# Investigation of RNA Hairpin Loop Folding with Time-Resolved Infrared Spectroscopy

A Dissertation

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

with a

Major in Chemistry

in the

College of Graduate Studies

University of Idaho

by

Aaron Lee Stancik

January 2012

Major Professor: Eric B. Brauns, Ph.D.

# AUTHORIZATION TO SUBMIT DISSERTATION

This dissertation of Aaron Lee Stancik, submitted for the degree of Doctor of Philosophy with a major in Chemistry and titled "Investigation of RNA Hairpin Loop Folding with Time-Resolved Infrared Spectroscopy," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor _____ Date _____
Eric B. Brauns

Committee Members _____Date _____
Ray von Wandruszka

_____ Date _____
W. Dan Edwards

_____ Date _____
F. Marty Ytreberg

Departmental Administrator _____ Date _____
Ray von Wandruszka

College Dean _____ Date _____
Scott Wood

Final Approval and Acceptance by the College of Graduate Studies

_____ Date _____
Jie Chen

## Abstract

Ribonucleic acids (RNAs) are a group of functional biopolymers central to the molecular underpinnings of life. To complete the many processes they mediate, RNAs must fold into precise three-dimensional structures. Hairpin loops are the most ubiquitous and basic structural elements present in all folded RNAs, and are the foundation upon which all complex tertiary structures are built. A hairpin loop forms when a single stranded RNA molecule folds back on itself creating a helical stem of paired bases capped by a loop. This work investigates the formation of UNCG hairpin loops with the sequence 5'-GC(UNCG)GC-3' (N = A, U, G, or C) using both equilibrium infrared (IR) and time-resolved IR spectroscopy. Equilibrium IR melting data were used to determine thermodynamic parameters. Melting temperatures ranged from 50 to 60°C, and enthalpies of unfolding were on the order of 100 kJ/mol. In the time-resolved work, temperature jumps of up to 20°C at 2.5°C increments were obtained with transient relaxation kinetics spanning nanoseconds to hundreds of microseconds. The relaxation kinetics for all of the oligomers studied were fit to first or second order exponentials. Multiple vibrational transitions were probed on each oligomer for fully folded and partially denatured structures. In the time-resolved limit, in contrast to equilibrium melting, RNA does not fold according to two-state behavior. These results are some of the first to show that RNA hairpins fold according to a rugged energy landscape, which contradicts their relatively simple nature. In addition, this work has proven that time-resolved IR spectroscopy is a powerful and novel tool for investigating the earliest events of RNA folding, the formation of the hairpin loop.

## Acknowledgements

A week before my father died he called me on the phone and gave me his best and final advice. He told me that no matter what happened to him, I should go all the way in college and get a doctorate degree. I was in my last undergraduate semesters when he passed. It was very important to him that I finish school because nobody in my family had attended college. My mother is a generous, selfless woman who managed to raise three kids by herself, despite many hardships. She always told me that anything in life is possible if you really put your mind to it. Without her I would have never graduated from high school, let alone higher education.

I owe many thanks to my wife who encouraged me to enroll in college and who sat through numerous chemistry courses by my side. She was great academic competition, lively conversation, and a good editor. For years she has supported my academic effort and given me companionship and love. Likewise, I want to thank my son for the great times we've had, and for all he has put up with during my college years. My wife and son have been my greatest motivation and foundation through this academic journey.

James Fromm, my high school chemistry teacher, coerced me into taking his course, turned me on to chemistry, and helped shape my career. He was my first, unofficial academic advisor. A handful of college professors stand out in my mind as exceptional: Donhong Zhang, Warren Hayman, Chein M. Wai, Peter R. Griffiths, W.D. Edwards, Ray von Wandruzska, and of course my patient research advisor, and mentor, Eric B. Brauns.

This work is dedicated to George A. Stancik Jr. and Dale D. Dinger

*"The roots of an education are bitter, but the fruit is sweet"- Aristotle*

*"Watch out for coyotes"- G.A.S*

## Table of contents

**Authorization to submit dissertation** **ii**

**Abstract** **iii**

**Acknowledgements** **iv**

**Dedications** **v**

**Table of contents** **vi**

**List of figures** **ix**

**List of tables** **xi**

**Chapter 1**

**Ribonucleic acids** **1**

*RNA* *2*

*The RNA folding problem* *3*

*RNA structure* *4*

Base pairing and base stacking—the building blocks of RNA structure 6

RNA structural hierarchy 8

Folding thermodynamics 10

Hairpin folding kinetics—what is known 11

*The folding landscape* *12*

*Research objectives* *13*

*Organization of dissertation* *15*

*Bibliography* *16*

**Chapter 2**

**Experimental foundations** **19**

*Introduction* *20*

*Infrared spectroscopy of RNA* *20*

*Equilibrium IR spectroscopy* *24*

*Time-resolved IR spectroscopy* *26*

*Oligonucleotides* *29*

*Bibliography* *30*

**Chapter 3**

**A simple asymmetric lineshape for fitting infrared absorption spectra** **31**

*Abstract* *32*

*Introduction* *32*

*Experimental* *33*

*Lineshape models* *34*

Asymmetric variation of the Lorentzian and Gaussian lineshapes 36

*Fitting experimental spectra* *38*

*Conclusion* *44*
*Bibliography* *45*

**Chapter 4**

**Investigating the thermodynamics of UNCG tetraloops using infrared spectroscopy** **46**
*Introduction* *47*
*Results* *47*
Equilibrium IR spectra 47
*UNCG tetraloop thermodynamics* *53*
*Discussion* *57*
*Conclusion* *60*
*Experimental* *60*
*Bibliography* *61*

**Chapter 5**

**Investigating the complex folding dynamics of UNCG hairpin loops with time-resolved IR spectroscopy** **62**
*Chapter overview* *63*
*Introduction* *63*
*Experimental approach* *65*
Data handling 66
*Results* *70*
*Discussion and conclusions* *78*
*Bibliography* *79*

**Chapter 6**

**Rearrangement of partially ordered stacked conformations contributes to the rugged energy landscape of a small RNA hairpin** **80**
*Abstract* *81*
*Introduction* *82*
*Experimental procedures* *85*
Sample preparation 85
FTIR spectroscopy 86
Time resolved IR spectroscopy 87
*Results* *89*
Equilibrium FTIR spectroscopy 89
Equilibrium melting 91
Time resolved spectroscopy 94
*Discussion* *98*
*Conclusion* *103*
*Bibliography* *105*

**Appendix A**

**Experimental protocols** **108**
*Nuclease controls* *108*
*Phosphate buffers* *109*
*Sample preparation* *111*
*IR cell construction* *112*
*Cell pathlengths* *114*
*Sample injection* *115*
*Equilibrium IR spectroscopy* *116*
*Time-resolved IR* *117*
Pump beam 117
Probe beam 118
Diode laser characterization 120
Temperature calibration for T-jump experiments 121
Overlapping pump and probe beams 123
Aligning the cell 124

**Appendix B**

**Transient data collection and handling** **126**

**Appendix C**

**Optics** **129**
*Beam waist measurement* *129*
*Raman conversion* *130*
*Off axis parabolic mirrors* *132*

**Copyrights** **133**

## List of figures

**Figure 1.1:** The 16S rRNA subunit and the 70S ribosome of E. Coli 5

**Figure 1.2:** Base pairing interactions 6

**Figure 1.3:** A UNCG hairpin tetraloop 9

**Figure 1.4:** RNA folding funnel energy diagram 13

**Figure 2.1:** UV-visible spectra of RNAs 21

**Figure 2.2:** IR spectra of the ribonucleotide monophosphates 22

**Figure 2.3:** Equilibrium infrared spectra of $tRNA^{phe}$ 23

**Figure 2.4:** Melting curve for $tRNA^{phe}$ at 1660 $cm^{-1}$ 25

**Figure 2.5:** Temperature jump relaxation kinetics 26

**Figure 2.6:** Laser induced temperature jump schematic 27

**Figure 3.1:** An asymmetric Lorentzian 37

**Figure 3.2:** Fitting of AMP IR abs spectra using a symmetric pseudo Voigt model 40

**Figure 3.3:** Fitting of AMP IR abs spectra using an asymmetric pseudo Voigt model 41

**Figure 4.1:** Equilibrium IR melting spectra of 5'-gcUNCGgc-3' RNA hairpins 50

**Figure 4.2:** Thermal difference spectra for the UNCG hairpins 51

**Figure 4.3:** Wavenumber specific IR melting curves 56

**Figure 5.1:** Kinetic relaxation data for 5'-gcUUCGgc-3' 68

**Figure 5.2:** Kinetic relaxation data for 5'-gcUACGgc-3' 69

**Figure 5.3:** Kinetic relaxation data for 5'-gcUGCGgc-3' 70

**Figure 5.4:** Activation plots and normalized amplitudes for UUCG 75

**Figure 5.5:** Activation plots and normalized amplitudes for UACG 76

**Figure 5.6:** Activation plots and normalized amplitudes for UGCG 77

**Figure 6.1:** Equilibrium IR melting spectra for UCCG 91

**Figure 6.2:** Normalized melting curves for UCCG 93

**Figure 6.3:** Temperature jump relaxation kinetics 95

**Figure A.1:** Custom IR cell for FTIR and time-resolved IR spectroscopies 114

**Figure A.2:** Laser induced temperature jump (pump-probe) schematic 118

**Figure A.3:** Monochromator scans of a tunable CW diode laser output 121

**Figure A.4:** Temperature dependent infrared spectra for $H_2O$ and $D_2O$ 123

**Figure A.5:** Transient absorbance data for a well aligned dummy cell 125

**Figure A.6:** Affect of severe laser cavitation on IR transient signals 126

**Figure A.7:** Time-resolved IR data handling 128

**Figure A.8:** Raman conversion data for the first stokes line at 1.9 μm 131

**Figure A.9:** Off-axis parabaloidal reflector 132

## List of tables

**Table 3.1:** Summary of curve fitting results 43

**Table 4.1:** Thermodynamic unfolding parameters for the UNCG hairpin loops 57

**Table 6.1:** Kinetic relaxation parameters for UCCG 98

# Chapter 1

## Ribonucleic acids

**RNA**

Ribonucleic acids (RNAs) are a group of structurally complex and functionally diverse, biological macromolecules that play a key role in numerous biochemical processes. Perhaps the most well known function of RNA is the role it plays in gene expression. This is highlighted in Crick's "central dogma of molecular biology". This states that DNA contains the genetic code, which is transcribed into RNA intermediates, and subsequently translated into proteins (the building materials of life). In essence, RNA is responsible for the transmission of genetic information and is a biopolymer central to the molecular underpinnings of life.

Although its role in gene expression is well known, recent evidence suggests that RNA plays an active role in gene regulation as well [1]. The presence of certain substrates triggers a conformational change in untranslated regions of mRNA that serves to turn on or turn off the translation of that particular gene [2]. These have been termed "riboswitches"[3]. In addition to its role in gene expression and regulation, RNA has been linked to some diseases including Huntington's disease, Fragile X syndrome, and the RNA viruses (HIV, hepatitis, influenza, etc) [4-9].

RNA may also provide unique insight in to the origins of life as we know it. The expression of genes (protein synthesis) described above is a cyclic process where all the steps are interdependent. This raises the question, "which came first: DNA, RNA, or proteins"? Some have theorized that RNA was the first biological macromolecule to emerge from the "primordial soup" billions of years ago [10]. This "RNA world" hypothesis is feasible because RNA can carry genetic information and catalyze its own replication [11, 12]. In short, RNA could have existed independently of DNA or proteins.

In order to complete the many biological processes it mediates, RNA must fold into a very precise three-dimensional structure. This "native structure" consists of many specific folds that are active sites for chemical catalysis, ligand binding, and protein recognition [13-17]. RNA's ability to *function* in these elaborate biochemical processes depends on its ability to *form* precise secondary and tertiary structures. Understanding how RNA folds into its biologically active form is the "RNA folding problem" [18].

**The RNA folding problem**

Why is the study of RNA folding important? In molecular biology function and form are intimately correlated. Accordingly, the precise three-dimensional conformation of an RNA molecule dictates its functionality. In terms of the fully folded, biologically active structures, a great deal is already known. However, relatively little is known about the forces that direct and govern the RNA *as it is folding*. Perhaps a more poignant question is, "*How* does RNA fold into its native configuration?" Answering this question is the ultimate goal of all RNA folding research [18].

The astounding thing about the folding process is that for any particular RNA molecule there is an enormously large manifold of *possible* three-dimensional conformations it can sample. If the structure folded by randomly sampling these potential conformations, it would take a time longer than the age of the universe ($10^{10}$ years) to find the correct structure [19]. In the case of a 76 nucleotide tRNA with 8 configurations per nucleotide, there are $8^{76}$ possible conformations. If it samples each of these structures in a time on the order of a molecular vibration ($10^{-13}$ s), it would take $10^{48}$ years to fold. Since life requires that RNA folding take place on biologically relevant

time scales, it cannot fold via random searching. This is referred to as Levinthal's paradox. In reality, RNA folds on time scales that are orders upon orders of magnitude faster than a random search predicts. For example, small secondary structures require nanoseconds to microseconds to fold while larger RNAs requires microseconds to minutes to fold [18]. This clearly indicates that there exist very specific pathways or mechanisms that guide an RNA molecule from an unfolded state to its folded, biologically active form. Determining these mechanisms is the goal of RNA folding research.

In principle, the instructions for folding are encoded in the primary nucleotide sequence [20]. In practice however, many external variables can contribute to fully realize RNA folding. *In vitro*, the folding problem can be dissected and analyzed in terms of separate contributing variables (i.e. ionic, thermal). *In vivo*, many coupled chemical reactions in various cellular sub compartments serve to direct folding [21, 22]. At present, we (as scientists) are not able to study how RNA folds within an actual living cell. This is still many years from physical observation. At best, we can design an experiment that captures the key features of the folding process and *postulate* details of a folding mechanism in living cells.

**RNA structure**

RNA is a single stranded polymer chain of nucleotide monomers connected by a ribose and phosphate backbone. The ribose and phosphate units are linked by successive 3'-5' phosphodiester bonds. The nucleotide monomers are derivatives of purine and pyrimidine bases; there are four of them: adenine (A), guanine (G), cytosine

(C) and uracil (U). An RNA strand can consist of anywhere from a few to thousands of nucleotide monomers (see **Figure 1.1**). The linear strand can fold into a vast number of conformations. DNA is also a polynucleotide but with two important differences. One is that DNA contains deoxyribose instead of ribose, and the other is that uracil is replaced by thymine (a different nucleotide base). Although RNA and DNA are chemically very similar, these two small distinctions amount to vast differences in structure and subsequent function. The primary role of DNA is to store genetic information. On the other hand, due to its vast structural diversity, RNA has much more varied biological functions.

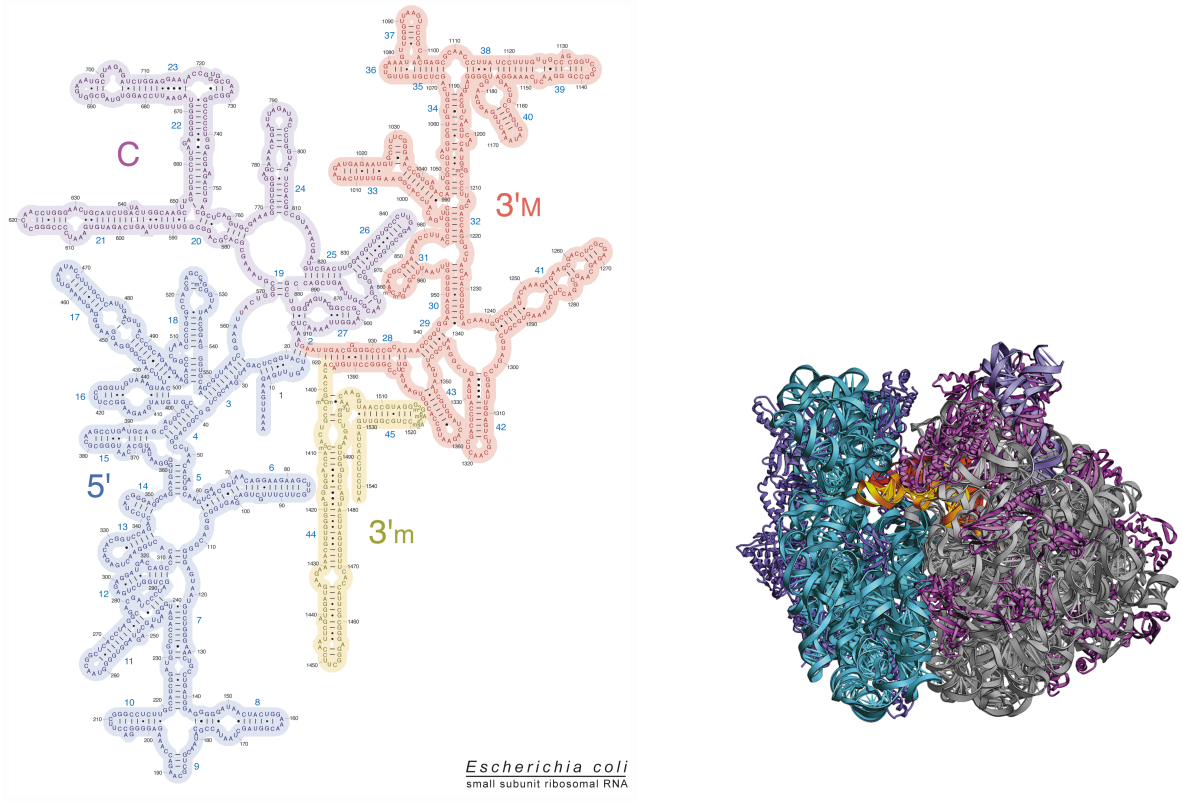


**Figure 1.1: The 16S rRNA subunit and 70S ribosome of E. Coli**

(Left) Two-dimensional helix diagram of the small 16S subunit ribosomal RNA of a much larger RNA-protein complex. (Right) Three-dimensional structure of a fully folded 70S ribosome RNA in its native form. Reprinted with permission from Dr. Harry Noller.

**Base pairing and base stacking—the building blocks of RNA structure**

At the foundation of all RNA secondary structures are base pairing and base stacking interactions. Combined, these interactions provide the energetic driving force for folding. Although both are equally important, base pairing will be discussed first. A base pair forms when two complementary nucleotides (i.e. AU or CG) interact in plane allowing polar functional groups (C=O, C=N, N-H) to interact and form hydrogen bonds (**Figure 1.2**). While an individual hydrogen bond (H-bond) is weak (~20 kJ/mol), AU base pairs form 2 and GC base pairs form 3. In addition, there are usually multiple base pairs in a given RNA structure. Analogous to the way that each subsequent pair of teeth on a zipper makes for a stronger linkage, the cumulative effect of multiple base pairing interactions provides the energetic driving force for RNA folding. Although, non Watson-Crick base pairs, miss-matched base pairs, and wobble base pairs do occur, Watson-Crick base pairs (AU and GC) maximize H-bonding interactions and are more structurally stable.

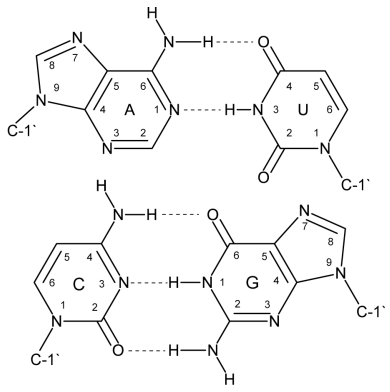


**Figure 1.2: Base pairing interactions**

Hydrogen bonding interactions between the four nucleotide bases stabilizes RNA structure and drives folding. Here Watson-Crick base pairs form between AU and CG base pairs.

In addition to base pairing interactions, there are also non-covalent base stacking interactions between adjacent bases that contribute to the overall structural stability of RNA. These interactions are due to π orbitals that extend above and below the plane of the nucleotide bases. Base stacking of parallel bases aligns these negative π orbitals, which then interact through dispersive forces. Although weaker than the H-bonds in base pairs, these forces play an equally important role in the folding process and in stabilizing RNA structure. In fact, results from Chen et al. have verified that a stable structure cannot form without at least one base stacking interaction [23].

**Role of counterions in RNA structure**

A discussion of RNA structure would not be complete without mentioning the important role that counterions play. At neutral pH, as in biological cells, the phosphate groups along RNA's backbone are predominately in the deprotonated ($PO_4^{2-}$) form and impart a large overall negative charge on RNA. Consequently, RNA structure is highly sensitive to ionic conditions [24-26]. The negatively charged phosphate groups along the entire chain interact with each other repulsively, forcing the chain into an extended unfolded configuration. In order for RNA to fold, a comparable positive charge is needed to shield this repulsive force [27, 28]. Cations such as $Na^+$ and $Mg^{2+}$ serve this function and are essential in the folding process. They initiate folding by diffusing into an RNAs electronegative field [20, 24] thereby shielding the negatively charged phosphates. Once repulsion is buffered, the formation of compact secondary structures can proceed. Accordingly, the earliest stages of folding are driven by counterion condensation.

**RNA structural hierarchy**

RNA structure is divided into three fundamental levels of organization. These are primary, secondary, and tertiary structures [13, 29, 30]. Primary structure simply refers to the sequence of nucleotide bases. Secondary structure consists mainly of local regions of contiguous base pairs (forming helices) that are "capped" by loops or bulges. While tertiary structures are of the result of interactions between distinct secondary elements. Since secondary structures are independently stable (i.e., they are stable in the absence of tertiary interactions), RNA folding tends to proceed hierarchically. In other words, tertiary structures are built from pre-formed secondary structural elements [30, 31]. Because of this, RNA folding can be resolved into distinct early and late events. Early events correspond to rapid collapse of the RNA chain into compact secondary structures that have little or no tertiary structure. In later steps, these collapsed structures rearrange through a slow conformational search to form the native structure [17, 32]. Since the events are necessarily coupled, understanding the early events is critical if we are to understand how RNAs fold into complex, biologically relevant structures. The goal of this project is to investigate the most fundamental of the earliest folding events; the formation of hairpin loops.

A hairpin loop forms when a single strand of RNA folds back on itself creating a loop of unpaired bases and a helical stem comprised of base pairs. Hairpin loops are the most common of all the possible secondary structural elements present in folded RNAs and are the foundation upon which all complex three-dimensional structures are built [29]. They are generally classified according to the number of unpaired bases in the loop and can contain as few as two (diloops) to more than eighteen (octadecaloops).

Tetraloops (4 unpaired bases in the loop) are by far the most common and are the focus of the work presented in this dissertation (**Figure 1.3**).

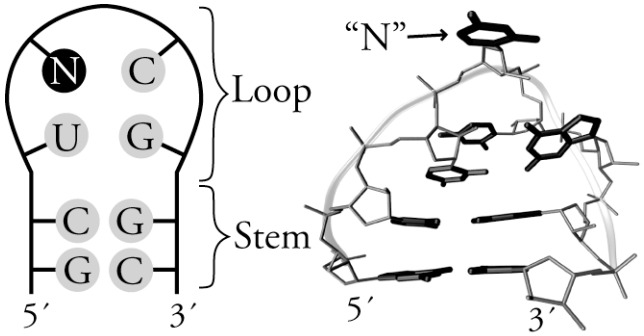


**Figure 1.3: A UNCG hairpin tetraloop**
(Left) Cartoon representation of a UNCG hairpin illustrating the loop and stem segments. (Right) A molecular structure showing the ribose-phosphate backbone as a ribbon and the N position directed out of the loop. In addition, base pairing and stacking interactions are depicted.

Although there are 256 possible sequence combinations of four nucleotides in a tetraloop, the sequence UNCG (N is any nucleotide) occurs at a statistically significantly higher rate than any other. One reason why the UNCG tetraloops are so ubiquitous is that they are known to form exceptionally stable structures suggesting that they serve as nucleation sites for folding of larger, more complex tertiary structures [33, 34]. In nature, most UNCG loops are closed by strong GC base pairs, which further stabilize the loop [23, 35]. For this reason, the primary objective of this work was to investigate the folding kinetics of oligonucleotides with the sequence GC(UNCG)GC. The choice of a two base pair stem is not trivial, these are the smallest RNAs that will form stable hairpins

[36, 37]. Choosing this sequence (i.e., the sequence with the smallest possible stem structure) allows us to focus on how the loop effects the folding kinetics.

The biological significance of the UNCG tetraloop can be inferred from its sequence conservation over many diverse species. Moreover, since folding is hierarchical and the UNCG hairpin tetraloop is the most common of all the secondary structures, one could argue that the formation of tetraloops with this sequence is one of the most (if not *the* most) fundamental steps in the overall folding of all biologically relevant RNAs.

**Folding thermodynamics**

Heating a sample of RNA to its melting temperature will unfold (denature) the ensemble of molecules giving a distribution of folded and unfolded structures. The change in Gibbs free energy ($\Delta G$) gives the ratio of folded to unfolded RNA. At the melting temperature ($\Delta G = 0$) there is a distribution of 50% folded and 50% unfolded structures (This is an oversimplification, there is really an ensemble of partially unfolded structures). Upon cooling, the RNA will fold back to its native conformation, which argues that folding occurs spontaneously ($\Delta G < 0$; below $T_m$). RNA folding is both exothermic and exergonic. As RNA folds energy is released as *heat* to the surroundings, and the internal energy (Gibbs free energy) of the RNA is minimized. In addition, RNA's entropy decreases as it folds, and is at a minimum in the native state ($\Delta S < 0$). This sounds like a *thermodynamic paradox*, until the enthalpic contribution of base pairing interactions is taken into account. H-bonding interactions energetically offset the unfavorable entropy loss, minimizing the Gibbs free energy. Desolvation by

surrounding water molecules also helps thermodynamically push folding, because it increases the entropy of the solution.

An unfolded RNA has much higher configurational entropy and can change conformations with more freedom of motion. The strand fluctuates freely between intrinsically unstructured conformations. Competing hydrogen bonds from solvating water molecules helps stabilize the unfolded chain. In the absence of counterions ("hypothetically speaking") the random coil will stay unfolded. In the presence of buffering counterions, the extended structure collapses allowing the strand to begin folding. These counterions contribute significantly to the thermodynamic stability of RNA secondary and tertiary structures and the thermodynamics of folding [16, 38]. After RNA contracts, intramolecular H-bonds begin forming between opposite bases zipping up the helical stem. Cumulative base pairing interactions provide the thermodynamic driving force for RNA folding to proceed spontaneously. The large enthalpic contribution from forming H-bonds mitigates the unfavorable, negative entropy change for folding and minimizes Gibbs free energy.

**Hairpin folding kinetics—what is known**

Hairpin folding begins with the collapse of the polyanionic strand [27, 39], and the nucleation of the loop. This initial "collapse" and nucleation occurs on very fast time scales (nanoseconds) as the ends of the strand come together. For hairpins with a long stem comprised of multiple base pairs, a number of mis-folded intermediates can form. This limits the time for the formation of hairpin loops, (10s to 100s of microseconds) as the mis-folded stems rearrange in the search for the native state. This is the slow

conformational search that has been well reported in the literature [17]. By choosing small oligomers containing just two closing base pairs, we purposefully minimize the number of conformational degrees of freedom. Now, when the strand collapses the only viable stem structure is the native conformation. A slow search is impossible, as the number of conformations is strictly limited since there is no possibility of mis-folded stem conformations. However, it still takes 10s to 100s of microseconds for the small hairpins to fold [23]. We are interested in *why* folding of hairpin loops takes so long. The motivation of our research is to explain why the folding of hairpin loops is so slow, and how base pairing and base stacking interactions contribute to the overall folding rate. What we intend to show in this study is that the loop contributes to the folding complexity and that the rate limiting step is not solely due to mis-folded stem structures. These results are new to the RNA folding community and represents novel and important information regarding hairpin loop folding and RNA structure in general.

**The folding landscape**

The thermodynamics and kinetics of RNA folding can be illustrated using a funnel shaped free energy diagram called an *energy landscape* (**Figure 1.4**). Energy varies vertically (i.e., it is at a minimum at the bottom of the funnel) while conformational entropy is proportional the width at a particular vertical coordinate (i.e., energy). Accordingly, a folding RNA moves down the funnel to its lowest energy native state. An important characteristic of the energy landscape for RNA is that it consists of many local energy minima—in other words, the landscape is quite "rugged" [20, 40]. These local minima represent subpopulations of mis-folded intermediates (both "on pathway"

and “off pathway”). Mechanistically, these local minima manifest as kinetic traps with barriers that must be overcome for an RNA to proceed to the native state. In the energy landscape picture, different sub-ensembles of unfolded RNA molecules can traverse different paths to the native state.

