Two-dimensional Infrared Spectroscopy of Nucleic Acids: Application to Tautomerism and DNA Aptamer Unfolding Dynamics

by

Chunte Sam Peng

B.S. Chemistry
University of California, Berkeley, 2008

Submitted to the Department of Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2014

© 2014 Massachusetts Institute of Technology. All rights reserved.

Signature redacted

Signature of Author: ________________

Chunte Sam Peng
April 25, 2014

Signature redacted

Certified by: ____________________________

Andrei Tokmakoff
Professor of Chemistry
Thesis Supervisor

Signature redacted

Accepted by: ____________________________

Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Studies
This doctoral thesis has been examined by a committee of the Department of Chemistry that included:

Professor John M. Essigmann
Chair

Professor Moungi G. Bawendi

Professor Andrei Tokmakoff
Thesis Supervisor

Signature redacted

Signature redacted

Signature redacted
Two-dimensional Infrared Spectroscopy of Nucleic Acids: Application to Tautomerism and DNA Aptamer Unfolding Dynamics

by

Chunte Sam Peng

Submitted to the Department of Chemistry on April 25, 2014
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

The structural dynamics of nucleic acids are intimately related to their biological functions; however, our ability to study these molecular dynamics has been largely impeded by the lack of techniques that possess both high time resolution and structural sensitivity. The motivation for the work in this thesis was to develop and apply two-dimensional infrared spectroscopy (2D IR) as a new experimental tool to investigate nucleic acid dynamics. Infrared spectroscopy is sensitive to structural changes of nucleic acids and 2D IR offers sub-picosecond time resolution. 2D IR spectroscopy is advantageous over the linear infrared absorption spectroscopy because the vibrational spectrum is spread onto two frequency axes, giving rise to the structurally sensitive cross-peaks. These cross-peaks allow the determination of vibrational couplings, which encode chemical bond connectivity, distance and orientation. However, 2D IR spectroscopy of nucleic acids is underdeveloped due to the difficulties in modeling highly delocalized and coupled vibrations of nucleobases. This thesis initiated the efforts to develop 2D IR spectroscopy of nucleic acids by first characterizing the 2D IR spectra and vibrational eigenstates of nucleobases, using a model of multiple anharmonically coupled oscillators. With pronounced cross-peaks existing between all the vibrations for a given nucleobase, 2D IR spectroscopy was shown to be capable of distinguishing between different tautomers, using pyridone as a model system. Coupled with a laser-induced temperature-jump (T-jump), 2D IR was used to monitor rapidly exchanging tautomers in real time under physiological conditions on the nanosecond timescale. Systematically characterizing the tautomer exchange rates as a function of various experimental variables lead to a two-state concerted mechanism involving bridging water wires for the lactam-lactim tautomerization of 6-chloro-2-pyridone. This method was then applied to study the tautomerism of a deoxycytidine analog, KP1212, which is an anti-HIV drug. Multiple tautomers, including the normally rare enol tautomers, were found under physiological conditions. This observation supports the rare tautomer hypothesis, which states that each tautomer displays a distinct base-pairing preference, eventually leading to mutations and population collapse of the HIV viruses. Beyond studies on the single
nucleotide level, 2D IR was used to characterize the structural dynamics of thrombin-binding aptamer (TBA), which is a 15mer DNA folded into a guanine-quadruplex (G-quadruplex). The 2D IR spectral signatures of G-quadruplex were established, and T-jump transient 2D IR was employed to investigate the unfolding dynamics of TBA. A mechanism of the early unfolding of TBA was proposed: A ~100 nanosecond response was attributed to the local deformation of the G-quadruplex, and a few-microsecond response was ascribed to be the fraying of the 3'-tail of TBA. This observation was consistent with a mechanism suggested by molecular dynamics simulations. Finally, the dissociation of double-stranded DNA formed by TBA and its complementary strand was found to be on the timescale of tens to hundreds of microseconds. The experiments in this thesis demonstrate the capability of 2D IR to investigate nucleic acid dynamics spanning a wide range of timescales.

Thesis Supervisor: Andrei Tokmakoff
Title: Professor of Chemistry
Acknowledgments

This thesis shall not start without thanking all the people who have helped and supported me throughout my graduate career.

I start with thanking my math and science teachers from high school who have helped me build a solid foundation and skills necessary for me to come this far. I thank my postdoc mentor from Berkeley, Oleg Kornilov, for encouraging me to take on projects that excite me the most, even if they are extremely challenging.

This thesis would have never been possible without my thesis advisor, Andrei Tokmakoff. I thank him for giving me the freedom to explore different projects during my early days in grad school, which ultimately led to the development of 2D IR spectroscopy on nucleic acids. Andrei's trust and confidence in me played a pivotal role, enabling me to persist in the unfamiliar territory of nucleic acid spectroscopy. His insightful suggestions and continuous encouragements have helped me through the frustrating moments. And his constant pursuit of perfection has inspired me to become the scientist I am today.

I am deeply indebted to the brilliant scientists surrounding me in the Tokmakoff group over the years. In particular I want to thank Kevin Jones, my graduate student mentor who was always patient with my questions and taught me everything I needed to know in order to operate the "boxcar" setup in the big lab. Kevin has a pair of magical hands that can always make the experiments work, and it was from him that I learned how to spot the weakest signal on top of a broad background of noise. Kevin was also invaluable in terms of keeping the spirits high in the office, with his unique humor and of course the Kevin Calendar. Ziad Ganim has led me into the computational side of spectroscopy and it was from him that I adopted the mindset of always trying to obtain a good match between the experiments and calculations. Joshua Lessing always reminded me that we are chemists (supposedly) by describing to me all the details of peptide synthesis. I overlapped with Poul Petersen only for a brief amount of time, but his advice on surviving through grad school with coffee has certainly served me well and given me a new morning ritual. Sean Roberts was an example of an extremely hard-working person who didn't appear to need sleep nor food. Although I never worked with Becky Nicodemus directly, she always said something encouraging to me. Krupa Ramasesha was the senior student who overlapped with me for the longest duration. Although she probably speaks the fastest in the group, she was always able to calm me down and give me useful advice and encouragement when I was stressed out. I really appreciate her support. Andrew Horning, Aritra Mandal and I are classmates who joined the Tokmakoff group at the same time. I thank both of them for working hard together on courses and qualifying exams during our first two years. Andrew's ability to take notes with time-stamps has never ceased to impress me. Aritra and I navigated through grad school together and reached the end of the tunnel at the same time, and I am really thankful for his help and encouragement along the way.

