UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# NUCLEAR MAGNETIC RESONANCE STUDY OF RIBONUCLEIC ACID TERTIARY STRUCTURES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

STEVEN HARRIS
Norman, Oklahoma
2011

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/319469

NUCLEAR MAGNETIC RESONANCE STUDY OF RIBONUCLEIC ACID TERTIARY STRUCTURES

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY

Dr. Susan J. Schroeder, Chair

Dr. Ben F. Holt III

Dr. Charles V. Rice

Dr. Wai Tak Yip

Dr. Elena I. Zgurskaya

© Copyright by STEVEN M. HARRIS 2011
All Rights Reserved.

## ACKNOWLEDGEMENTS PAGE

First, I would like to thank my advisor and research mentor, Dr. Susan J. Schroeder for sharing her knowledge, encouragement and most of all for believing in me. I am appreciative of the opportunities that were afforded to me that helped me become a better researcher, scientist and person. My success is dependent on my laboratory, oratorical as well as my writing skills. I am grateful that I was able to have her guidance in improving these aspects. I am also thankful to Drs. B. Holt III, C. Rice, W. T. Yip, and H. Zgurskaya for serving on my advisory committee.

I would like to thank all of the members of Dr. Schroeder's research group: Dr. Xiaobo Gu, Dr. Negin Mirhosseini, Koree Clanton-Arrowood, Mai-Thao Nguyen, Jonathan Stone, Samuel Bleckley, Gabriel Campbell, Abigail Overacre, Jessica Mackiewicz, and Joshua Malone. I am also thankful for the friendship of Gopal Abbineni, Garry Chapman, Jeffrey Hayle, and Jon Henrickson.

I would like to thank the National Science Foundation, Bridge to Doctorate Fellowship for the funding that I received as well as the continued support from their faculty and staff. I would also like to thank the Multicultural Engineering Program for their support as well.

Finally, I would like to thank all of my family and close friends who helped me along my journey. My mother and father, Carol and Melvin Harris, Sr., always provided for me in all aspects of life. My grandfather, who has continually been a positive role model in my life. My two brothers and sister as well as my extended family for the support and encouragement. My love, Kerriann Mitchell, and great friends Q. Hughes, C. Harlin, R. Jones, R. Allen, Q. Gentry, C. Rose, F. Robertson, B. Harris and L. Magwood.

## TABLE OF CONTENTS


Page

**List of Tables** viii

**List of Figures** ix

**Abstract** xi

**Chapter 1**

**Introduction**

1.1 Ribonucleic Acid at a Glance 1

1.2 Structure prediction algorithms 3

1.3 RNAstructure, Isostericity Matrices and the MC-Pipeline 6

1.4 Nuclear Magnetic Resonance 11

1.5 Crystallography and NMR Systems 14

1.6 The Φ29 pRNA E-loop Hairpin 16

1.7 Consecutive G-U Base Pairs at the Ends of Helices 23

**References** 29

**Chapter 2**

**Tertiary Structure Determination of the Φ29 E-loop Hairpin**

**2.1 Introduction** 36

**2.2 Materials and Methods**

2.2.1 RNA Preparation and Purification 39

2.2.2 UV Experiments 40

2.2.3 NMR Spectroscopy 40

2.2.4 NMR Restraints and Peak Assignments 41

2.2.5 Structure Calculations 42

2.2.6 Structure Predictions with MC-SYM 43

**2.3 Results**

2.3.1 UV Absorbance 44

2.3.2 $H_2O$ NOESY 44

2.3.3 $D_2O$ NOESY 47

2.3.4 DQFCOSY 55

2.3.5 $^{1}H$- $^{13}C$ HSQC 56

2.3.6 $^{1}H$-$^{31}P$ HETCOR 59

2.3.7 Hairpin Structure 61

**2.4 Discussion**

2.4.1 Structural Features of the Φ29 E-loop Hairpin 66

2.4.2 Comparison of Prohead and Ribosomal Hairpin Structures 67

2.4.3 Comparison of Φ29 NMR Hairpin Structure to MC-SYM Structure Predictions 68

2.4.4 Comparison of Hairpin Structures Determined by NMR to Predictions from Isostericity Matrices 74

**2.5 Conclusions** 76

**References** 78

**Chapter 3**

**Tertiary Structure Determination of Consecutive Terminal G-U pairs: 5'r(GUGUCGGUGU)$_2$**


| | | |
|---|---|---|
| **3.1** | **Introduction** | 88 |
| **3.2** | **Materials and Methods** | |
| 3.2.1 | RNA Preparation and Purification | 91 |
| 3.2.3 | NMR Spectroscopy | 91 |
| 3.2.4 | NMR Restraints and Peak Assignments | 93 |
| 3.2.5 | Structure Calculations | 94 |
| **3.3** | **Results** | |
| 3.3.1 | $H_2O$ NOESY | 95 |
| 3.3.2 | $D_2O$ NOESY | 99 |
| 3.3.3 | DQFCOSY | 108 |
| 3.3.4 | $^{1}H$- $^{13}C$ HSQC | 111 |
| 3.3.5 | $^{1}H$-$^{31}P$ HETCOR | 114 |
| 3.3.6 | $^{1}H$-$^{15}N$ HSQC | 117 |
| **3.4** | **Discussion** | |
| 3.4.1 | Structural Features of the 5'r(GUGUCGGUGU)$_2$ duplex | 118 |
| 3.4.2 | Quantitative Analysis of G-U Pairs in Various Helices | 122 |
| 3.4.3 | Comparison of the Solution Structure to MC-SYM Predictions | 125 |
| **3.5** | **Conclusions** | 133 |
| | **References** | 134 |

## LIST OF TABLES

### Chapter 2


| | | |
|---|---|---|
| 2.1 | Proton Resonance Assignments | 51 |
| 2.2 | NOE Assignments | 52 |
| 2.3 | Carbon and Phosphorus Resonances | 60 |
| 2.4 | Restraints and Ensemble Convergence for Φ29 E-Loop Hairpin | 64 |
| 2.5 | Preliminary Hairpin Refinements | 65 |
| 2.6 | Comparison of Prohead and Ribosomal Hairpin Structures | 70 |
| 2.7 | TSP Referencing Table | 71 |


### Chapter 3


| | | |
|---|---|---|
| 3.1 | Proton Resonance Table | 101 |
| 3.2 | Restraints for 5' r(**GUGU**CG**GUGU**)$_2$ | 102 |
| 3.3 | NOE Assignments | 103 |
| 3.4 | Carbon and Phosphorus Resonance Table | 116 |
| 3.5 | Helical Parameters for G-U Base Pairs | 128 |

## LIST OF FIGURES

### Chapter 1


1.1 Isostericity Matrices 8

1.2 Phage Φ29 Cryo-EM Data 20

1.3 Phage Φ29 pRNA Secondary Structures 21

1.4 Φ29 E-loop Secondary Structure 22

1.5 G-U Base Pairs 27


### Chapter 2


2.1 Hairpin Primary Sequences 37

2.2 E-loop Imino Proton Spectrum 45

2.3 2-Dimensional Imino Proton Spectrum 46

2.4 Aromatic Proton to H1’ Proton NOESY Sequential Walk 48

2.5 50 ms $D_2O$ NOESY Spectrum 49

2.6 NOESY Showing N+2 NOE 50

2.7 Double Quantum Filtered COSY Spectrum 55

2.8 $^{13}C$ Ribose HSQC Spectrum 57

2.9 $^{13}C$ Aromatic HSQC Spectrum 58

2.10 $^{31}P$ HETCOR Spectrum 59

2.11 Φ29 E-loop Hairpin Wall-eye Stereo Image 62

2.12 Φ29 E-loop Hairpin Overlay Images 63

2.13 Hairpin Tertiary Structure Comparisons 72

2.14 2-Dimensional Hairpin Comparisons and Contacts 73

2.15 Hairpin Backbone Overlays 74

**Chapter 3**


| | | |
|---|---|---|
| 3.1 | The Two Most Common G-U Base Pairs | 90 |
| 3.2 | Imino Proton $H_2O$ NOESY Spectrum | 97 |
| 3.3 | Exchangeable Proton $H_2O$ NOESY Spectrum | 98 |
| 3.4 | $D_2O$ NOESY Spectrum | 100 |
| 3.5 | DQF COSY Aromatic Spectrum | 108 |
| 3.6 | DQF COSY H1'-H2' Spectrum | 109 |
| 3.7 | Ribose Dihedral Angles | 110 |
| 3.8 | Sugar Pucker Conformations | 111 |
| 3.9 | $^{13}C$ Aromatic HSQC Spectrum | 112 |
| 3.10 | $^{13}C$ Ribose HSQC Spectrum | 113 |
| 3.11 | $^{31}P$ HETCOR Spectrum | 115 |
| 3.12 | $^{15}N$ Imino HSQC Spectrum | 117 |
| 3.13 | Stereo image of 5'r$(GUGUCGGUGU)_2$ Duplex | 120 |
| 3.14 | Lambda Angle | 124 |
| 3.15 | Solution and Predicted Structure Backbone Overlay | 127 |

## ABSTRACT

Ribonucleic acid (RNA) is ubiquitous to all organisms and a molecule that essential to all life forms. RNA is a versatile molecule with numerous cellular functions. With advances in nucleic acid sequencing, researchers are finding that there is an increasing number of RNAs that participate in non-coding cellular roles. The most logical means of discerning the function of these RNAs is to exploit the structure-function relationship that exists. The three-dimensional structure of RNA holds a wealth of information that gives insight into its molecular functioning scheme.

At this present time, the most common methods of solving the tertiary structure of RNA is x-ray crystallography and nuclear magnetic resonance spectroscopy. These two methods are very time consuming. If tertiary structures could be predicted from sequence, this would relieve the bottleneck in determine the function of RNAs. From the tertiary structure, better hypotheses could be formulated and less experimental work would need to be done to verify them.

Tertiary structure prediction of RNA is progressing but there is still room for improvement. Since prediction programs can predict Watson-Crick base pairs well, improvements must focus mainly on interactions involving non-canonical or unpaired bases. There is much more information to be learned from tertiary structures that will help improve the current prediction programs. Therefore, solving tertiary structures of RNA is of extreme importance in order to better the prediction ability from sequence.

This research presents the tertiary structure of two RNA sequences. The first is a RNA hairpin, which is one of the most common RNA secondary structures. Hairpins are believed to aid in the folding of large RNA and sometimes help keep the molecule from falling into kinetic traps. If this happens, the RNA molecule is usually in a non-functional state and is rendered useless. Improving our ability to predict hairpins will essentially help improve the prediction of larger molecules as well. This hairpin is a part of a nanomolecular motor that is used by a phage and incorporates the genome back into an empty viral capsid.

The *Bacillus subtilis* phage phi29 packaging motor requires prohead RNA for genome encapsidation. The NMR structure of the prohead RNA E-loop hairpin, r(5'AUUGAGUU), and MC-SYM predictions for prohead RNA E-loops provide a basis for comparative analysis of hairpin structures in prohead and ribosomal RNA. All the hairpins contain a U-turn motif but differ in the first non-canonical pair and backbone orientation. These structures provide benchmarks for further improvements in RNA structure predictions from sequence as well as provide the first piece of structural information about the phi29 pRNA molecule.

G-U base pairs are the most common non-canonical base pairs. They have been found to be conserved throughout numerous phylogenetic alignments and have been proven to be more than simply replacements for Watson-Crick base pairs. The tertiary structure that is presented in this research is a self-complimentary duplex 5'r(**GUGU**CG**GUGU**)$_2$ and is the first NMR solution structure with consecutive terminal G-U pairs. The stability of G-U pairs is

variable depending on where they are in the sequence. In the center of a duplex, they have been found to be destabilizing to the helix. On the ends of helices, they have been found to be thermodynamically stabilizing to the duplex. The tertiary structure of the duplex was found to be nearly A-form. This is a surprising result because it is comprised of mostly non-canonical base pairs. The helical parameters of the solution structure were then compared to those of internal G-U pairs to pinpoint the differences in base interactions and provide a hypothesis for the differences in stabilities of internal versus terminal G-U pairs. It was found that the internal parameters are, in fact, different from the terminal G-U pairs. These results were then compared to the helical parameters of the structure predicted by the tertiary structure prediction program, MC-SYM. Results show that the prediction has helical parameters that matched the trends seen for internal G-U base pairs even though the solution structure did not. This prediction results are not surprising since there were no previous solution structures containing consecutive terminal G-U base pairs. These results also show the importance of solving the tertiary structures of duplexes such as this in order to aid in the diversification of the databank of known tertiary structures.

# CHAPTER 1

# INTRODUCTION

## 1.1 Ribonucleic Acid at a Glance

Ribonucleic acid (RNA) is a molecule that is essential for the existence of all organisms. The cellular roles of RNA are copious and range from information transfer and regulation of gene expression to a host of other functions such as cleavage and ligation of nucleic acids to name a few (*1-5*). Much emphasis has been placed on studying RNA since the catalytic properties were discovered by Cech(*6*) and Altman(*7*) in the early 1980's proving that RNA was more than a carrier of genetic information. Research done on RNA has recently resulted in several Nobel Prize awards in Chemistry and Medicine. Those for Chemistry include Ramakrishnan(*8-11*), Steitz(*12-16*) and Yonath(*17-20*) in 2009 for studies of the structure and function of the ribosome and Kornberg(*21-24*) in 2006 for his studies of the molecular basis of eukaryotic transcription. In medicine, the 2009 prize was awarded to Blackburn(*25-28*), Greider(*29-32*) and Szostak(33-35) for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase and the 2006 prize went to Fire(*36-40*) and Mello(*36-37, 39, 41*) for their discovery of RNA interference - gene silencing by double-stranded RNA. This shows the impact and relevance that RNA research has presented to the sciences, and there are still new functions and applications of RNA that are yet to be discovered. The subsequent paragraphs and chapters will show the importance of studying the tertiary structures of the phi29 E-loop hairpin and

duplex containing terminal G-U base pairs as well as how it will impact the general sciences.

RNAs are intimately associated with a wide variety of biological functions, and these functions can be directly related to their three-dimensional or tertiary structures. Knowledge of RNA three-dimensional structure can help formulate better hypotheses of mechanisms of function and less experimental work to verify these hypotheses. Understanding the molecular basis of RNA structures is of extreme importance so that the function can be verified but requires a fundamental knowledge of the interactions that stabilize RNA duplexes. Furthermore, proper application of this knowledge can lead to new breakthroughs in chemistry, molecular biology, and genomics.

At this present time, RNA sequencing is abundant, but the “bottleneck” in the pathway from sequencing information to function is accurately discerning the tertiary structure. Since we are currently unable to accurately predict the active tertiary structure of RNA molecules from their primary sequences, we must rely mainly on the two proven methods of structure determination, Nuclear Magnetic Resonance Spectroscopy and X-ray Crystallography. Though there is current research being done that continues to push the limits of the aforementioned methods, these techniques will not be able to keep pace with the rapidly growing database of sequencing information. Thus, improving the structure prediction algorithms will prove beneficial in the process of designing experiments to explore RNA function. Even with low accuracy, these prediction algorithms can help suggest new experiments to perform that will help probe RNA structure as

well as give insights into the RNAs functional operating schemes. The research in the subsequent chapters focuses on improving our ability to predict tertiary structures by solving the solution structures of a hexa-nucleotide hairpin and a duplex with consecutive terminal G-U base pairs. These studies will provide new NMR solution structures as well as quantitative structural comparisons of predicted and solution structures. This information will aid in improving current prediction algorithms by pinpointing specific interactions that are consistently predicted incorrectly and by providing more solution structures for training the knowledge-based programs.

## 1.2 Structure prediction algorithms

There are many different types of structure prediction programs that base their predictions on various criteria and attempt to predict secondary and tertiary structures from the primary sequence. The most common methods of prediction include training the program on known, previously solved structures (knowledge-based), and others base their predictions on Turner thermodynamics nearest-neighbor models (physics-based). The latter technique seems to have the most potential because the predictions are based on a biophysical model and research shows that secondary structures can be predicted based on minimizing the free energy of the RNA molecule(*42-47*). The biophysical model also allows the measurement of smaller fundamental units as opposed to the knowledge-based programs that normally have a larger empirical unit. Watson-Crick base pairs are one of the most important motifs in RNA secondary structures(*48*) and

researchers have spent decades compiling the data that gives the physics based programs the ability to predict motifs containing canonical base pairs very well. But even G-U base pairs, which are the most common non-canonical base pair interaction, has yet to be well-understood. Some RNA sequences, especially those involving G-U base pairs, are difficult to predict because they are dependent on non-nearest-neighbor effects, such as location in the helix (*49-53*). These discrepancies, along with the fact that the thermodynamics tables for prediction algorithms are incomplete, lead to limitations in predictions of the free energy and the secondary structure.

Authors of the programs that predict structures understand that RNA molecules primary, secondary and tertiary structures follow an hierarchical arrangement. The energy gained upon forming the secondary structure is separable from that of the tertiary structure. Unlike proteins, the secondary structure is stable enough to stand alone so the tertiary structure can be viewed as an addition to the energy of the secondary structure. Thermodynamic data show that after the secondary structure of RNA is formed, additional tertiary interactions can come about with very little alteration of the secondary structure(*54*). Proteins generally gain a similar amount of free energy whether it is a secondary or tertiary interaction making the hierarchical approach impractical (*55*). The RNA sequence of bases, primary sequence, lays the foundation for the formation of the secondary structure, which consists of base pairing to form helices, hairpins, internal loops, junctions, bulge loops, etc. It has been shown that the secondary structure of RNA can be predicted successfully from

experimental thermodynamic data since the most likely secondary structure is the structure that has the lowest free energy. The thermodynamics behind the formation of the secondary structure are important in helping gain a better understanding of the structure-function relationships that exist in RNA molecules. Thermodynamics can help provide a basis for predicting secondary structure and overall stability, which can then assist in the determination of RNA function. Knowing this, it is imperative that the calculations that are used to predict the secondary structure are as accurate as possible because these structures are the intermediate step on which the tertiary structures are based. Current programs boast an average secondary structure prediction accuracy of 73% for the minimum free energy structure and 86% for the best sub-optimal structure on motifs fewer than 700 nucleotides. Prediction accuracy decreases as the number of nucleotides increase. The range of accuracies varies for different types of RNA. For RNase P, only 55% of the base pairs were predicted correctly, but 88% of base pairs were correctly predicted for group II introns(*42*). One addition that can help increase the accuracy of the predictions is to include experimental data such as chemical modification(*56*). This information aids prediction results by pinpointing the bases that are solvent accessible(*57*), which means they are in a single strand, at the end of a helix, or adjacent to a G-U base pair.

## 1.3 RNAstructure, Isostericity Matrices and the MC-Pipeline

As was previously stated there are both knowledge-based prediction programs such as CONTRAfold(*46*), and FARNA(*58*) and MC-SYM(*59*) (included in the MC-PIPELINE suite), as well as physics-based types like RNAStructure(*56*), Mfold(*60*), RNAfold(*61*), and Sfold(*45*). RNAStructure, isostericity matrices(*62*) and the MC-Pipeline were used as analytical and comparative tools on which to base the conclusions that were made in this research. RNAStructure is physics-based, isostericity matrices are knowledge-based and the MC-PIPELINE uses both physics-based (MC-FOLD) as well as knowledge-based (MC-SYM). These tools were chosen because they are at the forefront for prediction programs of the same type and help classify all base-to-base interactions, including non-canonical interactions which will aid in validating the results of these studies. MC-SYM was chosen because, unlike other tertiary structure prediction programs, the secondary structure predictions produced by MC-FOLD are completed using turner thermodynamics. The knowledge-based programs were also chosen because both programs were trained on data from the crystal structures of the ribosome that were deposited in the PDB. The comparisons of the experimental and predicted structures for the pRNA E-loop and ribosomal hairpins S2 and L11 (chapter 2) which are similar in sequence, test the applicability of MCSYM and isostericity matrices in predicting RNA structure. Similarly, the experimental structure of the duplex with the terminal GU pairs will be compared to single, double, and triple internal G-U pairs and analyzed using the same procedure.

RNAStructure is a Windows based re-implementation of one of the oldest secondary structure prediction programs, MFOLD which uses the Zuker algorithm. RNAStructure only requires the user to input the primary sequence, then a minimum free energy structure as well as a user chosen number of sub-optimal secondary structures are produced. This algorithm utilizes one of the most recent set of nearest neighbor parameters for determining the free energy at 37°C for an RNA secondary structure. The main assumptions that are utilized by free energy minimization prediction programs are that the free energies of stems depend on nearest-neighbor sequences (INN-HB) of base-pairs only and that free energies of stems and loops are additive(*63*). Therefore, adding the free energies of all of the helices, loops, etc. and minimizing them will produce the actual secondary structure. This program was used to help design the starting constructs in both the prohead RNA and G-U pair project and was also used to form a better understanding of some of the criteria that need more attention when attempting to predict secondary structures.

**Figure 1.1: Isostericity Matrices**

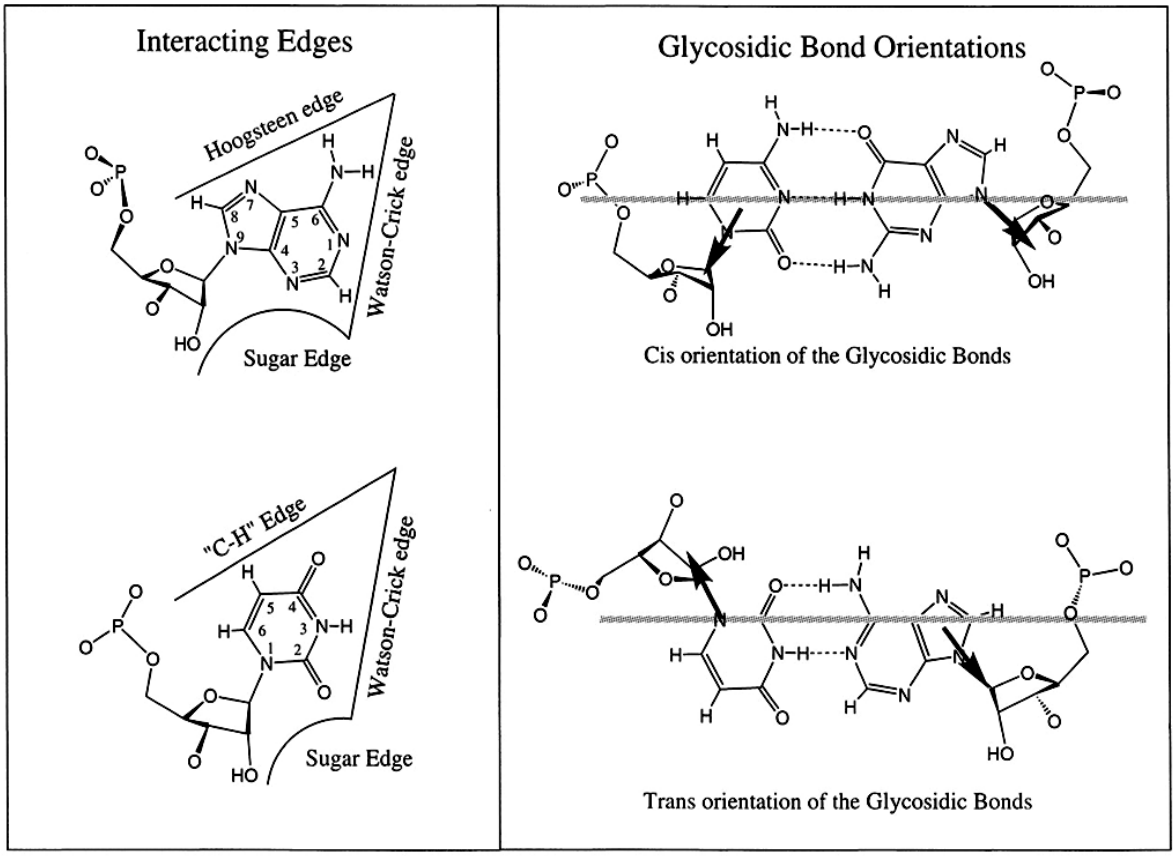



**Figure 1.1**: Isostericity Matrices. The left portion of figure shows the interacting edges that can participate in a base to base interaction. The right portion of the figure shows how determine the glycosidic bond orientation and if the bases are interacting in a *cis* or *trans* manner. Figure adopted from figure 1 Nucleic Acids Res. 2002 August 15; 30(16): 3497–3531.

Isostericity matrices are a concept that was proposed by *Leontis and Westhof* in 2002. In their studies, they used crystallography data from the Protein Data Bank (PDB) to attempt to classify all base-to-base interactions. In the isostericity matrices, three main criteria are used to unambiguously assign all interactions. The first is the edge or face through which the two bases interact. On each base there are three edges; the Watson-Crick (WC), Hoogsteen/C-H, and sugar edges (Figure 1.1). The W-C edge is used by two bases in a normal W-C base pair. The sugar edge covers the face of the base that is closest to the ribose and the Hoogsteen/C-H is the remaining opposing side. The authors give symbols to each side. The sugar edge is denoted by using a triangle. The Hoogsteen/C-H edge is a square, and the W-C edge is a circle. The next measure that is used in the classification is whether the bases are interacting in a *cis* or *trans* manner. In order to determine the orientation; first, a line is drawn planar to the bases (See Figure 1.1). Then, if the ribose are located on the same side of the line the orientation is considered *cis*, if not, then it is considered *trans*. The last and most important measure that is used in the classification is the C1' to C1' distance. In this case, the distance from the C1' atom on the ribose of one base to the C1' atom on the opposing base is measured. If two pairs of interacting bases have the same orientation, *cis* or *trans*, and the same C1' to C1' distance, they are said to be isosteric. The term is used because the isosteric bases will consume the same amount of space when placed into the helix. It is well known that the two canonical W-C base pairs are isosteric and thus can be stacked upon each other to form a regular A-form helix. At this time,

not all base to base interactions have been observed throughout all of the crystal structures that were analyzed from the PDB.

MC-PIPELINE is a knowledge-based software suite that includes programs that perform various tasks for RNA structure and structure prediction programs. The two programs from the suite that were used in this research are called MC-FOLD and MC-SYM these programs work in concert to sequentially produce secondary and tertiary structure predictions from primary sequence. The MC-FOLD algorithm produces the secondary structure and this output can then be directly put into the MC-SYM tertiary prediction program. Because this suite is knowledge-based, it also incorporates data that has been inputted into the PDB, more specifically, the X-ray crystallographic structure of the 23S ribosomal RNA of *Haloarcula marismortui*. From this information, they were able to define a new empirical unit, nucleotide cyclic motifs (NCMs), to base all of their predictions. A NCM is composed of at least three nucleotides and allows the program to classify and score all nucleotide interactions. Neighboring NCMs have at least one overlapping base pair and this feature helps give enough information about the base to base interactions to develop an efficient function for scoring. Another characteristic of the NCMs is that it naturally encompasses both the canonical and non-canonical interactions.

The predictive power of these programs was verified by comparing the tertiary structures that were predicted to experimentally determined 3D structures from sequence. The 3D structures were less than ~50 nucleotides and included small hairpins ranging in size from 4-7 base pair loops, internal ribosomal entry

sites, group II intron branchsites, hammerhead ribozymes as well as small telomerase domains. The predicted tertiary structures matched the minimum free energy structures in 11 of 13 structures (*59*). The data obtained from this research was analyzed and compared to the predictive power of the RNAStructure program, the isostericity matrices and the programs from the MC-PIPELINE software. This research will not only test the programs ability to predict structures, but it will also help identify areas of improvement.

## 1.4 Nuclear Magnetic Resonance (NMR)

The two most common tertiary structure determination techniques, NMR and x-ray crystallography, have advantages and disadvantages . NMR was chosen because it is more suitable for determining the structures of the small, dynamic RNAs, especially hairpins due to frequent structure disruption by crystal packing forces. Many researchers that use NMR know that there are a number of factors that limit the ability to solve the solution structure of a large molecule, over 12 kiloDaltons. These factors include proton overlap, changes in relaxation, and the ability to get a concentration of the sample into the NMR tube that is high enough to study(*64*). Proton overlap is a limitation partly because of the sheer number of protons, namely ribose protons, which resonate in ~1 ppm range causing severe overlap. This results in ambiguity in the peak assignment process so RNA structures larger than ~150 nucleotides are seldom attempted by NMR. In this research, the structures that were elucidated were only up to 17

nucleotides in length. The instruments that were used are easily capable of determining structures much larger in size.

Another reason why NMR was chosen over x-ray crystallography is because NMR gives a dynamic structure rather than a static structure. In NMR, the molecule is tumbling in solution and the range of motion is not limited by close contacts or the solvent. In crystallography, the molecule is packed into a crystal which brings up two major disadvantages that are associated with this technique. The first disadvantage and bottleneck in crystallography is actually getting a crystal. For small RNAs, this process can be very difficult as well as time consuming. In order to get a molecule to crystallize, it must form a well ordered array and the flexibility of RNA can make getting a crystal extremely complex. The second disadvantage, crystal packing forces, can be caused upon simply forming a crystal. This occurs when neighboring molecules interact with each other and sometimes help form the crystal. This occasionally causes the molecule to assume a non-native conformation, especially for small hairpins. The advantages afforded by NMR makes it more feasible for the structure determination of these small RNAs.

The NMR tertiary structure determination methods are well developed for biological macromolecules(*65-66*). The NMR data in this research were collected on Varian Inova 500 and 600 MHz spectrometers to gain structural information about the oligomers. The first experiment that is performed is the 1-dimensional imino proton experiment. Imino protons are exchangeable protons that are connected to the N-1 atom in a guanosine and the N-3 atom in the uracil

(Figure 1.5). These protons exchange with water, which causes their signal to broaden out and not be visible in normal, unbound conditions. When these protons are in a stable conformation, like a W-C base pair, they are protected from exchange and their signal can be observed. This experiment is also important because it allows preliminary identification of good spectra. Good spectra have very little overlap and show strong, sharp peaks that can be unambiguously assigned.

The next step is to perform a series of two-dimensional experiments so that the proton peaks may be assigned and used as distance and angle constraints. The Nuclear Overhauser Effect Spectroscopy, NOESY experiment is a through-space experiment that is performed to gain distance constraints for the protons in the structure. These experiments can be done with varying amounts of time for proton-proton interaction. This helps distinguish the strong, medium and weak interactions which are directly related to distance and used as constraints. The proton peaks are sequentially assigned and this process can be verified through other experiments.

The other two important experiments are the Correlation Spectroscopy , COSY, and Heteronuclear Single Quantum Coherence, HSQC. Both are through-bond experiments and give information on the type of covalent bond that is involved in the interaction. Unlike the NOESY experiments, they may involve the interaction of more than one type of nuclei. The COSY experiment can be used to determine strong covalent interactions and more specifically the H-5 to H-6 interaction on a pyrimidine base. This information distinguishes the purine

from the pyrimidine bases and aids in confirming the assignment of the NOESY spectra. Backbone angle information may also be obtained from the COSY experiment and used as constraints in the molecular modeling. The HSQC is very useful in helping distinguish the aromatic and the ribose protons from each other by differences in the carbons to which they are connected. Depending on the complexity of the structure, further experiments may need to be performed to gain additional constraints or assignments. This data is then compiled into constraint files for structure determination using the simulated annealing software, CNS.

## 1.5 Crystallography and NMR Systems (CNS)

Once data are collected from the NMR experiments, the next step is to perform computer simulations in which the data are collectively modeled into a tertiary structure. This was done using a software suite for macromolecular structure determination entitled Crystallography and NMR Systems (CNS) which was developed by an international collaboration of researchers that was led by Dr. Axel Brunger at Yale University(*67*). CNS is an updated version of the X-plor software that was developed earlier by Dr. Brunger. The authors of the program state that there are four goals of the program. The first is to create an easy and flexible framework that allows researchers the ability to explore new structure determination approaches. The next goal of the program is to improve the tools that allow the solution structure determination of large or difficult molecules. The authors also sought to develop models that aid in gaining macromolecular

information pertaining to structure and dynamics. And lastly, they wish to integrate all information gathered by the user into the structure determination process. The suite is a general system for NMR solution as well as x-ray crystallography structure determination of nucleic acids, proteins, carbohydrates, and their complexes. Other molecules such as water, ligands, and other ions can be included into the simulations. CNS is freeware and the source code is available in its entirety. The current CNS software is for UNIX-based operating systems and uses fortran, and C compilers to read “CNS language”, which is a symbolic type of language. Due to the accessibility of the source code, development of new algorithms can be easily done in the CNS language. CNS also has an HTML interface which provides a user-friendly means of operation and makes it easy for new users to perform standard simulations. It is available for most computing platforms; however it is only supported for a few (DEC, SGI, PC, HP).

CNS is a simulated annealing program or a computer algorithm that attempts to imitate the process of annealing. RNA annealing is when the oligomer is heated above its $T_m$ so that it can sample different conformations, then slowly brought down in temperature. As the RNA is cooled, this allows the RNA the opportunity to assume a lower energy state which usually produces a more optimal structure. In simulated annealing, computer algorithms imitate this process. To emulate heating, the kinetic energy of the oligomer is increased. This process also allows the oligomer to sample alternative conformations. To emulate cooling, kinetic energy is removed and the energy is minimized. This

process is repeated several times with several different starting structures. When the simulation is done properly, the structures will converge and look similar. The final structure is usually accepted as an average of the set of converged structures that agreed with all of the experimental data. CNS is an all inclusive program and only needs the assigned NMR data as well as structure files. These files are used in conjunction with the AMBER force field to complete the structure determination. The program allows many options pertaining to how the simulated annealing can be done and includes the options of incorporating J-coupling, NOE, chemical shift as well as dihedral angle restraints. Once the experimental and NMR data are collected, CNS will be used to produce and minimize the final tertiary structures.

### 1.6 The Φ29 pRNA E-loop Hairpin

The Φ29 bacteriophage is among a family of phages that that infect the bacterial host *Bacillus subtilis*. This family is comprised of 8 phages that have a linear double stranded DNA genome of 20 kilobases(*68*). During the packaging process, the genome is repackaged into an empty preformed viral capsid by means of an adenosine triphosphate (ATP) dependent nanomolecular motor(*69-72*). The Φ29 packaging motor has been well studied because it can undergo *in vitro* reconstitution so it can easily be used as a model for DNA packaging(*73-75*). It is unique in composition because unlike other motors it utilizes an RNA that is essential to the function of the motor. Many other phages use packaging motors but they are comprised of all proteins. This unique RNA, called prohead

RNA (pRNA), forms a pentameric ring along with other protein complexes to complete the motor (Figure 1.2)(*74, 76-79*) Figure 1.2 shows the cryo-EM data of the pRNA connected to the viral connector protein as well as the viral ATPase. Figure 1.2 also shows the location of the complete motor in relation to the capsid.

The pRNA molecules from different species in the family of phages are diverse in their primary sequence but share a highly conserved secondary structure (Figure 1.3)(*80*). Since the pRNA have the same function and share a conserved secondary structure, the hypothesis is that the tertiary structures of the pRNAs from the different species should also be conserved. To test this hypothesis, the tertiary structure of small portions of the pRNAs from different species will be determined by NMR and compared to one another. Using methods previously described in the introduction, the results of the structures will be compared on how isosterically similar they are. A quantitative comparison of how well the tertiary structure prediction program was able to calculate the structure will be done using an all-atom RMSD of the structures as well as the number of violated NOEs that were experimentally determined using NMR.

The portion of the pRNA that was chosen for tertiary structure determination was the E-loop hairpin (Fig. 1.4). The pRNA E-loop hairpin sequence is conserved in Φ29, M2/NF, and SF5 pRNA sequences (*80*). The pRNA E-loop hairpin interacts with gene product 10, the viral head-tail connector protein, a 290 amino acid monomer which forms a 422 kDa dodecameric complex (*81-85*). Mutations in the pRNA E-loop hairpin reduce protein binding to less than 30%, and mutation of the A nucleotide in the hairpin loop to a C

reduces packaging activity 95% (*86*), although substitution of this hairpin with a UUCG tetraloop retains some packaging activity *in vitro* (*87*). Nucleotides in the pRNA E-loop hairpin show protection from chemical modification when pRNA forms dimers and multimers (*88*). Thus, the pRNA E-loop hairpin is a conserved structural piece of the RNA and protein interactions in the phi29 packaging motor.

Questions pertaining to the solution structure of the E-loop hairpin arise upon introducing the fact that it is involved in a RNA-protein interaction. In some cases, an RNA-protein interaction can cause the RNA to assume a different three-dimensional shape upon protein binding (*89-90*). In other cases such as the S2 and L11 ribosomal hairpins, the RNA folds into a hairpin that is the same regardless of whether it is bound to the corresponding protein (*91-92*). In the case of the E-loop hairpin, knowing the tertiary structure will aid in improving structure prediction programs. Whether or not the three-dimensional conformation of the E-loop hairpin changes during the RNA-protein interaction, future studies of the RNA-protein interaction can use the NMR structure determined in this research as a reference point to ask questions about conformational changes in RNA-protein binding. These interactions will undergo vigorous evaluations and comparisons between solution structures and predicted structures in order to help further the ability to predict RNA structure. The tertiary structure has been deposited into the protein data bank (PDB) for further use by other researchers.

Solving the tertiary structure of the Φ29 E-loop hairpin has several implications that can help positively impact current structure prediction programs. First, in general the hairpin loop is one of the most commonly observed RNA structural motifs (*93*). In certain cases, hairpin formation is believed to help initiate the folding of the overall RNA and help keep it from falling into a kinetic trap(*93*). Other RNA hairpins are responsible for creating a platform for electrostatic RNA protein interactions because upon turning around in the hairpin, the backbone is sometimes left exposed, allowing the interactions the opportunity to occur.

The results from this project have already been used in a recent publication by Shu et al (*94*). The NMR tertiary structure of the Φ29 E-loop hairpin was used in a comparison to validate the results of a three-dimensional computer model of the pRNA dimer formation. Though the author believes that the structures agree very well, they state that the RMSD for the backbone atoms was ~3.3 Å. This is a considerably large amount for a RNA backbone comparison. The author attributed the disagreement to the hairpin bases that were not well defined by the NMR data. Conversely, the all-atom RMSD of the hairpin bases was less than 2 Å. The results from their computer modeling looked almost identical to the model that was predicted by MC-SYM before further refinement using experimental data. This is an example of why it's important to study small pieces carefully as benchmarks. The structure of the Φ29 E-loop hairpin will also form the foundation for further research pertaining to the function of the pRNA in the Φ29 packaging motor

**Figure 1.2: Phage Φ29 Cryo-EM Data**

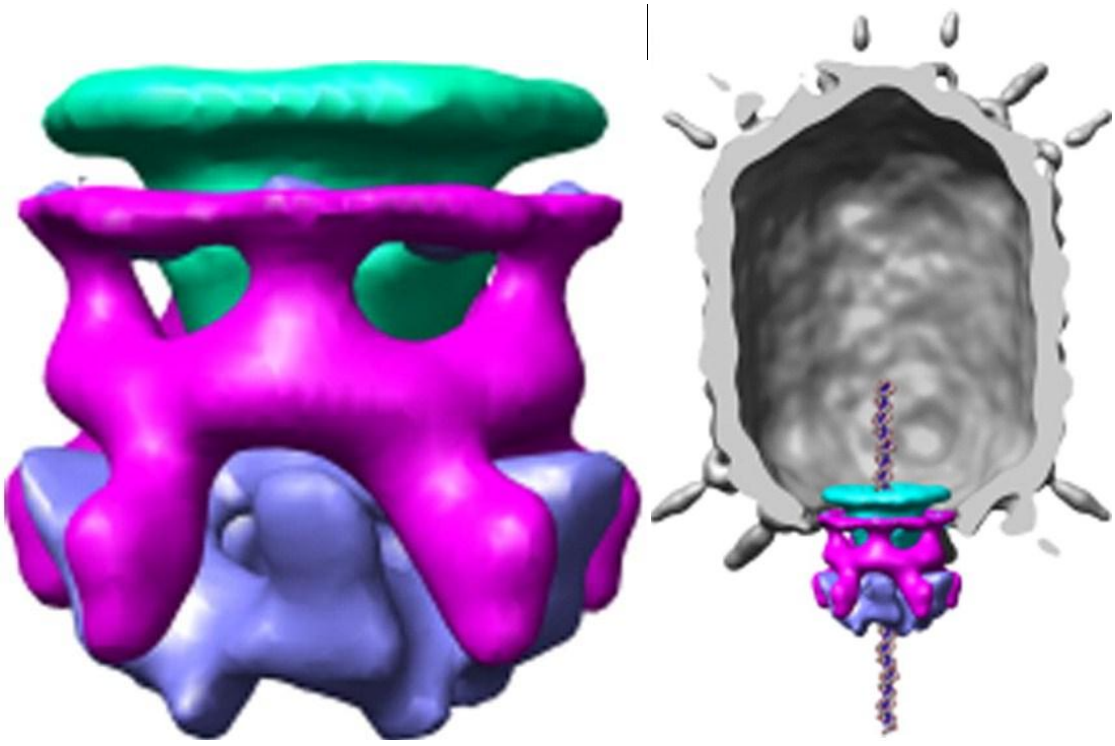


**Figure 1.2:** 18-21 Å cryo-EM data from Figure 6 Morais et al. 2008 Structure v 16 p 1270. The structure on the left shows the pRNA (magenta) in complex with the head-tail connector protein gp10 (green) as well as the protein ATPase gp16 (blue). On the right is the packaging motor in the context of the protein capsid.

