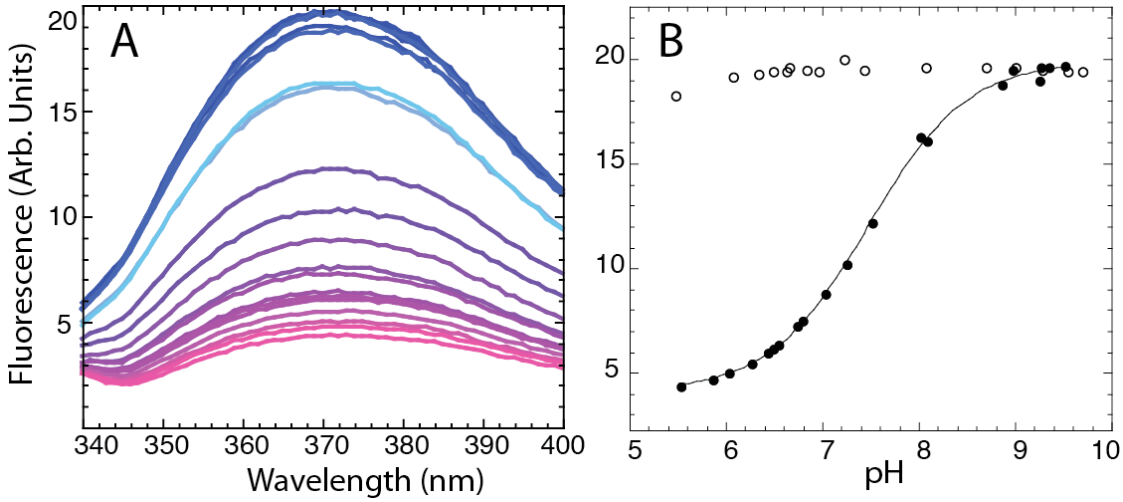


Figure **2-2**: Fluorescence-detected pH titration of 2AP5 DNA. (A) Overlaid emission spectra. Fluorescence increased with increasing pH, going from pink to blue colored curves. (B) pH titration of 2AP5 (•) overlaid with 2AP5-c (^{o}). Fluorescence intensities at 371 nm are plotted versus pH. The pK_a for this trial is 7.47 with a Hill coefficient of 0.85. The control construct 2AP5-c (o) showed no fluorescence change over the same pH range. For this plot, slit widths were adjusted separately for each construct as described in Material and Methods.

2.4.3 Fluorescence-Detected pH Titration of a Control Oligonucleotide

Next, we wished to confirm that the change in fluorescence observed in the pH titration was a result of protonation on the A^+•C base pair and not an artifact from quenching or a fluorescence impurity. A control oligonucleotide (2AP5-c) was derived from the pH-sensitive and highly fluorescent 2AP5 oligonucleotide, in which the A^+•C wobble pair was replaced with an A-T Watson-Crick pair (Figure 2-1B). This base pair substitution removed the potential for the oligonucleotide to ionize, therefore eliminated the potential for change in fluorescence over the pH range studied. As expected, no change in fluorescence was observed over the same pH range of 5.5 to 9.5 (Figures 2-2B and A-1). This confirmed that the change in fluorescence seen with 2AP5 was due to ionization of the A^+•C base pair and not due to quenching of fluorescence or a fluorescent impurity.

2.4.4 Determination of Optimal Placement of 2AP in DNA Helix

To determine the optimal positioning of the 2AP in the helix, we tested DNA oligonucleotides with 2AP in the fifth (2AP5) (Figure 2-1B), sixth (2AP6), and seventh (2AP7) positions from the 5'-end. The hairpin constructs used were similar in structure and sequence to those from a previous study.[18] In that study, high cooperativity in secondary structure formation was also observed when the protonation site was close to the terminus of the helix. Since we wished to avoid such complexity, only positions of 2AP 'above' the A^+•C base pair (i.e. closer to the hairpin loop) were investigated herein.

We needed to find the best positioning for 2AP in the helix in order for it to report the pK_a. of A in the A^+•C wobble. It was important that 2AP was in close proximity to the protonated base pair so that it could report on the pK_a but far enough away that it did not perturb the pK_a on which it was reporting. We moved the 2AP further above the protonating base pair. This proved ineffective, however. Having the 2AP two (2AP6) base pairs from the site of protonation gave a linear sloping response with a 1.3-fold change in fluorescence intensity over the pH range studied (Figure 2-3), which is negligible when compared to the 5-fold fluorescence change seen with 2AP5. Moving the 2AP three (2AP7) base pairs from the site of protonation gave no change in fluorescence (Figure 2-3). The lack of change in fluorescence in these DNA constructs is likely due to stacking of the 2AP in the G-C rich upper stem in both the protonated and unprotonated states. From these results, we conclude that positioning the 2AP immediately above the protonated base pair is optimal for reporting the pK_a in the A^+•C base pair in double-stranded DNA.

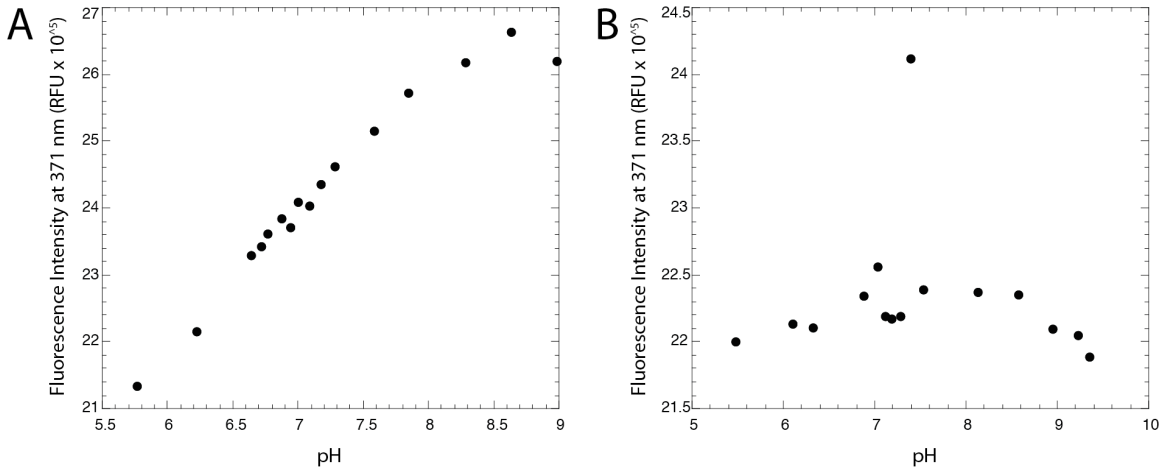


Figure **2-3**: Fluorescence-detected pH titration of 2AP6 and 2AP7 DNA oligonucleotides. Constructs were used to determine optimal positioning of 2AP in the helix. (A) 2AP6 pH titration. (B) 2AP7 pH titration. Either a linear change or no change in fluorescence with respect to pH was observed, making these oligonucleotides unsuitable for reporting the pK_a of the nearby A^+•C base pair. Samples were excited at 304 nm and emission spectra were collected from 325 to 400 nm. Fluorescence intensity at 371 nm was plotted versus the pH.

2.4.5 Validation of Fluorescence Method Using ^{31}P NMR

To test whether addition of 2AP perturbs the pK_a of the A^+•C wobble, we measured the pK_a with and without 2AP using our previously developed ^{31}P NMR method.[19] Three sequences were studied, two having the above 2AP-T orientation but differing in base pairing strength (2AP5 and 2AP5-b), and another in which this base pair was flipped to T-2AP (2AP15) (Table 2-1). For each sequence, two different NMR constructs were studied; each construct had a phosphorothioate label (referred to as 'PS') whose position is provided in Figure 2.1, while one construct also had a 2AP (referred to as 'PS+2AP'). (The phosphorothioate label isolates the ^{31}P resonance downfield, as previously described.[19]) The pK_a's obtained by NMR with and without 2AP are in good agreement (Table 2-1 and Figures A-2 to A-4). In particular, for the 2AP5 sequence introduction of 2AP gave a difference of only 0.31 pK_a units (7.92 and 7.61 for PS and

PS + 2AP, respectively), while in the background of 2AP5-b, a similarly small difference of just 0.21 pK_a units was observed (7.34 and 7.13 for PS and PS + 2AP, respectively). We also tested placing 2AP on the opposite side of the helix, in the T-2AP orientation in 2AP15, which otherwise has the same base pairs as 2AP5-b. Substitution of 2AP into 2AP15 had virtually no effect on the pKa observed by NMR, with pK_a's of 7.41 versus 7.46 for PS and PS + 2AP, respectively. Overall, substitution of 2AP into the helix did not significantly shift the pK_a.

To assess the accuracy of the fluorescence-detected pK_a's, we then compared the pK_a's determined by NMR and fluorescence (PS + 2AP and 2AP labels in Table 2-1). The largest difference between the pK_a's determined by NMR and fluorescence was only 0.26 units, in the case of 2AP15 (7.46 versus 7.20). For 2AP5, the pK_a determined by NMR was just 0.03 units higher than by fluorescence (7.61 versus 7.58), while in 2AP5-b the pK_a by NMR was 0.07 units lower (7.13 versus 7.20). Overall, the pK_a's determined for these three sequences by fluorescence and NMR are in very good agreement, regardless of where the 2AP was substituted in the helix or the strength of base pairing.

When comparing differences in pK_a's amongst constructs, it is of interest to note that changes in nearest neighbor strength affect pK_a shifting. In 2AP5, C-G and 2AP-T base pairs flank the A^+•C and the largest pK_a shift, to 7.58 as detected by fluorescence, is observed. Weakening the C-G base pair to A-T, in the case of 2AP5-b, shifts the pK_a to only 7.20. The lesser pK_a in the presence of AT is likely due to coupling of folding with pK_a change.

Table **2-1**: pK_a's of DNA as detected by NMR and fluorescence

Seq[a]	Label[b]	Method[c]	pK_a[d]	n[d,e]
2AP5	PS	NMR[f]	7.92 ± 0.04	1.03 ± 0.08
	PS+2AP	NMR[f]	7.61 ± 0.03	0.99 ± 0.08
	2AP	Fl[g]	7.58 ± 0.02	0.87 ± 0.03
2AP5-b	PS	NMR[f]	7.34 ± 0.03	0.82 ± 0.05
	PS+2AP	NMR[f]	7.13 ± 0.02	0.98 ± 0.05
	2AP	Fl	7.20 ± 0.03	0.64 ± 0.03
2AP15	PS	NMR	7.41 ± 0.02	0.91 ± 0.03
	PS+2AP	NMR[f]	7.46 ± 0.01	0.95 ± 0.03
	2AP	Fl	7.20 ± 0.03	0.73 ± 0.04

[a]Secondary structures are provided in Figure 1B. [b]'PS' refers to phosphorothioate label and '2AP' refers to 2-aminopurine. [c]NMR and fluorescence (Fl) methods are described in the SI. [d]Curves were fit to the Henderson-Hasselbalch equation (Equation 2-1) to determine pK_a's and Hill coefficients. [e]Hill coefficients were approximately 1, supporting a two-state model. [f]Noted pK_a's and Hill coefficients determined by NMR are an average of the fits of the two diastereomer peaks. [g]Values are the average of three trials. Error is from the fits.

2.4.6 Application of Fluorescence Method to Determine pK_a of BWYV RNA

Next, we tested whether the fluorescence method could be applied to tertiary structure in RNA by determining the pK_a for the BWYV RNA pseudoknot motif. Residue A9 is in close proximity to the protonated base $C8^+$ but makes no hydrogen bond contacts with the rest of the structure and was therefore substituted with 2AP. Titration of this RNA over a pH range of 6.8 to 9.8 led to an increase in fluorescence with pH, similar to that observed with the DNA oligonucleotides (Figure 2-5). Fitting of these data to equation 1 provided a Hill coefficient of 0.94 and a pK_a of 8.15, which agrees well with the previously reported pK_a of 8.4 ± 0.3, calculated using a thermodynamic linkage model under identical solution conditions.[12]

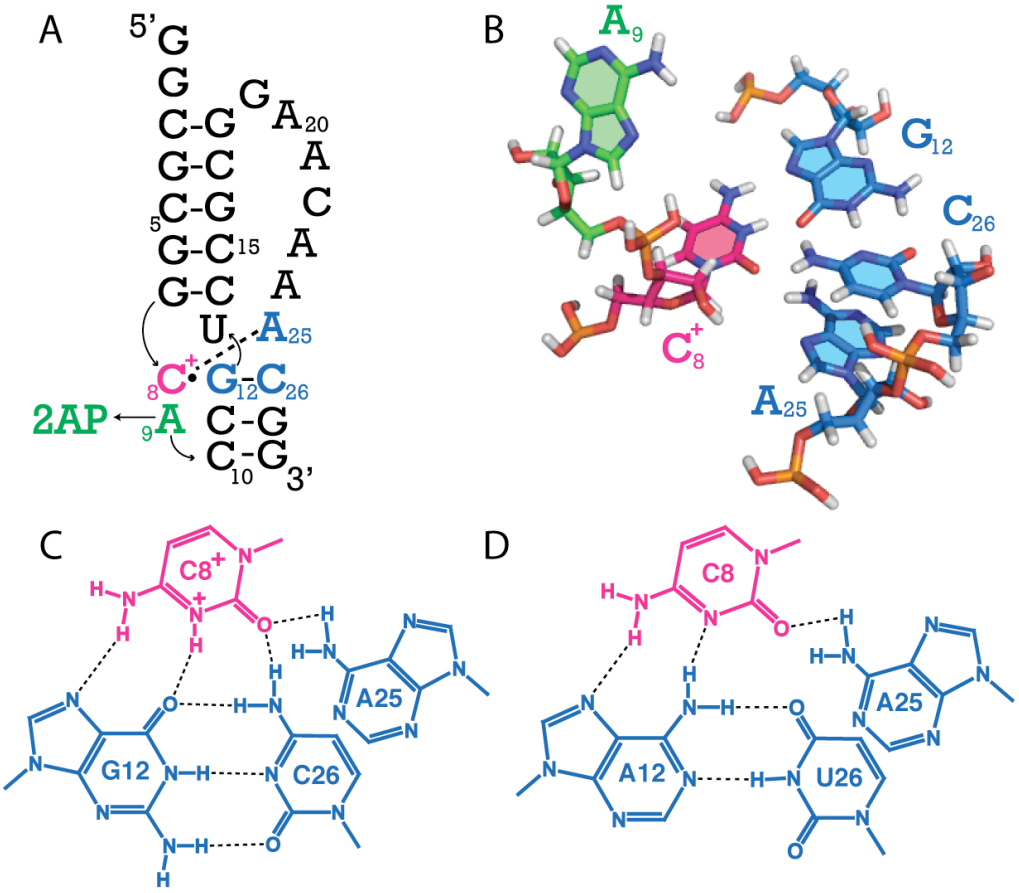


Figure **2-4**: Beet western yellows virus (BWYV) RNA. (A) Secondary structure of BWYV. The base quartet is shown with $C8^+$ (pink)-G12-A25-C26 (blue). A9 (green) was replaced with 2AP. (B) Crystal structure of the bases of interest (PDB 437D8). (C) Interactions in the base quartet with protonated C8. (D) Double mutant with G12A and C26U mutations to maintain structure with neutral C8.

2.4.7 Fluorescence-Detected pH titration of BWYV double mutant

As a control, we determined the pK_a of a double mutant (BWYV-dm) with G12A/C26U mutations (Figure 2-4D). This mutant, which swaps a GC Watson-Crick base pair for an AU, has the same tertiary structure as wild-type BWYV but lacks protonation at C8. A similar double mutant was used to characterize the HDV ribozyme.[20] Moreover, this mutant maintains A25, the fourth base of the quartet, which has shown to be critical for BWYV activity.[21] Lack of change in fluorescence in the BWYV-dm confirms that the pK_a determined in wild-type BWYV is due to protonation at C8 (Figure 2-5B). In addition, the intensity of BWYV-dm emission is nearly equal to that of wild-type at low pH under identical instrumental settings, as expected from the folding depicted in Figures 2-4C and 2-4D.

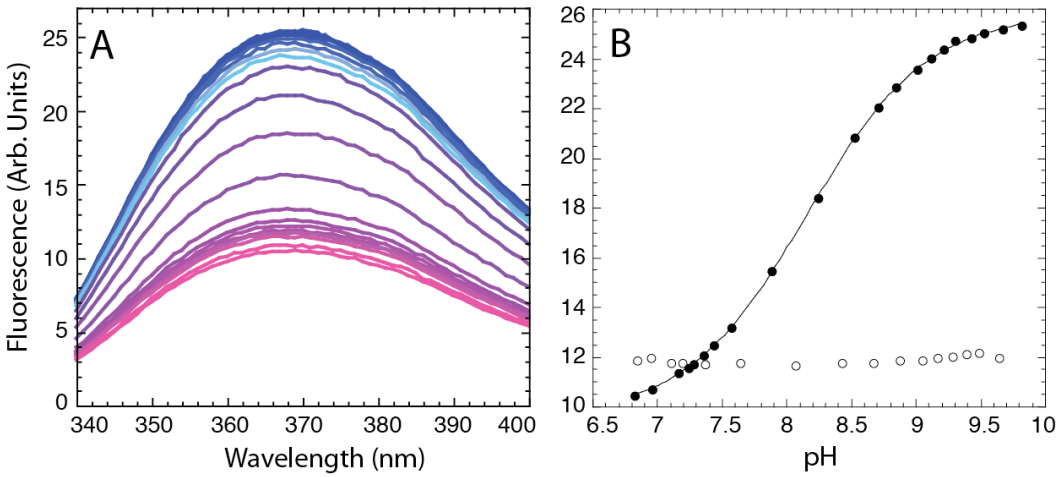


Figure **2-5**: Fluorescence-detected pH titrations of BWYV and BWYV double mutant control RNAs. (A) Overlaid emission spectra for BWYV. Fluorescence increased with increasing pH, going from pink to blue colored curves. (B) pH titration of BWYV RNA (•) overlaid with BWYV-dm RNA (o). Fluorescence intensities at 371 nm were plotted versus pH. The pK_a of $C8^+$ in the BWYV was 8.15 ± 0.01 with a Hill coefficient of 0.94 ± 0.01. The control construct showed no fluorescence change over the same pH range. For this plot, instrumental settings were kept the same for both constructs.

2.4.8 Efforts Towards High-Throughput pK_a Determination Using Fluorescence

Efforts were made towards adaptation of the fluorescence method for high-throughput determination of pK_a's in nucleic acids using a fluorescence plate reader. Using fluorescence to determine pK_a's is a significant improvement over the time and samples requirements for previously developed methods from our lab for pK_a determination using NMR[19] and Raman crystallography[22] and the use of fluorescence lends itself to utilization of a fluorescence plate reader to improve the method even further in a high-throughput nature. In principle, fluorescence plate readers allow for the analysis of a multitude of samples, consisting of different conditions such as salt and temperature or constructs in a short amount of time. Multiple plate readers were tested in an effort to develop the high-throughput aspect of this method, however, no instrument gave reliable, repeatable data and thus, a fluorometer was used for the duration of the method development as discussed earlier in this chapter.

Attempts to use a fluorescence plate reader began in Prof. Craig Cameron's lab using a Synergy4 plate reader. We were able to obtain one titration curve of high quality but with an unusually high Hill coefficient of 2.3. The data were not repeatable, mainly due to problems with alignment of the plate inside the instrument. When inserting the 96- or 384- well plate into the instrument's plate holder, there was a substantial amount of room for the plate to move around. Consistency of the data could not be ensured and fluorescence intensities varied widely (Figure A-7A) and we believe that this was the reason for the irreproducible data obtained. The low inherent fluorescence of 2AP likely contributed to the intensity fluctuation difficulties, as small fluctuations with a stronger fluorophore would probably not have been as detrimental to the use of the instrument.

Then, we looked for other plate readers to try to adapt the fluorescence method for high-throughput pK_a determination. We tried a SpectraMax M2 in Prof. Sarah Ades' lab. This instrument, although reliable, was not sensitive enough to detect the inherently low fluorescence

of 2AP (Figure A-7B). Higher sample concentration and different plates were used in an effort to detect 2AP fluorescence up to 25 μM 2AP DNA, but there was consistently no difference between 2AP-containing samples and water or buffer controls.

Additionally, we hypothesized that successful application of this method may require adding a standard to each sample with a non-overlapping pH-independent fluorophore and conducting experiments in a ratiometric approach to correct for discrepancies seen with the Synergy4 fluorescence plate reader. Alexa fluor 647 hydrazide was added to a mastermix of 2AP5 DNA and KCl, and fluorescence spectra were collected at various pH's. The ratio of 2AP fluorescence to Alexa fluor 647 hydrazide fluorescence was linear (Figure A-8A) and, like the fluorescence of 2AP, the ratio of 2AP to Alexa fluor 647 hydrazide fluorescence was irreproducible (Figure A-8B).

Ultimately, the application of this method to the use of a fluorescence plate reader will require access to a sensitive and properly calibrated and maintained instrument. Because 2AP has a relatively low fluorescence, the instrument needs to be sensitive enough to detect a change in fluorescence of the fluorophore. There are instruments that fit this criterium so the adaptation of this method for high-throughput pK_a determination is very feasible and could be used to rapidly examine effects of temperature, salt and sequence.

2.5 Conclusion

In summary, we have developed a method for pK_a determination using 2AP fluorescence that is simple and effective. Not only does 2AP fluorescence provide an accurate pK_a value, as verified by comparison to NMR the titration, it can be achieved in a fraction of the time, with much lower sample concentration, and with RNAs of any length. In particular, a typical NMR titration takes 14 to 36 hours of labor-intensive experiment time, depending on the strength of the instrument available. Not only is this time consuming, depending upon the conditions under

which the experiment are conducted, it can be detrimental to the sample integrity, especially for RNA. Fluorescence is also much more sensitive than NMR. A standard NMR experiment requires millimolar concentrations of nucleic acid, whereas a fluorescence experiment requires just micromolar amounts. Lastly, NMR is limited to RNAs less than 100 nucleotides, while fluorescence can be used on RNAs of any size if semisynthetic methods are used to introduce the modification.[23]

Addition of 2AP into a DNA hairpin perturbed the pK_a by negligible amounts, ranging from 0.05 to 0.31 pK_a units. In addition, introduction of 2AP into an RNA tertiary structure gave a pK_a that agrees with an earlier measurement from thermodynamic linkage. This lack of perturbation can be contrasted with a recent method for determining pK_a values of RNA through the introduction of 8-azaguanosine or 8-azaadenosine, where the pK_a of G or A is perturbed by 1.1 to 1.3 units.[24, 25] While the 8-azapurines are valuable for monitoring the pK_a of specific residues, our method is versatile in that it can be applied to a variety of nucleic acid systems, under many conditions, without the need for a pK_a correction.

Here, we demonstrated that both A and C protonated bases can be detected using a nearby 2AP; the 2AP itself can be in a double- or single-stranded segment; DNA or RNA can be used; and the pK_a can come from secondary or tertiary structure. In principle, this method should be applicable to determining shifted pK_a's in anionic deprotonated bases, as well. With the utilization of a fluorescence plate reader, the method has the potential to be high-throughput, allowing for analysis of multiple constructs or conditions in a short time.

2.6 Acknowledgments

We thank Scott Showalter, Alan Benesi and Wenbin Lu for assistance with NMR titrations, and Steven Benkovic and Michelle Spiering for assistance with fluorescence. We also thank Craig Cameron and Sarah Ades for use of their fluorescence plate readers in an attempt to

adapt the method to be high-throughput. We also thank John Alumasa for his assistance with the fluorescence plate reader in the Ades lab.

2.7 References

1. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer-Verlag, New York.

2. Izatt, R. M., Christensen, J. J., and Rytting, J. H. (1971) Sites and thermodynamic quantitities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and nucleotides, *Chem. Rev.* **71**, 439-481.

3. Legault, P., and Pardi, A. (1997) Unusual Dynamics and pKa Shift at the Active Site of a Lead-Dependent Ribozyme, *J. Am. Chem. Soc.* **119**, 6621-6628.

4. Wilcox, J. L., Ahluwalia, A. K., and Bevilacqua, P. C. (2011) Charged nucleobases and their potential for RNA catalysis, *Acc Chem Res* **44**, 1270-1279.

5. Chin, K., Sharp, K. A., Honig, B., and Pyle, A. M. (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function, *Nat. Struct. Biol.* **6**, 1055-1061.

6. Tang, C. L., Alexov, E., Pyle, A. M., and Honig, B. (2007) Calculation of pK(a)s in RNA: on the structural origins and functional roles of protonated nucleotides, *J. Mol. Biol.* **366**, 1475-1496.

7. Bevilacqua, P. C. (2008) Proton Transfer in Ribozyme Catalysis, in Ribozymes and RNA Catalysis (Lilley, D. M., and Eckstein, F., Eds.), pp 11-36, Royal Society of Chemistry, Cambridge.

8. Veeraraghavan, N., Ganguly, A., Chen, J. H., Bevilacqua, P. C., Hammes-Schiffer, S., and Golden, B. L. (2011) Metal binding motif in the active site of the HDV ribozyme binds divalent and monovalent ions, *Biochemistry* **50**, 2672-2682.

9. Su, L., Chen, L., Egli, M., Berger, J. M., and Rich, A. (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot, *Nat Struct Biol* **6**, 285-292.

10. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme, *Nature* **395**, 567-574.

11. Nixon, P. L., Rangan, A., Kim, Y. G., Rich, A., Hoffman, D. W., Hennig, M., and Giedroc, D. P. (2002) Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot, *J Mol Biol* **322**, 621-633.

12. Cornish, P. V., and Giedroc, D. P. (2006) Pairwise coupling analysis of helical junction hydrogen bonding interactions in luteoviral RNA pseudoknots, *Biochemistry* **45**, 11162-11171.

13. Ward, D. C., Reich, E., and Stryer, L. (1969) Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives, *J. Biol. Chem.* **244**, 1228-1237.

14. Nordlund, T. M., Andersson, S., Nilsson, L., Rigler, R., Graslund, A., and McLaughlin, L. W. (1989) Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence: fluorescence, molecular dynamics, and NMR studies, *Biochemistry* **28**, 9095-9103.

15. Jean, J. M., and Hall, K. B. (2001) 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking, *Proc. Natl. Acad. Sci. U S A* **98**, 37-41.

16. Hall, K. B. (2009) 2-aminopurine as a probe of RNA conformational transitions, *Methods Enzymol.* **469**, 269-285.

17. Jean, J. M., and Hall, K. B. (2001) 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking, *Proc Natl Acad Sci U S A* **98**, 37-41.

18. Siegfried, N. A., O'Hare, B., and Bevilacqua, P. C. (2010) Driving forces for nucleic acid pK(a) shifting in an A(+).C wobble: effects of helix position, temperature, and ionic strength, *Biochemistry* **49**, 3225-3236.

19. Moody, E. M., Brown, T. S., and Bevilacqua, P. C. (2004) Simple method for determining nucleobase pK(a) values by indirect labeling and demonstration of a pK(a) of neutrality in dsDNA, *J. Am. Chem. Soc.* **126**, 10200-10201.

20. Nakano, S., and Bevilacqua, P. C. (2007) Mechanistic characterization of the HDV genomic ribozyme: a mutant of the C41 motif provides insight into the positioning and thermodynamic linkage of metal ions and protons, *Biochemistry* **46**, 3001-3012.

21. Kim, Y. G., Su, L., Maas, S., O'Neill, A., and Rich, A. (1999) Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency, *Proc Natl Acad Sci U S A* **96**, 14234-14239.

22. Gong, B., Chen, J. H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. (2007) Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography, *J. Am. Chem. Soc.* **129**, 13335-13342.

23. Moore, M. J., and Sharp, P. A. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites, *Science* **256**, 992-997.

24. Liu, L., Cottrell, J. W., Scott, L. G., and Fedor, M. J. (2009) Direct measurement of the ionization state of an essential guanine in the hairpin ribozyme, *Nat Chem Biol* **5**, 351-357.

25. Cottrell, J. W., Scott, L. G., and Fedor, M. J. (2011) The pH dependence of hairpin ribozyme catalysis reflects ionization of an active site adenine, *J Biol Chem* **286**, 17658-17664.

Appendix A

Chapter 2 Supplemental Information

[Selected data were published in the Supporting Information for a paper entitled "A Simple Fluorescence Method for pK_a Determination in RNA and DNA Reveals Highly Shifted pK_a's" by Jennifer L. Wilcox and Philip C. Bevilacqua in *Journal of the American Chemical Society* **2013**, *135* 7390-7393.]

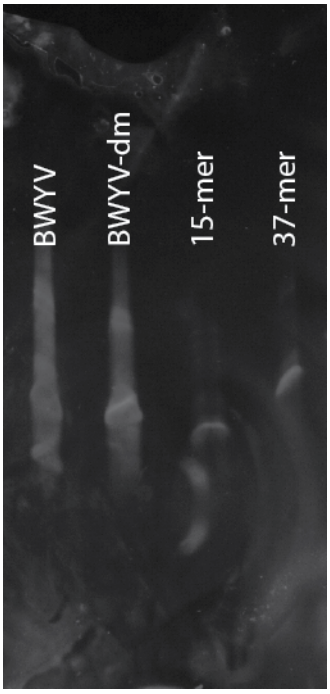


Figure **A-1**: Denaturing polyacrylamide gel electrophoresis with SYBR Gold staining was used to confirm the purity of BWYV and BWYV-dm RNA oligonucleotides. The 28 nucleotide RNAs were compared to standards of 15 and 37 nucleotides. Residual bands above the most prominent RNA bands were due to insufficient denaturation during gel electrophoresis. No degradation was observed.

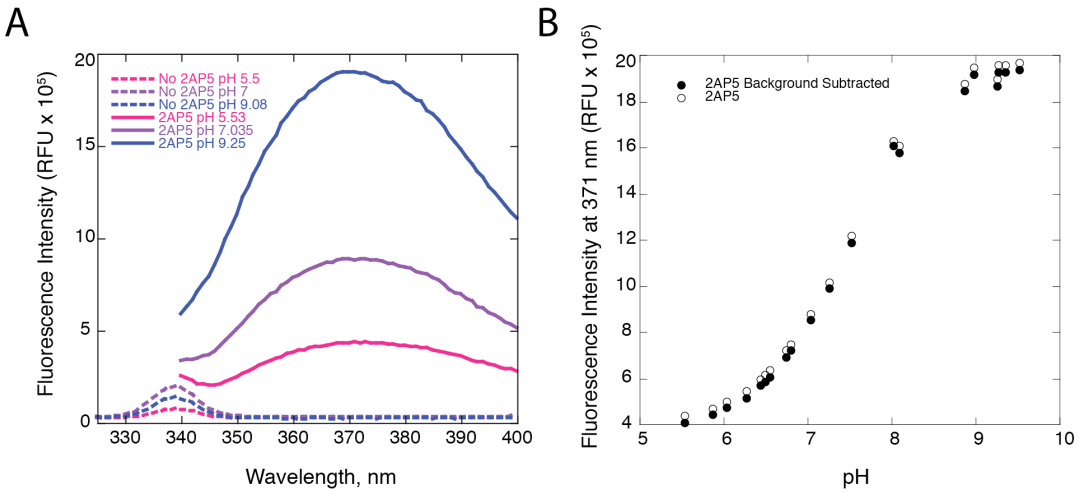


Figure **A-2**: Determination of fluorescence contribution from KCl background. (A) Samples were excited at 304 nm and emission spectra were collected from either 325 or 340 to 400 nm. Fluorescence from the background of 100 mM KCl (dashed lines) was determined to confirm that it did not contribute to the fluorescence of DNA samples (lines). (B) Background fluorescence emission of KCl at 371 nm was subtracted from the DNA samples at 371 nm (closed circles) and showed little to no difference from non-background subtracted DNA samples (open circles).

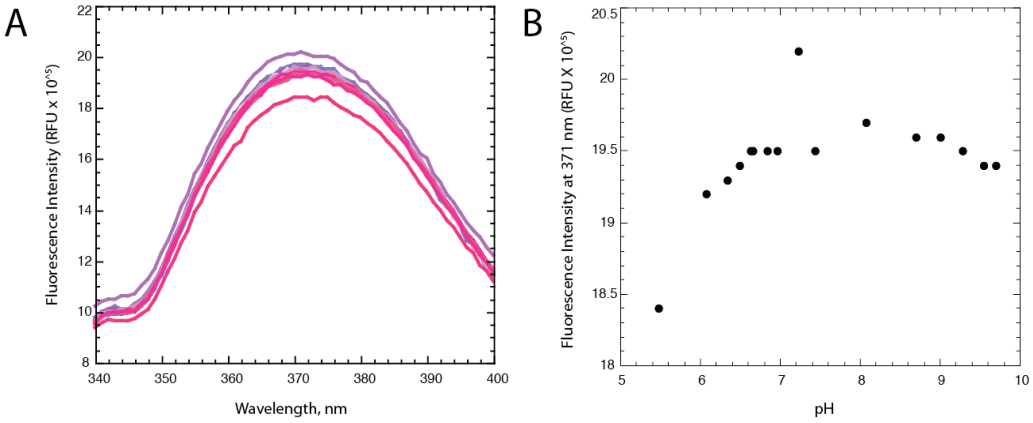


Figure **A-3**: Fluorescence-detected pH titration of 2AP5-c, the control oligonucleotide with an A-T Watson-Crick base pair replacing the A^{+}•C wobble pair found in 2AP5. (A) Overlaid emission spectra. Sample was excited at 304 nm and emission spectra were collected from 325 to 340 nm and overlaid. Color range is the same as used in Figure 2-2A. (B) Fluorescence-detected pH titration. Fluorescence intensity at 371 nm was plotted versus pH. No appreciable fluorescence change was found, thereby confirming that the fluorescence change seen in 2AP5 was due to protonation of the A^{+}•C base pair. These data are also provided as an overlay with 2AP5 in Figure 2-2B in the main text.

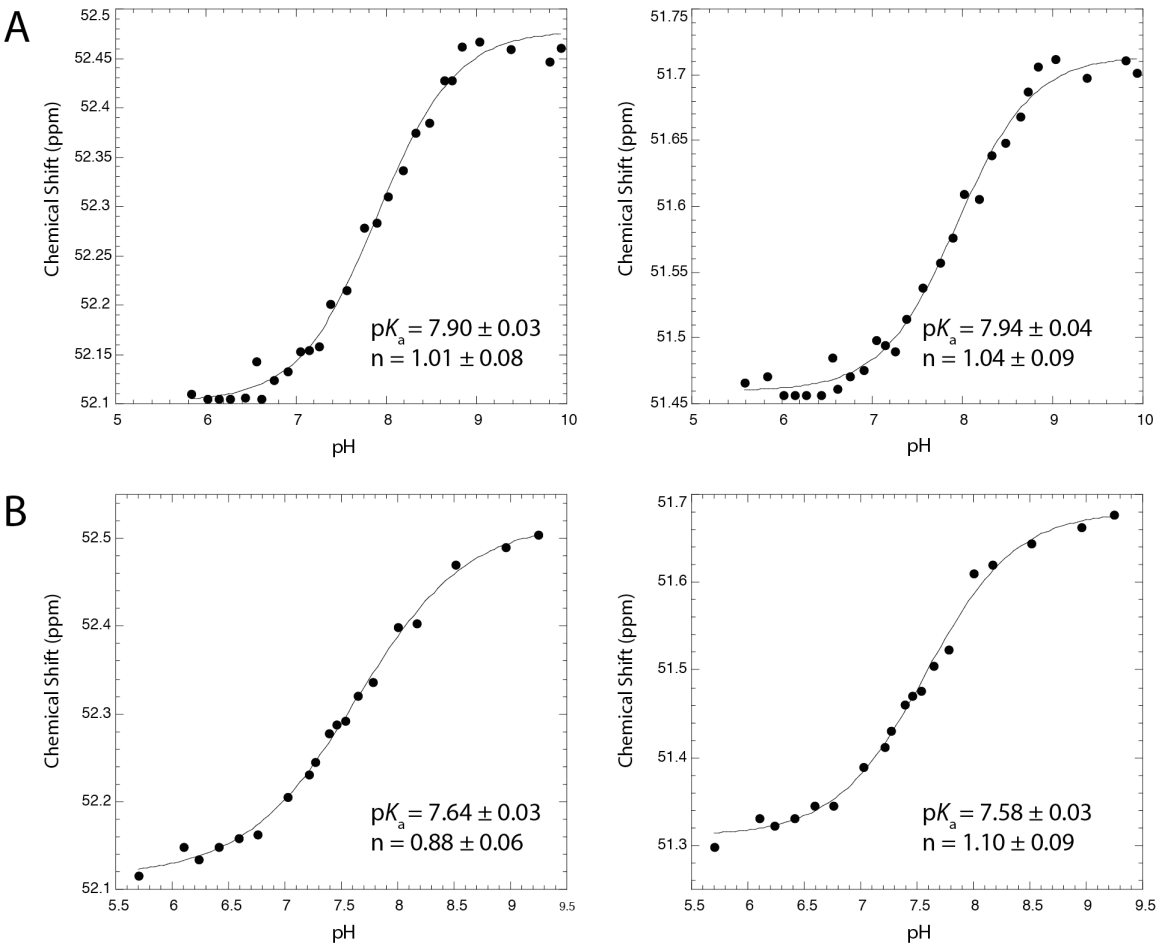


Figure **A-4**: ^{31}P NMR-detected pH titration of 2AP5. (A) pH titration of 2AP5 with A at position 5. This construct was used to test the change in pK_a caused by addition of 2AP. The two plots reflect the two phosphorothioate diastereomers. (B) pH titration of 2AP5 with 2AP at position 5. The two plots reflect the two phosphorothioate diastereomers. For both panels A and B, pK_a's and Hill coefficients for the two diastereomer peaks were averaged to obtain the values reported in Table 1.