The second generation of group members I worked with has been equally important. Carlos Baiz joined the lab around the time when Kevin handed to me the heavy burden of
running the entire big lab by myself. Carlos’ deep knowledge in ultrafast spectroscopy has made the transition much smoother, and he has since helped me with many of my experiments, especially the challenging T-jump experiments. I really appreciate his invaluable help and advice. Mike Reppert and Luigi De Marco joined the group at the same time. I admire their constant pursuit of the fundamental theory behind how everything works as they would argue for hours in the office in order to get all the equations worked out. I share their vision of how science should be done, and they have both provided me with valuable suggestions and support. Mike has always been my go-to guy for theory and computational problems, and I’m thankful for his help. I look forward to seeing Luigi apply broadband 2D IR to study nucleic acid dynamics in an “over-engineered” fashion and I’m certain it will be fabulous. Paul Stevenson took over the “boxcar” setup since we moved to Chicago and has really demonstrated his experimental talents. His ability to set up the system allowed me to focus on thesis writing and I’m really thankful for the hard work he has put in. I look forward to seeing him maintain the “boxcar” setup as the most reliable workhorse in the group and produce exciting results on membrane proteins. Recently I have been working closely with Paul Sanstead on exploring new possibilities of using 2D IR to study nucleic acid systems. I have no doubt of his future success as Paul has shown excellent work ethics and abilities to learn new things quickly. Ann Fitzpatrick is a relatively new postdoc but has quickly become the source of joy in the group. She has never hesitated to help others and I greatly appreciate her support and numerous proofreading on my thesis chapters. Ann has also built the “compact 2D IR” setup with Paul Sanstead and I’m eager to see the new science produced from this spectrometer. Martin Thämer and Denise Schach are two other new postdocs, and I thank them for their encouragement during my thesis writing and for sharing their grad school experiences. It has been a great pleasure to work with this group of excellent scientists and I will miss both Room 6-030 and E001.

I thank my collaborators from the Essigmann group: John Essigmann, Vipender Singh, Deyu Li, and Bogdan Fedeles for initiating the collaboration that really turned into multiple fruitful projects and allowed us to apply 2D IR on biologically relevant problems. The different perspectives brought by everyone have stimulated a lot of new ideas, and using a combination of various techniques to tackle a single problem has furthered our understanding of tautomerism in nucleic acid systems. I also thank Kimberly Hamad-Schifferli for introducing us to the thrombin-binding aptamer problem which motivated our work on studying DNA unfolding dynamics using 2D IR. Kim has provided much help in terms of the background knowledge and sample preparation of TBA.

I thank Professor Robert Silbey for being a brilliant scientist, excellent teacher, and a caring mentor. I had the honor to take his class in Quantum Mechanics and TA for him, which were wonderful experiences. I thank Professor James Fujimoto and Professor Franz Kaertner for their courses on ultrafast optics that taught me the essential ingredients to learn 2D IR spectroscopy, and I thank Prof. Fujimoto’s lessons on the expectations of a grad student that allowed me to understand my role as a first year student.

I thank two of my best friends from the same class at MIT Chemistry, Jennifer Yao and Jian Cui. We met at the HHMI TA training program for 5.111 in the summer of 2008. Besides the extreme excitement, I still remember the nervousness of transitioning to a new environment, especially a place like MIT. Having friends to navigate through the first few
months of uncertainty definitely made the transition easier, and our friendship lasts ever since. The occasional competition between our sections’ average exam score with Jen, the 5 am struggle on Fujimoto’s nonlinear optics problem sets with Jian, and many other interactions are all great memories of grad school.

I am grateful to the staff I’ve worked with both at MIT and the University of Chicago. Peter Guinta is always on top of purchase orders and reimbursements, which saves us a lot of efforts. His sense of humor also brings laughter to the office. John Phillips, the grandfather of GCIS, provides us with the most comfortable working environment, especially the office where I spent most of my time during the last year. I really appreciate how much he cares about us given the amount of trouble he went through just to give me an extra 2” of desk space (which was much needed!). Tanya Shpigel is extremely efficient in handling all sorts of matter and she has been tremendously helpful in making our lives easier.

I am thankful for my friends who have added so much fun to my life outside of the lab. I’ve met countless number of friends in the Taiwanese community through ROCSA. Every gathering event and every interaction with you has always reminded me of how great home is. The list of people goes on forever, but special acknowledgment goes to Hung-Wen Chen and Ya-Hui Chang, the ex-“co-President” of ROCSA when I served as the Vice Present during my third year in grad school. Together we have held various events, including the first Sports Day between the Taiwanese student associations from MIT and Harvard. We’ve also shared joyful memories at various restaurants, house parties, birthday celebrations, and trips to New York City, Montreal, and Orlando. I want to thank my roommate Syuan-Ming Guo who was also my high school classmate. I wouldn’t have gone on studying chemistry without his encouragement back in high school, and I couldn’t be more excited that several years later we are both pursuing a PhD in Chemistry at MIT! The countless number of nights of us cooking together, discussing chemistry and life has made the entire process much more enjoyable. I also acknowledge Shu-Wei Chang and Chia-Ching Chou for being friends whom I can talk to not just about interesting things in life, but also biomolecular dynamics (I know, this is the beauty of being at MIT).

As all of my friends know that basketball competes with lab work as my second life, I express my gratitude for my teammates in ROCSA basketball team (Yu-Chih Ko, Hung-Wen Chen, Paul Lin, Po-Yen Chen, Jonathan Leu, Jonathan Liu, Duck, Owen Chen, Chih-Chao Chuang, Henry Wei, Chien-Jen Lai, Hsiang-Chieh Lee, Yu-Chi Kuo) and Chemistry IM team (Jian Cui, Jeff Eliason, Russ Jensen, Gabe Juarez, Hyo Jun Han, Dave Song, Jon Medley, Wankyu Lee, Bryce Anderson). In the 2011-2012 season, with heart, sweat, grit-and-grind, the Chemistry IM team made it to the Finals after rounds of challenging competitions. Barely losing to the young and athletic group of undergrads didn’t take away any pride and honor we deserved. The following year, we had a perfect undefeated regular season record, albeit the eventual playoff loss, which was the perfect farewell gift before I moved to Chicago.

Recently I’ve met a group of new friends, the “HaHa Gang” (Harry Ha, Grace Lee, Yung-Tsen Chen, Chih-Chieh Hsu, Victor Lee, Shang-Huei Liang, Hsin-Yu Chen, Chien-Hung Lin), you have added much warmth to the unbearably cold Chicago winter. I also thank the family of my cousin Ching-Yu, whose husband happened to be doing a one-
year visiting scholar in the Math Department at the University of Chicago during my stay here. I was often invited to their home for dinner and spent time with the kids, David and Emmy, which has always helped me forget the pressure of thesis writing.

I thank Joyce, not only for her PAGE gel experiments shown in this thesis, but most importantly for her beloved company and continuous support, both in science and in life, since our days at Berkeley. We have grown and learned life lessons together over the years. It wasn’t an easy decision for you to leave the beautiful Bay Area, but I’m glad that we explored the equally wonderful Boston and New England area together. I’ve enjoyed every single one of our travels, be it a road trip with in-car karaoke or a bumpy Fung Wah bus ride to New York City. Your optimism, beautiful singing and piano playing have always cheered me up during the frustrating times of grad school. You’ve taught me how to enjoy and cherish this beautiful world by paying attention to even the tiniest detail in life such as budding flowers and fallen leaves. Hakuna Matata is the best phrase to describe you, and I thank you for sharing this wonderful philosophy with me.

Finally, I would like to express my gratitude for my family. I want to thank my parents for raising me up and always providing me with the best educational opportunities. Back in high school, both my brother and my parents strongly encouraged me to come to the US for college and that was how I began the first chapter of my exciting journey in physical chemistry.

The past six years have been an influential part of my life, and I truly appreciate everyone’s participation in any form.

THANK YOU ALL!