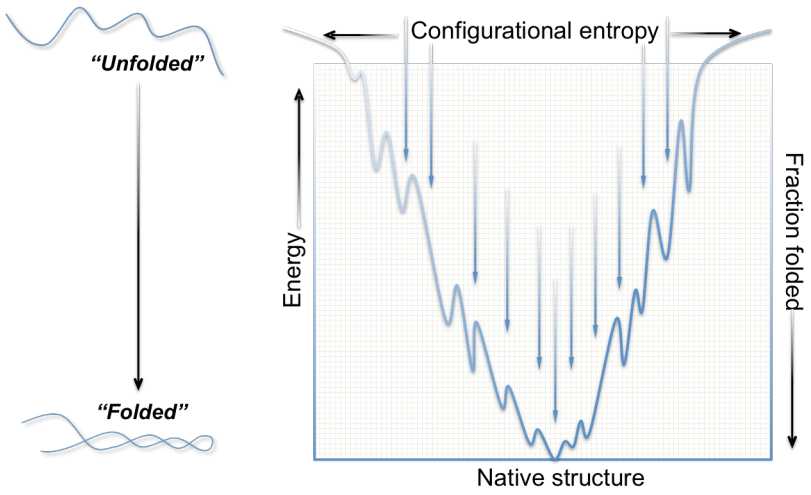


**Figure 1.4: RNA folding funnel energy diagram**
This image depicts a thermodynamic folding energy landscape and the conformational space available for an RNA folding to the native state.

## Research objectives

Determining an RNA folding mechanism is the overarching goal of any RNA folding research. Specifically, the objectives of this research are to investigate UNCG hairpin loop folding, and the role of various base substitutions at the N position on folding thermodynamics and kinetics. The first objective of this study was to thoroughly characterize the thermodynamics of UNCG tetraloops with equilibrium IR absorption spectroscopy. Equipped with this information, we were prepared to undertake our

main objective—investigating the folding kinetics. In this study, time-resolved infrared spectroscopy was used to probe intrinsic structural markers (vibrational bands) on the RNA hairpin loops subsequent to a laser-induced temperature jump (T-jump). The T-jump perturbs the thermal equilibrium allowing us to monitor the RNA as it relaxes to the new equilibrium position. Kinetic data were obtained for multiple vibrational transitions corresponding primarily to guanine base stacking and base pairing interactions. Each vibrational band we probe, interrogates a different structural interaction, and reports on distinct kinetics for that transition. With the ability to resolve early events at nucleotide resolution, our overall objective was to dissect the mechanism for hairpin loop formation.

Hairpin loops are ubiquitous secondary structural elements that, despite their relative structural simplicity, exhibit complex folding dynamics. This complexity arises from the presence of local minima in the folding landscape. One of the primary goals of this research was to figure out why hairpin loop folding is so complicated, and why folding proceeds so slowly. Studying small subunits of larger RNA structures was important because folding is hierarchical and the subunits are independently stable. Since folding is hierarchical, the study of loop formation (the first step) can be treated as separate from folding of a larger, more complex RNA (slow).

Recent data suggest that multiple probes are required to fully explore the rugged energy landscape of small RNA hairpins [41]. In addition, theoretical work using statistical mechanics also supports this theory [42]. The work in this dissertation is some of the first research to show that hairpin loops fold according to a rugged energy landscape similar to that of larger RNAs. Hairpins are relatively simple compared to

large RNAs, but still very complex on an atomic level (~2500 Da). Depending on which base was substituted into the loop, different thermodynamics and kinetics were observed. Remarkably, distinct kinetics were observed for different vibrational transitions on the same molecule. These results substantiated our two main research hypothesis: (1) Folding complexity stems from the earliest folding events (i.e. nanoseconds to microseconds), and (2) intra-loop interactions among the constituent bases influence the degree of complexity.

**Organization of dissertation**

This dissertation is organized into six chapters and appendices. The next chapter will discuss the experimental foundations (equilibrium IR and time-resolved IR spectroscopy of RNA) in detail. Chapter 3 deals briefly with curve fitting of complex biopolymer spectra. Chapter 4 describes the thermodynamics and equilibrium folding behavior of the different UNCG hairpin tetraloops, while Chapter 5 is dedicated to the time-resolved IR spectroscopy of the hairpins. Chapter 6 gives a detailed account of the thermodynamics and kinetics of the UCCG hairpin variant and suggests a folding mechanism. The appendices that follow give a comprehensive description of select experimental procedures (sample preparation, IR cell construction, spectroscopy, Raman conversion, and optics). These are topics that I hope will be helpful to future students.

**Bibliography**


[1] D. Conte and C. Mello, *Revealing the World of RNA Interface*. Nature, 431 (**2004**) 338-342.

[2] P.C. Bevilacqua and J.M. Blose, *Structures, kinetics, thermodynamics, and biological functions of RNA hairpins*. Annu. Rev. Biophys. Biomol. Struct., 59 (**2008**) 79-203.

[3] B.J. Tucker and R.R. Breaker, *Riboswitches as versatile gene control elements*. Current Opinion in Structural Biology, 15 (**2005**) 342-348.

[4] M. Broda, E. Kierzek, Z. Gdaniec, T. Kulinski and R. Kierzek, *Thermodynamic stability of RNA structures formed by CNG rinucleotide repeats. Implication for prediction of RNA structure.* Biochemistry, 44 (**2005**) 10873-10882.

[5] A. Jasinska, G. Michlewski, M. de Mezer, K. Sobczak, P. Kozlowski, M. Napierala and W.J. Kryzosiak, *Structures of trinucleotide repeats in human transcripts and their functional implications*. Nucleic Acids Research, 31 (**2003**) 5463-5468.

[6] G. Michlewski and W.J. Kryzosiak, *Molecular architecture of CAG repeats in human disease related transcripts*. Journal of Molecular Biology, 340 (**2004**) 665-679.

[7] A.D. Frankel and J.A. Young, *HIV-1: Fifteen proteins and an RNA*. Annu. Rev. Biophys. Biomol. Struct., 67 (**1998**) 1-25.

[8] K.J. Blight, A.A. Koykhalov and C.M. Rice, *Efficient initiation of HCV RNA replication in cell culture*. Science, 290 (**2000**) 1972.

[9] H. Huthoff and B. Berkhout, *Multiple secondary structure rearrangments during HIV-1 RNA dimerization*. Biochemistry, 41 (**2002**) 10439-10445.

[10] W. Gilbert, *The RNA world*. Nature, 319 (**1986**) 618.

[11] D.P. Bartel and P.N. Unrau, *Constructing an RNA world*. TCB, 9 (**1999**).

[12] T.R. Cech, *Crawling out of the RNA world*. Cell, 136 (**2009**) 599-602.

[13] R.T. Batey, R.P. Rambo and J.A. Doudna, *Tertiary motifs in RNA structure and folding*. Angew. Chem. Int. Ed., 38 (**1999**) 2326-2343.

[14] V.J. DeRose, *Metal ion binding to catalytic RNA molecules*. Current Opinion in Structural Biology, 13 (**2003**) 317-324.

[15] M. Pyle, *Metal ions in the structure and function of RNA*. J. Biol. Inorg. Chem., 7 (**2002**) 679-690.

[16] D.E. Draper, *A Guide to ions and RNA structure*. RNA, 10 (**2004**).

[17] G. Pljevaljeic, D. Klostermeier and D.P. Millar, *The tertiary structure of the hairpin ribozyme is formed through a slow conformational search*. Biochemistry, 44 (**2005**) 4870-4876.

[18] P.B. Moore, *The RNA folding Problem*. (**1999**) 381-401.

[19] C. Levinthal, *Are there pathways for protein folding?* Journal de Chimie Physique et de Physico Chimie Biologique, 65 (**1968**) 44-49.

[20] D. Thirumalai and S.A. Woodson, *Kinetics of folding of proteins and RNA*. Accounts of Chemical Research, 29 (**1996**) 433-439.

[21] R. Schroeder, R. Grossberer, A. Pichler and C. Waldsich, *RNA folding in vivo*. Current Opinion in Structural Biology, 12 (**2002**) 296-300.

[22] D. Herchlag, *RNA chaperones and the RNA folding problem*. The Journal of Biological Chemistry, 270 (**1995**) 20871-20874.

[23] W. Zhang and S. Chen, *Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states*. Biophysical Journal, 90 (**2006**) 778-787.

[24] D. Draper, D. Grilley and A.M. Soto, *Ions and RNA folding*. Annu. Rev. Biophys. Biomol. Struct., 34 (**2005**) 221-243.

[25] Y. Bai, M. Greenfield, K.J. Travers, V.B. Chu, J. Lipfert, S. Doniach and D. Hershlag, *Quantitative and comprehensive decompostion of the ion atmosphere around nucleic acids*. Journal of the American Chemical Society, 129 (**2007**).

[26] J. Vieregg, W. Cheng, C. Bustmante and I. Tinoco Jr., *Measurement of the effect of monovalent cations on RNA hairpin stability*. Journal of the American Chemical Society, 129 (**2007**) 14966-14973.

[27] D. Thirumalai, N. Lee, S.A. Woodson and D.K. Klimov, *Early events in RNA folding*. Annu. Rev. Phys. Chem, 52 (**2001**) 751-762.

[28] S. Chen, *RNA folding: Conformational statistics, folding kinetics, and ion electrostatics*. Annu. Rev. Biophys. Biomol. Struct., 37 (**2008**) 197-214.

[29] P.B. Moore, *Structural motifs in RNA*. Annu. Rev. Biophys. Biomol. Struct., 68 (**1999**).

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

[31] I. Tinoco Jr. and C. Bustamante, *How RNA folds*. Journal of Molecular Biology, 293 (**1999**) 271-281.

[32] B. Sclavi, M. Sullivan, M.R. Chance, M. Brenowitz and S.A. Woodson, *RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting*. Science, 279 (**1998**) 1940-1943.

[33] C. Cheong and H. Cheong, *RNA structure: Tetraloops*. (**2010**).

[34] V. Baumruk, C. Gouyette, T. Huynh-Dinh, J. Sun and M. Ghomi, *Comparison between CUUG and UUCG tetraloops: Thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy*. Nucleic Acids Research, 29 (**2001**) 4089-4096.

[35] V.P. Antao, S.Y. Lai and J. Tinoco, I., *A thermodynamic study of unusually stable RNA and DNA hairpins*. Nucleic Acids Research, 19 (**1991**) 5901-5905.

[36] M. Molinaro and I. Tinoco Jr., *Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: Thermodynamics and spectroscopic applications*. Nucleic Acids Research, 23 (**1995**) 3056-3063.

[37] M. Abdelkafi, N. Leulliot, V. Baumruk, V. Bednarova, P.Y. Turpin, A. Namane, C. Gouyette, T. Huynh-Dinh and M. Ghomi, *Structural features of the UCCG and UGGC tetraloops in very short hairpins as evidenced by optical spectroscopy*. Biochemistry, 37 (**1998**) 7878-7884.

[38] S.A. Woodson, *Metal ions and RNA folding: A highly charged topic with a dynamic future*. Current Opinion in Structural Biology, 9 (**2005**) 104-109.

[39] V.L. Murthy and G.D. Rose, *Is counterion delocalization responsible for collapse in RNA folding?* Biochemistry, 39 (**2000**) 14365-14370.

[40] R. Russell, X. Zhuang, H.P. Babcock, I.S. Millett, S. Doniach, S. Chu and D. Herschlag, *Exploring the folding landscape of a structured RNA*. PNAS, 99 (**2002**) 155-160.

[41] C. Hyeon and D. Thirumalai, *Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins*. Journal American Chemical Society, 130 (**2008**) 1538-1539.

[42] S. Chen and K.A. Dill, *RNA folding energy landscapes*. PNAS, 97 (**2000**) 646-651.

# Chapter 2

## Experimental foundations

## Introduction

The purpose of this chapter is to develop the theory of IR spectroscopy as it applies to the study of RNA folding. Both equilibrium and time-resolved IR spectroscopy are covered. In addition, this chapter details the experimental methods involved for both the equilibrium and time-resolved IR spectroscopy used in this work. First, the utility of IR spectroscopy is explained by comparison to other types of optical spectroscopic techniques (e.g., UV-vis and fluorescence). The need for a structurally specific technique (IR) is immediately established. Next, the importance of IR spectroscopy is recognized for its ability to resolve multiple, structurally specific vibrational bands. Finally, an overview of the experimental techniques is described.

## Infrared spectroscopy of RNA

To understand how IR spectroscopy is used to study RNA folding, it is necessary to reconsider the forces that are responsible for driving folding and stabilizing the secondary structures. Taking this information into consideration, we can then rationalize how IR spectroscopy is used to characterize and probe the specific interactions. As mentioned, RNA is a biopolymer consisting of a sequence of nucleotide bases connected by a ribose-phosphate backbone. Interactions between the nucleotide bases provide the structural foundation upon which the more complex native structures are built. Specifically, these intramolecular interactions are hydrogen bonds between complementary bases (i.e., base pairing) and base stacking interactions between adjacent bases **(refer to Figures 1.2 and 1.3).** In order to obtain information

about the folding process, it is necessary to employ a technique that can monitor these specific interactions on the time scale of interest.

A variety of techniques are available; each with their own advantages and disadvantages. For example, the nucleotide bases absorb strongly in the ultraviolet (UV) region of the spectrum. Unfortunately, UV spectroscopy offers little structural detail (**Figure 2.1**) [1-3]. Fluorescence spectroscopy provides more structurally specific information, but the native bases themselves are only weakly fluorescent requiring fluorescent probes to be incorporated into the RNA. This has a disadvantage in that the probes may disrupt the native RNA structure, which could render the results unreliable and difficult to interpret. Furthermore, the structural information obtained is limited to the region in the immediate vicinity of the probe.

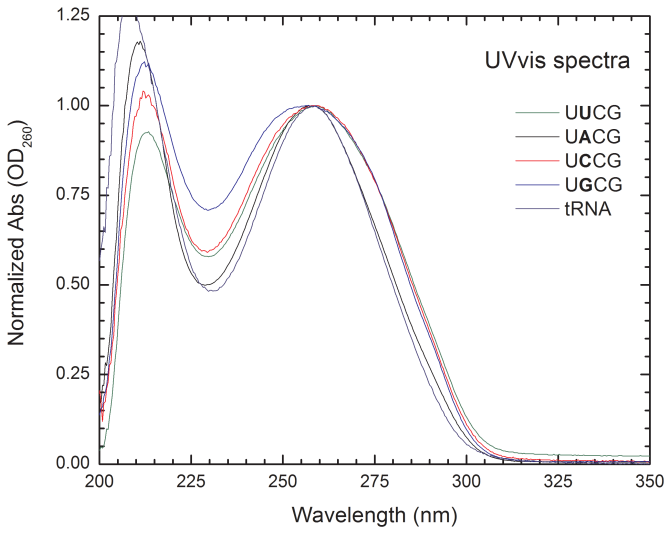


**Figure 2.1: UV-visible spectra of RNAs**

The four oligomers studied in this work as well as a large 76 nucleotide tRNA are shown. All the spectra are indistinguishable and exhibit a single broad band centered around 260 nm.

On the other hand, infrared (IR) spectroscopy offers excellent structural detail without the need to introduce non-native probes. For example, AU base pairs and GC base pairs are readily distinguished [4, 5]. The structural sensitivity of IR spectroscopy is due to the vibrational transitions of specific molecular groups on the nucleotide bases, which absorb strongly in the IR region between 1500-1800 $cm^{-1}$ (**Figure 2.2**) giving structural information and essential details into base pairing and base stacking interactions [4, 5].

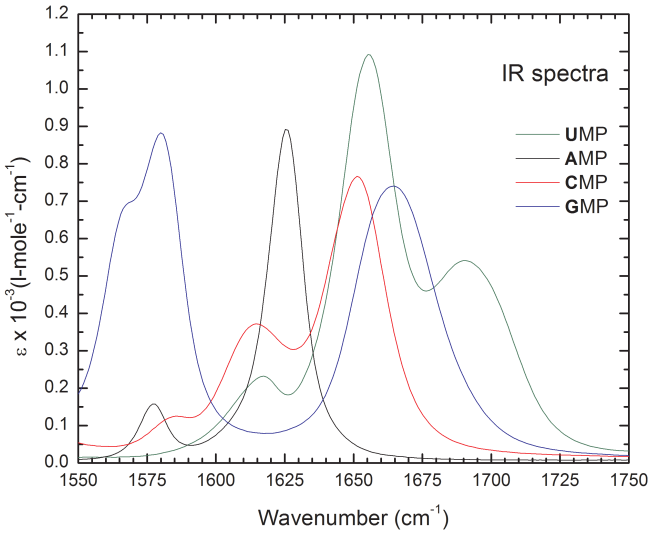


**Figure 2.2: IR spectra of the ribonucleotide monophosphates**
The infrared region between 1550 $cm^{-1}$ and 1750 $cm^{-1}$ is rich with vibrational information that reports on functional groups and specific structural interactions between the bases (band assignments described in the text).

These absorptions are due to the stretching vibrations of C=O, C=N, and C=C molecular groups [6-8]. The H-bonds between base pairs directly involve these groups **(refer to Figure 1.2).** When the H-bonds are altered (e.g., during folding or unfolding)

the changes are reflected in the IR spectrum. Similarly, dispersive interactions involved in base stacking also feature in the spectral region. These changes can manifest as band shifts, absorption changes, lineshape changes, and/or combinations thereof. Transitions at specific wavenumbers report on precise structural features between the associated bases. Absorption changes at 1575 $cm^{-1}$ are assigned to guanine base stacking interactions. Spectral changes at 1620 and 1660 $cm^{-1}$ correspond to AU and GC base pairing interactions, respectively [4]. Temperature dependent IR spectra for tRNA are shown in **Figure 2.3** and demonstrate the sensitivity of IR spectroscopy to monitor RNA folding and provide detail regarding the conformational state of the tRNA.

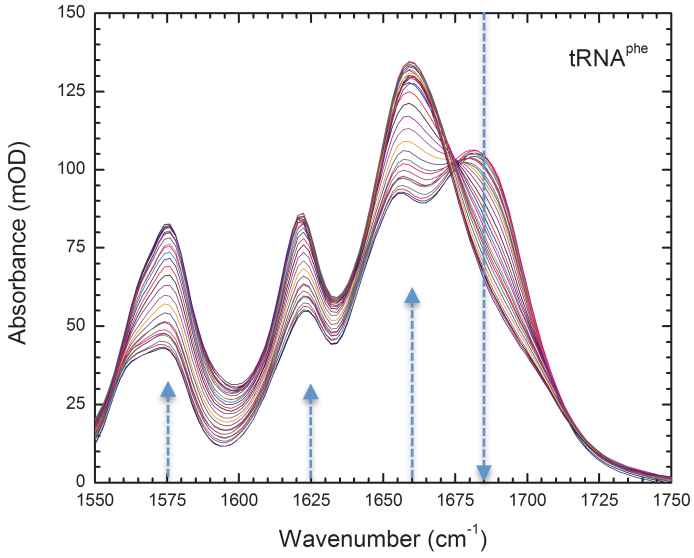


**Figure 2.3: Equilibrium infrared spectra of $tRNA^{phe}$**

Temperature dependent IR spectra of phenylalanine specific transfer RNA. The arrows indicate the direction of spectral changes with increasing temperature. Temperatures ranged from 20 to 90°C in 2 degree increments.

**Equilibrium IR spectroscopy**

Before time-resolved experiments were conducted, the equilibrium IR absorption spectra of each sample as a function of temperature were studied in detail. These spectra were collected using a Newport Model 8025 MIR FTIR spectrophotometer and involved performing melting experiments where IR spectra were recorded at thermal equilibrium for a range of temperatures from 20°C to 90°C in increments of 2°C. As the temperature increases, the RNA unfolds and its spectrum changes in accordance with the previous discussion. A series of temperature dependent IR spectra for tRNA was shown in **Figure 2.3** and illustrates the effect of increasing temperature on the sample. Increasing the temperature unfolds the RNA resulting in the observed absorption changes. The peak at 1575 $cm^{-1}$ due to a guanine C=N ring vibration increases in intensity and blue-shifts to 1580 $cm^{-1}$. The peak at 1625 $cm^{-1}$ due to an adenine C=N ring vibration increases in intensity and red-shifts to 1620 $cm^{-1}$. The broad band around 1660 $cm^{-1}$ is due to combined contributions from a guanine C6=O6 stretching vibration at 1669 $cm^{-1}$ and a cytosine C2=O2 stretch at 1652 $cm^{-1}$. As the GC base pairs melt, the guanine stretch shifts from 1688 $cm^{-1}$ to lower wavenumbers and the cytosine stretch at 1650 $cm^{-1}$ shifts to higher frequency. These transitions gives rise to the negative absorbance change at 1680 $cm^{-1}$ and a concomitant increase in intensity at 1660 $cm^{-1}$.

"Melt curves" are obtained by plotting the absorbance at specific wavenumbers against temperature. For example, **Figure 2.4** shows a melting curve for tRNA. By fitting these curves to a thermodynamic model, the melting temperature and other thermodynamic information can be obtained (the model will be discussed in Chapter 3).

Equipped with this information, we knew precisely at which vibrational frequencies and what temperatures to perform the T-jump experiments.

In addition to the hairpin tetraloops described in detail later in this work, IR melting experiments were also conducted on tRNA, the polynucleotides (polyA, polyU, polyG, polyC), the heteropolymers (polyA,U and polyG,C). For brevity, the polynucleotides will not be shown in this work. The individual monophosphates (AMP, UMP, GMP, and CMP) were also studied (their IR spectra were shown in **Figure 2.2**). The tRNA sample is well characterized and was used as an instrumental standard. All the other samples were used as standards to qualitatively and quantitatively distinguish between paired and unpaired bases allowing the extent of folding to be determined for an unknown sample.

All IR spectra were recorded in $D_2O$ because the IR spectrum of $H_2O$ has a strong vibrational transition at 1643.5 $cm^{-1}$ which overlaps, and obscures the IR spectra of the nucleotides [9]. The spectrum of $D_2O$ is shifted to lower frequencies allowing acquisition of data in this region.

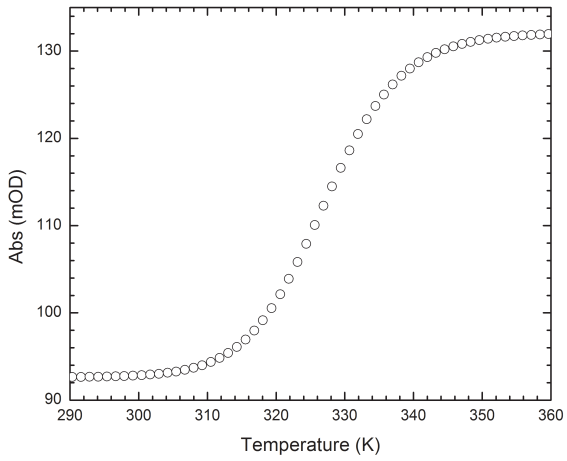


**Figure 2.4: Melting curve for $tRNA^{phe}$ at 1660 $cm^{-1}$**
Increasing temperature unfolds the tRNA resulting in sigmoidal melting behavior, which manifests as an increase in absorbance from 93 to 130 mOD.

**Time-resolved IR spectroscopy**

Initiating folding or unfolding on a sufficiently fast time scale is a difficult experimental hurdle to overcome. In this study a short laser pulse tuned to a specific vibrational transition of the solvent was used to rapidly heat a small volume of the sample solution. Upon rapid heating, the equilibrium between folded and unfolded RNA is displaced in favor of the unfolded structure (**Figure 2.5**). Using a second laser tuned to a specific vibrational frequency of the RNA, the relaxation kinetics are then probed. The heating method is called a laser induced temperature jump (T-jump). The temperature of the irradiated volume can be increased by as much as 20°C in as little as 10 ns. We have used this technique to study both large tRNAs and the small hairpins described in this work [10, 11]. For a detailed review of laser-induced temperature measurements of RNA folding see Brauns and Dyer [12].

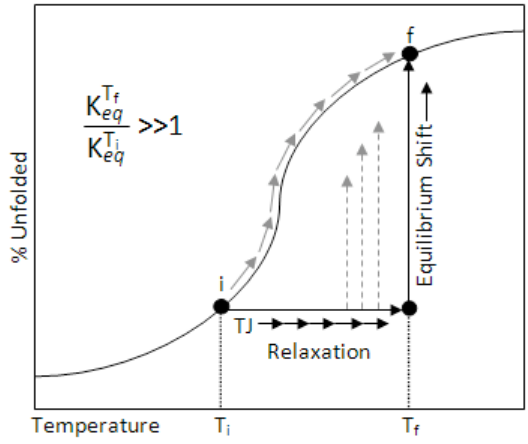


**Figure 2.5: Temperature jump relaxation kinetics**
The solvent responds instantaneously to the T-jump pulse and is completely done heating within ~50 ps. The RNA slowly "relaxes" (ns to ms) to this perturbation and its equilibrium shifts in favor of the unfolded state.

Essentially, this is a pump-probe experiment (**Figure 2.6**). The T-jump pulse at 1.9 μm is generated by Raman shifting the fundamental output of a Q-switched Nd:YAG operating at 1.064 μm and 10 Hz. The 1.064 μm pulses are passed through a 1 m long cell filled with $H_2$ gas at ~275 psi (see Appendix C). Raman scattering of the pump occurs and generates both Stokes and anti-Stokes lines. The first Stokes line is at 1.9 μm and is separated from the other wavelengths using a Pellin-Broca prism. The repetition rate of the 1.9 μm pulses remains at 10 Hz and the individual pulse energy is on the order of ~20 mJ. The pulses are directed toward the sample using gold coated mirrors. Before arriving at the sample, the pump passes through a polarizer and a long focal length lens. The polarizer can be rotated to finely adjust the pump energy (and therefore the magnitude of the T-jump) and the lens is used to adjust the pump diameter at the sample cell.

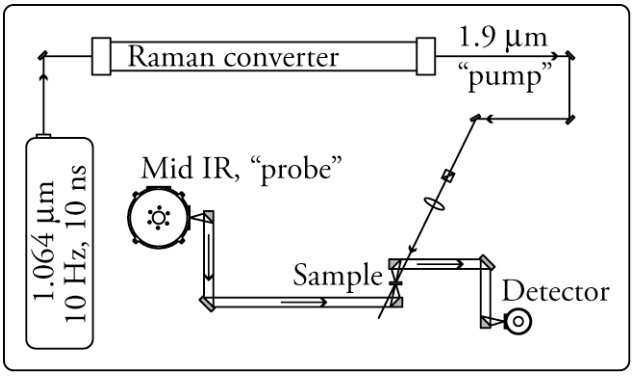


**Figure 2.6: Laser induced temperature jump schematic**
See text for details.

The probe beam is generated by a CW (continuous wave) diode laser tuned to the frequency corresponding to a vibrational mode of interest. The diode laser output is divergent and is first collected and collimated using a gold coated off axis paraboloidal reflector (OAP) with a focal distance of 1 inch and an off axis angle of 90°. A second OAP focuses the probe onto the sample and a third collects the transmitted radiation. The transmitted signal is finally refocused onto a mercury cadmium telluride (MCT) detector using a fourth OAP. The instrument response time is determined by the detector bandwidth or the laser pulse width. The detector signal is digitized using a 12-bit, high-speed data acquisition card. A LabVIEW program is used to collect the data and coordinate the instrumentation.

The sample cells are a custom design and are used for both the T-jump experiment and the equilibrium FTIR spectroscopy. They are composed of two $CaF_2$ windows separated by a Teflon spacer (~50μm). The windows are sandwiched within a copper housing to ensure good thermal contact. The Teflon spacer defines the optical path length and divides the cell into two compartments; one for the reference solution (buffer) and one for the sample (RNA + buffer). The copper housing is mounted on a thermally regulated copper platform for temperature control (using a re-circulating water bath). A thermocouple sensor mounted on the cell is used to control the bath temperature using a PID feedback loop.

The cell and platform are mounted on a 3-axis translation stage for precise positioning. The probe beam is focused tightly onto the cell (diameter < 50 μm) and the pump diameter is adjusted to be ~5× that of the probe (~1mm). An experiment is performed by digitizing the transient transmission signals subsequent to each T-jump

pulse. As a general rule, ~2000 transients are collected and averaged (at a rate of 10 transients per second). First, the experiment is performed on the reference side of the cell. The cell is then translated and the experiment is repeated on the sample side. Absorbance values are obtained from the transmission data, and by subtracting the reference from the sample, a kinetic profile of the RNA due to the temperature jump is obtained. The heated volume begins to cool in several hundred microseconds and limits the long time response of the experiment.

**Oligonucleotides**

The custom synthesized, HPLC purified oligomers were purchased from Integrated DNA technologies. Oligonucleotides having the sequence 5′-gcUNCGgc-3′ were studied. Each of the N substituted ribonucleobases (N=U, A, C, or G) were investigated. The oligomers were prepared at a concentration of ~2 mM (16 mM nucleotides) in 100 mM phosphate buffer at pH 7.2 and dialyzed a million fold. Next, they were lyophilized (4×) against $D_2O$ to remove labile protons, and reconstituted in $D_2O$. Finally, the samples were degased for three hours with a conventional table top vacuum pump. Dry work was done in a glove box purged with water free air. Well-established, standard protocols were followed for sample handling and preparation. For all other samples, and a comprehensive guide on sample preparation, see Appendix A.