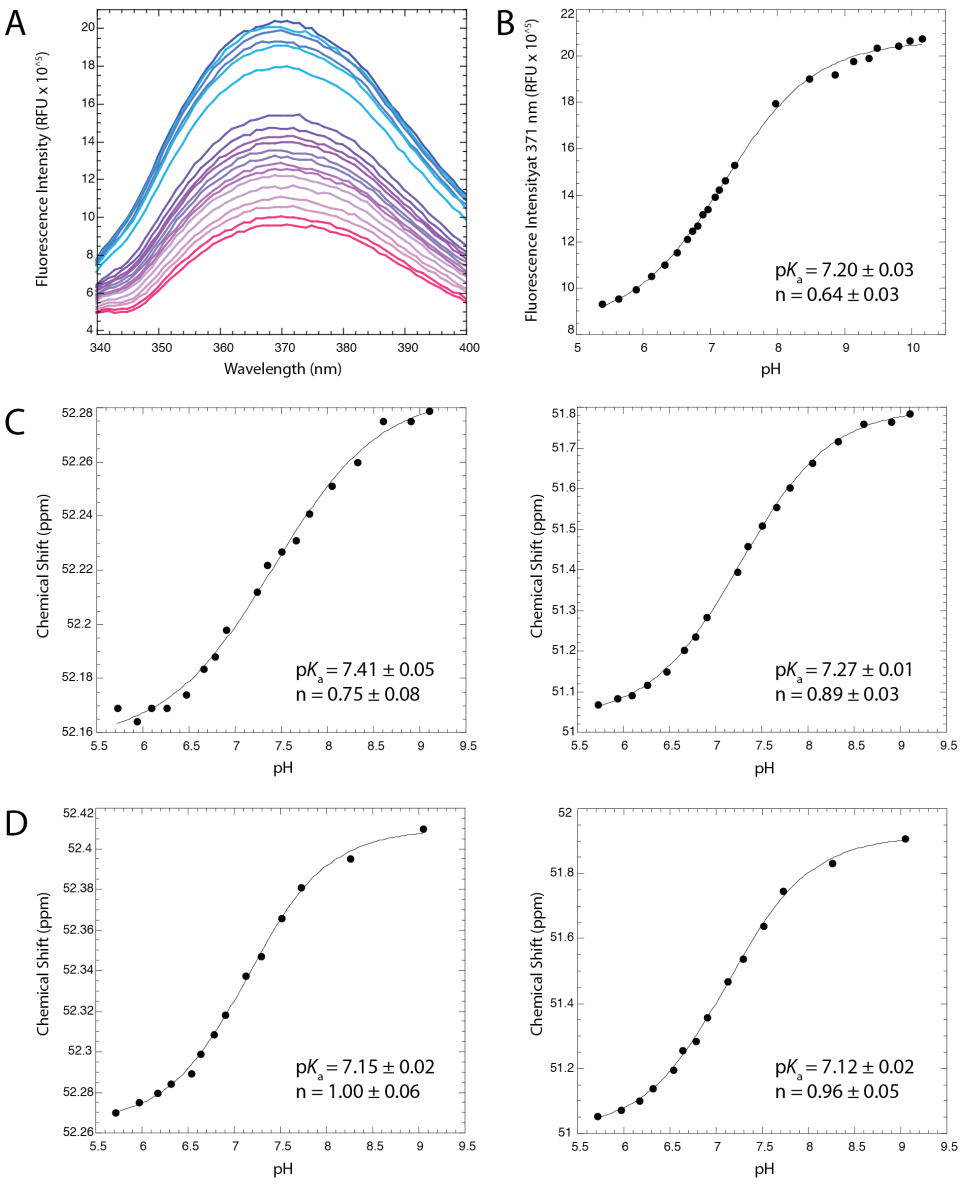


Figure **A-5**: pH titration of 2AP5-b, a DNA oligonucleotide used to test change in pK_a caused by sequence changes. (A) Overlaid emission spectra. The sample was excited at 304 nm and emission spectra were collected from 325 to 400 nm and overlaid. Color range is the same as used in Figure 2A. (B) Fluorescence-detected pH titration. (C) ^{31}P NMR-detected pH titration of 2AP5-b with A at position 5. This construct was used to test the change in pK_a caused by addition of 2AP. The two plots reflect the two phosphorothioate diastereomers. (D) ^{31}P NMR-detected pH titration of 2AP5-b with 2AP at position 5. The two plots reflect the two phosphorothioate diastereomers. For both panels C and D, pK_a's and Hill coefficients for the two diastereomer peaks were averaged to obtain the value reported in Table 1.

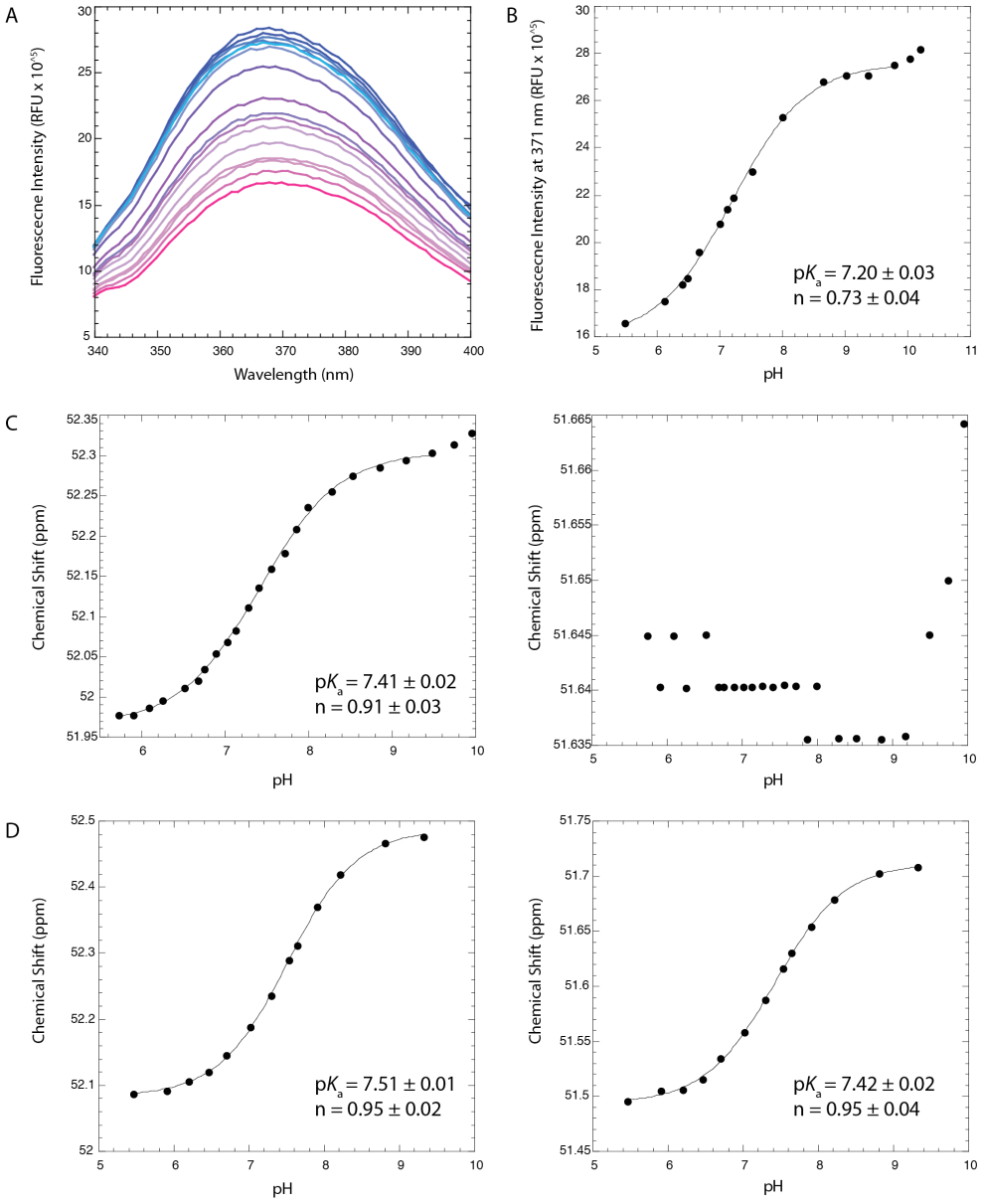


Figure **A-6**: pH titration of 2AP15, a DNA oligonucleotide used to observe changes in placement of 2AP in relation to the protonated base and its effect on pK_a reporting. (A) Overlaid emission spectra. The sample was excited at 304 nm and emission spectra were collected from 325 to 400 nm and overlaid. (B) Fluorescence-detected pH titration. Very high pH data deviated from the curve, likely because of alkaline denaturation, and were excluded from the fit. (C) ^{31}P NMR-detected pH titration of 2AP15 with A at position 15. This construct was used to test the change in pK_a caused by the addition of 2AP. No data could be obtained from one diastereomer peak (right panel). (D) ^{31}P NMR-detected pH titration of 2AP15 with 2AP at position 15. The two panels reflect the two phosphorothioate diastereomers. pK_a's and Hill coefficients for the two diastereomer peaks were averaged to obtain the value reported in Table 1.

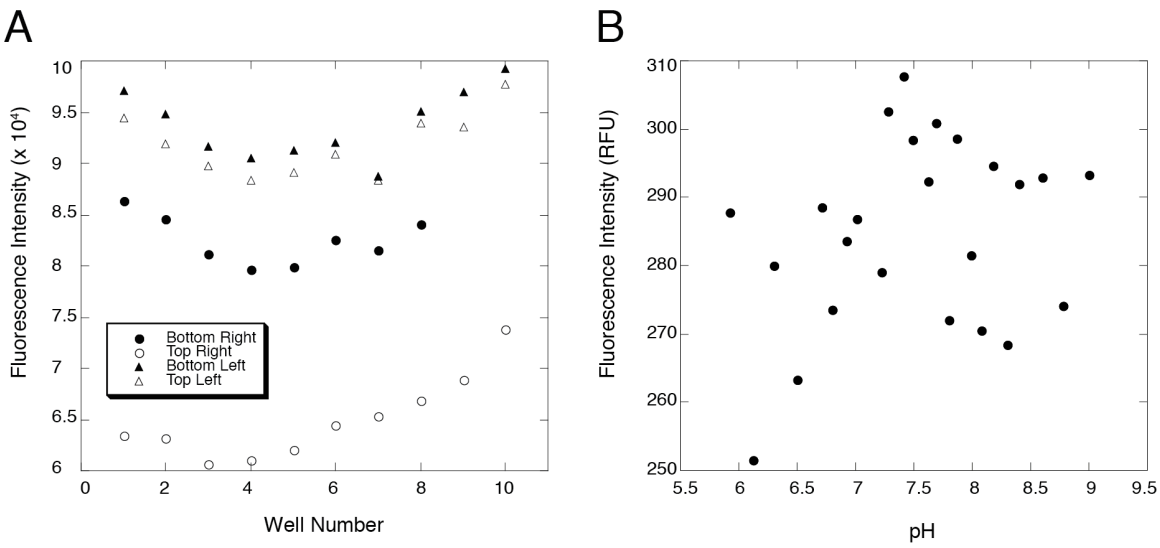


Figure **A-7**: Inconsistencies with 2AP5 fluorescence detected via fluorescence plate reader (A) Alignment problems with the Synergy4 fluorescence plate reader led to inconsistent data. Each well contained 10 μM 2AP5 DNA, 10 mM bis-tris-propane buffer, and 100 mM KCl and was thus expected to give identical results. However, this did not occur, as shown (left to right). The 96-well plate was positioned towards the bottom right, top right, bottom left and top left corners of the plate carrier and gave different fluorescence intensities depending on positioning (top to bottom). (B) Insufficient sensitivity of SpectraMax 2 fluorescence plate readers led to a substantial and problematic variation in sample intensities. Titration of 2AP5 was performed in the background of 100 mM KCl and 10 mM bis-tris-propane buffer in a 384-well plate.

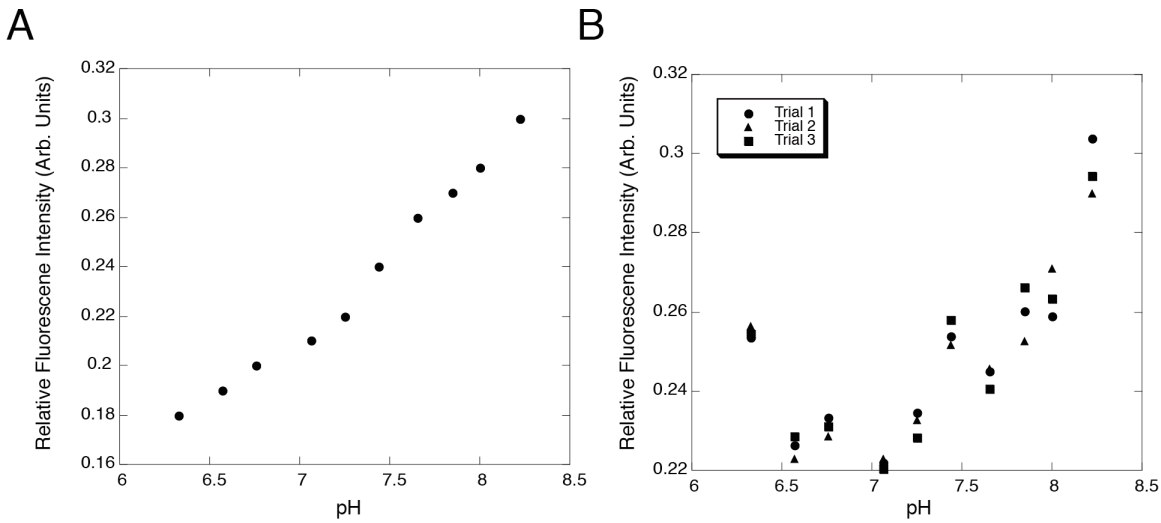


Figure **A-8**: Attempts to ratiometrically correct 2AP fluorescence discrepancies with Alexafluor 647 hydrazide fluorescence. Samples contained 100 mM KCl, 10 mM bis-tris-propane buffer, 10 µM 2AP5 DNA and 10 µM Alexafluor 647 hydrazide. 2AP was excited at 304 nm and spectra were collected from 325 to 400 nm. Alexafluor 647 hydrazide was excited at 647 nm and emission spectra were collected from 655 to 725 nm. Alexa fluor 647 hydrazide fluorescence intensities at 670 nm were used to correct for changes in 2AP fluorescence at 371 nm due to instrumental discrepancies. (A) Ratio of 2AP fluorescence to Alexa fluor 647 hydrazide fluorescence in 2AP5 DNA. (B) Ratio of 2AP fluorescence to Alexa fluor 647 hydrazide fluorescence in 2AP5 DNA in triplicate. Precision of fluorescence intensity decreased as pH increased.

Chapter 3

pK_a Shifting in Double-Stranded RNA is Highly Dependent Upon Nearest Neighbors and Bulge Positioning

[Portions of this chapter were submitted as a paper entitled "pK_a Shifting in Double-Stranded RNA is Highly Dependent Upon Nearest Neighbors and Bulge Positioning" by Jennifer L. Wilcox and Philip C. Bevilacqua to *Biochemistry* **2013**- Additional experiments were performed through revision of the manuscript and can be found in the submitted manuscript.]

3.1 Abstract

Shifting of pK_a's in RNA is important for biological processes, however the driving forces for shifting are not well understood. Herein, we seek to determine how structural environments surrounding protonated bases affect pK_a shifting in dsRNA. Using ^{31}P NMR, we determined the pK_a of adenine in an A^+•C base pair in various sequence and structural environments and found a significant dependence of pK_a shifting on the base pairing strength of nearest neighbors and location of a nearby bulge. These trends are used to estimate whether an A•C wobble is protonated in dsRNA processes, including A to I editing by ADAR.

3.2 Introduction

Single stranded DNA and RNA nucleobases are typically uncharged, with pK_a's of 3.5 and 4.2 for A and C, respectively, and 9.2 for G and T/U.[1] Watson-Crick base pairing shifts pK_a's even further from physiological pH since protons are engaged in hydrogen bonding.[2]

In some non-canonical base pairs, however, pK_a's are shifted towards neutrality,[3] thereby allowing these nucleobases to play important roles in biological processes, such as general acid-

base catalysis,[3] miRNA processing,[4-6] and programmed ribosomal frameshifting.[7] Although significant research has been done to determine how protonated bases participate in these processes, many protonated bases surely remain unidentified.

In this study, we investigate the effect the structural environment surrounding a protonated base pair has on pK_a shifting in RNA. Using ^{31}P NMR,[8] we analyze how changing the strength of nearest neighbor base pairing and the addition of a bulge of various proximities to an A^+•C wobble affects its pK_a.

3.3 Materials and Methods

3.3.1 Preparation of RNA Samples Used in Fluorescence Experiments

RNA oligonucleotides were synthesized by the manufacturer (Dharmacon Lafayette, Colorado) and deblocked according to manufacturer's protocol. Oligonucleotides were dialyzed as previously described.[9] RNA sequences were as follows, where '2' denotes 2-aminopurine substitutions:

AA^+2AP: 5′ CGA A2G ACU UCG GUC UCU CG

GA^+U: 5′ CGG AUG ACU UCG GUC 2CC CG

UA^+2AP: 5′ CGU A2G ACU UCG GUC UCA CG

GA^+2AP: 5′ CGG A2G ACU UCG GUC UCC CG

Nucleic acid concentrations for fluorescence-detected pH titrations were 10 μM. Titrations were conducted in the absence of buffer in a total volume of 3 mL in a 1 cm pathlength quartz cuvette. Omission of buffer allowed pH to be readily varied and removed a potential fluorescence artifact. We used 10 μM oligonucleotide, even though signal allowed lower concentrations, to allow the RNA to buffer itself to large pH changes during the titration.

RNA oligonucleotides were renatured prior to each experiment by heating to 90 °C for 2 min and cooling on the bench for 15 min. After renaturation, a small amount of KOH (~1 μL) was added to bring the pH up to the starting point of the experiment. Titrations involved the addition of small volumes of HCl (totaling ~5 μL HCl for the titration) of various concentrations. Upon each addition, the sample was mixed by capping the cuvette with parafilm and inverting, the pH was taken, a spectrum was collected, and the pH was retaken. The pH value used in the pK_a determination was the average of the values measured before and after the spectrum was collected, and were very similar to each other. The pH was monitored with a Hach IQ160 meter using an NMR tube micro probe. Before the start of each titration, the pH probe was calibrated with colorless pH 7 and 10 standards (BDH General- VWR), which were replaced periodically. (Non-colored pH standards were used to prevent any potential leaching of fluorescent contaminants into the probe and thus the sample.) During the high-pH portion of the titration, the pH was checked periodically with the pH 7 standard, which confirmed absence of drift. Around pH 7, the probe was recalibrated with colorless pH 4 and 7 standards and the titration continued.

3.3.2 Fluorescence-detected pH-titration of RNA

Fluorescence spectra were collected on a Horiba Jobin Yvon Fluoromax-4 and analyzed using FluorEssence and KaleidaGraph software. Samples were excited at 304 nm and emission spectra were collected from 325 to 400 nm. To maximize signal-to-noise for different samples, slit widths were adjusted so that the maximum fluorescence intensity was below the 3e+6 limit of the instrument. All fluorescence experiments were conducted at 25 °C or 37 °C. Fluorescence intensity at 371 nm was plotted versus pH to obtain a titration curve. All RNA curves were fit to a Henderson-Hasselbalch equation (3-1), and the pK_a and Hill coefficients were determined.

$$S = S_A + \frac{S_{HA} - S_A}{1 + 10^{n(pH - pK_a)}} \tag{3-1}$$

3.3.3 Preparation of RNA Samples Used in NMR Experiments

RNA oligonucleotides were synthesized by the manufacturer (Dharmacon Lafayette, Colorado) and deblocked according to manufacturer's protocol. Oligonucleotides were dialyzed as previously described.[9] RNA sequences were as follows:

WW:	5′ CGU AAG ACU UCG GUC UCA CG
SW:	5′ CGG AAG ACU UCG GUC UCC CG
SS:	5′ CGG ACG ACU UCG GUC GCC CG
Bulge-0 (SS′):	5′ CGG AGG ACU UCG GUC CCC CG
Bulge-1:	5′ CGG AGA CGC UUC GGC GCC CCG
Bulge-2:	5′ CGG AGC AGC UUC GGC GCC CCG

After dialysis, RNA was brought up in 500 µL. The sample contained between 0.88 and 1.26 mM RNA in 100 mM KCl and 10% D_2O. Titrations were conducted in the absence of buffer so as to allow for small pH changes upon addition of acid or base. Slight buffering of the system was provided by the high concentrations of RNA. The sample was renatured by heating at 90 °C for 2 min, then cooled on the bench top for 15 min. The pH was measured using a Hach IQ 160 meter with an NMR tube micro-probe. The probe was calibrated before pH determination with standards at pH 7 and 10 (BDH General- VWR), which were replaced periodically. After renaturation of the RNA, the pH was adjusted to the desired alkalinity with small volumes of KOH to begin the titration in the range of pH 9-10.

3.3.4 ^{31}P NMR-detected pH-titration of RNA

NMR spectra were collected on a Bruker AV-3-600 MHz NMR spectrophotometer using previously described parameters[10] and analyzed using TopSpin software. An internal coaxial tube containing 1% trimethyl phosphate (TMP) in 5% D_2O was used as a reference and set to 0 ppm for each spectrum. For each pH data point, the pH was determined, the internal coaxial tube was re-inserted and an NMR spectrum was collected at 25 °C (unless otherwise noted). Subsequently,

the coaxial tube was removed and the pH was re-determined. The pH value used in analysis was the average of the two recorded values. Small volumes of concentrated HCl were used to increase the acidity of the solution throughout the titration. While NMR spectra were being collected, the pH probe remained submerged in pH 7 standard. If the observed pH varied by more than 0.05 pH units from the standard value, the probe was recalibrated. Once the pH values being reported reached approximately 7, the probe was recalibrated for pH 4 and 7. For each pH value, the chemical shifts of the R_p and S_p phosphorothioate resonances were monitored as a function of pH and fit to the Henderson- Hasselbalch equation (3-1) using KaleidaGraph software, and the pK_a and Hill coefficient were determined.

3.3.5 ^{1}H NMR of RNA

To identify base pairs in bulged constructs, one dimensional (1D) imino proton (^{1}H) spectra were collected on a Bruker AV-3-600 MHz NMR spectrophotometer with a CTI cryoprobe. The temperature was 276 K to reduce proton exchange and promote optimal sharpness of peaks. Water suppression pulse programs were employed, as described previously.[10] Samples for ^{1}H NMR were ~0.25-0.5 mM RNA, 100 mM KCl and 10% D_2O. Data were analyzed with TopSpin software. D_2O reference peak was set to 5.0 ppm, an approximate value for 3 °C.[11]

3.3.6 UV monitored thermal denaturation

RNA (~1.5 μM) was melted in the background of 100 mm KCl and 10 mM buffer (bis-tris-propane pH 5.03, pH 8.79, pH 9.22 or MES pH 6.13; pH determined at room temperature) in 1 cm pathlength quartz cuvettes. Buffer pH's were chosen such that the pH was at least 1 unit removed from the pK_a of the construct to favor the protonated or deprotonated states as desired. Samples were renatured by heating to 90 °C for 2 min then cooled to room temperature for 15

min. UV monitored thermal denaturation experiments (melts) were conducted at 260 nm on a Gilford Response II spectrophotometer equipped with a temperature controlled microcuvette assembly. Absorbances were collected from 5 to 95 °C in 0.5 °C increments. Data were analyzed using KaleidaGraph software. Five-point smoothing was done prior to taking first derivatives. $\Delta G°$ values were calculated at 37, 50 and 70 °C to facilitate comparisons at various temperatures, and values at 50 °C are provided in Table 3-2. Errors in coupling free energy are typically 0.1 kcal/mol.[12]

3.3.7 Analysis of pri-miRNA nearest neighbors from miRBase

Human pri-miRNA's were analyzed using sequences in miRBase Release 19. Secondary structures were copied to a text file and inputted into an RNA Pair Analyzer program developed by Seth M. Morton, which outputted the number of A^+•C base pairs with their occurrence of nearest neighbors.[13]

3.4 Results and Discussion

Efforts in this chapter are further categorized based on their approach of fluorescence spectroscopy, NMR spectroscopy or bioinformatics. Initial efforts were made using 2-aminopurine (2AP) fluorescence to quantify pK_a shifting in RNA; however, the change in fluorescence observed upon protonation in model oligonucleotides was insufficient in determining the pK_a of an adenine in an A^+•C wobble. Further efforts were made using ^{31}P NMR to determine the sequence dependence and effect of a nearby bulge on pK_a shifting in RNA. We found that pK_a shifting was highly dependent upon nearest neighbor base pairing strength and bulge proximity. To investigate the potential role of protonated nucleobases in miRNA

processing, a bioinformatics approach was used to analyze A^+•C wobbles in human miRNAs. No preference for nearest neighbor strength was observed with human miRNAs in the miRBase.

3.4.1 2-Aminopurine-Detected pK_a Determination

Initial experiments described in this chapter utilized the fluorescence method previously developed and described in Chapter 2.[9] Secondary structures of RNA model oligonucleotides for fluorescence and NMR are presented in Figure 3-1. All RNA hairpins contain an A^+•C wobble in the stem where pK_a was being determined. For fluorescence-detected pH-titrations experiments, a 2AP-U base pair was positioned above a protonated base pair (Figure 3-1 top row) and the quenching of fluorescence upon protonation of adenine in the A^+•C wobble and base pair formation was used to determine the pK_a. However, for all RNA constructs tried, a very small change in fluorescence quenching upon protonation was seen, which made analysis of fluorescence-detected pH titrations impossible. Some initial dependence of pK_a shifting on nearest neighbor strength, temperature, and salt dependence can be inferred using fluorescence but, since the curves were not able to be fit with the highest degree of certainty, the pK_a's listed are just approximations and should be treated as such. Altering the temperature, salt and Mg^{2+} concentrations increased the pK_a. However, these environmental effects were not substantial in shifting the pK_a high enough to obtain a lower baseline and a full titration curve.

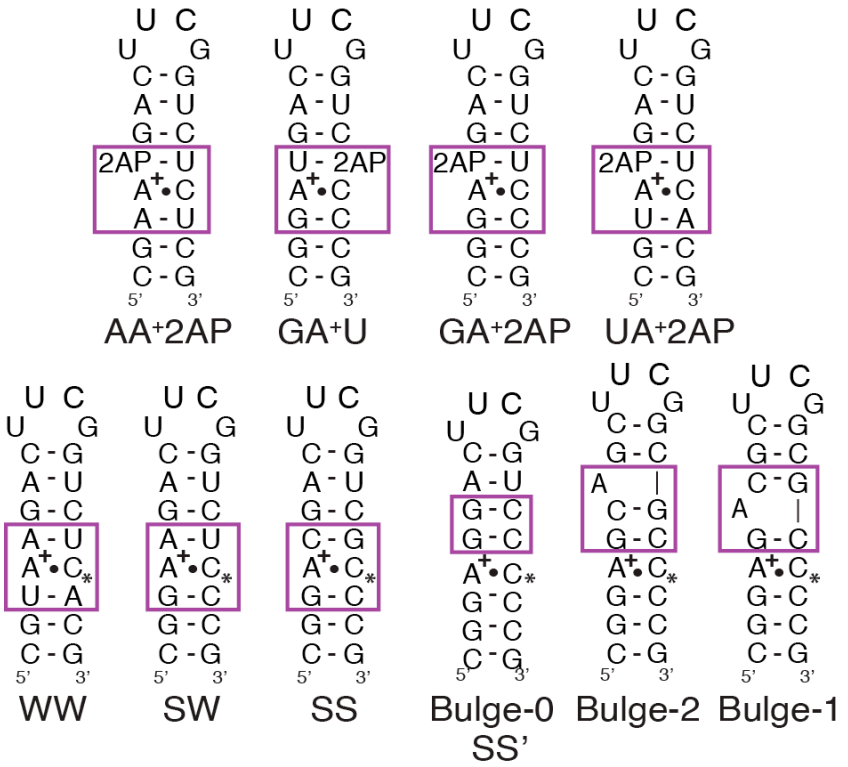


Figure **3-1**: Protonated RNA oligonucleotides used in 2-aminopurine fluorescence-detected (top) and ^{31}P-detected (bottom) pH-titrations. Initial experiments to test the dependence of pK_a on nearest neighbors were performed using the 2-aminopurine substituted RNA constructs shown (top row). Names denote the sequence of the 5' nearest neighbors of the A$^+$•C hairpin. To test the dependence of pK_a on nearest neighbors using ^{31}P NMR (bottom row), we used constructs WW (with two weak A-U base pairs), SW (with one strong G-C and one weak A-U base pair) and SS (with two strong G-C base pairs) neighboring the A$^+$•C wobble. To assess effects of a structural defect on the pK_a, we analyzed constructs with a bulge two (Bulge-2) or one base pair(s) away (Bulge-1) from the A$^+$•C wobble, along with an analogous RNA with no bulge (Bulge-0, also referred to as "SS' "). The asterisks denote position of the phosphorothioate substitution necessary for detection by ^{31}P NMR.

Fluorescence detected pH titrations began with AA$^+$2AP, an RNA with A-U and 2AP-U nearest neighbors (Figure 3-1). Determining a pK_a for adenine in the A$^+$•C wobble by fluorescence had significant difficulties. Initial titrations performed at 37 °C and in 140 mM KCl and 0.5 mM Mg^{2+} had a noisy upper baseline (Figure 3-2A). Extending the titration to pH 5.5 was incapable of producing a lower baseline, due to the apparently low pK_a of the RNA, although a pK_a of a similar construct with two A-U nearest neighbors and the 5' sequence UA$^+$A was

determined later in our study at 25 °C and in 100 mM KCl and 0 mM Mg^{2+} to be 6.51 by ^{31}P NMR. In further efforts, the temperature was reduced in case of the possibility of a low melting temperature for the oligonucleotide that could potentially cause structure deformation at 37 °C and as a means of increasing the pK_a to obtain a full titration curve with a lower baseline. Repeating the titration at 25 °C but maintaining the salt conditions (Figure 3-2B) somewhat decreased the noisy upper baseline observed in experiments at 37 °C, but continuing the titration down to pH 5 was still incapable of producing a lower baseline. Moreover lowering temperature from 37 to 25 °C decreased the pK_a, which came as a surprise. Maintaining the lower temperature of 25 °C and 100 mM KCl but removing Mg^{2+} (Figure 3-2C) resulted in the greatest reduction of noise in the upper baseline of the titration curve and the lowest error when fit to the Henderson-Hasselbalch equation. However, difficulties still remained in accurately determining a pK_a due to the lack of a lower baseline. Finally, omitting the Mg^{2+}, reducing the temperature to 25 °C, but increasing the KCl concentration to 500 mM (Figure 3-2D) shifted the pK_a lower and resulted in less of a lower baseline and greater error when fit to the Henderson- Hasselbalch equations. Lowering the pK_a upon increasing salt concentration is consistent with results of Siegfried *et al*,[10] as increasing the concentration of cations in solution disfavors the cationic adenine from forming, therefore inhibiting pK_a shifting.

For this AA^+2AP oligonucleotide, there were particular difficulties in both noise in the upper baseline and continuing the titration to low enough pH to obtain a fluorescence determination of the protonated state. Lowering the pH below 5 would have resulted in acid denaturation and inaccurate pK_a values.[14] Attempting to further shift the pK_a by reducing the temperature and adjusting the salt and Mg^{2+} concentrations were inefficient at shifting the pK_a high enough to allow for an accurate pK_a determination. An overlay of these titrations reveals that pK_a shifting in the AA^+2AP oligonucleotide is not significantly dependent upon temperature and Mg^{2+}, as only very small changes in fluorescence titrations curves were observed after decreasing

temperature and Mg^{2+} (Figure B-1). A 5-fold increase in K^+ had a moderate effect on lowering pK_a shifting in AA^+2AP, and was comparable to the effect of increasing K^+ in DNA.[10] As will be discussed further in this chapter, changing the temperature and the Mg^{2+} in a ^{31}P-detected pH-titration showed essentially no effect on an RNA with two strong G-C base pairs, indicating a difference in the means by which environmental conditions affect pK_a shifting based on nearest neighbor base pairing strength.

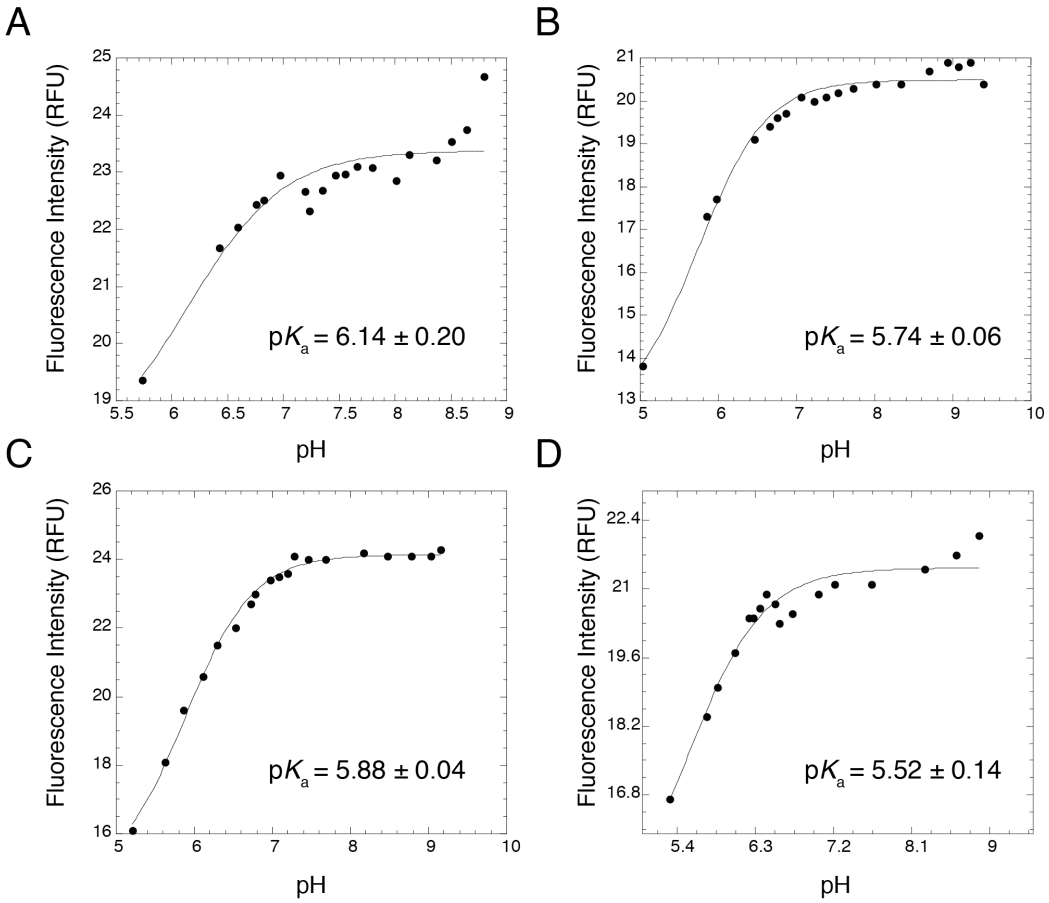


Figure **3-2**: 2-aminopurine fluorescence-detected pH-titrations of an AA^+2AP RNA oligonucleotide. Various reaction conditions were used in an attempt to increase the pK_a so that it could be accurately determined. Samples were excited at 304 nm and emission spectra were collected from 325 to 400 nm. Fluorescence intensities at 371 nm were plotted versus the pH to obtain titration curves. For all titrations, the Hill coefficient was forced to 1 for a two state system when fitting to the Henderson-Hasselbalch equation to determine the pK_a. pH titration performed at (A) 37 °C, 140 mM KCl, and 0.5 mM Mg^{2+}. (B) 25 °C, 140 mM KCl, and 0.5 mM Mg^{2+}. (C) 25 °C, 140 mM KCl, and 0 mM Mg^{2+}. (D) 25 °C, 500 mM KCl, and 0 mM Mg^{2+}.

After finding that decreasing temperature and salt was unable to substantially increase the pK_a of an RNA with A-U and 2AP-U nearest neighbors, we investigated pK_a shifting in an RNA oligonucleotide with G-C and 2AP-U nearest neighbors. We reasoned that the added strength from the G-C nearest neighbor might serve as a means to stabilize pK_a shifting and result in a full titration curve. Initial fluorescence-detected pH titrations of GA^+U (Figure 3-3A) indicated a non-two state system, as decreasing pH did not result in a titration curve as in previously studied constructs. Spectra were collected initially (as was done with previous experiments) and after allowing the sample to equilibrate to correct for any erroneous data due to incomplete equilibration. Comparable fluorescence intensities were obtained with and without equilibration (Figure 3-3B), indicating that the unusual data seen with the GA^+U construct was not due to incomplete equilibration and that collecting spectra without pausing for equilibration was sufficient for other constructs. As with all RNA oligonucleotides studied by fluorescence, the GA^+U oligonucleotide (Figure 3-1) has a 2AP-U base pair above the protonated A^+•C wobble. However, in the GA^+U construct, the 2AP is positioned above the cytosine, rather than the protonated adenine as in other RNA constructs. Although positioning the 2AP above cytosine was successful at reporting a pK_a in DNA,[9] it was not in RNA. Differences in geometries of A-form dsRNA and B-form dsDNA are most likely the cause of the lack of titration curve observed, as titrations of GA^+2AP and UA^+2AP gave at least partial titration curves (Figure 3-3C and D). Because a lower baseline, and thus a full titration curve, could not be obtained using 2AP fluorescence-detected pH-titrations for any of the 2AP-containing RNA sequences tried, ^{31}P NMR was used for the duration of this chapter to study the effect of nearest neighbor strength on pK_a shifting.