-Sam
Contents

CHAPTER 1 Introduction ................................................................................................. 15
  1.1 Dynamics of nucleic acids .................................................................................... 16
    1.1.1 Tautomerization ......................................................................................... 17
    1.1.2 Nucleic acid folding .................................................................................. 20
    1.1.3 Nucleic acid-protein recognition ............................................................... 22
  1.2 Dynamics experiments ........................................................................................ 25
  1.3 2D IR spectroscopy of nucleic acids ..................................................................... 26
    1.3.1 Comparison to 2D IR spectroscopy of proteins ............................................ 29
    1.3.2 Development of 2D IR spectroscopy of nucleic acids .................................. 31
  1.4 Thesis outline ...................................................................................................... 33
  1.5 References .......................................................................................................... 34

CHAPTER 2 Theoretical Background of 2D IR Spectroscopy ...................................... 41
  2.1 Introduction .......................................................................................................... 41
  2.2 Nonlinear response function formulism .............................................................. 42
    2.2.1 The third-order polarization .......................................................................... 42
    2.2.2 The third-order nonlinear response .............................................................. 44
    2.2.3 Phase-matching and time-ordering .............................................................. 46
  2.3 Six-level model system ........................................................................................ 47
    2.3.1 Peak position and sign ............................................................................... 49
    2.3.2 Lineshape .................................................................................................... 52
  2.4 Polarization-selective 2D IR spectroscopy ........................................................ 53
  2.5 Acknowledgments ............................................................................................... 62
  2.6 References .......................................................................................................... 63
CHAPTER 5 Identification of Lactam–Lactim Tautomers of Aromatic Heterocycles Using 2D IR Spectroscopy

5.1 Introduction

5.2 Tautomerism of pyridone derivatives in aqueous solution
   5.2.1 Equilibrium spectra and peak assignments
   5.2.2 DFT calculation: relative stabilities of the lactam-lactim tautomers
   5.2.3 Temperature dependence of 6-chloro-2-pyridone tautomerism
   5.2.4 Tautomerism of 4-pyrimidinone

5.3 Observation of oxythiamine tautomers
   5.3.1 Background of oxythiamine and its recognition by the TPP riboswitch
   5.3.2 Oxythiamine tautomers in the unbound form
   5.3.3 Tautomeric form of OxyTPP bound to the TPP riboswitch

5.4 Dimerization of pyridone in organic solvents
   5.4.1 Lactam dimers of 2-pyridone
   5.4.2 Lactim dimers of 6-chloro-2-pyridone

5.5 Acknowledgments

5.6 References

Appendix 5.A: Determination of equilibrium constant for 6-chloro-2-pyridone

CHAPTER 6 Direct Observation of Ground State Lactam–Lactim Tautomerization Using Temperature-Jump Transient 2D IR Spectroscopy

6.1 Introduction

6.2 Equilibrium measurements

6.3 Transient 2D IR spectroscopy
   6.3.1 Temperature dependence
   6.3.2 Solvent dependence
6.3.3 2-Chloro-4-pyridone ................................................................................... 178
6.3.4 pD dependence .......................................................................................... 182
6.4 Comparison with prior results ........................................................................ 184
6.5 Conclusion ........................................................................................................ 186
6.6 Acknowledgments ........................................................................................... 187
6.7 References ....................................................................................................... 188

CHAPTER 7 Monitoring Rapidly Interconverting Tautomers of an ...................
Anti-HIV Agent KP1212 ..................................................................................... 191
7.1 Introduction ....................................................................................................... 191
7.2 Results ............................................................................................................ 193
  7.2.1 Extracting the pKa’s from the pH-dependent FTIR spectra ...................... 193
  7.2.2 Separating tautomers using 2D IR spectra ............................................. 196
  7.2.3 Peak assignments based on DFT calculations ........................................ 202
  7.2.4 Tautomeric distribution probed by NMR spectroscopy ......................... 210
  7.2.5 Measuring the tautomerization kinetics with T-jump experiments ......... 211
7.3 Discussion ....................................................................................................... 213
  7.3.1 Proposed mechanism for the lethal mutagenesis of KP1212 ................. 213
  7.3.2 Connection between the tautomer populations and mutagenicity ......... 216
  7.3.3 Tautomerization kinetics ......................................................................... 217
7.4 Outlook ........................................................................................................... 219
7.5 Acknowledgments ........................................................................................... 219
7.6 References ....................................................................................................... 220
Appendix 7.A: SVD analysis on simulated FTIR ................................................. 223

CHAPTER 8 Vibrational Signatures of Base-pairing and KP1212
  Oligonucleotides .................................................................................................. 229
8.1 Introduction ....................................................................................................... 229
8.2 Vibrational signatures of Watson-Crick base-pairs ...................................... 229
  8.2.1 AT base-pairs .......................................................................................... 229
CHAPTER 9 Vibrational Signatures of G-quadruplex Revealed by 2D IR Spectroscopy

9.1 Introduction to G-quadruplex ................................................................. 257
9.2 Basic characterization of G-quadruplex .................................................. 259
9.2.1 Parallel and anti-parallel G-quadruplexes ......................................... 259
9.2.2 Anti and syn glycosidic configurations .............................................. 261
9.2.3 Basic characterization of thrombin-binding aptamer ......................... 262
9.3 Equilibrium IR spectroscopy of G-quadruplexes .................................... 266
9.3.1 Infrared absorption spectroscopy ...................................................... 266
9.3.2 2D IR spectroscopy ........................................................................... 278
9.4 Acknowledgements .................................................................................. 286
9.5 References .............................................................................................. 287

CHAPTER 10 Dynamics of TBA Unfolding and TBA-antidote Dissociation

10.1 Introduction ............................................................................................ 293
10.2 Equilibrium temperature-dependent FTIR spectra of TBA and tTBA .... 296
10.3 T-jump transient 2D IR of TBA unfolding ............................................. 300

Appendix 8.A: DNA oligomer sample preparation ........................................... 256
Curriculum Vitae
CHAPTER 1
Introduction

Nucleic acids, namely DNA and RNA, can adopt various three-dimensional structures that are stabilized by hydrogen bond (H-bond) and stacking interactions between the nucleic acid bases. H-bonds associate base-pairs in Watson-Crick, Hoogsteen, and Wobble geometries, which collectively lead to secondary structures such as helices, loops, hairpins, double helices, and junctions. These base-paired secondary structures in turn fold into well-defined tertiary structures such as pseudoknots, and G-quadruplexes. Their sequence dependent structural variation as well as flexibility are important components in biological processes. Beyond the storage and processing of genetic information in DNA replication, packaging, and transcription, there is an ever growing knowledge of how higher nucleic acid structures actively participate in biological processes. In particular, there has been a rapid growth of our knowledge about the biological functions of RNA in the past two decades. Initially, the ribosomal RNA was believed to merely serve as the scaffold for the essential ribosomal proteins, when in fact the ribosomal proteins actually provide the scaffold for the ribosomal RNA to function, such as catalyzing peptide bond formation. The stem and loops of nucleic acids have been found to bind to proteins and regulate their functions. There is increasing awareness of interplay of DNA, RNA, and proteins are essential in many different biological processes, and therefore there has been a big thrust in experimentally probing these interactions.

The ability to predict, manipulate, and monitor the atomic structures of these biomolecules has led to technological applications using artificial DNA self-assemblies, such as drug delivery and molecular devices. A predictive understanding of the interactions
between the nucleic acid bases is crucial to improve the design of DNA nano-constructs. Furthermore, there is an urgent need for testing and improvement of computational methods for simulating conformational dynamics of nucleic acids. In particular the force fields for nucleic acids remain significantly underdeveloped compared to proteins.