**Bibliography**


[1] V. Baurmruk, C. Gouyette, T. Huynh-Dinh, J. Sun and M. Ghomi, *Comparison Between CUUG and UUCG Tetraloops: Thermodynamic Stability and Structural Features Analyzed by UV Absorption and Vibrational Spectroscopy*. Nucleic Acids Research, 29 (**2001**) 4089-4096.

[2] M. Abdelkafi, N. Leulliot, V. Baumruk, V. Bednarova, P.Y. Turpin, A. Namane, C. Gouyette, T. Huynh-Dinh and M. Ghomi, *Structural features of the UCCG and UGGC tetraloops in very short hairpins as evidenced by optical spectroscopy*. Biochemistry, 37 (**1998**) 7878-7884.

[3] M. Abdelkafi, N. Leulliot, V. Baumruk, C.H. du Penhoat, O. Lampire, N. Bouchemal-Chibani, P. Goyer, A. Namane, C. Gouyette, T. Huynh-Dinh and L. Bednarova, *Common structural features of UUCG and UACG tetraloops in very short hairpins determined by UV absorption, Raman, IR, and NMR spectroscopies*. Journal Biomolecular Structure and Dynamics, 14 (**1997**) 579-593.

[4] G.J. Thomas Jr., *Determination of the base pairing content of ribonucleic acids by infrared spectroscopy*. Biopolymers, 7 (**1969**) 325-334.

[5] M. Banyay, M. Sarkar and A. Gräslund, *A library of IR bands of nucleic acids in solution*. Biophysical Chemistry, 104 (**2003**) 477-488.

[6] M. Banyay, M. Sarkar and A. Graslund, *A library of IR bands of nucleic acids in solution*. Biophysical Chemistry, 104 (**2003**) 477-488.

[7] E. Taillandier, and Liquier, J., *Handbook of Vibrational Spectroscopy*. (**2002**) 1-16.

[8] H.H. Mantsch and D. Chapman, *Infrared Spectroscopy of Biomolecules*. (**1996**) 131-158.

[9] S.Y. Venyaminov and F.G. Prendergast, *Water ($H_2O$ and $D_2O$) molar absorptivities in the 1000-4000 $cm^{-1}$ range and quantitative infrared spectroscopy of aqueous solutions*. Analytical Biochemsitry, 248 (**1997**) 234-245.

[10] E.B. Brauns and R.B. Dyer, *Time-Resolved infrared spectroscopy of RNA folding*. Biophysical Journal, 89 (**2005**) 3523-3530.

[11] A.S. Stancik and E.B. Brauns, *Rearrangement of partially ordered stacked conformations contributes to the rugged energy landscape of a small RNA hairpin*. Biochemistry, 11 (**2011**) 345.

[12] R.B. Dyer and E.B. Brauns, *Laser-induced temperature jump infrared measurements of RNA folding*. Methods in Enzymology, (**2009**).

## Chapter 3

## A simple asymmetric lineshape for fitting infrared absorption spectra

Stancik, A.L. and Brauns, E.B., *Vibrational Spectroscopy*, **2008**, 47, 66-69.

**Abstract**

Almost exclusively, lineshape functions used to model IR absorption peaks are based on symmetric frequency distributions. However, in complex systems such as large biological macromolecules in aqueous solutions, the distribution of vibrational frequencies may in fact be asymmetric. In this communication, we show that asymmetry can be introduced to the standard symmetric Lorentzian and Gaussian lineshapes using a simple, easy to implement method. Our technique involves replacing the static width parameter of the symmetric profiles with a smoothly varying function that is wavenumber dependent. In this way, the width varies across the IR band resulting in an asymmetric peak. In our model, the width varies sigmoidally with wavenumber. As a demonstration, we fit experimental spectra of adenosine 5′-monophosphate (AMP).

**Introduction**

IR absorption spectra of large biological macromolecules in aqueous solutions are frequently comprised of broad, overlapping peaks. When spectral congestion is substantial, the interpretation of the data is difficult. Curve fitting of IR absorption spectra is a useful technique that can be used to facilitate data interpretation [1]. In curve fitting, the experimental spectrum is modeled as a sum of individual spectral contributions. For obvious reasons, the choice of lineshape for the component spectra has a profound impact on the outcome of the fitting procedure. Mathematically, the lineshape is a statistical function that describes the distribution of frequencies within a given transition. Symmetric distribution functions are used almost exclusively.

Typically these functions are firmly grounded in theory and usually provide reliable results. However, in complex systems (e.g., large molecules in condensed phases), the frequency distributions may in fact be asymmetric. When symmetric lineshapes are used to fit asymmetric data, the results will be unsatisfactory. All too often, the temptation is to add additional symmetric components to the model to improve the fit. However, if the number of peaks in the model exceed the actual number of underlying transitions in the data, the results will be misrepresented. In this communication, we present a straightforward modification of well known symmetric lineshapes that introduces asymmetry.

**Experimental**

The disodium salt of adenosine 5′-monophosphate (AMP) was purchased from Sigma and used without further purification. Reported purity was greater than 99.0%. Samples were lyophilized against $D_2O$ three times to exchange the labile protons. The lyophilized samples were dissolved in 100 mM deuterated phosphate buffer (pD = 7.2) to a final nucleotide concentration of ~150 mM. Transmission infrared spectra were recorded on an Oriel MIR8025 FTIR spectrometer using custom built sample cells. The cells are comprised of two $CaF_2$ windows separated by a 50 μm thick Teflon spacer that defines the cell path length. The spacer also divides the cell into two compartments, one for the sample (AMP + buffer) and the other for the reference (buffer only). The $CaF_2$ windows are held together in a copper housing. The copper housing is then mounted to a temperature controlled platform. All spectra were recorded at 25 °C ± 0.01 °C and a

spectral resolution of 2 $cm^{-1}$. The interferograms were apodized using a triangular function prior to Fourier transformation.

**Lineshape models**

In principle, the number of functions that could be used to fit spectral data is limitless [2]. However, to be truly meaningful, the parameters of the lineshape model should have a direct correspondence with physical phenomena. For example, parameters such as peak width, peak area, and peak position are the relevant quantities when dealing with symmetric spectra. As we'll show shortly, one additional parameter is necessary to characterize asymmetry. The asymmetry parameter is an indication of the extent of asymmetry and the direction of skew (i.e., whether the spectrum is skewed toward higher or lower wavenumbers). In addition to having physically meaningful parameters, we would also argue that the lineshape functions are firmly rooted in theory. For example, an $n^{th}$ order polynomial can be used to fit some peaks with satisfactory results. However, such a model is purely empirical and has no theoretical basis.

There are three factors that contribute to the final lineshape of an IR band; Doppler broadening, radiation damping, and collision broadening. Doppler broadening results in Gaussian lineshapes [3, 4]

$$G(\nu) = \frac{A}{\gamma_0}\sqrt{\frac{4\ln 2}{\pi}}\exp\left[-4\ln 2\left(\frac{\nu-\nu_0}{\gamma_0}\right)^2\right] \qquad (3\text{-}1)$$

where $\nu$ is the frequency in wavenumbers, $\nu_0$ is the peak center, and $\gamma_0$ is the full width at half maximum (FWHM). The term in front of the exponential is a normalization constant where $A$ is the area under the peak. The latter two, radiation damping and collision broadening give Lorentzian distributions [3, 4]

$$L(\nu) = \frac{2A/\pi\gamma_0}{1 + 4\left(\frac{\nu - \nu_0}{\gamma_0}\right)^2} \qquad (3\text{-}2)$$

The parameters are defined as in Equation 3-1 and the numerator is the normalization constant.

Since all three effects act on molecules, lineshapes are usually a convolution of the two symmetric profiles; this is the so-called Voigt function [5]. A true Voigt function is cumbersome and impractical for use in fitting routines so a pseudo Voigt profile is frequently used. One form of a pseudo Voigt that we will use here treats each peak as a sum of fractional contributions of the Gaussian and Lorentzian shapes:

$$y(\nu) = fL(\nu) + (1 - f)G(\nu) \qquad (3\text{-}3)$$

Here, $f$ is the fraction of Lorentzian character contributing to the net line shape. For example, when $f = 1$, the shape is a pure Lorentzian and when $f = 0$, the shape is a pure Gaussian. For a symmetric shape, the first and third derivatives at the peak center ($\nu = \nu_0$) are both equal to zero. In other words, the peak maximum exactly corresponds to

the second derivative minimum. As will be shown shortly, this is not the case for an asymmetric band.

### Asymmetric variation of the Lorentzian and Gaussian lineshapes

When the width is not uniform across an entire band, an asymmetric profile results. Our approach is to replace $\gamma_0$ in Equations 3-1 and 3-2 with a well behaved function, $\gamma(\nu)$ to introduce asymmetry. In the symmetric limit, the lineshape will reduce to a pure Lorentzian or Gaussian profile. While numerous possibilities exist, we obtain the best results when the width is allowed to vary sigmoidally,

$$\gamma(\nu) = \frac{2\gamma_0}{1 + \exp\left[a(\nu - \nu_0)\right]} \qquad (3\text{-}4)$$

The parameter $a$ is a measure of asymmetry. Negative values of $a$ skew the spectrum toward higher wavenumbers while positive values of $a$ skew it toward lower wavenumbers. When $a$ is zero, Equation 3-4 reduces to $\gamma_0$ and the resulting band is a standard symmetric Gaussian or Lorentzian profile with a constant width (FWHM = $\gamma_0$). The sigmoid is advantageous since the width asymptotically approaches upper and lower bounds (i.e., 0 and 2 $\gamma_0$). In contrast, when other functions are used (e.g., linear) the width will increase and decrease without bound in opposite directions. If limits are not imposed, the width will eventually become non-physical (e.g., negative) or it will become so large that the peak will begin to increase after reaching a minimum. **Figure 3.1 a** shows the result of Equation 3-4 substituted for $\gamma_0$ in a Lorentzian profile. The

asymmetry parameter, $a$, is equal to $-1/5$. At $\nu = \nu_0$ the width is equal to $\gamma_0$ (see **Figure 3.1 b**).

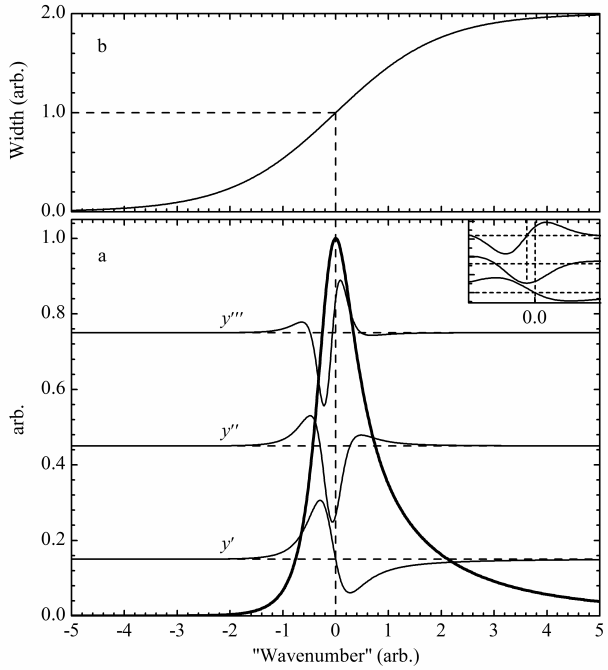


**Figure 3.1: Asymmetric Lorentzian**
(a) An asymmetric Lorentzian (thick black solid line) using Equation 4 for $\gamma_0$, where $\nu_0 = 0$, $\gamma_0 = 1$ and $a = -1/5$. The derivative spectra are shown on an arbitrary $y$ axis for comparison and labeled accordingly. The horizontal dashed lines indicate $y = 0$ for each derivative. The inset shows the derivative spectra on an expanded wavenumber scale (−0.5 to 0.5). (b) A plot of Equation 4, $\gamma(\nu)$.

A peak that appears to be asymmetric may in fact be comprised of two closely overlapped symmetric peaks. This is a condition that must be distinguished when dealing with asymmetric absorption bands. Up to a point, the two situations can be distinguished quantitatively. Consider a spectrum that is the sum of two symmetric component peaks of differing heights. When they are separated by a sufficient distance,

the second derivative shows two distinct minima. However, as one approaches the other (all other variables being held constant, only the peak center is changing), the second derivative minimum corresponding to the weaker peak diminishes*. The point at which the minimum is tangent to zero is known as the *detection limit* [6]. Stated more rigorously, the detection limit is the point at which the second and third derivatives are equal to zero at the same wavenumber coordinate. Below this limit, the second derivative spectra cannot resolve two peaks. Higher derivatives can be used, but this is impractical for real data. When such a situation arises, *a priori* knowledge of the spectrum in question may provide the only clarification.

**Fitting experimental spectra**

The impetus for this paper arose from our efforts to model IR absorption spectra of nucleic acids. In many cases, using symmetric lineshapes to model the spectra failed to give satisfactory results. We concluded that the frequencies must be distributed asymmetrically. The IR active moieties on a large molecule in solution can experience numerous "microenvironments" that impact the vibrational frequencies of those species. If the frequencies differ by a sufficient amount, multiple, distinct bands will be resolved. Conversely, if the differences are small, a single band results and the individual frequencies of each of the absorbing species are distributed about a central frequency. In many cases, the distribution is symmetric. However, it is possible for the symmetry to break down resulting in a single asymmetric band. This is more pronounced in large macromolecules where in homogeneities among the different

* Equivalently, the weaker peak could increase in height or sharpen relative to the stronger peak.

microenvironments may be more evident. To demonstrate this and the need to consider asymmetric lineshapes, we have chosen to fit FTIR absorption spectra of 5′-monophosphate (AMP). This is an excellent system to use since its IR absorption spectrum and underlying transitions are known.

The spectrum of AMP and its second derivative are shown in **Figure 3.2**. In agreement with what is expected based on previously published data, two transitions are observed [7]. A strong absorption centered around 1625 $cm^{-1}$ is due to C=N and C=C vibrations of the purine ring. The corresponding in-plane ring vibration is weaker and is centered around 1577 $cm^{-1}$. Initially, the spectra were fit using a symmetric pseudo Voigt model. The results are shown in **Figure 3.2** and tabulated in **Table 3.1**. While the fit is adequate, it clearly does not accurately model the spectra. In addition to the large residual, the second derivatives of the fit and the experimental data are substantially different, especially the region around the main peak (~1625 $cm^{-1}$). Quantitatively, the second derivative spectrum indicates only two peaks. However, we cannot unequivocally rule out the presence of a third peak based on these data alone. In fact, the addition of a third peak dramatically increases the agreement between the fit and the data (not shown). However, there is no precedent to indicate that a third transition is present [7]. We would conclude that, in reality, the strongest transition (and possibly to a lesser extent, the weak transition) is in fact asymmetric. **Figure 3.3** shows the fitting results using our asymmetric model. The quality of the fit is improved substantially when asymmetry is allowed for. In addition, the second derivative of the fit more closely reproduces the second derivative of the data. The results of both fits are tabulated in **Table 3.1**.

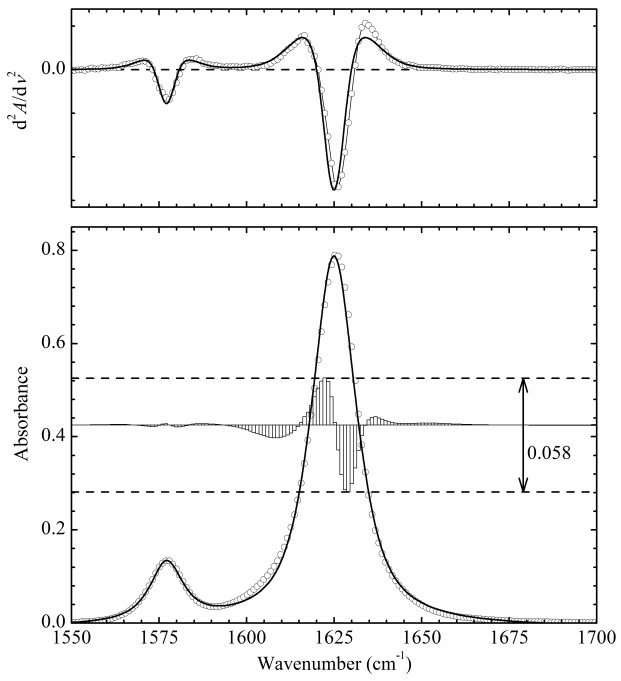


**Figure 3.2: Fitting of AMP IR absorption spectra using a symmetric pseudo Voigt model**

IR absorption spectra of AMP in $D_2O$ (lower plot, circles) and its second derivative (upper plot, circles with thin black solid line). The solid black line is a fit to a symmetric pseudo Voigt (Equation 3) profile (lower plot). The second derivative of the fit function is shown in the upper plot (solid black line). The residuals (fit – actual) are plotted on an expanded scale on the lower plot. The upper and lower limits are indicated.

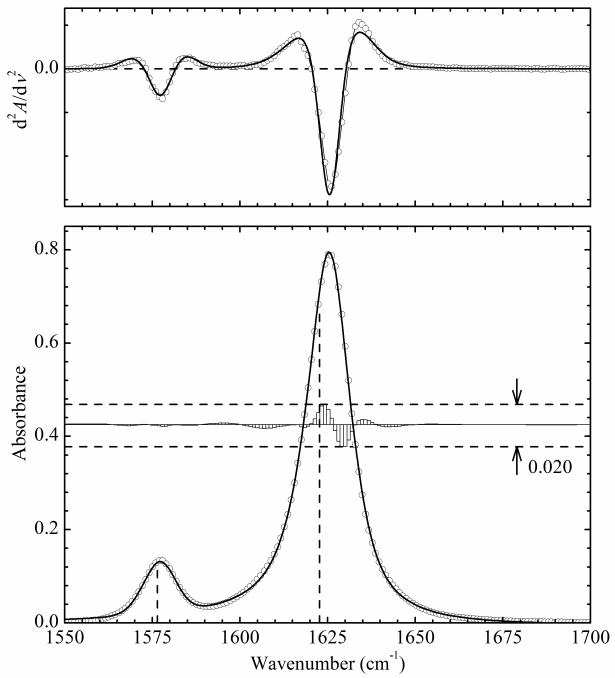


**Figure 3.3: Fitting of AMP IR absorption spectra using an asymmetric pseudo Voigt model**

IR absorption spectra of AMP in $D_2O$ (lower plot, circles) and its second derivative (upper plot, circles with thin black solid line). The solid black line in the lower plot is a fit to an asymmetric pseudo Voigt using Equation 4 for $\gamma_0$. The second derivative of the fit function is shown in the upper plot (solid black line). The residuals (fit – actual) are plotted on an expanded scale on the lower plot. The upper and lower limits are indicated. The vertical dashed lines show the mean wavenumber for each of the peaks, $\langle \nu \rangle$ = 1576.50 $cm^{-1}$ and 1622.75 $cm^{-1}$.

The introduction of asymmetry does more than improve the quality of the fit—there are significant practical implications as well. Most notably is the distinction between peak position and the mean wavenumber. The peak position, $\nu_0$, simply reports the wavenumber having the maximum absorbance. When dealing with asymmetric peaks, further insight can be obtained if one considers the mean wavenumber of each transition in addition to the peak maximum. The mean wavenumber (for each isolated band) is defined according to:

$$\langle \nu \rangle = \nu_0 + \frac{\int_{-5\gamma_0}^{+5\gamma_0} j\, y(j)\, dj}{\int_{-5\gamma_0}^{+5\gamma_0} y(j)\, dj} \qquad (3\text{-}5)$$

where $j = \nu - \nu_0$. The limits of integration, $\pm 5\gamma_0$, are chosen to encompass the majority of the band area. Without limits, the mean frequency of an asymmetric band would increase (or decrease) without bound. The mean wavenumber for a symmetric peak is simply equal to the maximum wavenumber, $\nu_0$. However, this is not the case for an asymmetric peak. If the peak is skewed toward higher wavenumbers, the mean wavenumber will be greater than the peak maximum. Likewise, if the peak is skewed toward lower wavenumbers, the mean will be less than the peak maximum. For example, our model indicates that the weakest peak is slightly asymmetric and its mean is shifted by less than a wavenumber (toward lower frequencies) from the peak position (**see Figure 3.3 and Table 3.1**). The asymmetry in the stronger transition is more pronounced and the mean is shifted by −2.69 $cm^{-1}$ from 1625.44 $cm^{-1}$ to 1622.75 $cm^{-1}$.

The most marked difference between the symmetric and asymmetric fits is the fraction of Lorentzian character, *f*, for the weak peak (~1577 $cm^{-1}$). In the symmetric model, this peak is 93% Lorentzian. When asymmetry is accounted for, the fraction of Lorentzian character decreases to 20%. The reason for this is that the asymmetry of the stronger peak is skewed toward shorter wavenumbers. As such, the tail of this peak on the short wavenumber side is more pronounced and contributes to the baseline of the weaker peak. One of the features that distinguishes a Gaussian from a Lorentzian is that the former drops off to zero very rapidly, while the latter approaches zero much more slowly. Thus, a Lorentzian peak with a substantial portion of its baseline negated will appear more "Gaussian". However, we note that other than this difference, the other parameters for the weaker peak are not impacted when asymmetry is introduced.

**Table 3.1** Summary of curve fitting results.

| | | Symmetric | Asymmetric |
|---|---|---|---|
| Peak 1 | $\nu_0/cm^{-1}$ | 1577.14 | 1577.14 |
| | $\langle\nu\rangle/cm^{-1}$ | – | 1576.50 |
| | $A$ | 2.26 | 1.41 |
| | $\gamma_0/cm^{-1}$ | 11.43 | 10.91 |
| | $a$ | – | 0.01 |
| | $f$ | 0.93 | 0.2 |
| Peak 2 | $\nu_0/cm^{-1}$ | 1624.97 | 1625.44 |
| | $\langle\nu\rangle/cm^{-1}$ | – | 1622.75 |
| | $A$ | 17.86 | 17.75 |
| | $\gamma_0/cm^{-1}$ | 15.30 | 15.04 |
| | $a$ | – | 0.01 |
| | $f$ | 0.84 | 0.86 |

## Conclusion

In short, we have established a straightforward method to account for asymmetric IR absorption lineshapes. By replacing the static width of a standard Lorentzian or Gaussian profile with a smoothly varying function, asymmetry is introduced. In our model, the width varies sigmoidally with wavenumber across the band. The model fits experimental spectra of AMP well. In addition, the sigmoid function asymptotically approaches upper and lower limits and therefore gives a physically meaningful result.

**Bibliography**

[1] J. Dodd, G. and L.K. DeNoyer, Curve-fitting: Modeling Spectra, in: J.M. Chalmers, P.R. Griffiths (Eds.) Handbook of Vibrational Spectroscopy, John Wiley & Sons Ltd., Chichester, 2002.

[2] V.B. Di Marco and G.G. Bombi, *Mathematical functions for the representation of chromatographic peaks*. Journal of Chromatography A, 931 (**2001**) 1-30.

[3] R.N. Jones, K.S. Seshadri, N.B.W. Jonathan and J.W. Hopkins, *A statistical approach to the analysis of infrared band profiles*. Canadian Journal of Chemistry, 41 (**1963**) 750-762.

[4] K.S. Seshadri and R.N. Jones, *The shapes and intensities of infrared absorption bands. A review*. Spectrochimica Acta, 19 (**1963**) 1013-1085.

[5] A. Belafhal, *The shape of spectral lines: widths and equivalent widths of the Voigt profile*. Optics Communications, 177 (**2000**) 111-118.

[6] B.G.M. Vandeginste and L. De Galan, *Critical evaluation of curve fitting in infrared spectrometry*. Analytical Chemistry, 47 (**1975**) 2124-2132.

[7] M. Banyay, M. Sarkar and A. Graslun, *A library of IR bands of nucleic acids in solution*. Biophysical Chemistry, 104 (**2003**) 477-488.

# Chapter 4

## Investigating the thermodynamics of UNCG tetraloops using infrared spectroscopy

## Introduction

In order to carry out meaningful time-resolved experiments, the equilibrium characteristics (i.e., the thermodynamics) of each sample must be thoroughly understood. The work described in this chapter details my efforts toward this goal. Briefly, temperature dependent IR spectra were obtained for each of the four UNCG tetraloops. From these data, melt curves (absorbance vs. temperature) at specific wavenumbers were plotted. By fitting these melting data to a two-state model (described in the text), thermodynamic information such as folding enthalpies and melting temperatures were obtained. Although the motivation for this work was to set the stage for the kinetic experiments, the results are surprisingly significant on their own merits. Because of the enhanced level of structural detail afforded by IR spectroscopy (described in Chapter 2), thermodynamics can be extracted that correspond to specific structural features (e.g., base pairing vs. base stacking, AU base pairs vs. GC base pairs, etc.). Consequently, far more insight into the folding mechanism is obtained.

## Results

### Equilibrium IR spectra

Equilibrium temperature dependent IR spectra for each of the UNCG hairpin loops are shown in **Figure 4.1**. In addition, difference spectra (**Figure 4.2)** are shown where the lowest temperature spectrum has been subtracted from each higher temperature spectrum. The difference spectra highlight the spectral features that are most sensitive

to the RNA's conformational changes. Consequently, these serve as the probe wavenumbers in subsequent experiments.

The sensitivity of IR spectroscopy to monitor RNA hairpin structure is immediately evident in the spectra shown in **Figures 4.1** and **4.2**. The unfolded IR spectra (i.e., the high temperature spectra) for all the oligomers are very similar. In the absence of secondary structural contacts the nucleotides can be treated as independent monomers separated by the ribose-phosphate backbone[1]. The IR spectrum of an unfolded polynucleotide is the sum of the IR spectra for each of the individual monomers weighted by their fractional compositions (refer to **Figure 2.2**). In contrast, the folded spectrum of each hairpin is quite distinct reflecting the structural contacts present.

As the temperature is raised, the RNA unfolds and the local environment surrounding the involved molecular groups is altered. This, in turn, changes their vibrational characteristics which manifest as changes in the IR spectra including intensity changes, band shifts, changes in bandwidths, or combinations thereof. In a folded hairpin, the C6=O6 stretch of guanine occurs near 1685 $cm^{-1}$. As it unfolds, this transition shifts to 1669 $cm^{-1}$ and increases in intensity. The two low-frequency peaks around 1564 and 1577 $cm^{-1}$ are a doublet due to a C=N ring vibration of guanine. Both red shift and increase in intensity as the hairpin unfolds. These changes manifest in the difference spectra as a strong feature at 1574 $cm^{-1}$. When the hairpin unfolds, base pairing and stacking interactions are disrupted. The proximity of the IR active group to the structural perturbation determines its relative sensitivity. Consequently, the spectral changes at 1574 $cm^{-1}$ and 1669 $cm^{-1}$ are weighted contributions from both pairing and stacking interactions. For example, the guanine C6=O6 group is directly

involved in a Watson-Crick base pair with cytosine. Therefore, changes at 1669 $cm^{-1}$ are primarily due to base pairing with a much smaller contribution from base stacking. In contrast, the changes at 1574 $cm^{-1}$ (due to guanine ring modes) are more sensitive to base stacking and, to a lesser extent, base pairing.

The C6=O6 group on guanine is directly involved in base pairing with cytosine. As the RNA unfolds electron density is redistributed on the guanine changing the dipole moment of the bond resulting in the observed IR changes. The changes are predominately due to the stem bases involved in pairing. Because the sequences in this study are small (only 8 bases), there is little ambiguity in these spectral assignments (by comparison, interpreting the IR spectra for a large tRNA, comprised of 76 bases, is somewhat more tenuous). In this case, the hairpin octamers contain only two base pairs in the stem. This restricts the changes observed in the difference spectra to these specific base pairing interactions and serves as a very specific indicator of loop closure.