**Figure 1.3: Phage Φ29 pRNA Secondary Structures**

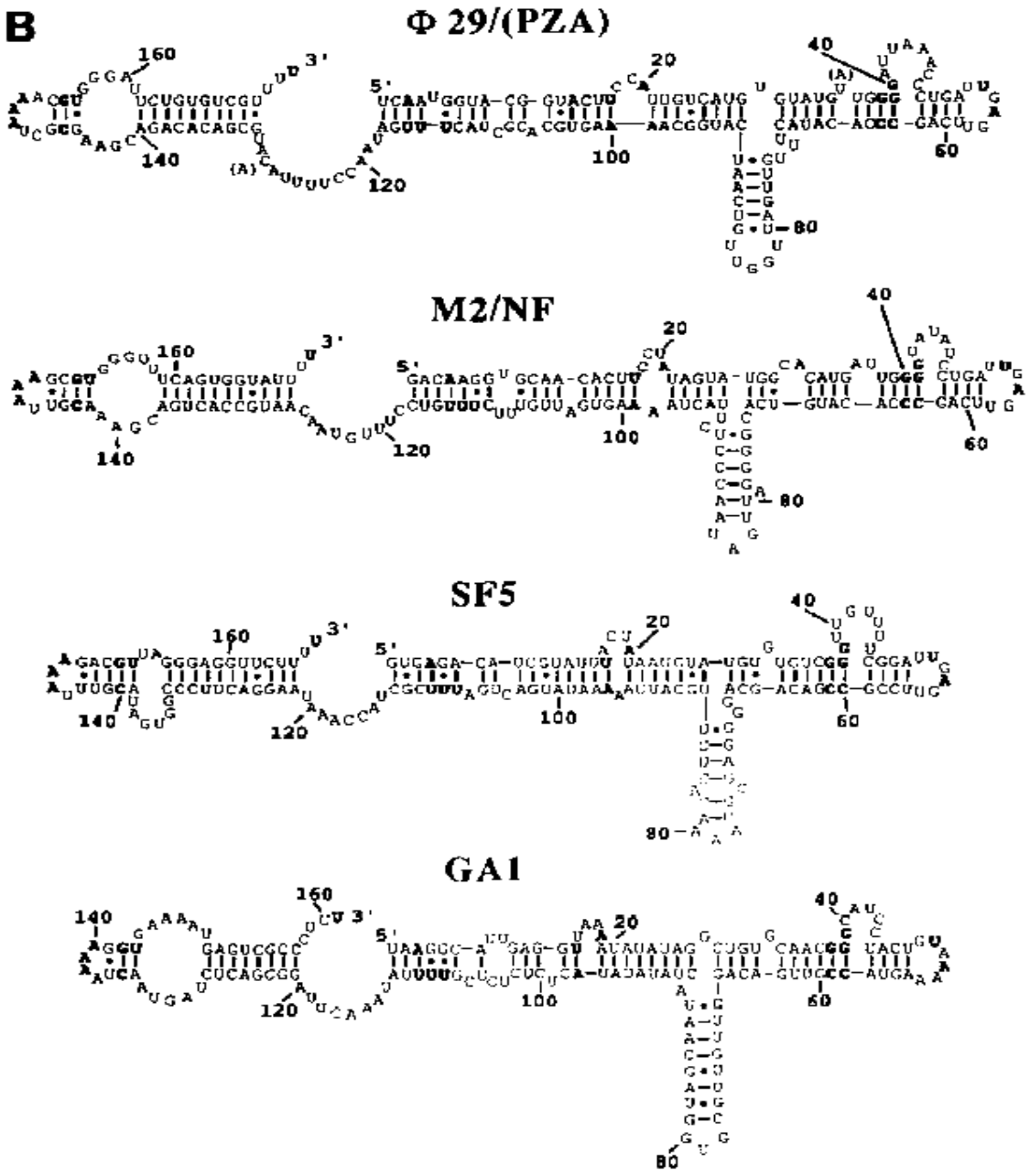



**Figure 1.3:** 1990 Bailey et. Al, J. Biol. Chem. Figure 3. This is the conserved secondary structure of the pRNA used by the different phages. The bold bases are conserved throughout all of the pRNA in the family of phages.

**Figure 1.4: Φ29 E-loop Secondary Structure**




**Figure 1.4:** The primary sequence of the pRNA in the predicted secondary structure of the E-loop hairpin. U3 is U50 in the full length pRNA. The bases in gray were added to increase the thermodynamic stability.

## 1.7 Consecutive GU Base Pairs at the Ends of Helices

There have been numerous cases through which the thermodynamics of Watson-Crick base pair formation have been studied (*44, 95-97*). The non-canonical G-U base pair is also well studied but stabilities have been shown to be idiosyncratic and context dependent. This makes the prediction of G-U pair stabilities and thermodynamics very complex. Optical melting is the most common means of measuring the stabilities of RNA duplexes and the protocols have been unchanged for many years. These experiments produce the raw data that are used to fill the thermodynamics tables that the physics-based algorithms use to make their predictions.

An important assumption in optical melting is that the RNA is in one of two states, duplex or single strand and the RNA must also only fold into one conformation. This is called the two-state assumption. The absorbance of the RNA is observed as it is being heated and the absorbance due to any intermediates or alternate conformations is assumed to be negligible. The results are then evaluated by two different calculations, the van't Hoff plot and the fit of the melting curves. When the calculations are compared, if the two results agree with less than 10-20% then they are in agreement with the two-state assumption. If not, then contributions to the absorbance coming from intermediates or alternate conformations affect the data and it cannot be used for thermodynamic predictions.

Data collected from these experiments include the $T_m$, $\Delta H$, $\Delta S$ and most importantly for thermodynamic predictions, $\Delta G$. The $\Delta G$ values are used

because the amount of error that is associated with the calculations is very low, ~2%. The results of these experiments are well-represented by the current nearest-neighbor model termed the individual nearest-neighbor (INN-HB) model (*63*). This model implies that the thermodynamic stability of a base pair is dependent on the adjacent base pair because of the strong influence of the stacking interactions. It has been shown that G-U base pairs are dependent on factors other than the adjoining base pairs and do not completely follow the INN-HB model(*98*).

G-U base pairs were proposed soon after the structure of DNA was discovered by Watson and Crick. Since, much work has been done to give values to the stabilities of these interactions(*50, 53, 98-100*). G-U base pairs are well studied and they have been found to be favorable on the ends of helices(*101*), but unfavorable in the middle of a helix(*100, 102*). There are still experiments that need to be done to improve the characterization of the G-U pair. The most common G-U base pair is the *syn* G-U wobble using the W-C faces of both bases(*95, 103*). The interaction makes two hydrogen bonds, one from the O-2-carbonyl of the Uracil to the Guanosine H1 imino proton and the other from the O-6-carbonyl of the Guanosine to the H3 imino proton of the Uracil. G-U base pairs are ubiquitous and extremely versatile possibly due to the formation of many different types of base pairs with differing stabilities. The sequence, adjacent base pair, and helical position are other factors that also alter the stability of G-U base pairs. They occur in many different types of RNA including rRNA, tRNA, and even catylitic RNA, most notably the 5' splice site of the group I

introns(*104-105*). Different instances where G-U pairs are used are continually being discovered.

The phenomenon in which small interfering RNA block the translation of certain messenger RNA (mRNA) via cleavage or translation inhibition is called RNA interference (RNAi)(*36-37*). During this process small RNAs called micro RNAs (miRNA) bind to the target mRNA and reduce or prevent its translation. The ability to regulate gene expression by use of miRNAs can possibly lead to new advances in the field of drug therapeutics as well as molecular biology. The thermodynamics of miRNA:mRNA interactions need to be well understood in order for researchers to utilize RNAi to its full potential. The binding and specificity of the miRNA determines the efficacy and mechanism of the RNAi(*106-108*). Generally, highly stable and perfectly bound miRNA follow the cleavage pathway while mismatched and intermediately stable interactions follow the inhibition pathway(*106*). In the case of the microRNA let-7, it is predicted to have two consecutive G-U base pairs that bind to the 3' untranslated region of RAS and HMGA mRNA in cancer regulatory pathways(*109-112*). These motifs include 5'A**GU**/3'U**UG**, 5'G**UU**/C**GG**, 5'G**GU**/3'C**UG**. When the optical melting studies were done on these motifs, the predictions were underestimated by 0.6-1.4 kcal/mol (*101*). These miscalculations directly affect the miRNA-mRNA target prediction programs and some programs even restrict the number of G-U pairs since they are calculated as unfavorable interactions(*113*).

Because the stability of the different GU pairs varies, the thermodynamic stability of interactions involving G-U pairs is poorly predicted. Much emphasis

has been placed on determining meaningful values as well as helical parameters for G-U pairs because they are so prevalent (*114-117*). G-U pairs have been shown to not be isosteric with a regular W-C base pair which causes a change in backbone orientation as well as stacking. Helix propagation by a G-U base pair gives differing amounts of stabilities. Even the location of the G-U pair in the helix cause differences in the stabilities gained. Recent studies have shown that consecutive G-U pairs at the ends of helices are up to 3.8 kcal/mol more stable than previously predicted(*101*). New models for predicting the stabilities of terminal G-U pairs were proposed based on this data, which will help improve the physics-based prediction programs and secondary structure predictions. As for the knowledge-based and tertiary structure programs, knowing the three-dimensional structure of RNAs with these types of interactions are essential because they can help identify the important factors that contribute to the stability and overall shape.

In this study, the tertiary structure of the duplex 5'r(**GUGU**CG**GUGU**)$_2$ will be determined. This sequence was chosen because of the large disagreement between the experimental data measured by optical melting and the stability predicted using the INN-HB model. This sequence was predicted with the INN-HB model to have no favorable thermodynamic stability and not form a stable duplex. When the optical melting experiments were performed, the oligomer formed a stable self-complimentary duplex with the consecutive terminal G-U pairs adding 3.1 kcal/mol worth of stability to the two G-C base pair core (*101*). The preliminary imino proton spectra showed good dispersion with no overlap as

well as very strong, sharp peaks. The tertiary structure of this duplex will be determined by NMR spectroscopy to identify interactions such as base stacking, hydrogen bonding, base orientation and other unknown interactions that may contribute to the stability of the terminal G-U base pairs. This structure will not only help pinpoint these interactions, it will give also give a three-dimensional benchmark for predicting the tertiary structure of RNAs containing consecutive terminal G-U pairs. Another aspect that this study will also aid in is discerning the helical parameters of G-U pairs at the ends of helices since very little is known about their ribose tendencies and preferred backbone orientation. This structure will also be added to the PDB for further use of other researchers.

In the subsequent chapters, the results and conclusions from the Φ29 E-loop hairpin as well as the duplex containing consecutive terminal G-U pairs will be reported. This research will not only aid in improving structure prediction algorithms, but it will lay the groundwork for further RNA structural research. These structures are benchmarks for validation of similar results and can significantly contribute to the database of NMR tertiary structures.

**Figure 1.5: G-U Base Pairs**

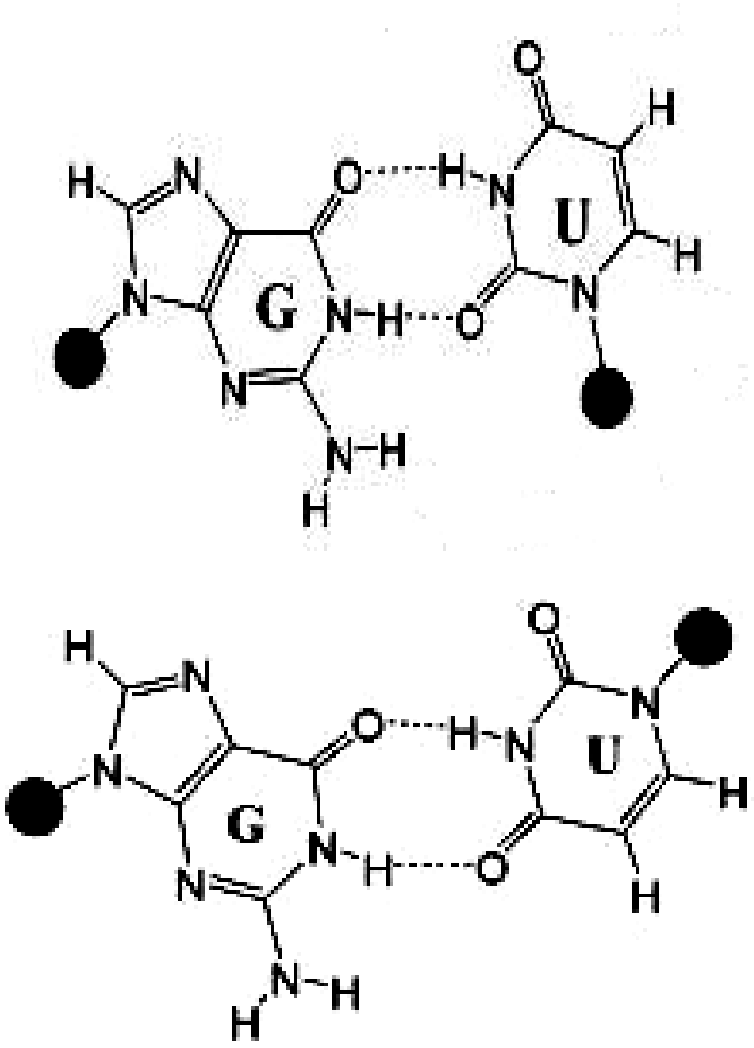



**Figure 1.5:** The two most common G-U base pair orientations. The wobble (Top), and reverse wobble (Bottom), make two hydrogen bonds involving the imino protons on the guanosine and uracil bases. Figures adopted from figure 2 and 3 Nucleic Acids Res. 2002 August 15; 30(16): 3497–3531.

1. Gestland, R., Cech, T., Atkins, J. (2006) *The RNA World*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
2. Sevignani, C., Calin, G., Siracusa, L., Croce, C. (2006) Mamalian Micro-RNA : A Small World for Fine Tuning Gene Expression, *Journal oF Mammalian Genome 17*, 189-202.
3. Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution, *Cell 123*, 1133-1146.
4. Seiwert, S. D., and Stuart, K. (1994) RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro, *Science 266*, 114.
5. Winkler, W. C. (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control, *Current opinion in chemical biology 9*, 594-602.
6. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena, *Cell 31*, 147-157.
7. Altman, S. (1990) Enzymatic cleavage of RNA by RNA, *Bioscience Reports 10*, 317-337.
8. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics, *Nature 407*, 340-348.
9. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation, *Cell 108*, 557-572.
10. Selmer, M., Dunham, C. M., Murphy IV, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA, *Science 313*, 1935.
11. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature 407*, 327-339.
12. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, *Science 289*, 905.
13. Hansen, J. L., Ippolito, J. A., Ban, N., Nissen, P., Moore, P. B., and Steitz, T. A. (2002) The structures of four macrolide antibiotics bound to the large ribosomal subunit, *Molecular cell 10*, 117-128.
14. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science 289*, 920.
15. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif, *Proceedings of the National Academy of Sciences of the United States of America 98*, 4899.
16. Steitz, J. A., and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3'terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli, *Proceedings of the National Academy of Sciences of the United States of America 72*, 4734.
17. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium, *Cell 107*, 679-688.
18. Yonath, A., Leonard, K., and Wittmann, H. (1987) A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction, *Science 236*, 813.

19. Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Glühmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., and Jacobi, C. (2001) Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3, *The EMBO Journal 20*, 1829-1839.
20. Bashan, A., Agmon, I., Zarivach, R., Schluenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., and Hansen, H. A. S. (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression, *Molecular cell 11*, 91-102.
21. Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism, *Science 288*, 640.
22. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution, *Science 292*, 1863.
23. Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H., and Kornberg, R. D. (1991) A mediator required for activation of RNA polymerase II transcription in vitro.
24. Lorch, Y., LaPointe, J. W., and Kornberg, R. D. (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones, *Cell 49*, 203-210.
25. Blackburn, E. H. (1991) Structure and function of telomeres, *Nature 350*, 569-573.
26. Blackburn, E. H. (2001) Switching and signaling at the telomere, *Cell 106*, 661-673.
27. Blackburn, E. H. (2000) Telomere states and cell fates, *Nature 408*, 53-56.
28. Epel, E. S., Blackburn, E. H., Lin, J., Dhabhar, F. S., Adler, N. E., Morrow, J. D., and Cawthon, R. M. (2004) Accelerated telomere shortening in response to life stress, *Proceedings of the National Academy of Sciences of the United States of America 101*, 17312.
29. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase RNA, *Cell 91*, 25-34.
30. Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., and Bacchetti, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity, *The EMBO Journal 11*, 1921.
31. Harley, C. B., Futcher, A. B., and Greider, C. W. (1990) Telomeres shorten during ageing of human fibroblasts.
32. Hemann, M. T., Strong, M. A., Hao, L. Y., and Greider, C. W. (2001) The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability, *Cell 107*, 67-77.
33. Blackburn, E., and Szostak, J. (1984) The molecular structure of centromeres and telomeres, *Annual review of biochemistry 53*, 163-194.
34. Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging, *Nature medicine 12*, 1133-1138.
35. Dunn, B., Szauter, P., Pardue, M. L., and Szostak, J. W. (1984) Transfer of yeast telomeres to linear plasmids by recombination, *Cell 39*, 191-201.
36. Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature 391*, 806-811.

37. Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing, *Cell 106*, 23-34.
38. Montgomery, M. K., Xu, S. Q., and Fire, A. (1998) RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans, *Proceedings of the National Academy of Sciences of the United States of America 95*, 15502.
39. Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C. C. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans, *Cell 99*, 123-132.
40. Timmons, L., and Fire, A. (1994) Specific interference by ingested dsRNA, *Science 263*, 802-805.
41. Mello, C. C., and Conte, D. (2004) Revealing the world of RNA interference, *Nature 431*, 338-342.
42. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure1, *Journal of Molecular Biology 288*, 911-940.
43. Gultyaev, A. P., Van Batenburg, F., and Pleij, C. W. A. (1995) The computer simulation of RNA folding pathways using a genetic algorithm, *Journal of Molecular Biology 250*, 37-51.
44. Borer, P. N., Dengler, B., and Tinoco, I. (1974) Stability of ribonucleic acid double-stranded helices, *Journal of Molecular Biology 86*, 843-853.
45. Ding, Y., Chan, C. Y., and Lawrence, C. E. (2004) Sfold web server for statistical folding and rational design of nucleic acids, *Nucleic Acids Research 32*, W135.
46. Do, C. B., Woods, D. A., and Batzoglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models, *Bioinformatics 22*, e90.
47. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures, *Monatshefte für Chemie/Chemical Monthly 125*, 167-188.
48. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) RNA structure prediction, *Annual Review of Biophysics and Biophysical Chemistry 17*, 167-192.
49. Clanton-Arrowood, K., McGurk, J., and Schroeder, S. J. (2008) 3 Terminal Nucleotides Determine Thermodynamic Stabilities of Mismatches at the Ends of RNA Helices†, *Biochemistry 47*, 13418-13427.
50. Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Free energy contributions of G-U and other terminal mismatches to helix stability, *Biochemistry 25*, 3209-3213.
51. Blanton S. Tolbert, S. D. K., Susan J. Schroeder, Thomas R. Krugh, Douglas H. Turner. (2006) NMR Structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: Probing the Structural Features That Shape the Thermodynamic Stability of GA Pairs, *Biochemistry*.
52. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) Nuclear Magnetic Resonance Spectroscopy and Molecular Modeling Reveal That Different Hydrogen Bonding Patterns Are Possible for GU Pairs: One Hydrogen Bond for Each GU Pair in r (GGCGUGCC) 2 and Two for Each GU Pair in r (GAGUGCUC) 2†,‡, *Biochemistry 39*, 8970-8982.
53. Schroeder, S. J., and Turner, D. H. (2001) Thermodynamic Stabilities of Internal Loops with GU Closing Pairs in RNA, *Biochemistry 40*, 11509-11517.

54. Banerjee, A. R., Jaeger, J. A., and Turner, D. H. (1993) Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure, *Biochemistry 32*, 153-163.
55. Tinoco, I. (1999) How RNA folds, *Journal of Molecular Biology 293*, 271-281.
56. David H. Mathews, M. D. D., Jessica L. Childs, Susan J. Schroeder, Michael Zuker, Douglas H. Turner. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. 101*, 7287–7292.
57. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) Probing the structure of RNAs in solution, *Nucleic Acids Research 15*, 9109.
58. Das, R., and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures, *Proceedings of the National Academy of Sciences 104*, 14664.
59. Marc Parisien, F. M. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature 452*, 51-55.
60. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Research 31*, 3406.
61. Hofacker, I. L. (2003) Vienna RNA secondary structure server, *Nucleic Acids Research 31*, 3429.
62. Leontis, N. B., Stombaugh, Jesse., Westhof, Eric. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Research 30*, 3497-3531.
63. Xia, T., SantaLucia Jr, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson- Crick Base Pairs†, *Biochemistry 37*, 14719-14735.
64. Jacobsen, N. E. (2007) *NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology.*, Wiley-Interscience, Hoboken, N.J.
65. (1993) *NMR of macromolecules: a practical approach*, Oxford University Press, Oxford, NY.
66. Terence N. Mitchell, B. C. (2004) *NMR-From Spectra to Structures*, Springer, Dortmund, Germany.
67. Axel T. Brüngera, P. D. A., G. Marius Clorec, Warren L. DeLanod, Piet Grose, Ralf W. Grosse-Kunstlevea,b, Jian-Sheng Jiangf, John Kuszewskic, Michael Nilgesg, Navraj S. Pannuh, Randy J. Readi, Luke M. Riceb, Thomas Simonsonj, Gregory L. Warrenb. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination, *Acta Cryst. D54*.
68. Wilfried J. J. Meijer, J. A. H., and Margarita Salas. (2001) phi 29 Family of Phages, *American Society for Microbiology 65*, 261-187.
69. Morais, M. C., Koti, J. S., Bowman, V. D., Reyes-Aldrete, E., Anderson, D. L., and Rossmann, M. G. (2008) Defining Molecular and Domain Boundaries in the Bacteriophage 29 DNA Packaging Motor, *Structure 16*, 1267-1274.
70. Simpson, A. A. T., Yizhi; Leiman, Petr G.; Badasso, Mohammed O.; He, Yongning; Jardine, Paul J.; Olson, Norman H.; Morais, Marc C.; Grimes, Shelley; Anderson, Dwight L.; Baker, Timothy S.; Rossmann, Michael G. (2000) Structure of the bacteriophage φ29 DNA packaging motor, *Nature 408*, 745-750.
71. Zhang, F., Lemieux, S., Wu, X., St.-Arnaud, D. McMurray, T., Major, F., Anderson, D. (1998) Function of Hexameric RNA in Packaging of Bacteriophage phi29 DNA in vitro, *Mol. Cell 2*, 141-147.

72. Guo, P. X., Erickson,S. and Anderson,D. (1987) A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA., *Science 236*, 690-694.
73. Atz R, M. S., Gao J, Anderson DL, Grimes S. (2007) Alanine scanning and Fe-BABE probing of the bacteriophage ø29 prohead RNA-connector interaction., *Journal of Molecular Biology 369*, 239-248.
74. PEIXUAN Guo, S. G., AND DWIGHT ANDERSON. (1986) A defined system for in vitro packaging of DNA-gp3 of the Bacillus subtilis bacteriophage Phi29, *Proc. Natl. Acad. Sci. 83*, 3505-3509.
75. Shu, D., Huang, LP., Hoeprich, S., Guo, P. (2005a) Construction of DNA-packaging phi29 RNA Monomers, Dimers, and Trimers with Variable Sizes and Shapes as Potential Parts for Nanodevices., *Journal of Nanoscience and Nanotechnology 3*, 295-302.
76. Grimes S, J. P., Anderson D. (2002) Bacteriophage phi 29 DNA packaging., *Advances in Virus Research 58*, 255-294.
77. Guo, P., Erickson, S., and Anderson, D. (1987) A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA, *Science 236*, 690.
78. M Trottier, C. Z., and P Guo. (1996) Complete inhibition of virion assembly in vivo with mutant procapsid RNA essential for phage phi 29 DNA packaging, *Journal of Virology 70*, 55-61.
79. Wichitwechkarn, J., Bailey,S., Bodley,J.W. and Anderson,D. (1989) Prohead RNA of bacteriophage phi 29: size, stoichiometry and biological activity., *Nucleic Acids Research 17*, 3459-3468.
80. Bailey, S., Wichitwechkarn, J., Johnson, D., Reilly, B. E., Anderson, D. L., Bodley, J. W. (1990) Phylogenetic Analysis and Secondary Structure of Bacillus subtilis Bacteriophage RNA Required for DNA Packaging., *Journal of Biological Chemistry 265*, 22365-22370.
81. Morais, M., Koti, J., Bowman, V., Reyes-Aldrete, E., Anderson, D.L., Rossman, M.G. (2008) Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor, *Structure 16*, 1267-1274.
82. Reid, R. J. D., Bodley, J.W., Anderson, D. (1994) Characterization of the Prohead -pRNA Interaction of Bacteriophage phi29, *Journal of Biological Chemistry 269*, 5157-5162.
83. Xiao, F., Moll, W.-D., Guo, S., Guo, P. (2005) Binding of pRNA to the N-terminal 14 amino acids of connector protein to bacteriophage phi29, *Nucl. Acids Res. 33*, 2640-2649.
84. Garver, K., Guo, P. (1997) Boundary of pRNA functional domains and minimum pRNA sequence requirement for specific connector binding and DNA packaging of phage phi29, *RNA 3*, 1068-1079.
85. Guasch, A., Pous, J., Ibarra, B., Gomis-Ruth, F.X., Valpuesta, J.M., Sousa, N., Carrascosa, J.L., Coll, M. (2002) Detailed Architecture of a DNA Translocating Machine: The High-resolution Structure of the Bacteriophage phi29 Connector Particle, *J. Mol. Biol. 315*, 663-676.
86. Reid, R., J.D., Bodley, J., W., Anderson, D. (1994b) Identification of Bacteriophage phi29 Prohead RNA Domains Necessary for in vitro DNA-gp3 Packaging, *J. Biol. Chem. 269*, 9084-9089.
87. Kitamura, A., Jardine, P.J., Anderson, D.L., Grimes, S., Matsuo, H. (2008) Analysis of intermolecular base pair formation of prohead RNA of the phage pi29 DNA packaging motor using NMR spectroscopy, *Nucl. Acids Res. 36*, 839-848.
88. Trottier, M., Mat-Arip, Y., Zhang, C., Chen, C., Sheng, S., Shao, Z., Guo, P. (2000) Probing the structure of monomers and dimers of the bacterial virus phi29 hexamer RNA complex by chemical modification, *RNA 6*, 1257-1266.

89. Williamson, J. R. (2000) Induced fit in RNA–protein recognition, *Nature Structural & Molecular Biology 7*, 834-837.
90. Draper, D. E. (1995) Protein-RNA recognition, *Annual review of biochemistry 64*, 593-620.
91. M. A. Fountain, M. J. S., T. R. Krugh, and D. H. Turner. (1996) Structural Features of a Six-Nucleotide RNA Hairpin Loop Found in Ribosomal RNA, *Biochemistry 35*, 6539-6548.
92. Zhang, H., Fountain, M., Krugh, T. (2001) Structural Characterization of a Six-Nucleotide RNA Hairpin Loop Found in *Escherichia coli*, r(UUAAGU), *Biochemistry 40*, 9879-9886.
93. Uhlenbeck, O. C. (1990) Tetraloops and RNA folding, *Nature 346*, 613.
94. Shu, D., Zhang, H., Petrenko, R., Meller, J., and Guo, P. (2010) Dual-Channel Single-Molecule Fluorescence Resonance Energy Transfer to Establish Distance Parameters for RNA Nanoparticles, *ACS nano*, 1676-1683.
95. Victor A. Bloomfield., D. M. C., Ignacio Tinoco (2000) *Nucleic Acids Structures, Properties and Functions*.
96. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability, *Proceedings of the National Academy of Sciences of the United States of America 83*, 9373.
97. Petersheim, M., and Turner, D. H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp, *Biochemistry 22*, 256-263.
98. He, L., Walter, E., and Turner, H. (1991) Nearest-Neighbor Parameters for GU Mismatches: 3tUG5'Is Destabilizing in the Contexts% if, YE: $, and $ Egy but Stabilizing in, *Biochemistry 30*, 46.
99. Turner, J. A. M. a. D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC)2 by Two-Dimensional NMR and Simulated Annealing, *Biochemistry 35*, 14077-14089.
100. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helixes, *Biochemistry 25*, 5755-5759.
101. Nguyen, M. T., and Schroeder, S. J. (2010) Consecutive Terminal GU Pairs Stabilize RNA Helices, *Biochemistry*.
102. McDowell, J. A., and Turner, D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC) 2 by Two-Dimensional NMR and Simulated Annealing†,‡, *Biochemistry 35*, 14077-14089.
103. Nagaswamy, U., Voss, N., Zhang, Z., and Fox, G. E. (2000) Database of non-canonical base pairs found in known RNA structures, *Nucleic Acids Research 28*, 375.
104. Cech, T. R. (1990) Self-splicing of group I introns, *Annual review of biochemistry 59*, 543-568.
105. Garriga, G., Lambowitz, A. M., Inoue, T., and Cech, T. R. (1986) Mechanism of recognition of the 5 splice site in self-splicing group I introns.
106. Haley, B., and Zamore, P. D. (2004) Kinetic analysis of the RNAi enzyme complex, *Nature Structural & Molecular Biology 11*, 599-606.
107. Doench, J. G., and Sharp, P. A. (2004) Specificity of microRNA target selection in translational repression, *Genes & development 18*, 504.

108. Ameres, S. L., Martinez, J., and Schroeder, R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC, *Cell 130*, 101-112.
109. Mayr, C., Hemann, M. T., and Bartel, D. P. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation, *Science 315*, 1576.
110. Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. J. (2005) RAS is regulated by the let-7 microRNA family, *Cell 120*, 635-647.
111. Lee, Y. S., and Dutta, A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene, *Genes & development 21*, 1025.
112. Shell, S., Park, S. M., Radjabi, A. R., Schickel, R., Kistner, E. O., Jewell, D. A., Feig, C., Lengyel, E., and Peter, M. E. (2007) Let-7 expression defines two differentiation stages of cancer, *Proceedings of the National Academy of Sciences 104*, 11400.
113. Ragan, C., Cloonan, N., Grimmond, S. M., Zuker, M., and Ragan, M. A. (2009) Transcriptome-wide prediction of miRNA targets in human and mouse using FASTH, *PloS one 4*, e5745.
114. Chen, X., McDowell, J., Kierzek, R., Krugh, T., and Turner, D. (2000) Nuclear Magnetic Resonance Spectroscopy and Molecular Modeling Reveal That Different Hydrogen Bonding Patterns Are Possible for GU Pairs: One Hydrogen Bond for Each GU Pair in r (GGCGUGCC) 2 and Two for Each GU Pair in r (GAGUGCUC) 2†,‡, *Biochemistry 39*, 8970-8982.
115. McDowell, J., and Turner, D. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC) 2 by Two-Dimensional NMR and Simulated Annealing†,‡, *Biochemistry 35*, 14077-14089.
116. Strobel, S. A., and Cech, T. R. (1995) Minor groove recognition of the conserved GU pair at the Tetrahymena ribozyme reaction site, *Science 267*, 675.
117. Limmera, S., and Reifa, B. (1996) NMR evidence for helix geometry modifications by a GU wobble base pair in the acceptor arm of E. coli tRNAAla, *FEBS Letters 385*, 15-20.

# CHAPTER 2

## TERTIARY STRUCTURE DETERMINATION OF THE Φ29 E-LOOP HAIRPIN

The following text is adapted from a publication in the American Chemical Society journal Biochemistry by Harris and Schroeder 2010 volume 49 pages 5989-5997.

### 2.1 Introduction

*Bacillus subtilis* phage Φ29 utilizes a nanomolecular motor to encapsidate the DNA genome into an empty viral capsid (*118-119*). The Φ29 motor contains an essential prohead RNA (pRNA) that works in concert with several proteins to complete the packaging process (*72, 74, 120*). Previous chemical modification and phylogenetic studies on pRNA show very little primary sequence conservation but a well conserved secondary structure (*68, 80, 121-122*). The pRNA E-loop hairpin sequence is conserved in Φ29, M2/NF, and SF5 pRNA sequences but varies in GA1 pRNA (Figure 2.1) (*80*). The pRNA E-loop hairpin interacts with gene product 10, the viral head-tail connector protein, a 290 amino acid monomer which forms a 422 kDa dodecameric complex (*81-85*). Mutations in the pRNA E-loop hairpin reduce protein binding to less than 30%, and mutation of the A nucleotide in the hairpin loop to a C reduces packaging activity 95% (*86*), although substitution of this hairpin with a UUCG tetraloop retains some packaging activity *in vitro* (*87*). Nucleotides in the pRNA E-loop hairpin show protection from chemical modification when pRNA forms dimers and

multimers (*88*). Thus, the pRNA E-loop hairpin is a conserved structural piece of the RNA and protein interactions in the phi29 packaging motor.

**Figure 2.1**: **Hairpin Primary Sequences**

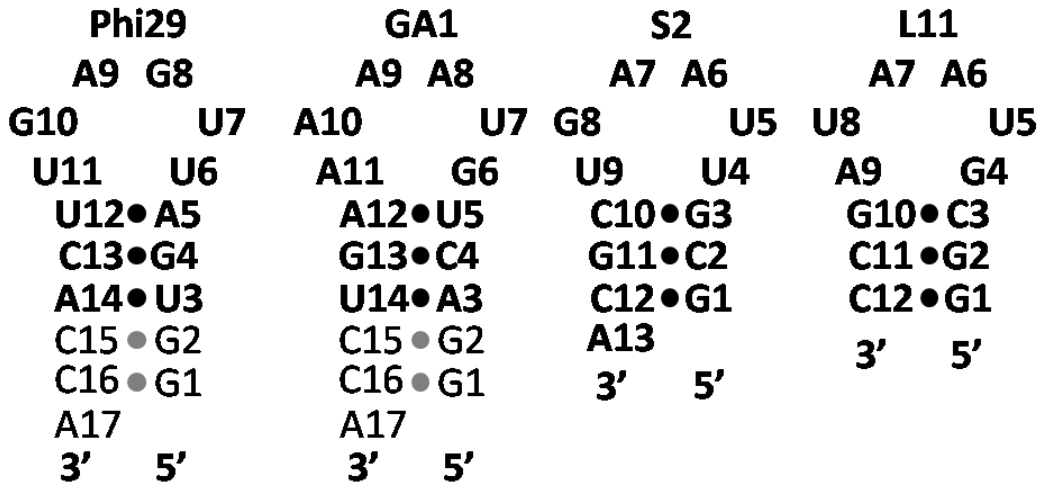



**Figure 2.1**: Primary sequence of the phi29 E-loop RNA hairpin. U3 is U50 in the phi29 RNA full-length sequence. Also shown are the GA-1 E-loop hairpin as well as the S2 and L11 ribosomal hairpins. Nucleotides in gray were added for stability of the NMR construct.

NMR studies of RNA hairpins provide benchmarks for methods to predict three-dimensional RNA structure from sequence. MC-SYM, a computational program that predicts RNA three-dimensional structure from sequence using symbolic programming and cyclic motifs, is trained on high resolution ribosome crystal structures (*59*). Isostericity matrices, which classify and predict base pairing interactions, are based on phylogenetic and structural studies of rRNA

and structures in the rfam database (*62, 123*). pRNA presents a good test of these prediction methods because the sequence conservation is unusually low but the secondary structure is well conserved.

The ribosomal hairpins S2 in the small subunit and L11 in the large subunit structures determined by both NMR and crystallography contain similar sequences to pRNA hairpins. The ribosomal hairpins have the same structure in both the small oligonucleotide model of the hairpin determined by NMR and in the context of the RNA-protein interactions determined by crystallography (*92, 124-129*). Thus, in this case, the three-dimensional structure of the ribosomal hairpins is determined by the RNA sequence and is not altered by protein and RNA tertiary interactions. Comparisons of the experimental and predicted structures for the pRNA E-loop and ribosomal hairpins S2 and L11 test the applicability of MC-SYM and isostericity matrices to predict RNA structure.

Both the ribosome and the phi29 packaging motor are large RNA-protein assemblies with complex architectures that support and direct biological catalysis. The S2, L11, and pRNA E-loop hairpins share a common structural function at junctions where RNA helices and proteins pack closely together through non-specific electrostatic interactions between residues such as arginine, lysine, glutamine, asparagine, and the RNA phosphate backbone. The high resolution ribosome crystal structures (*127-129*) provide an enormous database of RNA-protein structural interactions upon which structure prediction methods can be developed and tested. The small size of these six-nucleotide hairpins is advantageous for testing computational predictions and assessing the accuracy

of local three-dimensional structure prediction in detail. Accurate structure prediction from sequence can provide models to stimulate hypotheses of function in RNA-protein complexes. In the absence of high-resolution structural information for the complete phi29 packaging motor, atomic-resolution structure determination of parts of the motor can be modeled into the low-resolution cryo-electron structures in order to design experiments to determine the mechanism of the packing motor and the role of the essential pRNA. Atomic resolution crystal structures of proteins gp7, gp12, gp10, and gp13 have been determined and modeled into cryoEM maps of the prohead motor at 7.9 Å (*81, 130-136*). Models of pRNA hexamers and dimers have been computationally predicted (*137-138*), and experimental validation of prediction methods can increase confidence in the ability of these types of models to inform structure-function hypotheses. This research presents the first experimental atomic-resolution structure of a piece of the pRNA component of the phi29 packaging motor. Thus, the NMR structure of the pRNA E-loop hairpin contributes to structure prediction and modeling efforts that build from small pieces following an *aufbau* approach to RNA structure prediction and structure-function relationships.

## 2.2 Materials and Methods

### 2.2.1 RNA Preparation and Purification.

The RNA was synthesized and gel purified by Dharmacon, Inc. Protecting groups were removed using the supplied deprotecting buffer, a 1% acetic acid solution at pH 3.8. The RNA oligonucleotide was lyophilized, resuspended in 10

mM $NaPO_4$, 10 mM NaCl, and 0.5 mM $Na_2EDTA$ pH 5.5 NMR buffer, dialyzed in 5 mM NaCl, 5 mM $Na_3PO_4$, 0.25 mM $Na_2EDTA$ pH 6 buffer with a 1000 molecular weight cutoff membrane, and further purified using a Sephadex G-25 gel exclusion column. Oligomer purity was greater than 95%, proven by $^{32}P$ labeling and gel electrophoresis. The sample concentration was 1 mM on the basis of high temperature UV absorbance measurement. For NMR experiments in $D_2O$, the oligomer was exchanged with 99.9% $D_2O$ twice and once with 99.996% $D_2O$.

### 2.2.2 UV Experiments.

A Beckman DU 800 UV/Vis spectrometer was used to measure absorbance versus temperature melting curves at 260 nm from 20-90 $^oC$ with a heating rate of 1 $^oC$/min. The sample was in NMR buffer and the $T_m$ was calculated with Meltwin 3.5 (*99*) and the two-state model (*95*). Hairpin formation, rather than duplex formation, can be confirmed by a concentration-independent melting temperature.