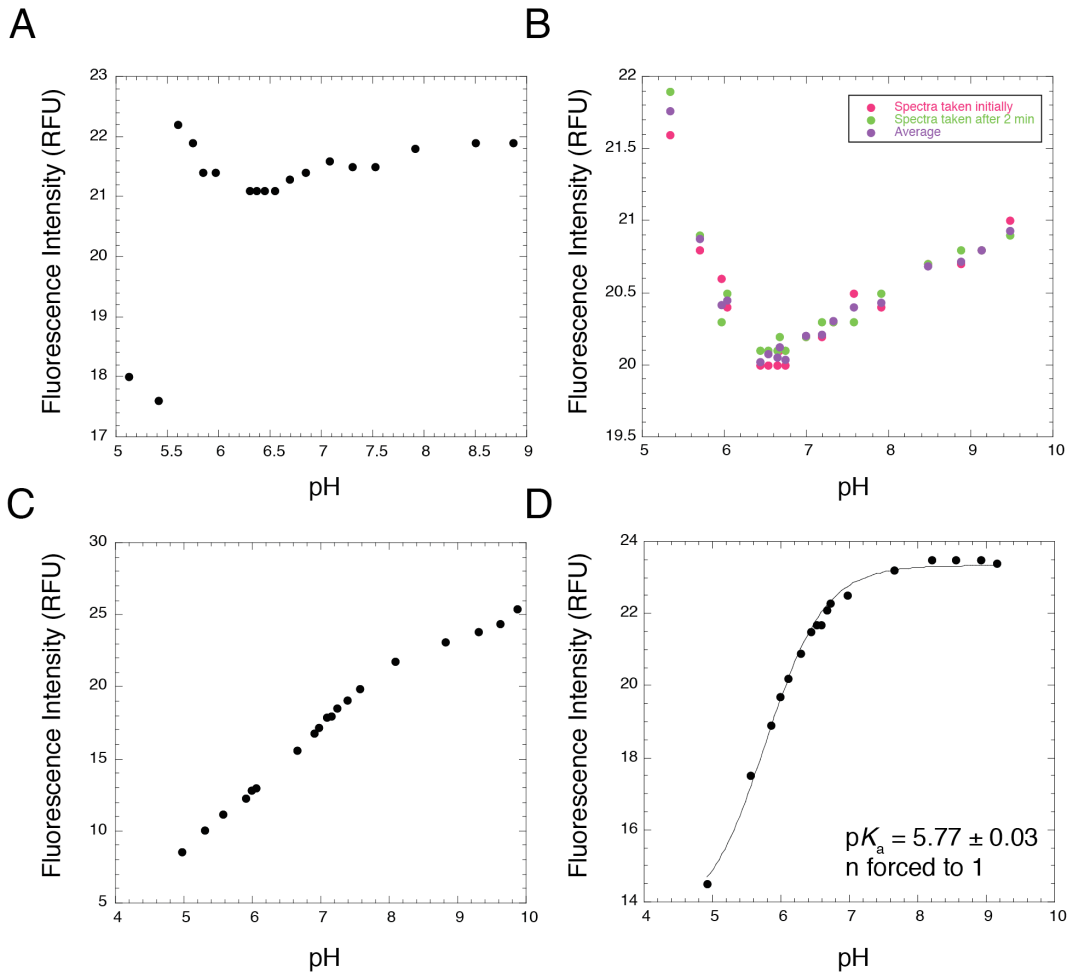


Figure **3-3**: 2-aminopurine fluorescence-detected pH-titrations. The samples were excited at 304 nm and emission spectra were collected from 325 to 400 nm. Fluorescence intensities at 371 nm were plotted versus pH to obtain titration curves. (A) $GA^{+}U$ RNA oligonucleotide. A similar construct with a 5' sequence of $GA^{+}A$ was determined to have a pK_a of 7.28 ± 0.08 by ^{31}P NMR. Because a pK_a could not be determined with the $GA^{+}U$ sample, the sample was re-titrated, taking spectra initially and after allowing the sample to equilibrate for 2 minutes (B), which showed no effect on fluorescence. (C) GA^{+}2AP RNA oligonucleotides. A similar construct with a 5' sequence of $GA^{+}A$ was determined to have a pK_a of 7.28 ± 0.08 by ^{31}P NMR. (D) UA^{+}2AP RNA oligonucleotide. A similar construct with a 5' sequence of $UA^{+}A$ was determined to have a pK_a of 6.51 ± 0.04 by ^{31}P NMR.

3.4.2 ^{31}P NMR-Detected pK_a Determination

To investigate the effect that nearest neighbor base pairing strength has on pK_a shifting, we determined the pK_a of adenine in an A^+•C wobble with two **w**eak (A-U) W-C base pairs (WW), one **s**trong (G-C) and one **w**eak (A-U) W-C base pair (SW), or two **s**trong (G-C) W-C base pairs (SS) as the nearest neighbors (Figure 3-1). Representative NMR titrations are shown in Figure 3-4 and pK_a and Hill coefficient values are provided in Table 3-1. An A^+•C wobble with two weak (A-U) base pairs as nearest neighbors (WW) had a pK_a of 6.51. This was shifted ~3 units upwards from free adenine, but was still the least shifted in this series. The combination of a strong (G-C) and a weak (A-U) base pair (SW) resulted in an intermediate pK_a of 7.28, suggesting a dependence of pK_a on nearest neighbor base-pairing strength. Continuing this trend, flanking the A^+•C base pair with two strong (G-C) base pairs (SS) gave the largest pK_a of 8.10. Notably, changing the nearest neighbors from two A-U base pairs to two G-C base pairs caused an additional pK_a shift of 1.6. We conclude that additional structural stability contributed by neighboring base pairs may enable a more shifted pK_a.

Overlay of the titration curves of the WW, SW, and SS RNA constructs reveals distinct differences in the dynamic range of the chemical shift (Figure 3-4). The WW curve shows an increase in normalized chemical shift upon deprotonation of 1.4%, the greatest of the three constructs, while the SS curve shows an increase of just 0.4%. Likewise the SW construct with the intermediate pK_a, has an intermediate dynamic range of 0.8%. We infer that the lesser dynamic ranges of the SW and especially SS constructs are due to the structural rigidity conferred by the stronger G-C base pairs. This rigidity limits the change in the chemical environment of the phosphorothioate reporter. Likewise, the two weak A-U nearest neighbors in the WW construct may promote flexibility and thus a larger difference in chemical environment between the protonated and deprotonated states, thus leading to a larger dynamic range. We see a similar trend with the dynamic range of fluorescence quenching in DNA oligonucleotides (Table B-1),[9] further

suggesting that decreased dynamic range is due to limited flexibility as a result of strengthening of nearest neighbors.

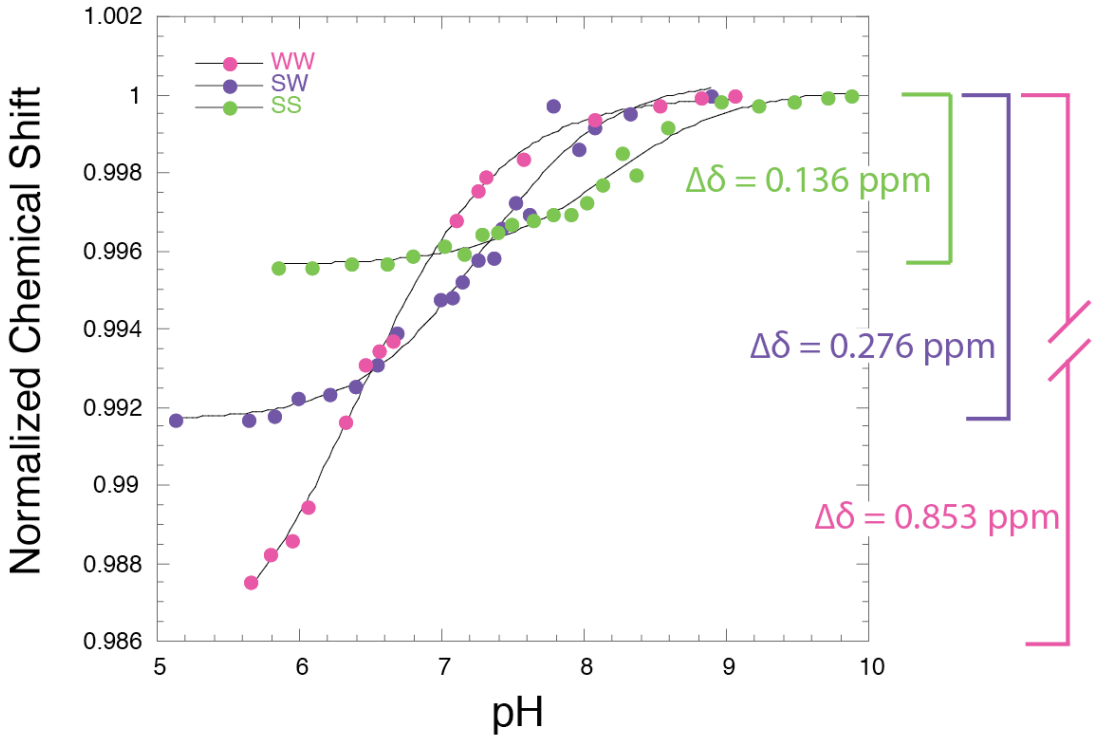


Figure **3-4**: Overlay of ^{31}P NMR-detected pH titrations of WW (pink), SW (purple), and SS (green) RNA constructs. Titration curves were fit to the Henderson- Hasselbalch equation. pK_a and Hill coefficient are provided in Table 1. Actual change in chemical shift of protonated and deprotonated states are given; values are from the raw NMR data. All values are at 100 mM KCl and 25 ^{o}C.

We compared the pK_a results above measured under our standard experimental conditions of 100 mM KCl and 25 °C to more biological conditions of 100 mM KCl and 37 °C or 0.5 mM Mg^{2+}. Titrations under these conditions showed no significant change in pK_a with temperature or magnesium concentration (Figure B-2), suggesting that the values in Table 3-1 are relevant to physiological conditions.

Table **3-1**: pK_a's of RNA oligonucleotides with varying nearest neighbors or proximity of bulges to the AC wobble

Construct	pK_a[a]	Hill coefficient[a]
WW	6.51 ± 0.04[b]	1.01 ± 0.07[b]
SW	7.28 ± 0.08[b]	1.00 ± 0.14[b]
SS	8.10 ± 0.06[b]	1.09 ± 0.16[b]
Bulge-0 (SS')	8.00 ± 0.06	1.34 ± 0.23
Bulge-2	7.46 ± 0.17	Forced to 1[d]
Bulge-1	No pK_a[c]	No Hill coefficient[c]

[a]Curves were fit to the Henderson- Hasselbalch equation (Eq S1) to determine pK_a's and Hill coefficients. Noted errors are from curve fits. [b]The pK_a's and Hill coefficients listed are the averages from the two diastereomer peaks. If not noted, only one peak showed sufficient change in signal to yield a pK_a. [c]No pK_a was reported due to a linear response to pH. [d]For better fitting, the Hill coefficient for Bulge-2 was forced to 1 in the Henderson- Hasselbalch equation. All values are at 100 mM KCl and 25 ^{o}C. Titrations are provided in Figure B-1 and Table 3-2

To understand the thermodynamic driving forces for the nearest neighbor dependence observed in pK_a shifting, UV-monitored thermal denaturation experiments were performed on the WW, SW and SS constructs. Thermodynamic parameters are provided in Table 3-2 and melts in Figures B4 and B5. In all constructs, a more favorable free energy was observed at the lower pH, consistent with protonation and base pair formation of the A^{+}•C. In addition, at both low and high pH, there is an increase in stability upon going from WW to SW to SS constructs, as found for pK_a shifting. Thus, addition of G-C base pairing contributes to the stability of the construct, as well as to shifting of the pK_a.

Table **3-2**: Thermodynamic parameters from UV monitored thermal denaturation

	T_m (° C)	ΔH (kcal/mol)	ΔS (eu)	$\Delta G°_{37}$ (kcal/mol)	$\Delta G°_{50}$ (kcal/mol)	$\Delta G°_{70}$ (kcal/mol)
WW - pH 5.03	66.4 ± 0.17	–49.8 ± 0.50	–146.6	–4.31	–2.40	0.53
WW - pH 8.79	68.2 ± 0.32	–28.6 ± 0.45	–83.8	–2.61	–1.52	0.15
SW - pH 6.13	77.7 ± 0.08	–62.8 ± 0.73	–179.1	–7.29	–4.97	–1.38
SW- pH 8.79	72.6 ± 0.07	–54.8 ± 0.58	–158.6	–5.64	–3.58	–0.41
SS - pH 6.13	89.7 ± 1.73	–53.2 ± 1.28	–146.6	–7.72	–5.81	–2.88
SS - pH 9.22	77.0 ± 0.04	–60.8 ± 0.45	–173.8	–6.95	–4.69	–1.21
SS' pH 6.13	87.2 ± 0.57	–60.7 ± 1.00	–168.5	–8.46	–6.27	–2.90
SS' pH 9.22	80.3 ± 0.08	–67.8 ± 1.52	–191.7	–8.31	–5.81	–1.98

So far, we have described pK_a and stability effects due to large differences in base pairing strength, going from weak to strong base pairs. We also wished to examine dependence of these parameters on sequences within the same types of base pairs. Parameters of SS were thus compared to SS', a construct having two G-C nearest neighbors but with a 5' strand sequence of GA^+G, rather than GA^+C in the SS construct. The pK_a's were the same within experimental error (Table 3-1), which is in line with their similar $\Delta G°_{50}$ and T_M values (Table 3-2).

In addition to different nearest neighbors, many systems have bulges proximal to AC wobbles. To examine the effect of disruption of helical structure near a protonated base, we determined the pK_a in three constructs: ones with a bulge one (Bulge-1) or two (Bulge-2) base pair(s) away from an A^+•C mismatch, and an analogous RNA construct with no bulge (Bulge-0) (Figure B-3). As described above, Bulge-0 (=SS') had a highly shifted pK_a of 8.00, and Bulge-2 had a shifted pK_a of 7.46. Thus, positioning the bulge two base pairs away from the A^+•C had a

minor effect on the pK_a, lowering it just 0.54 units from the perfectly double-stranded construct, with a value still above neutrality. Strikingly, no pK_a was observed for Bulge-1, which was further investigated using ^{1}H NMR.

Base pairing in Bulge-0, Bulge-1 and Bulge-2 constructs was analyzed by ^{1}H NMR at both low (~ 6) and high (~ 9) pH (Figure 3-5). G-C and A-U imino protons resonate at ~12-13 and ~13-15 ppm, respectively.[15] Of the three constructs analyzed, only Bulge-0 (Figure 3-1) contains an A-U base pair, resonating at 14.4 ppm in spectra at both pH. A resonance near 10 ppm at low and high pH is due to an imino proton from position 4 of the loop, and a resonance near 11.8 ppm at low pH is from an imino at position 1 of the loop.[16] These resonances confirm formation of hairpin rather than duplex.[17] Enhanced sharpening and resolution of peaks at low pH is indicative of base pairing adjacent to A^+•C wobbles,[15] and is observed in the low pH spectra of Bulge-0 and Bulge-2 constructs (Figure 3-5). Bulge-1, however, has distinctly fewer peaks, both at low and high pH, correlating to fewer Watson-Crick base pairs and the lack of pK_a observed in the ^{31}P NMR-detected pH titration. Positioning the bulge one base pair away from the A^+•C wobble is disruptive to the stem, apparently prohibiting formation of a G-C base pair below the bulge and above the A^+•C wobble. In addition, only four distinct peaks are seen in the NMR spectra of Bulge-1, with some lesser peaks evident. The large bulge in the middle of the stem most likely caused substantial breathing of the G-C base pairs above and below the bulge, eliminating peaks for these base pairs, as well.

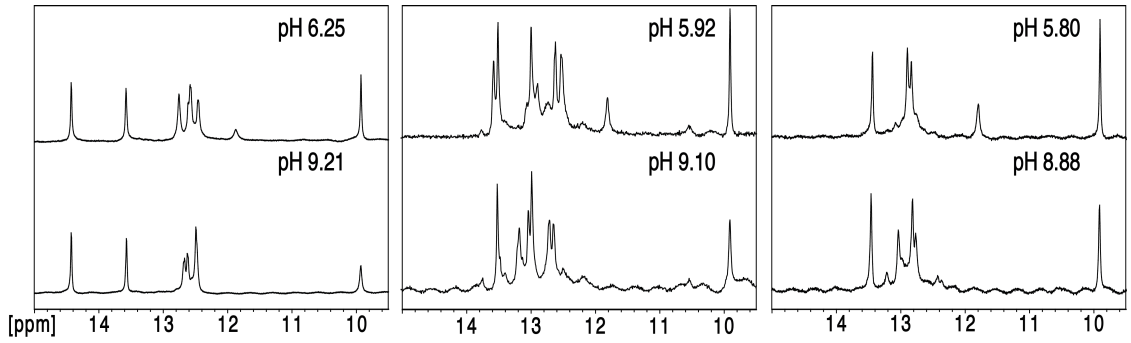


Figure **3-5**. ^{1}H NMR of Bulge-0 (left), Bulge-2 (middle), and Bulge-1 (right). Spectra were collected at 3 °C to limit exchange of imino protons at low (top) and high pH (bottom)

Appreciation of dsRNA's importance in biological processes such as regulation of gene expression by miRNA's,[4-6] RNA editing,[18-20] and ribosomal frameshifting[7] is steadily increasing. In RNA editing by ADARs, a preference for editing A's opposite C's has been reported.[18] Wong *et al.* indicated that the AC preference is due to sterics, selectivity, and the strength of base pairs formed. Positioning an A opposite a purine disfavors editing due to the bulk of the purine, while facile editing of an A-U base pair is disfavored because it would lead to non-selective editing. More recently, Kuttan and Bass observed a nearest neighbor dependence of editing the A opposite the C, in which an AC with two strong nearest neighbors is edited much more slowly than an AC with one strong and one weak nearest neighbor. They further showed that editing is correlated with the ability of the adenine to be flipped out.[19] We have shown that a construct in which an AC wobble is flanked by two strong nearest neighbors has a higher pK_a, suggesting that a higher pK_a may lead to less flipping and hence less editing. This possibility suggests that editing could be tuned by pK_a shifting.

3.4.3 Bioinformatics of miRBase to Investigate A^+•C Wobbles' Role in miRNA Processing

In gene regulation by miRNAs, protonation of adenine in an A^+•C base pair could affect miRNA processing. As previously shown, cleavage by Drosha is dependent upon location of a structure deformation at the cleavage site.[21] Protonation of adenine in an A^+•C base pair at the cleavage site might influence cleavage and miRNA processing. A bioinformatics survey of A•C's in miRNAs showed a large number located at the microprocessor cleavage site. Since the pK_a of an A^+•C is highly dependent upon nearest neighbors and structural changes, we reasoned that, to ensure a defect required for processing by the microprocessor, A^+•C's at the cut site should have weak nearest neighbors and nearby structural defects to further ensure deprotonation and cleavage. To alter miRNA processing in cancer cells with low pH,[22] the A^+•C could potentially be protonated, forming a base pair and removing defects required for cleavage.

For investigating the potential effect pK_a shifting has on miRNA processing, we employed a bioinformatics approach using RNA Pair Analyzer, a program that tallies base pairs in a secondary structure and their nearest neighbors.[13] All human pri-miRNAs in the miRBase (Release 19) were analyzed and data are compiled in Table B-2. Of all 2331 AC's found in the miRNAs, 3.4% were found with the 5' sequence GAG, which was only 0.8% more common than RNAs with the 5' sequence AAA (Table B-2). As determined previously by NMR, constructs with two strong G-C nearest neighbors have a pK_a 1.5 units higher than one with two weak A-U nearest neighbors. If nearest neighbor dependent pK_a shifting was responsible for reducing miRNA processing, we would expect all sequences with strong nearest neighbors to be most common as a means to influence pK_a shifting, base pairing and ultimately cleavage and miRNA processing. Furthermore, the percentage of AC's with two strong GC nearest neighbors regardless of sequence was 9.3%, only 0.8% higher than AC's with two weak A-U nearest neighbors (Table B-3). 14% of all AC's found in the miRBase had one strong GC and one weak A-U nearest neighbor, which, when accounting for the additional possible sequence combinations of a SW

construct is comparable to SS and WW sequences (1.8%, 2.3%, and 2.1%, respectively) (Table B-3) providing no apparent trends.

Since miRNAs demonstrate a lack of preference for nearest neighbor base pair strength, we investigated the possibility of control by bulge location. RNA's found in the bioinformatics study with two strong G-C nearest neighbors were categorized by their proximity to a bulge. miRNA's had either a bulge one base pair away from the A•C wobble (Bulge-1 like), two base pairs away (Bulge-2 like), or three or more base pairs away (Bulge-0 like). A slightly higher percentage of constructs had a bulge three or more base pairs away from the A•C wobble (54.1%) when compared to that of two base pairs away (23.9%) or one base pair away (21.9%) from the A•C wobble. Based on the pK_a's for bulged constructs determined by NMR earlier in this chapter, it can be estimated that 54.1% of AC wobbles with two strong GC nearest neighbors have a pK_a shifted above neutrality, as in the Bulge-0 construct. Bulge-2 like was determined to have a shifted pK_a towards neutrality by NMR, which indicates that 23.9% of AC's with two strong GC nearest neighbors have pK_a's near neutrality nearby as much as Bulge-0. Finally, based on the inhibition of protonated base pair formation in the Bulge-1 construct by NMR, we estimate that 21.9% of AC's with two strong GC nearest neighbors in Bulge-1 like constructs are not formed in miRNA's and instead result in bulge formation. These data indicate that more than half of AC's with two strong GC nearest neighbors in miRNA's are protonated, forming an A^+•C wobble, analogous to a GU wobble and can be tuned to alter miRNA processing. Furthermore, one fourth of AC's with two strong GCs are not only neutral and mismatched but have a large bulge encompassing the bulge, AC, and the GC in between the bulge and AC, as in the Bulge-1 construct. Since deformable sequences at the cut site and one helix turn away from the cut site are required for miRNA processing by Drosha,[21] we suggest tuning of pK_a's of AC's by bulge location to be a method to affect miRNA processing. Protonation of adenine in an A^+•C wobble would result in base pair formation, and remove the potential for deformation required for miRNA processing. Based on these bioinformatics data, approximately one fourth of AC's with

two G-C nearest neighbors have pK_a's at neutrality and have the potential to greatly aid in miRNA processing.

3.5 Conclusion

In this study, we have observed a significant dependence of pK_a shifting on nearest neighbors in RNA. Altering the nearest neighbors from two weak A-U base pairs to two strong G-C pairs increased the pK_a by 1.6 units, up to 8.1 as determined by ^{31}P NMR. 2AP fluorescence could not be used to determine the pK_a, however, as a lower baseline could not be obtained. Addition of a bulge two base pairs away from the protonated base pair reduced shifting of the pK_a by only 0.54 units and still showed a significantly shifted pK_a of 7.46. Positioning a bulge one base pair away, however, caused perturbation of helical structure and prohibited a protonated base pair from forming. Analysis of AC's in miRNA's showed a surprising lack of favorable nearest neighbors, although tuning of pK_a's by bulge proximity is likely. We hope that the trends presented in this study will help determine the contribution protonated bases make in various dsRNA-mediated processes.

3.6 Acknowledgements

The authors would like to thank the Penn State NMR facility including Emmanuel Hatzakis, Wenbin Luo, and John Litner, and Scott Showalter for assistance with ^{31}P and ^{1}H NMR experiments, We would also like to thank Seth M. Morton for development of the RNA Pair Analyzer program.

3.7 References

1. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer-Verlag, New York.

2. Legault, P., and Pardi, A. (1997) Unusual Dynamics and pKa Shift at the Active Site of a Lead-Dependent Ribozyme, *J. Am. Chem. Soc.* **119**, 6621-6628.

3. Wilcox, J. L., Ahluwalia, A. K., and Bevilacqua, P. C. (2011) Charged nucleobases and their potential for RNA catalysis, *Acc Chem Res* **44**, 1270-1279.

4. Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans, *Science* **294**, 858-862.

5. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs, *Science* **294**, 853-858.

6. Lee, R. C., and Ambros, V. (2001) An extensive class of small RNAs in Caenorhabditis elegans, *Science* **294**, 862-864.

7. Giedroc, D. P., Theimer, C. A., and Nixon, P. L. (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting, *J Mol Biol* **298**, 167-185.

8. Moody, E. M., Brown, T. S., and Bevilacqua, P. C. (2004) Simple method for determining nucleobase pK(a) values by indirect labeling and demonstration of a pK(a) of neutrality in dsDNA, *J. Am. Chem. Soc.* **126**, 10200-10201.

9. Wilcox, J. L., and Bevilacqua, P. C. (2013) A Simple Fluorescence Method for pK(a) Determination in RNA and DNA Reveals Highly Shifted pK(a)'s, *J. Am. Chem. Soc.* **135**, 7390-7393.

10. Siegfried, N. A., O'Hare, B., and Bevilacqua, P. C. (2010) Driving forces for nucleic acid pK(a) shifting in an A(+).C wobble: effects of helix position, temperature, and ionic strength, *Biochemistry* **49**, 3225-3236.

11. Gottlieb, H. E., Kotlyar, V., and Nudelman, A. (1997) NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities, *J Org Chem* **62**, 7512-7515.

12. Siegfried, N. A., and Bevilacqua, P. C. (2009) Thinking inside the box: designing, implementing, and interpreting thermodynamic cycles to dissect cooperativity in RNA and DNA folding, *Methods Enzymol* **455**, 365-393.

13. Morton, S. M. (2004) RNA Pair Analyzer, 1 ed., p Program to analyze RNA base pairs, State College.

14. Moody, E. M., Lecomte, J. T., and Bevilacqua, P. C. (2005) Linkage between proton binding and folding in RNA: a thermodynamic framework and its experimental application for investigating pKa shifting, *RNA* **11**, 157-172.

15. Allawi, H. T., and SantaLucia, J., Jr. (1998) Nearest-neighbor thermodynamics of internal A.C mismatches in DNA: sequence dependence and pH effects, *Biochemistry* **37**, 9435-9444.

16. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) Structure of an unusually stable RNA hairpin, *Biochemistry* **30**, 3280-3289.

17. Proctor, D. J., Kierzek, E., Kierzek, R., and Bevilacqua, P. C. (2003) Restricting the conformational heterogeneity of RNA by specific incorporation of 8-bromoguanosine, *J. Am. Chem. Soc.* **125**, 2390-2391.

18. Wong, S. K., Sato, S., and Lazinski, D. W. (2001) Substrate recognition by ADAR1 and ADAR2, *RNA* **7**, 846-858.

19. Kuttan, A., and Bass, B. L. (2012) Mechanistic insights into editing-site specificity of ADARs, *Proc Natl Acad Sci U S A* **109**, E3295-3304.

20. Goodman, R. A., Macbeth, M. R., and Beal, P. A. (2012) ADAR proteins: structure and catalytic mechanism, *Curr Top Microbiol Immunol* **353**, 1-33.

21. Quarles, K. A., Sahu, D., Havens, M. A., Forsyth, E. R., Wostenberg, C., Hastings, M. L., and Showalter, S. A. (2013) Ensemble Analysis of Primary MicroRNA Structure Reveals an Extensive Capacity To Deform near the Drosha Cleavage Site, *Biochemistry* **52**, 795-807.

22. Griffiths, J. R. (1991) Are cancer cells acidic?, *Br J Cancer* **64**, 425-427.

Appendix B

Chapter 3 Supplemental Information

[Selected data were submitted in the Supporting Information for a paper entitled "pK_a Shifting in Double-Stranded RNA is Highly Dependent Upon Nearest Neighbors and Bulge Positioning" by Jennifer L. Wilcox and Philip C. Bevilacqua in *Biochemistry* 2013. Additional experiments were performed through revision of the manuscript and can be found in the submitted manuscript.]

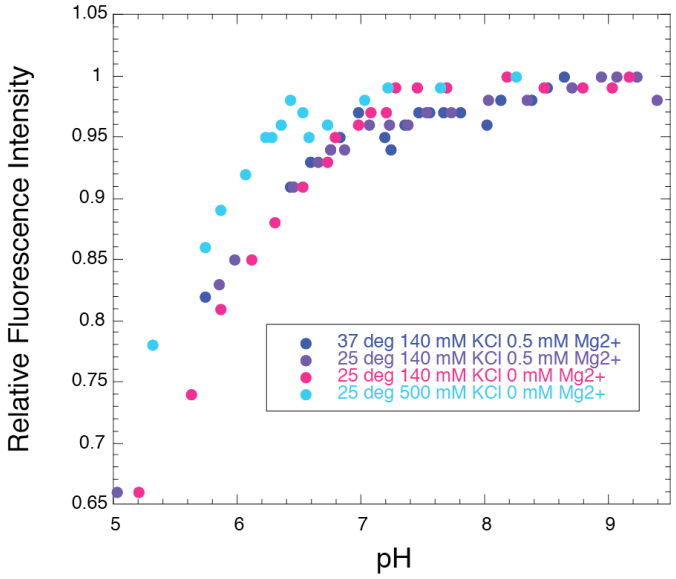


Figure **B-1**: Overlay of 2-aminopurine fluorescence-detected pH-titrations of an AA^+2AP RNA oligonucleotide. Non-overlaid titration curves are provided in Figure 3-2. Various reaction conditions were used in an attempt to increase the pK_a so that it could be accurately determined. Samples were excited at 304 nm and emission spectra were collected from 325 to 400 nm. Fluorescence intensities at 371 nm were plotted versus the pH to obtain titration curves. pH titration performed at 37 °C, 140 mM KCl, and 0.5 mM Mg^{2+} (blue), 25 °C, 140 mM KCl, and 0.5 mM Mg^{2+} (purple), 25 °C, 140 mM KCl, and 0 mM Mg^{2+} (pink), and 25 °C, 500 mM KCl, and 0 mM Mg^{2+} (cyan). Overlay of these plots shows the non-effect reducing the temperature from 37 to 25 °C (blue to purple curves) and reducing the magnesium concentration from 0.5 to 0 mM (purple to pink curves) has on pK_a shifting of the AA^+2AP oligonucleotide as little change in titrations curves is observed. Drastically increasing the K^+ concentration had the greatest effect on pK_a (pink to cyan), although it still showed only fairly minimal change in titration curves.

Table **B-1**: Dynamic range of NMR chemical shift (δ) change upon protonation of DNA and RNA oligonucleotides

	DNA Sequence	RNA Sequence	Dynamic Range of DNA (ppm)	Dynamic Range of RNA (ppm)
SS	CA^+G	GA^+C	0.21	0.14
SW	CA^+A	GA^+A	0.36	0.28
WW	TA^+T	UA^+A	0.45	0.86

For sequence comparison of DNA and RNA constructs, sequences of the 5' strand are provided. In each case, the A^+ noted is the ionizable nucleobases in an A^+•C wobble. Dynamic range of DNA and RNA were determined using the change in chemical shift upon protonation in an A^+•C wobble. Chemical shifts from ^{31}P NMR of DNA or RNA were normalized before comparison.

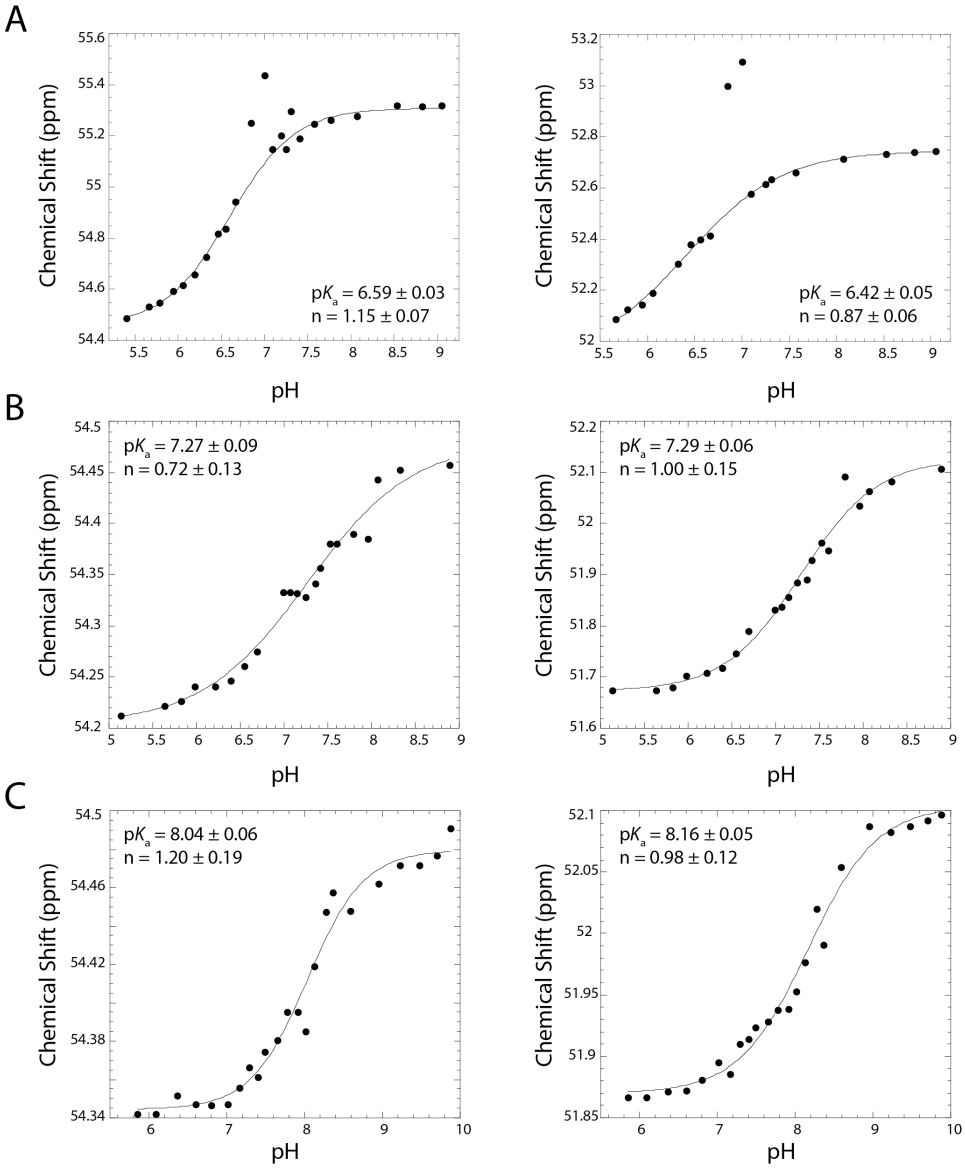


Figure **B-2**: Investigation of nearest neighbors on pK_a shifting using ^{31}P NMR-detected pH titrations. (A) pH titration of WW, a construct with two weak A-U Watson-Crick base pairs as nearest neighbors. The outlying points (three in left panel and two in right panel) were omitted in fitting of the data. (B) pH titration of SW, a construct with one strong G-C Watson-Crick base pair and one weak A-U Watson-Crick base pair as nearest neighbors. (C) pH titration of SS, a construct with two strong G-C Watson-Crick base pairs as nearest neighbors. For all panels, the two plots reflect the two diastereomer peaks that occur upon phosphorothioate substitution. Curves were fit to the Henderson- Hasselbalch equation to determine the pK_a and Hill coefficient. Secondary structures of RNA oligonucleotides are provided in Figure 1. All values were determined in 100 mM KCl and 25 ^{o}C

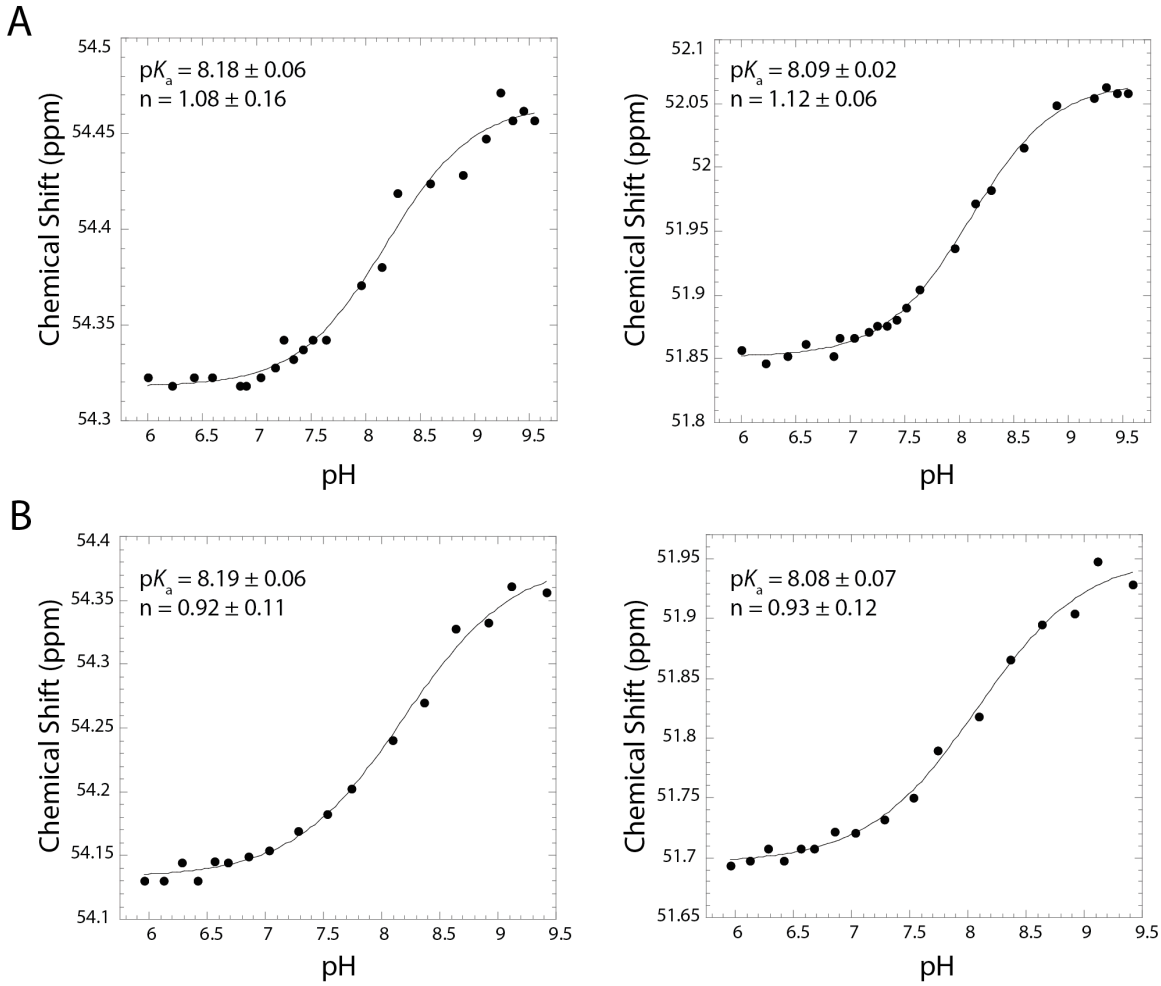


Figure **B-3**: Comparison of standard experimental conditions to biological conditions using ^{31}P NMR-detected pH titrations with oligonucleotide SS. (A) pH titration at biological magnesium concentrations (0.5 mM Mg^{2+}) (B) pH titration at 37 °C. For both panels, the two plots reflect the two diastereomer peaks after phosphorothioate substitution. Curves were fit to the Henderson-Hasselbalch equation to determine the pK_a and Hill coefficient. All values are in a background of 100 mM KCl.