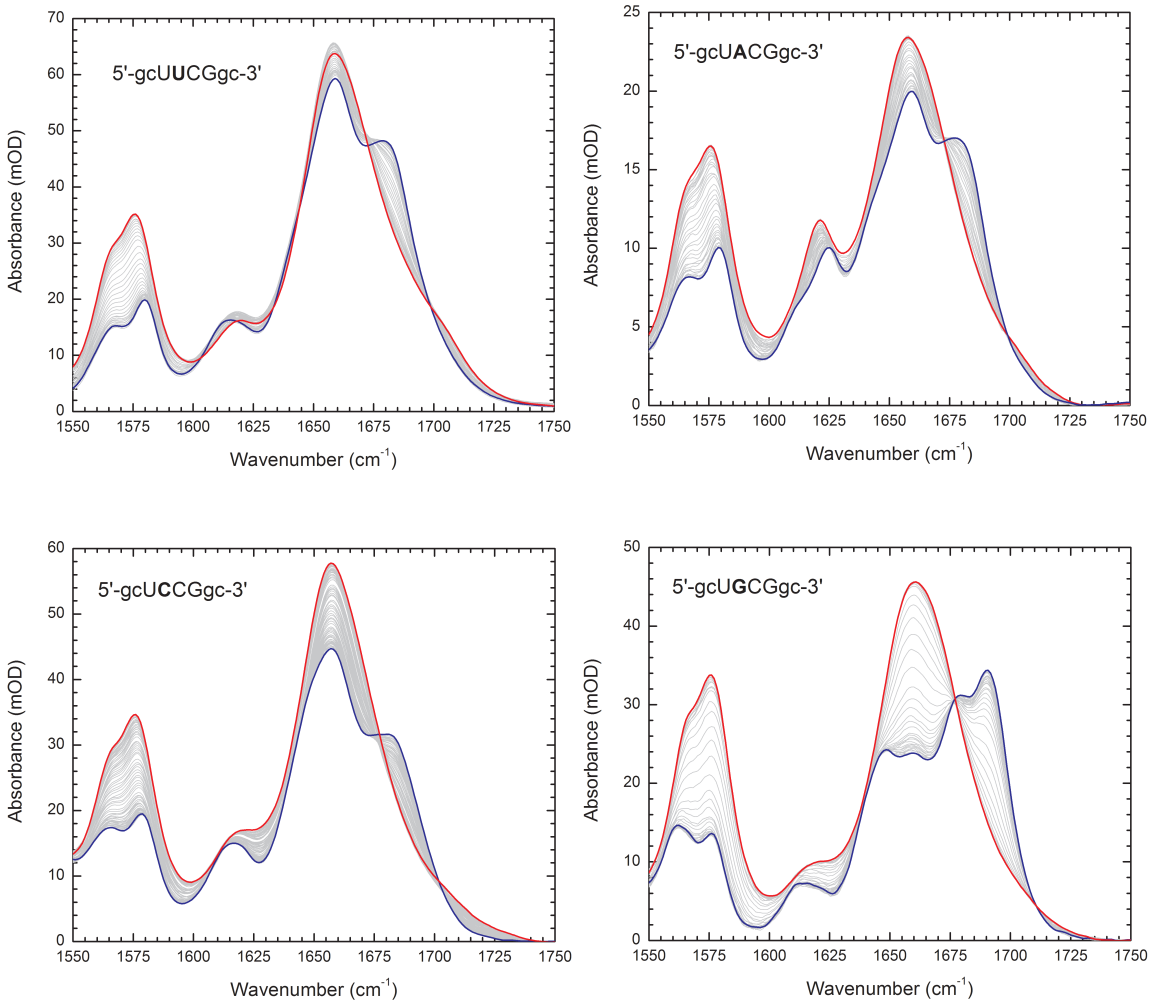


**Figure 4.1 Equilibrium IR melting spectra of 5'-gcUNCGgc-3' RNA hairpins**

Here N = U, A, C, and G as indicated. Spectra are collected for progressively higher temperatures at thermal equilibrium. The blue lines indicate the lowest temperature spectra at ~20 °C, and the red lines are the highest temperature spectra at ~90 °C.

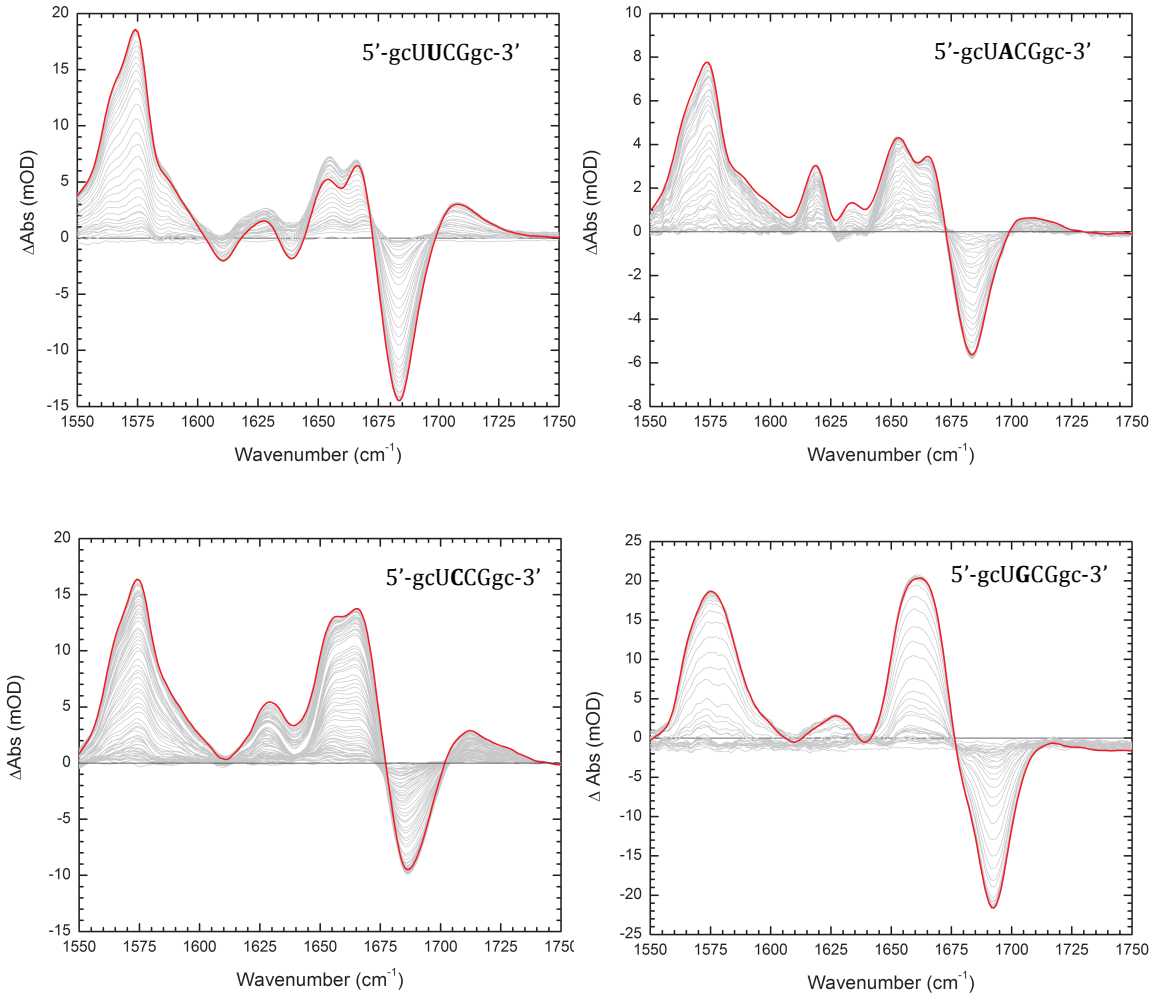


**Figure 4.2: Thermal difference spectra for the UNCG hairpins**

Difference spectra for the N = U, A, C, and G specific hairpin tetraloops (as labeled). Changes in the spectra report on specific structural transitions within the unfolding RNA hairpin. The transitions at these particular wavenumbers are then probed kinetically.

Previous Raman and NMR studies have confirmed that the oligomers all form A-type double helical stem structures, and that the N position of the tetraloop is effectively positioned out of the loop [2, 3]. This minimizes the N bases contribution to base pairing and stacking in the folded structure, but it is involved in single stranded stacking interactions during folding. Substituting a base that reports primarily on stacking (ring modes), or alternatively base pairing (stretching modes) allows these interactions to be directly interrogated during folding.

Unique among the hairpins studied is the UACG tetraloop. As the only oligomer that contains an adenine, it has a corresponding spectral feature around 1618 $cm^{-1}$ that is not observed in the other samples. This is a strong transition due to the C=C and C=N ring vibrations. When RNA folds, this band blue shifts and decreases in intensity. This transition is potentially very useful since it is "loop specific". In all the other oligomers, the base substituted in the N position of the loop is already present in the sequence. However, in the UACG tetraloop, the A in the loop is the only adenine present in the sequence. As a result, any observed spectral changes are due to structural changes affecting the ring vibrations on this base within the loop. This localized "probe" can help determine if the stem and the loop undergo simultaneous (two state) melting, or if they have different thermodynamics. In the time-resolved limit this adenine reports on loop folding dynamics, as distinguished from the stem, and can help resolve the different kinetic phases.

Before continuing, I should note that the spectral assignments just discussed are by no means absolute. For example, the transition near 1575 $cm^{-1}$ was described as an indicator of base stacking interactions. This does not mean that it reports on base

stacking only and is completely unaffected by base pairing. Because the molecular groups are connected by C–C bonds, when a guanine forms a base pair, the change in the dipole moment of the C=O group will be felt by the ring vibrations. As a result, the transition around 1575 $cm^{-1}$ could still change even if no stacking is occurring. However, this transition is most sensitive to stacking interactions and less sensitive to base pairing. The same is true for the base pairing transitions that were discussed. As a final comment, it is important to keep in mind that RNA folding is a concerted process and that base stacking and base pairing interactions probably do not occur independently of each other. This is just another piece of the complex folding energy landscape.

**UNCG tetraloop thermodynamics**

Thermodynamic information is obtained from the temperature dependent IR spectra by first plotting the absorbance values at specific wavenumbers (corresponding to the peaks in the difference spectra shown in **Figure 4.2**) as a function of temperature. A representative "melting curve" was shown in **Figure 2.4**. The shape of the curve is sigmoidal which suggests that these small hairpins exhibit two-state folding behavior in the equilibrium limit. This is typical and has been previously reported by multiple groups [4-6]. Because the IR spectra are recorded at thermal equilibrium, it is possible to extract thermodynamic information from the melting data. Here the data are analyzed according to a two-state model as described below. Absorbance is proportional to the fraction of unfolded hairpins, and the fraction of unfolded hairpins is related to the equilibrium constant by the following equation:

$$f_U = \frac{K_{eq}}{1 + K_{eq}}$$

where $f_U$ is the fraction unfolded, and $K_{eq}$ is the equilibrium constant for folding. The equilibrium constant is related to the Gibbs free energy, $\Delta G$, by the following equation:

$$K_{eq} = \exp\left(-\Delta G/RT\right)$$

Using this relation the enthalpy of unfolding ($\Delta H$), melting temperature ($T_m$), and change in heat capacity ($\Delta C_p$) are introduced by noting that the temperature dependence of the Gibbs free energy is given by the Gibbs-Helmoltz relationship:

$$\Delta G(T) = \Delta H(T)\left(1 - \frac{T}{T_m}\right) + \Delta C_p\left(T - T_m - T\ln\frac{T}{T_m}\right)$$

For this work, we held $\Delta C_p$ equal to zero to reduce the number of parameters in the thermodynamic model. Although this sacrifices some accuracy, it increases the reliability of the fit substantially, and the same conclusions can still be drawn. By combining the above relationships, we can fit the raw melting data to the following model that relates absorbance to the relevant thermodynamic variables:

$$A(T) = a + (a - b)\frac{e^{-\Delta G(T)/RT}}{1 + e^{-\Delta G(T)/RT}}$$

In the above equation, *a* and *b* are constants giving the lower and upper absorption limits, respectively. After the thermodynamic fits were obtained, the data sets were converted to fraction unfolded. This effectively normalizes the melt curves and facilitates comparison between different samples.

*Melting curves*

Normalized melt curves for all four tetraloops at each probe wavenumber are shown (**Figure 4.3)** along with the best fits to the two-state model (solid lines). For each hairpin, at least two transitions were probed (one corresponding to base pairing; ~1660 $cm^{-1}$ and one to base stacking; ~1574 $cm^{-1}$). However, for the UACG tetraloop, it was possible to probe a third transition (corresponding to the adenine ring vibration at (~1625 $cm^{-1}$). As mentioned previously, all the curves (i.e., for all samples at all probe wavenumbers) are sigmoidal which is the hallmark of two-state behavior. In addition, it bears mentioning that these experiments have been repeated multiple times and the results are highly reproducible. Error bars are not included in the plots because they would be the same size (or smaller) than the symbol diameters. The fit parameters (melting temperatures and unfolding enthalpies) are tabulated in **Table 4.1**.

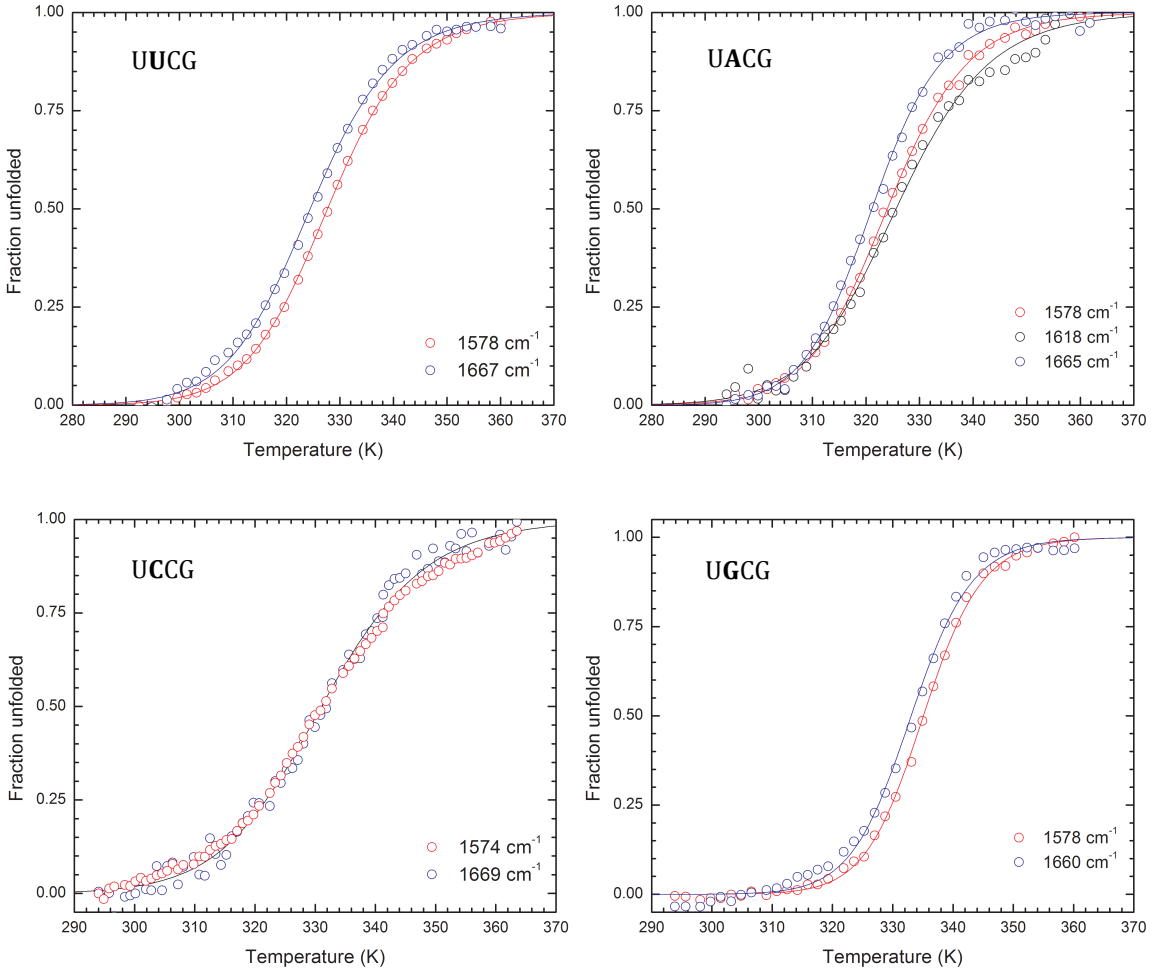


**Figure 4.3: Wavenumber specific IR melting curves**

Temperature dependent melting curves for the N = U, A, C, and G hairpin loop sequences. Melting behavior at different wavenumbers reporting on base stacking and base pairing interactions are shown. In addition, the graph for the N = A sequence shows an additional stacking interaction at 1618 $cm^{-1}$.

Perhaps the most noteworthy feature is that the melt curves recorded at different wavenumbers *do not* overlap for the N = U, A, and G tetraloops. By comparison, the curves are virtually indistinguishable for N = C. This suggests that the thermodynamics (via the two-state model) are *wavenumber dependent*, for the N = U, A, and G tetraloops. Looking at the values in **Table 4.1**, we see that the unfolding enthalpies are the same (for each tetraloop) and that the differences are from different melting temperatures.

| $T_m$ (°C)/oligo | U***A***CG | U***U***CG | U***C***CG | U***G***CG |
|---|---|---|---|---|
| 1575 $cm^{-1}$ | 50.8 ± 0.17 | 54.5 ± 0.10 | 57.2 ± 0.30 | 61.9 ± 0.12 |
| 1618 $cm^{-1}$ | 52.5 ± 0.57 | - | - | - |
| 1665 $cm^{-1}$ | 47.9 ± 0.23 | 51.3 ± 0.28 | 57.7 ± 0.34 | 59.9 ± 0.26 |
| $\Delta H_U$ (kJ/mol) | 116.3 ± 4.7 | 113 ± 2.7 | 104.6 ± 3.8 | 187 ± 6.1 |

**Table 4.1: Thermodynamic unfolding parameters for the UNCG hairpin loops**
Melting temperatures and enthalpies of unfolding for the series of oligonucleotides. Melting temperatures at different wavenumbers corresponding to specific structural interactions (pairing and stacking) are reported.

## Discussion

Overall, the thermodynamics reported in **Table 4.1** are in excellent agreement with known values for these hairpins [2, 4, 5, 7-10]. In this regard, our results are not particularly interesting and merely show that the samples are behaving as expected.

However, the true power of using IR spectroscopy to study RNA structure and dynamics lies in the ability to simultaneously probe multiple (intrinsic) vibrational transitions (each corresponding to different structural features) within a sample. The significance of our results becomes evident in the wavenumber dependence of the melting curves observed for the N = U, A, and G tetraloops. In effect, we are able to monitor a single event (unfolding) from different "perspectives". This would be analogous to filming a sporting event using multiple cameras placed at strategic positions along the playing field. While each camera captures the same event, the combined information from multiple cameras paints a far more detailed (and accurate) picture.

What these data suggest is that, even in the equilibrium limit, the folding of RNA hairpins is far more complex than previously thought. Most notably, the folding is not, strictly speaking, "two-state". Since a purely two-state transition would give the same thermodynamic parameters regardless of the probe wavenumber, the wavenumber dependence of the thermodynamics is strong evidence of a more involved folding mechanism.

What constitutes a "complex folding mechanism"? Generally speaking, folding complexity is a consequence of two possible phenomena. One is that folding proceeds sequentially along a single path but is interrupted by one or more intermediates (local minima in the energy landscape perspective). Another is a parallel mechanism where folding proceeds along at least two unique pathways that converge at the same folded state. In a parallel mechanism, each pathway can, in turn, be sequential. Our data seems to indicate a sequential mechanism according to the following argument.

Since enthalpy is a state function, its value depends only on the initial and final states of the system (i.e., it is independent of the pathway between the two states). For all the hairpins studied, the enthalpies were wavenumber independent which suggests that the folded and unfolded states were the same and therefore connected by a single pathway (note that this is not absolutely conclusive since it is possible for different states to be energetically degenerate). The support for the sequential mechanism lies in the wavenumber dependence of the melting temperatures which argues that different structural features (e.g., base pairs and base stacks) melt at different temperatures. Specifically, the base pairs "melt" (i.e., are disrupted) at a lower temperature than the base stacks for the N = U, A, and G tetraloops. In other words, as the temperature is increased, base pairs are disrupted and the loop begins to open up resulting in a structure (possibly an extended chain) that retains some of its native stacking contacts.

Substituting adenine into the N position allows loop specific structural interactions and thermodynamics to be interrogated (C=N ring vibrations). When the UACG tetraloop is fully folded, adenine is positioned out of the loop and does not experience pairing or stacking interactions [3]. As the hairpin stem unfolds and base pairs in the stem melt (47.9°C) the loop stays intact. Increasing the temperature another few degrees will finally melt the loop (52.5°C) and unfold the hairpin. This suggests that the adenine does not experience significant changes in its "stacking environment" until the stem first unzips. The adenine begins to feel stacking interactions when the loop melts, and it can then rotate in toward the stem bases. In the unfolded structure it forms single stranded stacking contacts which are likely stronger in the loop region. These forces help the strand collapse as the strand begins to fold.

## Conclusion

Although the motivation for performing a detailed thermodynamic analysis was to prepare for the subsequent time resolved experiments, the results were surprisingly significant in their own right. In addition to providing the information necessary to carry out meaningful kinetic experiments, the work presented in this Chapter offers considerable insight into the folding mechanism of RNA hairpin tetraloops. Even in the equilibrium limit, our results suggest that RNA hairpin loop folding is a complex process that does not proceed via a simple two-state mechanism. These results are further substantiated by the kinetic experiments that will be described in the next two chapters.

## Experimental

### *Sample Preparation and spectroscopy*

HPLC purified RNA oligonucleotides were purchased from Integrated DNA Technologies. Samples were prepared at a concentration of 2 mM (16 mM nucleotide) in 100 mM phosphate buffer at a pH of 7.2. The concentration of oligomer was sufficiently low to avoid duplex formation. Each spectrum was collected at 2 $cm^{-1}$ resolution and averaged over 512 scans. In order to minimize baseline drift sample and reference spectra were recorded alternately by translating the cell back and forth between scans. Experimental details are described in depth in Chapter 2 and the appendices.

**Bibliography**

[1] G.J. Thomas Jr., *Determination of the base pairing content of ribonucleic acids by infrared spectroscopy*. Biopolymers, 7 (**1969**) 325-334.

[2] V. Baurmruk, C. Gouyette, T. Huynh-Dinh, J. Sun and M. Ghomi, *Comparison Between CUUG and UUCG Tetraloops: Thermodynamic Stability and Structural Features Analyzed by UV Absorption and Vibrational Spectroscopy*. Nucleic Acids Research, 29 (**2001**) 4089-4096.

[3] M. Abdelkafi, N. Leulliot, V. Baumruk, C.H. du Penhoat, O. Lampire, N. Bouchemal-Chibani, P. Goyer, A. Namane, C. Gouyette, T. Huynh-Dinh and L. Bednarova, *Common structural features of UUCG and UACG tetraloops in very short hairpins determined by UV absorption, Raman, IR, and NMR spectroscopies*. Journal Biomolecular Structure and Dynamics, 14 (**1997**) 579-593.

[4] M. Molinaro and I. Tinoco Jr., *Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: Thermodynamics and spectroscopic applications*. Nucleic Acids Research, 23 (**1995**) 3056-3063.

[5] M. Abdelkafi, N. Leulliot, V. Baumruk, V. Bednarova, P.Y. Turpin, A. Namane, C. Gouyette, T. Huynh-Dinh and M. Ghomi, *Structural features of the UCCG and UGGC tetraloops in very short hairpins as evidenced by optical spectroscopy*. Biochemistry, 37 (**1998**) 7878-7884.

[6] V. Baumruk, C. Gouyette, T. Huynh-Dinh, J. Sun and M. Ghomi, *Comparison between CUUG and UUCG tetraloops: Thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy*. Nucleic Acids Research, 29 (**2001**) 4089-4096.

[7] D.J. Williams, Boots, J.L., and Hall, K.B., *Thermodynamics of 2'-Ribose Substitutions in UUCG Tetraloops*. RNA, 7 (**2001**) 44-53.

[8] K. Sarkar, K. Meister, A. Sethi and M. Gruebele, *Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe*. Biopysical Journal, 97.

[9] D. Proctor, H. Ma, E. Kierzek, R. Kierzek, M. Gruebele and P.C. Belvilacqua, *Folding thermodynamics and kinetics of YNMG RNA hairpins: Specific incorporation of 8-bromoguanosine leads to stabilization of the folding rate*. Biochemistry, 43 (**2004**) 14004-14014.

[10] V.P. Antao, S.Y. Lai and J. Tinoco, I., *A thermodynamic study of unusually stable RNA and DNA hairpins*. Nucleic Acids Research, 19 (**1991**) 5901-5905.

# Chapter 5

## Investigating the complex folding dynamics of UNCG hairpin loops with time-resolved IR spectroscopy

## Chapter overview

The main purpose of this chapter is to present the time-resolved kinetic relaxation data for the UNCG hairpin loops that were thermodynamically characterized in Chapter 4. First, the chapter will give a brief background on fast techniques and what has been done previously. Next, the chapter will develop the theory and application of time-resolved infrared spectroscopy and its utility for investigating the folding kinetics of fast folding, small RNA hairpins. Time-resolved kinetic relaxation data are presented for each UNCG hairpin. Multiple vibrational transitions reporting primarily on base stacking and base pairing interactions are interrogated. The kinetic results for each sequence are examined, and compared for different probe wavenumbers on the same molecule (and to the thermodynamic results). In addition, the hairpins will be contrasted across sequences and the inherent similarities and differences (i.e. complexities) expounded upon. In Chapter 6, a more detailed account of the folding mechanism for the UCCG hairpin variant will be given.

## Introduction

The first study to use IR spectroscopy to monitor RNA folding subsequent to a laser induced temperature jump was published by Brauns and Dyer in 2005 [1]. The technique is particularly powerful because IR spectroscopy is structurally specific and relies on intrinsic properties of the RNA, and the laser induced T-jump provides access to fast time scales. In the past, the study of fast kinetics was hindered by a lack of experimental methods that could monitor RNA conformations at sub-microsecond time scales with nucleotide resolution [2, 3]. Although time-resolved IR spectroscopy applied

to RNA folding is a novel technique, its application to protein folding has been well-established for some time [4].

The paper by Brauns and Dyer used the technique to investigate the folding kinetics of phenylalanine specific transfer RNA ($tRNA^{phe}$), a relatively large RNA molecule comprised of 76 nucleotides. In addition to demonstrating the utility of time-resolved IR spectroscopy to study RNA folding, the work uncovered previously unknown features in the folding mechanism of $tRNA^{phe}$. In their work, two vibrational transitions were probed; one at 1620 $cm^{-1}$ and the other at 1661 $cm^{-1}$. The former transition is assigned to a ring vibration of adenine and reports on the dynamics of AU base pairing interactions, while the latter is assigned to the C6=O6 stretching of guanine and reports on the dynamics of GC base pairing interactions. They observed three distinct kinetic phases at each investigated vibrational transition and found that the kinetics monitored at 1620 $cm^{-1}$ were slightly faster than those monitored at 1661 $cm^{-1}$.

In addition, the equilibrium data showed that the thermodynamics were slightly different at each wavenumber. Taken together, their results suggested that $tRNA^{phe}$ folds via a multi-step, parallel pathway mechanism [5]. From this work, time-resolved IR spectroscopy became an established method in the cache of tools available to RNA folding research. However, despite its advantages, its implementation requires a high level of experimental expertise (see Chapter 2) and we are the only group using it. As a result, we occupy a unique niche in the global effort to understand RNA folding.

In this chapter, the recently established technique of time-resolved IR spectroscopy is used to investigate the folding kinetics of a series of small RNA hairpin tetraloops. As discussed in previous chapters, the UNCG class of tetraloops are fundamental

components of all RNAs and there are unanswered questions about how they fold. For example, the characteristic time scale of end-to-end collisions for small hairpins has been shown to be on the order of nanoseconds [6, 7]. However, the formation of loops occurs on time scales ranging from 10s to 100s of microseconds—several orders of magnitude slower. A question that naturally arises is, "why is hairpin formation so slow?" We are interested in why loop formation is so much slower than the initial collapse, and what forces are responsible for stabilizing the folding loop. Time-resolved IR spectroscopy is extremely well-suited to address this question. The combination of fast time resolution with multiple probe wavelengths allows us to obtain a more detailed picture of the folding mechanism.

**Experimental approach**

Detailed experimental procedures are provided in Chapter 2 and in the Appendices; only an overview of the approach is provided here. The equilibrium IR spectroscopy discussed in Chapter 4 was done to characterize the thermal melting behavior of the UNCG hairpin loops so that meaningful kinetics experiments could be performed. By referencing this thermodynamic data (fraction unfolded and melting temperatures for key wavenumbers) the initial equilibrium temperatures and T-jump magnitudes can be determined. The initial temperature is chosen so that the RNA is partially unfolded (~10% unfolded). After equilibrating at this initial temperature, the thermal equilibrium is rapidly perturbed with a T-jump pulse. The magnitude of the T-jump is varied in ~2.5 degree increments until the final temperature is equal to the melting temperature (50% unfolded).