Traditional methods such as X-ray crystallography and NMR spectroscopy have been the major workhorses in studying nucleic acid structure and biophysical processes. Nevertheless, a large range of nucleic acid dynamics is not accessible by these techniques despite their excellent structural resolution. The work in this thesis is motivated by developing two-dimensional infrared spectroscopy (2D IR), an emerging technique for studying molecular dynamics, as a new and complementary tool to investigate the conformational dynamics of nucleic acids that are not easily accessible by traditional methods, as well as to bridge the gap between experiments and theory. The projects presented in this thesis aim to understand the fundamental physical principles underlying different nucleic acid dynamics, specifically tautomeration and DNA folding. This work involves development of the methodology using simple model systems and biophysical studies relevant to problems in drug design and bioengineering.

1.1 Dynamics of nucleic acids

The dynamics of nucleic acids span a wide range of timescales, from local fluctuations on the picosecond timescale to large tertiary structural changes that can take as long as minutes. Figure 1.1 shows an overview of various time-dependent processes and the techniques which can access the corresponding timescales. It is clear that an inverse correlation exists between the timescale of the motions and the length scale of the structures involved. Short-range local motions on the femtosecond to picosecond timescale include chemical bond vibrations, torsions, and librations, as well as fluctuations in base orientation and their associated H-bonds. Solvation dynamics, molecular rotation of bases and small oligonucleotides, translational diffusion over nanometer scales, single-stranded base-stacking, and addition of a base-pair to a double helix are on the nanosecond timescale. Tautomerization also occurs on this timescale, which is faster than the nucleobase diffusion into the polymerase active site and the rate of polymerase moving through a base-pair. Conformational changes to the secondary structure such as hairpins and loops are on the microsecond timescale. Double-strand association and global changes to the tertiary structure occur after secondary structure changes. This thesis presents studies of two categories of nucleic acid dynamics within the
diverse set of processes, tautomerization and DNA aptamer folding, which we describe in more detail below.

Figure 1.1: Overview of the range of time-scales experimentally observed for a variety of nucleic acid dynamics.

1.1.1 Tautomerization

Tautomerization refers to hydride rearrangement reactions within nucleobases, a specific class of proton-coupled electron transfer process. This process is often presented as the shift of a proton between basic nitrogens or oxygens of the nucleobase, which is accompanied by shifts in resonance structures within aromatic heterocycles. As a result different tautomers of a nucleobase will have marked different physical and chemical properties. Although the relevance of tautomerism to DNA nucleic acid biochemistry has been established for decades, there is a growing appreciation of the significant potential biological implications and breadth of processes influenced by tautomerism. Watson and Crick
noted that if nucleobases occasionally take on the non-canonical enol or imino tautomeric forms, the H-bond interactions can be altered, leading to formation of base-pair mismatches and spontaneous mutations.¹⁴-¹⁷ This rare-tautomer hypothesis has proved to be a compelling description of the chemical origins of mutations and genetic evolution, but definitive experimental evidence to support this proposal has been sparse. Recent crystal structure of a DNA polymerase with a C:A mismatch at the base incorporation site was reported, fifteen which suggested the presence of imino tautomer of either C or A. It has also been speculated that tautomerization plays a role in the catalysis and binding of RNA enzymes and aptamers.¹⁸,¹⁹ Small self-cleaving RNA enzymes such as Hammerhead, Varkud Satellite, Hairpin and GlnS are proposed to utilize either the ionized forms or minor enol or imino tautomers of conserved guanine to perform their catalytic functions.¹⁰,²⁰-²³

Despite the biological significance, it is challenging to study tautomerism under physiological conditions. One of the major limitations in identifying minor tautomers is the lack of experimental methods that are able to distinguish interconverting tautomeric species. Tautomers cannot be separated by simple chemical methods; rather, a non-contact in situ probe is required. For example, UV-vis spectroscopy has been used extensively.²⁴-²⁶ Electronic spectra, although relatively easy to acquire, are often broad and featureless, which complicates their assignment and the analysis when multiple tautomers co-exist. X-ray crystallography generally cannot determine accurate proton positions, limiting its use for directly identifying short-lived or weakly populated tautomers.¹⁹,²⁷ NMR spectroscopy can clearly distinguish tautomers since different atomic arrangements result in distinct chemical shifts. However, the observation of minor tautomers in aqueous solution and at room temperature becomes challenging because the exchange rates between tautomers can be rapid compared to the NMR time-scale and lead to motional narrowing.²⁸ Improvements on these spectroscopic techniques are crucial to understand tautomerization processes under physiological conditions. Finally, we note that quantum mechanical calculations are an important tool for comparing the relative stability of tautomeric systems;²⁹,³⁰ nevertheless, their predictions have not been tested extensively against experiment. Different levels of theory can lead to calculated energy differences of as much as 5 kcal/mol, and variation in the relative stability of species.³¹,³²

In this thesis, we developed 2D IR spectroscopy as a new tool to characterize tautomerization of aromatic heterocycles and applied this technique to two specific biological problems involving tautomerization. The first project investigated the tautomerization of oxythiamine (OxyT), which is an oxidized form of thiamine, commonly known as vitamin B1 (Figure 1.2a). Thiamine pyrophosphate (TPP) is an important cofactor in the metabolism of sugars and amides. The TPP riboswitch binds to TPP and negatively regulates the expression
of genes involved in the biosynthesis of TPP. X-ray crystallographic studies showed that both TPP and OxyTPP bind to the TPP riboswitch with similar H-bond interactions (donating a H-bond from the 4' position), suggesting that the bound OxyTPP exists in the rare enol tautomeric form. Nevertheless, given that the resolution of the crystal structure was not enough (2.6 Å) to determine the precise proton position, the exact tautomer bound to TPP is still under debate.

Figure 1.2: (a) Structures of thiamine (T) (top), oxythiamine (OxyT) (middle) and oxythiamine pyrophosphate (OxyTPP) (bottom). (b) Structures of 2'-deoxycytidine and KP1212. (c) Base-pairing schemes for different KP1212 tautomers.

The second project was aimed at studying the tautomerization of a deoxycytidine analog, 5,6-Dihydro-5-aza-2'-deoxycytidine (KP1212, Figure 1.2b). KP1212 is an anti-HIV drug candidate based on a pharmaceutical strategy called lethal mutagenesis. Antiviral drugs utilizing lethal mutagenesis are designed to increase viral mutation rates, which leads to viral population collapse. KP1212 is currently in phase IIa clinical trial and has been shown to induce A-to-G and G-to-A mutations in the genomes of replicating HIV. Despite its promising antiviral activity, the molecular mechanism by which KP1212 promotes mutagenesis has been elusive. The leading proposal is the rare tautomer hypothesis, which states that KP1212 exists in multiple tautomeric forms and each tautomer displays a distinct base-pairing preference during replication which eventually results in the observed mutations. For example, the canonical keto-amino form could base-pair with G while the keto-imino form could base-pair with A (Figure 1.2c).
A literature survey of these important biological problems clearly shows that there is an urgent need of new experimental technique to study tautomerization of nucleic acid bases and other aromatic heterocycles, especially under physiological conditions where the tautomerization rate increases and where water plays an active role in the proton-coupled electron transfer process. The major challenge is the lack of techniques that possesses both the time resolution and structural sensitivity to probe the rapidly exchanging tautomers.