### 2.2.3 NMR Spectroscopy.

Exchangeable proton spectra were taken at 1 $^oC$ using Watergate or SSnoesy pulse sequences. NOESY spectra were recorded at 50, 100, and 200 ms at 1 $^oC$ for exchangeable protons with a delay time of 1 s, 256 increments in the T1 dimension with 1024 data points and 128 scans per FID. Zero filling was used to finish the T1 matrix and had a 20 ppm spectral sweep width. Apodization was

done using sine bell shift weighting. The non-exchangeable proton spectra were recorded at 20 and 25 $^{o}$C using RNA tnnoesy pulse sequence with presaturation water suppression. NOESY spectra were collected at 50, 100, 150 and 400 ms at 25 $^{o}$C with a delay time of 1 s, 300 increments in the T1 dimension with 2048 data points and 30 scans per FID. Zero filling was used to finish the T1 matrix and the spectral sweep width was 8.5 ppm. Apodization was done using sine bell shift weighting. DQF-COSY experiments were performed at 25 $^{o}$C with 300 increments in the T1 dimension with 2048 data points and 48 scans per FID. The spectral sweep width was set to 8.5 ppm and the delay time was 1 s. Apodization was done using gaussian functions.

The $^{1}$H-$^{13}$C HSQC data were recorded using a spectral sweep width of 12 ppm for the $^{1}$H and 45 ppm for the $^{13}$C dimension with a 1 s delay time. There were 32 T1 increments and 500 scans with 2048 points. The data were apodized with a sine bell shifted function in both directions.

The $^{1}$H-$^{31}$P HETCOR spectra were collected using a Varian 600 MHz spectrometer with a 5 ppm spectral width in both the $^{1}$H and the $^{31}$P dimension Data was collected in 512 T1 increments, 128 scans, and 2048 data points. Apodization was done using sine bell shift weighting.

### 2.2.4 NMR Restraints and Peak Assignments.

Distance restraints from the non-exchangeable NOESY spectra were based on NOE crosspeaks in the 50, 100, and 150 ms experiments. Restraints were classified as strong, medium, or weak if the NOE occurred in the 50, 100, or

150 ms spectra, respectively. The 400 ms spectra were used to confirm NOE assignments that were weak in the 150 ms spectra but were not used directly as distance restraints. Backbone, chi and sugar pucker restraints for the stem nucleotides from the RNA backbone consortium (*139*) were applied during standard simulated annealing protocols (*51*). The backbone angles of the hairpin nucleotides were allowed to sample any conformation in order to fully explore conformational space with the exception of the syn G10 nucleotide. The backbone angles in the final structure ensemble are consistent with the observed crosspeaks in the COSY and $^{31}$P HETCOR experiments. The SPARKY assignment program was used for peak assignment, labeling, compilation of distance restraints and NOE lists (*140*). Table 2.4 contains the summary of NOE and dihedral restraints. Table 2.2 contains a list of all the NOE restraints used in the structure calculations.

### 2.2.5 Structure Calculations.

Structure calculations were performed with the Crystallography and NMR Systems (CNS) software version 1.2 (*141*) on a Dell Optiplex GX620. An unrefined, initial MCSYM structure was heated for 10 ps at 500 K without any restraints to generate 30 randomized starting structures (rmsd 6.1 Å). The RNA-DNA-all-atom parameter and topology files from CNS contain the force charges, atomic charges, atomic masses, and atom connectivity.

In the high temperature annealing stage, the temperature was raised to 20000 K with bond, angle, and improper restraints on and NOE and dihedral

angle restraints set to 150 kcal/mol Å$^2$ and 5 kcal/mol rad $^2$, respectively. The first slow-cool annealing stage was done using torsion dynamics with the electrostatics off except between hydrogens and other atoms, which were set to 10%. The temperature was cooled in 125 K temperature steps with the NOE restraints held at 150 kcal/mol Å$^2$ and the dihedral angle restraints set to 200 kcal/mol rad$^2$. The system was allowed to develop for 25 ps using 6000 MD steps at each temperature. The second slow cooled annealing stage was done in Cartesian space and allowed the system to continue from 2000 K to 0 K. The temperature steps were set to 15 K and allowed to develop for 15 ps while the van der Waals scale factor increased linearly from 1 to 4. All other parameters were unchanged during the second annealing stage. The final energy minimization was done by applying Powell energy minimization with the van der Waals scale factor at 1 and the electrostatic terms at 100% using 30 cycles of minimization each with 2000 MD steps. The scale factor for the Powell NOE and dihedral angles were 75 kcal/mol Å$^2$ and 400 kcal/mol rad$^2$, respectively. Planarity (150 kcal/mol Å$^2$) as well as hydrogen bond distance restraints were applied to the stem region of the hairpin. Pymol was used to analyze and visualize the structures (*142*).

### 2.2.6 Structure Predictions with MC-SYM.

MC-SYM structure predictions were done using the web-hosted MC-Fold | MC-Sym pipeline service for RNA secondary and tertiary structure prediction at: http://www.major.iric.ca/MC-Fold/. The MC-Fold program predicts secondary

structures first with the Zuker algorithm (*143*) and Turner thermodynamics (*144-145*), and then selects structures to be submitted to MC-Sym, which generates predicted tertiary structures. These structures were then energy minimized and scored with the AMBER force field (*146-148*).

## 2.3 Results

### 2.3.1 UV Absorbance.

The plot of absorbance versus temperature of the hairpin gives a normal distinctive two-state melting profile. The predicted and measured melting temperatures and free energies at 37 $^{o}$C are 70 $^{o}$C and 7.5 kcal/mol in 1 M NaCl and 67 $^{o}$C and 7.4 kcal/mol in 10 mM NaCl, respectively. These values are consistent within experimental error to estimates for the salt dependence of hairpin stabilities (*149-150*). The melting temperature was unperturbed by varying oligomer concentration, which is consistent with the RNA sequence forming a hairpin and not a duplex (*95*).

### 2.3.2 $H_2O$ NOESY.

In the imino proton spectra, only eight of the ten possible imino resonances were observed between 9-15 ppm (Figures 2.2 and 2.3). The G8 and G10 imino resonances were not seen. In the final model, G8 and G10 are solvent exposed and do not form hydrogen bonds, thus allowing imino protons to exchange with water. The two most downfield imino protons at 14.43 and 13.9 ppm were assigned to U12 and U3, respectively; the imino resonances at 12.32

and 11.71 ppm were assigned to G2 and G1, respectively. These assignments were based on chemical shifts and the pattern of NOEs expected for the stem duplex (*65, 151*). The remaining three imino protons at 11.63, 11.21, and 10.71 ppm were assigned to U11, U6, and U7, respectively. The strong NOE between U6 and U11 is consistent with the formation of a U-U mismatch base pair (*152-153*). These bases are also stacked on the closing A-U base pair and this is confirmed by the observation of both bases having NOEs to U12. The remaining imino proton was assigned to U7. U7H3 gives an NOE to U11 which is consistent with U7 stacking on the U6-U11 pair. Tables 2.3 and 2.4 consist of complete proton, carbon and phosphorus resonance assignments.

**Figure 2.2: E-loop Imino Proton Spectrum**

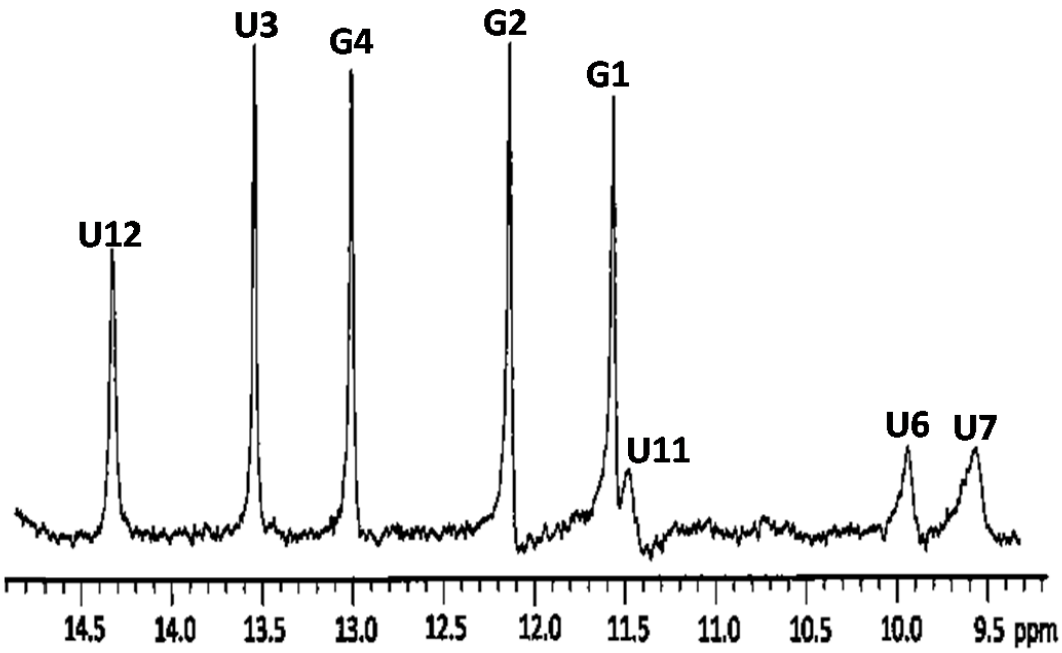



**Figure 2.2:** One dimensional imino proton spectrum of phi29 E-loop hairpin r(5'GGUGAUUGAGUUCACCA) collected at 1 $^{o}$C and 500 MHz.

**Figure 2.3: 2-Dimensional Imino Proton Spectrum**

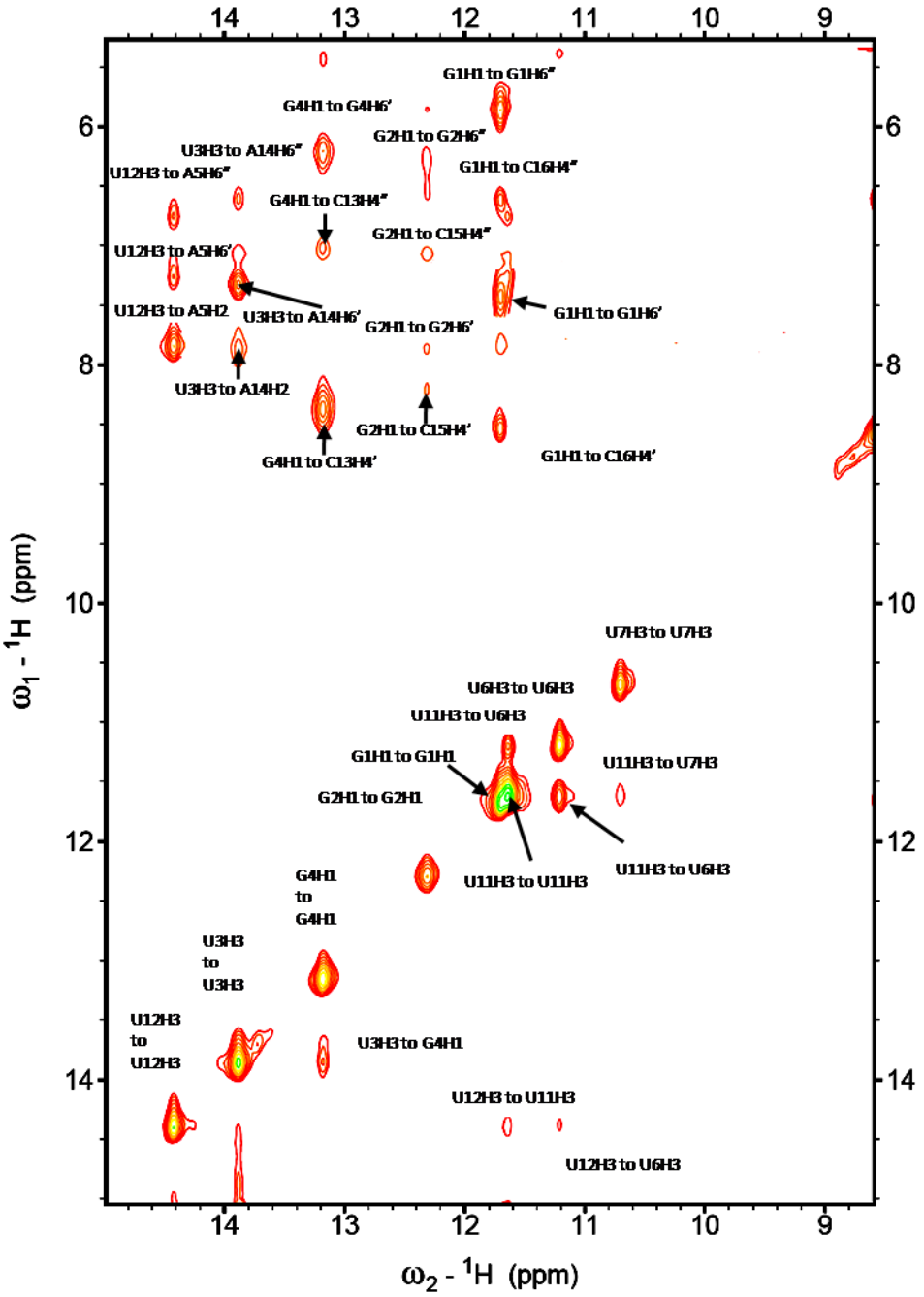



**Figure 2.3:** $H_2O$ NOESY showing exchangeable proton NOEs. Spectrum was taken at $1^{o}$ C in 10 mM NaCl, 10 mM $NaPO_4$, and 0.5mM $Na_2EDTA$ in $H_2O$ with 10% $D_2O$ and a 150 ms mixing time.

### 2.3.3 $D_2O$ NOESY.

Sequential assignment of ribose and aromatic protons was done using the $D_2O$ NOESY spectra and standard procedures (*65, 154*) (Figure 2.4, Table 2.1). A normal NOESY sequential walk is observed for bases G1-U7 and U12-A17. In the hairpin loop U7 shows no NOEs to G8, but U7 does show N + 2 NOEs to A9. Interestingly, NOEs between G8 and A9 occur. A9 shows N+2 NOEs to U11. These NOEs define the U-turn in the hairpin. The 50 ms NOESY shows a strong H1' to H8 NOE of similar intensity to that of a pyrimidine H5-H6 crosspeak (Figure 2.5). This NOE was assigned to G10. The possible assignment of this peak as an A9H1'-G10H8 NOE was considered much less likely because no NOE was observed between A9H2'-G10H8 even at 400 ms NOESY mixing times, which is long enough for either a direct NOE or secondary NOE to occur through a strong H1'-H2' NOE and a strong H1'-H8 NOE.

**Figure 2.4: Aromatic Proton to H1' Proton NOESY Sequential Walk**

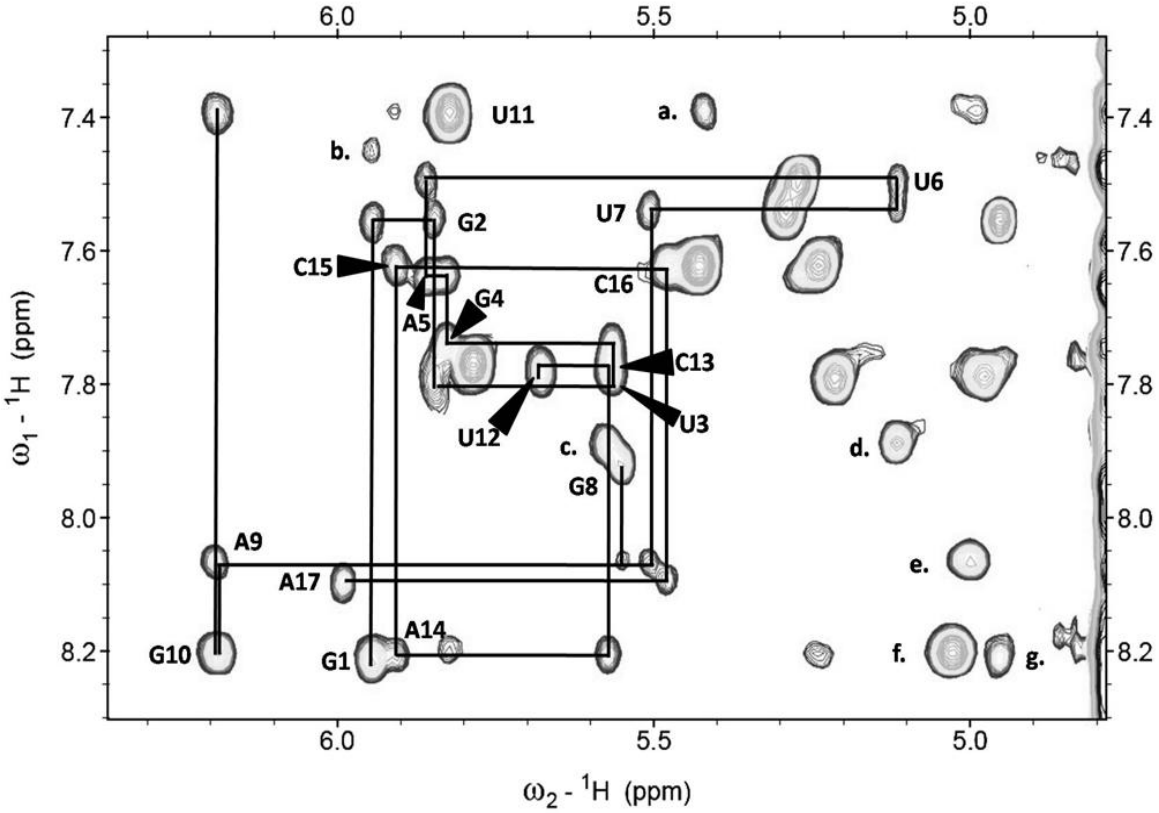



**Figure 2.4:** Aromatic proton to H1', and H5 region of the 400 ms $D_2O$ NOESY spectrum recorded at 20 °C for r(5'GGUGAUUGAGUUCACCA). The base to H1' NOESY walk is indicated by solid lines. Labels represent intranucleotide aromatic to H1' NOEs. The other labels are as follows: a. A14 H2 to C15 H1' b. A17 H2 to G1 H1' c. A5 H2 to C13 H1' d. A5 H2 to U6 H1' e. A9 H8 to A9 H3' f. G10 H8 to G10 H2' g. G1 H8 to G1 H2' h.G10 to H1'-H8

**Figure 2.5: 50 ms $D_2O$ NOESY Spectrum**

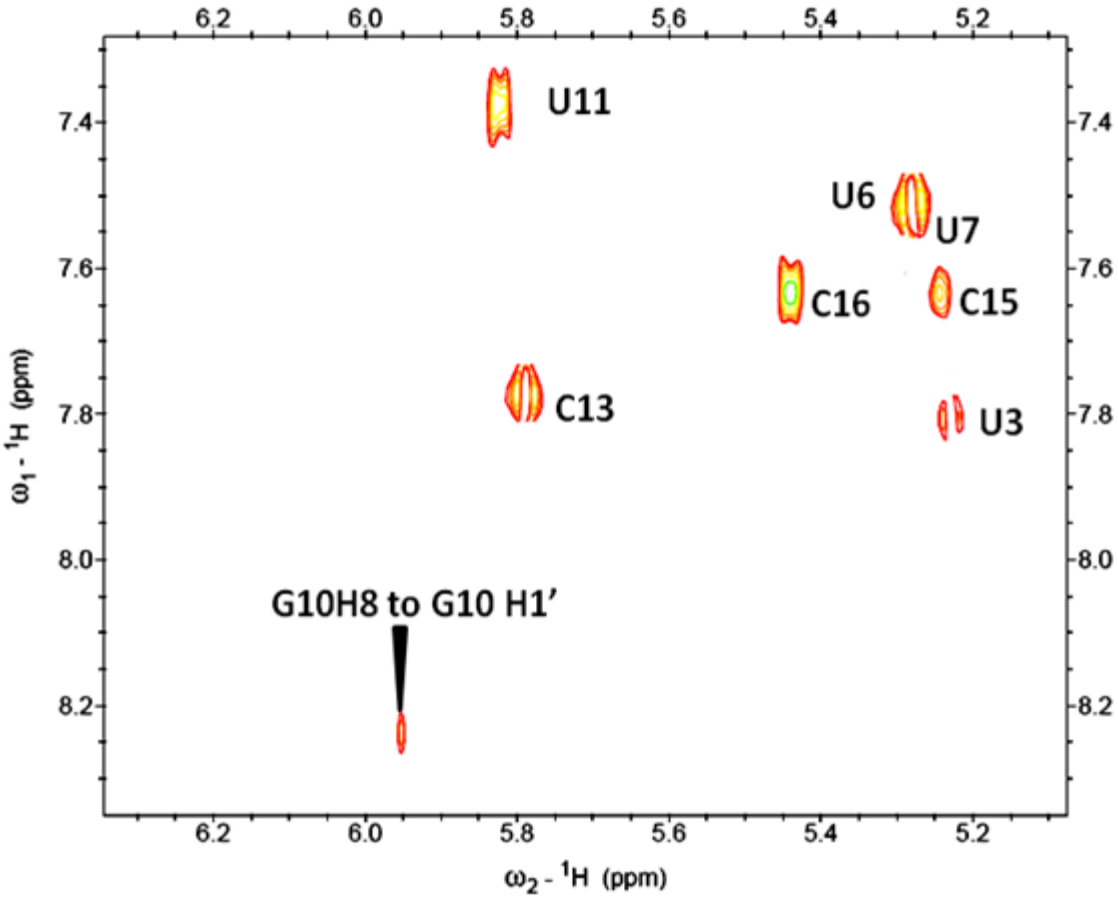



**Figure 2.5:** Aromatic proton to H1', and H5 region of the 50 ms $D_2O$ NOESY spectrum recorded at 20° C showing a strong H1' to H8 NOE of similar intensity to that of a pyrimidine H5-H6 crosspeak. NMR buffer contains 10mM NaCl, 10 mM $NaPO_4$, and 0.5mM $Na_2EDTA$ in $H_2O$ with 10% $D_2O$.

**Figure 2.6: NOESY Showing N+2 NOE**

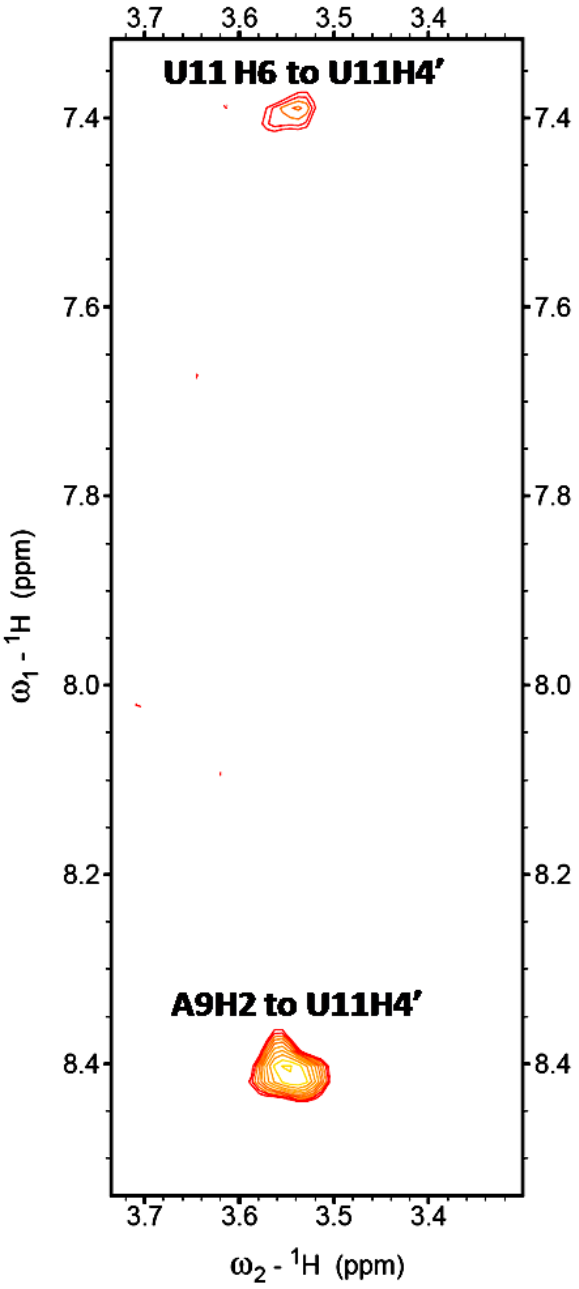



**Figure 2.6:** Section of the 400 ms NOESY at 20 $^{o}$C showing the NOE to A9 H2-U11 H4'

**Table 2.1: Proton Resonance Assignments**

| | H6/H8 | H2/H5 | H1' | H2' | H3' | H4' | H1/H3 imino | H2/H4/H6 amino |
|---|---|---|---|---|---|---|---|---|
| G1 | 8.215 | - | 5.945 | 4.953 | 4.612 | | 11.74 | 7.473/5.883 |
| G2 | 7.554 | - | 5.847 | 4.501 | 4.185 | 4.185 | 12.36 | 8.034/6.254 |
| U3 | 7.791 | 5.214 | 5.568 | 4.527 | 4.659 | | 13.86 | - |
| G4 | 7.739 | - | 5.831 | 4.591 | 4.687 | 4.166 | 13.16 | 8.419 |
| A5 | 7.639 | 7.888 | 5.859 | 4.566 | 4.286 | | - | 7.294/6.855 |
| U6 | 7.495 | 5.271 | 5.115 | 4.007 | 4.413 | 4.116 | 11.23 | - |
| U7 | 7.54 | 5.291 | 5.506 | 4.271 | 4.501 | | 10.72 | - |
| G8 | 7.919 | - | 5.55 | 4.603 | 4.283 | 4.030 | Not Assigned | 7.884/7.124 |
| A9 | 8.067 | 8.409 | 6.188 | 4.603 | 5.000 | | - | Not Assigned |
| G10 | 8.203 | - | 6.194 | 5.028 | 3.91 | 3.548 | Not Assigned | Not Assigned |
| U11 | 7.39 | 5.822 | 4.8 | 5.009 | 4.349 | 3.978 | 11.64 | - |
| U12 | 7.788 | 4.978 | 5.677 | 4.502 | 4.363 | 4.133 | 14.42 | - |
| C13 | 7.77 | 5.785 | 5.569 | 4.482 | 4.230 | | - | 8.419/7.027 |
| A14 | 8.203 | 7.389 | 5.907 | 4.468 | 4.137 | 3.992 | - | 7.953/6.596 |
| C15 | 7.623 | 5.242 | 5.427 | 4.22 | 4.38 | 4.068 | - | 8.272/7.002 |
| C16 | 7.623 | 5.428 | 5.478 | 4.396 | 4.509 | 4.075 | - | 8.616/6.665 |
| A17 | 8.098 | 7.450 | 5.990 | 4.066 | 4.308 | 4.260 | - | Not Assigned |

**Table 2.1:** Proton chemical shift assignments. The protons were referenced to water, which was referenced to 3-trimethylsilyl propionic acid in 10 mM phosphate buffer pH 6. The non exchangeable protons were assigned in experiments at 20 $^{o}$C and the exchangeable protons were assigned in experiments conducted at 1 $^{o}$C. G10 H5’ and G10 H5’’ resonate at 4.285 and 4.358 ppm, respectively, and no other H5’ protons were specifically assigned.

**Table 2.2: NOE assignments**

| | | |
|---|---|---|
| (G1 H22) | (G1 H1) | 3.0, 1.0, 1.5 |
| (G1 H22) | (G1 H21) | 2.1, 0.6, 0.9 |
| (G1 H21) | (G1 H1) | 3.0, 1.0, 1.5 |
| (G1 H1') | (G1 H2') | 2.1, 0.6, 0.9 |
| (G1 H1') | (G1 H3') | 4.0, 1.5, 2.0 |
| (G1 H2') | (G1 H3') | 2.1, 0.6, 0.9 |
| (G1 H8) | (G1 H1') | 4.0, 1.5, 2.0 |
| (G1 H8) | (G1 H2') | 3.0, 1.0, 1.5 |
| (G1 H8) | (G1 H3') | 4.0, 1.5, 2.0 |
| (G1 H8) | (G2 H8) | 4.0, 1.5, 2.0 |
| (G1 H1') | (G2 H8) | 4.0, 1.5, 2.0 |
| (G1 H2') | (G2 H8) | 3.0, 1.0, 1.5 |
| (G1 H3') | (G2 H8) | 4.0, 1.5, 2.0 |
| (G1 H1) | (C16 H42) | 3.0, 1.0, 1.5 |
| (G1 H1) | (C16 H41) | 3.0, 1.0, 1.5 |
| (G1 H1') | (A17 H2) | 4.0, 1.5, 2.0 |
| (G2 H22) | (G2 H1) | 3.0, 1.0, 1.5 |
| (G2 H22) | (G2 H21) | 2.1, 0.6, 0.9 |
| (G2 H21) | (G2 H1) | 3.0, 1.0, 1.5 |
| (G2 H1') | (G2 H2') | 2.1, 0.6, 0.9 |
| (G2 H2') | (G2 H3') | 2.1, 0.6, 0.9 |
| (G2 H8) | (G2 H1') | 4.0, 1.5, 2.0 |
| (G2 H8) | (G2 H2') | 3.0, 1.0, 1.5 |
| (G2 H8) | (G2 H3') | 4.0, 1.5, 2.0 |
| (G2 H2') | (U3 H5) | 4.0, 1.5, 2.0 |
| (G2 H1') | (U3 H6) | 4.0, 1.5, 2.0 |
| (G2 H2') | (U3 H6) | 3.0, 1.0, 1.5 |
| (G2 H1) | (C15 H42) | 3.0, 1.0, 1.5 |
| (G2 H1) | (C15 H41) | 3.0, 1.0, 1.5 |
| (U3 H1') | (U3 H2') | 2.1, 0.6, 0.9 |
| (U3 H6) | (U3 H1') | 4.0, 1.5, 2.0 |
| (U3 H6) | (U3 H3') | 4.0, 1.5, 2.0 |
| (U3 H6) | (U3 H5) | 2.1, 0.6, 0.9 |
| (U3 H6) | (U3 H2') | 2.1, 0.6, 0.9 |
| (U3 H3) | (G4 H1) | 3.0, 1.0, 1.5 |
| (U3 H1') | (G4 H8) | 4.0, 1.5, 2.0 |
| (U3 H2') | (G4 H8) | 3.0, 1.0, 1.5 |
| (U3 H3) | (A14 H2) | 3.0, 1.0, 1.5 |
| (G4 H1') | (G4 H2') | 2.1, 0.6, 0.9 |
| (G4 H2') | (G4 H3') | 2.1, 0.6, 0.9 |
| (G4 H8) | (G4 H1') | 4.0, 1.5, 2.0 |
| (G4 H8) | (G4 H2') | 3.0, 1.0, 1.5 |
| (G4 H8) | (G4 H3') | 4.0, 1.5, 2.0 |
| (G4 H8) | (G4 H4') | 4.0, 1.5, 2.0 |
| (G4 H1') | (A5 H8) | 4.0, 1.5, 2.0 |
| (G4 H2') | (A5 H8) | 3.0, 1.0, 1.5 |
| (G4 H3') | (A5 H8) | 4.0, 1.5, 2.0 |
| (G4 H1) | (C13 H41) | 3.0, 1.0, 1.5 |
| (G4 H1) | (C13 H42) | 3.0, 1.0, 1.5 |
| (A5 H1') | (A5 H2') | 2.1, 0.6, 0.9 |
| (A5 H1') | (A5 H3') | 3.0, 1.0, 1.5 |
| (A5 H62) | (A5 H61) | 2.1, 0.6, 0.9 |
| (A5 H2) | (A5 H1') | 4.0, 1.5, 2.0 |
| (A5 H8) | (A5 H1') | 4.0, 1.5, 2.0 |
| (A5 H8) | (A5 H3') | 3.0, 1.0, 1.5 |
| (A5 H2') | (U6 H5) | 4.0, 1.5, 2.0 |
| (A5 H3') | (U6 H5) | 4.0, 1.5, 2.0 |
| (A5 H2) | (U6 H1') | 4.0, 1.5, 2.0 |
| (A5 H1') | (U6 H6) | 4.0, 1.5, 2.0 |
| (A5 H2') | (U6 H6) | 3.0, 1.0, 1.5 |
| (A5 H3') | (U6 H6) | 4.0, 1.5, 2.0 |
| (A5 H2) | (U12 H3) | 3.0, 1.0, 1.5 |
| (A5 H62) | (U12 H3) | 3.0, 1.0, 1.5 |
| (A5 H61) | (U12 H3) | 3.0, 1.0, 1.5 |
| (A5 H2) | (C13 H1') | 4.0, 1.5, 2.0 |
| (U6 H1') | (U6 H2') | 2.1, 0.6, 0.9 |
| (U6 H1') | (U6 H3') | 4.0, 1.5, 2.0 |
| (U6 H6) | (U6 H1') | 4.0, 1.5, 2.0 |
| (U6 H6) | (U6 H2') | 3.0, 1.0, 1.5 |
| (U6 H6) | (U6 H3') | 4.0, 1.5, 2.0 |

| | | |
|---|---|---|
| (U6 H6) | (U6 H4') | 4.0, 1.5, 2.0 |
| (U6 H6) | (U6 H5) | 2.1, 0.6, 0.9 |
| (U6 H2') | (U7 H2') | 4.0, 1.0, 1.5 |
| (U6 H2') | (U7 H5) | 4.0, 1.0, 1.5 |
| (U6 H1') | (U7 H6) | 4.0, 1.5, 2.0 |
| (U6 H2') | (U7 H6) | 3.0, 1.0, 1.5 |
| (U6 H3) | (U11 H3) | 2.5, 0.7, 1.2 |
| (U6 H3) | (U12 H3) | 4.0, 1.0, 1.5 |
| (U7 H6) | (U7 H1') | 4.0, 1.5, 2.0 |
| (U7 H6) | (U7 H2') | 3.0, 1.0, 1.5 |
| (U7 H6) | (U7 H3') | 4.0, 1.5, 2.0 |
| (U7 H6) | (U7 H5) | 2.1, 0.6, 0.9 |
| (U7 H6) | (U7 H51) | 4.0, 1.5, 2.0 |
| (U7 H1') | (A9 H8) | 4.0, 1.5, 2.0 |
| (U7 H2') | (A9 H8) | 4.0, 1.5, 2.0 |
| (U7 H3') | (A9 H8) | 4.0, 1.5, 2.0 |
| (U7 H51) | (A9 H8) | 4.0, 1.5, 2.0 |
| (U7 H3) | (U11 H3) | 3.5, 0.8, 2.0 |
| (G8 H21) | (G8 H22) | 2.1, 0.6, 0.9 |
| (G8 H1') | (G8 H2') | 2.1, 0.6, 0.9 |
| (G8 H1') | (G8 H3') | 4.0, 1.5, 2.0 |
| (G8 H3') | (G8 H2') | 2.1, 0.6, 0.9 |
| (G8 H4') | (G8 H3') | 3.0, 1.0, 1.5 |
| (G8 H51) | (G8 H4') | 3.0, 1.0, 1.5 |
| (G8 H8) | (G8 H1') | 3.0, 1.0, 1.0 |
| (G8 H8) | (G8 H2') | 3.5, 1.0, 1.5 |
| (G8 H8) | (G8 H3') | 3.0, 1.0, 1.2 |
| (G8 H8) | (G8 H4') | 4.0, 1.0, 1.0 |
| (G8 H8) | (G8 H5') | 3.0, 1.0, 1.0 |
| (G8 H1') | (A9 H8) | 4.0, 1.5, 2.0 |
| (G8 H2') | (A9 H5') | 2.2, 1.0, 1.0 |
| (G8 H2') | (A9 H8) | 3.0, 1.0, 1.5 |
| (G8 H4') | (A9 H8) | 4.0, 1.5, 2.0 |
| (A9 H1) | (A9 H2') | 2.1, 0.6, 0.9 |
| (A9 H3') | (A9 H2') | 3.0, 1.0, 1.5 |

| | | |
|---|---|---|
| (A9 H8) | (A9 H1') | 4.0, 1.5, 2.0 |
| (A9 H8) | (A9 H2') | 3.0, 1.0, 1.5 |
| (A9 H8) | (A9 H3') | 2.1, 0.6, 0.6 |
| (A9 H2) | (U11 H4') | 4.0, 1.0, 2.0 |
| (G10 H1') | (G10 H2') | 2.1, 0.6, 0.9 |
| (G10 H1') | (G10 H3') | 4.0, 1.5, 2.0 |
| (G10 H1') | (G10 H52) | 4.0, 1.5, 2.0 |
| (G10 H1') | (G10 H51) | 4.0, 1.5, 2.0 |
| (G10 H3') | (G10 H4') | 3.0, 1.0, 1.5 |
| (G10 H51) | (G10 H3') | 4.0, 1.5, 2.0 |
| (G10 H51) | (G10 H4') | 3.0, 1.0, 1.5 |
| (G10 H8) | (G10 H1') | 4.0, 1.5, 2.0 |
| (G10 H8) | (G10 H2') | 3.0, 1.0, 1.5 |
| (G10 H8) | (G10 H51) | 3.0, 1.5, 1.5 |
| (G10 H8) | (U11 H5) | 4.0, 1.5, 2.0 |
| (G10 H1') | (U11 H6) | 4.0, 1.5, 2.0 |
| (G10 H2') | (U11 H5) | 3.0, 1.5, 2.0 |
| (G10 H2') | (U11 H6) | 3.0, 1.5, 1.0 |
| (G10 H3') | (U11 H6) | 4.0, 1.5, 2.0 |
| (G10 H52) | (U11 H6) | 4.0, 1.5, 2.0 |
| (G10 H51) | (U11 H6) | 4.0, 1.5, 2.0 |
| (G10 H52) | (U11 H5) | 4.0, 1.5, 2.0 |
| (G10 H51) | (U11 H5) | 4.0, 1.5, 2.0 |
| (G10 H8) | (U11 H4') | 4.0, 1.5, 2.0, |
| (G10 H8) | (U11 H5) | 4.0, 1.5, 2.0 |
| (U11 H2') | (U11 H3') | 3.0, 1.0, 1.5 |
| (U11 H6) | (U11 H2') | 3.0, 1.0, 1.5 |
| (U11 H6) | (U11 H3') | 4.0, 1.5, 2.0 |
| (U11 H6) | (U11 H4') | 4.0, 1.5, 2.0 |
| (U11 H6) | (U11 H5) | 2.1, 0.6, 0.9 |
| (U11 H3) | (U12 H3) | 3.0, 1.0, 1.5 |
| (U11 H4') | (U12 H6) | 4.0, 1.5, 2.0 |
| (U11 H2') | (U12 H4') | 4.0, 1.5, 2.0 |
| (U12 H1') | (U12 H2') | 2.1, 0.6, 0.9 |
| (U12 H1') | (U12 H3') | 4.0, 1.5, 2.0 |

| | | |
|---|---|---|
| (U12 H6) | (U12 H1') | 4.0, 1.5, 2.0 |
| (U12 H6) | (U12 H2') | 3.0, 1.0, 1.5 |
| (U12 H6) | (U12 H3') | 4.0, 1.5, 2.0 |
| (U12 H6) | (U12 H4') | 4.0, 1.5, 2.0 |
| (U12 H6) | (U12 H5) | 2.1, 0.6, 0.9 |
| (C13 H42) | (C13 H41) | 2.1, 0.6, 0.9 |
| (C13 H1') | (C13 H2') | 2.1, 0.6, 0.9 |
| (C13 H2') | (C13 H3') | 2.1, 0.6, 0.9 |
| (C13 H6) | (C13 H1') | 4.0, 1.5, 2.0 |
| (C13 H6) | (C13 H2') | 3.0, 1.0, 1.5 |
| (C13 H6) | (C13 H3') | 4.0, 1.5, 2.0 |
| (C13 H6) | (C13 H5) | 2.1, 0.6, 0.9 |
| (C13 H1') | (A14 H8) | 4.0, 1.5, 2.0 |
| (C13 H2') | (A14 H8) | 3.0, 1.0, 1.5 |
| (A14 H1') | (A14 H2') | 2.1, 0.6, 0.9 |
| (A14 H2) | (A14 H1') | 4.0, 1.5, 2.0 |
| (A14 H3') | (A14 H4') | 3.0, 1.0, 1.5 |
| (A14 H8) | (A14 H1') | 4.0, 1.5, 2.0 |
| (A14 H8) | (A14 H3') | 3.0, 1.0, 1.5 |
| (A14 H8) | (A14 H4') | 4.0, 1.5, 2.0 |
| (A14 H2') | (C15 H6) | 3.0, 1.0, 1.5 |
| (A14 H1') | (C15 H6) | 4.0, 1.5, 2.0 |
| (A14 H2) | (C15 H1') | 4.0, 1.5, 2.0 |
| (A14 H8) | (C15 H6) | 4.0, 1.5, 2.0 |
| (C15 H1') | (C15 H2') | 2.1, 0.6, 0.9 |
| (C15 H42) | (C15 H41) | 2.1, 0.6, 0.9 |
| (C15 H1') | (C15 H3') | 4.0, 1.5, 2.0 |
| (C15 H3') | (C15 H2') | 3.0, 1.0, 1.5 |
| (C15 H3') | (C15 H4') | 3.0, 1.0, 1.5 |
| (C15 H6) | (C15 H1') | 4.0, 1.5, 2.0 |
| (C15 H6) | (C15 H3') | 3.0, 1.0, 1.5 |
| (C15 H6) | (C15 H5) | 3.0, 1.0, 1.5 |
| (C15 H3') | (C16 H3') | 4.0, 1.5, 2.0 |
| (C15 H2') | (C16 H6) | 3.0, 1.0, 1.5 |
| (C16 H1') | (C16 H2') | 2.1, 0.6, 0.9 |
| (C16 H3') | (C16 H4') | 3.0, 1.0, 1.5 |
| (C16 H6) | (C16 H1') | 3.0, 1.0, 1.5 |
| (C16 H6) | (C16 H3') | 3.0, 1.0, 1.5 |
| (C16 H6) | (C16 H4') | 3.0, 1.0, 1.5 |
| (C16 H6) | (C16 H5) | 2.1, 0.6, 0.9 |
| (C16 H5) | (C16 H6) | 2.1, 0.6, 0.9 |
| (C16 H5) | (C16 H42) | 3.0, 1.0, 1.5 |
| (C16 H5) | (C16 H41) | 3.0, 1.0, 1.5 |
| (C16 H42) | (C16 H41) | 2.1, 0.6, 0.9 |
| (C16 H1') | (A17 H8) | 4.0, 1.5, 2.0 |
| (C16 H2') | (A17 H8) | 3.0, 1.0, 1.5 |
| (C16 H6) | (A17 H8) | 4.0, 1.5, 2.0 |
| (A17 H1') | (A17 H2') | 2.1, 0.6, 0.9 |
| (A17 H1') | (A17 H3') | 4.0, 1.5, 2.0 |
| (A17 H1') | (A17 H4') | 4.0, 1.5, 2.0 |
| (A17 H3') | (A17 H2') | 3.0, 1.0, 1.5 |
| (A17 H8) | (A17 H1') | 4.0, 1.5, 2.0 |
| (A17 H8) | (A17 H2') | 3.0, 1.0, 1.5 |
| (A17 H8) | (A17 H3') | 4.0, 1.5, 2.0 |
| (A17 H8) | (A17 H4') | 4.0, 1.5, 2.0 |

**Table 2.2:** NOEs used for molecular modeling with distance bins for restraints. The first number is the median distance. The second number is the maximum distance above the median, and the third number is the minimum below the median distance.