In this way, a series of kinetic profiles is obtained at different final temperatures (i.e., the temperature at which relaxation occurs). From these data, the temperature dependence of the kinetics is measured and activation characteristics can be determined. The same vibrational bands that were characterized in the equilibrium spectroscopy (Chapter 4) were probed kinetically. An experiment is performed by recording the transient absorptions of a sample and a reference. By taking the difference between these (i.e., sample minus reference) the transient absorption profile for the RNA is obtained.

**Data handling**

The raw kinetic traces for three of the UNCG hairpin loops (N = U, A, and G) are shown in **Figures 5.1, 5.2, and 5.3**. Multiple curves are shown for each representing different T-jump magnitudes. The data for N = C is treated separately in the next chapter. The solid red curves through the data in **Figures 5.1** through **5.3** are fits to the following exponential model:

$$\Delta A(t) = a_1\, e^{-t/\tau_1} + a_2\, e^{-t/\tau_2} + y_o(t) \qquad (5\text{-}1)$$

Here, $\tau_1$ and $\tau_2$ are the relaxation constants (i.e., characteristic relaxation times) and the corresponding amplitudes are $a_1$ and $a_2$. The third term, $y_o(t)$, accounts for the exponentially decaying background due to cooling of the heated volume. Although this has the form of an exponential, the associated parameters report on a characteristic of the solvent and have nothing to do with the relaxation of the RNA. All fit parameters

(with the exception of the background terms) are summarized graphically in **Figures 5.4**, **5.5**, and **5.6**. The solid lines through the time constants in **Figures 5.4 a**, **5.5 a**, and **5.6 a** are fits to the following equation:

$$\tau = \tau_o \exp\left[\frac{E_a}{R}\left(\frac{1}{T} - \frac{1}{T_o}\right)\right] \qquad (5\text{-}2)$$

where $T$ is the temperature, $R$ is the gas constant, $E_a$ is the activation energy, $T_o$ is a reference temperature, and $\tau_o$ is the relaxation constant at the reference temperature. Since it is somewhat arbitrary, we use the melting temperature for each sample as the reference temperature. Equation 5.2 is found by integrating the Arrhenius equation:

$$\frac{d\ln k}{dT} = \frac{E_a}{RT^2} \qquad (5\text{-}3)$$

and converting from rate constant ($k$) to relaxation constant ($\tau$) using $\tau = 1/k$. Although we do not report activation energies, Equation 5.2 shows that a negative slope corresponds to a positive activation energy (i.e., an "activated process") while a positive slope corresponds to a negative activation energy. The solid lines through the amplitudes in **Figures 5.4** through **5.6** only serve as a guide for the eye—a formal linear relationship is not implied.

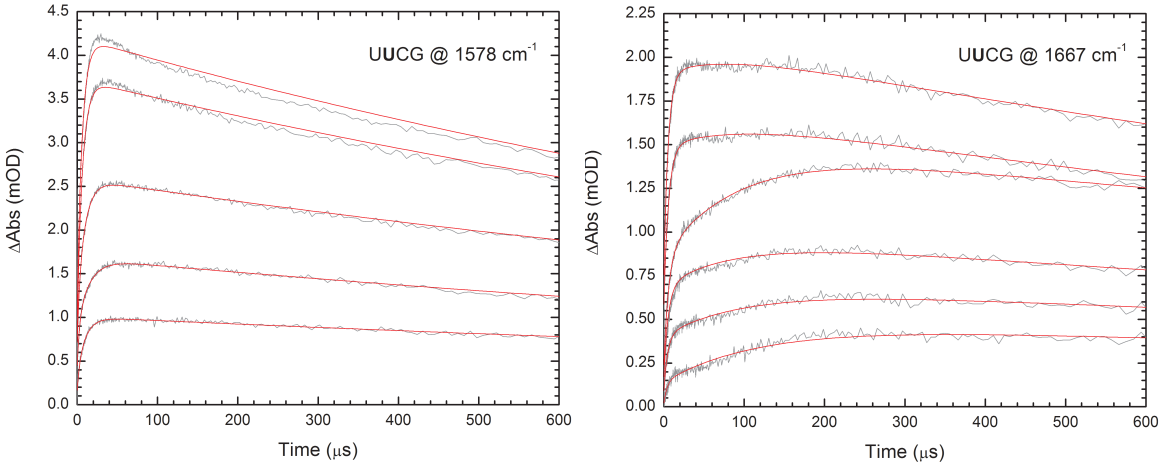


**Figure 5.1: Kinetic relaxation data for 5'-gcUUCGgc-3'**

Transient absorbance data collected in 2.5 degree increments starting below and T-jumping up to the melting temperature. Probed at 1578 $cm^{-1}$ ($T_m$ = 55°C), and 1667 $cm^{-1}$ ($T_m$ = 51°C). The solid red curves are fits to equation 5.1.

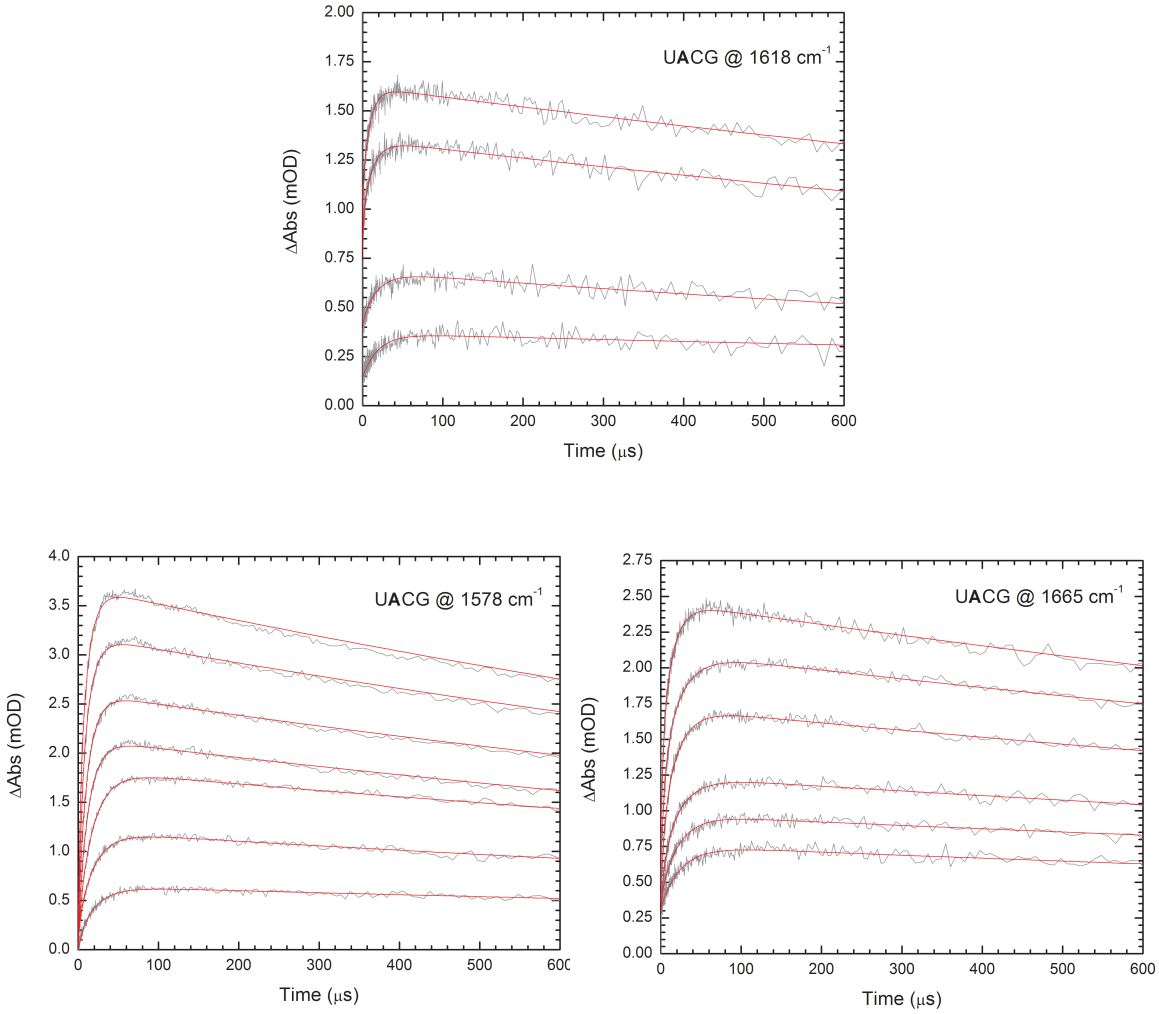


**Figure 5.2: Kinetic relaxation data for 5'-gcUACGgc-3'**

Transient absorbance data collected in 2.5 degree increments starting below and T-jumping up to the melting temperature. Probed at 1578 cm$^{-1}$ ($T_m$ = 51°C), 1618 cm$^{-1}$ ($T_m$ = 53°C), and 1665 cm$^{-1}$ ($T_m$ = 48°C). The solid red curves are fits to equation 5.1.

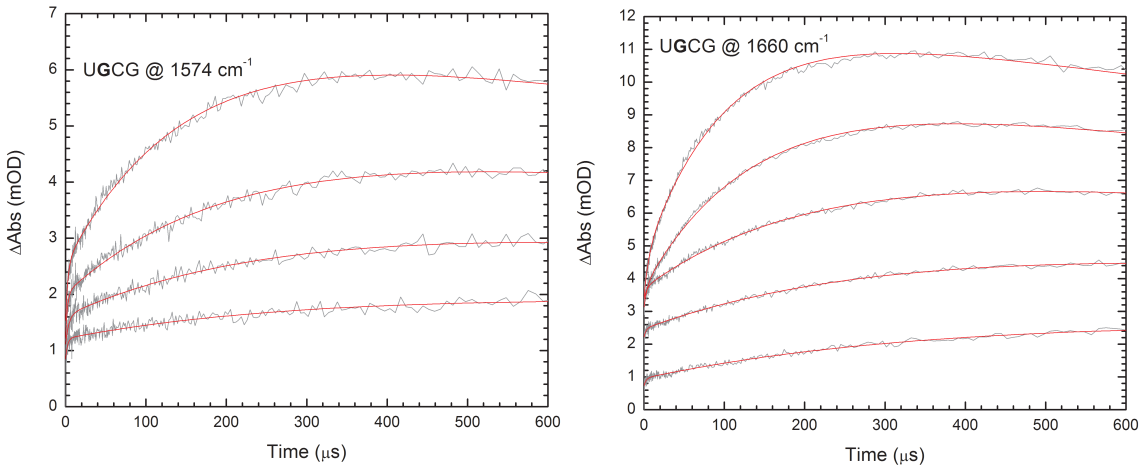


**Figure 5.3: Kinetic relaxation data for 5'-gcUGCGgc-3'**

Transient absorbance data collected in 2.5 degree increments starting below and T-jumping up to the melting temperature. Probed at 1574 $cm^{-1}$ ($T_m$ = 62°C), and 1660 $cm^{-1}$ ($T_m$ = 60°C). The solid red curves are fits to equation 5.1.

## Results

The UUCG hairpin (**Figure 5.1**) showed kinetic phases that were well resolved (in time) for base stacking (1574 $cm^{-1}$) and base pairing interactions (1667 $cm^{-1}$). The activation plots are shown in **Figure 5.4 a**. At higher temperatures, the fast characteristic relaxation components for base stacking and base pairing transitions were 1 and 5 μs respectively. Base pairing was five times slower and the data suggested a non-activated process with a likely entropic barrier. Even at lower temperatures, the slow relaxation component of each transition were separated by nearly an order of

magnitude with the stacking interactions relaxing in 15 μs and the base pairing relaxation taking 125 μs. In this case the separation in phases is pronounced and both transitions are activated processes. The relaxations become faster at higher temperatures where the separation of phases is larger than an order of magnitude.

The normalized amplitudes are plotted in the **Figure 5.4 b** and **c**. As temperature increases the amplitude $a_1$ for base stacking interactions, probed at 1574 $cm^{-1}$, decreased by 25% for $\tau_1$. This suggests that the fast component for the stacking transition becomes less significant with increasing temperature. Alternately, the amplitude $a_2$ of the slow component for base stacking interactions increases by 25%. In terms of the base pairing interactions probed at 1667 $cm^{-1}$ the opposite behavior was observed. The amplitude for the fast component ($\tau_1$) increases by 30%, and the slow component ($\tau_2$) decreases substantially by nearly 65%. Summing the normalized amplitudes (fraction of total amplitude) should result in a total amplitude of 1. Adding the amplitudes for stacking interactions and pairing interactions actually gives values of 0.8 and 1 respectively. The fact that the amplitudes for the interactions probed at 1574 $cm^{-1}$ do not sum to 1 indicates that there is an unresolved kinetics component in the data. This is observed as a fast sub-100 ns component in the raw data, and is substantiated by the normalized amplitude plots.

The most noteworthy feature of the UACG kinetics data (**Figure 5.2**) is the loop specific adenine stacking interaction (at 1618 $cm^{-1}$). This interaction showed the fastest relaxation kinetics of any of the other phases for all the other hairpin loops. The activation plots are shown in **Figure 5.5 a**. The fast component relaxes in 1 μs and at

the highest temperature relaxes in only 700 ns. This indicates that stacking interactions within the loop respond very fast in relation to the unfolding strand.

The fast component for guanine base stacking and base pairing relax five times slower (~5 μs) and are separated by several microseconds. Again, base stacking in the stem is faster than base pairing interactions, but only twice as fast at the highest temperature. The slow components for all transitions relax in 25 μs indicating that although the phases are initially well resolved, they converge at longer times. Overall, the UACG hairpin loop relaxes faster than the other sequences in only 10 μs.

The normalized amplitudes are plotted in **Figure 5.5 b, c** and **d**. As temperature increases, the amplitude $a_1$ for base stacking interactions, probed at 1578 $cm^{-1}$, decrease slightly by 10%. The amplitude $a_2$ stays constant for the slow component. Summing the amplitudes for this transition gives a value of 1, which accounts for the total observed kinetics. In this case, the amplitudes for transitions at 1618 $cm^{-1}$ and 1665 $cm^{-1}$ behave similarly. The $a_1$ values for both increase as a function of temperature by approximately 20%, and the $a_2$ values both decrease by approximately 25%. In both cases a fraction of the total amplitude is missing indicating a fast, unresolved component. In the data probed at 1618 $cm^{-1}$, the amplitudes sum to 0.5 indicating that a considerable portion of the kinetics is fast and unresolved. The amplitudes for the data probed at 1665 $cm^{-1}$ sum to 0.7 and show a missing kinetic phase 20% smaller than the data at 1618 $cm^{-1}$ (in the loop).

The slowest relaxation kinetics were observed for the UGCG tetraloop sequence (**Figure 5.3**), which relaxed on a time scale nearly five times slower than the next slowest phase on the N = U sequence. The activation plots are shown in **Figure 5.6 a**.

The relaxation constants were essentially the same at both vibrational bands for this sequence, although the fast and slow relaxation components were separated by over two orders of magnitude. This was the largest separation seen between the fast and slow phases. The fast components for both the base pairing and stacking interactions relax on the order of a few microseconds. The fast kinetic phase attributed to base pairing becomes five times slower with increasing temperature indicating an entropic barrier. The slow component for both interactions relaxes in 500 μs and become faster with increasing temperature, indicating an activated process. These results agree with the equilibrium data for the UGCG sequence, which showed that the hairpin loop was thermodynamically more stable and folding was more cooperative.

The normalized amplitudes are plotted in **Figure 5.6 b** and **c**. As temperature increases the amplitude $a_1$ for stacking interactions probed at 1574 $cm^{-1}$ stays essentially the same. On the other hand, the amplitude $a_2$ for the slow component increases by 30%, and contributes more at higher temperatures. Interestingly, the sum of the amplitudes is 0.6 at the lowest temperature and nearly 1 at the highest temperature. This suggests that as temperature is increased the unresolved kinetics phase present at lower temperatures becomes observable, and can be resolved in the kinetics data. The amplitudes for the base stacking interactions probed at 1660 $cm^{-1}$ do not show a measureable temperature dependence. The amplitude $a_1$ for the fast component is essentially the same at all temperatures, and likewise the amplitude $a_2$ for the slow component decreases by much less than 1%. Summing the fractional amplitudes for the stacking interactions gave a total amplitude of 0.6, while the sum of amplitudes for the pairing interactions is equal to 0.7. Again, in both cases a portion of

the kinetics remains unresolved and is on the same order of magnitude for both of the probe wavenumbers. Interestingly, the magnitude of the missing kinetics for the base pairing interactions is temperature independent and is 30% across the entire temperature range.

The UCCG hairpin variant is discussed in detail in the next chapter. However, the results will be briefly summarized here so that comparisons can be made with the other tetraloops. Similar to the other sequences, the UCCG tetraloop showed a fast sub-100 ns relaxation component that was not resolved. The kinetics for base stacking (1574 $cm^{-1}$) were bi-exponential at lower temperatures with relaxation constants of 3.6 and 15.6 μs. At higher temperatures, the relaxation became single exponential with a time constant of 3.1 μs. In contrast, the kinetics for base pairing (1669 $cm^{-1}$) were bi-exponential over the entire temperature range that was investigated. The fast relaxation at this wavelength was non-activated and had a time constant of 3.7 μs, while the time constant for the slow component went from 34.7 μs at low temperatures to 14.9 μs at higher temperatures.

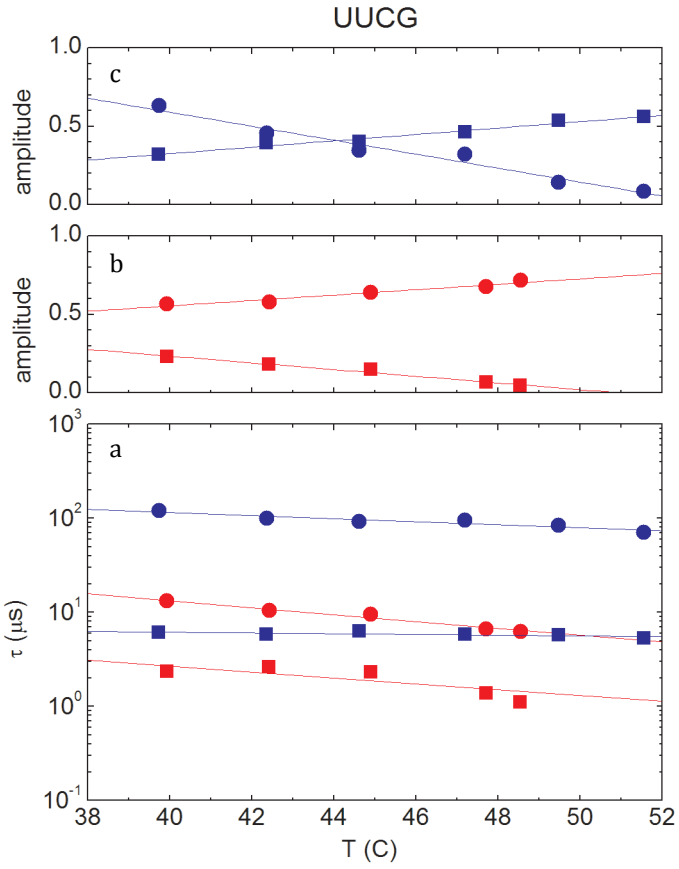


**Figure 5.4: Activation plots and normalized amplitudes for 5'-gcUUCGgc-3'**

Red data are for interactions probed at 1578 cm$^{-1}$ and blue data are for interactions probed at 1667 cm$^{-1}$. Squares are for the fast kinetic phase $\tau_1$ and $a_1$. Circles are for the slow kinetic phase $\tau_2$ and $a_2$. (a) Activation plots. (b) Normalized amplitudes for 1574 cm$^{-1}$ (c) Normalized amplitudes for 1667 cm$^{-1}$.

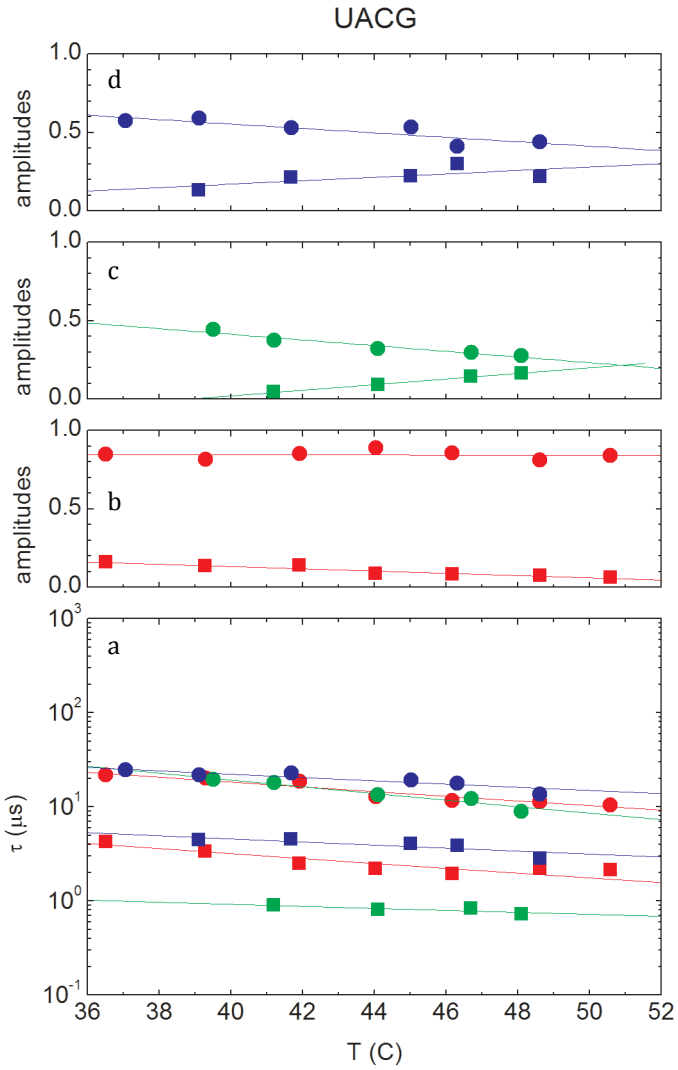


**Figure 5.5: Activation plots and normalized amplitudes for 5'-gcUACGgc-3'**

Red data are for interactions probed at 1578 $cm^{-1}$, green data are for interactions interrogated at 1618 $cm^{-1}$, and blue data are for interactions probed at 1665 $cm^{-1}$. Squares are for the fast kinetic phase $\tau_1$ and $a_1$. Circles are for the slow kinetic phase $\tau_2$ and $a_2$. (a) Activation plots. (b) Normalized amplitudes for 1578 $cm^{-1}$. (c) Normalized amplitudes for 1618 $cm^{-1}$. (d) Normalized amplitudes for 1665 $cm^{-1}$.

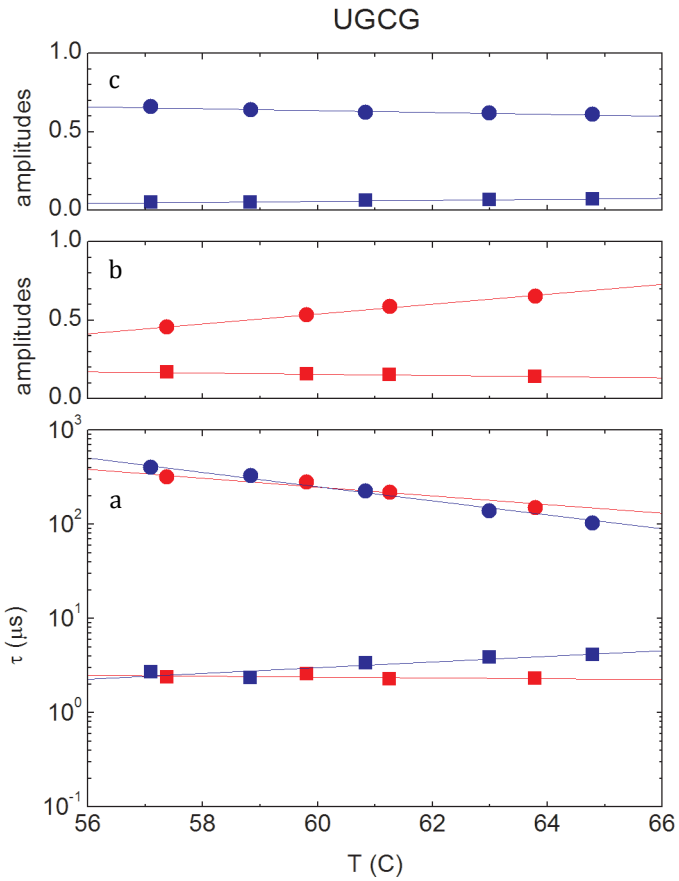


**Figure 5.6: Activation plots and normalized amplitudes for 5'-gcUGCGgc-3'**

Red data are for interactions probed at 1574 $cm^{-1}$, and blue data are for interactions probed at 1660 $cm^{-1}$. Squares are for the fast kinetic phase $\tau_1$ and $a_1$. Circles are for the slow kinetic phase $\tau_2$ and $a_2$. (a) Activation plots. (b) Normalized amplitudes for 1574 $cm^{-1}$. (c) Normalized amplitudes for 1660 $cm^{-1}$.

**Discussion and conclusions**

The equilibrium results from Chapter 4 suggested the possibility of a rugged landscape; the kinetics data presented in this chapter further substantiate those findings. These results reveal three significant features: (1) The relaxation kinetics for tetraloop formation are sequence dependent. In particular, the average rate of hairpin relaxation decreases in the order N = C, A, U and G. (2) The kinetics for each sequence are generally bi-exponential which argues that folding is non-two-state and proceeds through *at least* one intermediate. (3) The kinetics for each sequence are wavenumber dependent. Here, the kinetics for base stacking are (on average) faster than those for base pairing. In addition to validating the conclusion drawn from (2), this observation supports the sequential folding mechanism proposed in Chapter 4.

The goal of RNA folding research has traditionally been to determine the folding mechanism of RNA. However, our results argue that this approach may be overly simplistic. Even for relatively "simple" hairpin tetraloops, it is apparent that a single mechanism does not exist and each molecule must be considered separately. Furthermore, the mechanism for each RNA is quite complex making it difficult to determine. However, these assertions are significant in their own right. Armed with a greater appreciation for the true complexity of the problem, a more suitable goal can be formulated. At this stage, the underlying mechanisms for the N = A, U, and G sequences remain indeterminate and further experiments are warranted before more definitive conclusions can be drawn. As an example, the UCCG tetraloop has been explored in greater detail; these results are discussed in the next chapter.

**Bibliography**

[1] E.B. Brauns and R.B. Dyer, *Time-Resolved infrared spectroscopy of RNA folding*. Biophysical Journal, 89 (**2005**) 3523-3530.

[2] B. Sclavi, M. Sullivan, M.R. Chance, M. Brenowitz and S.A. Woodson, *RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting*. Science, 279 (**1998**) 1940-1943.

[3] D. Thirumalai and S.A. Woodson, *Kinetics of folding of proteins and RNA*. Accounts of Chemical Research, 29 (**1996**) 433-439.

[4] R.B. Dyer, F. Gai, W.H. Woodruff, R. Gilmanshin and R.H. Callender, *Infrared studies of fast events in protein folding*. Accounts of Chemical Research, 31 (**1998**) 709-716.

[5] R.B. Dyer and E.B. Brauns, *Laser-induced temperature jump infrared measurements of RNA folding*. Methods in Enzymology, (**2009**).

[6] Y. Shen, S.V. Kuznetsov and A. Ansari, *Loop dependence of the dynamics of DNA hairpins.* Journal of Physical Chemistry B., 105 (**2001**) 12202-12211.

[7] X. Wang and W.M. Nua, *Kinetics of end-to-end collision in short single-stranded nucleic acids.* Journal of the American Chemical Society, 126 (**2004**) 808-813.

# Chapter 6

## Rearrangement of partially ordered stacked conformations contributes to the rugged energy landscape of a small RNA hairpin

Stancik, A.L. and Brauns, E.B., *Biochemistry*, **2008**, 47, 10834-10840.

**Abstract**

We have studied the fast relaxation kinetics of a small RNA hairpin tetraloop using time resolved infrared spectroscopy. A laser induced temperature jump initiated the relaxation by rapidly perturbing the thermal equilibrium of the sample. We probed the relaxation kinetics at two different wavenumbers; 1574 $cm^{-1}$ and 1669 $cm^{-1}$. The latter is due to the C6=O6 carbonyl stretch of the base guanine and is a direct measure of guanine base pairing. The former is assigned to a ring vibration of guanine and tracks structure by sensing base stacking interactions. Overall, the kinetics at 1574 $cm^{-1}$ are faster than those observed at 1669 $cm^{-1}$. When relaxation occurs at the melting temperature, the kinetics at both wavenumbers are bi-exponential. When relaxation occurs at a temperature that is higher than the melting temperature, the data at 1669 $cm^{-1}$ is still bi-exponential while only a single *fast* phase is resolved in the data at 1574 $cm^{-1}$. The fast phases are in the microseconds range while the slower phases are in the tens of microseconds range. At both wavenumbers, a portion of the relaxation is not resolved indicating the existence of a very fast, sub−100 ns phase. Our results provide additional evidence that small, fast folding hairpin loops are characterized by a rugged energy landscape. Furthermore, our data suggest that single strand stacking interactions and stacking interactions in the loop contribute significantly to the ruggedness of the energy landscape. This work also demonstrates the utility of time resolved infrared spectroscopy to study RNA folding.