### 1.1.2 Nucleic acid folding

Biomolecular folding is a transition from an ensemble of large number of disordered states into a well-defined native structure. Thermodynamically, the folding process is a delicate balance between the gain of favorable enthalpy (by making native contacts) and the loss of configurational entropy. The folding problem of proteins and nucleic acids, especially RNA, has attracted great interests both experimentally and theoretically. A simple comparison to protein folding may imply that RNA folding be relatively straightforward to understand since there are twenty different amino acids while there are only four different nucleobases. The amino acid side chains are markedly different structures and can be polar, hydrophobic, or charged; while the nucleobases are structurally similar and are predominately hydrophobic, although they can form inter-base H-bonds. Nevertheless, RNA folding is also a remarkably complex process that involves the interplay of electrostatic interaction, conformational entropy, base-stacking and H-bond interactions. Just as proteins can adopt various secondary structures such as α-helices, β-sheets, β-turns, and random coil, RNA can also exist in helices, loops, hairpins, bulges, and junctions. Owing to the negative charges on the phosphate backbone, RNA folding is highly dependent on the ion type and concentration in the environment.

Although the driving forces for folding of proteins and nucleic acids are different (hydrophobic collapse for proteins and base-stacking for nucleic acids), multiple competing interactions involved in both folding problems cause their energy landscapes “rugged,” meaning that there are multiple local minima or kinetic traps separated by energy barriers. Besides energetic frustration, proteins and RNA experience topological frustrations that arise from the multiple structures that can be formed during the folding process. RNA folding has been described as a hierarchical process because the secondary structures are more stable than tertiary structures (the two steps can be experimentally separated by the presence of Mg$^{2+}$). The locally favorable secondary structures can pack in various ways to give different tertiary folds. As a consequence, an outstanding question remains in both the protein folding and
nucleic acid folding fields, which is how to accurately predict the tertiary structure from the primary sequence.\textsuperscript{13,14,44}

The hairpin structure is ubiquitous in single-stranded DNA and RNA and its thermodynamics and kinetics have been studied for more than four decades; nevertheless, our understanding of the kinetics and mechanism remains poor.\textsuperscript{13,45} The hairpin formation is often described with a two-state model with Arrhenius temperature dependence for the rates of formation and dissociation.\textsuperscript{14} However, there have been an increasing numbers of studies showing non-Arrhenius behavior and stretched exponential kinetics, therefore invoking the use of a rugged energy landscape or multi-state model to explain the data.\textsuperscript{44,46}

The early unfolding event of a DNA aptamer is studied in this thesis, which has been proposed to involve the fraying of a single-stranded tail.\textsuperscript{17} The specific nucleic acid system is thrombin-binding aptamer (TBA). As shown in Figure 1.3, TBA is a single-stranded DNA 15mer (5'-GGTTGGTGTGGTTGG-3') that folds into a chair-like structure which is stabilized by guanine-quadruplex (G-quadruplex). G-quadruplex consists of layers of G-quartet (Figure 1.3c), which is formed by four guanine bases connected by eight cyclic Hoogsteen H-bonds. The two G-quartets in TBA are linked by TT and TGT loops. Again, the folding of TBA has been described on the basis of the two-state model, the folded and unfolded states, without details of how the G-quadruplex is assembled and how the loops rearrange. Our goal in this thesis is to elucidate the dynamics and build a mechanistic description of how TBA folds.

In addition to understanding the fundamental physical driving forces for folding, it is also essential to study DNA aptamers being used in biomedical applications. TBA in particular has attracted particular attention as they are alternatives to traditional anticoagulant agents.\textsuperscript{48}
Figure 1.3: (a) X-ray crystal structure of TBA bound to thrombin (PDB code: 1HAO). The blue and magenta bases in TBA are guanine and thymine, respectively. (b) Chair form of thrombin-binding aptamer. The yellow rectangles represent the base moieties of guanine. (c) Structure of the G-quartet that is stabilized by Hoogsteen H-bond network and a potassium ion placed in the center.

The currently available anticoagulant drugs, mostly decades-old medications such as heparin, warfarin and aspirin, work to inhibit various proteases in the coagulation cascade. However, their clinical application has been limited by the common side effect of bleeding. TBA has been found to bind thrombin and inhibits its bioactivity of converting the soluble fibrinogen into insoluble fibrin. One of the most important advantages of using aptamers to inhibit thrombin is that an antidote can be easily designed by the use of its complementary strand, which forms a DNA duplex with TBA, releasing it from thrombin and reversing the drug effect. Moreover, DNA aptamers can be commercially synthesized and chemically modified, enabling conjugation to species such as nano-particles (NPs). One can then trigger the release of TBA by exciting at the surface plasmon resonance of gold NPs to induce melting and release a payload. Aptamer/antidote pairs have been employed to target thrombin both in vitro and in vivo. A better understanding of the folding of TBA would help engineer more effective next-generation anticoagulants.

1.1.3 Nucleic acid-protein recognition

The binding between thrombin and TBA falls in an example of molecular recognition—an important, long-standing problem in biophysics and ubiquitous process in molecular biology. Molecular recognition refers to the association of two molecules in a
specific and biologically functional form. This includes protein oligomerization, protein-protein interaction, enzyme-substrate binding, and protein-DNA interaction. Numerous studies have been performed on these complexes by characterizing their physical properties through changes in the amino acid composition, hydrophobicity, electrostatics, and H-bonds, or by comparing the interface sizes, planarity and complementarity. We are interested in studying the dynamics of how molecules recognize each other in solution, the associated conformational changes, and the free energy landscape of these processes.

To a first approximation, one could model the two molecules as spheres having uniform reactivity and calculate the upper bound for collision rate constant from the diffusion-limited association rate $k_a$, given by the Einstein-Smoluchowski equation,

$$k_a = 4\pi(D_A + D_B)(r_A + r_B)$$

where $D$ is the diffusion constant and $r$ is the radius of the spherical molecule. Since the diffusion constant is inversely proportional to the radius (a larger molecule is an easier target but also diffuses slower), the overall association rate constant does not depend on $r$. Typical values lie within the range of $10^9-10^{10}$ M$^{-1}$s$^{-1}$, and decrease by at least three to four orders of magnitude if stringent orientational constrains are imposed. This gives a timescale of $>0.1$-1 $\mu$s for two molecules encountering each other at 1 mM concentration.

Traditional models for molecular association includes “induced-fit” and “conformational selection”. When applied to TBA-thrombin binding, the “induced-fit” mechanism proposes that TBA makes contact with thrombin when in the unfolded form (random coil), and then binding takes place as TBA folds into the rigid chair-form which stabilizes the specific binding interactions. On the other hand, TBA may be required to fold into the chair-form before it can recognize thrombin in solution. This scenario is known as “conformational selection”, which states that molecular binding only occurs for preselected conformers that are in the native folded form.

However, the association rates observed from the experiments, especially for protein-DNA binding, are faster than expected from the diffusion model. The configurational search space in molecular recognition is intractably large. Beyond the global gradient of a free energy surface which guides formation of the desired configuration of the stable protein-DNA complex, one kinetic advantage which speeds up the association rate has been described through the “fly-casting” mechanism. This mechanism suggests that the search rate is enhanced via the long-range attractive interaction, and high affinity of the diffusing protein to the surface on DNA, leading to one or two dimensional diffusive sliding. In the case of thrombin-TBA binding, TBA must remain partially unfolded in order to allow an unstructured
extended chain to extend to a greater capture radius which weakly makes the initial contacts with thrombin through non-native interactions at a large distance, followed by folding as the TBA approach binding sites.