### 2.3.4 DQF-COSY.

An unusually strong and downfield shifted H1'-H2' NOE was observed and assigned to G10 which indicates partial C2' endo conformation (Figure 2.7). In the preliminary structure calculations, the G10 sugar pucker was unrestrained and the ensembles of structures showed a mixture of C2'endo and C3'endo. The structure was then further refined with restraints for G10 to be C2'endo in order to improve the convergence of the hairpin nucleotides. The COSY experiment also confirms that the sugar of A17 exhibits C2'- endo characteristics as expected for a 3' dangling end nucleotide (*125*).

**Figure 2.7: Double Quantum Filtered COSY Spectrum**

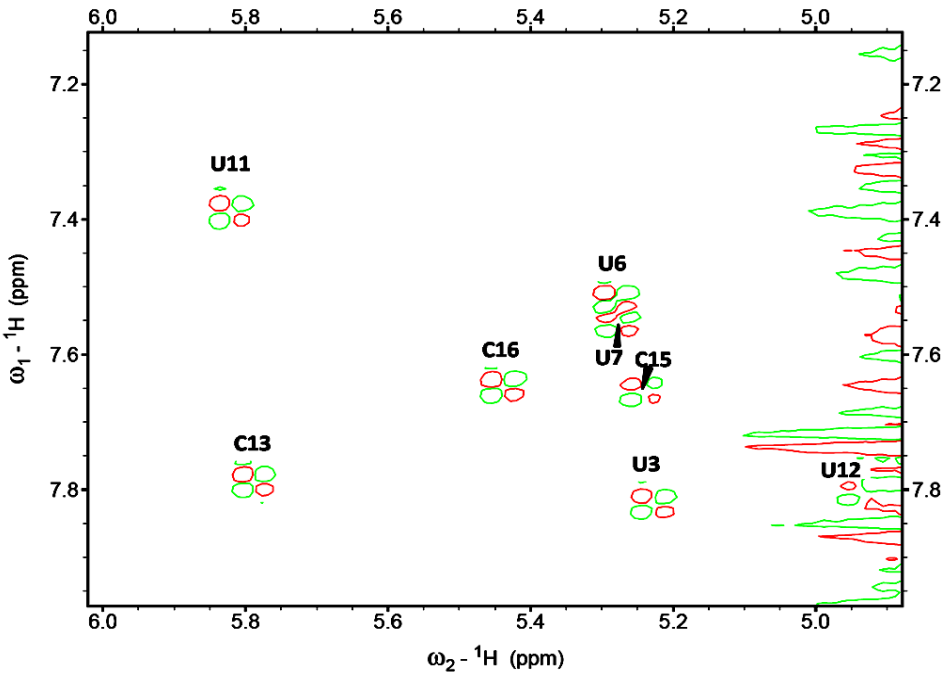



**Figure 2.7**: COSY spectrum showing H5-H6 pyrimidine crosspeaks. Spectrum was taken at 25° C in 10 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $H_2O$ with 10% $D_2O$.

### 2.3.5 $^{1}H$- $^{13}C$ HSQC.

The $^{1}H$-$^{13}C$ HSQC confirms key assignments for sugar proton resonances that were overlapped or had unusual chemical shifts. The G10 H1' and the A9 H1' both resonate at 6.18 ppm, respectively, but the carbon chemical shifts for A9 C1' and G10 C1' occur at 87.9 and 93.3 ppm, respectively (Figure 2.8). The U11 H1' resonates at ~4.8 ppm in the $^{1}H$-$^{13}C$ HSQC. G10 H2' and G1 H2' are shifted downfield relative to the HDO peak.

The H2 protons of A5, A9, A14, and A17 were distinguished from all other protons by their carbon chemical shift (Figure 2.9). Adenine H2 assignments were based on NOEs observed in both the $H_2O$ and $D_2O$ NOESY spectra. In the $H_2O$ NOESY, both A5 and A14 H2 protons give cross strand NOEs to U12 and U3 H3 protons, respectively. The A5 H2 proton also displays a cross-strand NOE to C13 H1'. The A9 H2 proton is shifted downfield and gives a NOE to its own H1' sugar proton as well as an N + 2 NOE to U11 H4'. The A14 H2 proton gives a NOE to its own H1' sugar proton and also to the C15 H1' proton. The A17 base was added to the end of the sequence to improve the stability of the duplex and increase the probability of forming the terminal base pair (*125*). The A17 H2 proton gives a cross-strand NOE to G1 H1', which verifies that A17 is stacked on the terminal G1-C16 base pair.

**Figure 2.8: $^{13}$C Ribose HSQC Spectrum**

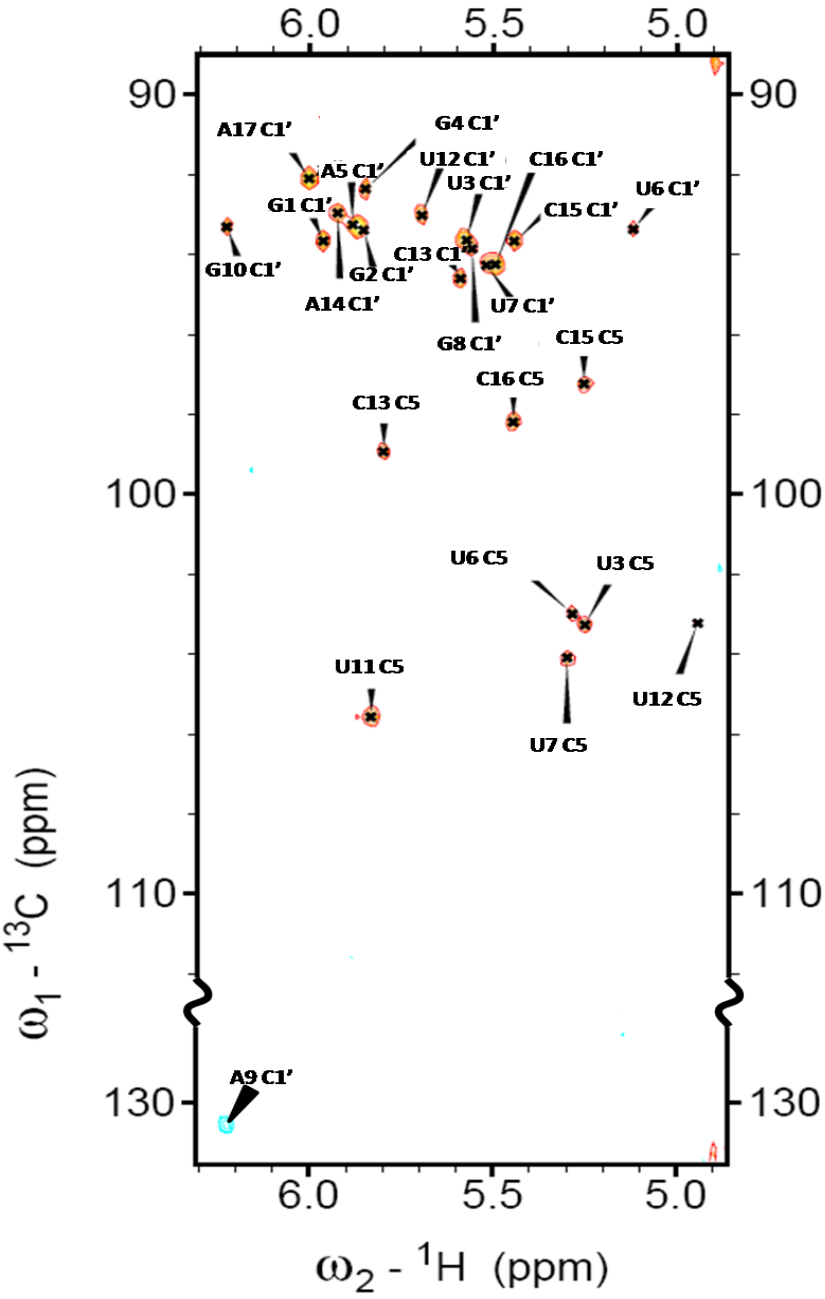



**Figure 2.8:** Ribose $^{13}$C HSQC showing the C1' and C5 region of the spectra. The A9 C1' resonance has been folded in and resonating at ~130 ppm. Spectrum was taken at 20$^{o}$ C in 10 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in 100% $D_2O$.

**Figure 2.9: $^{13}$C Aromatic HSQC Spectrum**

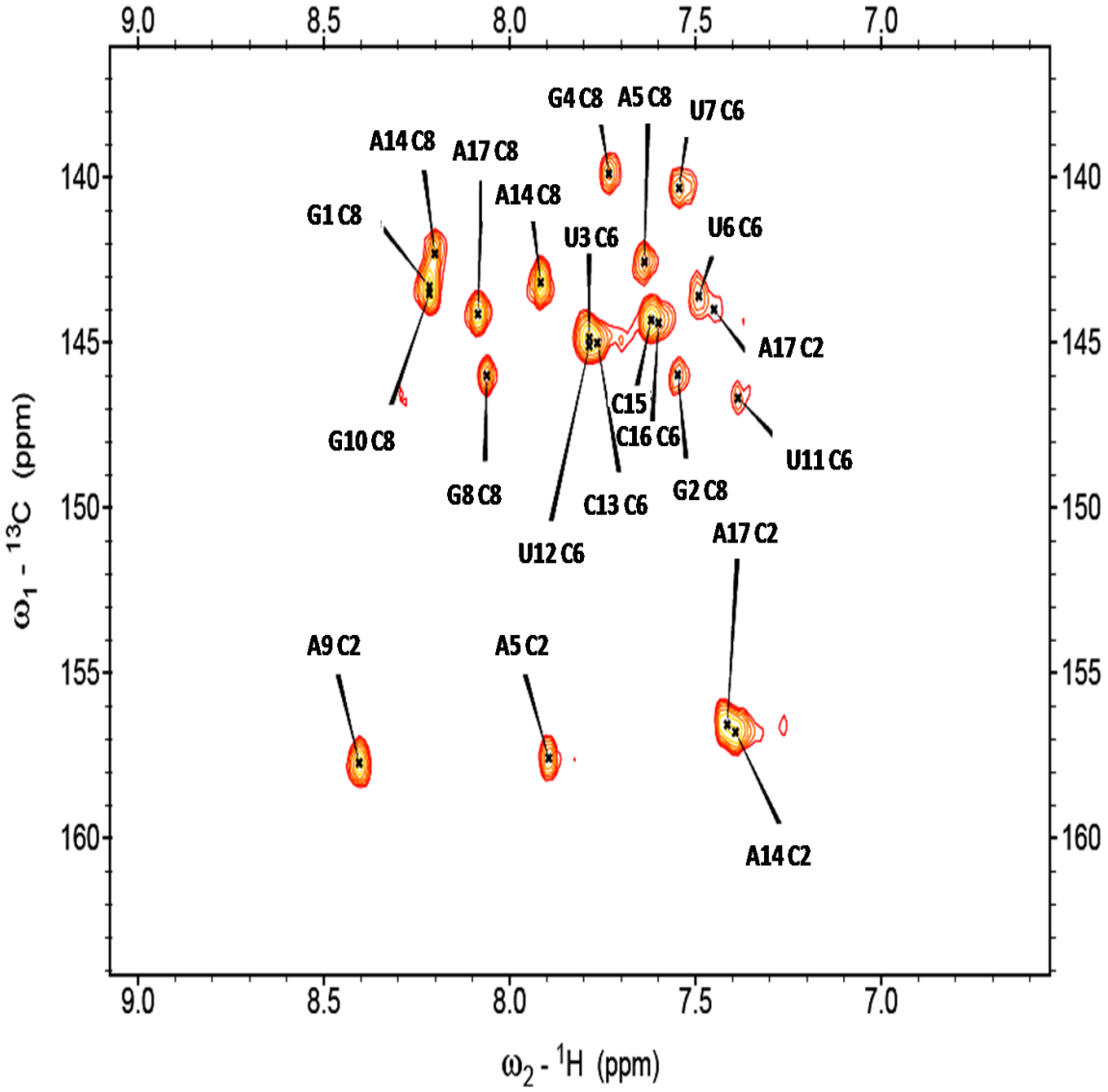



**Figure 2.9:** Aromatic $^{13}$C HSQC showing C2, C6 and C8 carbon resonances. Spectrum was taken at 20$^{o}$ C in 10 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in 100% $D_2O$.

### 2.3.6 $^{1}H$-$^{31}P$ HETCOR.

The phosphorus chemical shifts are within the normal range for A-form RNA except for A9 P and A17 P. In the final ensemble of structures, the A9 backbone angle ranges from -99.6 to -162.1 degrees. The A9 H3' is shifted downfield relative to the HDO peak (Figure 2.9). This assignment was also confirmed by the observation of a single peak having the carbon chemical shift of an H3' proton in the $^{1}H$-$^{13}C$ HSQC and an A9H8 to A9H3' NOE at 8.065 and 5 ppm in the100 ms $D_2O$ NOESY.

**Figure 2.9: $^{31}P$ HETCOR Spectrum**

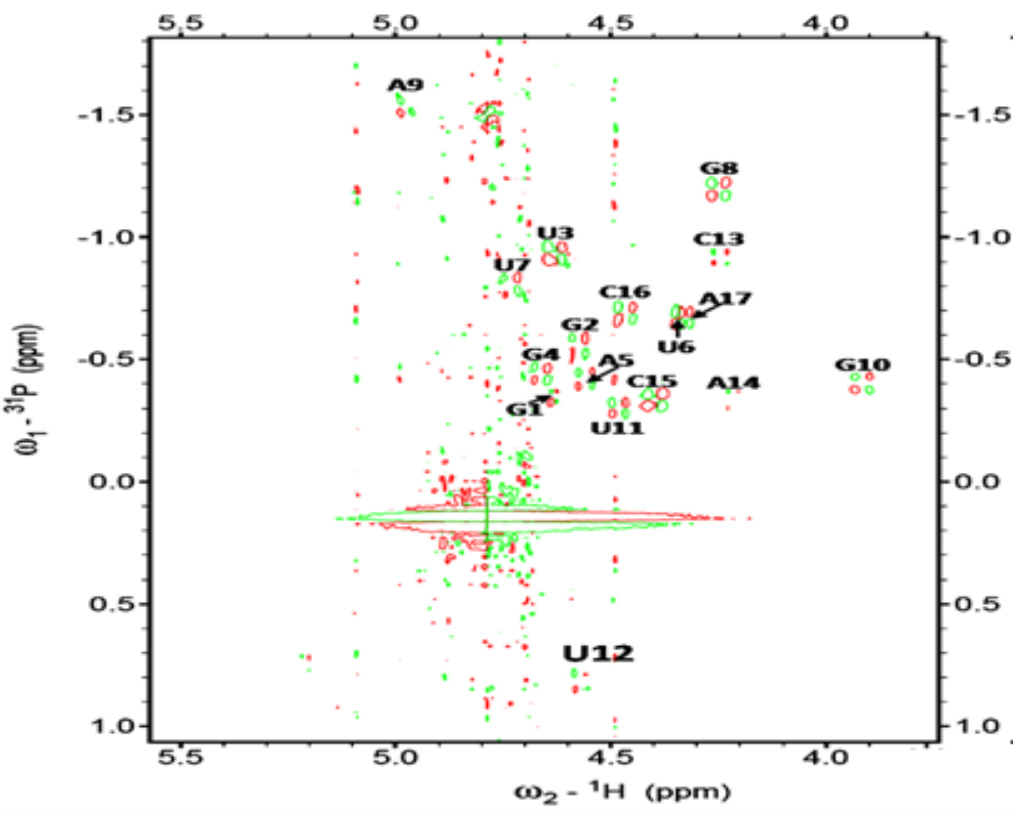



**Figure 2.9:** $^{31}P$ HETCOR showing C3' resonances. Spectrum was taken at 20$^{o}$ C in 10 mM NaCl, 10 mM $NaPO_4$, 0.5 mM $Na_2EDTA$ pH 6 in 100% $D_2O$.

**Table 2.3**: **Carbon and Phosphorus Resonances**

| | A C2 | C5 | C6/C8 | C1' | C2' | C3' | P |
|---|---|---|---|---|---|---|---|
| G1 | - | - | 143.3 | 93.67 | 71.9 | 74.7 | -0.3534 |
| G2 | - | - | 146 | 93.38 | 71.8 | 74.6 | -0.5613 |
| U3 | - | 103.3 | 144.9 | 93.63 | 71.7 | 75 | -0.9328 |
| G4 | - | - | 139.9 | 92.36 | 71.4 | 74.1 | -0.4403 |
| A5 | 157.6 | - | 142.6 | 93.25 | 71.3 | 74.2 | -0.4206 |
| U6 | - | 103 | 143.6 | 93.37 | 71.4 | 75.4 | -0.672 |
| U7 | - | 104.1 | 140.3 | 94.27 | 74.1 | 74.5 | -0.8072 |
| G8 | - | - | 146 | 93.86 | 71.7 | 77.2 | -1.189 |
| A9 | 157.7 | - | 143.2 | 87.9 | 73.1 | 74.5 | -1.536 |
| G10 | - | - | 143.3 | 93.31 | 71.8 | 78.5 | -0.4017 |
| U11 | - | 105.6 | 146.7 | | | 74.8 | -0.3003 |
| U12 | - | 103.2 | 144.9 | 92.97 | 71.1 | 73.5 | 0.8143 |
| C13 | - | 98.93 | 145 | 94.59 | 71.3 | 73.7 | -0.9186 |
| A14 | 156.6 | - | 142.3 | 92.95 | 71.5 | 74.9 | -0.3293 |
| C15 | - | 97.24 | 144.4 | 93.65 | 71.2 | 75.3 | -0.3391 |
| C16 | - | 98.18 | 144.3 | 94.25 | 72.1 | 74.9 | -0.6962 |
| A17 | 156.8/144 | - | 144.2 | 92.10 | 69 | 76.8 | -0.6742 |

**Table 2.3**: Carbon and phosphorus chemical shift assignments. The carbon resonances were referenced to an external standard and the phosphorus resonances were referenced to phosphate buffer at pH 6.

### 2.3.7 Hairpin Structure.

Figure 2.10 shows the UU pair, syn G10, and U-turn in the phi29 E-loop hairpin. The structure was calculated from 195 measured NOE restraints (Table 2.2) as well as hydrogen bonding and dihedral angle restraints for the A-form stem helix. Thirty-one structures were calculated and had an overall rmsd of 1.50 Å from starting structures with an rmsd of 6.08 Å. When G8. G10, and A17 are excluded from the rmsd calculation, the overall rmsd is 1.24 Å. Figure 2.11A shows the overlay of ten lowest energy structures. Thus, the ensemble of calculated structures converges well, has a low final energy, and contains no NOE violations. All the expected NOES in the average structure are observed in the NMR spectra. The hairpin turn is defined by a total of 33 internucleotide and 46 intranucleotide NOEs for six nucleotides and an all-atom RMSD of 1.76 Å (Table 2.5). The U-turn occurs in all members of the ensemble. Table 2.1 lists the number of intranucleotide NOEs, internucleotide NOEs, and the all-atom RMSD for each nucleotide. The UU pair is well-defined by 26 internucleotide NOEs and has an all-atom RMSD of 1.23 and 1.45 Å for U6 and U11, respectively. Figure 2.11B focuses on the UU pairs in the overlay of ten lowest energy structures. Fewer internucleotide NOEs define the positions of G8, A9, and G10, which is consistent with the Watson-Crick edges of these bases facing the solvent. G8 and G10 positions are defined by mainly intranucleotide NOEs. These nucleotides are likely to be somewhat structurally dynamic in solution. The partial overlap of H8 and C8 resonances for G8, A9, and G10 combined with the G8 and G10 imino protons in fast exchange with water precludes further T1

or residual dipolar coupling measurements to further define the positions and dynamics of these bases. G10 projects out of the hairpin and into the solvent, and the ensemble positions of G10 are consistent with the strong intranucleotide H1'-H8 NOE defining the syn conformation and the H1'-H2' COSY crosspeak defining the C2'endo sugar pucker.

**Figure 2.10: phi29 E-loop Hairpin Wall-eye Stereo Image**

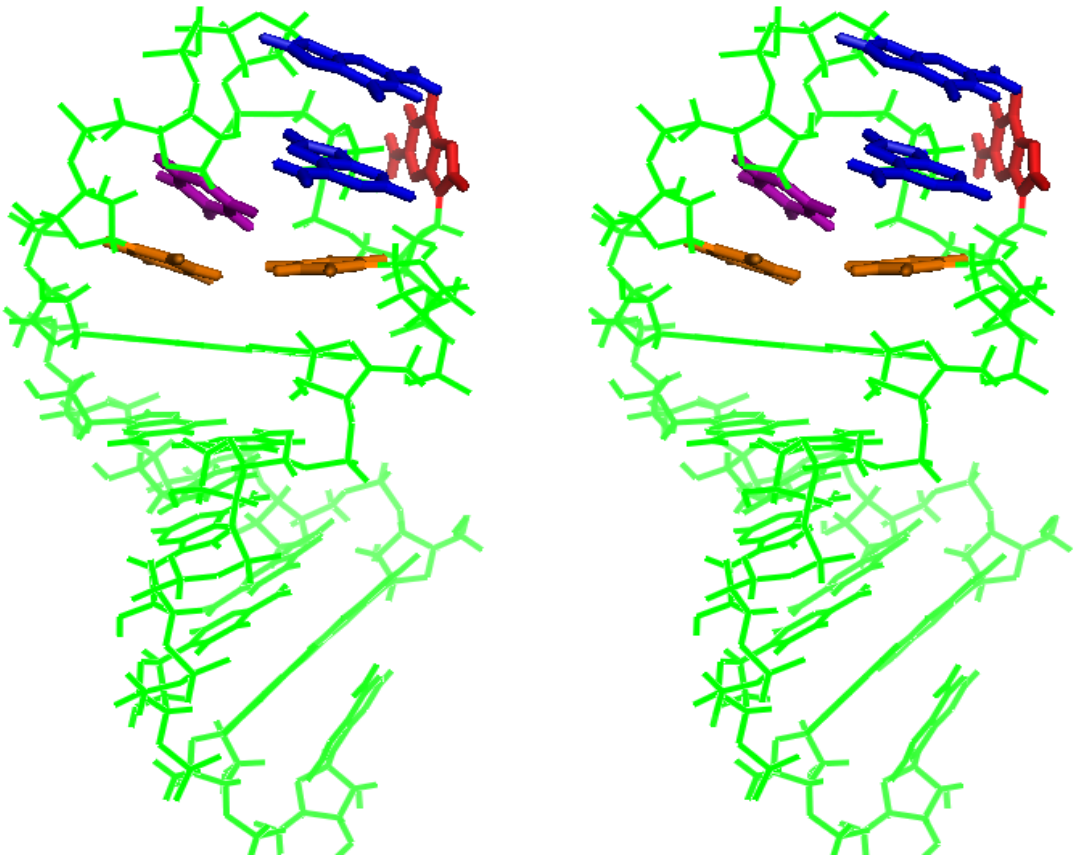


**Figure 2.10:** Wall-eye stereo image of the average structure of the prohead RNA E-loop hairpin, r(GGUGAUUGAGUUCACCA). The stem bases are green, U6 and U11 are orange, U7 is purple, G8 and A9 are blue and G10 is red.

**Figure 2.11**: **phi29 E-loop Hairpin Overlay Images**

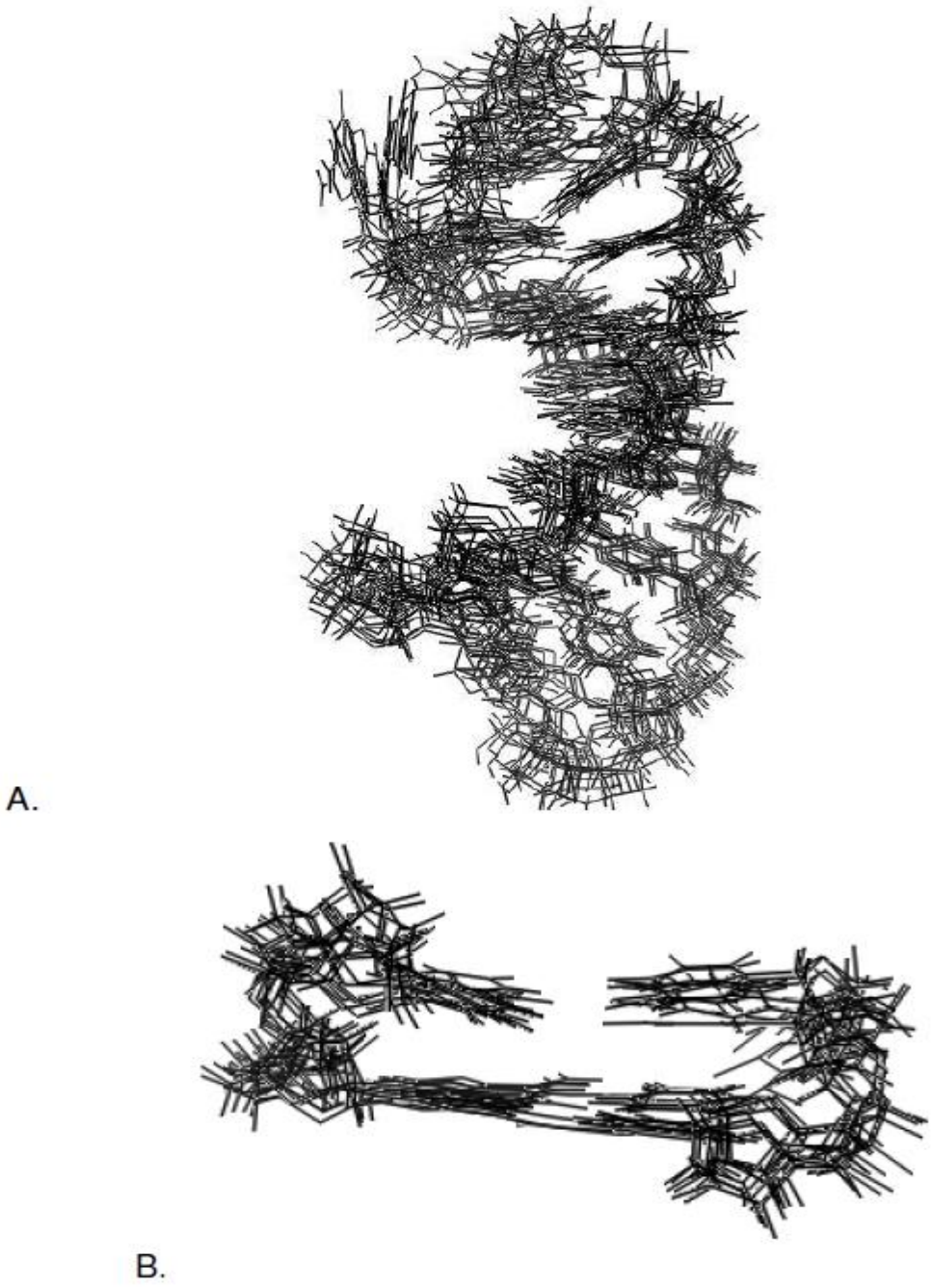



**Figure 2.11**: A. Overlay of the 10 lowest energy conformers from the ensemble of 31 structures. Note that the ribose backbone for nucleotides G8-U11 comes across the front of the hairpin. B. The U6-U11 base pair stacked on the A5-U12 pair in the overlay of the 10 lowest energy structures.

**Table 2.4: Restraints and Ensemble Convergence for Phi29 E-Loop Hairpin**

| Base | Intra | inter | Dihedral | RMSD (Å) |
|---|---|---|---|---|
| G1 | 8 | 7 | 3 | 1.76 |
| G2 | 8 | 9 | 5 | 1.38 |
| U3 | 5 | 7 | 5 | 1.05 |
| G4 | 6 | 8 | 5 | 1.00 |
| A5 | 7 | 13 | 3 | 1.08 |
| U6 | 7 | 12 | 0 | 1.23 |
| U7 | 5 | 9 | 0 | 1.33 |
| G8 | 11 | 4 | 0 | 3.49 |
| A9 | 5 | 11 | 0 | 1.83 |
| G10 | 10 | 9 | 2 | 2.34 |
| U11 | 5 | 17 | 0 | 1.45 |
| U12 | 7 | 7 | 3 | 1.35 |
| C13 | 7 | 5 | 5 | 1.08 |
| A14 | 6 | 7 | 5 | 0.92 |
| C15 | 8 | 8 | 5 | 0.90 |
| C16 | 10 | 7 | 5 | 1.02 |
| A17 | 8 | 4 | 3 | 2.23 |
| Total | 123 | 72 | 47 | 1.50 |

**Table 2.4:** Bases are numbered from the 5' to 3' direction. Intra refers to NOES between protons on the same nucleotide. Inter refers to NOES between protons on different nucleotides. Note that each NOE occurs between two protons, so the total internucleotide NOEs are the sum of the internucleotide NOEs for each nucleotide divided by 2. RMSD is an abbreviation for the all-atom root mean square deviation of the 31 structures in the final ensemble.

**Table 2.5: Preliminary Hairpin Refinements**

| | Preliminary Refinement | | | Additional Hairpin Restraints | | |
|---|---|---|---|---|---|---|
| Base | inter | Dihedral | RMSD (Å) | inter | Dihedral | RMSD (Å) |
| G1 | 7 | 3 | 1.71 | 7 | 3 | 1.76 |
| G2 | 9 | 5 | 1.22 | 9 | 5 | 1.38 |
| U3 | 7 | 5 | 1.08 | 7 | 5 | 1.05 |
| G4 | 8 | 5 | 1.06 | 8 | 5 | 1.00 |
| A5 | 13 | 3 | 1.28 | 13 | 3 | 1.08 |
| U6 | 12 | 0 | 1.41 | 12 | 0 | 1.23 |
| U7 | 9 | 0 | 1.63 | 9 | 0 | 1.33 |
| G8 | 3 | 0 | 4.05 | **4** | 0 | 3.49 |
| A9 | 8 | 0 | 2.21 | **11** | 0 | 1.83 |
| G10 | 9 | 0 | 2.98 | 9 | **2** | 2.34 |
| U11 | 15 | 0 | 1.63 | **17** | 0 | 1.45 |
| U12 | 7 | 3 | 1.44 | 7 | 3 | 1.35 |
| C13 | 5 | 5 | 0.99 | 5 | 5 | 1.08 |
| A14 | 7 | 5 | 0.85 | 7 | 5 | 0.92 |
| C15 | 8 | 5 | 0.88 | 8 | 5 | 0.90 |
| C16 | 7 | 5 | 1.20 | 7 | 5 | 1.02 |
| A17 | 4 | 3 | 2.03 | 4 | 3 | 2.23 |
| Total | 69 | 47 | 1.62 | 72 | 47 | 1.50 |

**Table 2.5:** The conservative preliminary refinement did not restrain any of the hairpin nucleotide backbone angles in order to fully explore conformational space. The additional hairpin restraints include G10 syn, G10 C2'endo, and 3 ribose internucleotide NOEs. These additional restraints improve slightly the rmsd of the structure but do not significantly change the main structural features of the hairpin. Bases are numbered from the 5' to 3' direction. All intranucleotide NOEs are listed in Table 2.1. "Inter" refers to NOES between protons on different nucleotides. Note that each NOE occurs between two protons, so the total internucleotide NOEs are the sum of the internucleotide NOEs for each nucleotide divided by 2. RMSD is an abbreviation for the all-atom root mean square deviation of the structures in the final ensemble.

## 2.4 Discussion

### 2.4.1 Structural Features of the phi29 E-loop Hairpin.

The U-U mismatch in the hairpin loop is confirmed by NOEs in the $H_2O$ NOESY (Figure 2.3). The U6 H6 proton follows the NOESY sequential walk observed in the $D_2O$ NOESY, which is consistent with U6 stacking on A5 (*95*). U6-U11 form a 2-Carbonyl-N3, 4-Carbonyl-N3 interaction or shifted *cis*-Watson-Crick, Watson-Crick (*123*). This type of UU pair is similar to that observed in the S2 hairpin, r(5'GCGUUAAGUCGCA), found in the 30s ribosome (*92, 126, 128*). The stacking of U7 on the preceding U-U mismatch, which is confirmed by the cross-strand NOE observed in the $H_2O$ NOESY, is also similar to the S2 hairpin. The phi29 E-loop hairpin has an A-U closing base pair. There is no observed effect in stacking the U-U pair on the closing A-U pair when compared to the S2 hairpin with a closing G-C base pair.

G10 does not pair within the loop but projects into solution. G10 has a syn conformation and a partial C2'endo sugar pucker revealed in 50 ms NOESY and COSY spectra, respectively. The internucleotide NOEs for G10 include only G10 H8 to U11 H4' and G10 H5', H5" to U11 H6. The few internucleotide NOEs to G10 and the NOEs between A9 H2 and U11 H4' confirm that the G10 base projects into solution (Figure 2.6). The A9H2-U11H4' NOE indicates that A9 stacks on the U11 sugar, and this stacking interaction causes an upfield shift of the U11 H1' resonance.

The uridine turn observed in the phi29 E-loop hairpin occurs in many hairpins (*155-156*). The characteristic UNR sequence occurs in the phi29 E-loop (5' U6-U7-G8-A9-G10-U11), the S2 hairpin, and the anticodon and T-loop of yeast tRNA (*157-158*). The U7 H1' and U7 H2' sugar protons both give N + 2 NOEs to the A9 H8 proton which causes a kink in the backbone and is consistent with the U-turn location between U7 and G8.

### 2.4.2 Comparison of Prohead and Ribosomal Hairpin Structures.

Comparisons of prohead RNA E loop hairpins and ribosomal RNA hairpins suggest that the NMR structure of the phi29 E loop will likely represent the functional structure of the RNA in the phi29 packaging motor. The S2 hairpin is similar in sequence to the bacteriophage phi29, M2/NF, and SF5 prohead E loop sequences (Figures 2.13, 2.14 b, and 2.14 c, Table 2.6, Figure 2.15). The L11 hairpin has sequence similarity to the GA1 prohead E loop sequence (Figures 2.13 and 2.14, Table 2.6, Figure 2.15). The hairpins in rRNA motifs that bind small ribosomal subunit protein S2 and large ribosomal subunit protein L11 have been studied by NMR as isolated hairpins in solution and by crystallography of entire ribosomal subunits (*92, 124-129*). The rRNA conformations are exactly the same in isolated hairpins and in rRNA-protein complexes. In these cases, a small section of the rRNA does represent the functional structure in the larger context of the ribosome. Thus, the rRNA sequence determines the hairpin structure, which does not change upon protein binding or tertiary RNA interactions.

The structures in Figures 2.13 and 2.14 show the conserved features in hairpins with similar sequences. All of the hairpins in the comparison are characterized by a non-canonical base pair stacked on the helix and a uridine stacked on the non-canonical base pair. All of the hairpins also incorporate the U-turn motif, which causes the two subsequent bases to stack upon one another. In the phi29 E-loop and the S2 hairpin, the G10 base is protruding into solution, but the GA-1 E-loop prediction and the L11 protein binding RNA have this base stacked under 5' purines. This difference in purine stacking may be the result of the different backbone orientation for the first cis UU or trans GA pair.

### 2.4.3 Comparison of Phi29 NMR Hairpin Structure to MC-SYM Structure Predictions.

The lowest energy MC-SYM predicted structure has an overall 2.48 Å RMSD with respect to the experimentally determined NMR structure of phi29 E-loop hairpin (Figure 2.14). Both structures have a UU pair, a U stacked on the UU pair, and a U-turn motif. The differences occur in the stacking of G10, the nucleotide with a syn conformation, and the subtleties of the UU pair hydrogen bonding. The energetically unfavorable syn conformation of G10 is balanced by a very favorable conformation of U6, U7, and U11 in the NMR structure. In contrast in the MC-SYM predicted structure, G10 is stacked in a favorable conformation and the UU pair is in a less favorable conformation. The predicted structure shows G10 stacked under G8 and A9 rather than extending into solution. The U-U mismatch in the MC-SYM prediction shows the N-3 imino

proton of the U bases directly facing each other. This orientation does not allow the C-2 or C-4 carbonyl close enough to the imino protons to hydrogen bond (4 Å). In the phi29 NMR structure, the shifted cis- Watson-Crick, Watson-Crick UU pair has two hydrogen bonds (2 carbonyl-N3, 4 carbonyl-N3). In the MC-SYM prediction, U7 is oriented in the syn conformation, which is not observed in the NMR structure. The U7 imino to U11 imino NOEs in the $H_2O$ NOESY define the U7 stacking orientation in the phi29 NMR structure. The MC-SYM predicted structure violates only two NOEs: U7H3-U11H3 and A9H2-U11H4'. The U7H3-U11H3 NOE is not compatible with a U7 syn conformation, and the A9H2-U11H4' NOE is not compatible with G10 stacked below the 5'purines. If these two NOEs are included in the MCSYM simulation, then the predicted structure has a 1.94 Å rmsd with the NMR structure (Figure 2.15).

A comparison of ribosomal hairpins and MC-SYM predictions demonstrates the generally accurate predictive ability of MC-SYM. The rmsd was 1.5 Å between the lowest energy structure for the MC-Sym predicted L11 hairpin and the crystal structure (Figure 2.14). The lowest energy S2 hairpin structure compared to the crystal structure has an 1.4 Á RMSD (Figure 2.15) for the loop nucleotides. These comparisons provide a basis for evaluating the MC-SYM predicted structures relative to the phi29 NMR structure and confidence in the predicted structure for GA1 E-loop hairpin.

Improvements in the force field parameters for RNA and inclusion of more diverse structures in the database could further improve MC-SYM predictions. The AMBER forcefield uses parameters optimized for the DNA phosphodiester

backbone and deoxyribose angles, rather than RNA backbone conformations (*146-148*). A more polarizeable force field optimized for RNA could improve predictions of the ways in which RNA hairpin backbones fold (*159-160*). As the diversity of experimentally determined RNA structures increases, the database for optimizing MC-SYM scoring functions expands, thus improving prediction capabilities for novel structures. While research to improve RNA force fields and structural databases continues, the incorporation of experimental data, such as NOEs, can refine MC-SYM predictions.