## Introduction

RNA folds in a hierarchical manner; small, independently stable secondary structural motifs fold first and are subsequently assembled into larger, three dimensional architectures [1-3]. Numerous types of structural motifs exist, but the hairpin loop is the most ubiquitous and is found in all RNA molecules [4, 5]. Understanding the formation of the hairpin loop is therefore an integral component of the overall RNA folding problem. A hairpin loop forms when a region of RNA folds back on itself forming a stem comprised of base pairs and a loop comprised of unpaired bases. Many loop sizes are found in nature ranging from diloops (2 unpaired bases in the loop) to tetradecaloops and larger. However, the tetraloop is by far the most common. Furthermore, tetraloops with the sequence UNCG (where N is any nucleotide) are the most common of all the tetraloops and occur at a statistically higher rate than other sequences in nature [4, 5]. The impetus for studying the folding kinetics of RNA tetraloops is clear.

The underlying folding dynamics of RNA tetraloops (and hairpins in general) are not nearly as straightforward as their apparent structural simplicity might suggest. As newer techniques are employed, increasing detail is uncovered and the true complexity is revealed. Recent evidence (experimental, theoretical, and computational) supports the notion that the energy landscape of even small hairpin loops is quite rugged and that simple two-state kinetics cannot accurately describe folding. For example, Ansari has developed a “configurational diffusion” model to describe the folding mechanism of small nucleic acid hairpins [6]. The recognition that hairpin formation was much slower that the rates of end to end collisions stressed the necessity of this model [7, 8].

Similarly, work by the Bevilacqua and Gruebele groups using mutants of well characterized tetraloops provides strong evidence in support of a rugged energy landscape [9-12]. An underlying theme in their work is that the location of the probe determines the observed folding kinetics. Essentially, populations of intermediate states are exposed by monitoring the kinetics from different locations within the molecule. A recent communication by Thirumalai and co-workers, explicitly states that multiple probes are required to explore the rugged energy landscape of RNA hairpins [13].

Folding complexity is attributed to misfolded traps and alternative populations. Both of which contribute to the ruggedness of the energy landscape. Recent theoretical work treats the non-native states as structures with incorrect or partially folded *stems*. However, it is possible that the non-native states could also be due to single stranded conformations having alternative base stacking orientations. Accounting for such states from a theoretical standpoint poses a challenge and has not yet been explicitly addressed. Since these theoretical models are often used to interpret experimental data, the contribution of single stranded intermediates has been neglected. It could also be said that the omission of single stranded intermediates from theoretical models is due to the lack of experimental evidence of their existence. We believe that the work presented here provides this evidence.

To explore the ruggedness of the folding landscape, two things are needed: (1) A method to initiate folding/unfolding on a sufficiently short time scale so that mechanistic nuances can be accessed. (2) The detection method must provide some type of structurally specific information. The first requirement can be satisfied by using

a laser induced temperature jump (T-jump) to initiate unfolding on nanosecond time scales (picosecond laser induced T-jumps are also possible). This is the method used in the experimental work cited above and is also the method that we employ in the work reported here. The second requirement is more difficult to fulfill. One way is to insert non-native probes at various locations within the molecule. This method provides localized kinetic information but is cumbersome (involving novel synthetic strategies) and may perturb the native structure to some small degree. An alternative solution is to exploit some intrinsic spectroscopic feature that is native to the molecule. For example, it is well known that UV-visible absorption in the 260 – 280 nm range is a reliable indicator of nucleic acid structure. However, probing specific structural locations within the molecule is not possible because this absorption band is broad and featureless. As a result, poorly populated (but not insignificant) intermediates or misfolded states may not be identifiable.

Our method mitigates some of these issues by probing the relaxation kinetics (subsequent to a laser induced T-jump) with infrared (IR) spectroscopy. IR spectroscopic transitions arise from vibrations of specific molecular groups within a molecule. For example, C=O, C=C, and C=N vibrations tend to have strong, distinguishable absorptions in the 1500 – 1700 $cm^{-1}$ range [14]. These groups are found in all nucleotide bases. When RNA undergoes a structural change, the local environment surrounding these groups is altered and spectral changes manifest. This is quite similar to the hypo- or hyper-chromic changes observed in the UV spectral region. However, there are several structural features that are present that can be monitored so that localized kinetic information can be obtained. In this way, we can exploit the

intrinsic properties of the molecule and monitor structurally specific kinetics. In earlier work, we demonstrated the utility of this technique by studying the folding of a larger tRNA molecule [15]. From this study we identified two distinct unfolded populations and postulated that tRNA folds via two parallel pathways each comprised of multiple steps. To our knowledge, this was the first demonstration of time resolved IR spectroscopy of RNA following a laser induced T-jump.

In the work reported here, we study the folding kinetics of the UCCG variant of the UNCG class of tetraloops. In addition to highlighting the utility or our technique, our present results provide further support that the folding landscape of a small hairpin loop is quite rugged and that stacking interactions contribute significantly to the ruggedness.

## Experimental procedures

### Sample preparation

The RNA oligonucleotide sequence used in this work is 5′-gcUCCGgc-3′. The bases written in uppercase are the unpaired bases found in the loop and those written in lowercase comprise the stem. The HPLC purified oligonucleotide was purchased from Integrated DNA Technologies. Samples were prepared by dialyzing against 100 mM phosphate buffer and then lyophilizing against $D_2O$ three times to remove labile protons. The final solution contained RNA at a concentration of ~15 mM (nucleotide) and a buffer pD = 7.2. This concentration is below the level needed to promote duplex formation [16, 17] and has been verified by our own dilution experiments.

The sample was placed in a custom built sample cell that was used for the T-jump experiments and the FTIR spectroscopy. Reliable comparisons between the two experimental components are facilitated by using the same cells for both. The cells are comprised of two $CaF_2$ windows separated by a Teflon spacer that defines the cell path length. The spacer also divides the cell into two compartments; one for the sample (RNA and buffer) and one for the reference (buffer alone). In this way, sample and reference measurements are obtained by simply translating the cell from side to side. For the experiments described here, the cell path length was 53 μm. The relative path lengths of the sample and reference compartments were determined by measuring interference fringes and were found to be within 0.3 μm of each other. The $CaF_2$ windows are then placed in a copper housing. The windows are secured by fastening a copper faceplate to the housing. The cell is then mounted on a larger copper block that is coupled to a circulating water bath for temperature control. The cell temperature is maintained to within ± 0.1°C.

**FTIR spectroscopy**

Equilibrium FTIR spectra as a function of temperature were recorded on a MIR 8025 FTIR spectrometer (Newport). The copper block with affixed cell was mounted to a computer controlled translation stage and placed in a custom built sample compartment. The stage allowed sample and reference spectra to be recorded without compromising the sample compartment purge. Temperature dependent spectra were recorded at 1°C increments from 20°C to 90°C. At each temperature, 256 co-added scans were recorded for the sample and reference. Rather than recording 256 scans for

the reference followed by 256 sample scans, the cell was translated back and forth so that sample and reference spectra were recorded alternately. This was done to minimize any long-term baseline drift.

Because of the spectral collection method just described, baseline deviations were essentially eliminated and no baseline adjustment of the final absorbance spectra was necessary. For presentation, the spectra were smoothed using a second order, 11 point Savitzky-Golay algorithm. Absorbance vs. temperature were plotted for two different wavenumbers, 1574 $cm^{-1}$ and 1669 $cm^{-1}$ to obtain melting curves. The melting curves were fit to a thermodynamic model (described in the main text). The melt curves shown in **Figure 6.2** were obtained from the *raw* data (not the smoothed data) and were normalized using parameters obtained from the fit.

**Time resolved IR spectroscopy**

The sample is probed by a CW diode laser tuned to either 1574 $cm^{-1}$ or 1669 $cm^{-1}$. Spectral resolution is provided by the narrow linewidth of the diode laser (< 0.5 $cm^{-1}$). The diode laser is focused onto the sample by a gold coated off axis paraboloidal mirror (90° off axis; parent focal length = 25.4 mm). The focused beam diameter is less than 100 μm. The transmitted probe radiation is collected by a second off axis paraboloidal mirror and then focused onto a 20 MHz MCT detector. The temperature jump (T-jump) pulse is generated by Raman shifting the fundamental output of a Q-switched Nd:YAG laser in a 1 m long Raman cell filled with $H_2$ gas at 500 psi. The laser operates at a repetition rate of 10 Hz and has a pulse temporal width of 10 ns. The first Stokes line is at a wavelength of 1.9 μm and serves as the T-jump pulse. A Pellin-Broca prism is used

to separate the 1.9 μm radiation from the Stokes, anti-Stokes, and residual Rayleigh radiation. The T-jump pulse is overlapped with the probe laser at the sample. A long focal length lens ($f$ = 250 mm) is used to control the diameter of the T-jump pulse. The pump pulse is adjusted to be roughly 5× the diameter of the probe laser at the point of overlap. This ensures uniform heating of the probe volume and minimizes the effects of drift.

$D_2O$ has a near IR absorption band centered around 1.9 μm. The $D_2O$ solvent transmits 75-80% of the T-jump pulse in our cells (path length = 53 μm). The energy from the fraction that is absorbed generates the temperature jump. The specific heat capacity of $D_2O$ is 4.22 J $K^{-1}g^{-1}$. Assuming 20% absorption and a sample volume of 3.9 × $10^{-5}$ mL, one can calculate that ~9 mJ of 1.9 μm radiation is required to generate a 10°C temperature jump. Since the thermal equilibration of the aqueous solvent occurs on a picosecond time scale, the temperature jump is instantaneous when induced by 10 ns laser pulse. Hence, the time resolution of the experiment is limited by the laser pulse width and/or the detector rise time (~20 ns combined).

Prior to collecting time resolved data, the transmission of the probe wavenumber through the reference as a function of temperature is measured. Using these data, the transient absorption through the reference is used as an internal thermometer to measure the magnitude of the T-jump.

The arrival of the T-jump pulse defines $t$ = 0 and triggers data collection. The transient absorption profile subsequent to the T-jump is detected by the MCT detector described above. The transient waveform from 0 to 1.4 ms is digitized at 5 ns intervals by a 14-bit high speed digitizer (100 MHz bandwidth, 200 MS/s). Each waveform is

averaged over 3000 laser shots. The time dependent absorption is calculated for each sample/reference pair. Finally, the time resolved absorbance of the sample is obtained by subtracting the reference absorption from the sample absorption.

Digitizing the waveforms at 5 ns increments over nearly 6 decades yields in excess of 260,000 data points. This has two drawbacks. One is that it is simply inconvenient; the excessive number of data points is cumbersome and impedes data analysis and display. The second drawback is more practical. Digitizing a long time span in small intervals leads to unnecessary point density at longer times. For example, at $t$ = 25 ns, the next point would be at 30 ns. However, at $t$ = 500 μs, the next point would be at 500.005 μs. Aside from being impractical, subsequent data analysis is heavily weighted toward longer times and could lead to erroneous results. We circumvent this by interpolating the transient absorption profiles (i.e., the time resolved data after the reference has been subtracted from the sample) using a logarithmically spaced time axis. Using this method, we reduce the number of data points to 3004. The interpolated data *exactly* overlays the original data.

Since the time resolution of our experimental setup is instrument limited to ~20 ns, we fit all of our kinetic data from 100 ns. This provides ample separation between instrumental artifacts and meaningful data.

## Results

### Equilibrium FTIR spectroscopy

Equilibrium FTIR spectra recorded at different temperatures are shown in **Figure 6.1**. The sensitivity of IR spectroscopy to RNA structure is immediately evident from

these data. In addition to hypo- and hyper-chromic changes, there are band shifts, measureable changes in bandwidths, and (with more involved analyses) changes in bandshape. Furthermore, most of these changes are linked to specific structural moieties within the hairpin so that structurally specific information can be extracted [14]. While the spectral features are somewhat broad and overlap to some degree, structural assignments can still be made. The spectral changes at 1574 $cm^{-1}$ and 1669 $cm^{-1}$ are the focus of the work reported here. These are assigned in the following paragraph.

In folded RNA, the C6=O6 stretch of guanine occurs near 1685 $cm^{-1}$. Upon melting, this transition shifts to 1669 $cm^{-1}$ and increases in intensity. The two low frequency peaks around 1564 $cm^{-1}$ and 1577 $cm^{-1}$ are a doublet due to ring vibrations of guanine. Both red shift and increase in intensity as the RNA unfolds. These changes manifest in the difference spectra as a strong feature at 1574 $cm^{-1}$. When the RNA unfolds, base pairing as well as base stacking interactions are disrupted. The proximity of the involved group to the environmental perturbation determines its sensitivity. Consequently, the spectral changes at 1574 $cm^{-1}$ and 1669 $cm^{-1}$ are weighted contributions from both pairing and stacking interactions. For example, the guanine C6=O6 is directly involved in a Watson-Crick base pair with cytosine. Therefore, chan ges at 1669 $cm^{-1}$ are primarily due to base pairing with a much smaller contribution from base stacking. In contrast, the changes at 1574 $cm^{-1}$ (due to guanine ring modes) are more sensitive to base stacking and, to a lesser extent, base pairing.

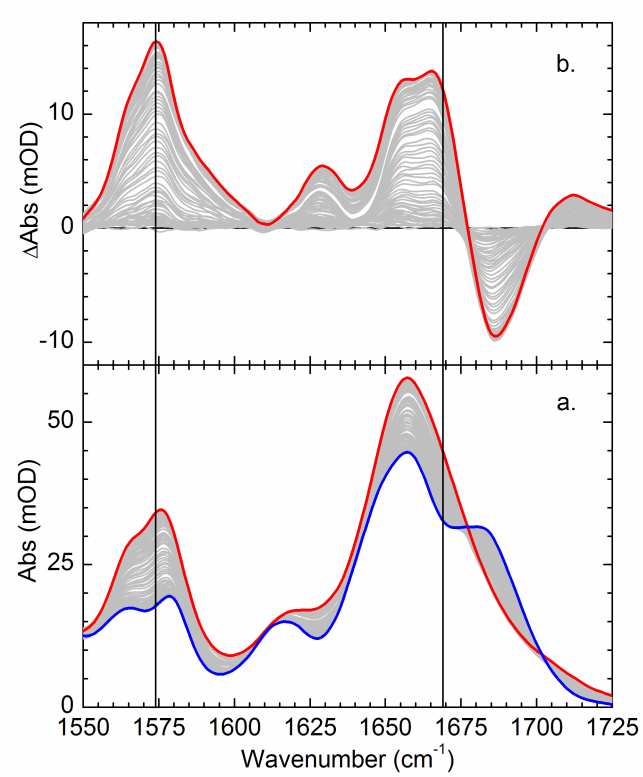


**Figure 6.1: Equilibrium IR melting spectra for UCCG**

(a) Equilibrium FTIR spectra and (b) corresponding difference spectra. In (a), the blue curve is the spectrum at the lowest temperature (20°C) and the red curve is the spectrum at the highest temperature (90°C). The difference spectra were calculated by subtracting the lowest temperature spectrum from the spectra at all other temperatures. The solid vertical lines are at 1574 $cm^{-1}$ and 1669 $cm^{-1}$; the probe wavenumbers.

## Equilibrium melting

Normalized melt curves for each probe wavenumber are shown in **Figure 6.2**. In the equilibrium limit, both curves show two state behavior and were analyzed accordingly. The absorbance is proportional to the fraction of unfolded RNA hairpins and the fraction of unfolded hairpins is related to the equilibrium constant according to:

$$f_U = \frac{K_{eq}}{1 + K_{eq}} \tag{6-1}$$

where $f_U$ is the fraction of unfolded hairpins and $K_{eq}$ is the equilibrium constant. Furthermore, the equilibrium constant is related to the Gibbs free energy of unfolding, $\Delta G$, by:

$$K_{eq} = \exp\left(-\Delta G / RT\right) \tag{6-2}$$

Finally, the enthalpy of unfolding ($\Delta H$), the melting temperature ($T_m$), and the change in heat capacity ($\Delta C_P$) are introduced into the model by noting that the temperature dependence of the Gibbs free energy is given by the Gibbs-Helmholtz relationship:

$$\Delta G(T) = \Delta H(T)\left[1 - \frac{T}{T_m}\right] + \Delta C_P\left[T - T_m - T\ln\frac{T}{T_m}\right] \tag{6-3}$$

However, we held $\Delta C_P$ constant at zero to reduce the number of parameters in the model. While a bit of accuracy is sacrificed by doing this, more general conclusions can still be drawn and the reliability of the fits are increased substantially. By combining the previous relationships, the raw melt curves were fit to the following model

$$A(T) = a + (b - a)\frac{e^{-\Delta G(T)/RT}}{1 + e^{-\Delta G(T)/RT}} \tag{6-4}$$

In this expression, *a* and *b* are constants that give the lower and upper absorption limits, respectively. After the fits were obtained, the data were converted to fraction unfolded vs. temperature so that the two sets of data could be overlapped and compared directly. These results are shown in **Figure 6.2**. Within the error limit, the melting behavior for both is identical. Both show a melting transition at 331 K (58°C). Both also have an unfolding enthalpy on the order of 100 kJ $mol^{-1}$. Although the data obtained from the equilibrium melting is useful, it is unremarkable without further analyses and/or experiments.

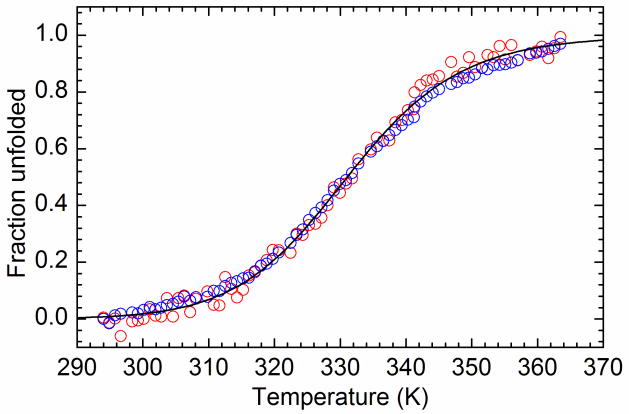


**Figure 6.2: Normalized melting curves for UCCG**

Normalized melting curves for the data recorded at 1574 $cm^{-1}$ (blue circles) and at 1669 $cm^{-1}$ (red circles). The solid line through the data points is a fit to the thermodynamic model described in the text.

**Time resolved spectroscopy**

Despite the similarities in the equilibrium data, the kinetic data reveal distinct differences. The kinetic data are shown in **Figure 6.3**. Two sets of kinetic traces were recorded at each wavenumber. One trace starts from an initial temperature of 49°C (lower panels) and jumps 9°C to the melting transition. The second trace at each wavenumber starts from the transition temperature (58°C) and jumps 9°C to a final temperature of 67°C. **Figure 6.2** shows that at a starting temperature of 49°C, ~24% of the RNA population is unfolded while at the melting transition the fraction unfolded is 50%. Likewise, 73% is unfolded at a final temperature of 67°C.

Before discussing the quantitative aspects of the kinetics, the more general features of the time resolved data will be described. A particularly prominent feature in all of the kinetic traces is the deviation from the zero baseline at early times. This is most pronounced in the data recorded at 1574 $cm^{-1}$. At the earliest resolvable times (down to 10 ns, data not shown), the transient absorption *does not* extrapolate to zero. Instead, it has a finite value between 2 and 3 mOD. There must be a very fast process that is virtually complete in less than 10–100 ns. Although not nearly as distinct, this is also seen in the data recorded at 1669 $cm^{-1}$.

The time at which the transient absorption reaches a maximum is another qualitative feature that is worthy of discussion. When relaxation occurs *at* the melting temperature (i.e., initial temperature of 49°C followed by a 9°C T-jump), the data at 1574 $cm^{-1}$ reaches a maximum by ~78 μs. By comparison however, the corresponding relaxation recorded at 1669 $cm^{-1}$ does not reach a maximum until nearly ~180 μs. Further insight can be gained by comparing the transient absorption maxima with the

equilibrium counterparts for the same temperature interval. For both wavenumbers, these are equal, indicating that the corresponding dynamics are able to fully relax to their equilibrium conformations within 78 μs (1574 $cm^{-1}$) and 180 μs (1669 $cm^{-1}$).

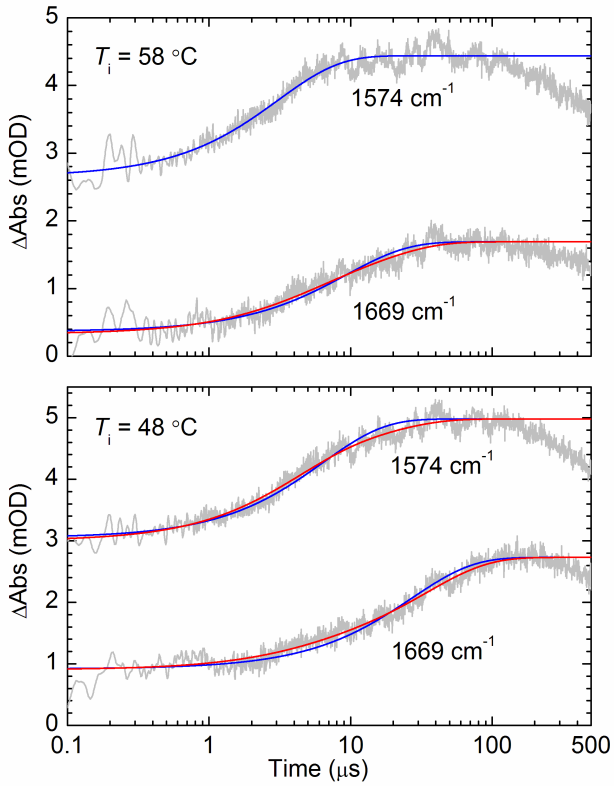


**Figure 6.3: Temperature jump relaxation kinetics**

The lower plot is the data starting from an initial temperature of 48°C and the upper plot is the data starting from an initial temperature of 58°C. The gray lines are the raw data, the blue lines are single exponential fits, and the red lines are double exponential fits. Note that the time axis is logarithmic. Also note that the data *have not* been scaled or offset in any way; the *y* axes values represent the raw data.

Similarly, when relaxation occurs at a temperature above the melting temperature (i.e., initial temperature of 58°C and jumping to 67°C), the transient absorption maxima are reached in 28 μs and 85 μs for the data recorded at 1574 $cm^{-1}$ and 1669 $cm^{-1}$, respectively. In both cases, they are faster than their lower temperature counterparts (as would be expected). Once again, the maxima for the data recorded at 1669 $cm^{-1}$ takes roughly 2 to 3 times longer to reach than the 1574 $cm^{-1}$ data. Comparison to the equilibrium data is not as revealing as in the previous case. At 1574 $cm^{-1}$, the transient absorption maximum and the equilibrium absorption maximum over the same temperature interval are equal. The same conclusion can be drawn as before; the corresponding dynamics are able to fully relax to their equilibrium conformation within 28 μs. However, this is not the case for the data at 1669 $cm^{-1}$. The transient absorption maximum is ~1 mOD *less* than its equilibrium maximum. However, we note that these data have the lowest signal to noise ratio of all the time resolved data. Because of this and the relatively small discrepancy (1 mOD), we are reserving further analysis until more detailed experiments can be performed.

With the exception of the high temperature data recorded at 1574 $cm^{-1}$, all traces are best fit to a double exponential (single exponentials are shown in the figures for comparison) with the following form

$$\Delta A(t) = a_1 \exp(-t/\tau_1) + a_2 \exp(-t/\tau_2) + y_0 \tag{6-5}$$

Here, $\tau_1$ and $\tau_2$ are the relaxation constants, $a_1$ and $a_2$ are their relative amplitudes, and $y_0$ is the maximum transient absorbance (this is slightly different than what was done in

the previous chapter, where $y_0$ was time dependent). The results are tabulated in **Table 6.1**. In the table, the amplitudes of each exponential has been converted to a fractional amplitude for easier comparison. This was done by assuming that at $t = 0$, $\Delta A(t)$ should equal zero. In accordance with this, $y_0$ represents the "full scale" and can be used to normalize the amplitudes. In addition, by assuming that the transients should extrapolate to zero at $t = 0$, we can quantify the magnitude of the "missing" kinetics mentioned earlier.

Perhaps the most noteworthy result is that the kinetics recorded at 1574 $cm^{-1}$ and 1669 $cm^{-1}$ are substantially different. As would be expected from the previous qualitative discussion, the kinetics recorded at 1574 $cm^{-1}$ are faster (overall) than those at 1669 $cm^{-1}$. At the melting temperature, both sets of data show bi-exponential behavior. However, there are marked differences. The resolvable kinetics recorded at 1574 $cm^{-1}$ are dominated by the faster of the two relaxation components (3.6 μs vs. 15.6 μs). In contrast, the transient data recorded at 1669 $cm^{-1}$ are dominated by the slower of the two relaxation components (34.7 μs vs. 3.8 μs). The un-resolvable fast component accounts for approximately ~60% of the relaxation recorded at 1574 $cm^{-1}$ while it accounts for roughly ~22% of the relaxation recorded at 1669 $cm^{-1}$.

At the higher relaxation temperature, the 1574 $cm^{-1}$ data are fit best by a single exponential. The slower component seen at the lower initial temperature vanishes and the resolvable relaxation is dominated by the faster component (3.1 μs). Interestingly, the fraction of the missing relaxation is still 60%. At the same higher relaxation temperature, the data recorded at 1669 $cm^{-1}$ is still bi-exponential and the majority of the resolvable relaxation is still dominated by the slower component. However, it is

nearly a factor of 2 faster than at the lower temperature (14.9 μs compared to 34.7 μs). The fastest of the resolvable phases is only slightly faster at 3.7 μs. However, its amplitude has increased from 10% to 31%.

**Table 6.1**

| $\nu$ ($cm^{-1}$) | $T_i$ (°C) | $a_1$ | $\tau_1$ (μs) | $a_2$ | $\tau_2$ (μs) |
|---|---|---|---|---|---|
| 1574 | 48 | 0.26 | 3.6 ± 0.2 | 0.14 | 15.6 ± 1.5 |
| | 58 | 0.40 | 3.1 ± 0.04 | - | - |
| 1669 | 48 | 0.10 | 3.8 ± 0.3 | 0.56 | 34.7 ± 0.4 |
| | 58 | 0.31 | 3.7 ± 0.7 | 0.49 | 14.9 ± 1.6 |

## Discussion

We probed the relaxation kinetics at two wavenumbers; 1574 $cm^{-1}$ and 1669 $cm^{-1}$. The former is due to a ring vibration of guanine and the latter is due to the C6=O6 of guanine. The C6=O6 stretch of guanine is sensitive to its hydrogen bonded state (i.e., base pairing). The free guanine has an absorption band around ~1685 $cm^{-1}$. When base pairing occurs, electron density is withdrawn from the carbonyl and the vibrational frequency red shifts and increases in intensity. However, the ring vibration around 1574 $cm^{-1}$ is not as straightforward to assign. There is little doubt that changes in the absorption around 1574 $cm^{-1}$ are sensitive to RNA structure. However, the exact nature of the changes is not as clear. The electron density changes that accompany the C6=O6 carbonyl are "felt" in the ring, thereby altering the vibrational characteristics of the ring mode. However, base stacking interactions would also affect this mode. While there is

no precedent for this specific assignment (to our knowledge), we are confident that base stacking interactions can be inferred from changes in this vibrational mode.

Recent theoretical treatments may provide molecular insight into our results [18-21]. By analyzing small hairpin sequences, it becomes feasible to enumerate possible intermediate structures. The statistical weights of the intermediates can be calculated by determining their energies from published data. A kinetic model can be constructed from these results by calculating the rates of interconversions between each structure (or cluster of related structures). This approach has been used quite successfully by Chen and co-workers with excellent results. Not only does their work predict folding intermediates, it also predicts the presence of alternative folding routes (kinetic heterogeneity). Chen's results are extremely significant and provide a much needed link between experimental data and molecular details. Unfortunately, his work cannot be used directly to model our data. However, indirect application of the model can still provide molecular insight.

According to Chen's model, base stacking interactions stabilize RNA structure; isolated base pairs do not. Hence, for a structure to be considered statistically relevant, it must contain at least one base stack (i.e., two adjacent base pairs). The absence of stabilizing stacking interactions renders other structures insignificant. This has two benefits. For one, the so-called Turner rules can be used to calculate the energies of the potential intermediates [22]. Secondly, by eliminating structures that lack base stacks, the configurational space is dramatically reduced and subsequent calculations become more tractable. By design, the model highlights stem interactions. Any intermediate

ensembles or alternate pathways that are identified are due to structural diversity in the stem.