The physical driving forces and details of the interactions for TBA-thrombin binding remain unclear. There are conflicting results regarding the specific interactions between thrombin and TBA, such as binding to either the TGT loop\textsuperscript{57} (Figure 1.4a) or the TT loops\textsuperscript{58} (Figure 1.4b). Recent crystallographic studies\textsuperscript{59} showed binding through the TT loops, but with a different orientation (Figure 1.4c). Clearly, further studies are necessary to resolve these uncertainties, and gain insight to the TBA-thrombin binding specificity.

![Figure 1.4: TBA binding to thrombin with either the GTG loop (a) or the TT loops with different orientations (b,c). Thymine bases that make contact with thrombin are shown explicitly and highlighted in magenta. Thrombin orientation is kept roughly the same for all three structures. The PDB structures used are 1HAP, 1HAO, and 4DIH for (a), (b), and (c), respectively.](image)

As a first step in this direction, we instead studied the association-dissociation kinetics of TBA-antidote, which is another molecular recognition problem. Experimental work probing the dynamics of the association of double-stranded DNA has been scarce, and often uses a simple two-state model to describe the reaction. We will show that a more sophisticated model which includes an intermediate state is necessary.
1.2 Dynamics experiments

It should be clear that we need to perform dynamics experiments instead of kinetics experiments. Kinetics experiments measure the rate of population change of reactants and products, essentially the stable states. On the contrary, dynamics experiments probe the time-dependent evolution of the system along a reaction coordinate, revealing the fine details of the conformational changes from the reactants to the transition states and the products.

In order to achieve this goal, we need to use a technique with both high time resolution and structural sensitivity. The high time resolution allows us to take snapshots during the reaction, otherwise we will only be measuring the average structure. The structural sensitivity is required to distinguish between different states. Unfortunately, as temporal and spatial resolution are conjugate variables, certain compromises have to be made, and different techniques may be better suited depending on the questions being asked.

In our case, 2D IR provides ultrafast sub-picosecond time resolution that is faster than the timescale of rapid tautomerization and secondary structural changes of nucleic acids (Figure 1.1). The choice of infrared spectroscopy is also for its sensitivity to structures of nucleic acids, such as base-pairing, sugar conformation, and glycosidic torsion angles. The vibrational IR and Raman spectra of nucleic acids and the general assignment of resonances have been known for decades. In the mid-IR range from 800–1800 cm⁻¹, there are four main spectral regions of interest that reflect nucleic acid conformations. (a) Bands originated from the in-plane base vibrations, such as C=O stretch, C=N stretch, and ND₂ bending modes are located in the 1500–1800 cm⁻¹ region. (b) The base-ribose vibrations that are sensitive to the glycosidic bond rotation have vibrational frequencies within the range of 1250–1500 cm⁻¹. (c) The phosphate-ribose vibrations which report on the backbone conformation lie within the frequency range of 1000–1250 cm⁻¹. (d) The ribose vibrations which provide reliable markers for the sugar puckering configurations can be observed in the 800–1000 cm⁻¹ region. As a result, infrared spectroscopy provides an excellent probe for studying the structural dynamics of various nucleic acid related systems. In this thesis, we will make use of the in-plane base vibration region (Figure 1.5a) as these vibrations are sensitive to H-bonding and base-stacking. Figure 1.5b illustrates examples of the vibrational motions involved in guanine carbonyl stretch and ring breathing modes.

However, having high temporal and structural resolution is still not enough to monitor fast chemical reactions. Chemical reactions never come to a halt, and they are always under equilibrium. Both the forward and reverse reactions occur at the same time and that the number of products formed is the same as the number of products broken down. The flux of
molecules undergoing barrier crossing events such as tautomerization and nucleic acid folding is very small under equilibrium, and therefore attempting to observe this process even with "high time resolution" is not plausible. In order to overcome this obstacle, it is necessary to synchronize these events. This can be achieved with rapid initiation techniques that drives the reaction to a non-equilibrium state transiently, therefore producing "excess reaction" in one direction which can be monitored experimentally.

Rapid initiation can be achieved with stopped-flow mixers whose time resolution can be as fast as 10 μs, photochemical triggering which has been used to study protein ligand binding or conformational changes on the ps timescale, and pressure-jump whose time resolution has been recently pushed to 0.7 μs. In this thesis, we used laser temperature-jump method that excites the water solvent absorption to increase the temperature by ~10 °C within 10 ns. Laser temperature-jump methods, pioneered by Manfred Eigen, have now been applied to study various chemical and biophysical processes.

1.3 2D IR spectroscopy of nucleic acids

Two-dimensional infrared spectroscopy is an emerging ultrafast vibrational spectroscopy that can be used to study molecular structure and dynamics of a wide variety of problems including hydrogen-bond fluctuations, chemical exchange kinetics, and vibrational energy transfer. 2D IR spectroscopy is the optical analog of 2D NMR spectroscopy. By spreading the vibrational spectrum over two frequency axes, 2D IR can measure the couplings between different molecular vibrations, as well as detect energy flow from one vibration to others. An example of 2D IR spectrum is shown in Figure 1.5b for the nucleobases guanosine monophosphate (GMP). 2D IR cross-peaks in the off-diagonal region correlate the excitation (ω1) and detection (ω2) frequencies and encode the vibrational couplings. The diagonal features in a 2D IR spectrum correspond to the IR absorption peaks in the FTIR spectrum. Each 2D IR peak consists of a doublet with a positive (red) peak due to the vibrational 0–1 transition, and a shifted negative (blue) peak originating from the 1–2 excited-state absorption.
Figure 1.5: (a) FTIR spectra of the five nucleobases. (b) 2D IR spectrum of GMP and two DFT calculated vibrational modes of guanosine G1 and G2, which correspond to the C=O stretch and ring mode, respectively. (c) 2D IR spectrum of equimolar mixture of dA and dT and one DFT calculated vibration of A ring mode in a AT base-pair.

In Figure 1.5b, it is clear that cross-peaks exist between the guanine C=O stretch (1662 cm⁻¹) and ring vibrations (1500–1600 cm⁻¹), revealing the delocalized nature of these in-plane base vibrations. By modeling the vibrational couplings measured by 2D IR spectroscopy, information on the chemical bond connectivity, orientation and distance can be obtained. The diagonal and anti-diagonal linewidths report on inhomogeneous and homogeneous broadening, respectively, which can be used to study the structural dynamics.

As mentioned previously, nucleobase in-plane vibrations in the 1500–1800 cm⁻¹ region include the carbonyl stretch and various ring breathing modes. Traditional spectral assignments of these vibrations are based on simple local modes such as C=O, C=C, and
C=N double bond stretches,\textsuperscript{90,64,76,78} whereas many computational studies suggest highly delocalized DNA vibrations.\textsuperscript{78-82} Clearly, the 2D IR spectra of nucleobases reveal significant amount of coupling between all the resonances for a given base. Chapter 4 will describe in details the characterization of the vibrational eigenstates of nucleobases.