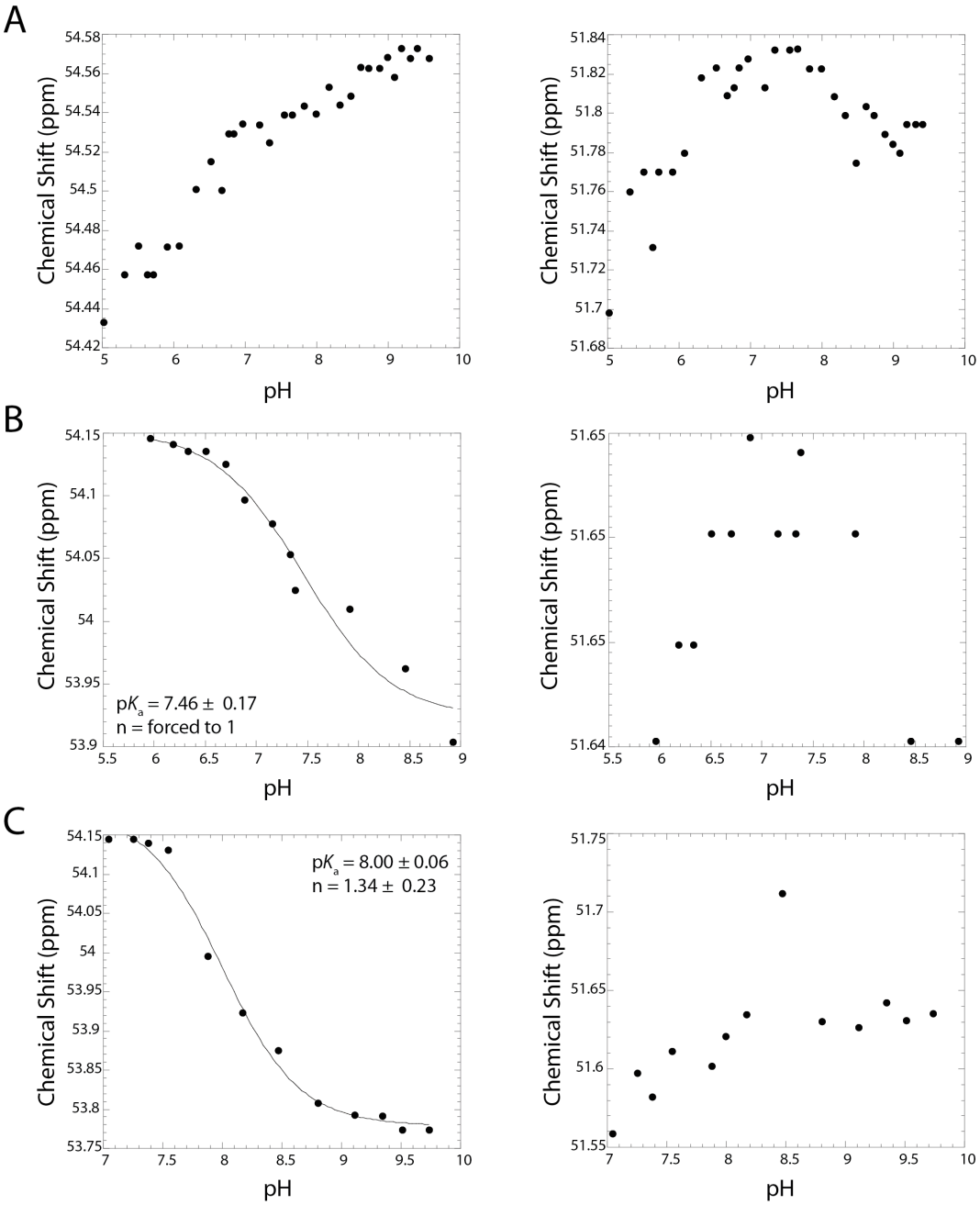


Figure **B-4**. Effect of bulge position on pK_a shifting using ^{31}P NMR-detected pH titrations. (A) pH titration of a construct containing a bulge one base pair away from A^+•C mismatch. (Bulge-1) (B) pH titration of a construct containing a bulge two base pairs away from A^+•C mismatch. (Bulge-2) (C) Analogous RNA construct with no bulge, referred to as Bulge-0 and SS'. For all panels, the two plots reflect the two diastereomer peaks after phosphorothioate substitution. Curves were fit to the Henderson- Hasselbalch equation to determine the pK_a and Hill coefficient. All values are at 100 mM KCl and 25 ^{o}C.

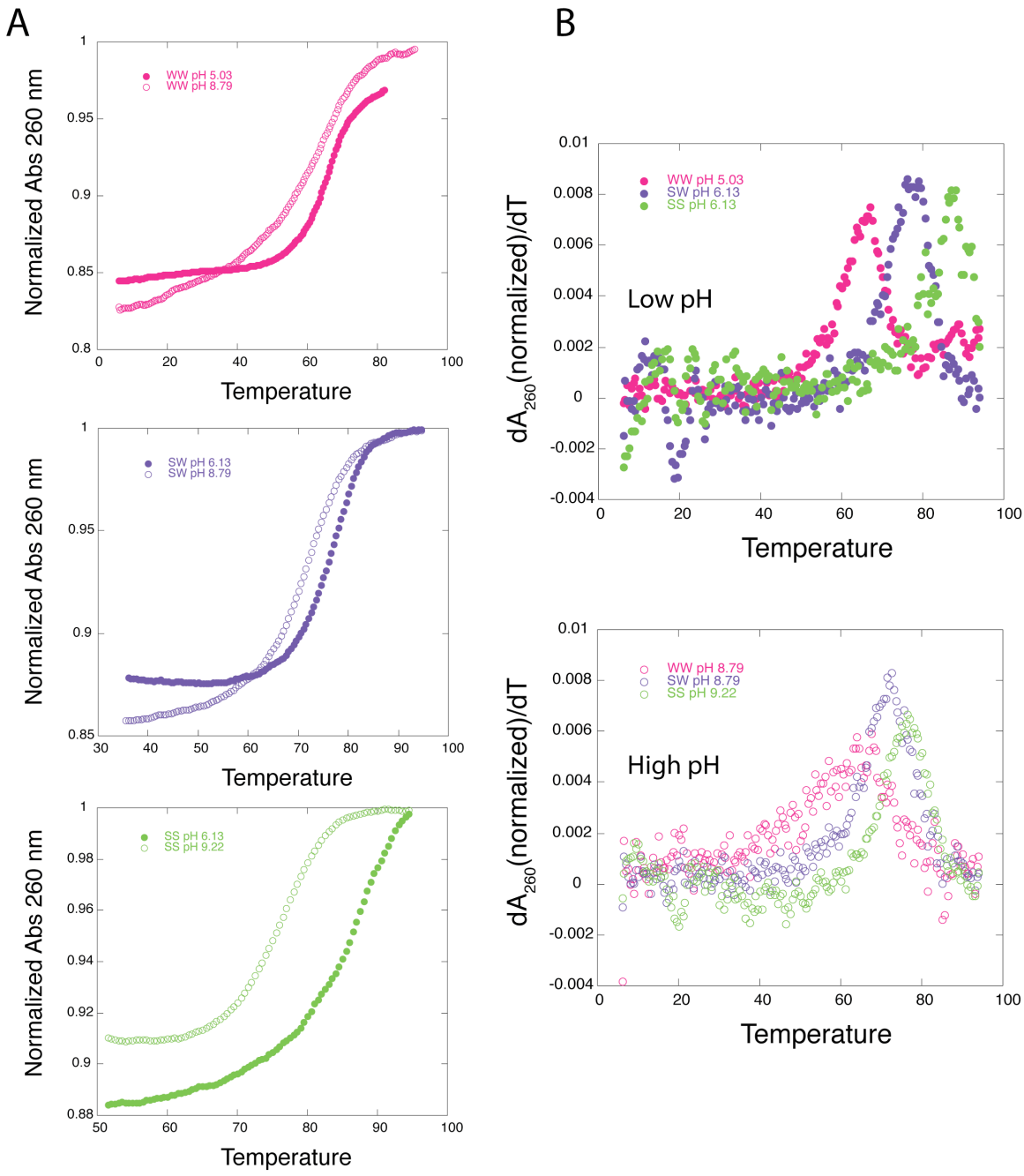


Figure **B-5**. Comparison of nearest neighbor constructs using UV-detected thermal denaturation experiments. Absorbance was detected at 260 nm for RNA constructs. Buffer pH was chosen based on the pK_a of the construct and was chosen so that it was >1 pH away from the pK_a determined by NMR. (A) Absorbance curves are as follows: WW (top panel-low pH 5.03, high pH 8.79), SW (middle panel-low pH 6.13, high pH 8.79), and SS (bottom panel-low pH 6.13, high pH 9.22) at low (•) and high (°) pH. (B) First derivatives of data for low pH (top panel) and high pH (bottom panel).

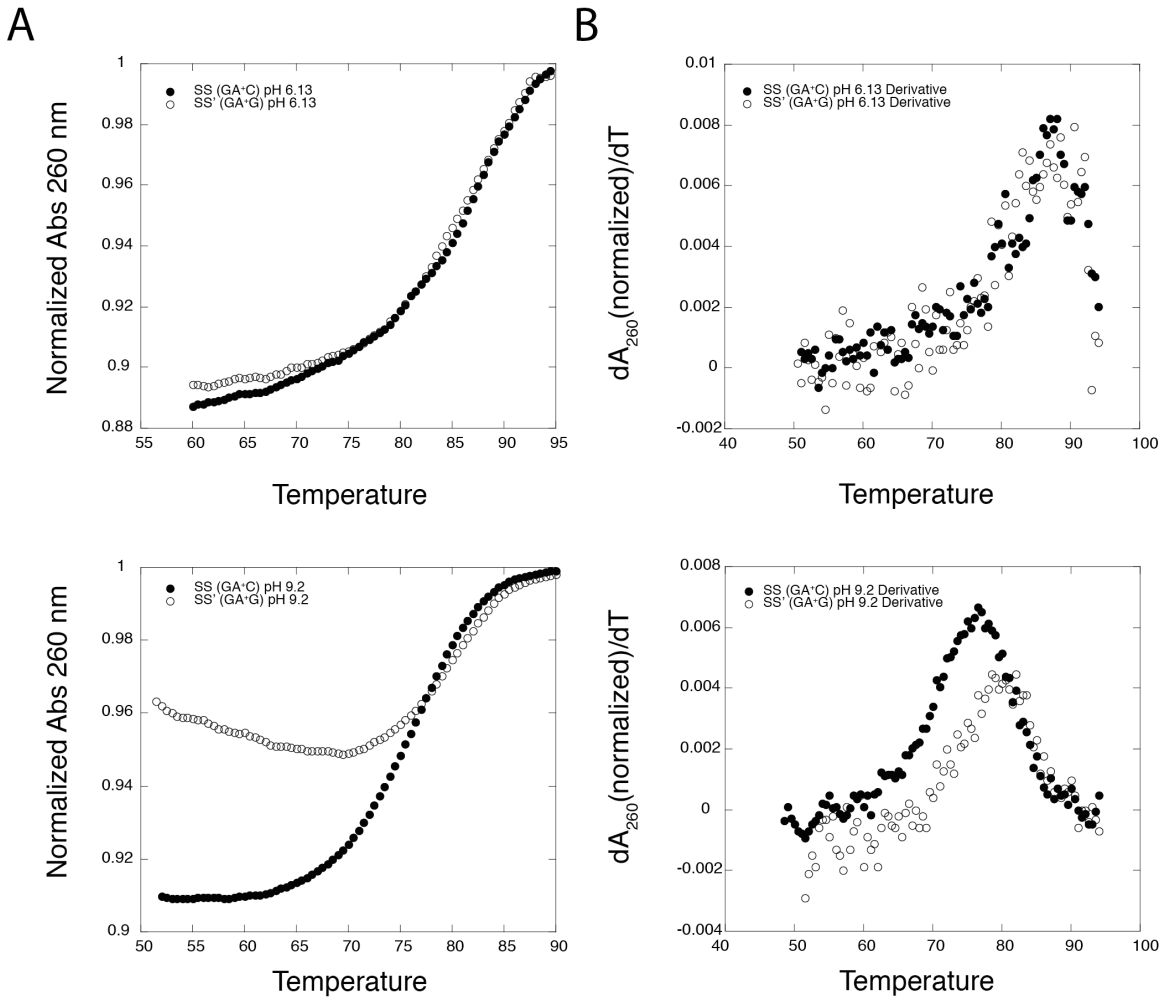


Figure **B-6.** Comparison of two RNA constructs each with strong base pair nearest neighbors using UV-detected thermal denaturation experiments. Absorbance was detected at 260 nm for RNA constructs. Buffer pH was chosen based on the pK_a of the construct and was chosen so that it was >1 pH away from the pK_a determined by NMR. (A) Absorbance curves were as follows: SS (with 5' sequence GA^+C) and SS' (with 5' sequence GA^+A) constructs at low (top panel) and high (bottom panel) pH. (B) First derivatives of data for low pH (top panel) and high pH (bottom panel).

Table **B-2**: Most common human pri-miRNA nearest neighbors.

Sequence[a]	Occurrence	Percentage of AC's[b]
GAG CCC	79	3.4%
AAA UCU	61	2.6%
GAU CCA	58	2.5%
UAG ACC	54	2.3%
AAG UCC	52	2.2%
CAG GCC	52	2.2%
AAU UCA	50	2.1%
GAA CCU	49	2.1%
UAU ACA	47	2.0%
GAC CCG	46	2.0%
UAA ACU	41	1.8%
CAC GCG	40	1.7%
CAA GCU	37	1.6%
UAC ACG	37	1.6%
AAC UCG	28	1.2%
CAU GCA	28	1.2%

[a]All sequences are listed in 5' to 3' for the top row. [b]The percentage noted is percent of 2331 AC's found

Table **B-3**: Most common human pri-miRNA nearest neighbors to AC's categorized by nearest neighbor base pair strength.

Nearest Neighbor Type[a]	Occurrence	Percentage[b]	Average Percentage[c]
WW	199	8.5%	2.1%
WS/SW	343	14.7%	1.8%
SS	217	9.3%	2.3%

[a]Nearest neighbors of miRNA's were categorized into WW (two weak A-U base pairs), SW (one weak A-U base pair and one strong G-C base pair), and SS (two strong G-C base pairs). [b]Percentage noted is of 2331 AC's found. The remaining AC's found have non-canonical nearest neighbors and were not investigated in this study. [c]Average percentage was determined by dividing the percentage of nearest neighbor type by the number of potential combinations for the nearest neighbor type. This was done to account for the additional possible base pair combinations for WS/SW constructs and to allow for comparison between WW, SW/WS and SS.

Table **B-4**: AC's with SS nearest neighbors categorized by their contexts.

Proximity of Bulge to A•C	Occurrence	Percentage[a]
Bulge-0 like	111	54.1%
Bulge-2 like	49	23.9%
Bulge-1 like	45	21.9%

miRNAs were categorized based on the proximity to a bulge. RNAs in were either Bulge-0 like, with a bulge three or more base pairs away from the A•C; Bulge-2 like, with a bulge two base pairs away from the A•C; or Bulge-1 like, with a bulge one base pair away from the A•C. Although all 217 SS RNAs were analyzed, only 205 were categorized as 12 RNAs were adjacent to a bulge.

Chapter 4

Chemical Method for Mapping Protonated Bases in Large RNA Systems

[Portions of the materials and methods section were published in part as a paper entitled "Mechanistic Analysis of the Hepatitis Delta Virus (HDV) Ribozyme: Methods for RNA Preparation, Structure Mapping, Solvent Isotope Effects, and Co-transcriptional Cleavage" by Durga M. Chadalavada, Andrea L. Cerrone-Szakal, Jennifer L. Wilcox, Nathan A. Siegfried and Philip C. Bevilacqua in *Methods in Molecular Biology* 2012 **848**: 21-40.]

4.1 Abstract

As described in earlier chapters, methods exist for quantifying pK_a shifting in nucleic acids, many of which were developed in our lab such as phosphorothioate NMR, Raman crystallography, and 2AP fluorescence. These methods are useful in determining the pK_a of a *known* nucleobase with a pK_a that is shifted towards neutrality. Identification of an unknown nucleobase with a shifted pK_a would be advantageous to determining mechanisms for general acid-base catalysis, RNA editing and potentially many other biological processes. Currently, no method is available for directly identifying which nucleobase is protonated. In this study, we describe efforts towards development of a chemical method for mapping protonated bases in large RNA systems. To identify protonated bases, attempts were made to selectively damage the protonated base. Then Sanger sequencing was employed to detect the damaged base, which proved to be problematic. Efforts were made to optimize sequencing reactions and gel electrophoresis and treatment with a chemical reagent under various conditions. Ultimately, mass

spectrometry was used to determine the viability of ammonium bisulfite as a means to damage protonated bases, which proved to not react with C^+ under present conditions.

4.2 Introduction

The pK_a's of the nucleobases are far from neutrality, with A and C having pK_a's of 3.5 and 4.2, respectively, and G and U or T with pK_a's of 9.2.[1] Upon Watson-Crick base pairing, the pK_a's of the nucleobases shift even *further* from neutrality.[2] But, there are some non-canonical base pairings in which the pK_a's of the nucleobases are shifted towards neutrality. These cases are of biological importance because they allow the nucleobases to have potential roles in general acid base catalysis, RNA editing, and other processes.[3]

As discussed in Chapters 1-3, spectroscopic methods exist, one of which was developed in Chapter 2, to determine the pK_a of a nucleobase in an RNA system.[4] Existing methods, such as fluorescence,[4] UV-Vis,[5] CD,[6] and Raman spectroscopy,[7, 8] however, lack the capability to directly determine which nucleobase is protonated in a large RNA system. There is a potential for identifying a protonated base with ^{13}C NMR,[9] but RNA size is limited to less than 80 nucleotides, which severely limits the application of this technique, as systems of interest are often much larger than 80 nucleotides. In systems where the identity of the protonated base is unknown, spectroscopic methods are useful in quantifying the pK_a but cannot report on the identity of ionizing nucleobase.[6] The ability to identify the ionizing nucleobase would clearly serve to answer many key questions regarding the specific mechanism of a multitude of biological processes. For example, programmed ribosomal frameshifting requires formation of a pseudoknot,[10] which is frequently a result of protonation, and ribozyme self-cleavage takes place through general acid-base catalysis.[11] In addition, possessing the ability to map protonated bases on RNA of any size, ultimately on a genome-wide level, could lead to the discovery of many new

protonated bases and motifs in RNA, and could help elucidate the roles they play in biological processes.

Chemical methods have been used to modify nucleic acids for almost half a century, and include DMS, DEPC, and bisulfite reagents, among many others.[12-17] Existing methods damage nucleobases in specific environments (i.e. single or double stranded, protected by proteins, and others). Identification of the damaged base is performed either by reverse transcription (RT) or fragmentation. If the chemical reagent damages the Watson-Crick face of the nucleobase, primer extension can be used to visualize stops one nucleobase downstream of the damaged nucleobase, as damage at the selected nucleobases would inhibit primer extension. If the reagent cleaves the phosphate backbone, the fragmentation results in RNA pieces of different lengths, which can also be separated based on size. With both RT and backbone cleavage, polyacrylamide gel electrophoresis (PAGE) or capillary electrophoresis (CE) is used for analysis. Challenges exist with chemical modification of nucleic acids, however, as genetic information is stored in stable, non-reactive pseudo-aromatic rings. Most existing methods for base damage thus require usage of highly concentrated solutions to damage the nucleobases. Development of a method to only map protonated bases requires a special reagent, strong enough to attack the pseudo-aromatic nucleobases but selective enough to only react with protonated bases. Following is a survey of some of the reagents of interest.

4.2.1 Sodium Borohydride

Sodium borohydride is a common reducing agent in organic chemistry. For example, Warren *et al* utilized its reducing properties to reduce and thereby identify a Schiff base in the mechanism of decarboxylation of acetoacetate decarboxylase, as shown in Figure 4-1A.[18] Because of the similarity of functional groups with a Schiff base and a protonated nucleic acid,

sodium borohydride has the potential as a chemical reagent to damage protonated bases. In this reaction, hydride attacks either C6 of cytidine (Figure 4-1B) or C2 of adenine (Figure 4-1C) to form a buckled, aliphatic product. Recently, this method was used to determine protonation in a C^{+}•U base pair in the 50S subunit of *Haloarcula marismortui* ribosome although further results were never presented.[19] Sodium borohydride is very reactive at low pH, due to reaction of hydride H^{-} with H^{+} to make H_2 gas (see Scheme 4-1 in Section 4.3.3), and it was estimated that the hydride loses its reactivity at pH 5.9 and 0 °C after 10 seconds.[18] Because of this, utilization of a more selective hydride, such as sodium cyanoborohydride, could potentially be a more viable option.[20]

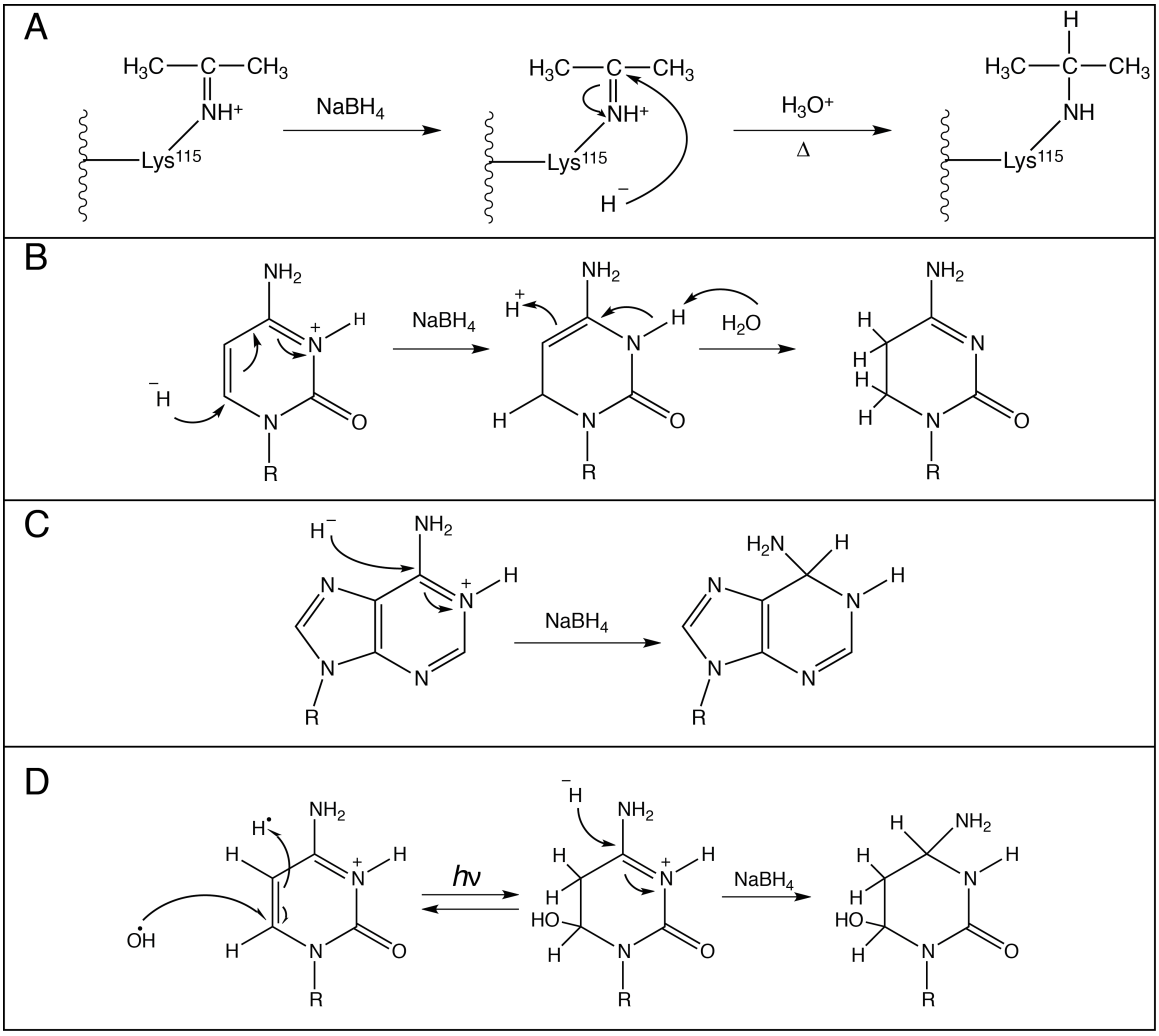


Figure **4-1**: Hydride and photohydration reactions. (A) Sodium borohydride was previously used to reduce a Schiff base to determine the mechanism of acetoacetate decarboxylase.[18] (B) Proposed mechanism for reduction of C^{+}.[19] The hydride attack is analogous to that in reduction of a Schiff base and begins with attack at C6. The product is buckled and aliphatic. (C) Proposed mechanism for reduction of A^{+}.[19] The hydride attacks C6, forming a buckled, aliphatic ring. (D) Proposed photohydration mechanism for reduction of C^{+}.[21] Photohydration proceeds similar to reduction by hydride. A hydroxide radical attacks the nucleobase at C6, resulting in a buckled intermediate. Subsequent treatment by sodium borohydride reduces the protonated nucleobase to form an aliphatic ring product.[21]

4.2.2 Photohydration

The effect of irradiation on nucleic acids has been utilized for almost half a century.[21] Photoinduced cross-linking of nucleic acids is commonly used to detect RNA-protein interactions in CLIP (cross-linking and immunoprecipitation) experiments.[22] In photohydration, a hydroxyl radical attacks cytidine at C6, resulting in reaction with a hydrogen free radical (Figure 4-1D). As photohydration is reversible, subsequent treatment with sodium borohydride attacks the photohydration product to form a stable, buckled aliphatic ring system.[21] Because of the increase in reactivity of cationic nucleobases, the photohydration reaction may preferentially occur in protonated bases over neutral nucleobases.

4.2.3 Hydrazine and Hydroxylamine

Chemical sequencing has depended on the reaction of hydrazine with C's and T's for decades.[12] Subsequent treatment with aniline causes cleavage of the phosphate backbone. Hydroxylamine and potassium permanganate have also been used to damage mismatched cytosines and thymines, respectively.[23] Both of these reactions begin by an attack on C6 by the lone pair of nitrogen (Figure 4-2). The hydrazine reaction proceeds with an attack from the secondary amino group on the same hydrazine molecule to create a bridged hydrazine substituted product (Figure 4-2A). Reaction with hydroxylamine, however, requires two molecules to attack, first at the C6 as with hydrazine and then at C4 to form the product shown (Figure 4-2B).[24] Both of these reactions progress faster at lower pH, indicating that they react mainly with protonated C.[24] Manipulation of these reactions to selectively damage protonated nucleobases would make effective reagents.

4.2.4 Sodium bisulfite

Most commonly, sodium bisulfite is used to identify 5-methyl-C in RNA systems.[13, 17, 25, 26] Because modification at the 5 position does not affect the Watson-Crick face, identification of this occurrence is challenging. Treatment of samples at low pH with sodium bisulfite results in immediate addition of a sulfite group to C's at C6, however reaction with 5-methyl-C progresses much slower and usually takes days to reach completion (Figure 4-2C).[27] Stopping the reaction when bisulfite treatment of C has reached completion but before 5-methyl-C has reacted allows for the selective treatment of C over 5-methyl-C. High pH treatment traps the reversible product and forms U. Sequencing of samples after bisulfite treatment reveals the transition of C to U for unmodified C's but remains a C for 5-methyl-C's. Since the bisulfite reaction begins with treatment of samples at low pH, causing protonation of C's, it has the potential to react with protonated nucleobases at neutrality and to identify pK_a-shifted bases in large RNA systems. However, adaptation of reaction conditions to preferentially react with protonated bases and trap the reaction intermediate is necessary for the exploitation of this method. Potentially, a two-step reaction mechanism, as shown in Figure 4-2D, can be employed to trap the bisulfite intermediate. A reagent such as hydroxylamine could be effective at reacting with the sulfite-substituted product to form a buckled, aliphatic ring system, which could be detected by reverse transcription. A two-step reaction with initial bisulfite reaction and subsequent hydroxylamine treatment would be preferred to treatment with hydroxylamine alone since substantially more is known about the reaction of bisulfite with nucleobases and it has more potential to preferentially react with protonated nucleobases. Although this mechanism is shown in Figure 4-2D with hydroxylamine, we reason that similar reagents with reactive amine groups would also be effective as a secondary reagent in the two-step bisulfite reaction.

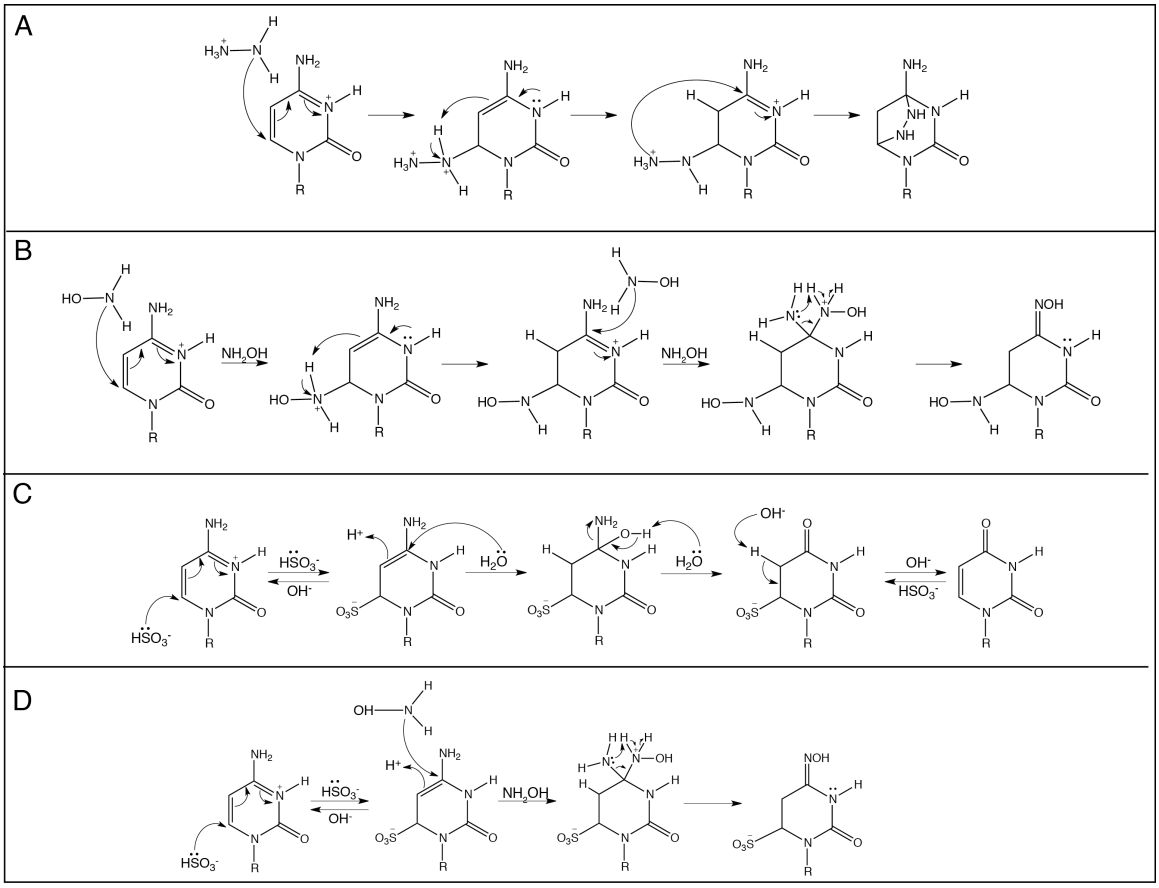


Figure **4-2**: Selected reactions with C^+. (A) Proposed mechanism of hydrazine reaction with C^+. Reaction requires an intramolecular attack of an amine after initial attack to form the hydrazine-substituted complex. (B) Proposed mechanism of hydroxylamine reaction with C+. This mechanism mirrors attack of hydrazine but requires an additional hydroxylamine molecule for the secondary attack of the hydroxylamine-substituted complex. (C) Mechanism of bisulfite sequencing.[25] Bisulfite is commonly used as a reagent to distinguish substituted C's in solution and begins with a low pH treatment of bisulfite, causing reaction of bisulfite with protonated C. This step is reversible and forms an unstable intermediate. Subsequent work-up at high pH leads to the formation of U. (D) Proposed mechanism of two-step bisulfite reaction. Because initial treatment with bisulfite produces a reversible intermediate, subsequent treatment with hydroxylamine reacts with C4, forming a stable aliphatic ring product.

In this study, efforts towards the development of a viable reagent for mapping protonated bases are presented. Various reagents and reaction conditions including sodium borohydride, sodium cyanoborohydride, hydroxylamine, ethylenediamine, and sodium bisulfite were explored as sources of nucleobase damage. Furthermore, gel electrophoresis and mass spectrometry are employed as analysis techniques. Chemical reagents have been tested on different RNA systems with known protonated bases of various biological importance. These systems include hepatitis delta virus (HDV) ribozyme (Figure 4-3A), which is a small, self-cleaving ribozyme with two protonated bases, C75 and C41: C75 has been proposed as the general acid in the cleavage mechanism of the ribozyme, and C41 participates in a base quartet and is required for proper folding of the ribozyme.[28] The constructs used in this study, however, were the cleaved 1/99 form of HDV, in which the pK_a of C75 is only slightly shifted (to ~5) but should not have been detectable in these experiments.[29] A mutant HDV, Δ3bp HDV, reduces stability in the highly structured P4 region of wild type HDV[30] and was used in a portion of the presented studies in this chapter (Figure 4-3B). The beet western yellows virus (BWYV) RNA (Figure 4-3D) is a viral RNA known for its usage of programmed ribosomal frameshifting through formation of a pseudoknot. A base quartet, similar to that of C41 in HDV,[28] is formed upon protonation of C8, which causes a kinetic barrier, and ribosomal pausing during translation.[31] The pK_a of C8 in the base quartet is 8.15, which was determined in Chapter 2 using fluorescence spectroscopy.[4] In addition to these biological RNA's, model oligonucleotide systems with common protonated base pair motifs (A^+•C and C^+•U) were also utilized in the efforts towards development of a method for mapping protonated bases (Figure 4-3C). Model oligonucleotides for the A^+•C motif were designed based on hairpins used in previous pK_a studies,[32] while those based on C^+•U were designed de novo.

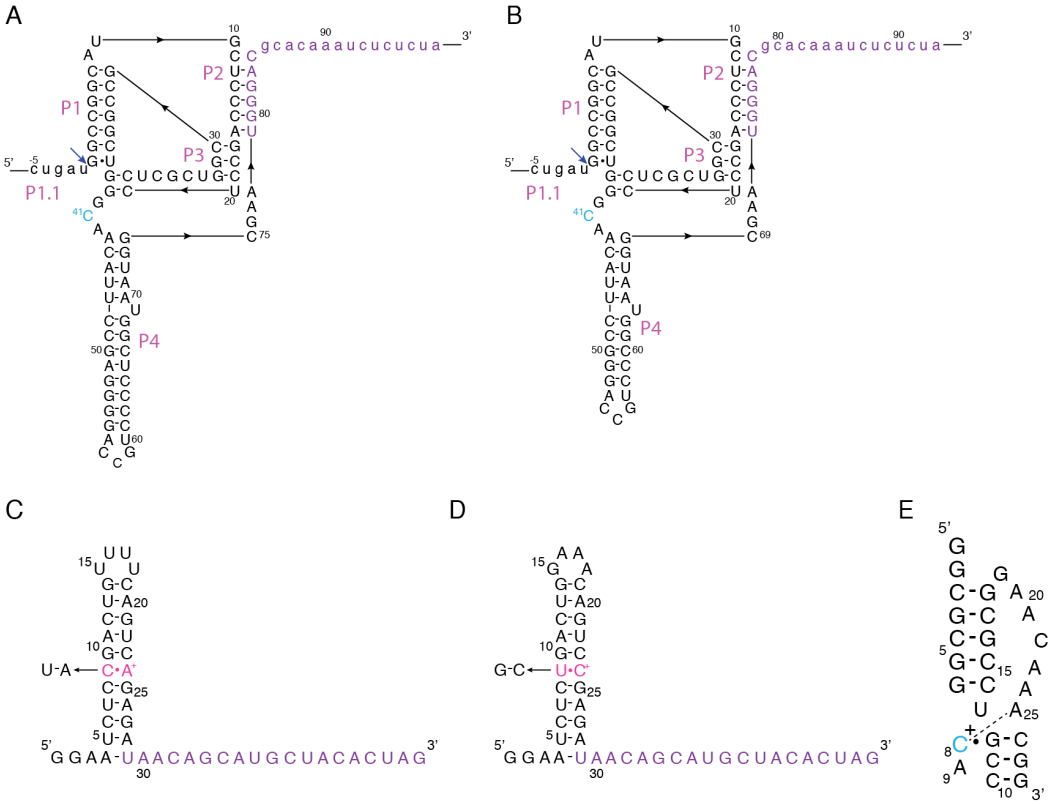


Figure **4-3**: Secondary structures of RNA's used throughout Chapter 4. (A) Hepatitis delta virus (HDV) ribozyme. Self-cleavage site indicated by blue arrow. Potentially protonated base C41 is shown in cyan. Primer binding site for RT shown in purple. (B) Mutant of hepatitis delta virus-Δ3bp HDV, which has three base pairs removed from the P4 region of the RNA. Numbering of the Δ3bp HDV differs from wild type HDV but the tertiary structure remains constant. C41 is potentially protonated and shown in cyan. (C) Model oligonucleotide system. Primer binding site shown in purple from 3' end. Protonated mismatch is located in the center of helix for stability. (D) Beet western yellows virus (BWYV) RNA. Protonation at C8 forms a base quartet with A25, C26 and G12 analogous to C41 in the HDV ribozyme.