**Table 2.6: Comparison of Prohead and Ribosomal Hairpin Structures**

| **Hairpin** | **First Pair** | **C2' endo** | **Syn** |
|---|---|---|---|
| phi29 NMR | UU cis-WC/WC | G10 | G10 |
| phi29 MC-SYM | UU cis-WC/WC | None | U7 |
| S2 NMR[a] | UU cis-WC/WC | G8,G10 | None |
| L11 NMR[b] | GA trans-S/H | U9,G10 | None |
| GA-1 MC-SYM | GA trans-S/H | None | None |

**Table 2.6**: The first pair conformations are described by the isostericity matrix nomenclature (*123*). C2' endo refers to hairpin nucleotides with the C2' endo sugar pucker conformation. Syn refers to the conformation of the glycosidic bond between the ribose and base. All numbering is consistent with the phi29 NMR construct. [a] ref (*126*),(*161*). [b] ref. (*125*).

**Table 2.7: TSP Referencing table**

| Temp ($^{o}$C) | $H_2O$ Resonance (ppm) |
|---|---|
| 1 | 5.054 |
| 5 | 5.007 |
| 10 | 4.952 |
| 15 | 4.895 |
| 20 | 4.842 |
| 25 | 4.789 |
| 30 | 4.738 |
| 35 | 4.687 |

**Table 2.7:** Proton spectra were referenced to water which was referenced to 3-(Trimethylsilyl)propionic-2,2,3,3-$d_4$ acid (TSP). Data were collected at different temperatures in 10mM NaCl, 10 mM $NaPO_4$, and 0.5 mM NaEDTA with 10% $D_2O$ at pH 6. TSPs chemical shift doesn't shift upon temperature change. Spectra were taken in 5$^{o}$C increments in order to determine the change in chemical shift of water at different temperatures. All spectra were referenced to the location of water using this data.

**Figure 2.13: Hairpin Tertiary Structure Comparisons**

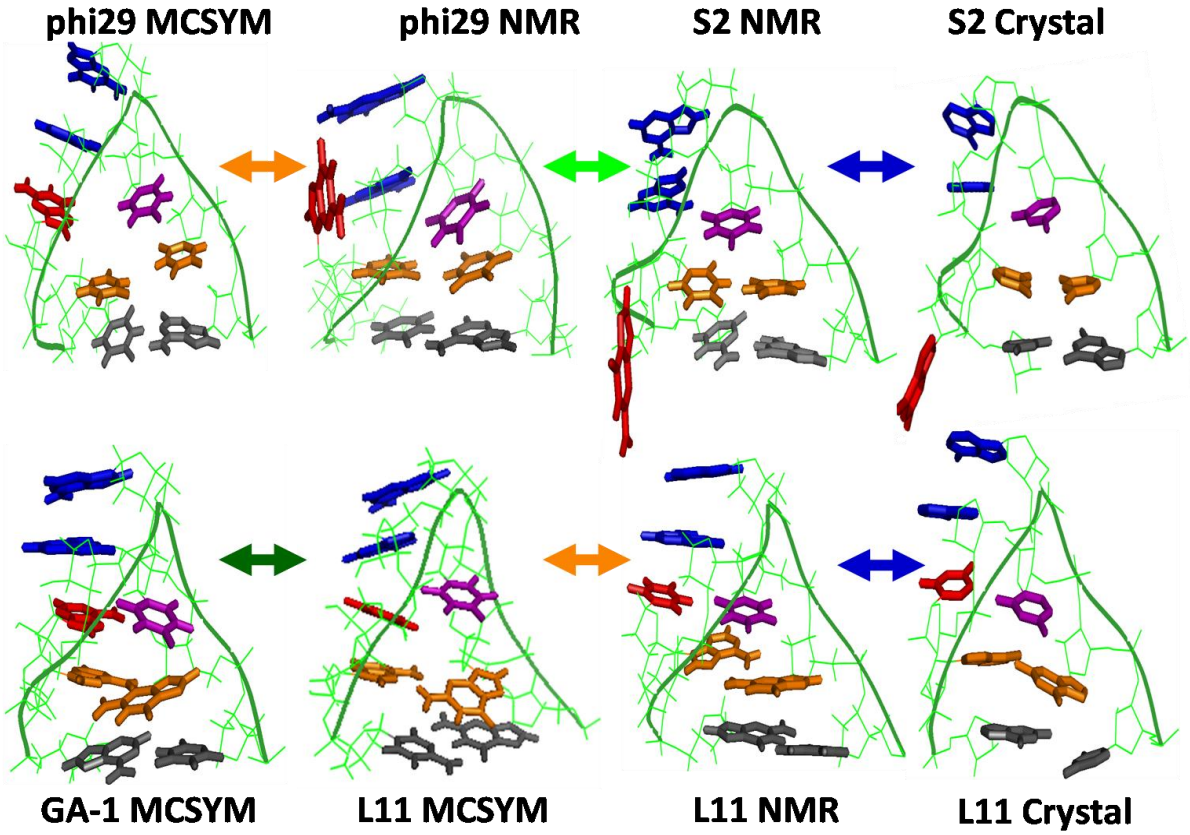



**Figure 2.13:** Comparison of prohead and ribosomal RNA hairpin three-dimensional structures. Blue arrows compare hairpins determined by NMR and crystallography. Orange arrows compare experimentally determined and predicted structures. The light and dark green arrows compare hairpins with different sequences and similar three-dimensional structures. The colors of the bases correspond to nucleotides in the same sequence position as the phi29 E-loop. The position of the G10 base in the S2 NMR hairpin is poorly defined and only one average position is shown. PDB entries 2WRJ (*162*) and 3HUW (*163*) were used to generate images of S2 and L11 hairpins in ribosome crystal structures.

**Figure 2.14: 2-Dimensional Hairpin Comparisons and Contacts**

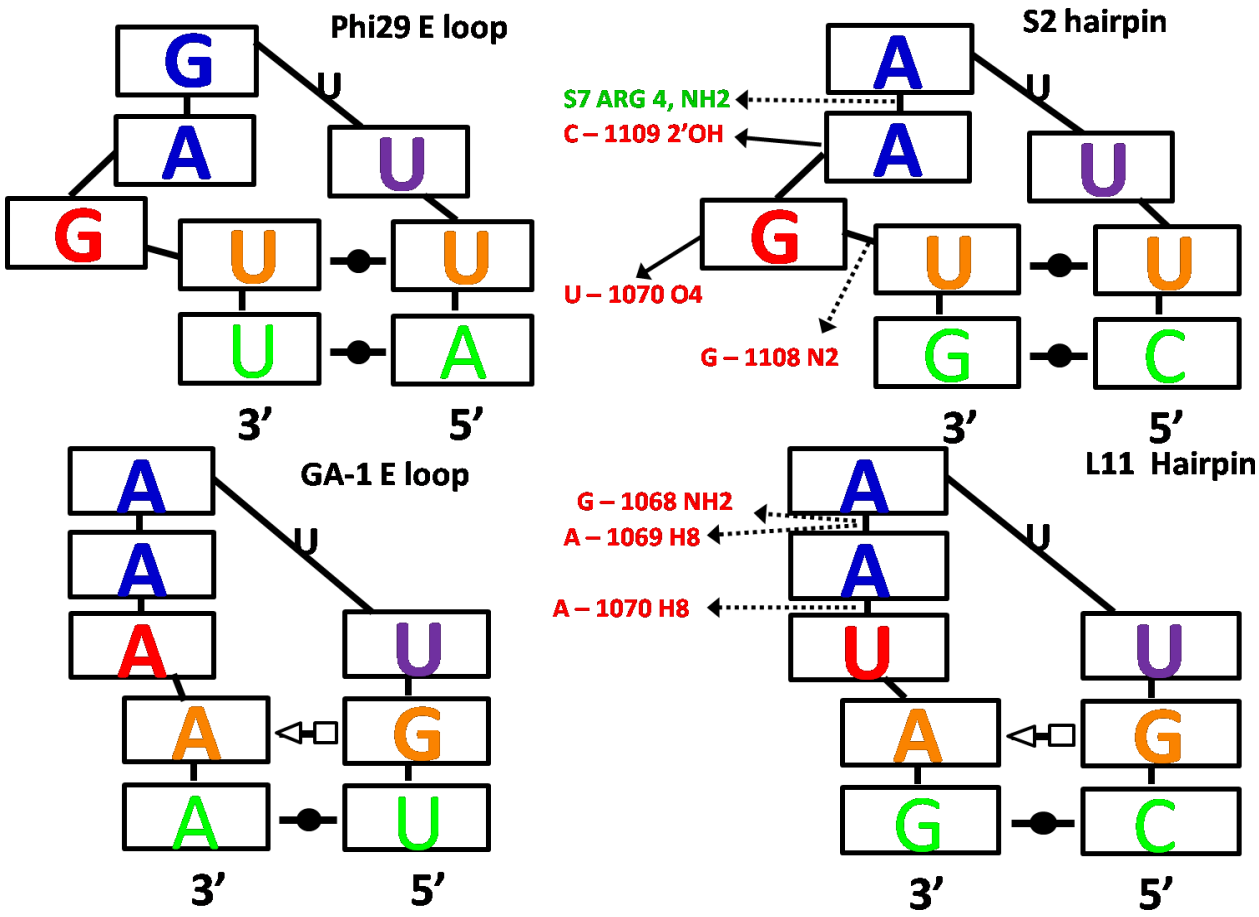



**Figure 2.14:** Diagram of base stacking in prohead and ribosomal RNA hairpins. Bases are represented by boxes, and the colors of the bases correspond to nucleotides in the same sequence position as the phi29 E-loop. Lines represent the ribose and phosphodiester backbone between bases. The U-turn location is indicated by a U. Arrows indicated hydrogen bonding in protein or RNA tertiary interactions in the crystal structures of ribosomal hairpins (*92, 124-129, 162-163*). The structure of the GA-1 E-loop hairpin was predicted by MC-SYM.

**Figure 2.15: Hairpin Backbone Overlays**

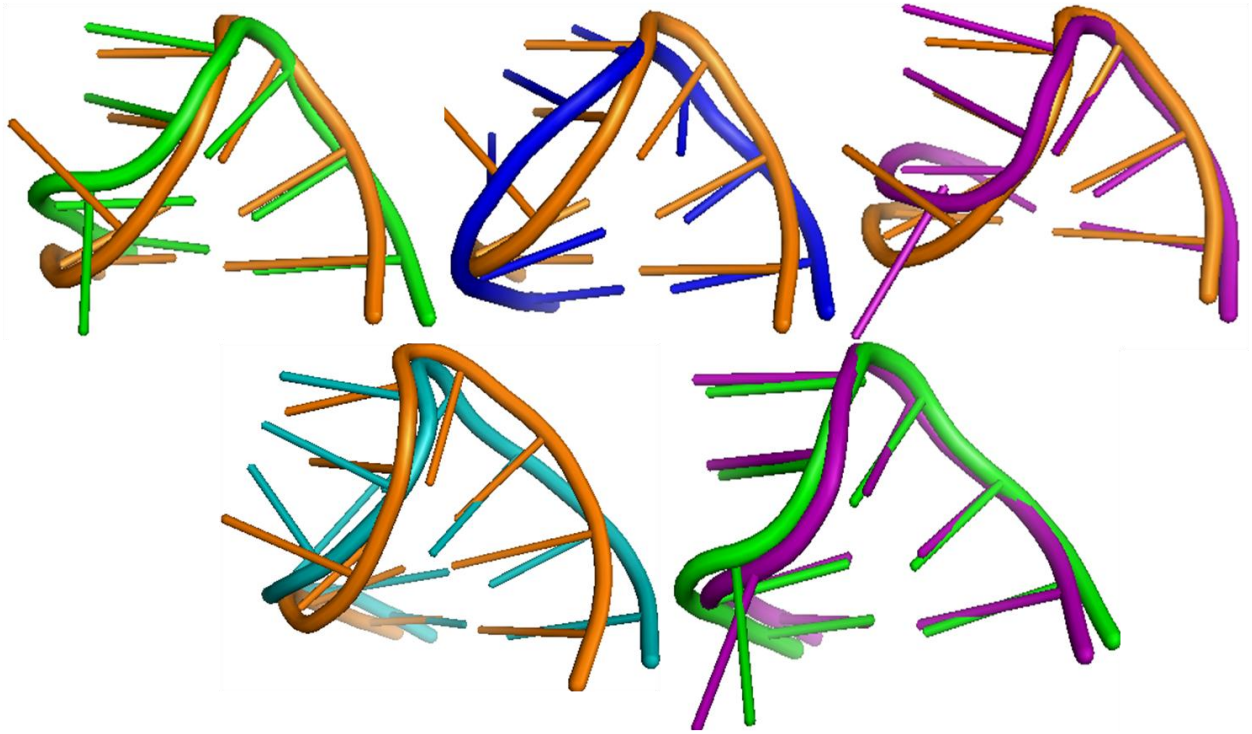


**Figure 2.15:** Comparison of the backbones of the different hairpin structures. The orange hairpin is the phi29 NMR structure. The green hairpin is the S2 NMR structure. The blue hairpin is the phi29 MC-SYM structure predicted with no NOEs. The purple hairpin is the phi29 MC-SYM structure predicted with the 2 NOEs that were violated in the predicted structure, A9H2-U11H4' and U7H3-U11H3. The teal hairpin is the L11 hairpin. Note that the position of G10 in the phi29 hairpin and its equivalent purine nucleotide is less well-defined by NMR data than other nucleotides; and thus the differences in the average position of this dynamic nucleotide that is projecting into solution reflect a lack of NMR restraints rather than a significant structural difference.

### 2.4.4 Comparison of Hairpin Structures Determined by NMR to Predictions from Isostericity Matrices.

Because the phi29 and GA1 pRNAs have a conserved secondary structure and function, one hypothesis may be that the tertiary structure may be conserved as well. Predicting base pair structures with isostericity matrices is based on the assumption that RNAs with similar functions will retain a similar structure. The comparisons of the experimental and predicted structures for ribosomal and prohead hairpins, however, suggest different structures for the phi29 and GA1 E-loop hairpins. The only known isosteric UU and GA pair forms hydrogen bonds only between the ribose edges and leaves the Watson-Crick edges solvent exposed, which is inconsistent with the NOEs between imino protons in U6 and U11 in the phi29 E-loop. There are no known isosteric pairs that conserve the orientation and C1' to C1' distance for UU and GA pairs with hydrogen bonding at the Watson-Crick edges. The UU pair in the phi29 E-loop is *cis*-Watson-Crick/Watson-Crick (2-carbonyl-N3, 4-carbonyl-N3) and the GA pair in the predicted GA1 E-loop is *trans*-Sugar/Hoogsteen (N7-amino, amino-N3 sheared) (*95, 123*). However, the experimentally determined phi29 UU pair has a longer than expected C1'-C1' distance of 10 Å, rather than the 8.1 Å C1'-C1' distance expected from isostericity matrices. The predicted GA1 GA pair has a C1'-C1' distance of 9.0 Å , which is similar to the 9.7 Å distance expected from isostericity matrices. Thus, both pRNA E-loop hairpins have C1'-C1' distances within 1 Å for the first pair. The significant difference occurs in the cis or trans backbone conformation of the first pair in the hairpin loop. This difference may

cause the G10 nucleotide to flip out in a syn conformation in the phi29 hairpin and the adenine to stack under the other adenines in the GA1 predicted structure. Although the overall backbone structure may not be conserved between the phi29 and GA1 E-loops, the U-turn and the uridine stacking on the first mismatch will likely be conserved. Differences between the gp10 proteins and surrounding RNA tertiary interactions may accommodate the different shapes of the phi29 and GA1 hairpins.

## 2.5 Conclusions

This research presents the NMR structure of the phi29 E-loop hairpin and compares the experimentally determined structure to predictions from the MC-SYM program and isostericity matrices. The comparisons between phi29 and GA1 and ribosomal hairpins provide benchmarks for testing RNA structure prediction methods. All the hairpins contain a well-predicted U-turn motif. The phi29 hairpin is more similar in structure and sequence to the ribosomal S2 hairpin than to predicted structures for the GA1 hairpin, which is similar in sequence and structure to the ribosomal L11 hairpin. Thus these two pRNA hairpins are not predicted to be isosteric despite sharing the same functional role in the same type of RNA. The first difference appears to be whether the first pair is a UU pair with cis backbone orientations or a GA pair with trans backbone orientations. The second difference is whether the G10 nucleotide in phi29 or its equivalent purine projects out into solution or stacks under the 5' purines. The differences between the phi29 NMR structure and the MC-SYM predictions

suggest improvements in estimating the stabilities of non-A-form backbone conformations and syn conformations is necessary for future improvements in RNA three-dimensional structure predictions.

ACKNOWLEDGEMENTS

The authors thank Scott Kennedy for $^{13}C$ and $^{31}P$ NMR data collection. The authors thank Marc Parisien for helpful instructions and discussions on MC-SYM.

## REFERENCES


1. Gestland, R., Cech, T., Atkins, J. (2006) *The RNA World*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
2. Sevignani, C., Calin, G., Siracusa, L., Croce, C. (2006) Mamalian Micro-RNA : A Small World for Fine Tuning Gene Expression, *Journal oF Mammalian Genome 17*, 189-202.
3. Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution, *Cell 123*, 1133-1146.
4. Seiwert, S. D., and Stuart, K. (1994) RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro, *Science 266*, 114.
5. Winkler, W. C. (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control, *Current opinion in chemical biology 9*, 594-602.
6. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena, *Cell 31*, 147-157.
7. Altman, S. (1990) Enzymatic cleavage of RNA by RNA, *Bioscience Reports 10*, 317-337.
8. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics, *Nature 407*, 340-348.
9. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation, *Cell 108*, 557-572.
10. Selmer, M., Dunham, C. M., Murphy IV, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA, *Science 313*, 1935.
11. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature 407*, 327-339.
12. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, *Science 289*, 905.
13. Hansen, J. L., Ippolito, J. A., Ban, N., Nissen, P., Moore, P. B., and Steitz, T. A. (2002) The structures of four macrolide antibiotics bound to the large ribosomal subunit, *Molecular cell 10*, 117-128.
14. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science 289*, 920.
15. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif, *Proceedings of the National Academy of Sciences of the United States of America 98*, 4899.
16. Steitz, J. A., and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3'terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli, *Proceedings of the National Academy of Sciences of the United States of America 72*, 4734.
17. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium, *Cell 107*, 679-688.
18. Yonath, A., Leonard, K., and Wittmann, H. (1987) A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction, *Science 236*, 813.

19. Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Glühmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., and Jacobi, C. (2001) Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3, *The EMBO Journal 20*, 1829-1839.
20. Bashan, A., Agmon, I., Zarivach, R., Schluenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., and Hansen, H. A. S. (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression, *Molecular cell 11*, 91-102.
21. Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism, *Science 288*, 640.
22. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution, *Science 292*, 1863.
23. Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H., and Kornberg, R. D. (1991) A mediator required for activation of RNA polymerase II transcription in vitro.
24. Lorch, Y., LaPointe, J. W., and Kornberg, R. D. (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones, *Cell 49*, 203-210.
25. Blackburn, E. H. (1991) Structure and function of telomeres, *Nature 350*, 569-573.
26. Blackburn, E. H. (2001) Switching and signaling at the telomere, *Cell 106*, 661-673.
27. Blackburn, E. H. (2000) Telomere states and cell fates, *Nature 408*, 53-56.
28. Epel, E. S., Blackburn, E. H., Lin, J., Dhabhar, F. S., Adler, N. E., Morrow, J. D., and Cawthon, R. M. (2004) Accelerated telomere shortening in response to life stress, *Proceedings of the National Academy of Sciences of the United States of America 101*, 17312.
29. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase RNA, *Cell 91*, 25-34.
30. Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., and Bacchetti, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity, *The EMBO Journal 11*, 1921.
31. Harley, C. B., Futcher, A. B., and Greider, C. W. (1990) Telomeres shorten during ageing of human fibroblasts.
32. Hemann, M. T., Strong, M. A., Hao, L. Y., and Greider, C. W. (2001) The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability, *Cell 107*, 67-77.
33. Blackburn, E., and Szostak, J. (1984) The molecular structure of centromeres and telomeres, *Annual review of biochemistry 53*, 163-194.
34. Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging, *Nature medicine 12*, 1133-1138.
35. Dunn, B., Szauter, P., Pardue, M. L., and Szostak, J. W. (1984) Transfer of yeast telomeres to linear plasmids by recombination, *Cell 39*, 191-201.
36. Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature 391*, 806-811.

37. Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing, *Cell 106*, 23-34.
38. Montgomery, M. K., Xu, S. Q., and Fire, A. (1998) RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans, *Proceedings of the National Academy of Sciences of the United States of America 95*, 15502.
39. Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C. C. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans, *Cell 99*, 123-132.
40. Timmons, L., and Fire, A. (1994) Specific interference by ingested dsRNA, *Science 263*, 802-805.
41. Mello, C. C., and Conte, D. (2004) Revealing the world of RNA interference, *Nature 431*, 338-342.
42. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure1, *Journal of Molecular Biology 288*, 911-940.
43. Gultyaev, A. P., Van Batenburg, F., and Pleij, C. W. A. (1995) The computer simulation of RNA folding pathways using a genetic algorithm, *Journal of Molecular Biology 250*, 37-51.
44. Borer, P. N., Dengler, B., and Tinoco, I. (1974) Stability of ribonucleic acid double-stranded helices, *Journal of Molecular Biology 86*, 843-853.
45. Ding, Y., Chan, C. Y., and Lawrence, C. E. (2004) Sfold web server for statistical folding and rational design of nucleic acids, *Nucleic Acids Research 32*, W135.
46. Do, C. B., Woods, D. A., and Batzoglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models, *Bioinformatics 22*, e90.
47. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures, *Monatshefte für Chemie/Chemical Monthly 125*, 167-188.
48. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) RNA structure prediction, *Annual Review of Biophysics and Biophysical Chemistry 17*, 167-192.
49. Clanton-Arrowood, K., McGurk, J., and Schroeder, S. J. (2008) 3 Terminal Nucleotides Determine Thermodynamic Stabilities of Mismatches at the Ends of RNA Helices†, *Biochemistry 47*, 13418-13427.
50. Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Free energy contributions of G-U and other terminal mismatches to helix stability, *Biochemistry 25*, 3209-3213.
51. Blanton S. Tolbert, S. D. K., Susan J. Schroeder, Thomas R. Krugh, Douglas H. Turner. (2006) NMR Structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: Probing the Structural Features That Shape the Thermodynamic Stability of GA Pairs, *Biochemistry*.
52. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) Nuclear Magnetic Resonance Spectroscopy and Molecular Modeling Reveal That Different Hydrogen Bonding Patterns Are Possible for GU Pairs: One Hydrogen Bond for Each GU Pair in r (GGCGUGCC) 2 and Two for Each GU Pair in r (GAGUGCUC) 2†,‡, *Biochemistry 39*, 8970-8982.
53. Schroeder, S. J., and Turner, D. H. (2001) Thermodynamic Stabilities of Internal Loops with GU Closing Pairs in RNA, *Biochemistry 40*, 11509-11517.

54. Banerjee, A. R., Jaeger, J. A., and Turner, D. H. (1993) Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure, *Biochemistry 32*, 153-163.
55. Tinoco, I. (1999) How RNA folds, *Journal of Molecular Biology 293*, 271-281.
56. David H. Mathews, M. D. D., Jessica L. Childs, Susan J. Schroeder, Michael Zuker, Douglas H. Turner. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. 101*, 7287–7292.
57. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) Probing the structure of RNAs in solution, *Nucleic Acids Research 15*, 9109.
58. Das, R., and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures, *Proceedings of the National Academy of Sciences 104*, 14664.
59. Marc Parisien, F. M. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature 452*, 51-55.
60. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Research 31*, 3406.
61. Hofacker, I. L. (2003) Vienna RNA secondary structure server, *Nucleic Acids Research 31*, 3429.
62. Leontis, N. B., Stombaugh, Jesse., Westhof, Eric. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Research 30*, 3497-3531.
63. Xia, T., SantaLucia Jr, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson- Crick Base Pairs†, *Biochemistry 37*, 14719-14735.
64. Jacobsen, N. E. (2007) *NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology.*, Wiley-Interscience, Hoboken, N.J.
65. (1993) *NMR of macromolecules: a practical approach*, Oxford University Press, Oxford, NY.
66. Terence N. Mitchell, B. C. (2004) *NMR-From Spectra to Structures*, Springer, Dortmund, Germany.
67. Axel T. Brüngera, P. D. A., G. Marius Clorec, Warren L. DeLanod, Piet Grose, Ralf W. Grosse-Kunstlevea,b, Jian-Sheng Jiangf, John Kuszewskic, Michael Nilgesg, Navraj S. Pannuh, Randy J. Readi, Luke M. Riceb, Thomas Simonsonj, Gregory L. Warrenb. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination, *Acta Cryst. D54*.
68. Wilfried J. J. Meijer, J. A. H., and Margarita Salas. (2001) phi 29 Family of Phages, *American Society for Microbiology 65*, 261-187.
69. Morais, M. C., Koti, J. S., Bowman, V. D., Reyes-Aldrete, E., Anderson, D. L., and Rossmann, M. G. (2008) Defining Molecular and Domain Boundaries in the Bacteriophage 29 DNA Packaging Motor, *Structure 16*, 1267-1274.
70. Simpson, A. A. T., Yizhi; Leiman, Petr G.; Badasso, Mohammed O.; He, Yongning; Jardine, Paul J.; Olson, Norman H.; Morais, Marc C.; Grimes, Shelley; Anderson, Dwight L.; Baker, Timothy S.; Rossmann, Michael G. (2000) Structure of the bacteriophage ϕ29 DNA packaging motor, *Nature 408*, 745-750.
71. Zhang, F., Lemieux, S., Wu, X., St.-Arnaud, D. McMurray, T., Major, F., Anderson, D. (1998) Function of Hexameric RNA in Packaging of Bacteriophage phi29 DNA in vitro, *Mol. Cell 2*, 141-147.

72. Guo, P. X., Erickson,S. and Anderson,D. (1987) A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA., *Science 236*, 690-694.
73. Atz R, M. S., Gao J, Anderson DL, Grimes S. (2007) Alanine scanning and Fe-BABE probing of the bacteriophage ø29 prohead RNA-connector interaction., *Journal of Molecular Biology 369*, 239-248.
74. PEIXUAN Guo, S. G., AND DWIGHT ANDERSON. (1986) A defined system for in vitro packaging of DNA-gp3 of the Bacillus subtilis bacteriophage Phi29, *Proc. Natl. Acad. Sci. 83*, 3505-3509.
75. Shu, D., Huang, LP., Hoeprich, S., Guo, P. (2005a) Construction of DNA-packaging phi29 RNA Monomers, Dimers, and Trimers with Variable Sizes and Shapes as Potential Parts for Nanodevices., *Journal of Nanoscience and Nanotechnology 3*, 295-302.
76. Grimes S, J. P., Anderson D. (2002) Bacteriophage phi 29 DNA packaging., *Advances in Virus Research 58*, 255-294.
77. Guo, P., Erickson, S., and Anderson, D. (1987) A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA, *Science 236*, 690.
78. M Trottier, C. Z., and P Guo. (1996) Complete inhibition of virion assembly in vivo with mutant procapsid RNA essential for phage phi 29 DNA packaging, *Journal of Virology 70*, 55-61.
79. Wichitwechkarn, J., Bailey,S., Bodley,J.W. and Anderson,D. (1989) Prohead RNA of bacteriophage phi 29: size, stoichiometry and biological activity., *Nucleic Acids Research 17*, 3459-3468.
80. Bailey, S., Wichitwechkarn, J., Johnson, D., Reilly, B. E., Anderson, D. L., Bodley, J. W. (1990) Phylogenetic Analysis and Secondary Structure of Bacillus subtilis Bacteriophage RNA Required for DNA Packaging., *Journal of Biological Chemistry 265*, 22365-22370.
81. Morais, M., Koti, J., Bowman, V., Reyes-Aldrete, E., Anderson, D.L., Rossman, M.G. (2008) Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor, *Structure 16*, 1267-1274.
82. Reid, R. J. D., Bodley, J.W., Anderson, D. (1994) Characterization of the Prohead -pRNA Interaction of Bacteriophage phi29, *Journal of Biological Chemistry 269*, 5157-5162.
83. Xiao, F., Moll, W.-D., Guo, S., Guo, P. (2005) Binding of pRNA to the N-terminal 14 amino acids of connector protein to bacteriophage phi29, *Nucl. Acids Res. 33*, 2640-2649.
84. Garver, K., Guo, P. (1997) Boundary of pRNA functional domains and minimum pRNA sequence requirement for specific connector binding and DNA packaging of phage phi29, *RNA 3*, 1068-1079.
85. Guasch, A., Pous, J., Ibarra, B., Gomis-Ruth, F.X., Valpuesta, J.M., Sousa, N., Carrascosa, J.L., Coll, M. (2002) Detailed Architecture of a DNA Translocating Machine: The High-resolution Structure of the Bacteriophage phi29 Connector Particle, *J. Mol. Biol. 315*, 663-676.
86. Reid, R., J.D., Bodley, J., W., Anderson, D. (1994b) Identification of Bacteriophage phi29 Prohead RNA Domains Necessary for in vitro DNA-gp3 Packaging, *J. Biol. Chem. 269*, 9084-9089.
87. Kitamura, A., Jardine, P.J., Anderson, D.L., Grimes, S., Matsuo, H. (2008) Analysis of intermolecular base pair formation of prohead RNA of the phage pi29 DNA packaging motor using NMR spectroscopy, *Nucl. Acids Res. 36*, 839-848.
88. Trottier, M., Mat-Arip, Y., Zhang, C., Chen, C., Sheng, S., Shao, Z., Guo, P. (2000) Probing the structure of monomers and dimers of the bacterial virus phi29 hexamer RNA complex by chemical modification, *RNA 6*, 1257-1266.

89. Williamson, J. R. (2000) Induced fit in RNA–protein recognition, *Nature Structural & Molecular Biology 7*, 834-837.
90. Draper, D. E. (1995) Protein-RNA recognition, *Annual review of biochemistry 64*, 593-620.
91. M. A. Fountain, M. J. S., T. R. Krugh, and D. H. Turner. (1996) Structural Features of a Six-Nucleotide RNA Hairpin Loop Found in Ribosomal RNA, *Biochemistry 35*, 6539-6548.
92. Zhang, H., Fountain, M., Krugh, T. (2001) Structural Characterization of a Six-Nucleotide RNA Hairpin Loop Found in *Escherichia coli*, r(UUAAGU), *Biochemistry 40*, 9879-9886.
93. Uhlenbeck, O. C. (1990) Tetraloops and RNA folding, *Nature 346*, 613.
94. Shu, D., Zhang, H., Petrenko, R., Meller, J., and Guo, P. (2010) Dual-Channel Single-Molecule Fluorescence Resonance Energy Transfer to Establish Distance Parameters for RNA Nanoparticles, *ACS nano*, 1676-1683.
95. Victor A. Bloomfield., D. M. C., Ignacio Tinoco (2000) *Nucleic Acids Structures, Properties and Functions*.
96. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability, *Proceedings of the National Academy of Sciences of the United States of America 83*, 9373.
97. Petersheim, M., and Turner, D. H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp, *Biochemistry 22*, 256-263.
98. He, L., Walter, E., and Turner, H. (1991) Nearest-Neighbor Parameters for GU Mismatches: 3tUG5'Is Destabilizing in the Contexts% if, YE: $, and $ Egy but Stabilizing in, *Biochemistry 30*, 46.
99. Turner, J. A. M. a. D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC)2 by Two-Dimensional NMR and Simulated Annealing, *Biochemistry 35*, 14077-14089.
100. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helixes, *Biochemistry 25*, 5755-5759.
101. Nguyen, M. T., and Schroeder, S. J. (2010) Consecutive Terminal GU Pairs Stabilize RNA Helices, *Biochemistry*.
102. McDowell, J. A., and Turner, D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC) 2 by Two-Dimensional NMR and Simulated Annealing†,‡, *Biochemistry 35*, 14077-14089.
103. Nagaswamy, U., Voss, N., Zhang, Z., and Fox, G. E. (2000) Database of non-canonical base pairs found in known RNA structures, *Nucleic Acids Research 28*, 375.
104. Cech, T. R. (1990) Self-splicing of group I introns, *Annual review of biochemistry 59*, 543-568.
105. Garriga, G., Lambowitz, A. M., Inoue, T., and Cech, T. R. (1986) Mechanism of recognition of the 5 splice site in self-splicing group I introns.
106. Haley, B., and Zamore, P. D. (2004) Kinetic analysis of the RNAi enzyme complex, *Nature Structural & Molecular Biology 11*, 599-606.
107. Doench, J. G., and Sharp, P. A. (2004) Specificity of microRNA target selection in translational repression, *Genes & development 18*, 504.
108. Ameres, S. L., Martinez, J., and Schroeder, R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC, *Cell 130*, 101-112.

109. Mayr, C., Hemann, M. T., and Bartel, D. P. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation, *Science 315*, 1576.
110. Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. J. (2005) RAS is regulated by the let-7 microRNA family, *Cell 120*, 635-647.
111. Lee, Y. S., and Dutta, A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene, *Genes & development 21*, 1025.
112. Shell, S., Park, S. M., Radjabi, A. R., Schickel, R., Kistner, E. O., Jewell, D. A., Feig, C., Lengyel, E., and Peter, M. E. (2007) Let-7 expression defines two differentiation stages of cancer, *Proceedings of the National Academy of Sciences 104*, 11400.
113. Ragan, C., Cloonan, N., Grimmond, S. M., Zuker, M., and Ragan, M. A. (2009) Transcriptome-wide prediction of miRNA targets in human and mouse using FASTH, *PloS one 4*, e5745.
114. Chen, X., McDowell, J., Kierzek, R., Krugh, T., and Turner, D. (2000) Nuclear Magnetic Resonance Spectroscopy and Molecular Modeling Reveal That Different Hydrogen Bonding Patterns Are Possible for GU Pairs: One Hydrogen Bond for Each GU Pair in r (GGCGUGCC) 2 and Two for Each GU Pair in r (GAGUGCUC) 2†,‡, *Biochemistry 39*, 8970-8982.
115. McDowell, J., and Turner, D. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC) 2 by Two-Dimensional NMR and Simulated Annealing†,‡, *Biochemistry 35*, 14077-14089.
116. Strobel, S. A., and Cech, T. R. (1995) Minor groove recognition of the conserved GU pair at the Tetrahymena ribozyme reaction site, *Science 267*, 675.
117. Limmera, S., and Reifa, B. (1996) NMR evidence for helix geometry modifications by a GU wobble base pair in the acceptor arm of E. coli tRNAAla, *FEBS Letters 385*, 15-20.
118. Moffitt, J. R., Chemla. Y.R., Aathavean. K., Grimes, S., Jardine, P.J., Anderson, D.L., Bustamante, C. (2009) Intersubunit coordination in a homomeric ring ATPase, *Nature 457*, 446-451.
119. Lee, T. J., Zhang, H., Chang, C.L., Savran, C., Guo, P. (2009) Engineering of the fluorescent-energy conversion arn of phi29 DNA packaging motor for single molecule studies, *Small 5*, 2453-2459.
120. Peixuan Guo, S. B., James W. Bodley, Dwight Anderson (1987) Characterization of the small RNA of the bacteriophage ø29 DNA packaging machine, *Nucleic Acids Research 15*, 7081-7090.
121. Zhang, F. a. A., D. (1998) In vitro selection of bacteriophage phi29 prohead RNA aptamers for prohead binding., *J. Biol. Chem. 273*, 2947–2953.
122. Chen, C., Zhang, C., Guo, P. (1999) Sequence requirements for hand-in-hand interaction in formation of RNA dimers and hexamers to gear phi29 DNA translocation motor, *RNA 5*, 805-818.
123. Stombaugh, J., Zirbel, C.L., Leontis, N.B. (2009) Frequency and Isostericity of RNA Base Pairs, *Nucl. Acids Res. 37*, 2294-2312.
124. Graeme L. Conn, D. E. D., Eaton E. Lattman, Apostolos G. Gittis. (1999) Crystal Structure of a Conserved Ribosomal Protein-RNA Complex, *Science 284*, 1171.
125. Fountain, M. A., Serra, M.J., Krugh, T.R., Turner, D.H. (1996) Structural Features of a Six-Nucleotide RNA Hairpin Loop found in Ribosomal RNA, *Biochemistry 35*, 6539-6548.
126. Shengrong Huang, Y.-X. W. a. D. E. D. (1996) Structure of a Hexanucleotide RNA Hairpin Loop Conserved in Ribosomal RNAs, *Journal of Molecular Biology 258*, 308-321.

127. Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A. (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4A, *Science 289*, 905-919.
128. Wimberly, B. T., Broderson, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature 407*, 327-339.
129. Selmer, M., Dunham, C.M., Murphy, F.V., Weizlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., Ramakrishnan, V. (2006) Structure of the 70S Ribosome Complexed with mRNA and tRNA, *Science 313*, 1935.
130. Morais, M. C., Tao, Y., Olson, N.H., Grimes, S., Jardine, P., Anderson, D.L., Baker, T.S., Rossman, M.G. (2001) Cryoelectron-Microscopy Image Reconstruction of Symmetry Mismatches in Bacteriophage phi29, *Journal of Structural Biology 135*, 38-46.
131. Morais, M. C., Kanamaru, S., Badasso, M.O., Kotis, S.J., Owen, B.A., McMurray, C.T., Anderson, D.L., Rossman, M.G. (2003) Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly, *Nature Structural and Molecular Biology 10*, 572-576.
132. Simpson, A. A., Tao, Y., Leiman, P.G., Badasso, M.O., He, Y., Jardine P.J., Olson, N.H., Morais, M.C., Grimes, S., Anderson, D.L., Baker, T.L., Rossman, M.G. (2000) Structure of the Bacteriophage phi29 DNA Packaging Motor, *Nature 408*, 745-750.
133. Morais, M. C., Choi, K.H., Kotis, J.S., Chipman, P.R., Anderson, D.L., Rossman, M.G. (2005) Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29, *Molecular Cell 18*, 149-159.
134. Gausch, A., Pous, J., Ibarra, B., Gomis-Ruth, F.X., Valpuesta, J.M., Sousa, N., Carrascosa, J.L., Coll, M. (2002) Detailed Architecture of a DNA Translocating Machine: The High Resolution Crystal Structure of the Bacteriophage Phi29 Connector Particle, *Journal of Molecular Biology 315*, 663-676.
135. Xiang, Y., Leiman, P.G., Li, L., Grimes, S., Anderson, D.L., Rossman, M.G. (2009) Crystallographic insights into the autocatalytic assembly mechansim of a bacteriophage tail spike, *Mol. Cell 34*, 375-386.
136. Xiang, M., M.C., Cohen, D.N., Bowman, V.D., Anderson, D.L., Rossman M.G. (2008) Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage phi29 tail, *Proc. Natl. Acad. Sci. USA 105*, 9552-9557.
137. Zhang, F., Lemieux, S., Wu, X., St.-Arnaud, D., McMurray, C., Major, F., Anderson, D. (1998a) Function of Hexameric RNA in Packaging of Bacteriophage phi29 DNA In Vitro, *Molecular Cell 2*, 141-147.
138. Hoeprich, S., Guo, P. (2002) Computer Modeling of Three-dimensional Structure of DNA-packaging RNA (pRNA) Monomer, Dimer, and Hexamer of Phi29 DNA Packaging Motor, *J. of Biol. Chem. 277*, 20794-20803.
139. Jane S. Richardson, B. S., Laura W. Murray, et al. (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution), *RNA 14*, 465-481.
140. Kneller, T. D. G. a. D. G. SPARKY 3.
141. Brunger, A. T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kuntsleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998) Crystallography and NMR System: A new software suite for macromolecular structure determination, *Acta Cryst. D54*, 905-921.
142. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, *DeLano Scientific*.
143. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule, *Science 244*, 48-52.