Since the in-plane base vibrations are sensitive to H-bonding, when a base-pair forms, couplings between the vibrations from individual bases will give rise to pronounced cross-peaks. In addition, the vibrational coupling will change the eigenstates of these coupled vibrations, resulting in variations in frequency, intensity and lineshapes. Figure 1.5c shows the 2D IR spectrum of equimolar mixture of dA\textsubscript{11} and dT\textsubscript{11}. Cross-peaks highlighted with magenta boxes originate from the coupling between the adenine ring vibration at 1621 cm\textsuperscript{-1} and the ring and C=O stretches of thymine at 1641, 1663, and 1695 cm\textsuperscript{-1}. The appearance of cross-peaks in 2D IR spectra of base-pairs offer an additional level of structural discrimination compared to the often congested FTIR spectra. Chapter 8 will provide more details on the vibrational signatures of base-pairing.

The observation that all in-plane vibrational modes give rise to cross-peaks for a given nucleobases also inspired the development of 2D IR spectroscopy as a new tool to study tautomerism. Figure 1.6 shows a cartoon illustration of this approach. For nucleobase A which has two vibrations in the spectral region of interests (Figure 1.6 left), two peaks are observed in the FTIR spectrum and two diagonal along with cross-peaks between them are measured in the 2D IR spectrum. In a second scenario (Figure 1.6 middle) where the system contains a mixture of nucleobases A and B, and each has only one vibration in this spectral region. The FTIR spectrum will again display two peaks, which will be indistinguishable from the first case. However, the 2D IR spectrum will only show two diagonal peaks, revealing the presence of two chemically distinct species. As a consequence, 2D IR cross-peak patterns enable the separation of two tautomers A and B coexisting in rapid equilibrium (Figure 1.6 right).
Figure 1.6: Cartoon 2D IR spectra illustrating how different tautomers can be distinguished by 2D IR cross-peaks.

1.3.1 Comparison to 2D IR spectroscopy of proteins

In the last few years, 2D IR spectroscopy has transitioned from characterizing the vibrational eigenstates of small molecules to dynamics experiments of larger-scale and more complicated systems such as proteins. We would like to briefly discuss the development of protein 2D IR spectroscopy and contrast it with DNA spectroscopy.

Protein 2D IR spectroscopy has utilized mostly the amide I vibration, which is primarily the amide C=O stretch (Figure 1.7b). The amide I band is sensitive to the protein secondary structure, and the well-established structure-frequency relation is shown in Figure 1.7a. When proteins form secondary structures, the amide I vibration is delocalized across the protein backbone and give rise to characteristic IR signatures. Anti-parallel \( \beta \)-sheets have two IR active modes: a lower frequency \( \nu_\perp \) mode at \( \sim 1630-1640 \) \( \text{cm}^{-1} \) and a higher frequency \( \nu_\parallel \) mode at \( \sim 1680 \) \( \text{cm}^{-1} \). \( \alpha \)-helices also has two modes, \( \nu_\parallel \) and a much weaker \( \nu_\perp \), at \( \sim 1650 \) \( \text{cm}^{-1} \). As these vibrations have broad lineshapes, it is challenging to discriminate proteins with different secondary structure content using the FTIR spectra. Figure 1.7c shows the FTIR of two different proteins: Con A which consists of primarily \( \beta \)-sheet and myoglobin which is mostly \( \alpha \)-helical. The two FTIR spectra are remarkably similar and featureless. However, their 2D IR spectra provide discernible features. The 2D IR spectrum of Con A shows that couplings between the two \( \beta \)-sheet modes \( \nu_\perp \) and \( \nu_\parallel \) result in cross-peak ridges.
extending along $\omega_3 = 1630$ and 1680 cm$^{-1}$, therefore giving an overall “Z” shaped spectrum.$^{83}$ In contrast, the 2D IR spectrum of myoglobin, having only the $\nu_A$ mode, exhibits a “figure 8” lineshape.

Figure 1.7: Protein amide I spectroscopy. (a) Empirical secondary structure-frequency relation. (b) Visualization of the IR active $\alpha$-helix and $\beta$-sheet modes. Colors indicate the phase of the vibrations. Blue and red are 180° out-of-phase. (c) Structure, FTIR and 2D IR spectra of Con A and myoglobin. Figure adapted from ref. [83] with permission.

2D IR spectra of proteins provide better structural resolution compared to the linear FTIR spectra, however, interpreting the broad lineshape can still be challenging. Over the last few years, significant development of the spectroscopic maps that relates protein structure to spectral features have greatly improved our ability to gain extra information from the 2D IR spectra. These maps use classical force field simulations to calculate vibrational frequencies and couplings and ultimately the 2D IR spectra. Qualitative agreement between the experimental and theoretical spectra have been achieved in some cases (depending on the
proteins or peptides. MD simulations provide exquisitely detailed structural information at atomic resolutions. Combining these two computational approach, one can simulate 2D IR spectra for various structures of the same protein and use them to interpret experimental spectra, opening up the possibility to study conformational changes during protein folding. Obviously, substantial improvements can still be made on both the experimental and theoretical fronts of protein 2D IR spectroscopy, but the general approach is appealing and a similar method development in nucleic acid 2D IR spectroscopy will be tremendously valuable.

1.3.2 Development of 2D IR spectroscopy of nucleic acids

2D IR spectroscopy on nucleic acids is relatively undeveloped, and much remains to be understood in order to use it as a practical tool for studying the structure and dynamics of nucleic acids. As mentioned in the previous section, developing a spectroscopic map that calculates the spectrum for a given structure will facilitate the interpretation of experimental spectra immensely; however, to date, there is no accurate theoretical model that can simulate the experimental spectra in terms frequency, intensity and lineshapes. The major challenge lies in the fact that these nucleobase ring vibrations are heavily delocalized across the entire ring and simple dipole treatment like the one used in protein spectroscopy is not immediately applicable. As a result, even with the extensive body of work on the vibrational spectroscopy of nucleic acids that exist, a description of the vibrational motions assigned to different resonances, and an understanding of how these are influenced by interactions in secondary structures remain elusive.

Moreover, distinguishing the secondary structures of nucleic acids using the in-plane base vibrations has not been realized. Liquier and Taillandier discussed the experimental spectral features that can be used as marker bands to distinguish between the A, B and Z double-helical conformations, but mostly using the phosphate and deoxyribose vibrations. The advantage of using the in-plane base vibration is the much higher oscillator strength. Figure 1.5c shows that pronounced cross-peaks are observed when base-pairs form. Since different secondary structures adopt different H-bond and base-stacking configurations, the in-plane base vibrations should be sensitive to these changes. Theoretical calculations on the vibrational dynamics of DNA by Cho and coworkers investigated the vibrational basis modes and couplings of Watson-Crick base-paired DNA, as well as the effects of different DNA conformations on the exciton delocalization. The authors have simulated 2D IR spectra for different forms of AT and GC oligonucleotides and found that, for example, the spectra vary significantly for the A, B and Z forms of AT base-pairs, demonstrating the potential
structural sensitivity of 2D IR spectroscopy. Nevertheless, these predications have not been tested experimentally.