4.3 Materials and Methods

4.3.1 Preparation of 1/99 HDV, model oligonucleotides, and BWYV RNAs

Plasmids were generated from pUC19 and purified using a Maxiprep (Qiagen). Isolated DNA was sequenced and then digested using *Bfa*I. The -54/99 construct was transcribed (Ambion), at which point the hepatitis delta virus (HDV) ribozyme underwent self-cleavage to form the 1/99 RNA construct, which was then gel purified to separate from the -54/-1 construct on a denaturing 8% polyacrylamide gel. For model oligonucleotides, DNAs were synthesized, deblocked and desalted by the manufacturer (Integrated DNA Technologies, Coralville, Iowa). Purity of oligonucleotides was tested by the manufacturer using electrospray ionization mass spectrometry (ESI-MS). Model DNA oligonucleotides were transcribed (Ambion) and gel purified. Secondary structures for constructs used in the study are provided in Figure 4-3. BWYV RNA was synthesized by the manufacturer (Dharmacon-Thermo Scientific) and deblocked according to the manufacturer's protocol. The RNA was dialyzed against 100 mM KCl for 8 h and H_2O for 16 h.

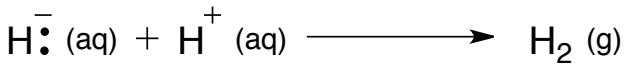
4.3.2 Modification with Chemical Reagents and Irradiation

Sodium borohydride reactions were modeled loosely after two protocols:

Warren[18]: $NaBH_4$ was prepared in a background of sodium hydroxide in various concentrations. 1 mM $NaBH_4$ in 0.03 mM NaOH, 5 mM $NaBH_4$ in 0.15 mM NaOH, and 10 mM $NaBH_4$ in 0.3 mM NaOH.

Kubarenko[19]: $NaBH_4$ was prepared in the absence of sodium hydroxide. Final concentrations were 5 mM-100 mM and varied for each experiment.

For both methods, solutions were made fresh before each addition to renatured buffered RNA solutions. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8). Buffering capabilities were confirmed in 5 mM $NaBH_4$ by spotting pH paper. Upon addition, reactions progressed for 30 min on ice as suggested in literature[18, 19] before performing RT as with sequencing samples. After initial attempts with one Warren method with a background of NaOH, the method presented by Kubarenko was primarily followed. When preparing sodium borohydride solutions, bubbles formed due to formation of H_2 gas. Bubbles continued to form as the sodium borohydride solution was added to RNA, indicating that the hydride was still present in the solution and was available to react with the RNA.



Scheme **4-1**: Reaction of hydride under acidic conditions. Because hydride reacts readily under acidic conditions, solutions were prepared fresh and quickly before addition to RNA.

For sodium cyanoborohydride reactions, solutions were prepared as with sodium borohydride reagent. Solutions were made fresh before each addition, and concentrations of $NaBH_3CN$ were comparable to that of $NaBH_4$.

Sodium bisulfite was prepared by first preparing a solution of sodium sulfite (EMD Millipore- minimum 98 % pure) and lowering the pH with glacial acetic acid from pH ~8.5 to form the bisulfite form at pH 5.5. Standard bisulfite sequencing requires treatment with 3.6 M sodium bisulfite at pH 5 but the protocol was modified to selectively damage protonated nucleobases.[17] Sodium sulfite was prepared in a 1.5 M stock solution and titrated to pH 5.5 with glacial acetic acid. It should be noted that significant difficulties persisted in preparing highly concentrated samples as there were troubles getting sodium sulfite to dissolve even though the solution was not approaching solubility limits. Concentrations ranged from 0.03 M to 1.5 M in reactions. As with hydride reagents, sodium bisulfite solutions were made fresh before addition to

renatured RNA. Unlike the hydride reactions, however, sodium bisulfite was pHed to the desired reaction, omitting usage of a buffer. Sodium bisulfite reactions were either performed in one- or two-step reactions. One-step reactions required treatment of samples with various concentrations of sodium bisulfite for 60 min at room temperature before performing RT. For two-step bisulfite reactions, RNA was treated with various concentrations of sodium bisulfite (30-120 mM final) and reacted for 30 min at room temperature, at which point the secondary reagent (sodium borohydride, aniline, hydroxylamine, ethylenediamine) was added and reacted for an additional 30 min at room temperature. Initial bisulfite treatment was carried out at ~pH 5.5 after titrating sodium bisulfite to pH 5.5 with glacial acetic acid. Secondary reactions were not buffered or titrated to pH 5.5 and the pH for these reactions was not determined. The bisulfite was not removed throughout the reaction and remained present during the secondary reaction.

Scheme **4-2**: Decomposition reaction of bisulfite. Because bisulfite readily breaks down to water and sulfur dioxide, sodium bisulfite is prepared fresh before addition to RNA.

For photohydration experiments, samples were placed directly on a Saran Wrap covered hand-held 254 nm UV-lamp and irradiated for various lengths of time (typically 1 to 10 min). A cardboard box was placed over the Saran Wrap-covered handheld UV lamp to block out all other light while the samples were being irradiated. Samples were buffered at pH 6, 7, or 8. Photohydration reactions were performed in conjunction with hydride addition to trap reversible photohydration products. In these experiments, samples were irradiated for 10’ and samples were treated with sodium borohydride (25 mM).

4.3.3 Reverse Transcription and Sequencing of 1/99 HDV and Model Systems

Primers were designed for RT. Because of difficulties performing dideoxy sequencing, primer length was varied to optimize primer binding and RT. DNA primers (synthesized, deblocked and desalted by manufacturer- Integrated DNA Technologies, Coralville, Iowa) were radiolabeled using γ ^{32}P-ATP and purified using a Sephadex G25 column. Primer sequences were as follows:

```
RTP1        5′ CTA GTG TAG CAT GCT GTT A
RTP1′       5′ CTA GTG TAG CAT GCT GTT ATC TC
RTPHDV      5′ TAG AGA GAT TTG TGC GTC CCA
RTPHDV52    5′ TAG AGA GAT TTG TGC GTC CCA TTC GCC ATT ACC GAG GGG
            ACG GUC CCC TCG G
```

RT primers RTP and RTP' were used with model oligonucleotides and RTPHDV and RTPHDV52 were used with HDV. RT was performed according to manufacturer's (New England Biolabs,) protocol using M-MuLV reverse transcriptase. An annealing mix (AM) 1/99 HDV (0.25 μM), 5' end labeled DNA primer (5 nM) and dNTP mix (1.25 mM each dNTP- pH 7.0 with Tris HCl) were heated at 75 °C for 3 min and cooled to room temperature (15 min on bench top). Regardless of reaction conditions, RNA was always renatured prior to sequencing. A buffer enzyme mix (BEM) was prepared consisting of 2.5X RT buffer and 2.5X M-Mulv reverse transcriptase. Five solutions, each containing 2 μL of AM, 2 μL of BEM, and 1 μL of ddNTP (Roche) or no ddNTP were heated for 60 min. Although reaction conditions were varied throughout experiments to determine optimal reaction conditions, standard conditions included RT at 42 °C and 0.5 mM each ddNTP. The reaction was quenched by adding 2X formamide loading buffer, heated to 90 °C for 2 min, and loaded on a denaturing polyacrylamide gel

immediately, at which point the position of stops with RT were visualized. Samples were fractionated using polyacrylamide gel electrophoresis with 8.4 M urea for ultra denaturing gel conditions.

In one effort to obtain optimal sequencing results, we attempted to degrade the RNA prior to PAGE. This was done by one of two ways after RT: either RNases were added or the RNA was boiled in alkali. RNases were used as a cocktail, which was made up of 1 μg/mL RNase A and 1 U/μL RNase T1. The RT reaction was stopped using 2 equivalents of EDTA (1 μL of 30 mM EDTA), heated at 85 °C for 5 min, then cooled to room temperature for 15 min on bench top, then added 1/5 the volume (1 μL) of the RNase cocktail, heated to 50 °C for 30 min, then added 7 μL of 2X formamide loading buffer, heated the samples at 85 °C for 3 min then loaded immediately on a gel. Alternatively, samples were boiled for 30 min in a large beaker of boiling water before loading samples immediately on a gel. To obtain better sequencing results, 7-deaza-dGTP (New England Biolabs) was used in one sequencing reaction since this can weaken self-structure of the DNA.[33] New dNTP mixes were made with 10 mM each dNTP replacing dGTP with 7-deaza-dGTP. Neither addition of RNases, boiling in alkali, or substitution of 7-deaza-dGTP improved sequencing and therefore were not used after initial attempts.

4.3.4 UV Monitored Thermal Denaturation (Melts)

RNA (~1.5 μM) was melted in the background of 10 mM bis-tris-propane buffer (pH 5.5 or 7.5; pH determined at room temperature) in 1 cm pathlength quartz cuvettes. Samples were renatured by heating to 90 °C for 2 min then cooled to room temperature for 15 min. UV monitored thermal denaturation experiments (melts) were conducted at 260 nm on a Gilford Response II spectrophotometer equipped with a temperature controlled microcuvette assembly.

Absorbances were collected from 5 to 95 °C in 0.5 °C increments. Data were analyzed using KaleidaGraph software. Eleven-point smoothing was done prior to taking first derivatives.

4.3.5 Analysis by ^{1}H NMR and UV-Vis

^{1}H NMR was performed on CMP before and after bisulfite treatment to confirm bisulfite addition. Cytidine 5'-monophosphate disodium salt was brought up in an unbuffered 10 mM stock solution in 100 % D_2O. The 10 mM stock was then treated with 1 M sodium bisulfite (pH ~4) for a final concentration of 5 M CMP and 500 mM sodium bisulfite. Samples were allowed to react at 4 °C for 2 hr. NMR spectra were collected at room temperature on a Bruker-Biospin Avance 360 with a 5mm quadranuclear probe. Absorbance spectra were collected of solutions with 100 µM CMP and 1 M sodium bisulfite pH 5 with glacial acetic acid using a Beckman Coulter DU 650 Spectrophotometer.

4.3.6 Analysis by Mass Spectrometry

For analysis by mass spectrometry, ammonium bisulfite was used as an alternative to sodium bisulfite, as ammonium is preferred for mass spectrometry. The ammonium sulfite solution (Sigma Aldrich) was titrated to pH 6.5 with glacial acetic acid. Ammonium bisulfite (1.5 M) or water control was added to renatured beet western yellows virus (BWYV) RNA samples and reacted at room temperature for 60 min at pH 6.5. I developed a new way to prepare RNA samples for mass spectrometry that worked very well using a two-step dialysis procedure. Samples were dialyzed for 8 hr against ammonium acetate (2.5 M) and di-H_2O for 36 hr and submitted for analysis. Negative ion mass spectra of RNA samples were acquired on a Thermo LTQ Orbitrap Velos mass spectrometer. It should be noted that, for treatment of model

oligonucleotides and HDV, sodium bisulfite was titrated to pH 5.5 whereas ammonium sulfite was pHed to 6.5 for treatment of BWYV RNA. This was due to the known pK_a of BWYV RNA for ammonium bisulfite treatment of 8.15. In samples with pK_a's around neutrality, sodium bisulfite was pHed to 5.5 to ensure complete protonation.

4.4 Results

The first portion of the Results section discusses efforts to optimize RT and Sanger dideoxy sequencing for the 1/99 construct of HDV and model oligonucleotides. Problems remained throughout the duration of this chapter with sequencing. The second portion discusses efforts to chemically damage protonated bases in the 1/99 HDV construct or the model oligonucleotides using sodium borohydride, sodium cyanoborohydride, photohydration, sodium bisulfite, and two-step reactions of sodium bisulfite with sodium borohydride, aniline, hydroxylamine, and ethylenediamine. Through the utilization of NMR and UV-Vis spectroscopies, efforts were made towards characterization of the bisulfite treated product and are discussed in the third section of this chapter. Lastly, the fourth section discusses efforts made using mass spectrometry to analyze samples after ammonium bisulfite treatment.

4.4.1 Determination of Optimal Sequencing Conditions

As with most assays, a significant challenge is varying assay conditions in order for it to be a viable method of detection. Because of this, efforts were made in altering RT, sequencing, and gel electrophoresis conditions in an effort to obtain clear sequencing of HDV RNA and model RNA oligonucleotides. In each experiment, one condition was optimized to determine how it affected sequencing quality. Unless otherwise noted, standard experimental conditions included

RT at 42 °C with 0.2 mM ddNTP concentration and unmodified dNTPs, heating of samples to 90 °C for 2 minutes before loading them on an 8.4 M urea gel and running of gels at 100 W.

To determine the optimal temperature for RT using M-MuLV reverse transcriptase, RT was performed at 37 °C, 42 °C and 55 °C (Figure 4-4). The temperature of RT recommended by the manufacturer was 42 °C. Performing RT at 55 °C resulted in a large amount of undesired stops as evident from the dark bands observed in the blank control sequencing lane. Sequencing at 37 °C resulted in lighter bands overall, which made sequencing difficult. Sequencing at 42 °C produced the clearest sequencing with both the least undesired stops and most distinguishable sequencing results. Carrying out RT at 42 °C may be able to melt out some structure, improving sequencing from the sequencing performed at 37 °C. RT at 55 °C, however, may encourage dissociation of the primer-RNA complex, resulting in premature stops and a high background with sequencing. For further experiments, RT was performed at 42 °C.

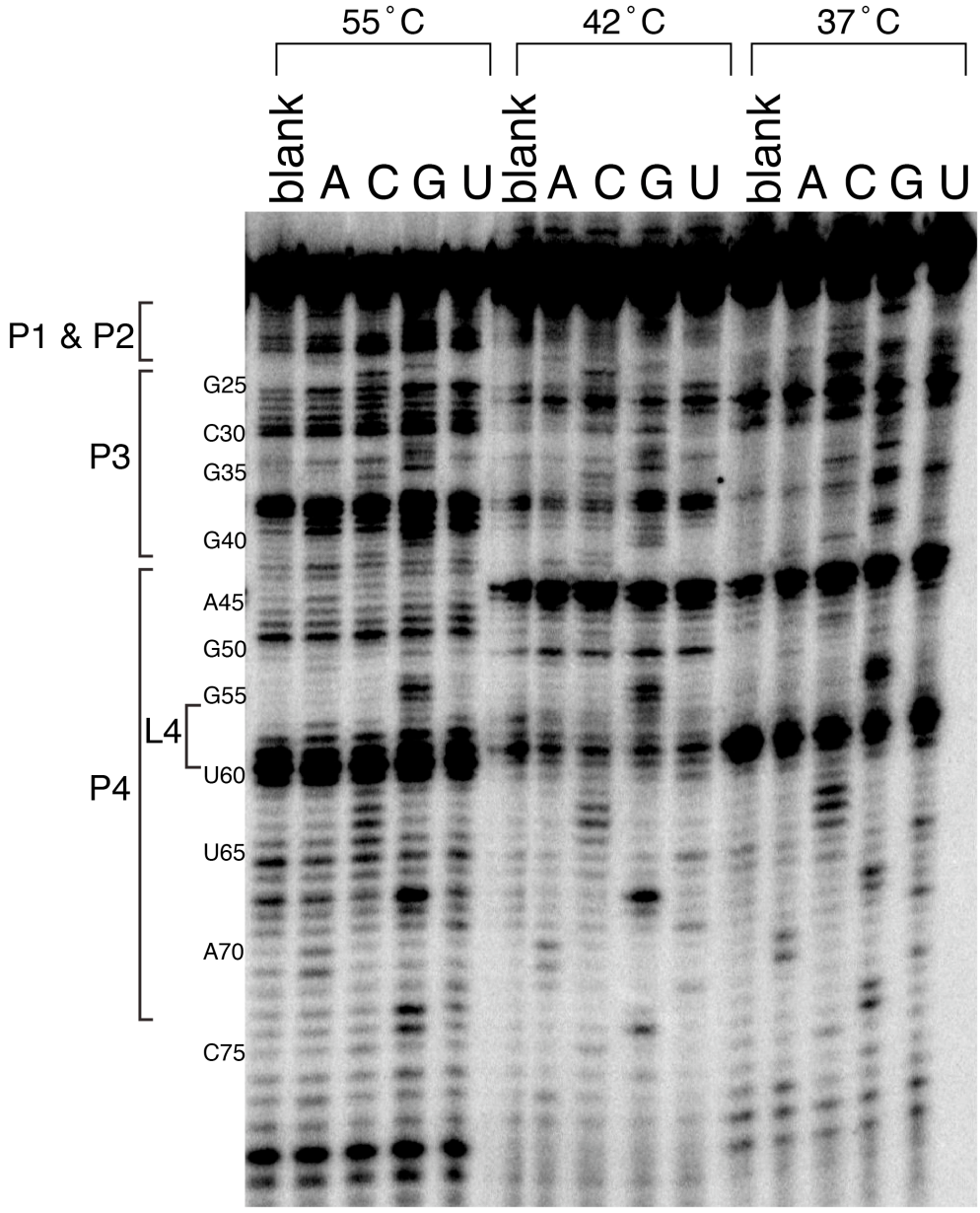


Figure **4-4**: Sanger dideoxy sequencing of HDV. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. A blank with 0 mM ddNTP was run to control for structural stops. RT was performed at 37 °C, 42 °C and 55 °C for 60 min to determine the optimal temperature for RT. Before loading on a gel, samples were heated to 90 °C for 2 min. The gel was run at 100 W. (7-23-08)

For efficient sequencing, the ddNTP concentration needs to be high enough to cause stops during RT but low enough to ensure a large percentage of full-length primer is produced. With optimal ddNTP concentrations, sharp, clear bands with full extension of primer should be observed. Concentrations of ddNTP's were varied (Figure 4-5) from 0.1 to 0.4 mM ddNTP. Increasing the ddNTP concentration to 0.2 mM resulted in the clearest sequencing with sharp bands and low background. Sequencing at 0.1 mM ddNTP was slightly inefficient and resulted in somewhat light, fuzzy bands. Sequencing at 0.4 mM, however, was also unsuccessful, as a slightly high background was observed. Sequencing with 0.2 mM facilitated an optimal number of stops during RT and eliminated undesired stops. For further experiments, 0.2 mM ddNTP concentration was used.

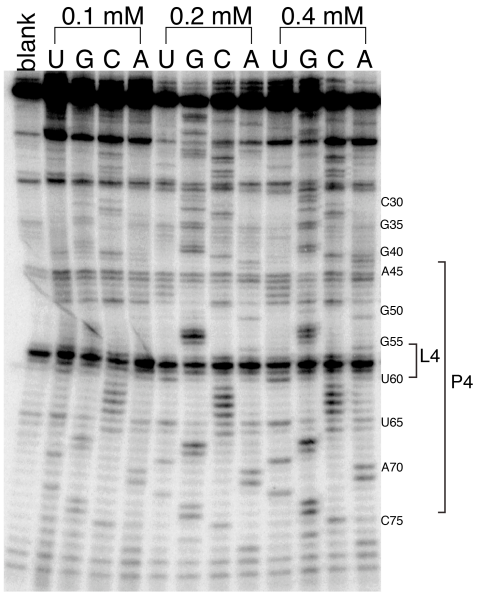


Figure 4-**5**: Determination of optimal ddNTP concentration for RT of HDV. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, and 1.25 mM each dNTP. 0.1 mM, 0.2 mM and 0.4 mM ddNTP concentrations were used to optimize RNA sequencing. A blank with 0 mM ddNTP was run to control for structural stops. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run at 100 W. (8-1-08).

After altering the temperature and ddNTP concentration, sequencing was still troublesome. One possible explanation to why sequencing remained fuzzy was incomplete dissociation of the RNA-DNA primer complex. If RNA was still interacting with the newly extended primer during PAGE, achieving sharp bands for sequencing would be difficult. To test this hypothesis, RNase T1 and A were added in an RNase cocktail to sequencing solutions following RT (Figure 4-6). Addition of RNases would degrade remaining RNA after completion of RT and eliminate interaction with extended primer. Treatment with RNases, however, had little effect. Bands remained fuzzy and sequencing was still problematic. It was concluded that sequencing difficulties were not due to RNA DNA interactions and RNases were not used in further experiments.

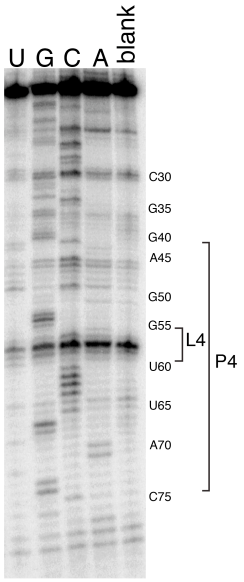


Figure **4-6**: Sequencing HDV with RNase treatment. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.2 mM ddNTP. A blank with 0 mM ddNTP was run to control for structural stops. RT was performed at 42 °C for 60 minutes. RNase T1 and A were added to sequencing solutions after RT to degrade RNA and eliminate sequencing difficulties due to RNA-DNA primer complex formation. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run at 100 W. (8-1-08)

7-deaza-dGTP is commonly used in PCR amplification with constructs with high GC content to disrupt undesired interactions between the nucleobases.[34] Addition of 7-deaza-dGTP can result in clearer RT and lower background. To determine if 7-deaza-dGTP could aid in sequencing of HDV, dNTP mixes were prepared with 7-deaza-dGTP substituting completely for dGTP (Figure 4-7). Substitution of 7-deaza-dGTP decreased sequencing quality, causing bands to smear. However, the blank control sequencing lane had minimal background stops, suggesting that 7-deaza-dGTP was successful in improving RT quality. For future experiments, 7-deaza-dGTP was not used in dNTP mix.

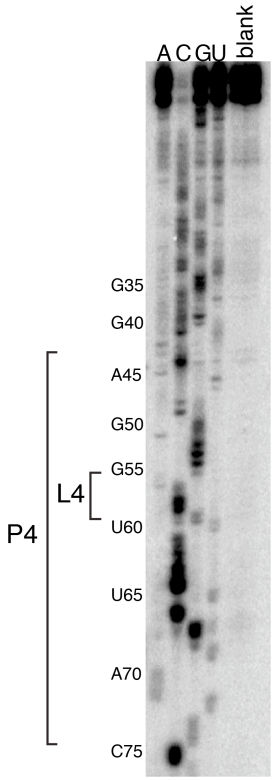


Figure **4-7**: Sequencing with 7-deaza-dGTP. Sequencing HDV with RNase treatment. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, and 0.2 mM ddNTP. dNTP mixes were substituted with 7-deaza-dGTP (1.25 mM each dNTP) as a potential method to improve sequencing quality. A blank with 0 mM ddNTP was run to control for structural stops. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run at 100 W. (1-28-09)

After varying reaction conditions to improve the quality of RT and sequencing, we reasoned that difficulties were most likely originating from the highly structured HDV constructs used in sequencing experiments. Since HDV is so structured, especially in the GC rich P4 region, poor sequencing may have been due to incomplete denaturation during either RT or gel electrophoresis. Further efforts were focused on improvement of the quality of sequencing by optimizing denaturation conditions.

To ensure complete denaturation while performing gel electrophoresis, the power of the gel was increased and a Plexiglas shield was positioned in front of the glass plates to act as an insulator. Although not attached to the gel apparatus, the shield was angled so that it was lying flat, pressing directly up against the outer glass plate used in PAGE. For pre-running of the gel, the power was increased incrementally; from 100, 105 then 110 watts. During this time, the temperature of the outer gel plate reached 55 °C, as determined using a thermo-couple. The power was then turned up to 125 and finally 150 watts. Samples were loaded when the temperature of the outer glass plate was 65 °C. The power was adjusted throughout the running period to maintain a constant temperature of 62 °C. When other lab members attempted this increase in power their gel plates cracked, so this variation was attempted with extreme caution. My plates, however, did not crack and increasing the temperature of gel running increased quality of sequencing (Figure 4-8). It was possible to sequence up to C15, however, sharp bands were still unattainable. Although increasing the temperature of PAGE did not result in sharp sequencing bands, reduction of background in sequencing suggests that some incomplete denaturation was occurring during gel electrophoresis. Further experiments were run at 100 W with a Plexiglas shield for insulation. Using a Plexiglas shield without increasing the power insulated the gel and increased the temperature without the risk of cracking of plates and loss of data.

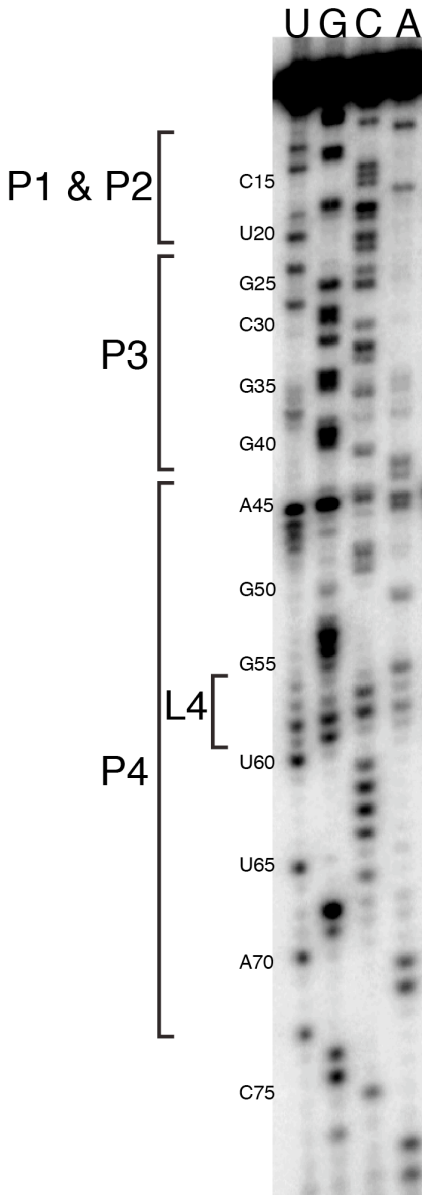


Figure **4-8**: Sequencing of HDV with increased power during PAGE. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.2 mM ddNTP. A blank with 0 mM ddNTP was run to control for structural stops. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run at 150 W and a Plexiglas shield was used for insulation. The outer gel plate remained at 62 °C while the gel was running (8-27-09).

In addition to incomplete denaturation during gel electrophoresis, sequencing difficulties could have originated from incomplete denaturation of HDV during RT. Because of this, we explored methods of destabilizing structure formation beginning with a mutant HDV oligonucleotide. Laurie Heinicke (Bevilacqua lab) previously used an HDV mutant with a deletion of three base pairs in the P4 region (Δ3bpHDV) (Figure 4-3B). Because most of the problems with sequencing were thought to be a result of structural stops in the P4 region, we hoped that reducing the stability of this region would increase sequencing quality. Deletion of three base pairs in P4 allowed for better sequencing in the P4 region, with significant reduction of structural stops (Figure 4-9). It did not, however, increase sharpness of bands other than those in the P4 region, so further attempts to increase sequencing quality were investigated. Since slight improvements were observed with the Δ3bpHDV construct, it was used in conjunction with wild type HDV in some future studies.

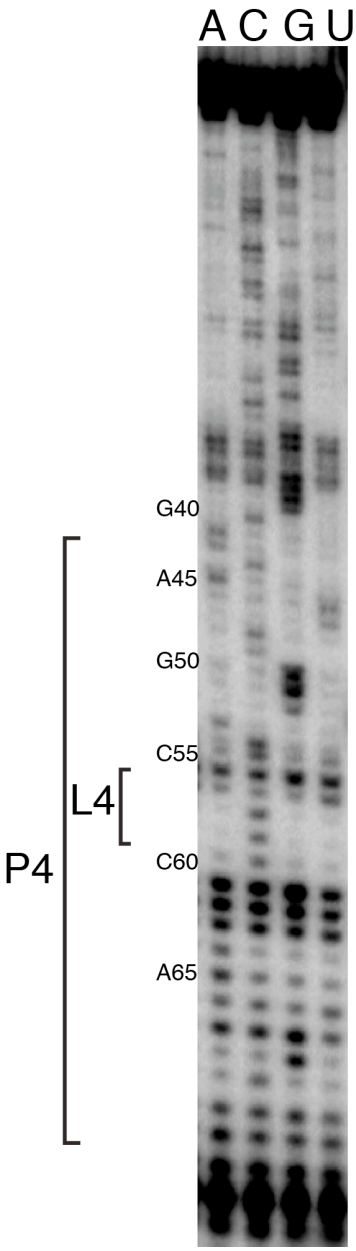


Figure **4-9**: Sequencing of Δ3bpHDV, which removes three base pairs in the P4 region of HDV. This construct was pursued as a potential solution to incomplete denaturation of WT HDV. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 μM RNA, 0.02 μM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run at 100 W and a Plexiglas shield was used for insulation. (9-3-09).

All sequencing gels were run with 8.4 M urea to ensure maximum denaturation conditions. To further optimize denaturation conditions, the concentration of urea was increased to 10 M and gel stocks were prepared fresh as to prevent any loss of denaturation activity after preparation. In addition to increasing the urea concentration, the gel was pre-run for 4 hours to guarantee high gel temperature when samples were loaded. After pre-running the gel, it seemed to melt and wells combined. The gel was unusable and further attempts to optimize denaturation of samples were investigated.

In previous attempts to improve denaturation, all samples were heated to 90 °C for two minutes prior to loading them on a gel. To make certain that the DNA was fully denatured, in one instance samples were boiled for 30 minutes prior to gel loading at pH 8.3 in M-MuLV buffer. In addition to ensuring complete denaturation of DNA, boiling samples would also degrade any remaining RNA in samples. The rationale for degradation of RNA by boiling was similar to RNase treatment, eliminating the potential for smearing of gel bands due to DNA-RNA interactions, but the thought was that base denaturation might be more effective. However, examining the difference in sequencing lanes between those heated under standard conditions and those boiled for 30 minutes (Figure 4-10) showed no improvement in boiled samples. Boiling of samples actually increased background in sequencing lanes and caused fuzzier sequencing bands. Samples were heated at 90 °C for 2 minutes before loading in future experiments.

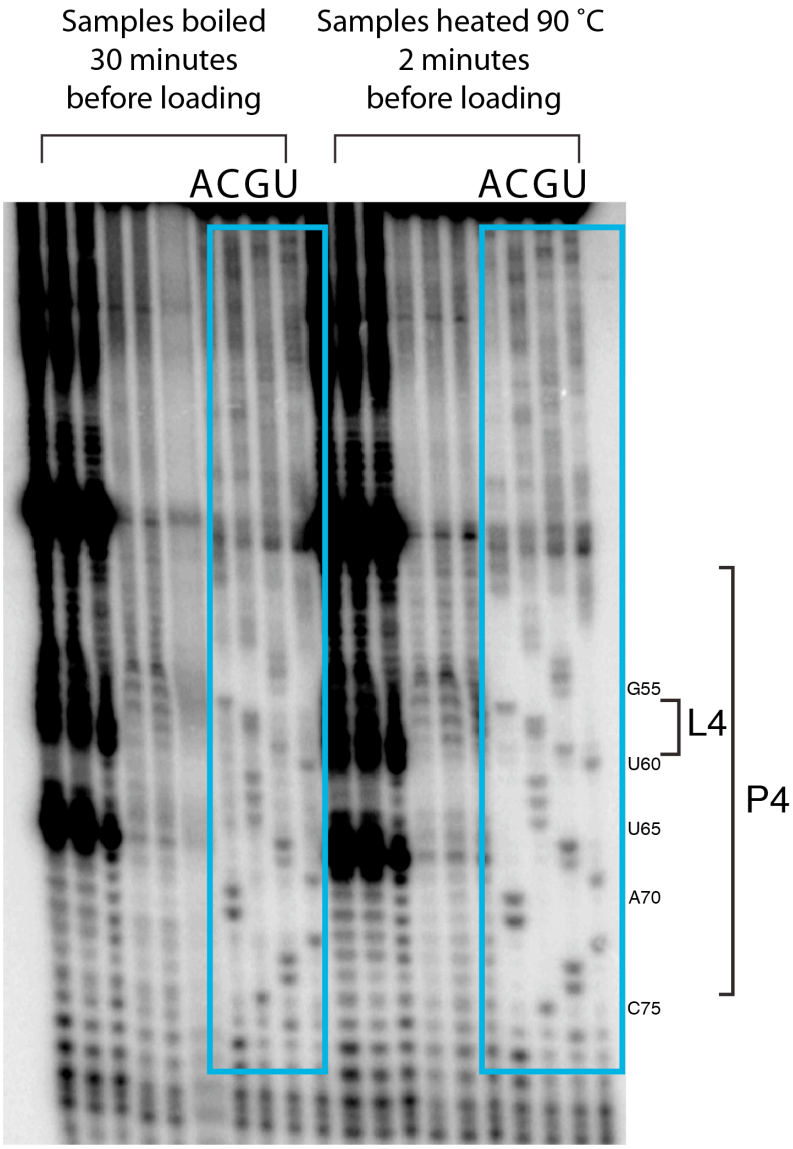


Figure **4-10**: Sequencing of HDV after boiling for 30 minutes. This construct was pursued as a potential solution to incomplete denaturation of WT HDV. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 μM RNA, 0.02 μM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Prior to loading samples for gel electrophoresis, samples were either boiled for 30 minutes at pH 8.3 or heated to 90 °C for 2 minutes. Gel was run at 100 W and a Plexiglas shield was used for insulation. Sequencing lanes are outlined in blue. (10-8-09)

Further efforts were made towards improving sequencing by reducing challenges with sequencing P4. A longer RT primer (RTPHDV52) was used in efforts to alleviate structural challenges in the highly stable P4 region of HDV. This 52-nucleotide primer extends through the P4 region when annealed to U47. Because it extends beyond C75, a limitation for using this longer primer was the reduction of information that can be obtained with RT. However, the longer primer still has the potential to report on protonation at C41. Additionally, kinase of this primer required gel purification because it was significantly larger than the 20 nt primer previously used. Utilization of a longer primer almost completely eliminated sequencing quality and visualization of distinct sequencing bands was impossible (Figure 4-11). Increasing the length of the primer could have decreased its ability to anneal to the RNA, hindering sequencing.

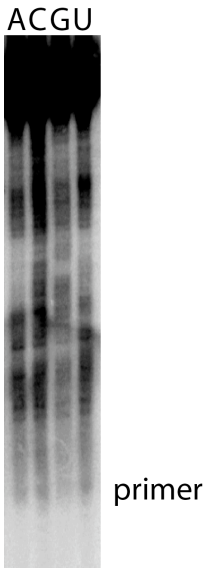


Figure **4-11**: Sequencing of HDV with 52-nucleotide primer. This primer extended to U47 and was used in an attempt to increase sequencing through the RNA upstream of the P4 region. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 μM RNA, 0.02 μM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run at 100 W and a Plexiglas shield was used for insulation. (10-08-09)

One contributing factor for fuzzy bands during sequencing could have been inconsistent migration of molecules of the same length DNA. Denaturing gel electrophoresis is a method for separation based on size; however, proper separation requires an initial thin, horizontal band of sample prior to separation. This would ensure complete separation by size. One could imagine a molecule of a larger size, originating in the bottom of the well, traveling with a molecule of a slightly smaller size that originated at the top of a well. This would result in smeary bands containing molecules of various sized RNAs. If these molecules were located next to each other horizontally rather than stacked vertically prior to gel electrophoresis, the samples would be completely separated.

Previous experiments were performed using a comb with small wells to allow for use of small sample volumes with gel electrophoresis loading samples of ~2 μL. To ensure a thin, sharp band for sequencing, we increased the size of the comb to increase the well size. We used combs with 8 × 5-mm wells and loaded between 3 and 5 μL of sample to allow for a think sample band to form in a wider well during PAGE. Loading small volumes of sample enables the sample to travel consistently horizontally throughout the gel and decreases smearing. Sequencing samples using a larger comb gave sharp, distinguishable bands (Figure 4-12 left). Sequencing was easily achievable to G25, although some difficulties remained when sequencing the region near 41, most likely due to RNA fold-back after primer extension of P4.

Because of the significant improvement of sequencing clarity with larger wells, gels were run with 8 × 5-mm wells in future experiments. The extent to which clear sequencing was obtained could not be reproduced, however, and efforts continued to optimize sequencing (Figure 4-12 right).

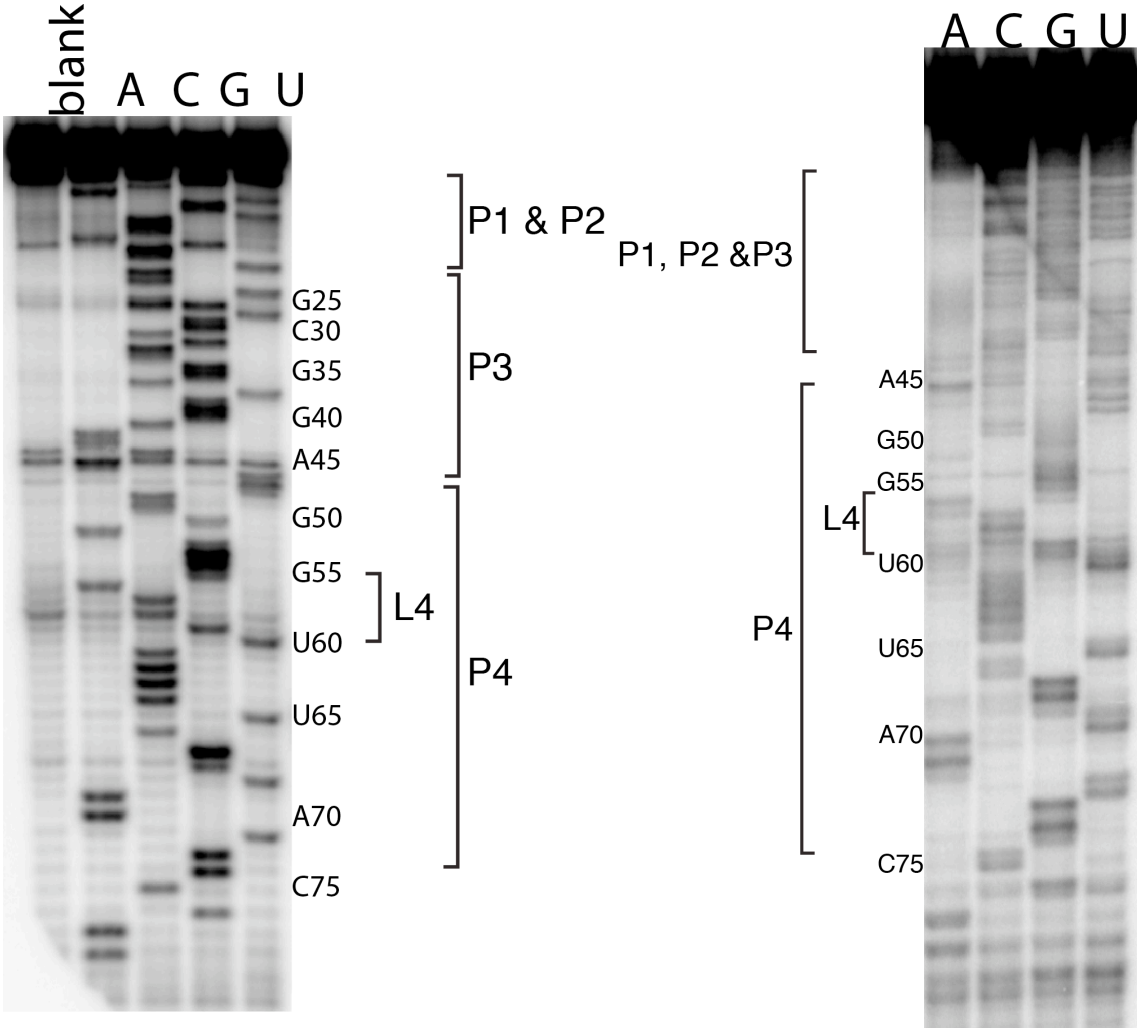


Figure **4-12**: Sequencing of HDV with larger wells. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 μM RNA, 0.02 μM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Clear sequencing was obtained after using a comb with larger wells but loading small sample volumes (left) but could not be repeated (right). (10-23-09 left; 10-31-09 right)

Optimal visualization of bands in gel electrophoresis requires quick and efficient drying of the gel after electrophoresis. Inadequate gel drying could lead to diffusion of bands and make obtaining clear sequencing bands impossible. To ensure that gels were being sufficiently dried, settings on the gel dryer were optimized. All experiments were performed using an oil vacuum pump. Pressure of the vacuum pump was determined to be 75 Torr, which was close to the expected value of 100 Torr. A thermo-couple was used to test the temperature on the dryer to confirm that the gel was being accurately heated while drying. The temperature increased to maintain a constant drying temperature of 80 °C. An additional cool-down period was added to the standard drying procedure. The heat on the gel dryer was turned off and the gel was allowed to dry with vacuum pressure in the absence of heat for 20 minutes. Sequencing bands were moderately sharper (Figure 4-13) but problems remained throughout sequencing.