144. Xia, T., Santa Lucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., Turner, D.H. (1998) Thermodynamic Parameters for an Expanded Nearest-neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs, *Biochemistry 37*, 14719-14735.
145. Mathews, D. H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. USA 101*, 7287-7292.
146. JUNMEI WANG, P. C., PETER A. KOLLMAN. (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?, *Journal of Computational Chemistry 21*, 1049-1074.
147. JUNMEI WANG, R. M. W., JAMES W. CALDWELL, PETER A. KOLLMAN, DAVID A. CASE. (2004) Development and Testing of a General Amber Force Field, *Journal of Computational Chemistry 25*, 1157-1174.
148. Scott J. Weiner, P. A. K., Dzung T. Nguyen, David A. Case,. (1986) An All Atom Force Field for Simulations of Proteins and Nucleic Acids, *Journal of Computational Chemistry 7*, 230-252.
149. Williams, D. J., Hall, K.B. (1996) Thermodynamic Comparison of the Salt Dependence of Natural RNA Hairpins and RNA Hairpins with Non-Nucleotide Spacers, *Biochemistry 35*, 14665-14670.
150. Vieregg, J., Cheng, W., Bustamante, C., Tinoco, I., Jr. (2007) Measurement of the effect of monovalent cations on RNA hairpin stability, *J. Am. Chem. Soc. 129*, 14966-14973.
151. Lukavsky, P. (2007) Basic Principles of RNA NMR Spectroscopy, In *Structure and Biophysics – New Technologies for Current Challenges in Biology and Beyond*, Springer Netherlands.
152. Schroeder, S. J., Turner, D.H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, *Biochemistry 39*, 9257-9274.
153. Santa Lucia, J., Jr., Kierzek, R., 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.
154. Wuthrich, K. ( 1986) *NMR of Proteins and Nucleic Acids*, New York.
155. Robin R. Gutell, J. J. C., Danielle Konings, Daniel Gautheret. (2000) Predicting U-turns in Ribosomal RNA with Comparative Sequence Analysis, *journal of Molecular Biology 300*, 791-803.
156. Fiona M Jucker, A. P. (1995) GNRA tetraloops make a U-turn, *RNA 1*, 219-222.
157. Ashrah, S. S., Ansari, G., Guenther, R., Sochacka, E., Maliewicz, A., Agris, P.F. (1999) The uridine in "U-turn": Contributions to tRNA-ribosomal binding, *RNA 5*, 503-511.
158. G. J. Quigley, A. R. (1976) Structural domains of transfer RNA molecules, *Science 194*, 796-806.
159. Yildirim, I., Turner, D.H. (2005) RNA CHallenges for Computational Chemists, *Biochemistry 44*, 13225-13234.
160. Yildirim, I., Stern, H.A., Sponer, J., Spackova, N., Turner, D.H. (2009) Effects of Restrained Sampling Space and Nonplanar Amino Groups on free-Energy Predictions for NRA with IMino and Sheared Tandem GA Base Pairs Flanked by a GC, CG, iGiC, or iCiG Base Pairs, *J. Chem. Theory Comput. 5*, 2088-2100.
161. Zhang, H., Fountain, M., Krugh, T. (2001) Structural characterization of a six-nucleotide RNA hairpin found in Eschericia coli, r(UUAAGU), *Biochemistry 40*, 9879-9886.

162. Gao, Y.-G., Selmer, M., Dunham, C.M., Weixlbaumer, A., Kelley, A.C., Ramakrishnan, V. (2009) the structure fo the Ribosome with Elongation FAcor G Trapped in the Posttranslocational State, *Science 326*, 694-699.
163. Blaha, G., Stanley, R.E., Steitz, T.A. (2009) Formation of the First Peptide Bond: the Structure of EF-P Bound to the 70S Ribosome, *Science 325*, 966-970.

**CHAPTER 3**

# Tertiary Structure Determination of Consecutive Terminal G-U pairs: 5'r(GUGUCGGUGU)$_2$

## 3.1 Introduction.

The "Wobble Hypothesis" proposed by Crick in 1966 suggested the existence of the G-U base pair with hopes that it would help explain degeneracy in the genetic code (*1*). Since then, numerous cases have been discovered in which the G-U base pair has been shown to have structural as well as functional roles. Notable cases include the group I intron of *Tetrahymena Thermophila (2-4)* and tRNA recognition (*5-6*). G-U pairs have also been found to be conserved ribosomal RNA of numerous organisms (*7*). In the case of the group I intron, there is a conserved mismatch that places the splice site in the correct orientation so that it can contact the intron's catalytic core, and the G-U pair also stabilizes the ribozyme's transition state (*3*). The *Escherichia coli* alanine tRNA utilizes the G-U pair as the recognition site for acceptor stem identification by aminoacyl tRNA synthetase. A single G-U wobble pair has been shown to suffice as the recognition factor of the alanine tRNA by its associated synthetase (*6*).

The G-U pair is the most common base pair besides the Watson-Crick canonical base pairs and makes up almost 50% of non-canonical base pairs and mismatches (*8-11*). Statistical studies and sequence alignments have shown that G-U mismatches are often conserved (2, 5, 7). These data suggest that they may play a specific structural or functional role and are not simply mismatched

replacements of Watson-Crick base pairs. Regardless of their frequent occurrences, there have been very few studies on the three-dimensional structures of RNA containing consecutive terminal G-U base pairs. G-U base pairs are well studied and they have been found to be unfavorable in the middle of a helix(*10, 12-13*) but favorable on the ends of helices (*14-15*). Recent thermodynamic studies have shown that consecutive terminal G-U motifs add stability to the duplexes and are more stable than previously predicted (*14*). The errors in thermodynamic data as well as in structure predictions show that there is a drastic need to improve the prediction parameters as well as solve tertiary structures containing these motifs. Knowing the fundamentals of three-dimensional structures for G-U base pairs would significantly aid in understanding the molecular basis of RNA stability and function. Furthermore, knowledge of the tertiary structures of G-U mismatches is therefore important for improving RNA structure prediction.

**Figure 3.1: The Two Most Common G-U Base Pairs**

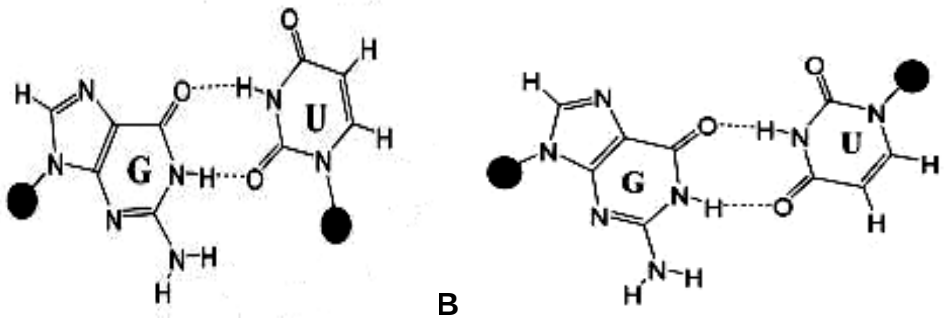



**Figure 3.1:** The two most common G-U base pair orientations. The wobble (A), and reverse wobble (B), make two hydrogen bonds involving the imino protons on the guanosine and uracil bases. Figures adopted from figure 2 and 3 Nucleic Acids Res. 2002 August 15; 30(16): 3497–3531.

In this study, the tertiary structure of the duplex 5'r(**GUGU**CG**GUGU**)$_2$ has been determined to identify interactions such as base stacking, hydrogen bonding, base orientation and other unknown interactions that may contribute to the stability of the terminal G-U base pairs. This tertiary structure has been contrasted to internal single and double G-U base pairs in an effort to pinpoint the similarities and differences between internal and terminal G-U base pairs.

Thermodynamic as well as UV studies have been previously performed for this particular duplex (*14*). Those studies show that the previous nearest neighbor parameters for calculating the free energy did not hold true for consecutive G-U base pairs at the ends of helices. Those studies also showed that the predicted free energy of this duplex was miscalculated by 4.0 kcal/mol and the stability that was gained due to the consecutive terminal G-U pairs was

3.1 kcal/mol. Preliminary 1-dimensional imino proton spectra show good peak dispersion and very little overlap which was suitable for tertiary structure determination by NMR.

## 3.2 Materials and Methods.

### 3.2.1 RNA Preparation and Purification.

The RNA was synthesized and gel purified by Dharmacon, Inc. Protecting groups were removed using the supplied deprotecting buffer, a 1% acetic acid solution at pH 3.8. The RNA oligonucleotide was lyophilized, resuspended in 10 mM $NaPO_4$, 50 mM NaCl, and 0.5 mM $Na_2EDTA$ pH 6 NMR buffer, dialyzed in 25 mM NaCl, 5 mM $Na_3PO_4$, 0.25 mM $Na_2EDTA$ pH 6 buffer with a 1000 molecular weight cutoff membrane, and further purified using a Sephadex G-25 gel exclusion column. Samples were also subjected to Sep-Pak C-18 reverse-phased columns for further purification. Oligomer purity was greater than 95%, proven by $^{32}P$ labeling and gel electrophoresis. The sample concentration was 2.1 mM on the basis of high temperature UV absorbance measurement. For NMR experiments in $D_2O$, the oligomer was exchanged with 99.9% $D_2O$ twice and once with 99.996% $D_2O$.

### 3.2.3 NMR Spectroscopy.

Exchangeable proton spectra were taken at $1^o$ C using Watergate or SSnoesy pulse sequences. NOESY spectra were recorded at 50, 100, and 150 ms at $1^o$ C for exchangeable protons with a delay time of 1 s, 256 increments in

the T1 dimension with 2048 data points and 128 scans per FID. Zero filling was used to finish the T1 matrix and had a 20 ppm spectral sweep width. Apodization was done using sine bell weighting. The non-exchangeable proton spectra were recorded at $1^{o}$ C using RNA tnnoesy pulse sequence with presaturation water suppression. NOESY spectra were collected at 50, 100, 200 and 400 ms at $1^{o}$ C with a delay time of 1 s, 300 increments in the T1 dimension with 2048 data points and 30 scans per FID. The 50 ms NOESY was performed using a Varian Inova 600 MHz spectrometer and all others were performed on a Varian Inova 500 MHz spectrometer. Zero filling was used to finish the T1 matrix and the spectral sweep width was 8.5 ppm. Apodization was done using sine bell weighting. DQF-COSY experiments were performed using a Varian 600 MHz spectrometer at $1^{o}$ C with 300 increments in the T1 dimension with 2048 data points and 48 scans per FID. The spectral sweep width was set to 8.5 ppm and the delay time was 1 s. Apodization was done using gaussian functions.

The $^{13}$C HSQC data were recorded at $1^{o}$ C using a spectral sweep width of 12 ppm for the $^{1}$H and 45 ppm for the $^{13}$C dimension with a 1 s delay time. There were 32 T1 increments and 500 scans with 2048 points. The data were apodized with a sine bell shifted function in both directions.

The $^{31}$P HETCOR spectra were collected at $1^{o}$ C with a 5 ppm spectral width in both the $^{1}$H and the $^{31}$P dimension. Data was collected in 512 T1 increments, 128 scans, and 2048 data points. Apodization was done using sine bell shift weighting.

### 3.2.4 NMR Restraints and Peak Assignments.

Distance restraints from the non-exchangeable NOESY spectra were based on NOE crosspeaks in the 50, 100, and 150 ms experiments. Restraints were classified as strong, medium, or weak if the NOE occurred in the 50, 100, or 150 ms spectra, respectively. The 200 ms spectra were used to confirm NOE assignments that were weak in the 150 ms spectra but were not used directly as distance restraints. Backbone, chi and sugar pucker restraints for the central G-C base pairs from the RNA backbone consortium (*16*) were applied during standard simulated annealing protocols (*17*). All other backbone angles were allowed to sample any conformation within two times its normal standard deviation in order to explore conformational space that is similar to A-form RNA, as indicated by the results of COSY and $^{31}$P HETCOR experiments. The sugar puckers of the four terminal bases, G1, U2, G9, and U10, were modeled using three different criteria since the data show that these bases assume altering conformations by the appearance of H1'-H2' crosspeaks in the DQF-COSY. These ribose were restrained to be C2'-endo, C3'-endo as well as unrestrained to test the likelihood of each. All bases were constrained to be in the anti conformation as shown by weak H1'-H8 intranucleotide NOEs. The SPARKY assignment program was used for peak assignment, labeling, compilation of distance restraints and NOE lists (*18*). Table 3.2 contains the summary of NOE and dihedral restraints. Table 3.3 contains a list of all the NOE restraints used in the structure calculations.

### 3.2.5 Structure Calculations.

Structure calculations were performed with the Crystallography and NMR Systems (CNS) software version 1.3 (*19*) on a Dell Optiplex GX620. An unrefined, initial MCSYM structure was heated for 1.5 ps at 20000 K without any restraints to generate 30 randomized starting structures (rmsd 8 Å). The RNA-DNA-all-atom parameter and topology files from CNS contain the force charges, atomic charges, atomic masses, and atom connectivity.

In the high temperature annealing stage, the temperature was raised to 20000 K with 4000 steps for 5 ps with bond, angle, and improper restraints on and NOE and dihedral angle restraints set to 150 kcal/mol $Å^2$ and 5 kcal/mol $rad^2$, respectively. The first slow-cool annealing stage was done using torsion dynamics with the electrostatics off except between hydrogens and other atoms, which were set to 10%. The temperature was cooled in 125 K temperature steps with the NOE restraints held at 150 kcal/mol $Å^2$ and the dihedral angle restraints set to 200 kcal/mol $rad^2$. The system was allowed to develop for 90 ps using 6000 MD steps at each temperature. The second slow cooled annealing stage was done in Cartesian space and allowed the system to continue from 2000 K to 0 K. The temperature steps were set to 15 K and allowed to develop for 20 ps while the van der Waals scale factor increased linearly from 1 to 4. All other parameters were unchanged during the second annealing stage. The final energy minimization was done by applying Powell energy minimization with the van der Waals scale factor at 1 and the electrostatic terms at 100% using 30 cycles of minimization each with 2000 MD steps. The scale factor for the Powell

NOE and dihedral angles were 75 kcal/mol $Å^2$ and 400 kcal/mol $rad^2$, respectively. Pymol was used to analyze and visualize the structures (*20*).

## 3.3 Results.

### 3.3.1 H2O NOESY.

In the imino proton spectra, imino peaks were observed between 9-15 ppm for all bases except for the four terminal bases; G1, U2, G9 and U10 (Figures 3.2, 3.3). In the preliminary model, the G1 and U2 as well as the G9 and U10 are stacked but allow the imino protons to exchange with water. The imino peaks for these bases are therefore not seen. For the other bases, the imino protons resonate at 12.41, 12.17, 11.86, 11.50, and 11.13 ppm and were assigned to bases G6, U4, U8, G7 and G3, respectively. These assignments were based on chemical shifts and the pattern of NOEs expected for Watson-Crick G-C and non-canonical G-U base pairs (*21-22*). The strong NOEs between U4-G17 and G7-U14 in the duplex are consistent with the formation of either a wobble or reverse wobble G-U base pair (*23-24*). These bases are stacked on the central G-C base pairs and this is confirmed by the observation of the bases having a continuous NOESY walk through the whole duplex as well as cross-strand imino to imino proton NOEs to the N+1 base. Similar interactions are observed for G3 and U8 bases but they are of less intensity than that of the U4-G17 and G7-U14 interactions. These bases are also stacked on the previous G-U base pair and this is confirmed by the observation of a similar pattern of

NOEs to that of the U4-G7 interaction. Tables 3.3 and 3.4 consist of complete proton, carbon and phosphorus resonance assignments.

Other important imino proton NOEs include G3 H1 to G17 H1 and G17 H1 to U18 H3. Very important cross-strand ribose to imino NOEs that helped define the duplex were U4 H1', H2', and H3' to G17 H1 as well as C5 H1' to G17 H1. In addition, C5 H2' and H3' to G16 H1 as well as G16H1'-U4H3 are other cross-strand ribose to imino NOE. Important intrastrand NOEs include G3H21 to U4H1', G7 H1 to U8H1', G7H21 to U8H1', G17H1 to U18H1' and G17H21 to U18H1'. These NOEs help determine how these bases stack upon one another when forming the helix and can be seen in Figure 3.3.

**Figure 3.2: Imino Proton $H_2O$ NOESY Spectrum**

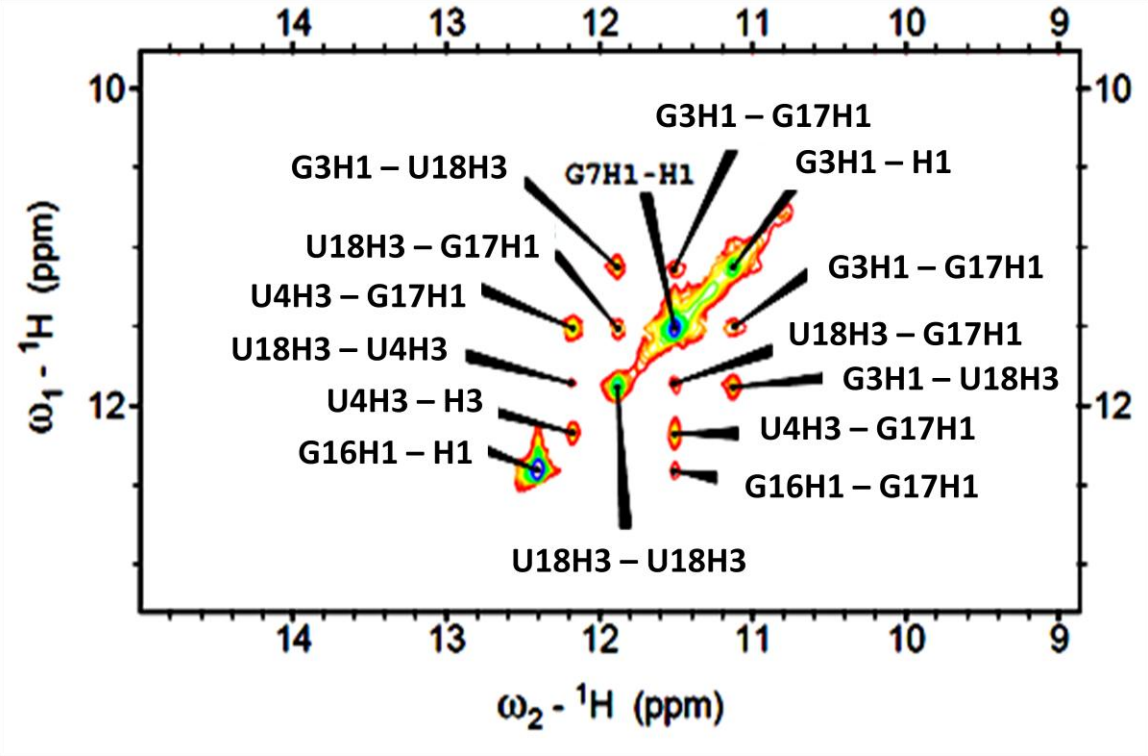



**Figure 3.2:** $H_2O$ NOESY showing imino to imino exchangeable proton NOEs. Spectrum was taken at $1^{o}$ C in 50 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $H_2O$ with 10% $D_2O$.

**Figure 3.3: Exchangeable Proton $H_2O$ NOESY Spectrum**

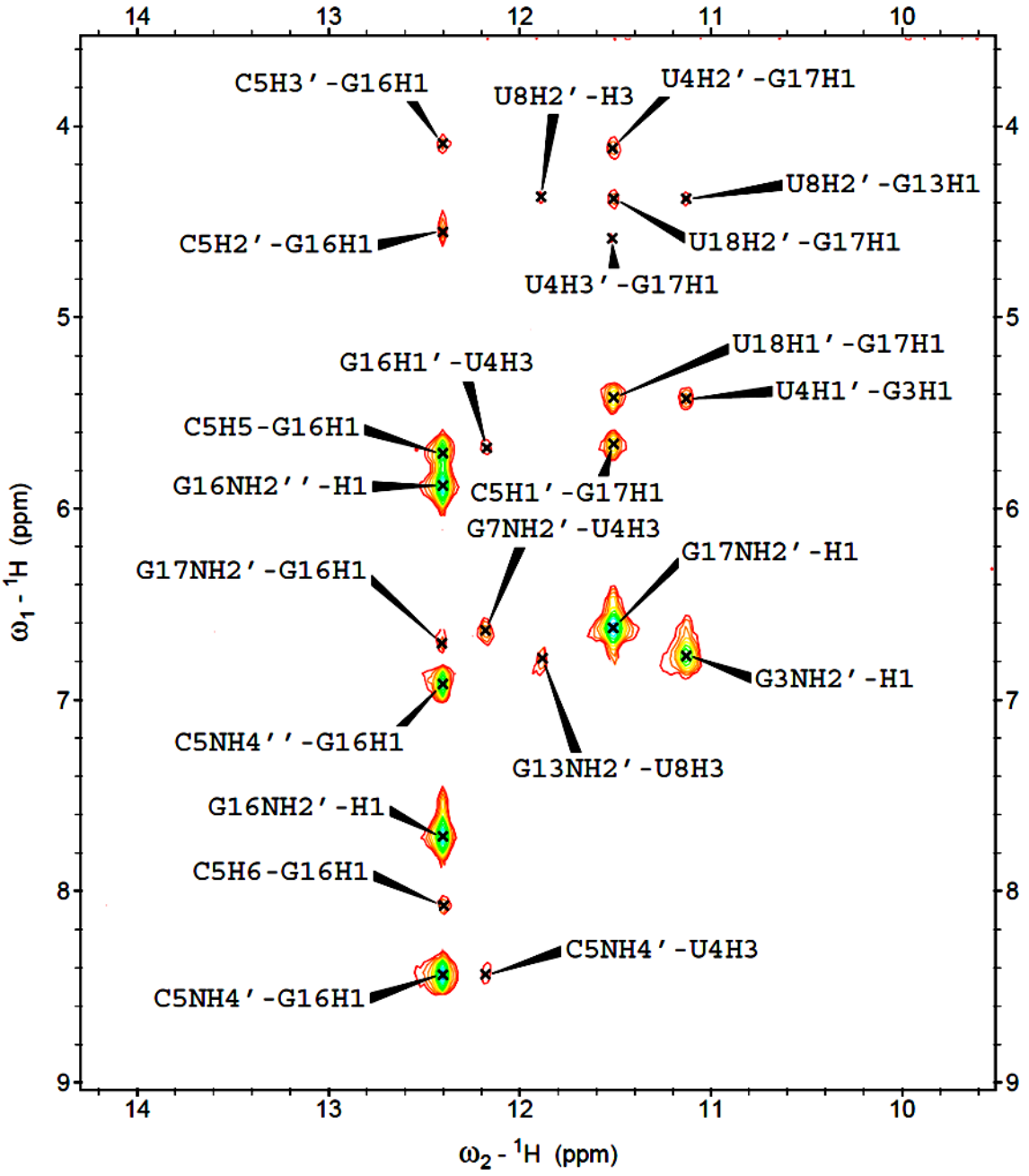



**Figure 3.3:** $H_2O$ NOESY showing exchangeable proton NOEs. Spectrum was taken at $1^{o}$ C in 50 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $H_2O$ with 10% $D_2O$.

### 3.3.3 $D_2O$ NOESY.

Sequential assignment of ribose and aromatic protons was done using the $D_2O$ NOESY spectra and standard procedures (*22, 25*) (Figure 3.4, Table 3.1). A normal NOESY sequential walk is observed for the complete sequence without any breaks in the walk. There are no strong H1' to H8 intranucleotide crosspeaks. This suggests that all bases are in the anti conformation. No other unusual NOEs were observed.

**Figure 3.4: $D_2O$ NOESY Spectrum**

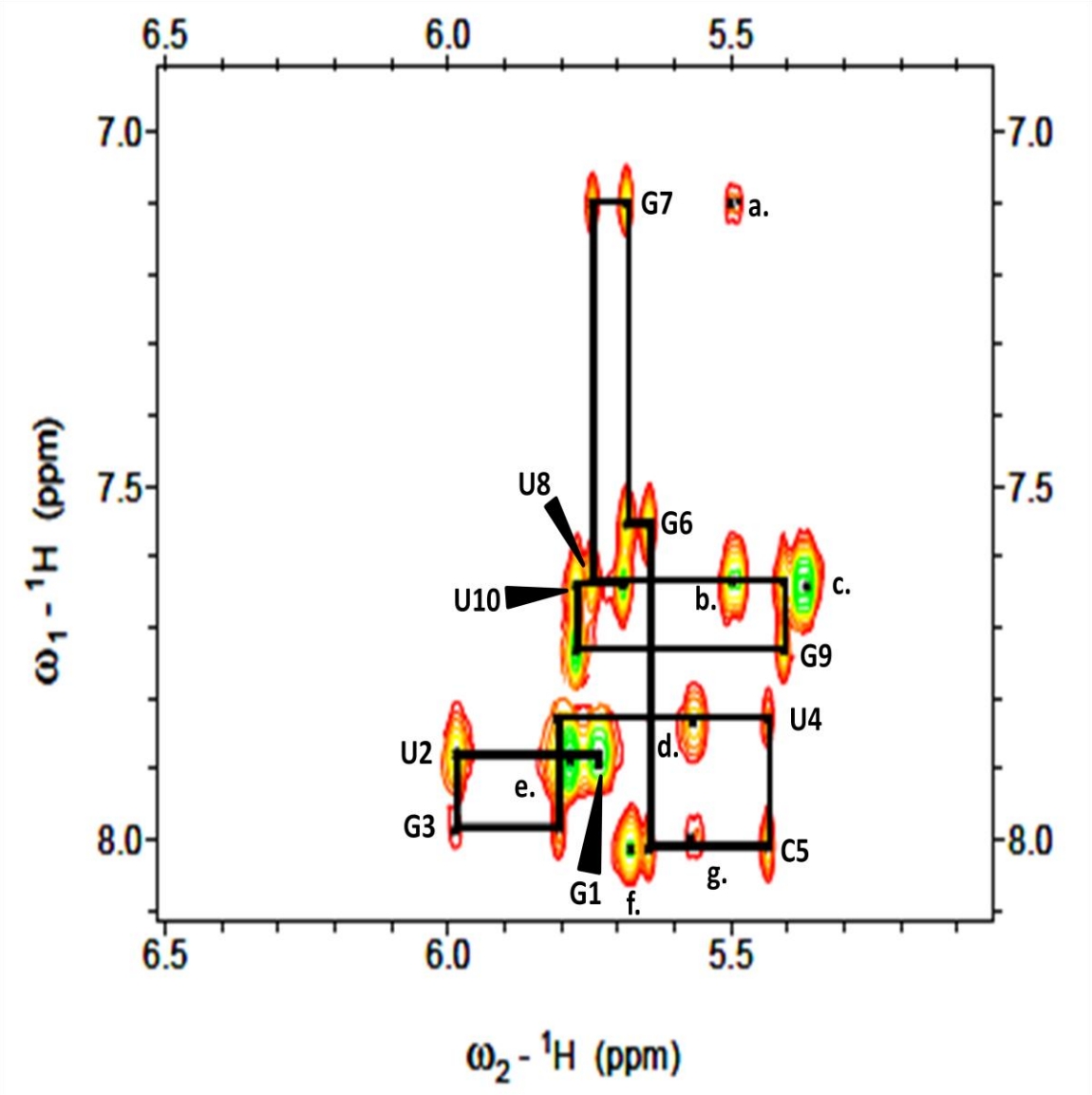



**Figure 3.4:** Base to H1' and H5 region of the 400 ms D2O NOESY spectrum recorded at 1o C for 5' r(**GUGU**CG**GUGU**)2. The base-H1' NOESY walk is indicated by solid lines. Labels represent intranucleotide aromatic-H1' NOEs. The other labels are as follows: (a.) G7 H8–U8 H5 (b.) U8 H5-H6 (c.) U10H5-H6 (d.) U4 H5-H6 (e.) U2 H5-H6 (f.) C5 H5-H6 (g.) G3 H8-U4 H5

**Table 3.1: Proton Resonance Table**

| | H6/H8 | H5 | H1' | H2' | H3' | H4' | H5'/5" | H1/H3 | H2/H4/H6 Amino |
|---|---|---|---|---|---|---|---|---|---|
| G1 | 7.89 | - | 5.77 | 4.82 | 4.62 | 4.34 | 3.83/3.94 | Not Assigned | Not Assigned |
| U2 | 7.88 | 5.79 | 5.99 | 4.5 | 4.75 | 4.57 | 4.35/4.19 | Not Assigned | - |
| G3 | 7.99 | - | 5.81 | 4.94 | 4.35 | 4.21 | 4.59/4.36 | 11.13 | 6.77 |
| U4 | 7.83 | 5.56 | 5.43 | 4.12 | 4.61 | 4.36 | 4.58/4.45 | 12.17 | - |
| C5 | 8.01 | 5.68 | 5.66 | 4.56 | 4.53 | 4.34 | 4.38/4.08 | - | 8.43/6.92 |
| G6 | 7.55 | - | 5.68 | 4.54 | 4.63 | 4.5 | 4.11/4.43 | 12.41 | 7.71 |
| G7 | 7.1 | - | 5.75 | 4.73 | 4.52 | 4.44 | 4.24/4.05 | 11.5 | 6.62 |
| U8 | 7.63 | 5.49 | 5.41 | 4.34 | 4.46 | 4.3 | 4.16/4.04 | 11.86 | - |
| G9 | 7.74 | - | 5.78 | 4.53 | 4.25 | 4.43 | 4.19/4.09 | Not Assigned | Not Assigned |
| U10 | 7.64 | 5.37 | 5.69 | 4.09 | 4.18 | 4.36 | 4.53/4.43 | Not Assigned | - |

**Table 3.1:** Proton chemical shift assignments. The protons were referenced to water, which was referenced to 3-trimethylsilyl propionic acid in 10 mM phosphate buffer pH 6. All protons were assigned in experiments at $1^{o}$ C.

| **Table 3.2:** Restraints for 5' r(**GUGU**CG**GUGU**)$_2$ | | | |
|---|---|---|---|
| **Base** | **Intra** | **Inter** | **Dihedral** |
| G1 | 17 | 3 | 3 |
| U2 | 19 | 7 | 5 |
| G3 | 10 | 15 | 5 |
| U4 | 12 | 19 | 5 |
| C5 | 14 | 15 | 5 |
| G6 | 12 | 12 | 5 |
| G7 | 14 | 11 | 5 |
| U8 | 12 | 17 | 5 |
| G9 | 10 | 10 | 5 |
| U10 | 10 | 5 | 3 |
| **Total** | 260 | 114 | 92 |

**Table 3.2:** Bases are numbered in the 5’ to 3’ direction. Intra refers to NOEs between protons on the same nucleotide and are synonymous for each strand (i.e. G1 has the same NOEs as G11, U2 to U12 and so forth). The total number of Intranucleotide NOEs is doubled because the oligomer forms a self-complimentary duplex. Inter refers to NOEs between protons on different nucleotides. Note that each NOE occurs between two protons, so the total internucleotide NOEs are the sum of the internucleotide NOEs for each nucleotide divided by 2. Since there are two strands, there is no need to divide by 2