Our model system is quite small and contains only a single base stack in the stem of the native structure. Folding intermediates in our system simply cannot be due to misfolded or partially folded stems. Enumeration of "possible" intermediates for our system shows that alternative base pairings (with or without stacks) would result in structures that are energetically highly unfavorable. Other than partially folded native hairpins (either of the stem native base pairs formed), any other structure would have either non-native base pairs, small loops (e.g., di- or triloops), or a combination of the two. Despite the lack of stem diversity, we still see kinetic intermediates and evidence of kinetic heterogeneity. From this we conclude that loop interactions and single strand stacking interactions must also play a key role in the folding mechanism.

Indeed, favorable stacking interactions between the bases within the loop contribute to the stability of the native structure. This has been known for some time and is accounted for in some models [23, 24]. However, these models are still largely dominated by nearest neighbor interactions and the loop contribution remains somewhat phenomenological; it improves the overall model accuracy but provides little additional molecular insight. It is quite feasible however, that these intra-loop and single strand stacking interactions are strong enough to produce local energy minima as well as possible routes for alternate folding populations.

We explain our data by treating stacking and base pairing as separate degrees of freedom. Earlier work has shown that base stacking occurs in the 0.1 – 1 μs range [25, 26]. Accordingly, the un-resolvable fast phases in our data are attributed to this.

However, the resolvable fast phases occur on the microsecond time scale. In addition, these show only a small temperature dependence, indicating that the corresponding process is opposed by a small activation energy and that the barrier is predominantly entropic. This is consistent with stacking *rearrangement* previously observed by Porschke [27]. For a hairpin to form, the ends must come together requiring that the chain bend somewhat sharply. For this to occur, the adjacent bases must be oriented to permit favorable stacking interactions. This partial conformational ordering explains the unfavorable entropy and results in microsecond relaxation dynamics. In this regard, the partially stacked intermediate is "on-pathway" since it connects the native and unfolded states. However, this does not imply that only native intramolecular interactions stabilize the intermediate. The data presented here are insufficient to make such an assignment.

In contrast, the slow phases behave differently. At 1669 $cm^{-1}$, this phase becomes almost twice as fast as the temperature is increased; corresponding to an activated process. Based on this observation and our equilibrium thermodynamic data, we believe that this is consistent with stem formation. In addition, its amplitude decreases slightly as the temperature is raised. Assuming that the fast phase corresponds to stacking rearrangement, then at higher temperatures the chain is more mobile and more configurations become available. Thus, the entropic barrier is lowered. In so doing, the chain configuration is closer to its equilibrium state when the ends come together. Thereby, the amount of structural relaxation that must occur for the stem to form is less. Hence, its amplitude is decreased. The same argument is used to explain the disappearance of the slow phase recorded at 1574 $cm^{-1}$ on increasing temperature.

However, since stacking precedes base pairing, the ring vibration "equilibrates" faster than the carbonyl vibration so the effects of stem formation at the higher temperature are not felt as strongly at 1574 $cm^{-1}$.

In the previous work of Gruebele and Bevilacqua, single exponential relaxation kinetics are observed near the melting temperature [9]. A fast phase appears below the melting temperature and a slow phase appears above it. Our data is different. We see bi-exponential kinetics at the melting temperature for both wavenumbers. However, above the melting temperature, the kinetics at 1574 $cm^{-1}$ are single exponential. We do not have kinetic data below the melting temperature for comparison but we can make comparisons with the relaxations at and above the melting temperature.

They see a main kinetic phase that is due to the transition from the native ensemble to the unfolded ensemble. Near the melting temperature, only these two states are populated enough to be observed experimentally and the relaxation is single exponential. They argue that at higher temperatures, an ensemble of off pathway, stem sensitive structures is populated. These serve as misfolded traps that are separated from the unfolded state by a large barrier. Consequently, the a slow kinetic phase is observed. Below the melting temperature, an ensemble comprised of hairpins with partially folded (native) stems is populated. These are on pathway intermediates and the rate of interconversion between them and the native state is fast. This results in bi-exponential kinetics with a fast phase in addition to the main phase.

Rather than assume that our results are in opposition, we believe that there are marked similarities and that the differences provide further evidence that the energy landscape of small RNA hairpins is extremely rugged. The differences are accounted for

by recognizing that the same process viewed using different experiments can produce different results. Each type of experiment is sensitive to slightly different phenomena. Previous work of Dyer and co-workers is a good example [28]. T-jump relaxation kinetics of apomyoglobin probed using tryptophan fluorescence revealed mono-exponential kinetics while the same system under the same conditions, studied using IR, revealed bi-exponential kinetics. The differences are not in conflict. Rather, they highlight the sensitivity of different types of measurements. In addition, these differences can be interpreted as further evidence of a relaxation process governed by a rugged energy landscape. In this regard, different types of measurements are complementary and a variety of experiments are needed to fully characterize the process.

In terms of parallels, the main kinetic phase observed by Gruebele and our "slow" phase are likely to be analogous. Both are activated and the rates are similar. We both assign this phase as the interconversion between the native and unfolded states. However, we further assign this to the formation of stem base pairs. They attribute their fast phase to the interconversion between two closely related ensembles. Whereas, we assign our fast phase to stacking rearrangement; motion that precedes stem formation. These are not mutually exclusive observations.

**Conclusion**

Using a fast folding, relatively "simple" hairpin structure, our results show that the most fundamental aspect of RNA folding (hairpin formation) is a complex process. Multiple kinetic phases are observed that indicate the presence of local energy minima

between the unfolded and native states. This supports the notion that the energy landscape is rugged. Furthermore, we monitor folding using two different (but intrinsic) spectral markers. This is akin to probing a multiply labeled RNA molecule. Our data show that the observed kinetics depend upon the probe wavenumber. Since each probe wavenumber is assigned to a structural aspect of the sample, we can discern alternative populations. Probe dependent kinetics suggest that kinetic heterogeneity is also a characteristic of RNA hairpin formation. In fact, we observe substantially different kinetics at the two wavenumbers probed. Overall, the kinetics recorded at 1574 $cm^{-1}$ are roughly twice as fast as those recorded at 1669 $cm^{-1}$.

We interpret our findings as evidence that misfolding in the stem is insufficient to explain the folding complexity. Our results show that loop interactions and/or single strand stacking fluctuations also contribute to the folding complexity. The earliest stages of hairpin folding involve base stacking. Before the chain can collapse permitting loop closure, the stacks rearrange leading to partial conformational ordering. With the bases properly oriented relative to their neighbors, the chain can collapse and zipping of the stem base pairs can occur.

**Bibliography**


[1] P. Brion and E. Westhof, *Hierarchy and dynamics of RNA folding*. Annual Review of Biophysics and Biomolecular Structure, 26 (**1997**) 113-137.

[2] L.W. Kwok, I. Shcherbakova, J.S. Lamb, H.Y. Park, K. Andresen, H. Smith, M. Brenowitz and L. Pollack, *Concordant exploration of the kinetics of RNA folding from global and local perspectives*. Journal of molecular biology, 355 (**2006**) 282-293.

[3] J. SantaLucia, Jr. and D.H. Turner, *Measuring the thermodynamics of RNA secondary structure formation*. Biopolymers, 44 (**1997**) 309-319.

[4] G. Varani, *Exceptionally stable nucleic acid hairpins*. Annual Review of Biophysics and Biomolecular Structure, 24 (**1995**) 379-404.

[5] C.R. Woese, S. Winker and R.R. Gutell, *Architecture of ribosomal RNA: Constraints on the sequence of "tetra-loops"*. Proceedings of the National Academy of Sciences of the United States of America, 87 (**1990**) 8467-8471.

[6] A. Ansari, S.V. Kuznetsov and Y. Shen, *Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins*. Proceedings of the National Academy of Sciences of the United States of America, 98 (**2001**) 7771-7776.

[7] Y. Shen, S.V. Kuznetsov and A. Ansari, *Loop dependence of the dynamics of DNA hairpins.* Journal of Physical Chemistry B., 105 (**2001**) 12202-12211.

[8] X. Wang and W.M. Nua, *Kinetics of end-to-end collision in short single-stranded nucleic acids.* Journal of the American Chemical Society, 126 (**2004**) 808-813.

[9] H. Ma, D.J. Proctor, E. Kierzek, R. Kierzek, P.C. Bevilacqua and M. Gruebele, *Exploring the energy landscape of a small RNA hairpin*. Journal of the American Chemical Society, 128 (**2006**) 1523-1530.

[10] E.M. Moody, J.C. Feerrar and P.C. Bevilacqua, *Evidence that folding of an RNA tetraloop hairpin is less cooperative than Its DNA counterpart*. Biochemistry, 43 (**2004**) 7992-7998.

[11] D.J. Proctor, H. Ma, E. Kierzek, R. Kierzek, M. Gruebele and P.C. Bevilacqua, *Folding thermodynamics and kinetics of YNMG RNA hairpins: Specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate*. Biochemistry, 43 (**2004**) 14004-14014.

[12] N.A. Siegfried, S.L. Metzger and P.C. Bevilacqua, *Folding Cooperativity in RNA and DNA Is Dependent on Position in the Helix*. Biochemistry, 46 (**2007**) 172-181.

[13] C. Hyeon and D. Thirumalai, *Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins*. Journal of the American Chemical Society, 130 (**2008**) 1538-1539.

[14] M. Banyay, M. Sarkar and A. Gräslund, *A library of IR bands of nucleic acids in solution*. Biophysical Chemistry, 104 (**2003**) 477-488.

[15] E.B. Brauns and R.B. Dyer, *Time-resolved infrared spectroscopy of RNA folding*. Biophysical Journal, 89 (**2005**) 3523-3530.

[16] M. Abdelkafi, N. Leulliot, V. Baumruk, V. Bednarova, P.Y. Turpin, A. Namane, C. Gouyette, T. Huynh-Dinh and M. Ghomi, *Structural features of the UCCG and UGGC tetraloops in very short hairpins as evidenced by optical spectroscopy*. Biochemistry, 37 (**1998**) 7878-7884.

[17] V. Baurmruk, C. Gouyette, T. Huynh-Dinh, J. Sun and M. Ghomi, *Comparison Between CUUG and UUCG Tetraloops: Thermodynamic Stability and Structural Features Analyzed by UV Absorption and Vibrational Spectroscopy*. Nucleic Acids Research, 29 (**2001**) 4089-4096.

[18] W. Zhang and S.J. Chen, *RNA hairpin-folding kinetics*. Proceedings of the National Academy of Sciences of the United States of America, 99 (**2002**) 1931-1936.

[19] W. Zhang and S.J. Chen, *Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis*. Biophysical Journal, 90 (**2006**) 765-777.

[20] W. Zhang and S.J. Chen, *Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states*. Biophysical Journal, 90 (**2006**) 778-787.

[21] W. Zhang and S.J. Chen, *Master equation approach to finding the rate-limiting steps in biopolymer folding*. Journal of Chemical Physics, 118 (**2003**) 3413-3420.

[22] M.J. Serra and D.H. Turner, *Predicting thermodynamic properties of RNA*. Methods in Enzymology, 259 (**1989**) 242-261.

[23] V. Ivanov, Y. Zeng and G. Zocchi, *Statistical mechanics of base stacking and pairing in DNA melting*. Physical Review E., 70 (**2004**) 051907.

[24] S.V. Kuznetsov, Y. Shen, A.S. Benight and A. Ansari, *A seminflexible polymer model applied to loop formation in DNA hairpins*. Biophysical Journal, 81 (**2001**) 2864-2875.

[25] T.G. Dewey and D.H. Turner, *Laser temperature-jump study of stacking in adenylic acid polymers*. Biochemistry, 18 (**1979**) 5757-5762.

[26] D. Porschke, *Molecular states in single-stranded adenylate chains by relaxation analysis*. Biopolymers, 17 (**1978**) 315-323.

[27] M. Menger, F. Eckstein and D. Porschke, *Dynamics of the RNA hairpin GNRA tetraloop*. Biochemistry, 39 (**2000**) 4500-4507.

[28] M. Gulotta, R. Gilmanshin, T.C. Buscher, R.H. Callender and R.B. Dyer, *Core formation in apomyoglobin: Probing the upper reaches of the folding energy landscape*. Biochemistry, 40 (**2001**) 5137-5143.

# Appendix A

## Experimental protocols

This work entails many experimental aspects from sample preparation, cell construction, acquisition of IR melting spectra and data workup, to the final time-resolved IR experiment and subsequent curve fitting. The material in this Appendix is a comprehensive guide describing the experimental procedures.

### Nuclease controls

Stringent nuclease controls were employed in every step of the RNA sample preparation to prevent ribonuclease (RNase) contamination. Buffers and samples were brought up in milli-Q certified RNase free water and prepared in an established RNase free work area. Counters were regularly decontaminated using 10% bleach solution. Pipette tips and eppendorf tubes were certified RNase, DNase, and human DNA free. All Nalgene bottles were cleaned using RNase Away® and then rinsed with nuclease free water. Glassware and metal utensils were baked in an oven at 190°C for 24 hours. The pH electrode(s) were cleaned prior to each use with Ambion's Electrozap™ and then rinsed with nuclease free water.

RNase contamination was routinely monitored with IDT's RNase Alert™ test kit. This analysis is performed with a sensitive, fluorescent RNA oligomer that detects pico-molar quantities of RNase enzymes. Both the 3' and 5' ends of the molecule are functionalized, one with a fluorescent probe, and the other with a corresponding quenching group. When the oligomer is intact, fluorescence is quenched and this gives a negative result for RNase contamination. If RNases are present, the oligomer will be

cleaved by the contaminating ribonucleases causing the RNA probe to fluoresce. Contamination can be assessed qualitatively with a UV lamp (short -wavelength) and quantitatively with a fluorimeter. In this way we were able to regularly evaluate RNase contamination while only using small volumes (5 µL) of the precious RNA samples. Following these strict nuclease controls, contamination was consistently avoided.

**Phosphate buffers**

Care must be taken when preparing phosphate buffers in order to calculate the appropriate molar ratios of mono and dibasic sodium phosphates. The literature value for the p$K_a$ of phosphoric acid is typically listed as 7.2. This would make the calculation of a 100 mM phosphate buffer at a biological pH of 7.2 seem trivial. Adding equal molar concentrations of the conjugate acid and base should give a pH close to 7.2. Experimentally, this is not the case. Mixing equal molar concentrations of the mono and dibasic salts actually gives a solution with a pH of 6.86 at 100 mM. This is the true concentration dependent p$K_a$, which inherently takes activity coefficients into account. The experimental value can also be found referenced in the literature, in fact a range of values (from 6.7-7.2) are reported depending on the particular source. The value of 7.2 for the p$K_a$ is a "theoretical" value calculated without considering activity coefficients. Neglecting activities in this case leads to a definite deviation in the p$K_a$ value and must be considered when using the Henderson-Hasselbalch equation to calculate relative salt ratios.

The conjugate acid ($NaH_2PO4$) and base ($Na_2HPO_4$) concentrations for a 100 mM phosphate buffer at pH 7.2 are calculated below:

$$\text{pH} = \text{p}K_a + \log\frac{[\text{A}^-]}{[\text{HA}]}$$

$$7.2 = 6.865 + \log\frac{[\text{A}^-]}{[\text{HA}]}$$

$$2.163 = \frac{[\text{A}^-]}{[\text{HA}]}$$

For a buffer concentration of 100 mM:

$$100\ \text{mM} = [\text{A}^-] + [\text{HA}]$$

$$100\ \text{mM} = 2.163[\text{HA}] + [\text{HA}]$$

$$[\text{HA}] = \mathbf{31.6\ mM}\ NaH_2PO_4$$

$$100\ \text{mM} = [\text{A}^-] + 31.4\ \text{mM}$$

$$[\text{A}^-] = \mathbf{68.4\ mM}\ Na_2HPO_4$$

To prepare a 100 mM buffer solution from stock solutions of 0.5 M $NaH_2PO_4$ and 0.5 M $Na_2HPO_4$, dissolve the following volumes of the stock solutions into one liter of water and adjust to pH 7.2 with 5 M NaOH and 5 M HCl:

$$NaH_2PO_4: \quad \frac{0.0316\ \text{mol}}{\text{L}} \times \frac{1000\ \text{mL}}{0.5\ \text{mol}} = \mathbf{63.2\ mL\ stock/L}$$

$$Na_2HPO_4: \quad \frac{0.0684\ \text{mol}}{\text{L}} \times \frac{1000\ \text{mL}}{0.5\ \text{mol}} = \mathbf{136.8\ mL\ stock/L}$$

**Sample preparation**

Prepare phosphate buffers as previously described, or an alternate buffer (e.g., Tris, Tris + EDTA, etc). RNA samples are first brought up in buffer to the desired concentration (~15 mM nucleotides). Do not bring RNA up in pure water (or $D_2O$), as lyophilizing becomes very difficult do to electrostatics. The solution melts even after freezing at liquid nitrogen temperatures (77 K) and under strong vacuum (0.03 mBar). Precious RNA samples can be lost if they melt during freeze-drying. Lyophilize off water (or $D_2O$) from *buffered* RNA solutions to ensure proper freeze-drying.

Take an aliquot (~125 μL) for FTIR and time-resolved IR spectroscopic analysis. Store the remaining stock solution in a freezer at -20°C. Dialyze the small aliquot volume using a Tube-O-DIALYZER™ with a molecular weight cut off of 1000 Da. Dialyze the sample with a 1000 fold dilution factor two times, which gives a total dilution factor of one million fold. Change the dialysis buffer after 6 hours. UV-vis measurements can now be taken to quantify the purity and concentration of the final RNA solution.

Immerse the RNA in liquid nitrogen for about one minute, and lyophilize off the water leaving the RNA and buffering salts. The freeze-drying process takes 6-24 hours depending on sample volumes. Next, while working in a dry box (just a large Plexiglas box purged with dry air), dissolve the sample in $D_2O$ using twice the original aliquot volume and vortex briefly. Lyophilize against $D_2O$ four times to eliminate water and to ensure complete proton exchange on the RNA. Reconstitute the final solution by bringing it up to the original aliquot volume. The resulting solution will have a pD ~7.2 at the originally calculated buffer and oligomer concentrations. This entire process

should be done in parallel with a matching buffer solution for use as a spectroscopic reference.

Transfer the reconstituted solutions (reference and sample) into individual 0.3 mL capacity Wheaton V-vials with septum caps and degas both solutions. Degas under vacuum using a conventional tabletop vacuum pump adapted with a syringe. Inject the metal needle tip through the septum. Be careful not to insert the needle into the solution, which would cause loss of sample.

**IR cell construction**

This section describes the preparation and construction of a custom IR sample cell. The same cells are used in both the equilibrium and time-resolved IR experiments. The $CaF_2$ windows must be extremely clean and defect free. The windows are cleaned by gently rubbing with Liquinox under running DI water (be sure to use a clean pair of nitrile gloves when doing this). Dry by placing the window on a lens tissue that is itself on top of a paper towel. This draws off the water without scratching the windows. Once dry, place the windows onto a lens tissue in the bottom of a petri dish. Fill the petri dish with hexanes. Using calipers and a folded lens tissue, dip the tissue in the solvent and *very gently* wipe off both surfaces several times. Submerge the window under the hexanes to dissolve any liberated oils or residual silicone grease. Repeat this process several times. Since this step is to remove residual silicone grease on old windows, you can skip this step if you are using new windows.

Silicone grease can be difficult to remove and may require multiple aqueous and organic cleaning cycles. Repeat the procedures until no residual silicone can be detected

(viewing the windows under different lighting conditions and from various angles can sometimes allow one to see unwanted residue and imperfections). Once the windows have been scrupulously cleaned, blow off the windows with compressed air and wipe them off gently (i.e., a light single swipe) with a new lens tissue and HPLC grade methanol.

Prepare the IR cell (**Figure A.1**) for sample injection by placing the back plate of the copper housing on the lab bench. Place a black rubber O-ring in the housing and then set the $CaF_2$ window (no holes) on top of the O-ring. Next, grease a custom cut Teflon spacer (~55 μm thickness) with a very small amount of silicone grease (a little goes a very long way), front and back, and very carefully place this washer on the cell window ensuring not to contaminate the viewing surface with grease (a pair of precision tweezers is helpful). Place the other $CaF_2$ window (4 injection holes) on top of the spacer with the holes in the proper orientation. Next, place a custom cut Buna red rubber washer on top of the windows and cap the cell housing with the front copper plate. The windows should now be sandwiched between the front and back copper plates, and separated by the Teflon spacer. Secure the cell with 4 socket cap screws (2-56 threading) tightening in a star pattern until gently snug.

Place the assembled cell in an oven at 95°C for 15 minutes to allow the grease to evenly coat the Teflon spacer. Remove the cell and finish tightening in a star pattern until fully snug. Take care not to over tighten as the windows are very brittle and will crack easily.

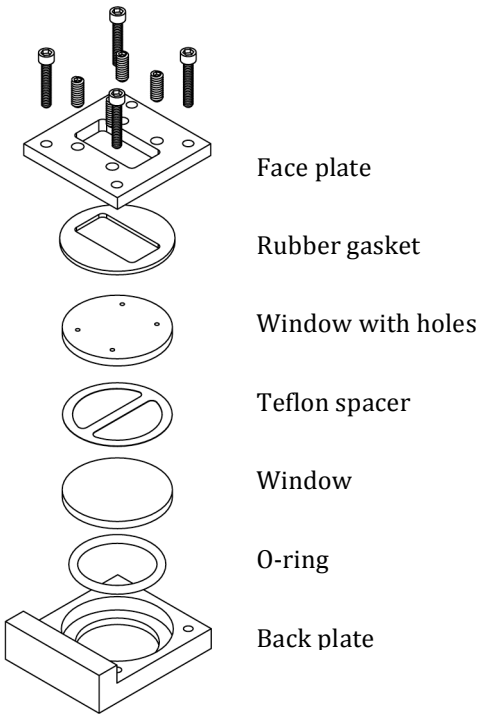


**Figure A.1: Custom IR cell for FTIR and time-resolved IR spectroscopies**
See text for details. Schematic provided by Giovanni Crosland.

## Cell pathlengths

Measure cell path lengths using the fringe counting method. Take the distance on the interferogram from the center burst to the shoulder bands. At 2 $cm^{-1}$ resolution there are 16384 points and the centerburst will be at 8192. For a path length of ~57 μm, the interference fringes will be at approximately 8012 and 8372. The difference between the center burst and each fringe is 180. Taking the difference divided by two and multiplying by the wavelength of the HeNe at 0.6328 μm gives a pathlength of 57 μm.

The calculation is as follows:

$$\frac{180}{2} \times 0.6328\ \mu m = 57 \mu m$$

Teflon is made to a tolerance of ±5 μm, which seems undesirable as our tolerances need to be within ± 0.3 μm. However, Teflon is somewhat malleable and can usually be compressed to give pathlengths within ±0.1 μm (often exact). Other polytetraflouroethylenes can also be used that have better tolerances (e.g., ±0.5 μm) such as fluorinated ethylene propylene (FEP), but the material is rigid and resists compression giving path length differences on the order of the nominal tolerance. Ultimately, Teflon is the best choice as the pathlenghts can be precisely adjusted.

**Sample injection**

Inject the IR cell with the reference solution first, followed by the sample solution in order to not contaminate the reference solution with RNA. This should be done in a dry box purged with water free air. Place a needle tip in the exit port of the reference side to serve as a vent. Wet a 50 μL Luer tip Hamilton gas chromatography syringe with buffer solution and then slowly empty the solution back into its vial to expel air. Draw ~25 μL of buffer into the syringe and smoothly inject the reference side of the cell with the buffer. The solution will exit through the venting syringe tip. When working with the more precious sample, every effort should be made to save this portion. Place setscrews (4-40 threading) in the injection and exit ports and tighten until snug. Repeat the entire

procedure with the sample solution (RNA + buffer) on the sample side of the IR cell. Inspect both cell compartments to make sure the injection went smoothly and the viewing windows do not show signs of air bubbles. Make sure the window surfaces are scrupulously clean and have no imperfections (scratching etc.) in the viewing area. Blow off the windows with canned or house air. The cell is now ready for use.

**Equilibrium IR spectroscopy**

The FTIR spectrometer used in this work was a Newport MIR 8025 Fourier Transform IR spectrometer. Although the spectrometer was a turnkey instrument, the melting experiment requires significant modification of the system. The overall IR melting experiment couples the FTIR spectrometer with a recirculating temperature bath, a micron resolution computer controlled translational stage, a dry-box purged with water free and carbon dioxide free air, and a thermocouple to monitor the cell temperature. The entire melting experiment is automated via LabVIEW software programmed to coordinate the components and control the experiment.

The entire system is very useful for performing these RNA melting experiments and collecting high quality spectral data. However, there has been one major complication that arises during the execution of the experiment. This is due to a flaw in the Newport software drivers that they (Newport) have been unwilling to address. Essentially, each spectrum that is acquired remains in the computer's memory even after the data has been saved. During the course of a melting experiment, thousands of spectra can be acquired due to co-adding over many different temperatures. Eventually, this consumes the available RAM and crashes the computer. To mitigate this problem, data is collected

in four sets consisting of 11 temperatures each. Between each set, the program is shut down (this clears the memory) and restarted. The temperature sets are collected in two degree increments from 20-40°C, 42-62°C, 64-84°C, and 86-90°C.

## Time-resolved IR

### Pump beam

The Nd:YAG (neodynium: yttrium aluminum garnet) laser is a reliable, stable, high output (6 W) instrument producing 1.064 μm pulses at 10 ns nominal pulse width. The laser used in this work was the Surelite model by Continuum. The high energy beam is easy to visualize with an IR viewer or burn paper which facilitates alignment. After aligning the YAG fundamental through the Raman converter (see **Figure A.2**), block the anti-stokes (visible) scattering with a colored glass filter and allow the remaining beam to pass through a Pellin-Broca prism (**Figure A.2 a**). The prism separates the residual YAG fundamental and the Stokes lines. A beam block placed around 10 inches from the prism allows the first Stokes line (the desired wavelength for the T-jump pulse) to pass and blocks the residual YAG fundamental and higher order Stokes lines. The T-jump pulse is then passed through a polarizer (**Figure A.2 c**) (to control the pulse intensity) and a long focal length (~200 mm) lens (**Figure A.2 d**) (to control the beam diameter) before passing through the sample. After the sample, be sure to terminate the beam with a beam block. Focus the beam a few inches beyond the sample. Focusing the beam too tight on the sample cell can damage the $CaF_2$ windows. The beam diameter should be ~1 mm at the sample cell.

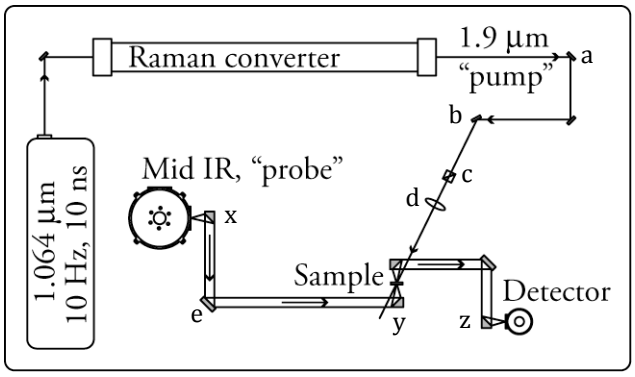


**Figure A.2: Laser induced temperature jump (pump-probe) schematic**
See text for details.

**Probe beam**

The probe beam is a continuous wave (CW) lead-salt (Pb:Se) diode laser that can be tuned between 1550 and 1750 $cm^{-1}$. Working with diode lasers is challenging due to the poor beam quality and propensity to mode-hop. The output of a diode laser is highly divergent and must be collimated prior to use. This is done with a gold coated off-axis parabolic mirror (1 inch diameter, 90° off-axis, 1 inch parent focal length). Aligning an off-axis parabolic mirror (OAP) is not trivial (see Appendix C) and requires a translational stage with three translational degrees of freedom and at least two angular (tilt) adjustments.

To collimate the beam remove the flat mirror **(Figure A.2 e**) just after the first collimating OAP (**Figure A.2 x**). Roughly align a *temporary* OAP and MCT detector at the far edge of the laser table directly down field from the collimating OAP (perpendicular to the diode). Adjust both until you can see a measurable signal, and

then maximize the signal on the detector. You will have to adjust the height and orientation of the OAP as well as the distance between the OAP and detector. Check the collimation of the diode beam with a circular aperture at multiple points along the beam. This shows if the beam is divergent, convergent or collimated. Adjust the first OAP to better collimate the beam (see Appendix C). Each time the collimating OAP is adjusted, the temporary OAP will need to be realigned to maximize the signal. Check the diode beam collimation again with a circular aperture and repeat the alignment sequence. This will need to be done multiple times until the beam is well collimated, and the signal at the detector is maximized. This indicates that all the light being emitted by the diode is being captured by the detector element.