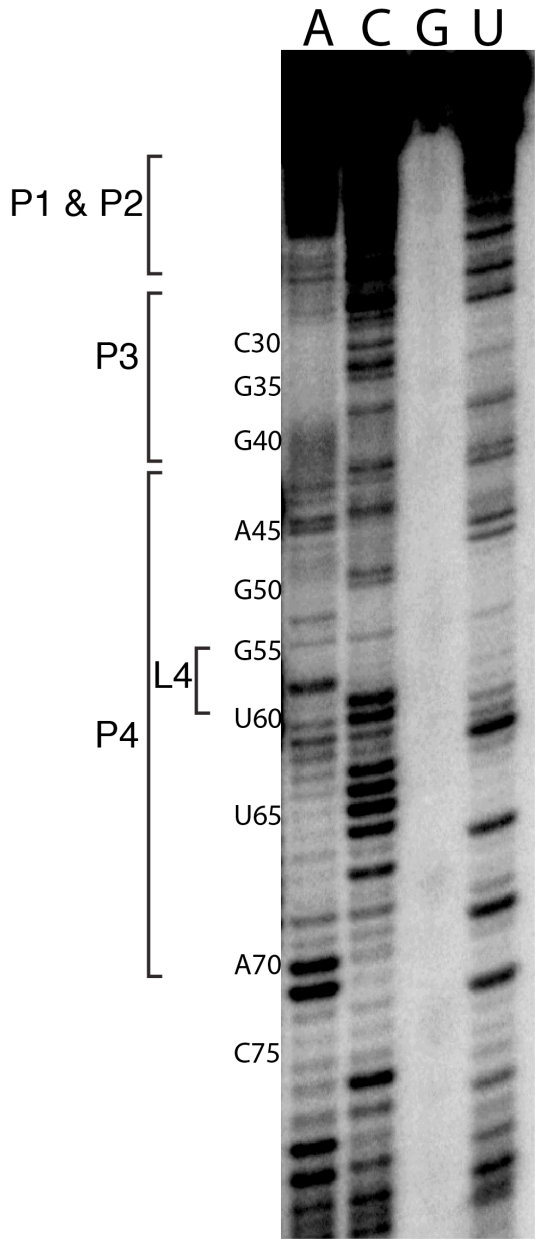


Figure **4-13**: Effect of increased gel drying temperature and addition of cool-down period on sequencing of HDV. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Gel was dried at high temperature with a 20 minute cool down period.

After difficulties encountered in attempting to optimize RT and sequencing of HDV, efforts were made towards sequencing smaller, less structurally complex model oligonucleotides. Focus was moved towards sequencing model hairpin oligonucleotides with an A^{+}•C wobble in the middle of the stem or a non-ionizable A-U control base pair (Figure 4-3). Sequencing was first performed with primers of various lengths to determine optimal reaction conditions for sequencing. Primer RTP1, a 19 nucleotide DNA, extended only through the designed primer-binding site, to U28. Primer RTP1' was longer and extended to G25. This primer was designed to be too long, however, as chemical damage at A24 would be evident by a stop at G25. Theoretically, damage at A24 would still prohibit primer extension at G25 with RTP1', however, complications typically arise the sequencing of nucleotides immediately upstream of the primer binding site and analysis in this region should be avoided. Sequencing of model oligonucleotides resulted in clear, sharp, distinguishable bands (Figure 4-14). Similar to complexities of sequencing P4 in HDV, difficulties arose after sequencing past the loop region, which was most likely due to RNA fold-back after primer extension of the hairpin's 3' strand, but because the area of interest with protonation was located in the 3' strand of the hairpin, the constructs designed were suitable for analysis by PAGE. Additionally, model oligonucleotides were designed with a C^{+}•U and non-ionizable G-C control base pairs, which are discussed in Section 4-#.

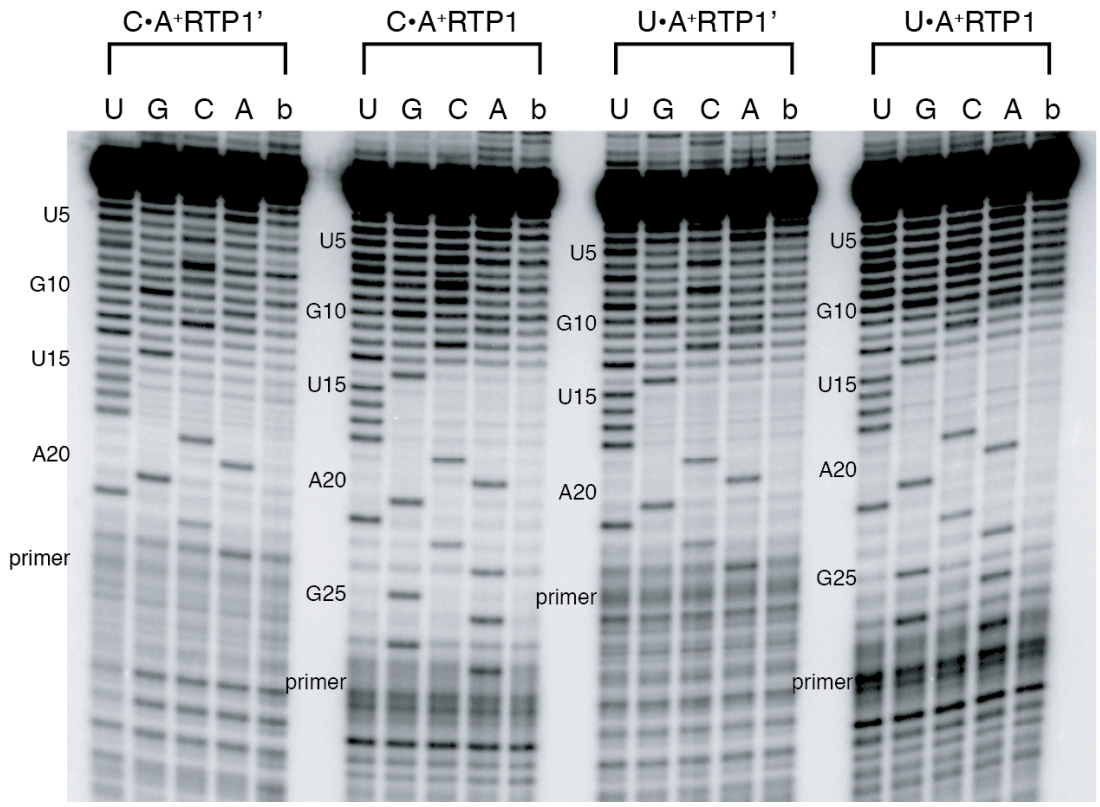


Figure **4-14**: Sequencing of model oligonucleotides. Secondary structures of model oligonucleotides are in Figure 4-3. Primer length was varied to obtain optimal sequencing. RTP1 is a 19 nucleotide DNA, extended to U28. RTP1' is longer, a 23 nucleotide primer and extended to G25. C^{+}•A oligonucleotides contain a C^{+}•A wobble between C9 and A26. U-A oligonucleotides contain a non-ionizable control U-A base pair substitution for C^{+}•A. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. (11-13-09)

Table **4-1**: Summary of efforts towards optimization of conditions for RT

Optimization	Result	Conditions for Future Experiments
Varying RT temperature	Clearest sequencing was obtained with RT at 42 °C	RT was performed at 42 °C
Varying ddNTP concentration	Best sequencing was observed with 0.2 mM each ddNTP	Sequencing was performed with 0.2 mM ddTP's
Treatment of RNA with RNases after RT	RNase treatment increased background	RNase treatment was not used
Replacement of dGTP with 7-deaza-dGTP	No 7-deaza-dGTP	7-deaza-dGTP was not used
Increasing power and temperature of gel to increase degree of denaturation	Moderate improvement	Gels were run at 100 W with a Plexiglas plate for insulation
Replacement of HDV with Δ3bp HDV construct	Better sequencing throughout P4 but still difficult throughout the oligonucleotide	Both WT HDV and Δ3bp HDV were used
Increase in urea from 8.4 to 10 M and prerunning time to 4 h	Gel melted	Urea concentration was 8.4 M and prerunning time was 2 h
Boiling samples for 30 min before loading them on the gel	Increased background	Samples were heated to 90 °C for 2 min before loading
Replacement of primer with 52-mer RTPHDV52	Sequencing quality was decreased-possibly because the primer was unable to anneal	52-mer was not used
Comb with larger wells	Nice, clear sequencing	A comb with larger wells was used
Increase temperature and vacuum pressure of gel dryer	Moderate improvement	Temperature of the gel dryer and vacuum pressure was increased
Use of model oligonucleotides	Clear sequencing was obtained with difficulties sequencing the 5' strand of the model oligonucleotides	Model oligonucleotides were used

4.4.2 Efforts towards chemical damage of RNA by various reagents

While optimizing conditions for RT, efforts were made towards determination of suitable reagent for damaging protonated bases. The goal of these experiments was to preferentially damage a protonated nucleobase using a selected reagent, prohibiting base pairing of the damaged nucleobases and primer during RT and visualization of the shortened product by gel electrophoresis. A stop during RT would be observed one nucleotide downstream of the protonation site using gel electrophoresis. In HDV, this stop would be located at A42, one nucleotide downstream of C^+41. Since the cleaved 1/99 HDV constructs were used in these experiments, protonation at C75 was not expected.[29] In model oligonucleotides, protonation at either A24 in A^+•C constructs or C24 in C^+•U constructs would result in a stop at G25. Control oligonucleotides with non-ionizable U-A and C-G base pairs were used as confirmation of specific reaction due to protonation.

Reagent experiments either required the use of buffered solutions, in which we expected to see higher reactivity at lower pH due to protonation of selected nucleobases, or with a pHed reagent in the absence of buffer, in which we expected to observe a higher reactivity with addition of reagent when compared to a control sample with no addition of reagent. Additionally, since some photohydration or bisulfite intermediates are reversible, a two-step method utilizing a primary reagent to damage the protonated base with a secondary reagent was used to trap reversible, unstable intermediates was used. For all analysis methods, Sanger (dideoxy) sequencing was utilized to identify locations of RT stops. Since experiments for the development of a chemical reagent to damage protonated bases were carried out simultaneously with optimization of sequencing experiments, some experimental conditions were not optimized

throughout reagent investigation. This portion of the Results section is categorized by the reagent being tested.

4.4.2.1 Sodium Borohydride Treatment of HDV

Chemical damage experiments began with treatment with sodium borohydride because of the known ability for sodium borohydride to reduce a Schiff base. Concentrations for borohydride treatments were initially based on literature concentrations.[18, 19] Per Warren *et al*, reactions were performed in the background of sodium hydroxide.[18] Since low pH samples contain protonated nucleobases and are more reactive than high pH samples, low pH samples treated with sodium borohydride resulted in darker bands than high pH samples as expected (Figure 4-15), however we expected bands to be localized A42, one nucleotide downstream of the proposed protonation site of C41. Stops were seen at terminal base pairs of looped region (Figure 4-16, pink boxes), which is not uncommon as nucleotides such as A43 and G55 have more single stranded nature because of their location at the end of a helix. Surprisingly, the majority of stops observed indicate damage of base paired nucleotides (Figure 4-16, green boxes), even those base paired in the highly structured P4 region of HDV. Higher reactivity at nucleotides C49, G52, and C62 suggests some flexibility of the helices at these locations to enable reactivity with hydride. Bands corresponding to G35, C36, A43, C44, C66 and G67 were observed throughout the treatment conditions and are inferred as structural stops. In future experiments, 0 mM control samples were run to account for stops due complications with RT. In addition, higher background damage was observed with higher concentration of sodium borohydride. No increase in quality of results was observed with the addition of sodium hydroxide so it was eliminated for further experiments.

Based on these data, it was concluded that sodium borohydride was too reactive of a reagent to be utilized. An excess of sodium borohydride caused non-specific damage of

nucleobases, resulting in a higher background with higher concentrations of sodium borohydride. The potential for hydride reactivity with H^+ led to hesitation on lowering the concentration of sodium borohydride. At this point, the focus was switched to a more selective, less reactive hydride source with the use of sodium cyanoborohydride.

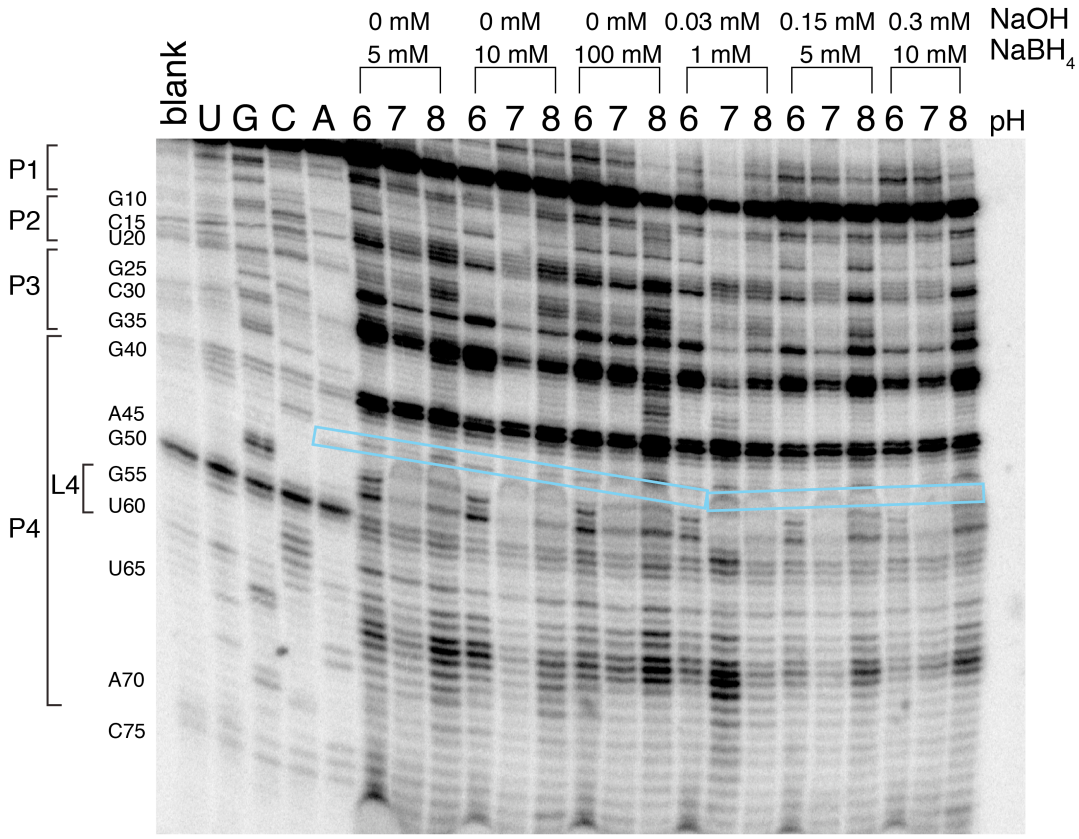


Figure **4-15**: Treatment of HDV RNA with sodium borohydride. Samples were treated with sodium borohydride in the absence[19] or the presence of sodium hydroxide,[18] as per literature protocols, on ice for one hour at which point RT was performed. . Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes, then cooled to room temperature for 15 minutes for renaturation. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at C41 would result in increase in intensity at A42 (blue box). Reaction at C41 would result in increase in intensity at A42 (blue box). (8-13-08)

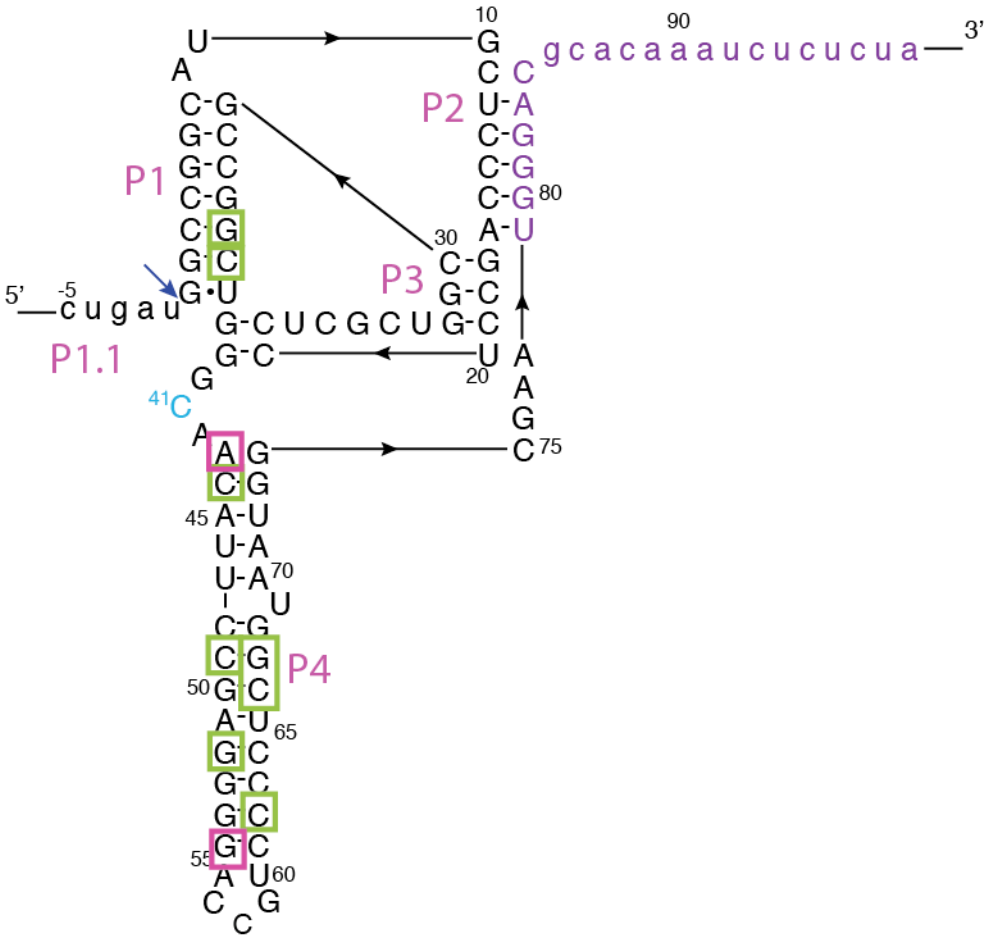


Figure **4-16**: Treatment of HDV RNA with sodium borohydride mapped to HDV secondary structure. Stops were observed in the gel in Figure 4-15 corresponding to the boxed region in the HDV secondary structure. Green boxes correspond to double stranded damage at nucleotides in the base paired regions of reactivity and pink boxes correspond to damage at nucleotides in the terminal base pairs of helices. No reactivity was observed for C41, the protonated base in HDV, shown in blue. (8-13-08)

4.4.2.2 Sodium Cyanoborohydride Treatment of HDV

Because of the high reactivity of sodium borohydride and the higher background damage observed in samples treated with a higher concentration of sodium borohydride, the possibility remained that the hydride could be too reactive to selectively damage protonated nucleobases. To eliminate this possibility, we employed sodium cyanoborohydride, a less reactive, and more selective hydride source (Figure 4-17). Interestingly, a bell shaped reactivity dependence on pH was observed in both samples treated with sodium borohydride and sodium cyanoborohydride, indicating that samples were more reactive at pH 7 rather than pH 6 as originally expected. Specifically, stops corresponding to U23, C24 and G25 were darkest at pH 7 after treatment with sodium borohydride and sodium cyanoborohydride (Figures 4-17 and 4-18, green boxes). This bell shaped reactivity was not observed in samples treated with sodium borohydride previously (Figure 4-15). As hydride is consumed under acidic conditions (Scheme 4-1), it could have been reacted at pH 6, leaving less hydride to react with RNA, whereas at pH 7, less hydride has been consumed by acid, allowing a higher reaction with RNA, resulting in higher intensity stops in pH 7 samples. At pH 8, there would have been less protonated RNA to react causing the rate to go back down. In nucleotides A43, C44 and A45, consistently high reactivity was observed (Figures 4-17 and 4-18, pink boxes). These nucleotides correspond to the terminus of P4 and are most likely due to complex structure formation and complications of RT. C61 and G59 (Figure 4-18, blue boxes) showed a high reactivity at pH 7 with sodium cyanoborohydride and at pH 8 with sodium borohydride (Figure 4-17). As expected, darker bands were observed with sodium borohydride than sodium cyanoborohydride because, as a less reactive hydride source, sodium cyanoborohydride is more selective in its reaction with nucleobases.

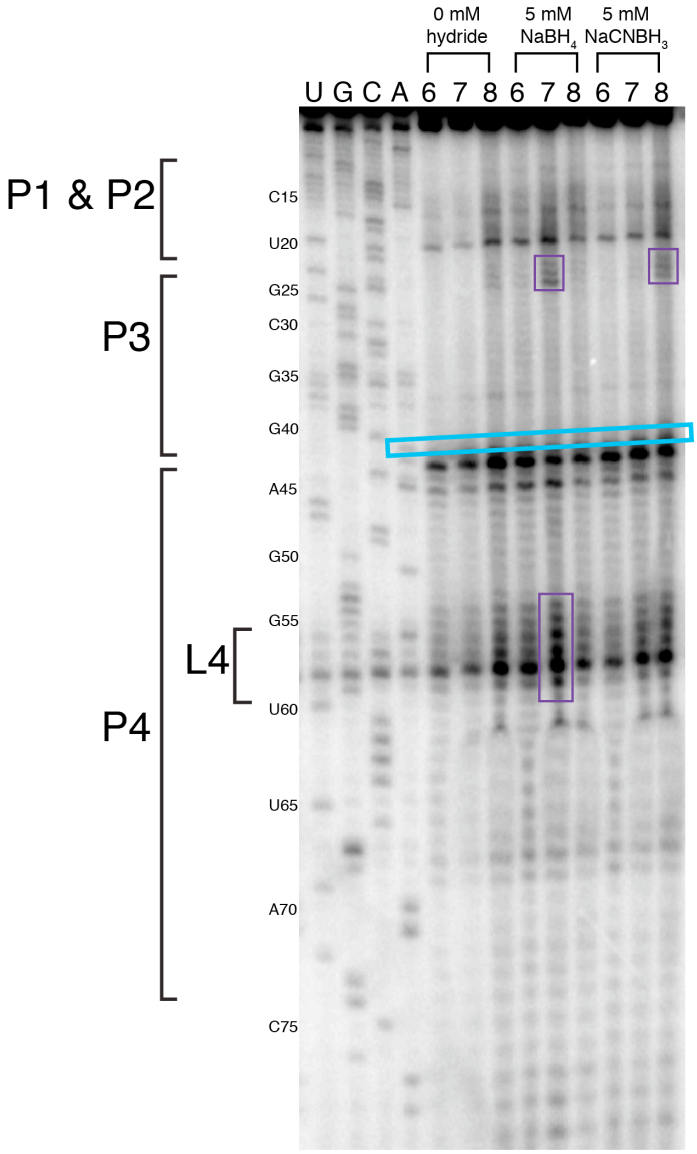


Figure **4-17** Sodium borohydride and sodium cyanoborohydride treatment of HDV. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 5 mM sodium borohydride or sodium cyanoborohydride on ice for 1 hour. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at C41 would result in increase in intensity at A42 (blue box). Reaction at C41 would result in increase in intensity at A42 (blue box). (5-8-09)

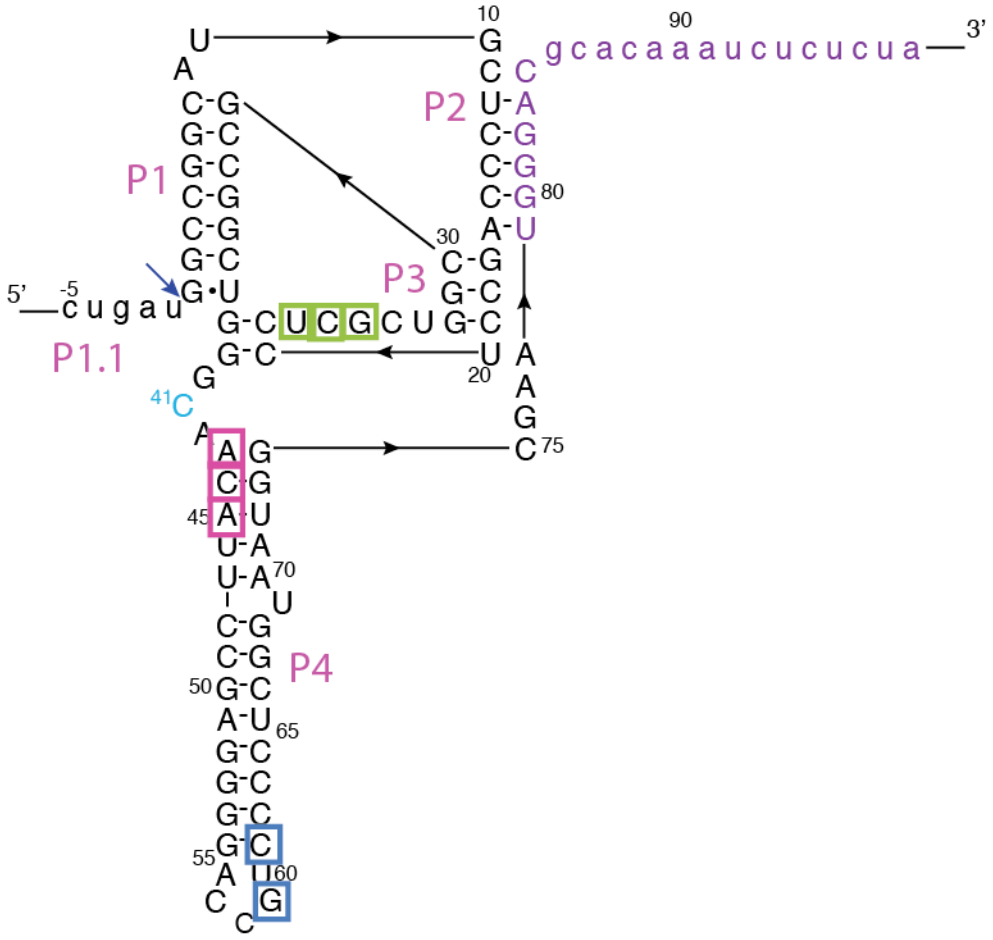


Figure **4-18** Sodium borohydride and sodium cyanoborohydride treatment of HDV RNA mapped to secondary structure. Stops were observed in the gel in Figure 4-17 corresponding to the boxed region in the HDV secondary structure. Green boxes correspond to bands with the highest reactivity observed at pH 7, pink boxes correspond to consistently high reactivity, and blue boxes correspond to high reactivity observed at pH 7 with sodium borohydride and at pH 8 with sodium cyanoborohydride. No reactivity was observed for C41, the protonated base in HDV, shown in blue. (5-8-09)

Initial sodium borohydride and sodium cyanoborohydride treatments were performed on ice to slow the reaction of hydride non-specifically, as performed in the enzyme literature.[18] The temperature of hydride reactions was varied in an effort to optimize reaction conditions. Samples were treated with sodium borohydride or sodium cyanoborohydride at 0 °C, 25 °C, and 37 °C for one hour. (Figure 4-19) As with previous experiments, more reaction was observed with sodium borohydride than sodium cyanoborohydride. Previous experiments showed higher reactivity at pH 7 (Figure 4-17) while we expected higher reactivity to occur at pH 6, due to protonation of the A^{+}•C wobble. Results of these hydride reactions showed more reactivity at pH 8 (Figure 4-19). Observed bands, however, do not correspond to stops downstream of protonation site. In Figure 4-19, the majority of stops are due to structural complexities of RT and are not localized to A42. Because less structural stops were observed with the hydride treatment was performed on ice, future experiments with hydride treatment were carried out on ice.

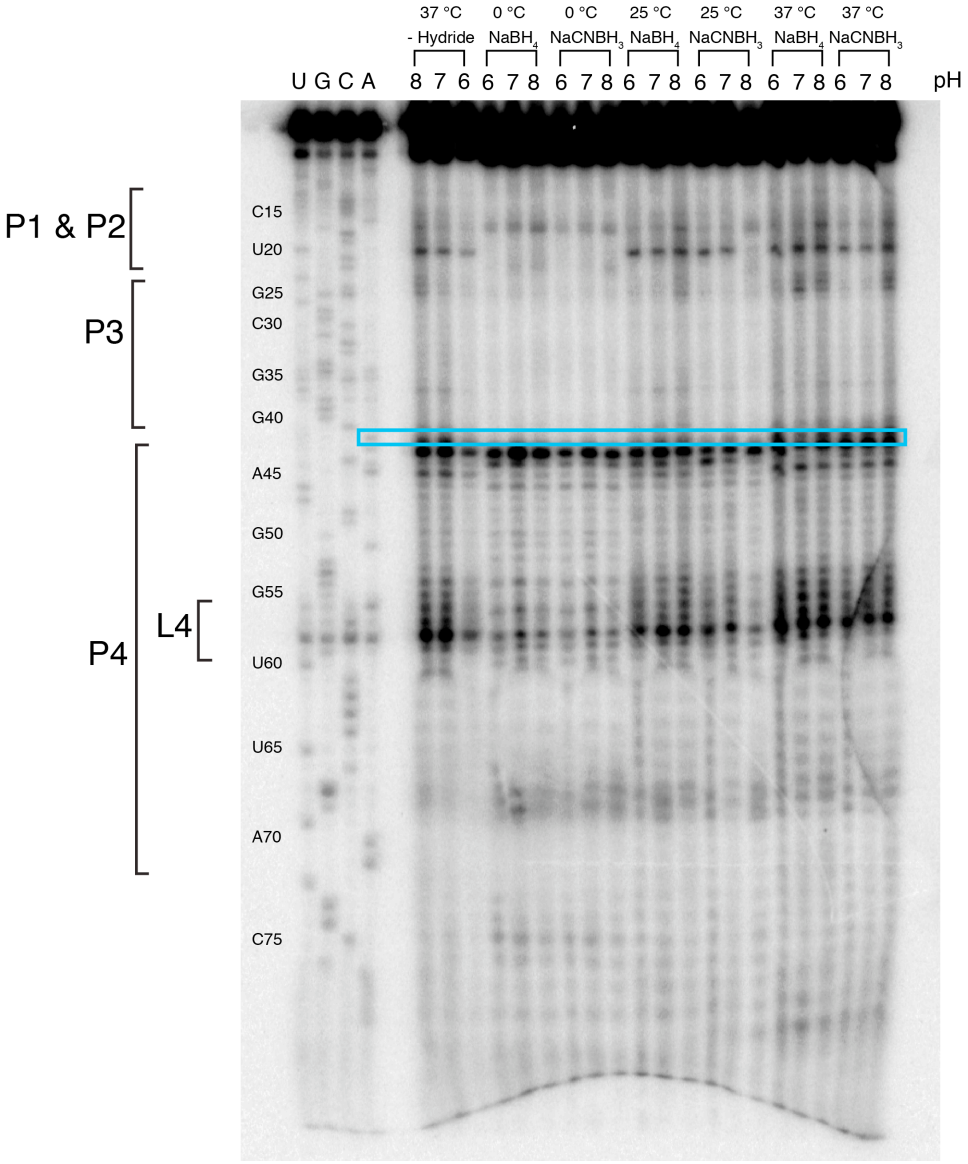


Figure **4-19**: Treatment of HDV RNA with sodium borohydride or sodium cyanoborohydride at various temperatures. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 5 mM sodium borohydride or sodium cyanoborohydride at 0 °C, 25 °C, and 37 °C for 1 hour. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at C41 would result in increase in intensity at A42 (blue box). (5-12-09)

4.4.2.3 Sodium Cyanoborohydride Treatment of Model Oligonucleotides

As in attempts to optimize sequencing, the use of smaller less complex oligonucleotides removed the complexities associated with HDV, therefore, efforts for developing a reagent for damaging a protonated base were made using model oligonucleotides. Initial model oligonucleotides (Figure 4-3) contain an A^+•C wobble or a non-ionizable A-U base pair as a control in the stem of a hairpin. Upon comparison of sodium borohydride treatment of model oligonucleotides with a protonated A^+•C wobble and control oligonucleotides with a non-ionizable A-U base pair, we expected to see more reaction in protonated oligonucleotides due to the addition of a protonation site (Figure 4-20). However, both control experiments of no hydride treatment of A^+•C model oligonucleotides and sodium borohydride treatment of A-U model oligonucleotides showed similar bands, thus the difference was not due to protonation. Background stops were higher with sodium borohydride treatment of RNA than in control samples but were not localized at G25, which would indicate protonation at A24. Treatment of model oligonucleotides with sodium cyanoborohydride showed comparable lack of reactivity at G25 in both protonated A^+•C oligonucleotides and non-ionizable A-U controls (Figure 4-21). Variation of sodium borohydride concentration and comparison to treatment of A-U control oligonucleotides showed slightly more reactivity when treating with 25 mM sodium borohydride at low pH (Figure 4-22). Higher reactivity was observed primarily at U22 and C23 with 25 mM sodium borohydride at pH 6, which was substantially more reactive than those observed in U-A non-ionizable control constructs and untreated control lanes. These nucleotides correspond to G21 and U22, two and three nucleotides upstream of protonated A24 suggesting no increased reactivity at protonated A's.

In both sequencing and sodium cyanoborohydride reactions, additional bands traveling slower than the full-length model oligonucleotides were observed. Analysis of the sequence of this additional RNA portion did not reveal its identity. Since sequencing of the hairpin region was accurate, use of the model oligonucleotides for the development of a reagent for mapping protonated bases continued.