## Table 3.3: NOE Assignments

| | | | | | | |
|---|---|---|---|---|---|---|
| G1 H1' | G1 H2' | 2.1 0.6 0.9 | | U2 H5 | U2 H6 | 2.1 0.6 0.9 |
| G1 H1' | G1 H3' | 4.0 1.5 2.0 | | U2 H6 | G3 H8 | 4.0 1.0 2.0 |
| G1 H1' | G1 H4' | 4.0 1.5 2.0 | | G3 H1 | U4 H1' | 4.0 1.5 2.0 |
| G1 H1' | G1 H5' | 4.0 1.5 2.0 | | G3 H1 | G17 H1 | 3.0 1.0 1.5 |
| G1 H1' | G1 H5'' | 4.0 1.5 2.0 | | G3 H1 | U18 H3 | 2.1 0.6 0.9 |
| G1 H1' | G1 H8 | 3.0 1.0 1.5 | | G3 H1' | G3 H2' | 2.1 0.6 0.9 |
| G1 H2' | G1 H3' | 2.1 0.6 0.9 | | G3 H1' | U4 H5 | 4.0 1.5 2.0 |
| G1 H2' | G1 H4' | 4.0 1.5 2.0 | | G3 H1' | U4 H6 | 4.0 1.5 2.0 |
| G1 H2' | G1 H5' | 4.0 1.5 2.0 | | G3 H1' | G3 H8 | 4.0 1.5 2.0 |
| G1 H2' | G1 H5'' | 4.0 1.5 2.0 | | G3 H2' | G3 H3' | 2.1 0.6 0.9 |
| G1 H2' | G1 H8 | 4.0 1.5 2.0 | | G3 H2' | G3 H4' | 4.0 1.5 2.0 |
| G1 H2' | U2 H5 | 4.0 1.5 2.0 | | G3 H2' | G3 H5' | 4.0 1.5 2.0 |
| G1 H2' | U2 H6 | 2.6 0.4 0.9 | | G3 H2' | G3 H8 | 3.0 1.0 1.5 |
| G1 H3' | G1 H5' | 3.0 1.0 1.5 | | G3 H2' | U4 H5 | 4.0 1.5 2.0 |
| G1 H3' | G1 H5'' | 3.0 1.0 1.5 | | G3 H2' | U4 H6 | 3.0 1.0 1.5 |
| G1 H3' | G1 H8 | 4.0 1.5 2.0 | | G3 H21 | G3 H1 | 2.5 0.6 0.9 |
| G1 H3' | U2 H6 | 4.0 1.5 2.0 | | G3 H21 | U4 H1' | 4.0 1.5 2.0 |
| G1 H4' | G1 H5' | 2.1 0.6 0.9 | | G3 H21 | U18 H3 | 4.0 1.5 2.0 |
| G1 H4' | G1 H5'' | 2.1 0.6 0.9 | | G3 H3' | G3 H4' | 2.5 0.6 0.9 |
| G1 H5' | G1 H5'' | 2.1 0.6 0.9 | | G3 H3' | U4 H6 | 4.0 1.5 2.0 |
| U2 H1' | U2 H2' | 2.1 0.6 0.9 | | G3 H3' | G3 H8 | 4.0 1.5 2.0 |
| U2 H1' | U2 H3' | 4.0 1.5 2.0 | | G3 H5' | G3 H5'' | 2.1 0.6 0.9 |
| U2 H1' | U2 H4' | 4.0 1.5 2.0 | | G3 H8 | U4 H5 | 4.0 1.5 2.0 |
| U2 H1' | U2 H5' | 4.0 1.5 2.0 | | G3 H8 | U4 H6 | 4.0 1.0 2.0 |
| U2 H1' | U2 H5'' | 4.0 1.5 2.0 | | U4 H1' | U4 H2' | 2.1 0.6 0.9 |
| U2 H1' | U2 H6 | 3.0 1.0 1.5 | | U4 H1' | U4 H3' | 3.0 1.0 1.5 |
| U2 H1' | G3 H8 | 4.0 1.5 2.0 | | U4 H1' | C5 H5 | 4.0 1.5 2.0 |
| U2 H2' | U2 H3' | 2.1 0.6 0.9 | | U4 H1' | U4 H6 | 3.0 1.0 1.5 |
| U2 H2' | U2 H4' | 4.0 1.5 2.0 | | U4 H1' | C5 H6 | 4.0 1.5 2.0 |
| U2 H2' | U2 H5' | 4.0 1.5 2.0 | | U4 H1' | G17 H1 | 4.0 1.5 2.0 |
| U2 H2' | U2 H6 | 2.6 0.4 0.9 | | U4 H2' | U4 H3' | 2.1 0.6 0.9 |
| U2 H2' | G3 H8 | 2.6 0.4 0.9 | | U4 H2' | U4 H4' | 4.0 1.5 2.0 |
| U2 H3' | U2 H4' | 4.0 1.5 2.0 | | U4 H2' | C5 H5 | 4.0 1.5 2.0 |
| U2 H3' | U2 H5' | 4.0 1.5 2.0 | | U4 H2' | U4 H6 | 4.0 1.5 2.0 |
| U2 H3' | U2 H5'' | 4.0 1.5 2.0 | | U4 H2' | C5 H6 | 2.6 0.4 0.9 |
| U2 H3' | U2 H6 | 4.0 1.5 2.0 | | U4 H3 | G17 H1 | 2.1 0.6 0.9 |
| U2 H3' | G3 H8 | 4.0 1.5 2.0 | | U4 H3 | G17 H21 | 4.0 1.5 2.0 |
| U2 H4' | U2 H5' | 4.0 1.5 2.0 | | U4 H3 | U18 H3 | 3.0 1.0 1.5 |
| U2 H4' | U2 H5'' | 3.0 1.0 1.5 | | U4 H3' | U4 H4' | 2.5 0.6 0.9 |
| U2 H5' | U2 H5'' | 2.1 0.6 0.9 | | U4 H3' | U4 H6 | 3.0 1.0 1.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| U4 H3' | C5 H6 | 4.0 1.5 2.0 | | G6 H3' | G6 H4' | 2.5 0.6 0.9 |
| U4 H4' | U4 H5' | 2.1 0.6 0.9 | | G6 H3' | G6 H5' | 4.0 1.5 2.0 |
| U4 H4' | U4 H5'' | 3.0 1.0 1.5 | | G6 H3' | G6 H8 | 4.0 1.5 2.0 |
| U4 H4' | U4 H6 | 4.0 1.5 2.0 | | G6 H4' | G6 H5' | 2.1 0.6 0.9 |
| U4 H5' | U4 H5'' | 2.1 0.6 0.9 | | G6 H4' | G6 H5'' | 3.0 1.0 1.5 |
| U4 H5 | U4 H6 | 2.1 0.6 0.9 | | G6 H8 | G7 H8 | 4.0 1.0 2.0 |
| U4 H6 | C5 H6 | 4.0 1.0 2.0 | | G7 H1 | G7 H21 | 2.5 0.6 0.9 |
| C5 H1' | C5 H2' | 2.1 0.6 0.9 | | G7 H1 | U8 H1' | 3.0 1.0 1.5 |
| C5 H1' | C5 H3' | 3.0 1.0 1.5 | | G7 H1 | U8 H3 | 4.0 1.5 2.0 |
| C5 H1' | C5 H6 | 3.0 1.0 1.5 | | G7 H1' | G7 H2' | 2.1 0.6 0.9 |
| C5 H1' | G6 H8 | 3.0 1.0 1.5 | | G7 H1' | G7 H5' | 4.0 1.5 2.0 |
| C5 H1' | G17 H1 | 4.0 1.5 2.0 | | G7 H1' | G7 H5'' | 4.0 1.5 2.0 |
| C5 H2' | C5 H3' | 2.1 0.6 0.9 | | G7 H1' | G7 H8 | 4.0 1.5 2.0 |
| C5 H2' | C5 H4' | 3.0 1.0 1.5 | | G7 H1' | U8 H5 | 4.0 1.5 2.0 |
| C5 H2' | C5 H6 | 2.5 0.6 0.9 | | G7 H1' | U8 H6 | 3.0 1.0 1.5 |
| C5 H2' | G6 H8 | 2.5 0.6 0.9 | | G7 H2' | G7 H3' | 2.1 0.6 0.9 |
| C5 H2' | G16 H1 | 4.0 1.5 2.0 | | G7 H2' | G7 H4' | 4.0 1.5 2.0 |
| C5 H3' | C5 H4' | 2.5 0.6 0.9 | | G7 H2' | G7 H5' | 4.0 1.5 2.0 |
| C5 H3' | C5 H6 | 4.0 1.5 2.0 | | G7 H2' | G7 H5'' | 4.0 1.5 2.0 |
| C5 H3' | G6 H1 | 4.0 1.5 2.0 | | G7 H2' | G7 H8 | 2.6 0.4 0.9 |
| C5 H41 | C5 H5 | 3.0 1.0 1.5 | | G7 H2' | U8 H5 | 4.0 1.5 2.0 |
| C5 H41 | C5 H42 | 1.2 0.6 0.9 | | G7 H2' | U8 H6 | 2.6 0.4 0.9 |
| C5 H41 | G16 H1 | 3.0 1.0 1.5 | | G7 H21 | U8 H1' | 4.0 1.5 2.0 |
| C5 H42 | C5 H5 | 2.7 0.6 1.1 | | G7 H3' | G7 H4' | 2.5 0.6 0.9 |
| C5 H42 | G16 H1 | 3.6 1.0 1.5 | | G7 H3' | G7 H8 | 4.0 1.5 2.0 |
| C5 H4' | C5 H5'' | 3.0 1.0 1.5 | | G7 H4' | G7 H5' | 2.1 0.6 0.9 |
| C5 H5' | C5 H5'' | 2.1 0.6 0.9 | | G7 H4' | G7 H5'' | 3.0 1.0 1.5 |
| C5 H5 | C5 H6 | 2.1 0.6 0.9 | | G7 H5' | G7 H5'' | 2.1 0.6 0.9 |
| C5 H5 | G6 H1 | 4.0 1.5 2.0 | | G7 H8 | U8 H5 | 4.0 1.5 2.0 |
| C5 H6 | G6 H8 | 4.0 1.0 2.0 | | G7 H8 | U8 H6 | 4.0 1.0 2.0 |
| G6 H1 | G6 H21 | 3.0 1.0 1.5 | | U8 H1' | U8 H2' | 2.1 0.6 0.9 |
| G6 H1 | G7 H1 | 3.0 1.0 1.5 | | U8 H1' | U8 H6 | 4.0 1.5 2.0 |
| G6 H1' | G6 H2' | 2.1 0.6 0.9 | | U8 H1' | G9 H1' | 4.0 1.5 2.0 |
| G6 H1' | G6 H3' | 4.0 1.5 2.0 | | U8 H1' | G9 H8 | 4.0 1.5 2.0 |
| G6 H1' | G6 H4' | 4.0 1.5 2.0 | | U8 H2' | U8 H3' | 2.1 0.6 0.9 |
| G6 H1' | G6 H8 | 3.0 1.0 1.5 | | U8 H2' | U8 H6 | 2.6 0.4 0.9 |
| G6 H1' | G7 H8 | 3.0 1.0 1.5 | | U8 H2' | G9 H8 | 2.6 0.4 0.9 |
| G6 H2' | G6 H3' | 2.1 0.6 0.9 | | U8 H3' | U8 H4' | 2.1 0.6 0.9 |
| G6 H2' | G6 H8 | 2.6 0.4 0.9 | | U8 H3' | U8 H5' | 4.0 1.5 2.0 |
| G6 H2' | G7 H8 | 2.6 0.4 0.9 | | U8 H3' | U8 H5'' | 4.0 1.5 2.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| U8 H3' | U8 H6 | 4.0 1.5 2.0 | | G11 H2' | G11 H5' | 4.0 1.5 2.0 |
| U8 H3' | G9 H8 | 4.0 1.5 2.0 | | G11 H2' | G11 H5'' | 4.0 1.5 2.0 |
| U8 H4' | U8 H5' | 2.1 0.6 0.9 | | G11 H2' | G11 H8 | 4.0 1.5 2.0 |
| U8 H4' | U8 H5'' | 3.0 1.0 1.5 | | G11 H2' | U12 H5 | 4.0 1.5 2.0 |
| U8 H5' | U8 H5'' | 2.1 0.6 0.9 | | G11 H2' | U12 H6 | 2.6 0.4 0.9 |
| U8 H5 | U8 H6 | 2.1 0.6 0.9 | | G11 H3' | G11 H5' | 3.0 1.0 1.5 |
| U8 H6 | G9 H8 | 4.0 1.0 2.0 | | G11 H3' | G11 H5'' | 3.0 1.0 1.5 |
| G9 H1' | G9 H2' | 2.1 0.6 0.9 | | G11 H3' | G11 H8 | 4.0 1.5 2.0 |
| G9 H1' | G9 H4' | 4.0 1.5 2.0 | | G11 H3' | U12 H6 | 4.0 1.5 2.0 |
| G9 H1' | G9 H5' | 4.0 1.5 2.0 | | G11 H4' | G11 H5' | 2.1 0.6 0.9 |
| G9 H1' | G9 H8 | 3.0 1.0 1.5 | | G11 H4' | G11 H5'' | 2.1 0.6 0.9 |
| G9 H1' | U10 H5 | 4.0 1.5 2.0 | | G11 H5' | G11 H5'' | 2.1 0.6 0.9 |
| G9 H1' | U10 H6 | 4.0 1.5 2.0 | | U12 H1' | U12 H2' | 2.1 0.6 0.9 |
| G9 H2' | G9 H4' | 4.0 1.5 2.0 | | U12 H1' | U12 H3' | 4.0 1.5 2.0 |
| G9 H2' | G9 H8 | 2.1 0.6 0.9 | | U12 H1' | U12 H4' | 4.0 1.5 2.0 |
| G9 H2' | U10 H5 | 4.0 1.5 2.0 | | U12 H1' | U12 H5' | 4.0 1.5 2.0 |
| G9 H2' | U10 H6 | 2.1 0.6 0.9 | | U12 H1' | U12 H5'' | 4.0 1.5 2.0 |
| G9 H3' | G9 H5' | 4.0 1.5 2.0 | | U12 H1' | U12 H6 | 3.0 1.0 1.5 |
| G9 H4' | G9 H5' | 3.0 1.0 1.5 | | U12 H1' | G13 H8 | 4.0 1.5 2.0 |
| G9 H4' | G9 H5'' | 4.0 1.5 2.0 | | U12 H2' | U12 H3' | 2.1 0.6 0.9 |
| G9 H5' | G9 H5'' | 2.1 0.6 0.9 | | U12 H2' | U12 H4' | 4.0 1.5 2.0 |
| G9 H8 | U10 H6 | 4.0 1.0 2.0 | | U12 H2' | U12 H5' | 4.0 1.5 2.0 |
| U10 H1' | U10 H2' | 2.1 0.6 0.9 | | U12 H2' | U12 H6 | 2.6 0.4 0.9 |
| U10 H1' | U10 H5 | 4.0 1.5 2.0 | | U12 H2' | G13 H8 | 2.6 0.4 0.9 |
| U10 H1' | U10 H6 | 4.0 1.5 2.0 | | U12 H3' | U12 H4' | 2.5 0.6 0.9 |
| U10 H2' | U10 H3' | 2.1 0.6 0.9 | | U12 H3' | U12 H5' | 4.0 1.5 2.0 |
| U10 H2' | U10 H4' | 4.0 1.5 2.0 | | U12 H3' | U12 H5'' | 4.0 1.5 2.0 |
| U10 H2' | U10 H6 | 2.6 0.4 0.9 | | U12 H3' | U12 H6 | 4.0 1.5 2.0 |
| U10 H3' | U10 H4' | 2.5 0.6 0.9 | | U12 H3' | G13 H8 | 4.0 1.5 2.0 |
| U10 H3' | U10 H6 | 4.0 1.5 2.0 | | U12 H4' | U12 H5' | 3.0 1.0 1.5 |
| U10 H5 | U10 H6 | 2.1 0.6 0.9 | | U12 H4' | U12 H5'' | 3.0 1.0 1.5 |
| U10 H5' | U10 H5'' | 2.1 0.6 0.9 | | U12 H5' | U12 H5'' | 2.1 0.6 0.9 |
| G11 H1' | G11 H2' | 2.1 0.6 0.9 | | U12 H5 | U12 H6 | 2.1 0.6 0.9 |
| G11 H1' | G11 H3' | 4.0 1.5 2.0 | | U12 H6 | G13 H8 | 4.0 1.0 2.0 |
| G11 H1' | G11 H4' | 4.0 1.5 2.0 | | G13 H1 | U14 H1' | 4.0 1.5 2.0 |
| G11 H1' | G11 H5' | 4.0 1.5 2.0 | | G13 H1 | G7 H1 | 4.0 1.5 2.0 |
| G11 H1' | G11 H5'' | 4.0 1.5 2.0 | | G13 H1 | U8 H3 | 2.1 0.6 0.9 |
| G11 H1' | G11 H8 | 4.0 1.5 2.0 | | G13 H1' | G13 H2' | 2.1 0.6 0.9 |
| G11 H2' | G11 H3' | 4.0 1.5 2.0 | | G13 H1' | U14 H5 | 4.0 1.5 2.0 |
| G11 H2' | G11 H4' | 4.0 1.5 2.0 | | G13 H1' | U14 H6 | 3.0 1.0 1.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| G13 H1' | G13 H8 | 3.0 1.0 1.5 | | C15 H1' | C15 H6 | 3.0 1.0 1.5 |
| G13 H2' | G13 H3' | 2.1 0.6 0.9 | | C15 H1' | G16 H8 | 3.0 1.0 1.5 |
| G13 H2' | G13 H4' | 4.0 1.5 2.0 | | C15 H1' | G7 H1 | 4.0 1.5 2.0 |
| G13 H2' | G13 H5' | 4.0 1.5 2.0 | | C15 H2' | C15 H3' | 2.1 0.6 0.9 |
| G13 H2' | G13 H8 | 3.0 1.0 1.5 | | C15 H2' | C15 H4' | 4.0 1.5 2.0 |
| G13 H2' | U14 H5 | 4.0 1.5 2.0 | | C15 H2' | C15 H6 | 2.6 0.4 0.9 |
| G13 H2' | U14 H6 | 3.0 1.0 1.5 | | C15 H2' | G16 H8 | 2.6 0.4 0.9 |
| G13 H21 | G13 H1 | 2.5 0.6 0.9 | | C15 H2' | G6 H1 | 4.0 1.5 2.0 |
| G13 H21 | U14 H1' | 4.0 1.5 2.0 | | C15 H3' | C15 H4' | 2.5 0.6 0.9 |
| G13 H21 | U8 H3 | 3.0 1.0 1.5 | | C15 H3' | C15 H6 | 4.0 1.5 2.0 |
| G13 H3' | G13 H4' | 2.5 0.6 0.9 | | C15 H41 | C15 H5 | 3.0 1.0 1.5 |
| G13 H3' | U14 H6 | 4.0 1.5 2.0 | | C15 H41 | C15 H42 | 1.2 0.6 0.9 |
| G13 H3' | G13 H8 | 4.0 1.5 2.0 | | C15 H41 | G6 H1 | 3.0 1.0 1.5 |
| G13 H5' | G13 H5'' | 2.1 0.6 0.9 | | C15 H42 | C15 H5 | 2.7 1.0 1.5 |
| G13 H8 | U14 H5 | 4.0 1.5 2.0 | | C15 H42 | G6 H1 | 3.6 1.0 1.5 |
| G13 H8 | U14 H6 | 4.0 1.0 2.0 | | C15 H4' | C15 H5'' | 3.0 1.0 1.5 |
| U14 H1' | U14 H2' | 2.1 0.6 0.9 | | C15 H5' | C15 H5'' | 2.1 0.6 0.9 |
| U14 H1' | U14 H3' | 3.0 1.0 1.5 | | C15 H5 | C15 H6 | 2.1 0.6 0.9 |
| U14 H1' | C15 H5 | 4.0 1.5 2.0 | | C15 H6 | G16 H8 | 4.0 1.0 2.0 |
| U14 H1' | U14 H6 | 4.0 1.5 2.0 | | G16 H1 | G16 H21 | 2.1 0.6 0.9 |
| U14 H1' | C15 H6 | 4.0 1.5 2.0 | | G16 H1 | G17 H1 | 3.0 1.0 1.5 |
| U14 H1' | G7 H1 | 4.0 1.5 2.0 | | G16 H1' | G16 H2' | 2.1 0.6 0.9 |
| U14 H2' | U14 H3' | 2.1 0.6 0.9 | | G16 H1' | G16 H3' | 4.0 1.5 2.0 |
| U14 H2' | U14 H4' | 4.0 1.5 2.0 | | G16 H1' | G16 H8 | 3.0 1.0 1.5 |
| U14 H2' | C15 H5 | 4.0 1.5 2.0 | | G16 H1' | G17 H8 | 3.0 1.0 1.5 |
| U14 H2' | U14 H6 | 4.0 1.5 2.0 | | G16 H2' | G16 H3' | 2.1 0.6 0.9 |
| U14 H2' | C15 H6 | 2.6 0.4 0.9 | | G16 H2' | G16 H8 | 2.6 0.4 0.9 |
| U14 H3 | G7 H1 | 2.1 0.6 0.9 | | G16 H2' | G17 H8 | 2.6 0.4 0.9 |
| U14 H3 | G7 H21 | 4.0 1.5 2.0 | | G16 H3' | G16 H4' | 2.5 0.6 0.9 |
| U14 H3 | U8 H3 | 3.0 1.0 1.5 | | G16 H3' | G16 H5' | 4.0 1.5 2.0 |
| U14 H3' | U14 H4' | 2.5 0.6 0.9 | | G16 H3' | G16 H8 | 4.0 1.5 2.0 |
| U14 H3' | U14 H6 | 3.0 1.0 1.5 | | G16 H4' | G16 H5' | 2.1 0.6 0.9 |
| U14 H3' | C15 H6 | 4.0 1.5 2.0 | | G16 H4' | G16 H5'' | 3.0 1.0 1.5 |
| U14 H4' | U14 H5' | 2.1 0.6 0.9 | | G16 H8 | G17 H8 | 4.0 1.0 2.0 |
| U14 H4' | U14 H5'' | 3.0 1.0 1.5 | | G17 H1 | G17 H21 | 2.5 0.6 0.9 |
| U14 H5' | U14 H5'' | 2.1 0.6 0.9 | | G17 H1 | U18 H1' | 4.0 1.5 2.0 |
| U14 H5 | U14 H6 | 2.1 0.6 0.9 | | G17 H1 | U18 H3 | 3.0 1.0 1.5 |
| U14 H6 | C15 H6 | 4.0 1.0 2.0 | | G17 H1' | G17 H2' | 2.1 0.6 0.9 |
| C15 H1' | C15 H2' | 2.1 0.6 0.9 | | G17 H1' | G17 H5' | 4.0 1.5 2.0 |
| C15 H1' | C15 H3' | 4.0 1.5 2.0 | | G17 H1' | G17 H5'' | 4.0 1.5 2.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| G17 H1' | G17 H8 | 4.0 1.5 2.0 | | U18 H4' | U18 H5'' | 3.0 1.0 1.5 |
| G17 H1' | U18 H5 | 4.0 1.5 2.0 | | U18 H5' | U18 H5'' | 2.1 0.6 0.9 |
| G17 H1' | U18 H6 | 4.0 1.5 2.0 | | U18 H5 | U18 H6 | 2.1 0.6 0.9 |
| G17 H2' | G17 H3' | 2.1 0.6 0.9 | | U18 H6 | G19 H8 | 4.0 1.0 2.0 |
| G17 H2' | G17 H4' | 4.0 1.5 2.0 | | G19 H1' | G19 H2' | 2.1 0.6 0.9 |
| G17 H2' | G17 H5' | 4.0 1.5 2.0 | | G19 H1' | G19 H4' | 4.0 1.5 2.0 |
| G17 H2' | G17 H5'' | 4.0 1.5 2.0 | | G19 H1' | G19 H5' | 4.0 1.5 2.0 |
| G17 H2' | G17 H8 | 2.6 0.4 0.9 | | G19 H1' | G19 H8 | 3.0 1.0 1.5 |
| G17 H2' | U18 H5 | 4.0 1.5 2.0 | | G19 H1' | U20 H5 | 4.0 1.5 2.0 |
| G17 H2' | U18 H6 | 2.6 0.4 0.9 | | G19 H1' | U20 H6 | 4.0 1.5 2.0 |
| G17 H21 | U18 H1' | 4.0 1.5 2.0 | | G19 H2' | G19 H4' | 4.0 1.5 2.0 |
| G17 H3' | G17 H4' | 2.5 0.6 0.9 | | G19 H2' | G19 H8 | 2.6 0.4 0.9 |
| G17 H3' | G17 H8 | 4.0 1.5 2.0 | | G19 H2' | U20 H5 | 4.0 1.5 2.0 |
| G17 H4' | G17 H5' | 2.1 0.6 0.9 | | G19 H2' | U20 H6 | 2.6 0.4 0.9 |
| G17 H4' | G17 H5'' | 3.0 1.0 1.5 | | G19 H3' | G19 H5' | 4.0 1.5 2.0 |
| G17 H5' | G17 H5'' | 2.1 0.6 0.9 | | G19 H4' | G19 H5' | 3.0 1.0 1.5 |
| G17 H8 | U18 H5 | 4.0 1.5 2.0 | | G19 H4' | G19 H5'' | 3.0 1.0 1.5 |
| G17 H8 | U18 H6 | 4.0 1.0 2.0 | | G19 H5' | G19 H5'' | 2.1 0.6 0.0 |
| U18 H1' | U18 H2' | 2.1 0.6 0.9 | | G19 H8 | U20 H6 | 4.0 1.0 2.0 |
| U18 H1' | U18 H6 | 4.0 1.5 2.0 | | U20 H1' | U20 H2' | 2.1 0.6 0.9 |
| U18 H1' | G19 H8 | 4.0 1.5 2.0 | | U20 H1' | U20 H5 | 4.0 1.5 2.0 |
| U18 H2' | U18 H3' | 2.1 0.6 0.9 | | U20 H1' | U20 H6 | 4.0 1.5 2.0 |
| U18 H2' | U18 H6 | 2.6 0.4 0.9 | | U20 H2' | U20 H3' | 2.1 0.6 0.9 |
| U18 H2' | G19 H8 | 2.6 0.4 0.9 | | U20 H2' | U20 H4' | 4.0 1.5 2.0 |
| U18 H3' | U18 H4' | 2.5 0.6 0.9 | | U20 H2' | U20 H6 | 2.1 0.6 0.9 |
| U18 H3' | U18 H5' | 4.0 1.5 2.0 | | U20 H3' | U20 H4' | 2.5 0.6 0.9 |
| U18 H3' | U18 H5'' | 4.0 1.5 2.0 | | U20 H3' | U20 H6 | 4.0 1.5 2.0 |
| U18 H3' | U18 H6 | 4.0 1.5 2.0 | | U20 H5 | U20 H6 | 2.1 0.6 0.9 |
| U18 H3' | G19 H8 | 4.0 1.5 2.0 | | U20 H5' | U20 H5'' | 2.1 0.6 0.9 |
| U18 H4' | U18 H5' | 2.1 0.6 0.9 | | | | |

**Table 3.3:** NOEs used for molecular modeling with distance bins for restraints. The first number is the median distance. The second number is the maximum distance below the median, and the third number is the maximum above the median distance.

### 3.3.4 DQF-COSY.

The COSY spectra show the five pyrimidine H5-H6 crosspeaks which is expected for the duplex (Figure 3.5). The spectra also show four unusually strong H1' to H2' COSY crosspeaks in the spectra and were assigned to G1, U2, G9 and U10. This information indicates partial C2' endo conformation (Figure 3.6). These bases were allowed to be all C2'-endo, all C3'-endo, or unrestrained in the simulated annealing protocols.

**Figure 3.5: DQF COSY Aromatic Spectrum**

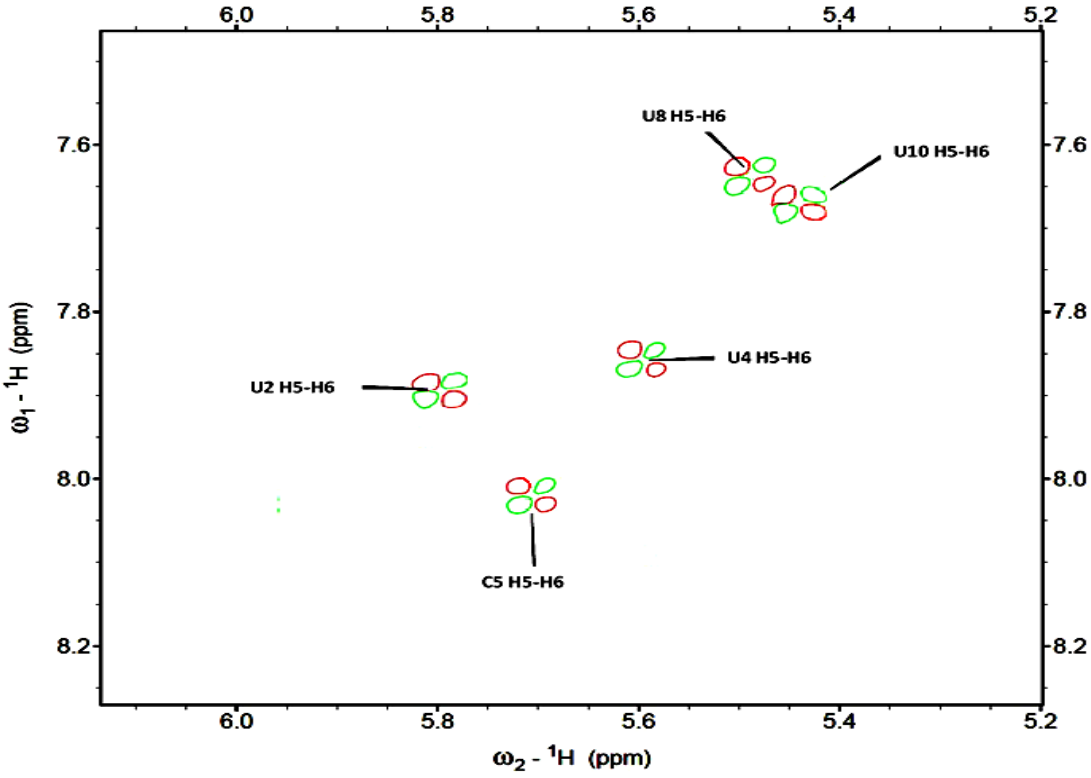



**Figure 3.5:** COSY spectrum showing H5-H6 pyrimidine crosspeaks. Spectrum was taken at 1$^{o}$ C in 50mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $D_2O$.

**Figure 3.6: DQF COSY H1'-H2' Spectrum**

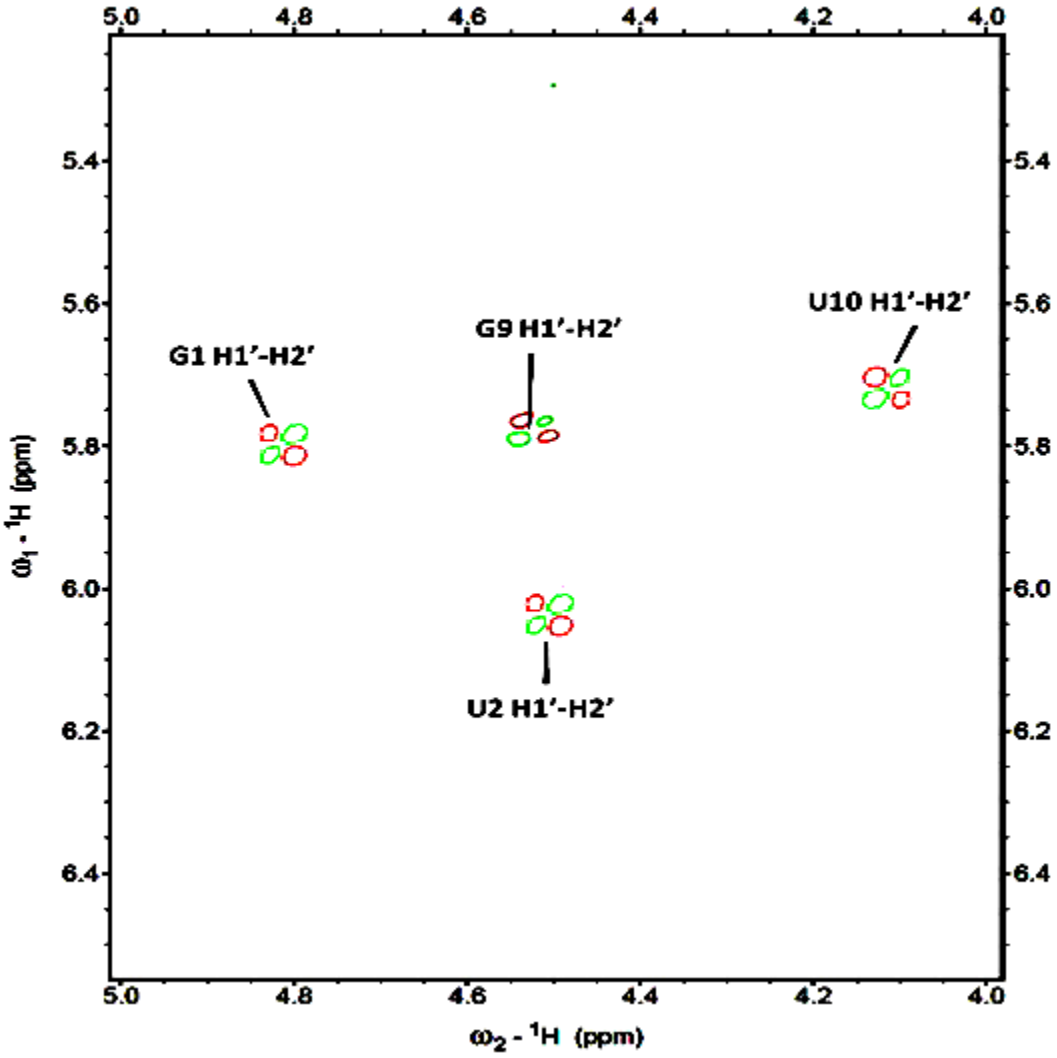



**Figure 3.6:** DQF COSY spectrum showing four very strong H1'-H2' crosspeaks. This shows that these bases assume partial C2'-endo conformation.

**Figure 3.7: Ribose Dihedral Angles**

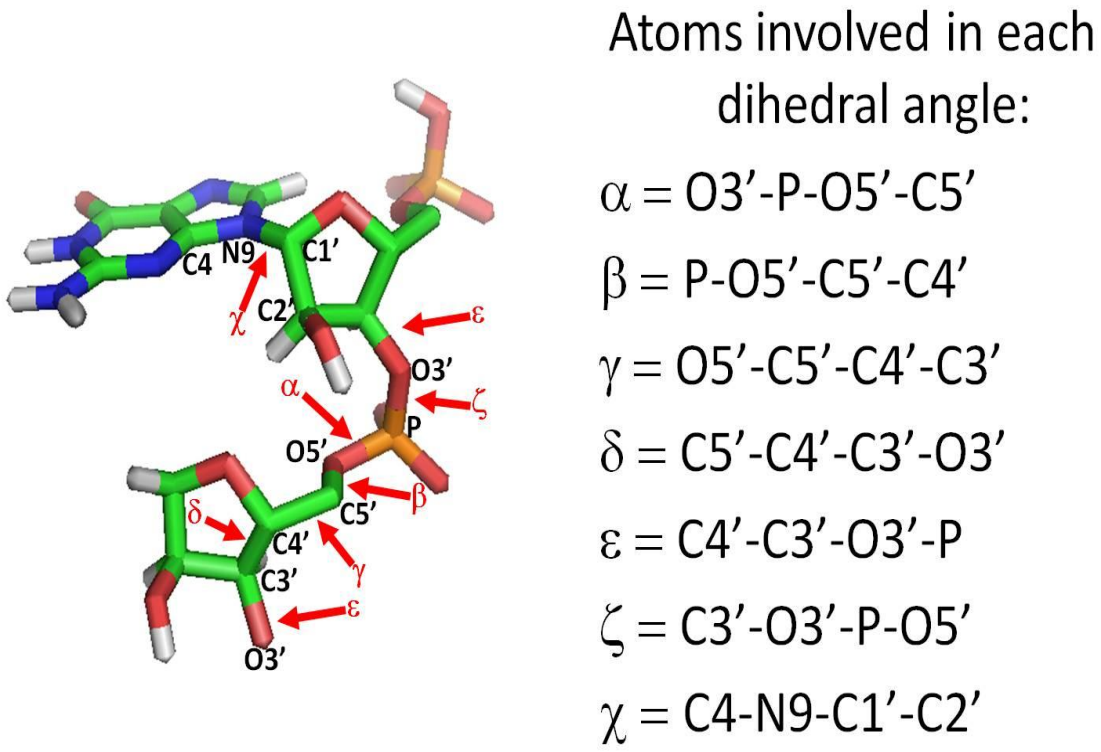
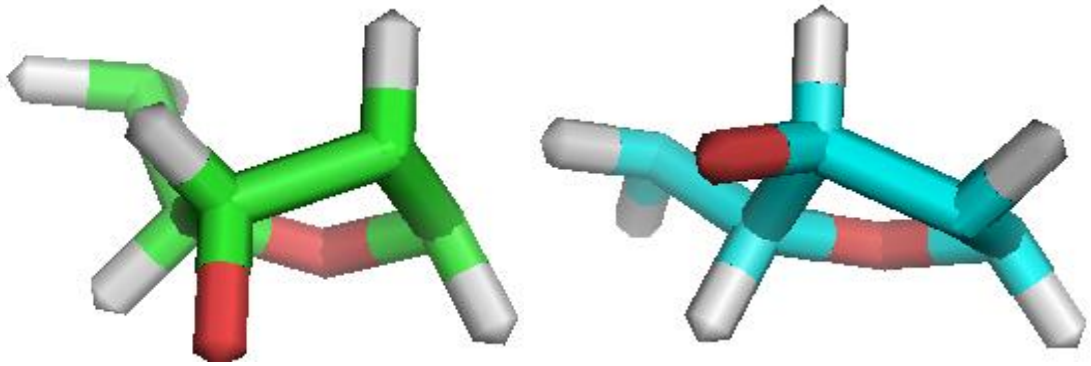



**Figure 3.7:** The ribose dihedral angles are made up of three bonds between four atoms. Ribose protons have been removed so that the angles could be seen. The nomenclature and the atoms involved in each dihedral angle are shown.

**Figure 3.8: Sugar Pucker Conformations**



**A** **B**

**Figure 3.8:** The most common sugar pucker conformation of the ribose in an RNA backbone is C3'-endo (A). Dynamic bases can cause the ribose to also assume the C2'-endo conformational state (B) as well.

### 3.3.5 $^{13}$C HSQC.

The carbon chemical shifts were useful in helping distinguish overlapped aromatic as well as ribose resonances. In the proton spectra the G1 H8 and U2 H6 resonances are overlapped but in figure 3.9 you can clearly see that there are two aromatic resonances at 7.88 ppm. In the ribose H1' region, there are 7 resonances that fall in a 0.3 ppm range in the proton dimension. All 7 of these resonances can be unambiguously assigned using the dispersion in the carbon dimension (Figure 3.10).

**Figure 3.9: $^{13}C$ Aromatic HSQC Spectrum**

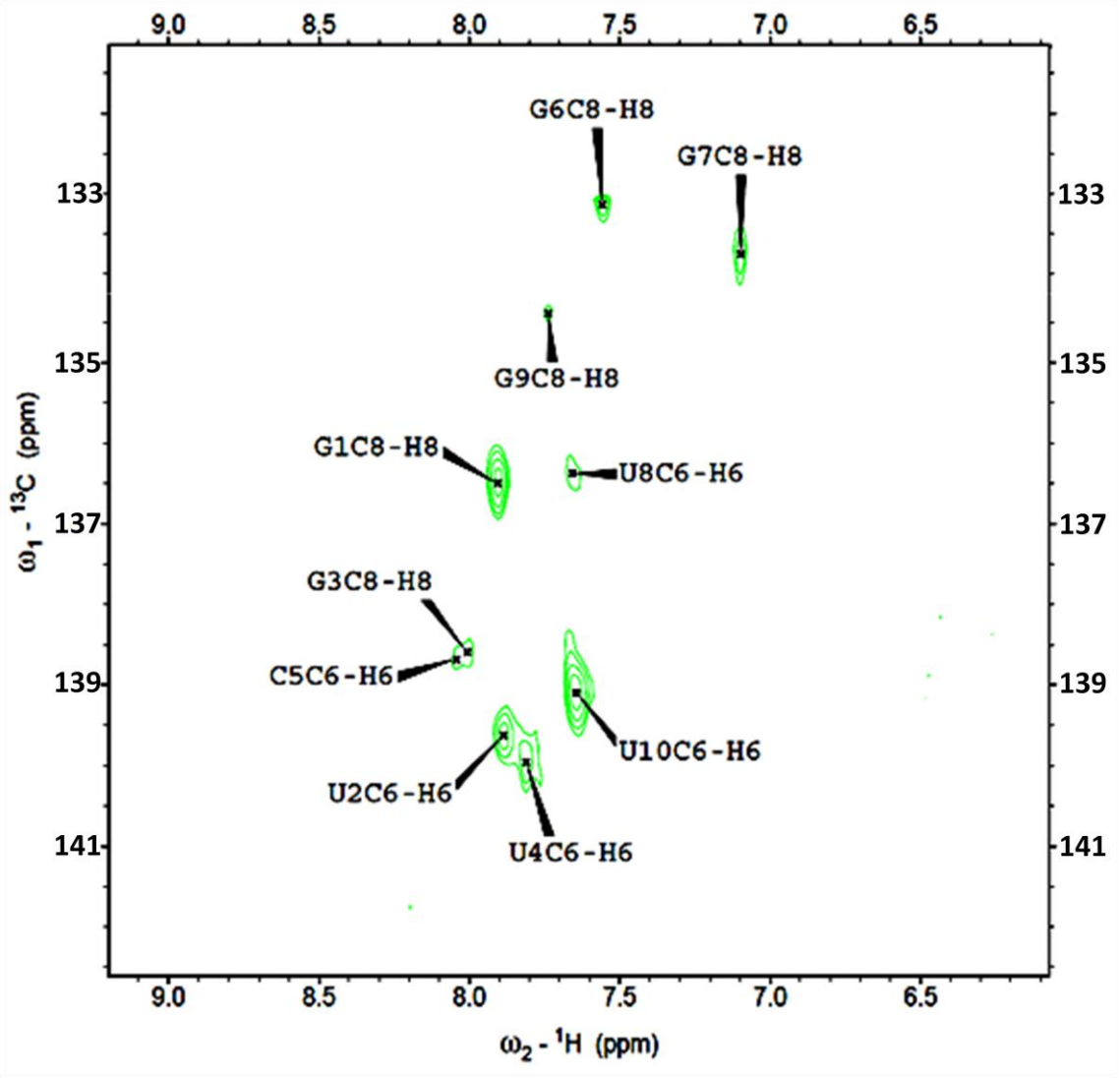



**Figure 3.9:** $^{13}C$ Ribose HSQC spectrum showing the carbon chemical shift dispersion that allows the unambiguous assignment of the aromatic proton resonances. Spectrum was taken at $1^{o}$ C in 50mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $D_2O$.

**Figure 3.10: $^{13}$C Ribose HSQC Spectrum**

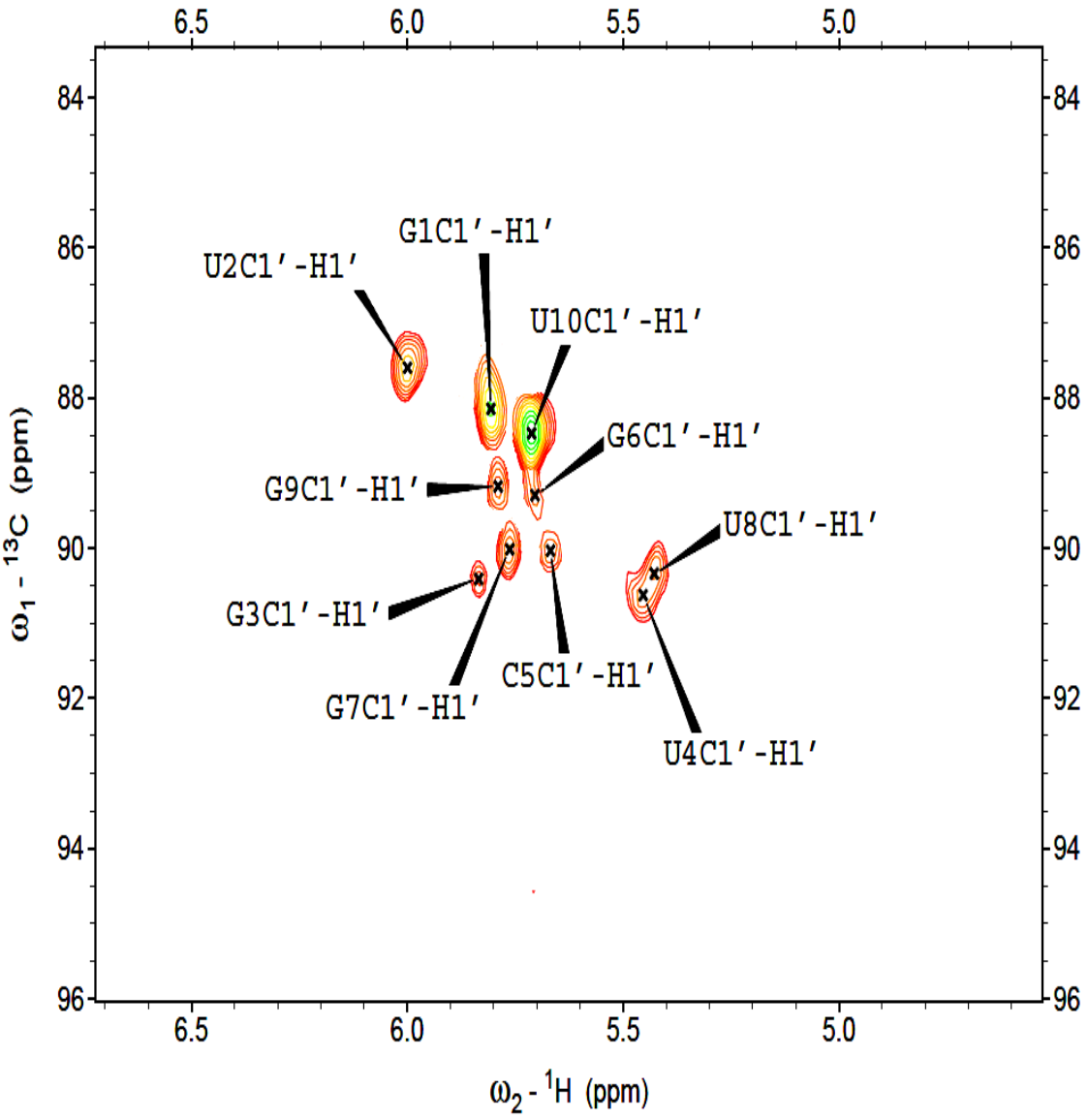



**Figure 3.10:** $^{13}$C Ribose HSQC spectrum showing the carbon chemical shift dispersion that allows the unambiguous assignment of the H1' ribose protons. Spectrum was taken at 1°C in 50mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $D_2O$.

### 3.3.6 $^{31}$P HETCOR.

The phosphorus chemical shifts aid in the unambiguous assignment of the H3' ribose protons. In the spectrum these protons fall within a 1 ppm range which is common for normal A-form RNA. Since none are outside of this range, this is consistent with the final ensemble of structures in near A-form orientation. The value of the coupling constants were measured and compared to A-form RNA. These coupling constants are directly affected by the dihedral orientation of the epsilon angle. Two angles, G3 and G9, were found to have a 28 and 33 percent higher coupling constants than that of A-form RNA, respectively. This information can be seen in figure 3.11. All carbon and phosphorus chemical shifts are located in table 3.4

**Figure 3.11: $^{31}$P HETCOR Spectrum**

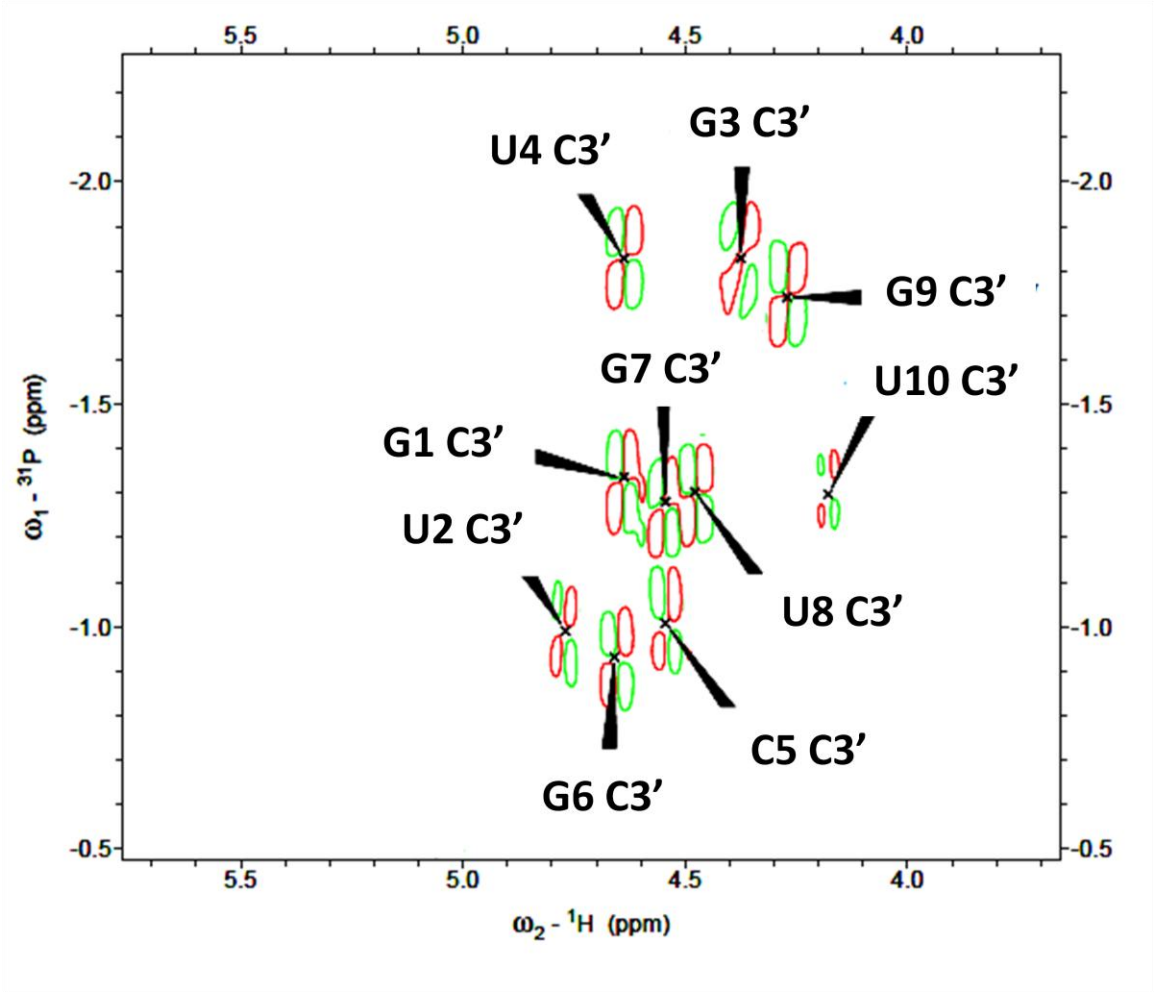



**Figure 3.11:** $^{31}$P HETCOR Spectrum showing the ribose C3' resonances. Spectrum was taken at 1°C in 50 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $D_2O$.