Knowing the full intensity of the diode is useful when aligning the beam through the optical setup so that losses can be evaluated. Once the diode beam is well collimated, and the full intensity is determined, the flat mirror can be replaced and the probe beam aligned through the rest of the optical set-up. Guide the probe beam through rest of the system until you can detect a signal at the MCT. Using the previously determined full output of the diode, adjust the signal until it is optimized at the detector. Adjust the OAP just before the MCT (**Figure A.2 z**) to maximize the signal, expecting losses from each reflective optic (~3%). Using a 50 μm circular aperture adjust the beam waist of the diode where it crosses the sample. A well-collimated beam with a beam diameter of 12 mm can be focused to ~50 μm (see Appendix C). The tighter the beam waist at the cell, the better the beam overlap will be. Good beam overlap in the laser interaction volume helps minimize T-jump drift and increases signal to noise ratios.

**Diode laser characterization**

The wavelength of the diode laser is tuned by varying its temperature and current (see **Figure A.3**). The temperature/current behavior must be characterized prior to use and repeated periodically since the diode output is prone to mode-hopping. First, ensure that the monochromator is properly calibrated by checking the wavelength of a HeNe laser which has a wavelength at precisely 632.8 nm. Since we're interested in IR wavelengths, the $9^{th}$ order reflection of the HeNe at 1755 $cm^{-1}$ should be used. If the reading is far off, the monochromator will need to be calibrated. Once this is done, the diode laser can be characterized by independently tuning the temperature and the current. Prior to doing this, be sure to check the diode laser manual for the appropriate temperature and current ranges for the particular diode in use. Using inappropriate temperatures and/or currents can damage the diode.

Begin by setting the temperature on the diode controller and sweeping the current until a signal is detected on the fast MCT detector. Once a reasonable signal is detected (a few Volts), insert the flat mirror just before the sample into the optical setup and guide the diode beam into the monochromator and measure the wavelength. Incrementally increase the current on the diode laser's controller again until the fast MCT indicates an intensity fluctuation or mode hop between wavenumbers, and scan the corresponding diode output. Continue this process across the entire current range. Once this is done, increase the temperature and repeat. To obtain an overview of the modal behavior of the laser, varying the temperature in 5 degree increments works well. However, the temperature can be varied in much smaller increments to finely tune the diode to produce a desirable output. In general, except for instances of reverse

mode hopping, the output energy (wavenumbers) will increase as a function of current and temperature.

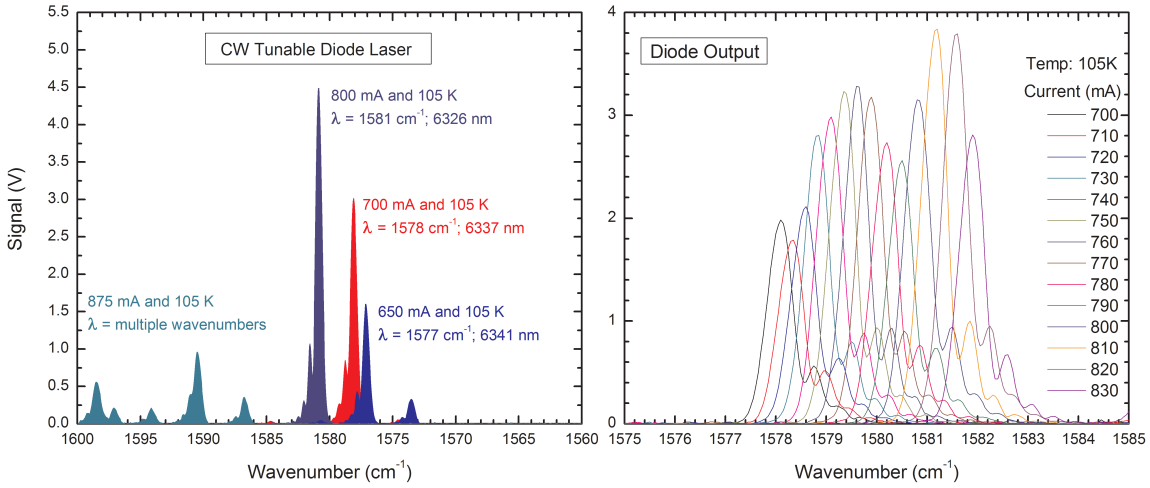


**Figure A.3: Monochromator scans of a tunable CW diode laser output**
(Left) The current is varied and the temperature is held constant. Notice the modal quality deteriorate and mode hop at higher current. (Right) The output is continuously tunable through a few wavenumbers, before mode hopping.

### Temperature calibration for T-jump experiments

The transmission of the probe beam through the solvent ($D_2O$ or deuterated buffer for our experiments) is temperature dependent (see **Figure A.4 a**) and is used as an "internal thermometer" to measure the magnitude of the T-jump. Consequently, a temperature calibration must be performed prior to each time-resolved experiment. It is very important to note that this procedure be repeated every time a new sample is used—even if the probe wavelength is the same. This is because the temperature dependence is not very pronounced and even subtle differences between samples can

lead to errors if the same calibration is used. Moreover, it is good practice to repeat this procedure for the same sample if the cell has been moved or if a significant amount of time has elapsed since the last calibration. Also note that the temperature calibration is only performed on the reference side of the cell (notice in **Figure A.4 b** $I_o = I_r$).

A calibration is performed by measuring the temperature dependent transmission of the probe in ~10°C increments from ~30 – 70°C. Be sure to record the temperature of the cell (using the thermocouple) and *not* the temperature of the recirculating bath—these can differ by several degrees. At each temperature set point let the cell equilibrate for ~10 minutes (or until the cell temperature is constant) before measuring the probe intensity. Measure *both* the probe intensity in the absence of the cell ($I_d$) and through the cell ($I_r$). Make sure the cell temperature doesn't drift during the intensity measurement. Repeat these measurements for each temperature. Plot the transmission values ($I_r/I_d$) against the cell temperature and fit the data to a line. The fit equation relates the change in transmittance ($\Delta I$) with the change in temperature ($\Delta T$), and the slope ($m$) where the magnitude of a temperature jump is calculated as:

$$\Delta T = \frac{\Delta I_r}{m I_d}$$

**Figure A.4 b** shows a typical temperature dependent calibration curve. After the calibration, make sure to re-anneal the RNA by heating the cell to ~90°C and letting it slowly return to room temperature.

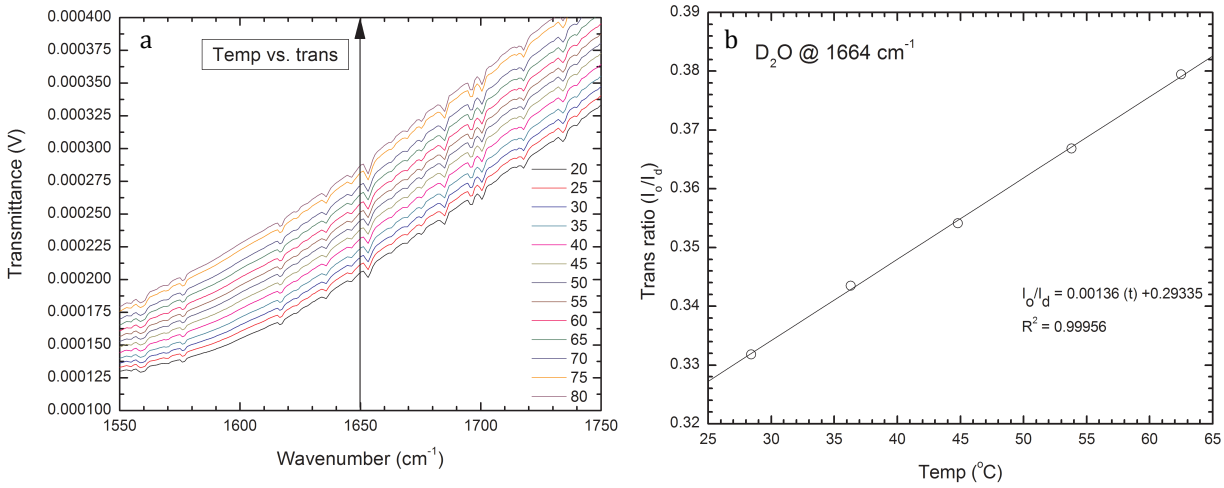


**Figure A.4: Temperature dependent infrared spectra for $H_2O$ and $D_2O$**
(a) The IR signal increases as a function of temperature over the entire region of interest. Temperature dependent spectral changes are indicated by the arrow. (b) A temperature calibration curve for $D_2O$ at 1664 $cm^{-1}$ with a linear fit and equation. Note: Here $I_o = I_r$.

## Overlapping pump and probe beams

Both the pump and the probe beam must be well behaved in order to ensure tight overlap in the laser interaction volume, reliable transient collection, and to maintain the integrity of the $CaF_2$ cell windows. The diameter of the pump beam is controlled by adjusting the position of a long focal length lens and should be ~5 times larger than probe. Place a pinhole (~100 μm diameter) at the sample cell position and align the probe beam so that it passes through it. The focusing OAP just before the sample (**Figure A.2 y**) is used to carefully adjust the probe beam diameter ("waist"). The procedure for measuring the beam waist is provided in Appendix C.

Once the beam diameter is minimized (~50 μm), the pump beam is aligned so that it overlaps it. To overlap the pump beam across the pinhole, first turn down the intensity

by rotating the polarizer. This will minimize damage to the aperture, and reduce the intensity of reflections, which can burn (ablate) nearby optics. Carefully adjust the mirror just before the polarizer (**Figure A.2 b**) until flashing from the laser can be seen centered on the pinhole. Now the pump beam and probe beams are roughly overlapped. Quickly turn off the pump beam to minimize further damage to the pinhole.

To finely tune overlap, the actual temperature jump signal needs to be maximized. Remove the pinhole and replace it with the IR cell. Turn on the pump beam and monitor the transient signal through the cell. Translate the cell forward or backwards (parallel to the probe beam) until the temperature jump signal is maximized on the oscilloscope. Now carefully adjust the flat mirror (**Figure A.2 b**) just before the polarizer and long focal length lens to finely tune the T-jump signal until it is optimized. When the signal is optimized, the polarizer can be rotated to increase the magnitude of the T-jump. Do this slowly while monitoring the cell for damage. A well-aligned instrument should give T-jumps on the order of 20°C.

### Aligning the cell

Proper cell alignment is critical for collecting quality data. Even a slight error in cell position can result in significant absorbance errors between the sample and reference. First, the cell must be aligned so that it is perpendicular to the probe beam. Second, and more importantly, it must be mounted so that translating it from side to side (i.e., between the sample and reference) does not change the cell's position relative to the pump/probe overlap. To align the stage, use a "dummy cell" containing just $D_2O$ (or deuterated buffer) in both cell compartments.

Using a dial indicator, rotate the cell so that when the stage is translated the indicator shows no forward or backward movement. Next, loosen the set screws on the stage mounts and rotate the entire apparatus (stage and cell) so that the HeNe retroreflection overlaps back through (just horizontally) the alignment apertures. Next, optimize the cell position in the beam cross sectional area as described above. To test the alignment, perform transient absorbance experiments on both sides of the cell and calculate the absorbance difference. If the cell is well aligned is should give an absorbance difference equal to zero (i.e., both sides of the cell experience the same T-jump magnitude) as indicated in **Figure A.5**.

Once the dummy cell is aligned, replace it with a sample cell. Rotate the sample cell until the HeNe retroreflection is overlapped with the alignment apertures, check the positional deviation with the dial indicator, and adjust as necessary. Finally, optimize the T-jump signal on the reference side and proceed with the experiment.

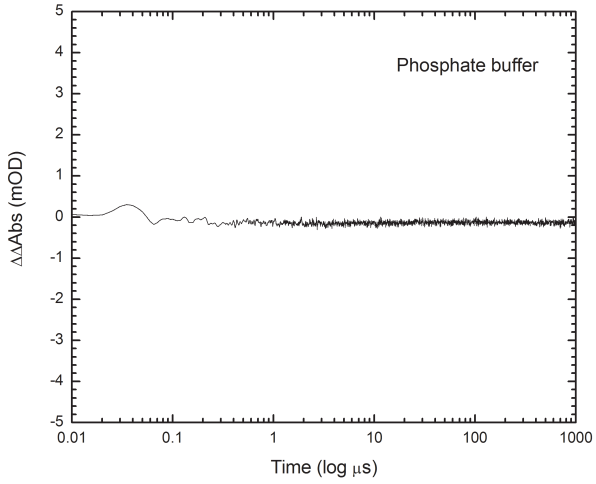


**Figure A.5: Transient absorbance data for a well aligned dummy cell**
A time-resolved absorbance difference centered on zero (±0.5 mOD) indicates that the "dummy" cell is well aligned.

# Appendix B

## Transient data collection and handling

Temperature jump measurements are collected for both the reference (buffer) and sample (buffer + RNA) sides of the IR cell. Starting with the reference side, measure the probe intensity through the cell ($I_r$). Next, turn on the pump beam and measure the T-jump magnitude on the oscilloscope. Once this is done, collect an average of ~2500 laser shots. Repeat this procedure for the sample side ($I_s$). Make sure to measure the full intensity of the diode in the absence of the cell ($I_d$) for each set of measurements. This value is needed to calculate the magnitude of a T-jump. As a final note, be on the look-out for "cavitation". This can occur when samples are not properly degased or the pump laser is not well-behaved (i.e., focused too tight, poor modal quality, hot spots) causing the formation of gas bubbles. The result is a large artifact that renders the data unusable as can be seen in **Figure A.6**.

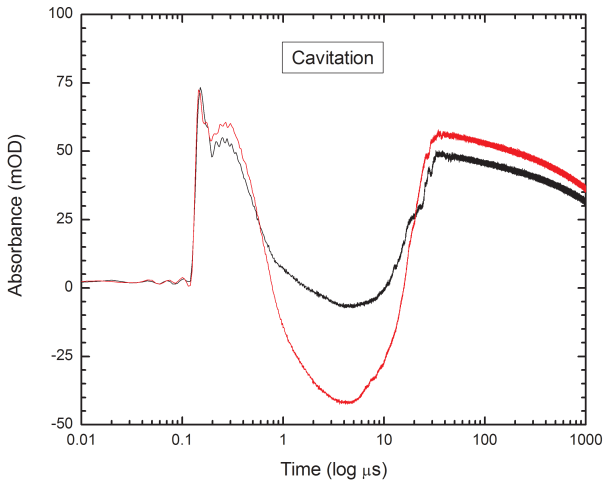


**Figure A.6: Affect of severe laser cavitation on IR transient signals**
Laser-induced photoacoustic cavitation is detected between ~1 to 10 μs and manifests as a rapid decrease in signal. Note: the y-axis above should be labeled as transient intensity (mV).

Raw transient IR intensity data for both the reference ($\Delta I_r$) and the sample ($\Delta I_s$) are shifted in time and baseline adjusted. The time shift adjusts the data so that the point at which the pump interacts with the sample is $t = 0$. The baseline adjustment ensures that the signal intensity prior to $t = 0$ is at zero. Since the digitized signals are collected over 262,000 points at 5 ns resolution, the point density is unnecessarily high (especially at longer times). This is quite cumbersome and can complicate data handling. To mitigate this, the data are interpolated to a logarithmically spaced time axis. The interpolated data for each (**Figure A.7 a**) can be converted to absorbance (**Figure A.7 b**) using the following equation:

$$\Delta A = -\log\left(\frac{I_o + \Delta I_o}{I_o}\right)$$

where $I_o$ is the steady state signal through the sample or reference solution ($I_o = I_r$ or $I_s$), and $\Delta I_o$ is the transient IR data subsequent to the laser induced T-jump ($\Delta I_o = \Delta I_r$ or $\Delta I_s$). The difference of these transient absorbance data gives the absorbance response of the RNA (**Figure A.7 c** and **d**):

$$\Delta A_{rna} = \Delta A_{sam} - \Delta A_{ref}$$

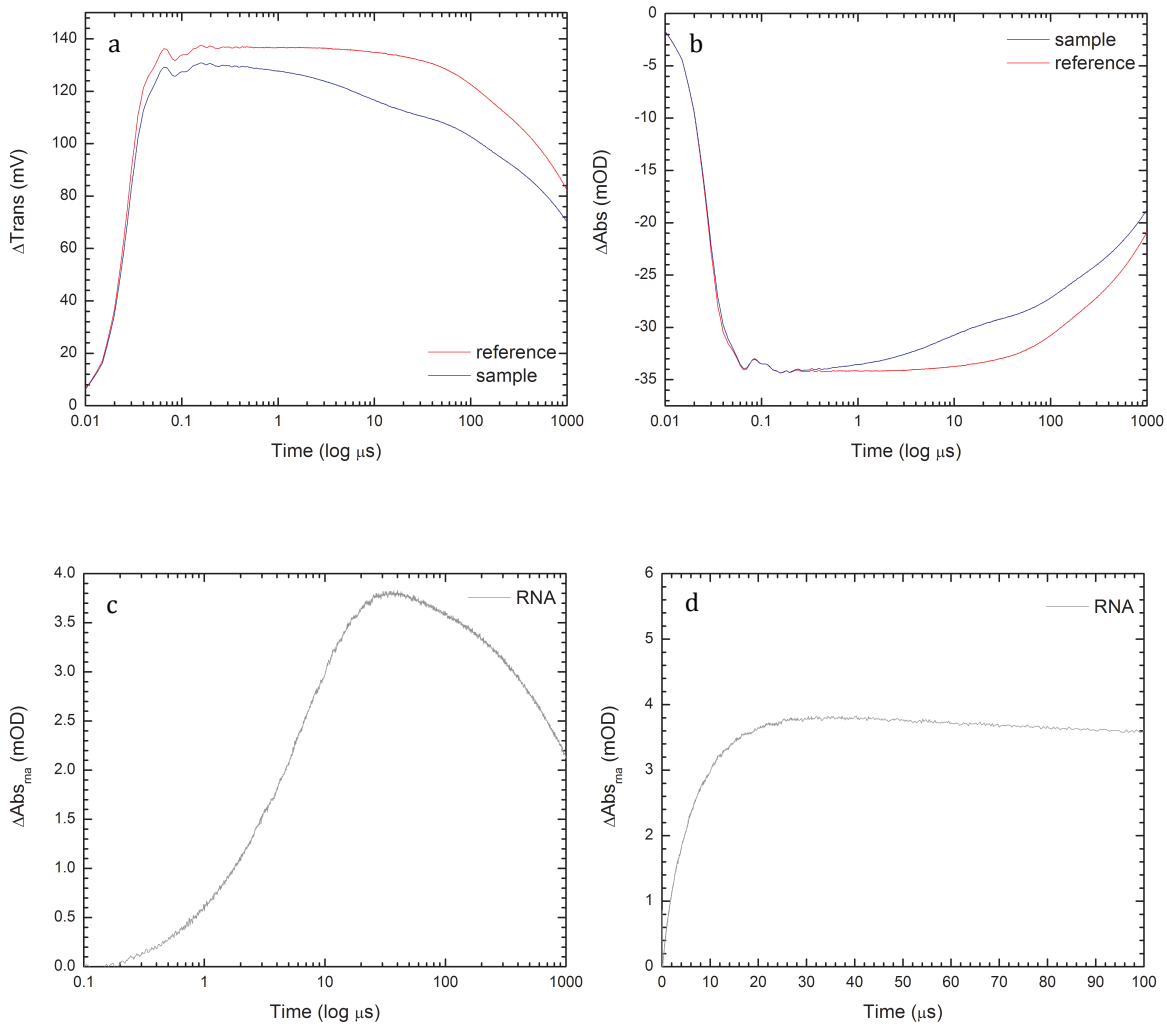


**Figure A.7: Time-resolved IR data handling**

(a) Adjusted and interpolated transient signals, and (b) the resultant transient absorbance data. (c) Transient RNA absorbance plotted on a logarithmic, and (d) linear time axis from 0-100 μs.

# Appendix C

## Optics

### Beam waist measurement

The beam waist can be easily calculated by measuring the transmission ($I/I_o$ where $I$ is the intensity through the pinhole and $I_o$ is the intensity in the absence of the pinhole) of the beam through an aperture of a known diameter. The beam diameter, $w$, is related to the aperture diameter, $a$, and the transmission according to the following equation:

$$\frac{I}{I_o} = 1 - e^{-2a^2/w^2}$$

This procedure is particularly convenient for measuring the beam waist of a focused beam. An alternative procedure (this one is particularly convenient for larger beams) to measure the beam waist is the knife-edge method. Mount a razor blade on translational stage and place it perpendicular to the beam. Translate the stage until the razor just begins to cut (attenuate) the beam. Measure the intensity on an oscilloscope. Move the razor in ~5 μm increments and measure the signal intensity at each position. Repeat this procedure until the entire beam is blocked and no light gets to the detector. Plot the intensity as a function of knife-edge displacement. If a beam profile is assumed (e.g., Gaussian), the beam diameter can be calculated from these data.

To calculate the theoretical spot size of a beam (i.e., how tight it can be focused) use the following equation:

$$d_o \cong \frac{2f\lambda}{D}$$

Where $d_o$ is the theoretical spot size, $f$ is the focal length of the focusing optic, $\lambda$ is the wavelength of light, and $D$ is the beam diameter before focusing. For example, if a beam of frequency 1650 $cm^{-1}$ (6.061 μm) with a diameter of 12 mm is focused with an optic having a focal length of 50 mm, its theoretical spot size is equal to 50.51 μm.

**Raman conversion**

Overall, a Nd:YAG laser produces a dependable, well-behaved pump beam with a Guassian profile. Unless done carefully, Raman conversion can deteriorate the beam quality resulting in hot spots than can damage optics and cell windows. Placing a Galilean telescope after the converter that expands (and then re-collimates) the pump beam can be helpful in minimizing the effect of potential hot spots. Another problem arises from feedback due to retroreflections through the converter. Feedback can minimized by changing the launch angle of the YAG fundamental as it enters the Raman converter. An incident launch angle perpendicular (normal) to the converter windows gives the most feedback. Skewing the launch angle slightly minimizes the feedback. Perhaps the most important aspect of Raman conversion is the pressure of the $H_2$ gas inside the converter. In addition, some commercial Raman converters (ours included)

have a fan to circulate the gas inside the converter. Although the fan tends to increase conversion efficiency at high gas pressures, it also leads to more pronounced hot spots. **Figure A.8** shows the output of the Raman converter as a function of gas pressure (in power and percent conversion), and with the fan on or off. Ultimately, the optimal conditions for the Raman convertor used in this work were at a pressure of 250-300 psi with the fan off.

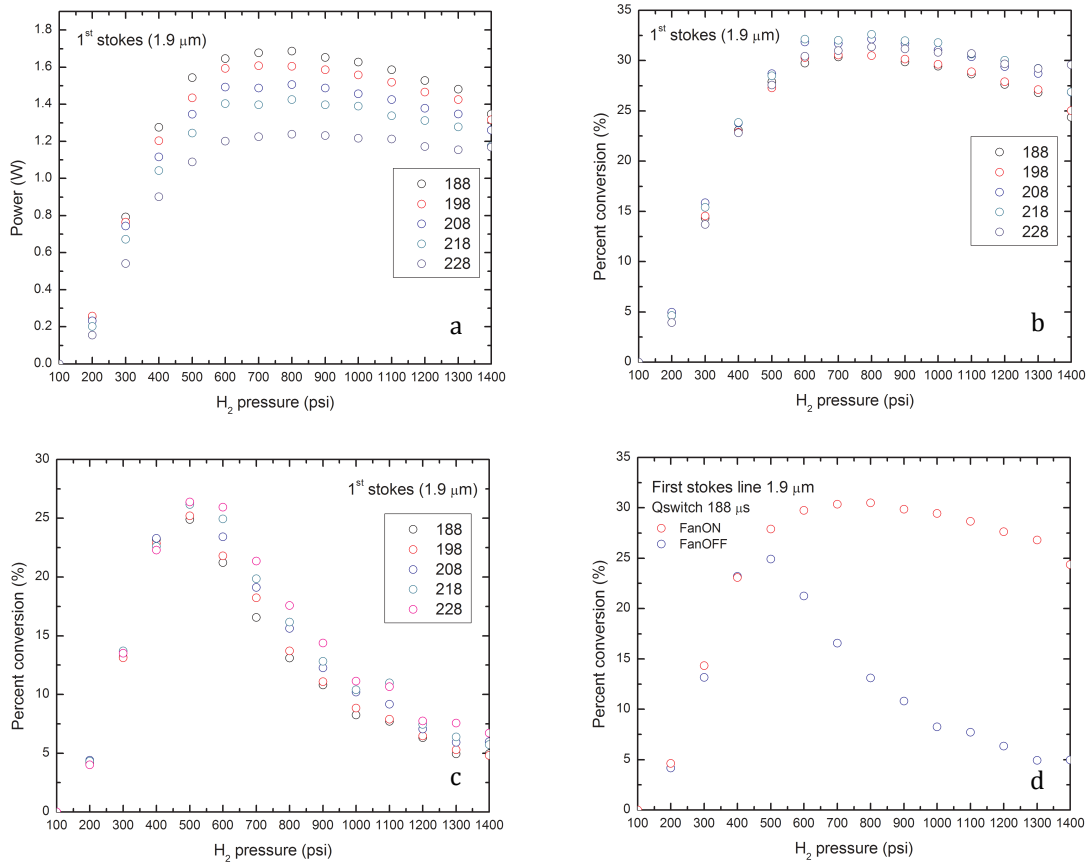


**Figure A.8: Raman conversion data for the first stokes line at 1.9 μm**
(a) Power of 1.9 μm as a function of $H_2$ pressure [fan on], and (b) the resultant percent conversion. (c) Percent conversion with the fan off, and (c) the fan both on and off. The colored symbols indicate different Q-switch values.

### Off axis parabolic mirrors

Off axis parabolic mirrors (OAPs) can be challenging to align. **Figure A.9** is a ray-tracing diagram showing the effect of various translational movements on the reflection of a divergent beam. This diagram can be used as a general guide to aligning (i.e., collimating, focusing etc.) OAPs.

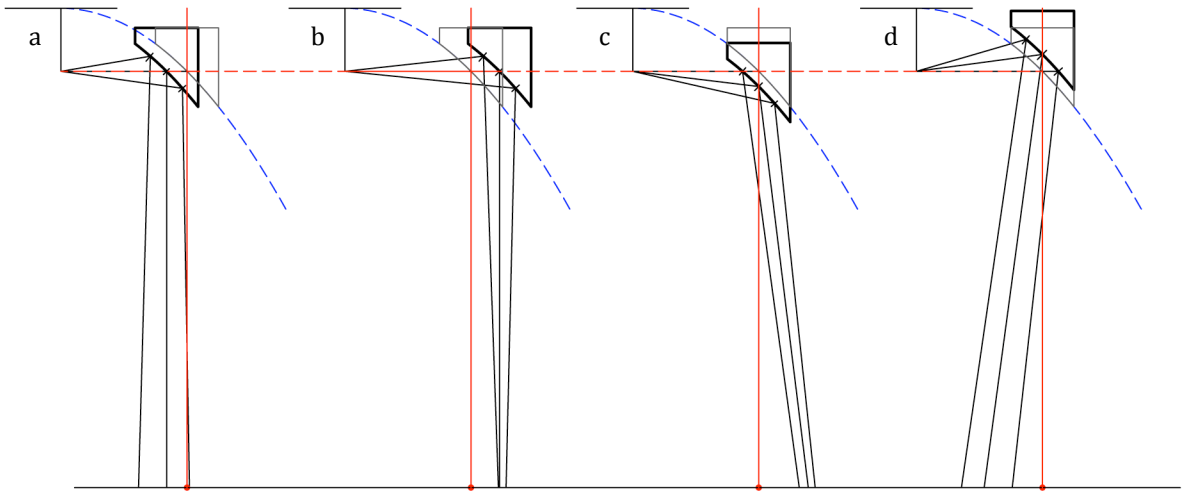


**Figure A.9: Off-axis parabaloidal reflector**

The dashed blue line shows the parent parabola. The light gray outline shows the OAP in the proper position relative to the divergent beam to collimate it. The darker black outline shows the OAP moved relative to the "correct" position (indicated by the light gray outline). (a) The effect of moving the OAP toward the source. (b) The effect of moving it away from the source. (c) The effect of moving it forward along the optical axis. (d) The effect of moving it backward along the optical axis.

# Copyrights

### RNA images

Harry Noller [harry@nuvolari.ucsc.edu]
To:stan5201@vandals.uidaho.edu

Aaron,

You are most welcome to use any of our rRNA or ribosome images in any way you like.

Best,
Harry Noller

### Manuscript 1



### Manuscript 2