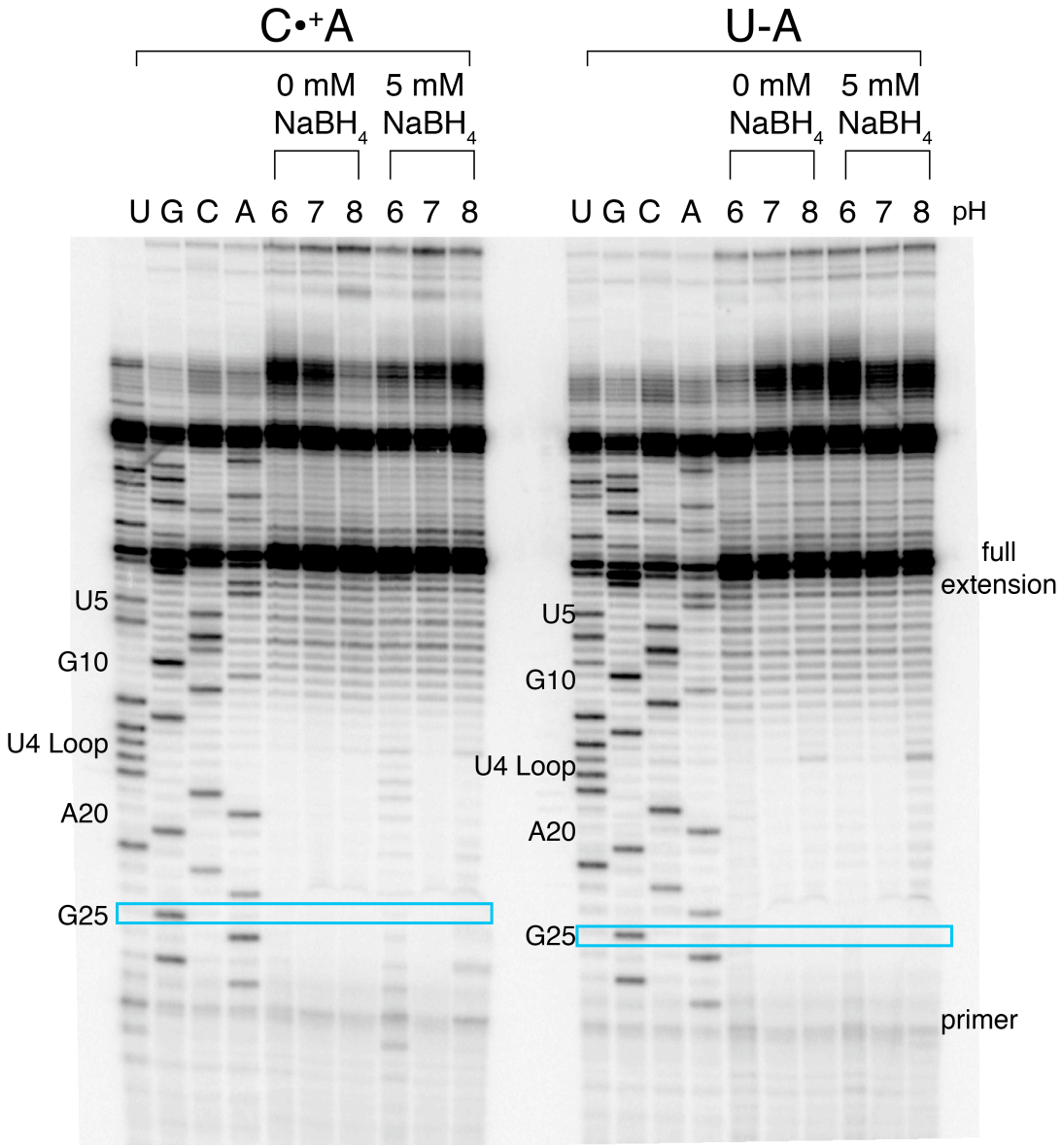


Figure **4-20**: Treatment of model oligonucleotides with sodium borohydride. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 5 mM sodium borohydride at 37 °C for 30 minutes. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP. RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at A24 would result in increase in intensity at G25 (blue box). (1-14-10)

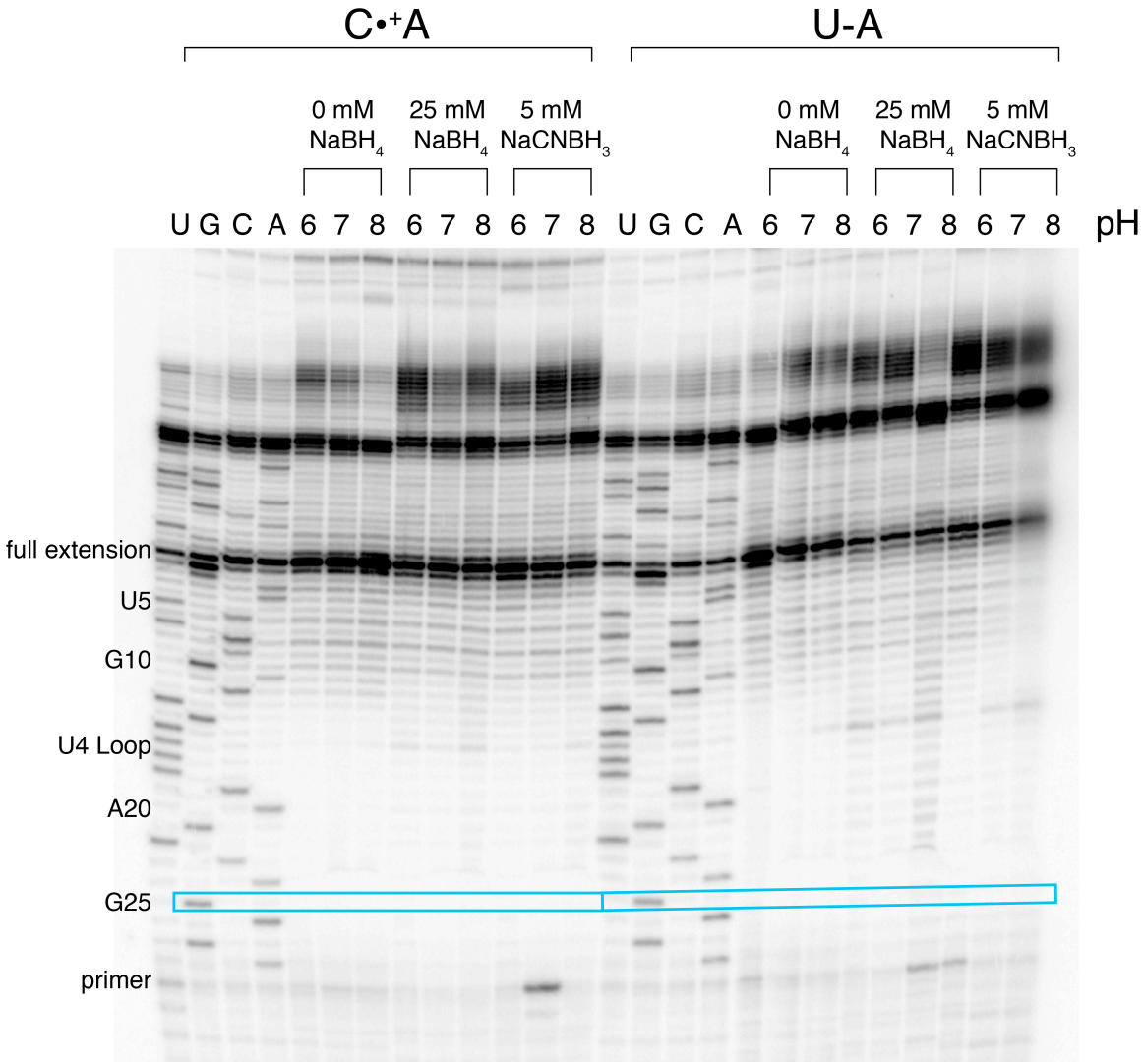


Figure **4-21**: Treatment of model oligonucleotides with sodium borohydride or sodium cyanoborohydride. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 25 mM sodium borohydride or 5 mM sodium cyanoborohydride at 37 °C for 30 minutes. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at A24 would result in increase in intensity at G25 (blue box). (1-15-10)

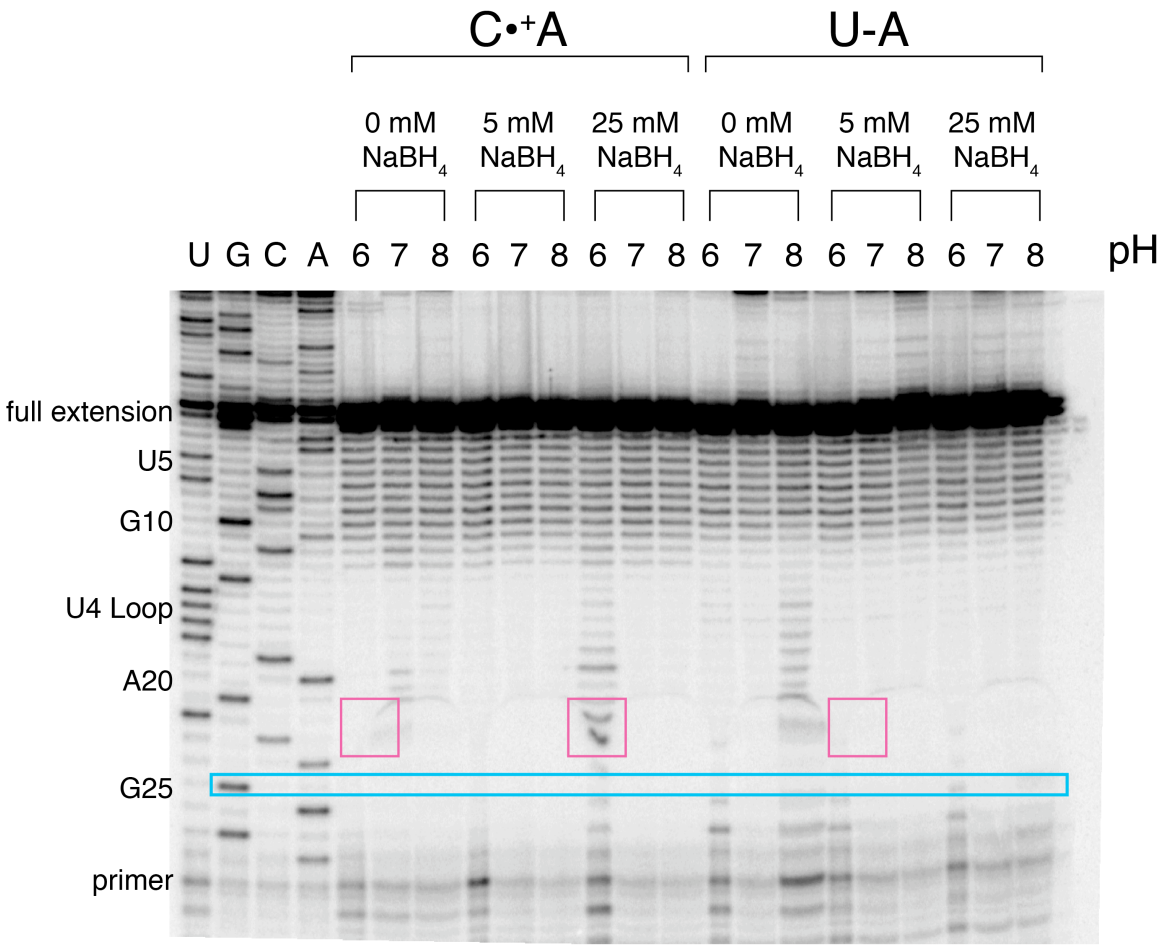
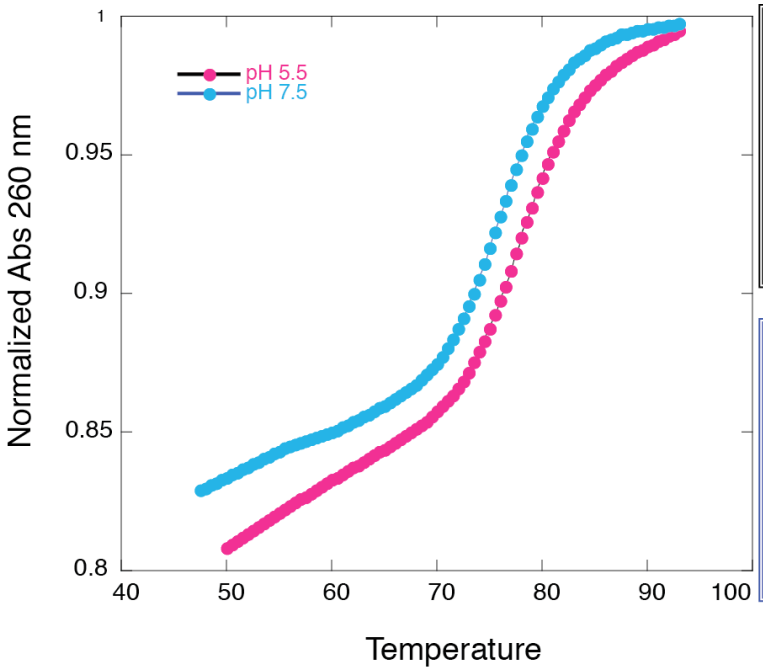


Figure **4-22**: Treatment of model oligonucleotides with sodium borohydride. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 5 mM or 25 mM sodium borohydride for 30 minutes on ice. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at A24 would result in increase in intensity at G25 (blue box). Higher reactivity was observed at bands corresponding to reactivity at G21 and U22. Reactive areas are outlined in pink. The gel was cropped to focus on the fully extended hairpin region. (1-26-10)

4.4.2.4 Confirmation of Protonation of C^+•U Model Oligonucleotides by UV-Melts

Model oligonucleotides with an A^+•C wobble and were based off of constructs used in previous pK_a studies.[32] As most known DNA reactive reagents preferentially react with pyrimidines, we expanded our potential for damage to protonated C nucleobases. Model oligonucleotides were designed to include a C^+•U protonated base pair and a non-ionizable C-G as a control (Figure 4-3). Protonation of C^+•U was confirmed by pH dependent thermal denaturation melting experiments (Figure 4-23). Thermal denaturation experiments at pH 5.5 revealed a higher melting temperature than pH 7.5, indicating protonation and base pair formation of C^+•U at pH 5.5. The melting temperatures were only slightly stabilized with increasing pH, suggesting a minimally shifted pK_a, however they are comparable to stabilized melting temperatures upon protonation in Chapter 3 in similar RNA hairpin constructs.



y = Melt(2e-5, 1, 2e-5, 0.85...		
	Value	Error
a	0.0018613	5.2924e-05
b	0.8223	0.004777
c	0.002266	1.2753e-05
d	0.69611	0.0007554
f	-100.59	0.99247
g	78.041	0.038787
Chisq	1.4372e-05	NA
R^2	0.99996	NA

y = Melt(2e-5, 1, 2e-5, 0.85...		
	Value	Error
a	0.0006589	4.3103e-05
b	0.93676	0.0038684
c	0.0016076	1.2536e-05
d	0.75386	0.00071195
f	-92.754	0.74849
g	76.377	0.032305
Chisq	1.5063e-05	NA
R^2	0.99996	NA

Figure **4-23**: Thermal denaturation U^{+}•C oligonucleotides. U^{+}•C oligonucleotides were renatured at pH 5.5 or 7.5 with bis-tris-propane. Absorbances were measured at 260 nm. Curves were fit to a two-state model to determine the thermodynamic parameters: pH 5.5- T_m=78.0 °C, ΔH = -100.6 kcal/mol, ΔS = -286.4 eu, ΔG°_{37} = -11.6 kcal/mol; pH 7.5- T_m=76.4 °C, ΔH = -92.6 kcal/mol, ΔS = -265.4 eu, ΔG°_{37} = -10.4 kcal/mol.

Results after treatment of sodium borohydride with C^+•U oligonucleotides were comparable to those seen in A^+•C oligonucleotides. When treated with 100 mM sodium borohydride, higher reactivity was observed for all oligonucleotides and did not show a preference for treatment of protonated oligonucleotides (Figure 4-24). Treatment with lower concentration of sodium borohydride showed more reaction at low pH, however, similar results were seen on control experiments of no hydride treatment and treatment of C-G control oligonucleotide (Figure 4-25). Lack of difference between protonated and non-ionizable controls suggests that reaction was not occurring preferentially with protonated nucleobases.

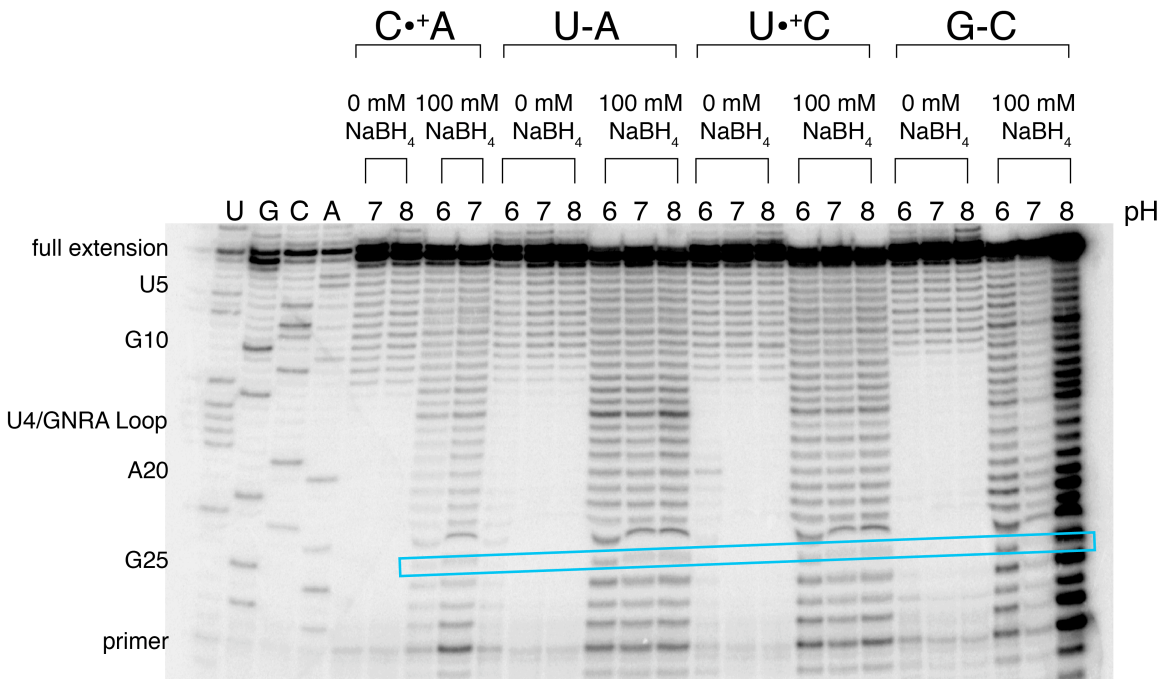


Figure **4-24**: Treatment of C^{+}•A and U^{+}•C oligonucleotides and U-A and G-C non-ionizable control oligonucleotides with 100 mM sodium borohydride. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 100 mM sodium borohydride for 30 minutes on ice. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at A or C 24 would result in increase in intensity at G25 (blue box). The gel was cropped to focus on the fully extended hairpin region. (2-4-10)

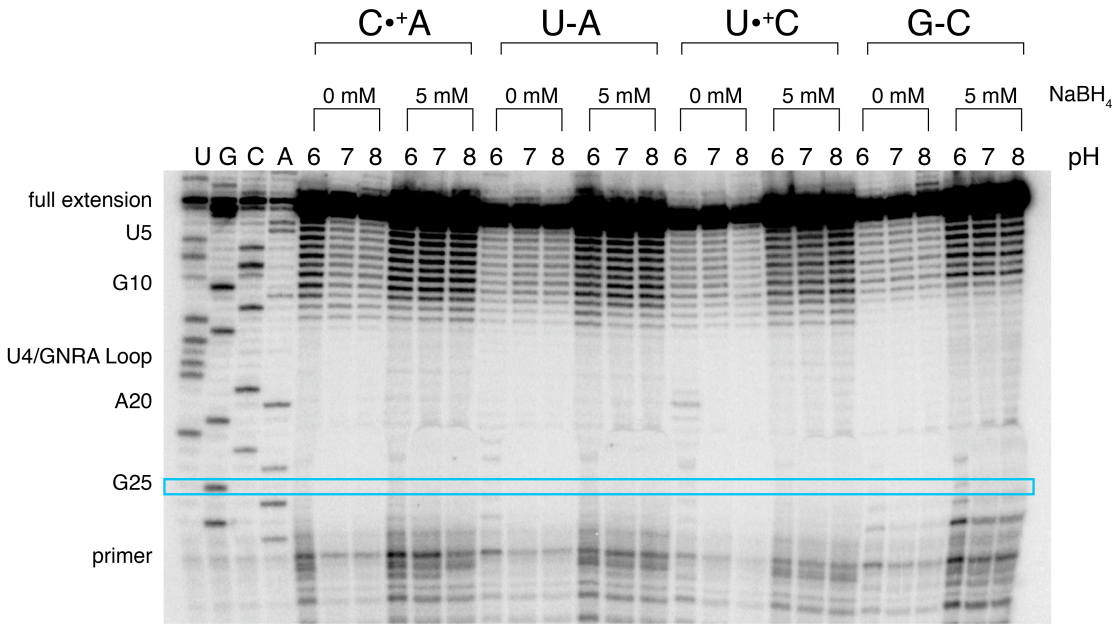


Figure **4-25**: Treatment of C^{+}•A and U^{+}•C oligonucleotides and U-A and G-C non-ionizable control oligonucleotides with 5 mM sodium borohydride. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 5 mM sodium borohydride for 30 minutes on ice. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at A or C 24 would result in increase in intensity at G25 (blue box). The gel was cropped to focus on the fully extended hairpin region. (2-17-10)

4.4.2.5 Photohydration Treatment of HDV

In addition to chemical reagents, photohydration was performed on HDV in a manner analogous to RNA cross-linking. Because of the proposed increased reactivity of protonated bases, it was hypothesized that irradiation would preferentially cause photohydration of protonated nucleobases. For photohydration, a control experiment was done to ensure temperature was maintained throughout irradiation. Using a thermocouple, it was determined that the temperature of the sample increased by less than 10 °C after ten minutes of continuous irradiation. Samples began at 18 °C and were heated to just 27.3 °C after ten minutes. Irradiation time was initially varied to determine optimal length of irradiation (Figure 4-26). There was increased reaction after ten minutes of irradiation, however substantially less full extension of HDV DNA primer was seen, suggesting that irradiation could have led to degradation of samples rather than photohydration. Small volumes of sequencing samples and reaction samples were mixed after RT to confirm that reaction samples were running at the same speed as sequencing samples and thus correct identification of bands in reaction lanes. Sequencing samples, photohydration samples and mixed samples ran at the same speed. Since difficulties still remained in achieving sharp bands when sequencing HDV, the focus at this point was moved to performing photohydration on model oligonucleotides

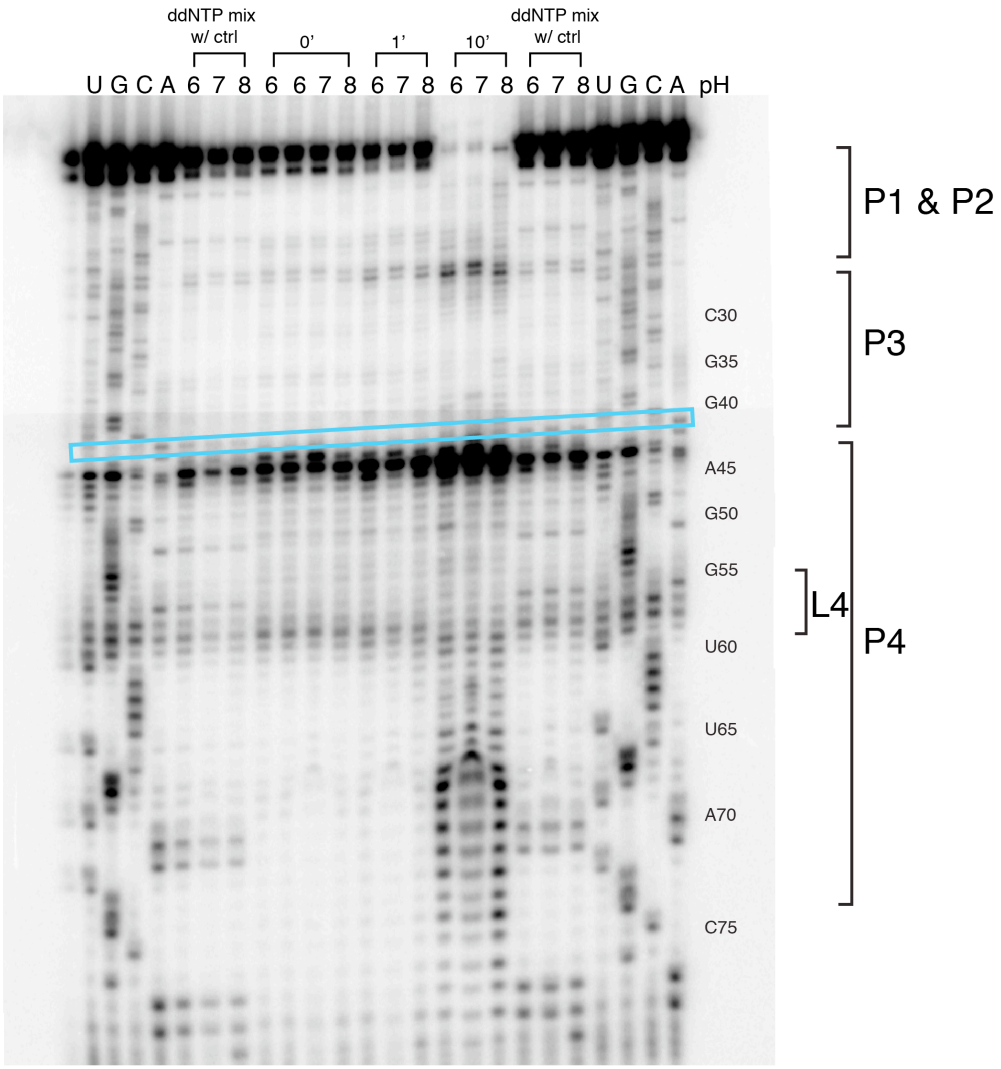


Figure **4-26**: Photohydration of HDV RNA. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then irradiated for one or ten minutes with a 254 nm UV-handheld lamp. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. ddUTP sequencing reaction was mixed with photohydration reaction samples to ensure all samples were migrating at the same speed throughout the gel. Reaction at C41 would result in increase in intensity at A42 (blue box). (8-24-09)

4.4.2.6 Photohydration Treatment of Model Oligonucleotides

Model oligonucleotides with an A^+•C or C^+•U base pair or non-ionizable control oligonucleotides were irradiated with UV light for ten minutes and compared to untreated samples (Figure 4-27 left). As seen in hydride experiments, higher reactivity occurred at low pH but there was no preference for reaction of protonated oligonucleotides over control oligonucleotides. Dark bands were observed in the loop region for both oligonucleotides, most likely due to crosslinking in the U4 tetraloop as U's crosslink readily and result in the formation of uracil dimers.[35] C^+•U and C-G oligonucleotides were designed with a GA3 tetraloop rather than a U4 tetraloop in A^+•C and A-U oligonucleotides. Replacement of the U4 tetraloop with a GA3 tetraloop eliminated potential for crosslinking and bands after irradiation. The dark bands corresponding to the U4 loop in A^+•C and A-U oligonucleotides almost completely disappeared in C^+•U and C-G oligonucleotides with a GA3 tetraloop. Irradiation of C^+•U and C-G oligonucleotides showed slightly higher reactivity in C-G control oligonucleotides. Additionally, a two-step photohydration and hydride reaction was employed to trap any reactive intermediates but had comparable effects to separate photohydration and hydride treatments suggesting that combining the two reactions did not increase reactivity with protonated nucleobases. Because specificity of reactivity was not observed with protonated nucleobases after irradiation, further attempts to develop a method for mapping protonated bases were made through the use of chemical reagents.

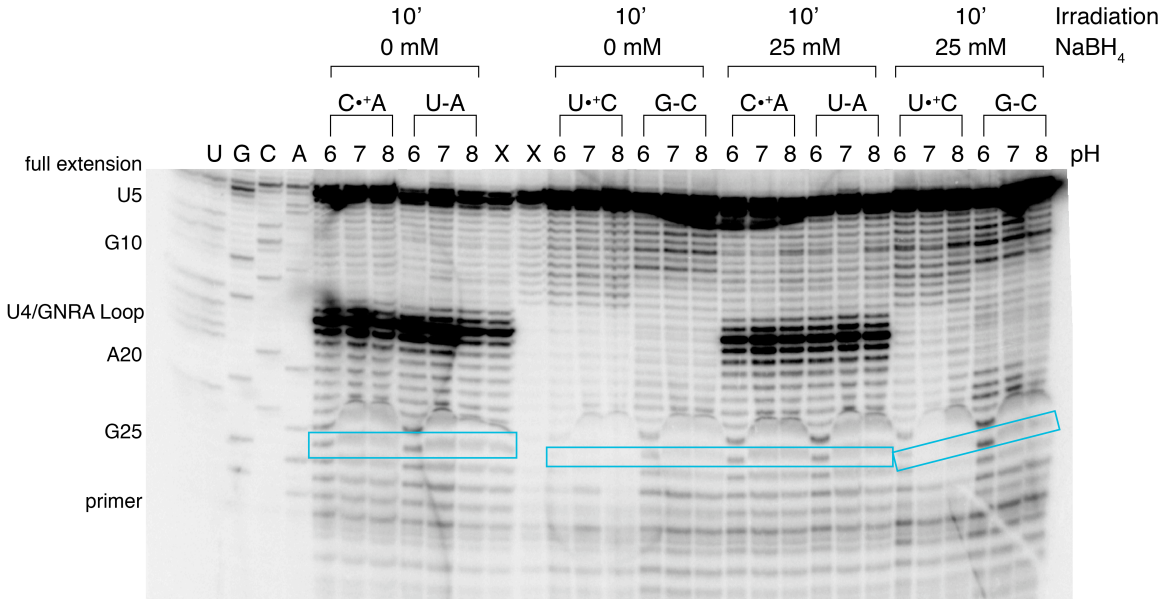


Figure **4-27**: Photohydration of C^{+}•A, U-A, U^{+}•C and G-C model oligonucleotides with two-step photohydration and sodium bisulfite reaction. Samples were buffered by 25 mM MES (pH 6) or HEPES (pH 7 and 8), heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 0 mM or 25 mM sodium borohydride for 30 minutes on ice then irradiated with a 254 nm hand-held UV lamp for 10 minutes. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at A or C 24 would result in increase in intensity at G25 (blue box). (2-4-10)

4.4.2.7 Sodium Bisulfite Treatment of HDV and Model Oligonucleotides

Initial efforts were made using sodium bisulfite, a reagent commonly used for conversion of C's to U's in bisulfite sequencing. Results after treatment of HDV with bisulfite were discouraging (Figure 4-28). A dark band was present in samples treated with 0.025 M sodium bisulfite and was darker than a band at the same location in control samples with no bisulfite treatment. However, this band corresponded to A45, indicating damage at C44, a terminally base paired nucleotide in the P4 region of HDV. No increase in intensity was observed in bands corresponding to C41 (Figure 4-28- blue box). Treatment with 0.25 M and 1 M sodium bisulfite showed degradation of RNA as a high amount of primer was evident.

After treatment of model oligonucleotides with bisulfite (Figure 4-29), higher reactivity was observed at A24 and C24, the sites of protonation (Figure 4-29- pink box). If protonation were to have occurred in these systems, we would have expected to observe stops at G25, one nucleotide downstream of the protonated base (Figure 4-29 blue box). Highest reactivity was observed in the control 0 M bisulfite treatment of the C^+•U oligonucleotide, which could have been due to misloading of the gel, as a structural stop at C24 would have been evident in reaction lanes as well as the control lane. As observed in previous gels, a higher reactivity was observed in the U4 loop of the C^+•A and U-A oligonucleotides (Figure 4-29 purple boxes). After replacing the U4 loop with a GA3 loop when designing the U^+•C and G-C oligonucleotides, no reactivity was observed in the loop region, indicating that reaction at the U4 loop was uridine specific rather than structural stops. Treatment with 1.5 M sodium bisulfite in all constructs resulted in deformation of samples during gel electrophoresis, most likely due to the high salt concentration in those samples. The untreated G-C control oligonucleotide also exhibited the effects of high salt treatment, which could be attributed to significant overflow from the well of the 1.5 M treated

C^{+}•U oligonucleotide into the well of the untreated control of the C-G control oligonucleotide (Figure 4-29 green box).

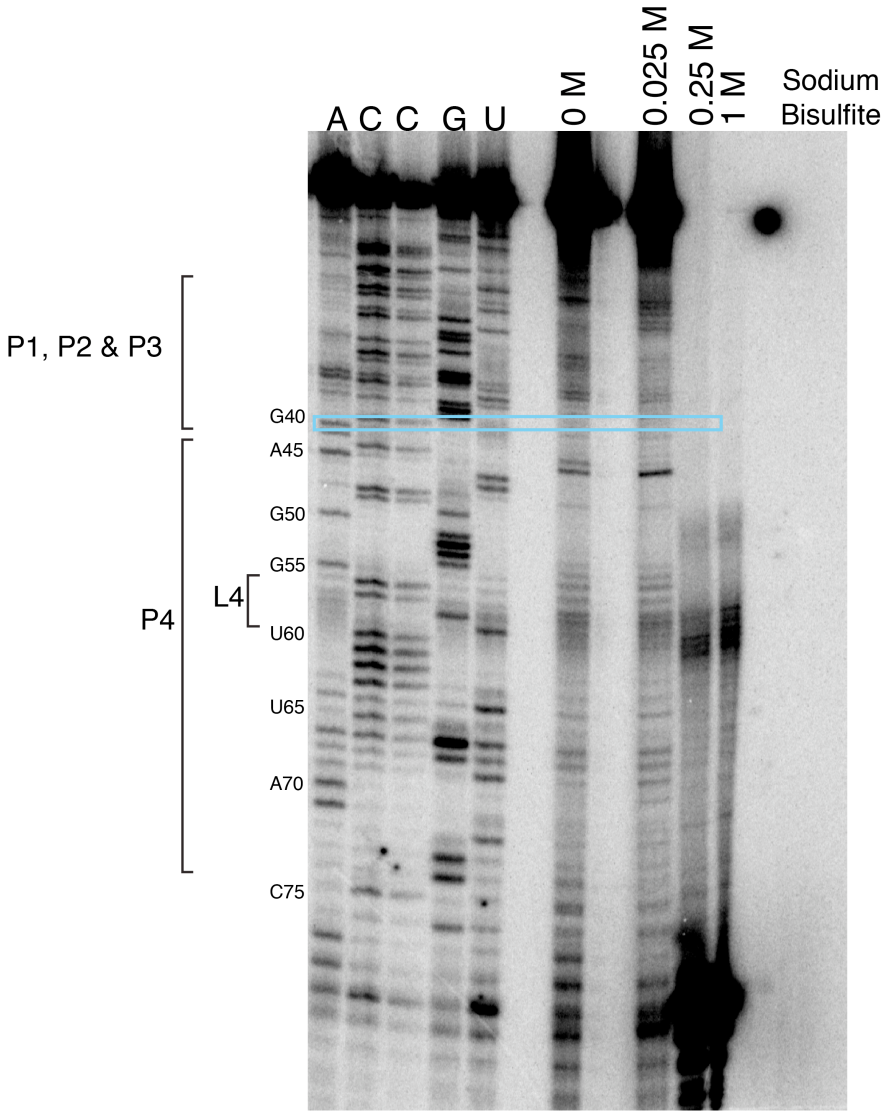


Figure **4-28**: Sodium bisulfite treatment of HDV. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 0.025 M, 0.25 M or 1M sodium bisulfite pH 5.53 at room temperature for 1 hour. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. Reaction at C41 would result in increase in intensity at A42 (blue box). (12-1-12)

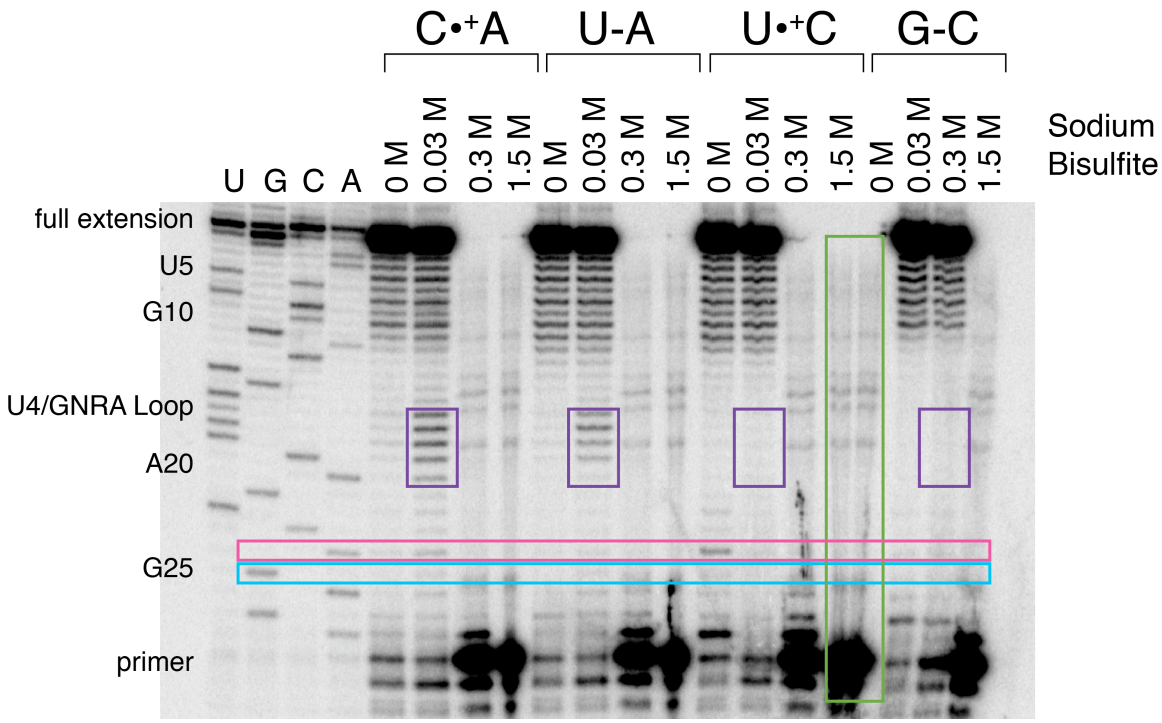


Figure **4-29**: Sodium bisulfite treatment of C^{+}•A, U-A, U^{+}•C and G-C model oligonucleotides. Samples were heated to 90 °C for 2 minutes then cooled to room temperature for 15 minutes for renaturation then treated with 0 M, 0.03 M, 0.3 M or 1.5 M sodium bisulfite pH 5.5 for 60 minutes on ice. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. (2-26-10)

4.4.2.8 Two-Step Sodium Bisulfite Treatment of Model Oligonucleotides

Because of the lack of clear results from hydride and photohydration experiments, alternate reagents were utilized to determine damage to model oligonucleotides. Expected results included reaction at A24 in A^{+}•C oligonucleotides which would be evident by a stop at G25. Since the bisulfite sequencing reaction is reversible before increasing pH to convert to U, it is possible that bisulfite treatment was damaging protonated bases reversibly and damaged products were being converted into more stable reactants. We used a two-step procedure using bisulfite as the primary reagent and varied the secondary reagent with the intent of trapping the reversible intermediate. When ethylenediamine was used as the secondary reagent, primer extension was prohibited and degradation of RNA occurred independent of when over the course of the bisulfite reaction it was added (Figure 4-30). Both addition of ethylenediamine at the beginning of treatment with bisulfite and 30 minutes after the start of the bisulfite reaction caused degradation to RNA, as evident by the single band present in gel electrophoresis corresponding to DNA primer.

Treatment of samples with two-step reactions with sodium borohydride as the secondary reagent showed varied reactivity (Figure 4-31 left). A higher activity was observed at high sodium bisulfite concentration in all constructs in the 3' strand of the model oligonucleotide hairpin. For nucleotides further upstream, higher reactivity was observed at lower concentrations of sodium bisulfite. Full primer extension decreased as sodium bisulfite concentration increased, explaining the increased intensity of bands in the 5' strand with lower bisulfite concentrations. Although the extent of damage remained constant independent of location in the hairpin, less primer was able to extend to report on the damage in the 5' strand in higher concentrations of

bisulfite since increased bisulfite concentration increased damage, giving the impression of less reactivity.

Increased reactivity was also seen in samples treated with a high concentration of bisulfite with two-step aniline treatment (Figure 4-31 right). Dark bands were seen in the U4 tetraloop region of C^{+}•A and U-A model oligonucleotides with 60 mM bisulfite treatment; but changing the U4 tetraloop to a GA3 tetraloop in U^{+}•C and G-C model oligonucleotides almost completely eliminated this effect, indicating a pyrimidine exclusive effect.