**Table 3.4: Carbon and Phosphorus Resonance Table**

| | C5 | C6/C8 | C1' | C2' | C3' | O3P |
|---|---|---|---|---|---|---|
| G1 | - | 135.5 | 88.33 | 73.4 | 80.4 | -1.33 |
| U2 | 101.9 | 139.63 | 87.51 | 73.3 | 76.3 | -0.99 |
| G3 | - | 138.6 | 90.48 | 73.2 | 79.5 | -1.83 |
| U4 | 101.1 | 139.96 | 90.63 | 73.8 | 77.5 | -1.83 |
| C5 | 101.6 | 138.69 | 89.98 | 74.4 | 80.2 | -1.27 |
| G6 | - | 133.1 | 89.29 | 73.7 | 76.9 | -0.93 |
| G7 | - | 133.8 | 90.01 | 74.6 | 78.4 | -1 |
| U8 | 101.4 | 138.6 | 90.36 | 74 | 79.1 | -1.3 |
| G9 | - | 134.45 | 89.21 | 73.8 | 75.7 | -1.74 |
| U10 | 100.8 | 138.1 | 88.56 | 75 | 79 | -1.31 |

**Table 3.4:** Carbon and phosphorus chemical shift assignments. The carbon resonances were referenced to an external standard and the phosphorus resonances were referenced to phosphate buffer at pH 6.

### 3.3.7 $^{15}$N HSQC

The nitrogen HSQC allowed for the assignment of the imino proton peaks since they could not be determined by proton chemical shift alone. The nitrogen chemical shift helped distinguish the uracil imino protons from the guanine imino protons. The spectrum confirms that the imino proton from base G6 resonates the most downfield at 12.41 ppm. The two uracil imino protons form U4 and U8 resonate at 12.17 and 11.86, respectively. Then G7 and G3 resonate at 11.5 and 11.13 ppm, respectively. This information can be seen in figure 3.12.

**Figure 3.12: $^{15}$N Imino HSQC Spectrum**

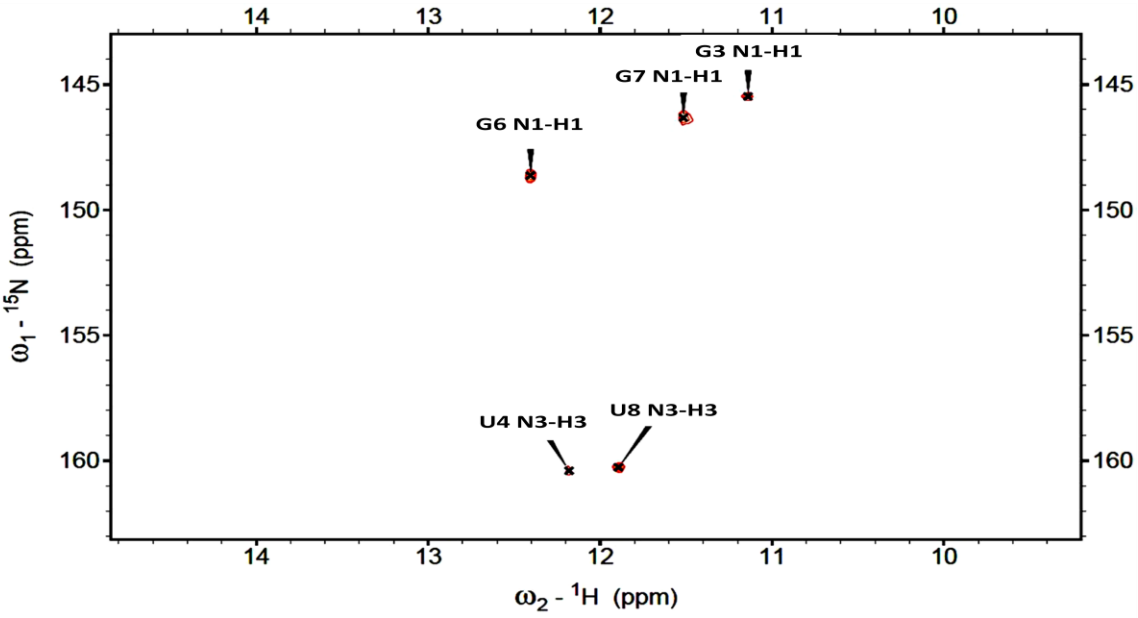



**Figure 3.12:** $^{15}$N Imino HSQC Spectrum showing the imino protons that are shielded from exchange with water. This suggests that these five imino protons are most likely in a stable base pair conformation. Spectrum was taken at 1$^{o}$C in 50 mM NaCl, 10 mM $NaPO_4$, and 0.5 mM $Na_2EDTA$ in $H_2O$ with 10% $D_2O$.

## 3.4 Discussion

### 3.4.1 Structural Features of the 5'r(GUGUCGGUGU)$_2$ duplex.

A complete NOESY sequential walk for all bases in the duplex can be performed. This shows that the base stacking is in a manner similar to that of a normal A-form helix. Four G-U base pairs involving bases 3, 4, 7 and 8 are confirmed by the imino NOEs in the $H_2O$ NOESY spectrum (Figures 3.2, 3.3). However, the orientation of the nucleotides was not determined until the simulated annealing was performed. The two most common G-U base pairs are the wobble and reverse wobble orientations. Both of these orientations would satisfy the NOEs observed in the $H_2O$ NOESY spectrum. In the simulated annealing, the bases were left unrestrained and were allowed to sample any conformation. All of the bases were found to be in a wobble orientation without restraints so they were then restrained to meet this requirement in order to improve convergence. The four terminal bases, 1, 2, 9, and 10 did not show any evidence of base pairing since their respective imino protons were not observed in the $H_2O$ NOESY spectrum. Due to the complete NOESY walk observed in the $D_2O$ NOESY spectrum, these bases most likely stacked on the neighboring G-U pair. In the final model, the two central G-C pairs form normal cis-W-C base pairs with A-form backbone dihedral angles. The G-U pairs formed by bases 3, 4, 7, and 8 are in cis-wobble base pairs using the W-C faces of the bases and have near A-form backbone orientations as well. The four terminal bases 1, 2, 9, and 10 stacked on the end of the helix in an orientation that a NOESY could be observed. The sugar pucker of the four terminal bases 1, 2, 9, and 10 were

shown to have partial C2'-endo conformation as seen in the COSY spectrum (Figure 3.6). Figure 3.8 shows the two most common sugar conformations assumed by RNA ribose molecules. These bases were restrained to meet this C2'-endo after simulations restraining the bases to be C2'-endo, C3'-endo, or unrestrained were completed. These simulations showed that none of the constraints were violated when all of the bases were restrained to be either C2'-endo or unrestrained. The simulations that were performed with all bases restrained to be C3'-endo could satisfy all NOE restraints but, either, the *beta* or *epsilon* (Figure 3.7) dihedral angle restraint in base G3 or G9 could not. Upon review of the unrestrained structural ensemble, the terminal bases G1 and U10 were found to have C2'-endo conformation in 8 of 10 structures and bases U2 and G9 were C2'-endo in 6 of 10 structures. The final model is therefore consistent with all of the NMR data. Figure 3.13 shows two images of the 5'r(**GUGU**CG**GUGU**)$_2$ solution structure.

**Figure 3.13: Image of 5'r(GUGUCGGUGU)$_2$ Duplex**

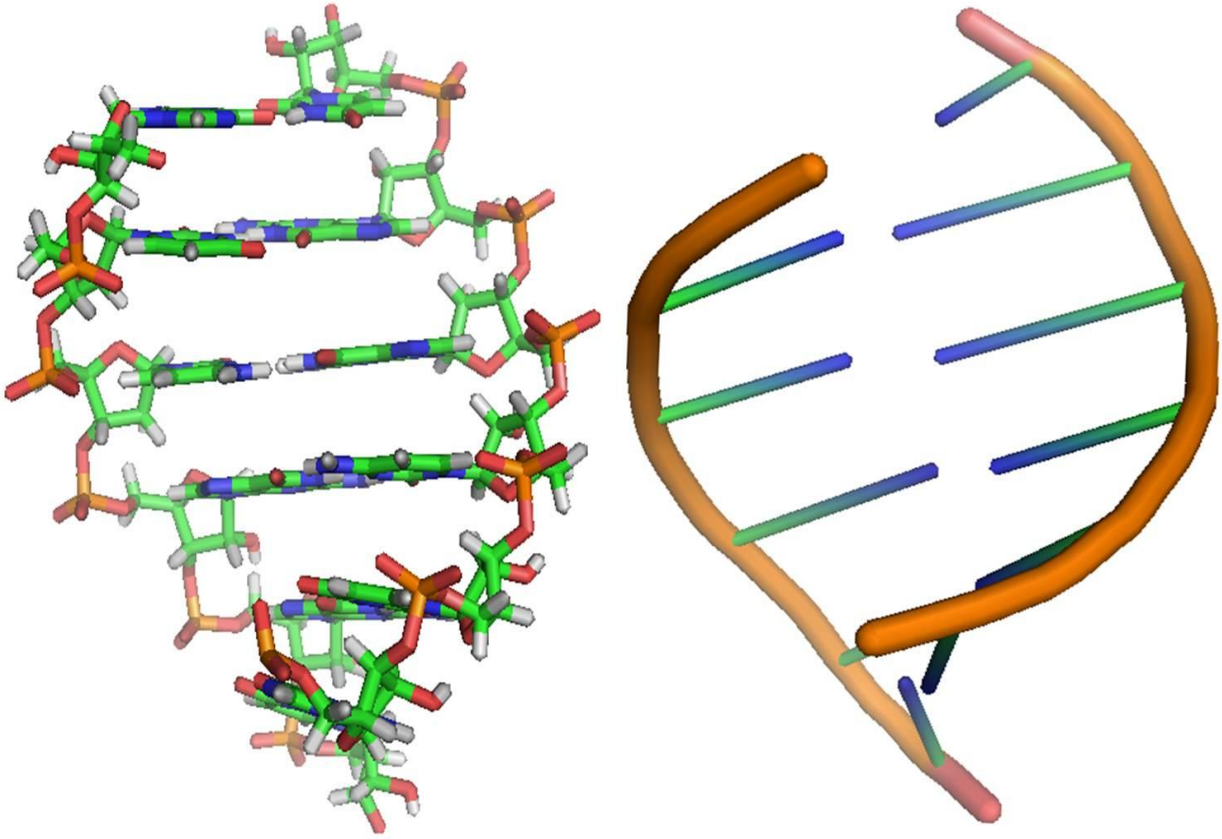


**Figure 3.13:** Image of 5'r(**GUGU**CG**GUGU**)$_2$ duplex (A) with the backbone shown in orange cartoon orientation (B).

**Table 3.5: Exchangeable Proton NOEs and Convergence of Base Pairs**

| Base Pair | Intrastrand | Interstrand | R.M.S.D. |
|---|---|---|---|
| G1-U20 | 0 | 0 | 5.435 |
| U2-G19 | 0 | 0 | 3.794 |
| G3-U18 | 3 | 3 | 2.066 |
| U4-G17 | 4 | 3 | 1.557 |
| C5-G16 | 2 | 4 | 1.343 |
| G6-C15 | 2 | 4 | 1.226 |
| G7-U14 | 4 | 3 | 1.049 |
| U8-G13 | 3 | 3 | 1.035 |
| G9-U12 | 0 | 0 | 1.967 |
| U10-G11 | 0 | 0 | 3.106 |

**Table 3.5:** The exchangeable protons help define the pairing orientation as well as the stacking. The intrastrand NOEs are the NOEs in the same strand and the interstrand NOEs are NOEs between bases on different strands. The number of NOEs closely correlates to the R.M.S.D. which shows how well defined the three dimensional location of each base pair is.

### 3.4.2 Quantitative Analysis of G-U Pairs in Various Helices.

The helical parameters for each base were determined using X3DNA (25). The helical data from crystal and NMR structures containing all combinations of single and double G-U pairs that have been solved were analyzed. The data includes parameters such as dihedral angles, helical rise, propeller twisting, and buckling as well as various others. Upon scouring the data, it became apparent that two parameters, the lambda angle and helical twist, varied greatly when comparing canonical base pairs to G-U pairs. The lambda angle refers to the virtual angle between C1'-pyrimidine-N1 or C1'-purine-N9 glycosidic bonds and the base-pair C1'-C1' line as shown in Figure 3.14. For canonical base pairs, the lambda angle is nearly equal at $55^{o} \pm 5^{o}$ and is not dependent on the base. In a G-U wobble base pair, the lambda angle for the two bases differs greatly. The guanine has a lambda angle of nearly $45^{o} \pm 6^{o}$ while the uracil has a larger lambda angle of $69^{o} \pm 6^{o}$ in order to compensate for the small guanine lambda angle. This variation in lambda for internal base pairs can be directly related to an alteration of the helical twist. Helical twist refers to the amount in degrees that the phosphate backbone rotates upon helix propagation. The helical twist per base for a canonical A-form duplex is $33^{o} \pm 3^{o}$. Unlike canonical base pairs, the helical twist in G-U base pairs is dependent on the lambda angle. When a uracil, with its large lambda angle, is in the 5' strand (lambda (I) in Table 3.5) the base pair step will have a helical twist of $42^{o} \pm 4^{o}$. If the uracil is in the complimentary strand then the following step will have the same altered degree helical twist. In most cases, a compensation for the larger degree of helical twist is made either

in the previous base pair step, the following step or both. In these particular compensation steps, the helical twist is $25^{o} \pm 4^{o}$ which is considerably lower than that of canonical base steps. This observation holds true for all single and double internal G-U pairs.

The aforementioned differences in helical twist cause a disruption to the helix and could be the reason that G-U pairs are unfavorable when in the center of a helix. The compensation steps that involve a smaller degree of helical twist seem to be needed in order to match the average amount of helical twist that is obtained from canonical base pairs. For instance, if a helix were propagated by two canonical base pairs, on average, the helix would turn $66^{o}$. The total amount of helical twist would be the same if the helix was propagated by a normal G-U pair followed by a compensation step ($67^{o}$). In contrast, consecutive terminal G-U base pairs have the ability to freely adopt the most favorable conformation and not return to normal A-form orientation. These results may be the explanation of why consecutive terminal G-U pairs add stability to the duplex and internal G-U pairs are not energetically favorable.

**Figure 3.14: Lambda Angle**

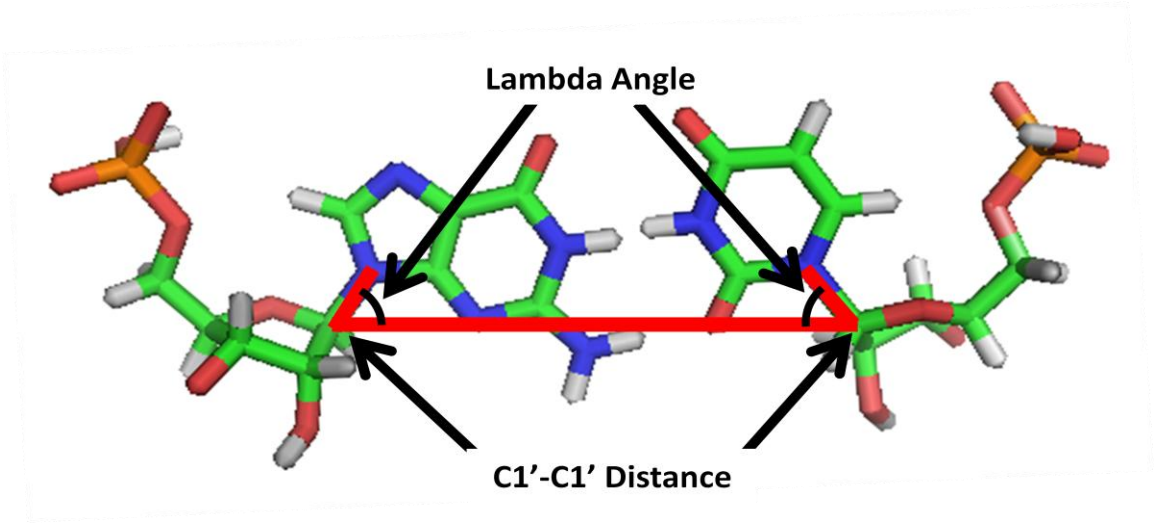



**Figure 3.14:** The lambda angle is a virtual angle between C1'-pyrimidine-N1 or C1'-purine-N9 glycosidic bonds and the base-pair C1'-C1' line. It is represented by the black arc between the glycosidic bond and the virtual C1’-C1’ line which are shown in red.

### 3.4.3 Comparison of the Solution Structure to MC-SYM Predictions.

The three-dimensional structure of the sequence 5'r(**GUGU**CG**GUGU**)$_2$ was predicted using the online MC-SYM tertiary structure prediction program (http://www.major.iric.ca/MC-Sym/). The structures were scored using the AMBER '99 force-field (*26*) and sorted based on lowest free energy. The prediction shows that the structure will form a near A-form duplex. The base pairs in the structure consist of normal Watson-Crick G-C base pairs for the central core and all other base pairs forming cis-wobble G-U base pairs. This structure does not violate any NMR data except that the terminal bases are all in C3'-endo conformation. Even so, the solution and predicted models are still in very good agreement and only have a 2.504 Å all-atom rmsd. For bases 3-8, the all atom RMSD is only 1.38 Å. Figure 3.15 is a cartoon overlay of the backbones of the two models showing the subtle differences in the backbone and suggesting that the most of the difference in the structures are a result of the less defined terminal bases. The MC-SYM structure also does not violate any NOEs or any other experimental data.

Upon performing further helical analysis, differences in helix rotation can be seen between the solution structure and the MC-SYM prediction. The lambda angles between the two structures are in agreement but the helical twist differs significantly. The helical twist for the consecutive terminal G-U pairs in the solution structure does not follow the same trend as the single and double internal base pair. Though the uracil still has a large lambda angle, the terminal G-U pairs generally have a larger degree of helical twist and do not have

compensation steps. The MC-SYM prediction has helical twist parameters similar to those of the internal G-U pairs. MC-SYM predicts that the duplex will have a large helical twist with compensation steps that are dependent on lambda. These errors stem from the lack of known structures that contain consecutive terminal base pairs. Since there is a difference in the helical parameters for internal and terminal base pairs, the error in the prediction is not surprising. Additional studies involving consecutive terminal G-U pairs should be performed to further investigate and verify the results from the helical analysis. As of now, these results give conclusive evidence that the helical parameters of internal and terminal base pairs differ and predictions need to vary accordingly.

**Figure 3.15: Solution and Predicted Structure Backbone Overlay**

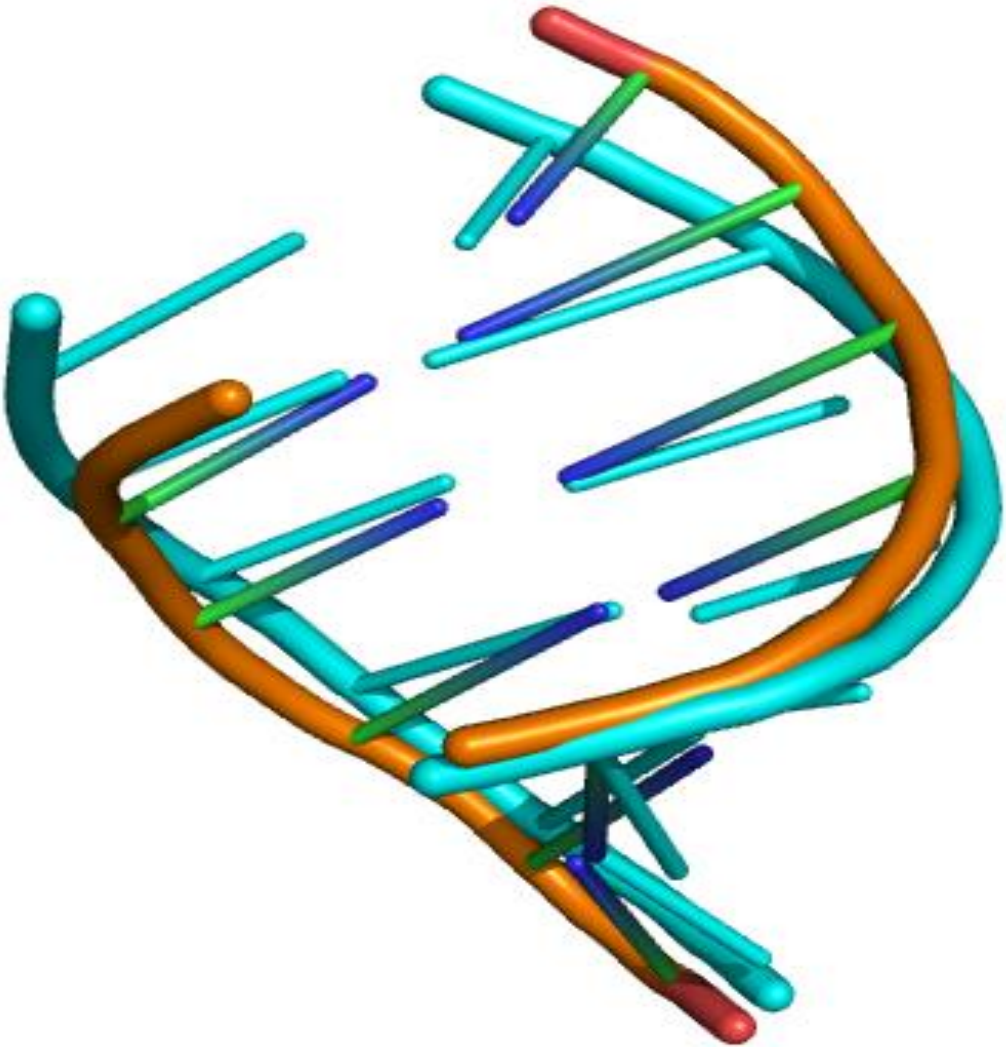


**Figure 3.15:** Backbone overlay for bases 3-8 of the solution structure (orange) and the structure predicted using MC-SYM (teal). The solution structure was calculated using 374 NOEs and 92 dihedral angle restraints. The all-atom rmsd of the two structures was 1.38 Å

## Table 3.5: Helical Parameters of G-U Base Pairs

| **CTGUP** | Base Pair | lambda(I) | lambda(II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-U | 41.5 | 71.8 | | | |
| | 2 U-G | 64.3 | 43 | | 1 GU/GU | 43.19 |
| | 3 C-G | 56.2 | 50.7 | | 2 UC/GG | 31.78 |
| | 4 G-C | 50.9 | 57.2 | | 3 CG/CG | 29.1 |
| | 5 G-U | 45.6 | 68.4 | | 4 GG/UC | 32.14 |
| | 6 U-G | 69.6 | 56.9 | | 5 GU/GU | 39.82 |

| **MC-SYM** | Base Pair | lambda(I) | lambda(II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 2 U-G | 64.7 | 40.7 | | 1 GU/GU | 50.18 |
| | 3 G-U | 41.3 | 66.5 | | 2 UG/UG | 15.49 |
| | 4 U-G | 64.7 | 40.6 | | 3 GU/GU | 51.64 |
| | 5 C-G | 51.6 | 54.2 | | 4 UC/GG | 24.51 |
| | 6 G-C | 52.4 | 52.4 | | 5 CG/CG | 33.57 |
| | 7 G-U | 41.4 | 65.6 | | 6 GG/UC | 22.54 |
| | 8 U-G | 64.7 | 40.7 | | 7 GU/GU | 51.33 |
| | 9 G-U | 41.3 | 66.5 | | 8 UG/UG | 15.65 |

| **2PXB** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 55.2 | 56.5 | | | |
| | 2 G-C | 52 | 56.7 | | 1 GG/CC | 36.42 |
| | 3 U-G | 70.2 | 43 | | 2 GU/GC | 46.32 |
| | 4 G-U | 55 | 55.4 | | 3 UG/UG | 25.19 |
| | 5 C-G | 55.6 | 54.6 | | 4 GC/GU | 33.96 |
| | 6 U-A | 53.2 | 53.9 | | 5 CU/AG | 34.05 |

| **2PXD** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 57.5 | 46.9 | | | |
| | 2 G-C | 53.3 | 52.8 | | 1 GG/CC | 31.25 |
| | 3 G-U | 43.5 | 63.9 | | 2 GG/UC | 31.39 |
| | 4 G-U | 46.2 | 61 | | 3 GG/UU | 37.29 |
| | 5 C-G | 59.1 | 52.4 | | 4 GC/GU | 40.13 |
| | 6 U-A | 55.3 | 51.7 | | 5 CU/AG | 32.33 |

| **2PXE** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 56.7 | 50.8 | | | |
| | 2 G-C | 52.2 | 56.4 | | 1 GG/CC | 35.97 |
| | 3 G-U | 42.7 | 66.6 | | 2 GG/UC | 29.77 |
| | 4 U-G | 67.2 | 43.5 | | 3 GU/GU | 47.65 |
| | 5 C-G | 54.4 | 55.3 | | 4 UC/GG | 27.11 |
| | 6 U-A | 53.7 | 51.8 | | 5 CU/AG | 33.68 |

| **2PXF** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 54.1 | 55.5 | | | |
| | 2 G-C | 53.2 | 53 | | 1 GG/CC | 37.85 |
| | 3 G-U | 40 | 72.3 | | 2 GG/UC | 24.46 |
| | 4 G-U | 45.3 | 64 | | 3 GG/UU | 38.83 |
| | 5 G-C | 55.4 | 56 | | 4 GG/CU | 37.02 |
| | 6 U-A | 52.5 | 51.6 | | 5 GU/AC | 36.96 |

| **2PXK** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 58.1 | 54.3 | | | |
| | 2 G-C | 46.6 | 60 | | 1 GG/CC | 28.77 |
| | 3 G-U | 44 | 64.6 | | 2 GG/UC | 37.64 |
| | 4 U-G | 68 | 49.5 | | 3 GU/GU | 43.84 |
| | 5 G-C | 55.4 | 63.9 | | 4 UG/CG | 23.45 |
| | 6 U-A | 53.4 | 53 | | 5 GU/AC | 38.71 |

| **2PXL** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-U | 43.9 | 70.3 | | | |
| | 2 G-U | 43.5 | 72.5 | | 1 GG/UU | 34.55 |
| | 3 G-C | 55.6 | 57.6 | | 2 GG/CU | 38.85 |
| | 4 U-A | 57.6 | 52.5 | | 3 GU/AC | 36.11 |
| | 5 G-C | 53.3 | 54.9 | | 4 UG/CA | 26.72 |
| | 6 U-G | 67.5 | 43.1 | | 5 GU/GC | 43.99 |
| | 7 U-G | 68.3 | 42.2 | | 6 UU/GG | 36.85 |
| | 8 U-A | 58 | 56.2 | | 7 UU/AG | 23.96 |

| **2PXP** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 55.1 | 56.1 | | | |
| | 2 G-C | 51 | 57.6 | | 1 GG/CC | 33.42 |
| | 3 G-U | 42.3 | 70 | | 2 GG/UC | 30.95 |
| | 4 G-C | 54.8 | 51.6 | | 3 GG/CU | 41.29 |
| | 5 C-G | 59.5 | 51.2 | | 4 GC/GC | 33.41 |
| | 6 U-A | 54.4 | 53.7 | | 5 CU/AG | 33.5 |

| **2PXQ** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 56.8 | 57.8 | | | |
| | 2 G-C | 48.9 | 56.2 | | 1 GG/CC | 36.32 |
| | 3 G-U | 40.3 | 69.8 | | 2 GG/UC | 28.97 |
| | 4 C-G | 57.8 | 56.3 | | 3 GC/GU | 44.91 |
| | 5 C-G | 59.1 | 50.2 | | 4 CC/GG | 36.56 |
| | 6 U-A | 53.8 | 56.7 | | 5 CU/AG | 29 |

| **2PXT** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 51.7 | 52.9 | | | |
| | 2 C-G | 56.1 | 52.9 | | 1 GC/GC | 38.77 |
| | 3 U-G | 68.8 | 48.7 | | 2 CU/GG | 41.19 |
| | 4 G-C | 53.2 | 55.6 | | 3 UG/CG | 27.41 |
| | 5 C-G | 55.6 | 52.9 | | 4 GC/GC | 33.56 |
| | 6 U-A | 54.2 | 53.6 | | 5 CU/AG | 35.48 |

| **2PXU** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 49.8 | 59.7 | | | |
| | 2 G-C | 51 | 53 | | 1 GG/CC | 38.75 |
| | 3 U-G | 70.6 | 45.8 | | 2 GU/GC | 44.34 |
| | 4 G-C | 53.7 | 54.4 | | 3 UG/CG | 26.77 |
| | 5 C-G | 58.6 | 49.5 | | 4 GC/GC | 36.15 |
| | 6 U-A | 54.1 | 54 | | 5 CU/AG | 32.73 |

| **2PXV** | Base Pair | lambda (I) | lambda (II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 52.1 | 57.2 | | | |
| | 2 G-C | 49.9 | 53.2 | | 1 GG/CC | 37.47 |
| | 3 U-G | 66.9 | 38.8 | | 2 GU/GC | 44.74 |
| | 4 G-U | 40.8 | 69.7 | | 3 UG/UG | 18.06 |
| | 5 G-C | 51.5 | 62 | | 4 GG/CU | 34.35 |
| | 6 U-A | 58.1 | 54 | | 5 GU/AC | 44.16 |

| **1IKD** | Base Pair | lambda(I) | lambda(II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 53.7 | 56.2 | | | |
| | 2 G-C | 52.6 | 57.5 | | 1 GG/CC | 33.59 |
| | 3 G-U | 45.3 | 70.3 | | 2 GG/UC | 27.52 |
| | 4 G-C | 51.5 | 56.9 | | 3 GG/CU | 54.71 |
| | 5 C-G | 57.3 | 51.9 | | 4 GC/GC | 31.59 |
| | 6 U-A | 54.4 | 58.7 | | 5 CU/AG | 29.15 |
| | 7 C-G | 59.1 | 49.7 | | 6 UC/GA | 34.77 |
| | 8 U+G | 108.1 | 11.7 | | 7 CU/GG | 89.13 |

| **1EKD** | Base Pair | lambda(I) | lambda(II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 52.4 | 56.4 | | | |
| | 2 G-C | 54.4 | 56.4 | | 1 GG/CC | 30.48 |
| | 3 C-G | 58.8 | 52.7 | | 2 GC/GC | 33.78 |
| | 4 G-U | 58.2 | 87 | | 3 CG/UG | 22.24 |
| | 5 U-G | 87.3 | 58.1 | | 4 GU/GU | 46.1 |
| | 6 G-C | 53 | 58.7 | | 5 UG/CG | 22.3 |
| | 7 C-G | 56.7 | 54.4 | | 6 GC/GC | 33.69 |
| | 8 C-G | 56.4 | 52.6 | | 7 CC/GG | 30.05 |

| **1EKA** | Base Pair | lambda(I) | lambda(II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 1 G-C | 51.7 | 56.9 | | | |
| | 2 A-U | 57.6 | 53.9 | | 1 GA/UC | 32.73 |
| | 3 G-C | 54.2 | 55.7 | | 2 AG/CU | 29.1 |
| | 4 U-G | 71.5 | 47.4 | | 3 GU/GC | 38.59 |
| | 5 G-U | 46.7 | 70.3 | | 4 UG/UG | 15.71 |
| | 6 C-G | 55.7 | 54.3 | | 5 GC/GU | 38.85 |
| | 7 U-A | 53.7 | 57.9 | | 6 CU/AG | 29.16 |
| | 8 C-G | 58.1 | 52.4 | | 7 UC/GA | 32.61 |

| **2D4R** | Base Pair | lambda(I) | lambda(II) | | Step | Twist |
|---|---|---|---|---|---|---|
| | 4 A-C | 44.5 | 74.6 | | 3 GA/CC | 25.15 |
| | 5 C-A | 53.1 | 71.6 | | 4 AC/AC | 35.34 |
| | 6 C-G | 59.5 | 53.6 | | 5 CC/GA | 43.18 |
| | 7 U-G | 72.1 | 42.1 | | 6 CU/GG | 39.25 |
| | 8 C-G | 51.5 | 56.4 | | 7 UC/GG | 18.8 |
| | 9 C-G | 53 | 54.1 | | 8 CC/GG | 32.2 |
| | 10C-G | 53.3 | 56.1 | | 9 CC/GG | 30.89 |
| | 11G-C | 55.5 | 54.5 | | 10CG/CG | 34.93 |

**Table 3.5:** Helical parameters for various helices containing G-U base pairs. The G-U pairs are highlighted in yellow. The pattern of helical twist can be followed by the location of the uracil, large lambda angle. A uracil or large lambda angle in the lead strand (lambda I) leads to a large amount of helical twist in the same base pair step. A large lambda angle in the complimentary strand (lambda II) leads to a large helical twist in the following base pair step. CTGUP is the data from solution structure of the consecutive terminal G-U pair (CTGUP) duplex 5'r(**GUGU**CG**GUGU**). MC-SYM is the prediction of the consecutive terminal G-U pair (CTGUP) duplex. Step refers to the 5'-3' base connectivity. Twist is the amount, in degrees, each base pair turns per step. 1EKA, 1EKD (13), 1IKD (27), 2D4R (28), 2PXB, 2PXD, 2PXE, 2PXF, 2PXK, 2PXL, 2PXP, 2PXQ, 2PXT, 2PXU, 2PXV (29)

## 3.5 Conclusions.

This research presents the NMR structure of the duplex 5' r(**GUGU**CG**GUGU**)$_2$ and compares it to the structure predicted by the tertiary structure prediction program MC-SYM. This solution structure is the first of its kind in that there have been no NMR structures solved with consecutive terminal G-U base pairs. Even with the majority of the structure comprised of non-canonical base pairs, the structure surprisingly still forms a near A-form duplex. Though the atom placement is nearly identical between the solution structure and the MC-SYM prediction, the structures differ in helical rotation by an average of $18^{o}$ per base pair step. These errors show the need for solving the tertiary structures of similar duplexes.

This work also does a quantitative analysis of internal G-U pairs as opposed to consecutive terminal pairs and suggests reasons for the differences in energetics between the two types. As in most cases, this research gives rise to other questions. One of these questions is the applicability of these results to other wobble base pairs. While new questions are proposed, new information has also been obtained. These results provide benchmarks for studying as well as predicting consecutive terminal G-U base pairs and aids in the diversification of the database of known RNA structures.

1. Crick, F. (1966) Codon--anticodon pairing: the wobble hypothesis, *Journal of Molecular Biology 19*, 548-555.
2. Cech, T. R. (1990) Self-splicing of group I introns, *Annual review of biochemistry 59*, 543-568.
3. Strobel, S. A., and Cech, T. R. (1995) Minor groove recognition of the conserved GU pair at the Tetrahymena ribozyme reaction site, *Science 267*, 675.
4. Michel, F., and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis, *Journal of Molecular Biology 216*, 585-610.
5. Hou, Y. M., and Schimmel, P. (1988) A simple structural feature is a major determinant of the identity of a transfer RNA.
6. McClain, W. H., and Foss, K. (1988) Changing the identity of a tRNA by introducing a GU wobble pair near the 3'acceptor end, *Science 240*, 793.
7. Stern, S., Weiser, B., and Noller, H. F. (1988) Model for the three-dimensional folding of 16 S ribosomal RNA* 1, *Journal of Molecular Biology 204*, 447-481.
8. Gutell, R. R. (1993) Collection of small subunit (16S-and 16S-like) ribosomal RNA structures, *Nucleic Acids Research 21*, 3051.
9. Nagaswamy, U., Voss, N., Zhang, Z., and Fox, G. E. (2000) Database of non-canonical base pairs found in known RNA structures, *Nucleic Acids Research 28*, 375.
10. McDowell, J. A., and Turner, D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC) 2 by Two-Dimensional NMR and Simulated Annealing†,‡, *Biochemistry 35*, 14077-14089.
11. Gutell, R. R. (1994) Collection of small subunit (16S-and 16S-like) ribosomal RNA structures: 1994, *Nucleic Acids Research 22*, 3502.
12. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helixes, *Biochemistry 25*, 5755-5759.
13. Chen, X., McDowell, J., Kierzek, R., Krugh, T., and Turner, D. (2000) Nuclear Magnetic Resonance Spectroscopy and Molecular Modeling Reveal That Different Hydrogen Bonding Patterns Are Possible for GU Pairs: One Hydrogen Bond for Each GU Pair in r (GGCGUGCC) 2 and Two for Each GU Pair in r (GAGUGCUC) 2†,‡, *Biochemistry 39*, 8970-8982.
14. Nguyen, M. T., and Schroeder, S. J. (2010) Consecutive Terminal GU Pairs Stabilize RNA Helices, *Biochemistry*.
15. Victor A. Bloomfield., D. M. C., Ignacio Tinoco (2000) *Nucleic Acids Structures, Properties and Functions*.
16. Jane S. Richardson, B. S., Laura W. Murray, et al. (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution), *RNA 14*, 465-481.
17. Blanton S. Tolbert, S. D. K., Susan J. Schroeder, Thomas R. Krugh, Douglas H. Turner. (2006) NMR Structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: Probing the Structural Features That Shape the Thermodynamic Stability of GA Pairs, *Biochemistry*.
18. Kneller, T. D. G. a. D. G. SPARKY 3.
19. Brunger, A. T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kuntsleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998) Crystallography and NMR System: A new software suite for macromolecular structure determination, *Acta Cryst. D54*, 905-921.

20. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, *DeLano Scientific*.
21. Lukavsky, P. (2007) Basic Principles of RNA NMR Spectroscopy, In *Structure and Biophysics – New Technologies for Current Challenges in Biology and Beyond*, Springer Netherlands.
22. (1993) *NMR of macromolecules: a practical approach*, Oxford University Press, Oxford, NY.
23. Schroeder, S. J., Turner, D.H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, *Biochemistry 39*, 9257-9274.
24. Santa Lucia, J., Jr., Kierzek, R., 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.
25. Wuthrich, K. ( 1986) *NMR of Proteins and Nucleic Acids*, New York.
26. Case, D., Pearlman, D., Caldwell, J., Cheatham III, T., Ross, W., Simmerling, C., Darden, T., Merz, K., Stanton, R., and Cheng, A. (1999) AMBER 99, *University of California, San Francisco*.
27. Ramos, A., and Varani, G. (1997) Structure of the acceptor stem of Escherichia coli tRNAAla: role of the G3· U70 base pair in synthetase recognition, *Nucleic Acids Research 25*, 2083.
28. Nakabayashi, M., Shibata, N., Komori, H., Ueda, Y., Iino, H., Ebihara, A., Kuramitsu, S., and Higuchi, Y. (2005) Structure of a conserved hypothetical protein, TTHA0849 from Thermus thermophilus HB8, at 2.4 A resolution: a putative member of the StAR-related lipid-transfer (START) domain superfamily, *Acta Crystallographica Section F: Structural Biology and Crystallization Communications 61*, 1027-1031.
29. Keel, A. Y., Rambo, R. P., Batey, R. T., and Kieft, J. S. (2007) A general strategy to solve the phase problem in RNA crystallography, *Structure 15*, 761-772.