Treatment of hydroxylamine and diluted ethylenediamine showed similar results to two-step treatments with sodium borohydride and aniline (Figure 4-32), indicating that increased bisulfite prohibited primer extension regardless of identity of the secondary reagent.

Most importantly, with all reagents, at lower concentration, there was no increase in reactivity in protonated oligonucleotides over non-ionizable control oligonucleotides. Based on these preliminary results, either bisulfite was not able to preferentially damage a protonated base or the secondary reagents were incapable of trapping the reversible intermediate under the reaction conditions used. The two-step procedure with sodium bisulfite as the primary reagent has potential to selectively damage a protonated nucleobase. Experiments utilizing this method will be discussed in detail later in this chapter.

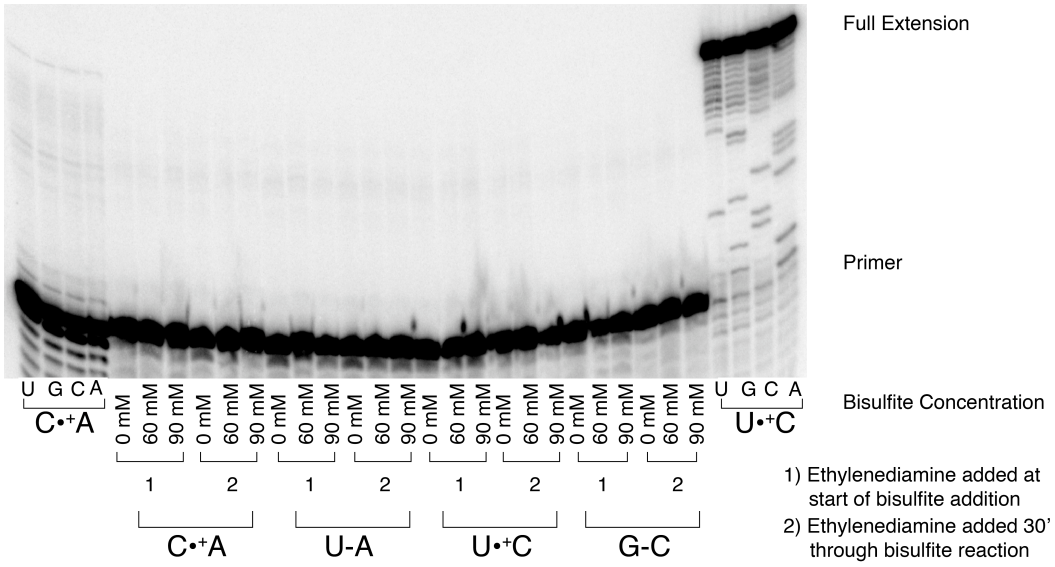


Figure **4-30**: Two-step treatment of model oligonucleotides with sodium bisulfite and ethylenediamine. Samples were heated to 90 °C for 2 minutes, cooled to room temperature for 15 minutes for renaturation then treated with 0, 60 or 90 mM sodium bisulfite pH 5.5 at room temperature for 30 minutes before treatment with ethylenediamine for 30 minutes at room temperature. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. (3-30-10)

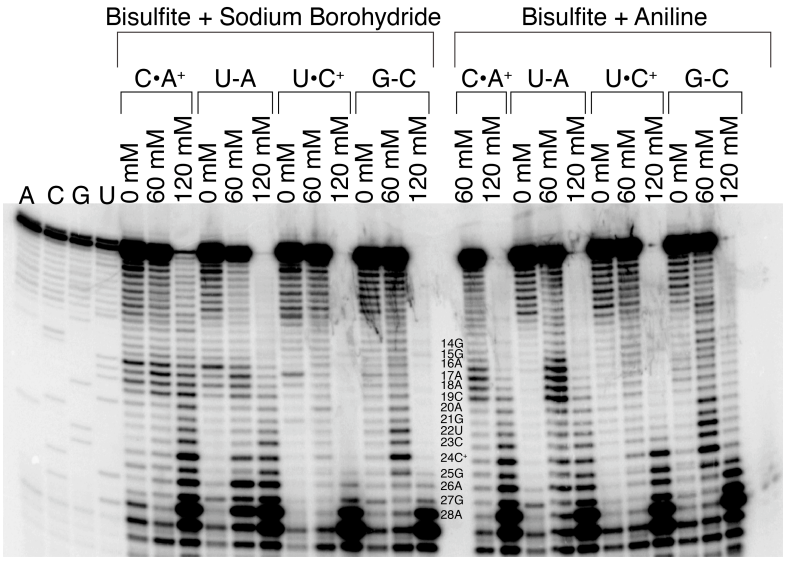


Figure **4-31**: Two-step treatment of model oligonucleotides with sodium bisulfite and sodium borohydride or sodium bisulfite and aniline. Samples were heated to 90 °C for 2 minutes, cooled to room temperature for 15 minutes for renaturation then treated with 0, 60 or 90 mM sodium bisulfite pH 5.5 at room temperature for 30 minutes before treatment with sodium borohydride or aniline for 30 minutes at room temperature. Final concentrations of RT solutions were 0.1 µM RNA, 0.02 µM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. (4-27-10)

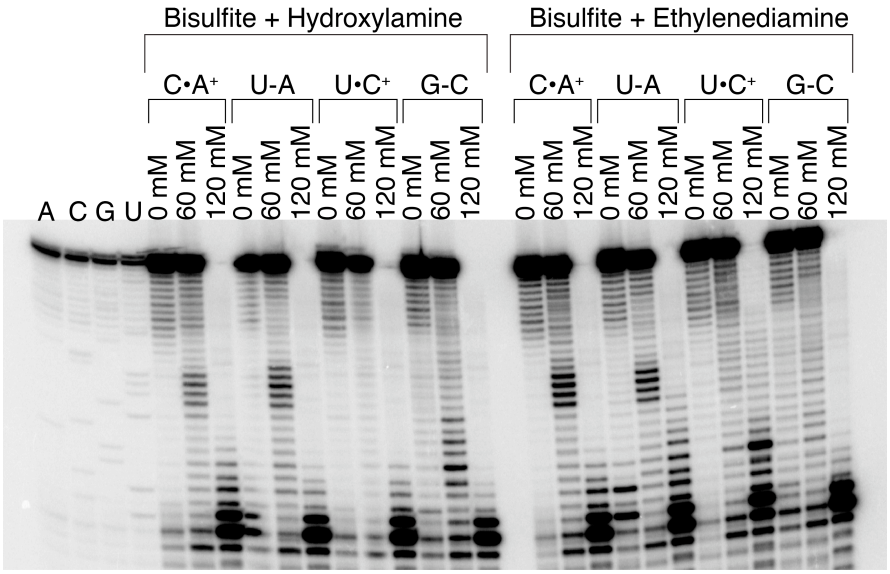


Figure **4-32**: Two-step treatment of model oligonucleotides with sodium bisulfite and hydroxylamine or diluted ethylenediamine. Samples were heated to 90 °C for 2 minutes, cooled to room temperature for 15 minutes for renaturation then treated with 0, 60 or 90 mM sodium bisulfite pH 5.5 at room temperature for 30 minutes before treatment with hydroxylamine or diluted ethylenediamine for 30 minutes at room temperature. Final concentrations of RT solutions were 0.1 μM RNA, 0.02 μM ^{32}P radiolabeled DNA primer, 1.25 mM each dNTP, and 0.02 mM ddNTP (ddNTP in sequencing lanes only). RT was performed at 42 °C for 60 minutes. Before loading on a gel, samples were heated to 90 °C for 2 min. Gel was run with large wells at 100 W and a Plexiglas shield was used for insulation. (4-27-10)

4.4.3 ^{1}H NMR and UV-Vis Spectroscopies for Visualization of Reagent Addition

NMR and UV-Vis spectroscopies were employed to visualize the addition of sodium bisulfite to CMP through characterization of the bisulfite treated product. The addition of bisulfite to CMP would replace H6 with SO_3^-, removing the proton peak in the bisulfite treated samples. Spectra of CMP in the absence of sodium bisulfite were clean and suitable for detection. Distinct peaks for H5 and H6 were observed (Figure 4-33 top panel). After addition of sodium bisulfite at pH 4, peaks corresponding to H5 and H6 were still present, resonating at approximately 8 ppm

(Figure 4-33 bottom panel). Although these reactions suggest that bisulfite was unable to permanently damage CMP, experimental procedures were problematic. Although molar concentration of bisulfite is recommended for bisulfite sequencing, millimolar concentrations of sodium bisulfite were used for treatment of millimolar concentrations of CMP, inhibiting the bisulfite reaction. The pH of CMP before the treatment of sodium bisulfite was not determined, adding complexity to comparison of the two spectra. The use of UV-Vis spectroscopy was employed, however, addition of bisulfite significantly distorted the absorbance spectrum of CMP and the disruption of aromatic nature after addition of bisulfite could not be confirmed (data not shown).

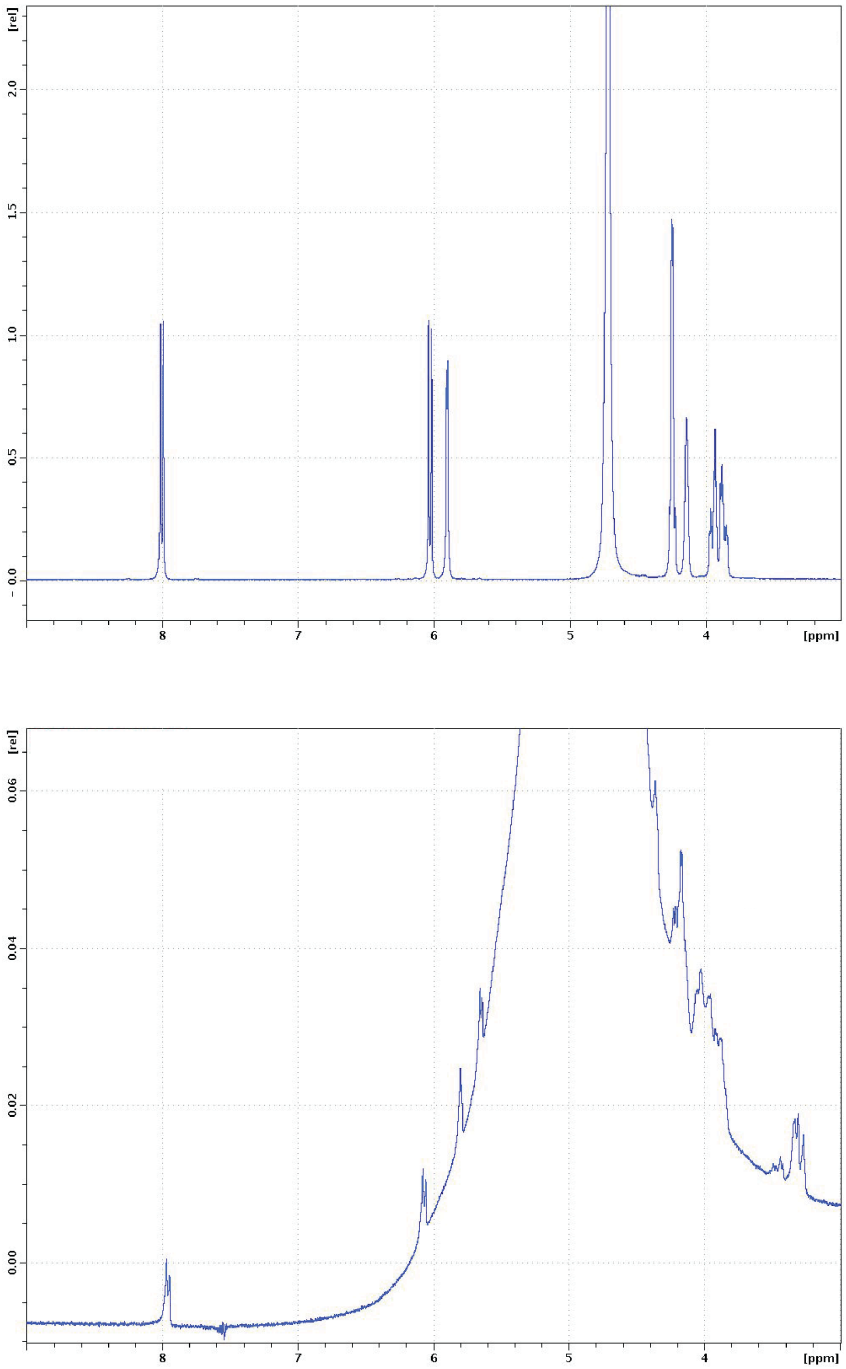


Figure **4-33**: ^{1}H NMR spectra of sodium bisulfite addition to CMP. ^{1}H NMR spectra of CMP were obtained before (top) and after (bottom) treatment with sodium bisulfite, pH 4. Disappearance of the doublet at approximately 8 ppm would indicate addition of sodium bisulfite and a successful reaction.

4.4.4 Mass Spectrometry analysis of Bisulfite treated samples

To limit the variables with detection by RT, analysis by mass spectrometry was pursued. Samples of beet western yellows virus (BWYV) RNA were analyzed by fluorescence spectrometry previously (Chapter 2) and a pK_a of 8.15 was determined. Four samples were prepared for analysis using mass spectrometry. Samples 1 and 2 contained beet western yellows virus (BWYV) RNA with a protonated C at C8. Samples 3 and 4 contained a double mutant of BWYV (BWYV-dm). Mutating a G-C base pair to an A-U maintains tertiary structure in the RNA but removes the potential for ionization. If bisulfite is able to attack and covalently bond to the protonated base, it should be evident in an addition of 79.1 amu (=+SO_3-H) in the mass spectra. Although, mass spectra of high quality were obtained, they showed no mass change between BWYV treated with water and ammonium bisulfite (Figure 4-34), indicating no covalent bond formation upon bisulfite addition.

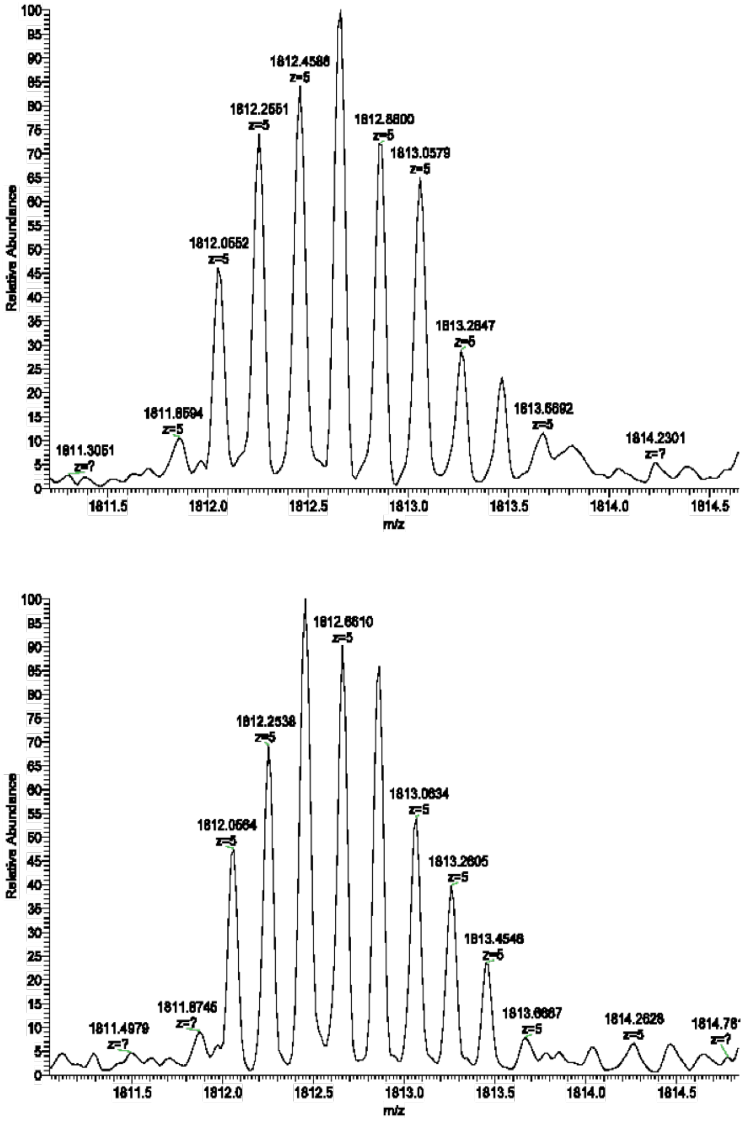


Figure **4-34**: Mass spectra of beet western yellows virus RNA oligonucleotides. Samples were either treated with water (A) as a control of ammonium bisulfite (B). For purification for analysis by mass spectrometry, samples were dialyzed against 2.5 mM ammonium acetate and water. Identical masses were obtained for the water and bisulfite treated samples and equal to the predicted mass of BWYV RNA, indicating that bisulfite did not covalently bond to the BWYV RNA. Very clean, high quality spectra were obtained.

4.5 Discussion

4.5.1 Optimization of Sequencing Conditions

Throughout sequencing with HDV and model oligonucleotides, challenges in sequencing arose from gel electrophoresis. Although background was significantly decreased when using model oligonucleotides, sharp, distinguishable bands were still scarce. It is important to note that although sequencing was possible in the majority of the gels presented, sharp, high-resolution bands are necessary for confidently identifying a stop in reaction lanes. HDV was too structured to sequence the RNA. Sequencing was typically clear through the P2, but sequencing through and past P4 was very challenging because C41 is located in P1.1, clear sequencing throughout P4 is crucial for visualization of pH dependent stops downstream of C41. We reasoned that problems with visualization of results most likely occurred during gel electrophoresis. To ensure samples were completely denatured, when fractionated by gel electrophoresis, samples were boiled before loading, urea concentration was increased, the power while running the gel was increased, and the gel was insulated to ensure even heating. Small improvements were observed throughout optimization of gel electrophoresis but the major improvement occurred after changing the comb used. Replacing the smaller well comb with a comb with larger wells but maintaining the loaded sample volume allowed for thinner bands to travel evenly throughout the gel. Drastic improvements in sequencing were observed when the comb was replaced. Although slight improvements remained after first using a comb with larger wells, such a drastic improvement could not be reproduced. The best conditions included sequencing at 0.2 mM ddNTP concentration, RT with unmodified dGTP, gels run at 100 W with a Plexiglas shield for insulation and a comb with large wells and a small sample volume loaded.

Efforts for improvement of RT mainly focused on minimizing structure in P4 to obtain clear sequencing. The temperature of RT was increased to destabilize P4, but difficulties in sequencing the G-C rich stem remained. Increasing the temperature of RT increased background in sequencing the other regions of HDV. To visualize stops in close proximity to C41, a longer primer was tried. The 52-nucleotide primer was designed to disrupt structure in P4 as it base paired with C48 (Figure 4-1). Unfortunately, primer extension could not be obtained with the longer primer, indicating that annealing the primer to HDV was unsuccessful and sequencing with the longer primer could not be established. An HDV mutant that deleted 3 G-C base pairs from P4 was obtained and sequenced in an effort to minimize structural obstacles in P4. Deletion of the 3 G-C base pairs in the mutant HDV allowed for a slight increase in sequencing quality but only moderately.

After working with HDV and constantly facing challenges in sequencing P4, smaller model oligonucleotides were designed. Because much of the research on protonated nucleobases in the Bevilacqua lab utilized an A^+•C wobble, constructs were first designed with an A^+•C in the middle of a hairpin. As a control, model oligonucleotides were designed with a non-ionizable A-U base pair replacing the A^+•C wobble. Model oligonucleotides were also designed with a C^+•U base pair after concluding that literature reagents used to modify nucleotides primarily targeted pyrimidines. Additionally, a non-ionizable C-G model oligonucleotide was designed as a control to replace the C^+•U oligonucleotide. These model oligonucleotides significantly improved ease of sequencing. As with P4 in HDV, sequencing difficulties were observed on the 5' strand of the stem after RT of the loop. In model oligonucleotides, however, protonation was designed to occur in the 3' strand and so, it if occurred, could still be easily visualized even with the difficulties in sequencing the 5' strand.

It can be concluded after attempting to improve sequencing quality through varying reaction, gel running, and gel drying conditions, sequencing quality is dependent upon all

examined variables. Clear, crisp sequencing was observed once and could not be reproduced, indicating that additional factors, other than those tested, contributed to the quality of sequencing. Ultimately, gel electrophoresis was replaced with mass spectrometry as an analysis technique because of the large amount of unknown variables that could be contributing to sequencing quality.

4.5.2 Efforts towards Chemical Damage of RNA

Because of the difficulties associated with sequencing and gel electrophoresis, results of chemical damage were ambiguous. Initial treatment with sodium borohydride showed higher reactivity at low pH but stops corresponded to double-stranded nucleotides that should not have been reactive. A bell shaped dependence was observed in further experiments. Use of a less reactive hydride in sodium cyanoborohydride gave less background but inconclusive results. Experiments using sodium bisulfite showed some reaction at the protonation site in model oligonucleotides but only slightly more than that of control non-ionizable oligonucleotides. Furthermore, efforts to trap reversible intermediates using a two-step reaction procedure after bisulfite treatment proved unsuccessful. Overall, results of chemical treatment were inconclusive.

Problems could have arisen from the design of constructs for chemical experiments. Initial investigation into a chemical reagent for damaging protonated bases began using HDV with protonated C41. Although C75 is protonated in the precleaved form as the proposed general acid in general acid-base catalysis, the cleaved 1/99 form used in these experiments contains C75 with a pK_a of approximately 4.5.[29] Additionally, C41 participates in a base quartet,[28] making damage by a chemical reagent more difficult than one participating in simple secondary base pairing. Although the development of a reagent would require the ability to permeate tertiary structure such as the base quartet formed in HDV, determination of a viable reagent for damaging

protonated nucleobases using such a complex system is not ideal. At this point, efforts were made to determine a chemical reagent using smaller, less complex model oligonucleotide systems. Hairpins containing an A^+•C wobble were used in previous pK_a studies[32] and constructs were modeled off of hairpins used previously. Because chemical reagents such as hydrazine, hydroxylamine and bisulfite react preferentially with pyrimidines, we designed oligonucleotides with a C^+•U wobble and pH dependent melts were performed to confirm protonation. Ultimately, efforts towards development of a chemical reagent could have been hindered due to development of constructs for experiments rather than an inability of reagents to preferentially damage protonated bases.

Efforts were made towards utilization of mass spectrometry to replace analysis of samples by gel electrophoresis and overcome difficulties presented with sequencing. Utilization of mass spectrometry eliminated RT and visualization of stops by gel electrophoresis. Initial samples were treated with ammonium bisulfite and analyzed by mass spectrometry, in the hopes that addition mass corresponding to the mass of a bisulfite addition would be evident. Although analysis by mass spectrometry could only determine if the bisulfite addition was present and not the location of occurrence, we had planned to employ an RNase digestion technique in a manner analogous to protein digestion for sequencing.

4.5.3 Analysis of Bisulfite by Mass Spectrometry

After facing significant difficulties in PAGE as an analysis technique, efforts were made in adapting mass spectrometry for analysis. Mass spectra, however, revealed identical masses for BWYV treated with ammonium bisulfite and control samples treated with water. It also showed no difference in spectra for protonated BWYV and a non-ionizable double mutant BWYV RNA. As with bisulfite treatment of oligonucleotides analyzed by gel electrophoresis, the concern

remains in the reversibility of the first steps of bisulfite conversion of C's to U's, and the possibility that unstable intermediates return to stable pseudo-aromatic reactants. Although these results did not support the potential of bisulfite as a viable reagent to damage protonated bases with my new protocol of dialyzing samples in ammonium acetate, clear spectra were obtained. An initial purification procedure requiring dialysis against concentrated ammonium acetate followed by a long dialysis against water sufficiently purified RNA samples for analysis by mass spectrometry, which not only aids in the analysis of samples to develop a method for mapping protonated bases, it enables other RNA samples to be analyzed by mass spectrometry throughout the RNA field.

4.6 Future Directions

In addition to gel electrophoresis, mass spectrometry and capillary electrophoresis should be employed as analysis methods. Since photohydration and bisulfite reactions are reversible, pursuing a two-step method in which a secondary reagent is used to trap the reversible intermediate would be advantageous. In addition to the reagents presented, hydrazine and hydroxylamine should be tried as potential regents.

4.6.1 Optimal Oligonucleotides for Method Development

Because obtaining clear sequencing, in which bands were sharp enough for identification of reactive sites with complete confidence, could not be obtained regularly on HDV, analysis of reaction gels was difficult and unreliable. Future experiments should begin on smaller, less complex systems such as nucleotides or model oligonucleotide systems in which reagents can be developed without the difficulties of more complex systems such as HDV. Beginning with

nucleotides and nucleosides will significantly simplify confirmation of damage as nucleotides and nucleosides are much smaller, less complex systems with which to work and pK_a's of nucleotides and nucleosides are known.

4.6.2 Mass Spectrometry

In future efforts to map protonated bases in large RNA systems, the use of mass spectrometry as an analysis technique will significantly aid in confirmation of damage of a protonated base. Identification of addition location would require RNase treatment to fragment the RNA and analyze smaller pieces, analogous to a trypsin digest. However, replacing the previous analysis methods with mass spectrometry greatly increases the experiment cost. Identifying suitable reagents by trial and error by mass spectrometry will be costly and inefficient. Given the right approach and background knowledge of organic reactions, the potential for development of a method for mapping protonated bases remains.

Future experiments should begin with the utilization of mass spectrometry as an analysis method as conditions have been developed for purification of ammonium bisulfite treated samples. Because of the reversibility of the bisulfite reaction with protonated nucleobases, a two-step treatment with analysis by mass spectrometry should also be utilized in future experiments.

4.6.3 Capillary Electrophoresis

For the development of an electrophoresis based detection method, capillary electrophoresis (CE) is a potential detection method that should be considered in the future. CE requires performing RT with fluorescent primers and detection by a microfiber capillary electrophoresis chip[36] and has been commonly used for DNA sequencing since it was first

developed. Utilization of CE for sequencing would eliminate troubles associated with gel electrophoresis sequencing and allow for rapid determination of a viable reagent for damaging protonated nucleobases.

4.6.4 Additional Reagents

A more complete examination into the viability of reagents used previously should be performed, as buffering capacity of hydride experiments was not determined at hydride concentrations greater than 5 mM. Additionally, future experiments will focus on the use of hydrazine and hydroxylamine as potential reagents for selective damage of protonated bases. Their amine functionalities would enable reaction with protonated bases (Figure 4-2). Additionally, combining sodium bisulfite and hydroxylamine or hydrazine reagents in a two-step reaction would utilize the extensive knowledge of bisulfite reactions with nucleic acids while trapping the sulfite-substituted reversible intermediate.

4.7 Conclusion

We presented significant progress towards the development of a method for mapping protonated nucleobases in large RNA systems. There were difficulties in using gel electrophoresis as an analysis method, specifically when analyzing HDV. Smaller, less complex model oligonucleotides were analyzed by gel electrophoresis with greater success. Inconclusive results were presented after treatment of HDV and model oligonucleotides with sodium borohydride, sodium cyanoborohydride, sodium bisulfite and two-step reactions with sodium bisulfite and sodium borohydride, aniline, ethylenediamine, or hydroxylamine. Additionally, mass

spectrometry was employed as an analysis technique and a protocol for purification of oligonucleotides for mass spectrometry was developed.

Ultimately, the development of a reagent to map protonated bases in large RNA systems is possible. Future experiments should be performed on less complex systems to avoid any unnecessary difficulties due to persistent structure formation. In addition to gel electrophoresis, mass spectrometry and capillary electrophoresis should be pursued as analysis techniques. The suitable reagent for mapping protonated nucleobases needs to be strong enough to penetrate the pseudo-aromaticity of the nucleobases while also specific enough to preferentially damage protonated nucleobases. The potential for the presented reagents to efficiently damage protonated nucleobases still remains, however, a two-step treatment to trap reversible intermediates may be necessary. The development of a reagent for mapping protonated nucleobases is very important, as it has the potential to determine the mechanism of numerous known biological processes that require protonated nucleobases, such as general acid-base catalysis and programmed ribosomal frameshifting. It also has the potential for identifying protonated nucleobases in unknown systems

4.8 Acknowledgements

Mass spectrometry experiments were funded by a Huck Facility Core Seed Grant by the Huck Institutes of the Life Sciences. I would like to thank the entire mass spectrometry facility at The Pennsylvania State University, especially Tatiana Laremore. Tania was instrumental in developing the protocol for mass spectrometry of nucleic acids and a tremendous joy to work with. I would like to thank Dr. Laurie Heinicke for assistance preparing the hepatitis delta virus (HDV) RNA and for donating the HDV Δ3bp RNA, Dr. Adam McGraw for assistance with sequencing conditions, Amber Bisch and Dr. Mark Nilson for assistance obtaining select chemicals, Dr. Andrey Krasilnikov for assistance with RNase experiments, and Dr. Squire

Booker, Dr. Kenneth Feldman, and Dr. Kent Gates at The University of Missouri for conversations regarding bioorganic mechanisms.

4.9 References

1. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer-Verlag, New York.

2. Legault, P., and Pardi, A. (1997) Unusual Dynamics and pKa Shift at the Active Site of a Lead-Dependent Ribozyme, *J. Am. Chem. Soc.* **119**, 6621-6628.

3. Wilcox, J. L., Ahluwalia, A. K., and Bevilacqua, P. C. (2011) Charged nucleobases and their potential for RNA catalysis, *Acc Chem Res* **44**, 1270-1279.

4. Wilcox, J. L., and Bevilacqua, P. C. (2013) A Simple Fluorescence Method for pK(a) Determination in RNA and DNA Reveals Highly Shifted pK(a)'s, *J. Am. Chem. Soc.* **135**, 7390-7393.

5. Moody, E. M., Lecomte, J. T., and Bevilacqua, P. C. (2005) Linkage between proton binding and folding in RNA: a thermodynamic framework and its experimental application for investigating pKa shifting, *RNA* **11**, 157-172.

6. Abeysirigunawardena, S. C., and Chow, C. S. (2008) pH-dependent structural changes of helix 69 from Escherichia coli 23S ribosomal RNA, *RNA* **14**, 782-792.

7. Gong, B., Chen, J. H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. (2007) Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography, *J. Am. Chem. Soc.* **129**, 13335-13342.

8. Guo, M., Spitale, R. C., Volpini, R., Krucinska, J., Cristalli, G., Carey, P. R., and Wedekind, J. E. (2009) Direct Raman measurement of an elevated base pKa in the active site of a small ribozyme in a precatalytic conformation, *J. Am. Chem. Soc.* **131**, 12908-12909.

9. Legault, P., and Pardi, A. (1994) In situ Probing of Adenine Protonation in RNA by 13C NMR, *J. Am. Chem. Soc.* **116**, 8390-8391.

10. Chamorro, M., Parkin, N., and Varmus, H. E. (1992) An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA, *Proc Natl Acad Sci U S A* **89**, 713-717.

11. Bevilacqua, P. C., Brown, T. S., Nakano, S., and Yajima, R. (2004) Catalytic roles for proton transfer and protonation in ribozymes, *Biopolymers* **73**, 90-109.

12. Maxam, A. M., and Gilbert, W. (1977) A new method for sequencing DNA, *Proc Natl Acad Sci U S A* **74**, 560-564.

13. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) Probing the structure of RNAs in solution, *Nucleic Acids Res* **15**, 9109-9128.

14. Conway, L., and Wickens, M. (1989) Modification interference analysis of reactions using RNA substrates, *Methods Enzymol* **180**, 369-379.

15. Peattie, D. A. (1979) Direct chemical method for sequencing RNA, *Proc Natl Acad Sci U S A* **76**, 1760-1764.

16. Weeks, K. M., and Crothers, D. M. (1993) Major groove accessibility of RNA, *Science* **261**, 1574-1577.

17. Pappas, J. J., Toulouse, A., and Bradley, W. E. (2009) A modified protocol for bisulfite genomic sequencing of difficult samples, *Biol Proced Online* **11**, 99-112.

18. Warren, S., Zerner, B., and Westheimer, F. H. (1966) Acetoacetate decarboxylase. Identification of lysine at the active site, *Biochemistry* **5**, 817-823.

19. Kubarenko, A. V., Sergiev, P. V., Bogdanov, A. A., Brimacombe, R., and Dontsova, O. A. (2001) A protonated base pair participating in rRNA tertiary structural interactions, *Nucleic Acids Res* **29**, 5067-5070.

20. Smith, M. (2002) Organic synthesis, 2nd ed., McGraw-Hill, Boston.

21. Miller, N., and Cerutti, P. (1968) Structure of the Photohydration Products of Cytidine and Uridine, *Proc Natl Acad Sci U S A* **59**, 34-38.

22. Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R. B. (2003) CLIP identifies Nova-regulated RNA networks in the brain, *Science* **302**, 1212-1215.

23. Tabone, T., Sallmann, G., Chiotis, M., Law, M., and Cotton, R. (2006) Chemical cleavage of mismatch (CCM) to locate base mismatches in heteroduplex DNA, *Nat Protoc* **1**, 2297-2304.

24. Johnston, B. H., and David M.J. Lilley, J. E. D. (1992) Hydroxylamine and methoxylamine as probes of DNA structure, in Methods in Enzymology, pp 180-194, Academic Press.

25. Hayatsu, H. (1976) Bisulfite modification of nucleic acids and their constituents, *Prog Nucleic Acid Res Mol Biol* **16**, 75-124.

26. Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994) High sensitivity mapping of methylated cytosines, *Nucleic Acids Res* **22**, 2990-2997.

27. Wang, R. Y., Gehrke, C. W., and Ehrlich, M. (1980) Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues, *Nucleic Acids Res* **8**, 4777-4790.

28. Wadkins, T. S., Shih, I., Perrotta, A. T., and Been, M. D. (2001) A pH-sensitive RNA tertiary interaction affects self-cleavage activity of the HDV ribozymes in the absence of added divalent metal ion, *J Mol Biol* **305**, 1045-1055.

29. Luptak, A., Ferre-D'Amare, A. R., Zhou, K., Zilm, K. W., and Doudna, J. A. (2001) Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme, *J. Am. Chem. Soc.* **123**, 8447-8452.

30. Heinicke, L. A., and Bevilacqua, P. C. (2012) Activation of PKR by RNA misfolding: HDV ribozyme dimers activate PKR, *RNA*.

31. Su, L., Chen, L., Egli, M., Berger, J. M., and Rich, A. (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot, *Nat Struct Biol* **6**, 285-292.

32. Siegfried, N. A., O'Hare, B., and Bevilacqua, P. C. (2010) Driving forces for nucleic acid pK(a) shifting in an A(+).C wobble: effects of helix position, temperature, and ionic strength, *Biochemistry* **49**, 3225-3236.

33. Dierick, H., Stul, M., De Kelver, W., Marynen, P., and Cassiman, J. J. (1993) Incorporation of dITP or 7-deaza dGTP during PCR improves sequencing of the product, *Nucleic Acids Res* **21**, 4427-4428.

34. McConlogue, L., Brow, M. A., and Innis, M. A. (1988) Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine, *Nucleic Acids Res* **16**, 9869.

35. Goodsell, D. S. (2001) The molecular perspective: ultraviolet light and pyrimidine dimers, *Oncologist* **6**, 298-299.

36. Woolley, A. T., and Mathies, R. A. (1995) Ultra-High-Speed DNA Sequencing Using Capillary Electrophoresis Chips, *Analytical Chemistry* **67**, 3676-3680.

VITA

Jennifer L. Wilcox

Education

The Pennsylvania State University — Dec. 2013
University of Scranton — May 2007

Publications

Jennifer L. Wilcox, Bevilacqua, P.C. pK_a Shifting in Double Stranded RNA is Highly Influenced by Nearest Neighbors and Bulge Locations. (*submitted*).

Jennifer L. Wilcox, Bevilacqua P.C. A Simple Fluorescence Method for pK_a Determination in RNA and DNA Reveals Highly Shifted pK_a's. *J. Am. Chem. Soc.* **2013**, 135, 7390-7393.

Jennifer L. Wilcox**,** Ahluwalia, A.K., Bevilacqua, P.C. Charged Nucleobases and their Potential for RNA Catalysis. *Accounts Chem. Res.* **2011,** 44, 1270-1279. (*peer reviewed*)

Chadalavada, D. M., Cerrone-Szakal, A. L., Jennifer L. Wilcox, Siegfried, N. A., Bevilacqua P.C. Mechanistic Analysis of the Hepatitis Delta Virus (HDV) Ribozyme: Methods for RNA preparation, Structure Mapping, Solvent Isotope Effects, and Co-Transcriptional Cleavage. *Methods Mol. Bio.* **2012**, 848, 21-40.

Selected Honors

Outstanding Oral Presentation, Rustbelt RNA Meeting — October 2012
RNA Society Meeting Annual Travel Award — May 2012
Penn State Chemistry Graduate Student Leadership and Service Award — December 2011
Dan H. Waugh Memorial Teaching Award — November 2009
Penn State Chemistry Graduate Student Travel Award — 2009 and 2012

Presentations

Jennifer L. Wilcox, Bevilacqua, P.C. "Protonated Bases in Nucleic Acids- Where and Why Do They Occur?" *Speaker Presenter* at the Pennsylvania State University Chemistry Department Lion Lecture Series. University Park, PA, December 5, 2012

Jennifer L. Wilcox, Bevilacqua, P.C. "A Simple Fluorescence Method for pK_a Determination in RNA and DNA Reveals Highly Shifted pK_a's" *Speaker Presenter* at the Rustbelt RNA Meeting. Dayton, OH, October 19-20, 2012- Outstanding Oral Presentation

Jennifer L. Wilcox, Bevilacqua, P.C. "Use of Fluorescence Spectroscopy for High-Throughput Quantification of pK_a Shifting in RNA" *Poster Presenter* at the RNA Society Annual Meeting. Ann Arbor, MI, May 29-June 3, 2012

Jennifer L. Wilcox, Bevilacqua, P.C. "Investigating the Roles of Protonated Bases in Nucleic Acids" *Speaker Presenter* at the Pennsylvania State University Center for RNA Biology RNA Club Series. University Park, PA, April 20, 2012

Jennifer L. Wilcox, Siegfried, N.A., Bevilacqua, P.C. "Determination of Driving Forces for pK_a Shifting in RNA and DNA Model Systems" *Poster Presenter* at the North Carolina RNA Society Symposium on RNA Biology VIII: RNA Tool and Target. Research Triangle Park, NC, October 16-17, 2009.