On the experimental front, Krummel and Zanni reported the first 2D IR studies of Watson-Crick guanine-cytosine (G-C) base pairs and found that both electrostatics and base stacking effects are important in modeling the inter-base couplings between carbonyl stretches. More recently, Elsaesser and coworkers have performed 2D IR experiments to characterize the vibrational couplings between NH and NH₂ stretching modes in DNA single base-pair or oligomers in non-aqueous solvents, or at different hydration levels. These experiments illustrate the rich spectral information contained in the 2D IR spectra of nucleic acids and the potential for various applications in biophysics, such as DNA-water and DNA-protein interactions. For example, distinction between the proposed mechanisms for thrombin-TBA recognition can be made possible using protein-DNA cross-peaks, taking the advantage of the distinct vibrational signatures of different nucleotides, protein backbone and side-chain resonances. When necessary, isotope labeling such as ¹³C,¹⁸O for the protein backbone, or ¹³C,¹⁵N,¹⁸O for the nucleotides can help reduce the spectral congestion and enable site-specific information. Combining 2D IR spectroscopy with T-jump experiments offers a unique capability to probe the coupled folding/binding structural dynamics. Nevertheless, much work remains to be done before these dreams can be realized. A systematic characterization of the 2D IR spectra of nucleobases, base-pairs, secondary structures, and improvement of MD simulations, force fields, and theoretical spectral modeling are essential next steps.
1.4 Thesis outline

The necessary theoretical background and experimental methods are given in Chapter 2 and Chapter 3, respectively. Chapter 4 describes the characterization of the 2D IR spectra of the in-plane base vibrations of nucleobases. We then applied the knowledge gained in Chapter 4 to study the tautomerism of aromatic heterocycles in Chapter 5. We first demonstrated the methodology of separating tautomers of a model system, pyridone derivatives. We then studied the tautomerism of oxythiamine. In Chapter 6, we further extended the capability of 2D IR in investigating tautomerism by performing T-jump transient 2D IR experiments to monitor the tautomer exchange on the nanosecond timescale. We found that the lactam-lactim tautomerization occurs through a two-state concerted mechanism facilitated by bridging water molecules. Chapter 7 describes another fascinating biological application—tautomerization of an anti-HIV drug, KP1212. After discovering that KP1212 can exist in multiple tautomeric states (including the rare enol forms) under physiological conditions, we focus our efforts on directly observing mispair between adenine and KP1212 tautomers in DNA oligomers, which is detailed in Chapter 8. Chapter 9 discusses the IR signatures of G-quadruplex formation, in addition to characterization using traditional methods. With the spectral assignments made in Chapter 9, in Chapter 10 we investigated the early unfolding events of TBA and the association-dissociation of the TBA-antidote duplex using T-jump transient 2D IR spectroscopy.

The experiments presented in this thesis show great promise for studying conformational dynamics of nucleic acids using 2D IR spectroscopy, and hopefully will stimulate continuous efforts in the development of advanced theoretical modeling that can complement and help interpret the complex multidimensional spectra. We believe that a dramatic advancement in DNA spectroscopy will be made from these experiments, in much the same way that 2D IR revolutionized peptide and protein spectroscopy.
1.5 References


CHAPTER 2
Theoretical Background of 2D IR Spectroscopy

2.1 Introduction

Two-dimensional infrared spectroscopy is an emerging molecular vibrational spectroscopy for biophysical sciences. In 2D IR experiments, sequences of ultrafast IR laser pulses are employed to excite and probe molecular vibrations. The resulting 2D spectrum correlates the vibrational excitation frequency with the detection frequency. 2D IR spectra provide more information content compared to the traditional linear FTIR spectra, which are usually congested by inhomogeneously broadened peaks. The peak positions, amplitudes, signs, and lineshapes of the cross-peaks in the 2D IR spectrum unveil properties of the molecular structure such as chemical bond connectivity, distance, and orientation, as well as characterize the vibrational eigenstates of the system. Because of the more comprehensive spectral features contained in 2D IR spectra, 2D IR spectroscopy has gained increasing popularity over the previously used third-order nonlinear spectroscopy such as pump-probe, photon echo, and transient grating experiments. In this chapter, we describe the theoretical background of 2D IR spectroscopy following the treatment of Sung and Silbey.\textsuperscript{1,2}
2.2 Nonlinear response function formulism

2.2.1 The third-order polarization

The following analysis will be based on the pulse sequence and phase matching conditions shown in Figure 2.1. Assuming no other external fields are present, the total external electric field can be written as the sum of the five fields,

\[ E_{ex}(\vec{r},t) = \sum_{j=1}^{3} e_j(t) \cdot e^{i\vec{k}_j \cdot \vec{x} - \omega_j t} + c.c \]  

(2.1)

where \( e_j(t) \) is the temporal envelope. In this chapter, four-wave mixing (4WM) type of 2D IR experiments will be presented, thus three pulses created from a single laser source using beam-splitters and controlled time-delay stages have the same frequency, \( \omega_0 \). The effective Hamiltonian of the system is given by

\[ H_{eef}(t) = H_M + H_{int} = (H_S + H_B) + H_{int} \]  

(2.2)

where \( H_S \), \( H_B \), and \( H_{int} \) are the Hamiltonians of the system, the bath, and the laser interaction with the molecular vibrational motions. \( H_S \) contains a number of coupled vibrational coordinates, and can be expressed in a basis in the local modes, \( \tilde{3} \) normal modes, or eigenstates, depending on the information of interest. The Hamiltonian for the bath

\[ E_{3} E_{2} E_{1} \]

\[ \tau_{1} \tau_{2} \tau_{3} \]

\[ k_{sig} = k_{1} + k_{2} + k_{3} \]

Figure 2.1: (a) Non-collinear "boxcar" phase-matching geometry for the three input fields in 2D IR spectroscopy. After the sample, a mask is used to allow only the heterodyne signal of the third-order nonlinear signal and the local oscillator (LO) to go through. (b) The pulse sequence and time variables for a 2D IR experiment. After the first light-matter interaction, the system is prepared in a coherence state during \( \tau_1 \), the evolution time. The second interaction puts the system in a population state during \( \tau_2 \), the waiting time. The third interaction excites the system again into a coherence state during \( \tau_3 \), the detection time. Finally the emitted nonlinear signal is heterodyne with the LO. (c) Showing a clear view of the phase-matching condition and the wave vector of the nonlinear signal, \( k_{sig} = k_1 + k_2 + k_3 \).
describes the dissipation on vibrational modes to the harmonic heat bath. Together with the system Hamiltonian, they form the material Hamiltonian, $H_{\text{mat}}$. Finally, the interaction Hamiltonian describes the coupling of the system to classical external radiation fields. In the case of 2D IR experiments, it is given as

$$H_{\text{int}} = -M(Q) \cdot E_{\text{ex}} = -\sum_{a,b} |a\rangle \mu^{a,b} \cdot E |b\rangle$$

(2.3)

where $M(Q)$ represents the dipole operator of the coupled vibrational coordinates $Q$, and $\mu^{a,b} = \langle a | M(Q) | b \rangle$ are the transition dipole matrix elements, connecting two eigenstates $|a\rangle$ and $|b\rangle$. In linear optics, the macroscopic polarization $P$ depends linearly on the electric field $E$; however, in nonlinear optics where the electric fields are high enough, nonlinear dependence has to be taken into consideration. The nonlinear polarization induced by the strong field is given by the expectation value of the dipole operator:

$$P(t) = \langle \mu(t) \cdot \rho_g(t) \rangle$$

(2.4)

where $\rho_g(t)$ is the ground state density matrix, which follows the Liouville equation:

$$\frac{\partial}{\partial t} \rho_g(t) = -\frac{i}{\hbar} \left[ H_{\text{mat}} + H_{\text{int}}, \rho_g(t) \right]$$

(2.5)

Integrating the Liouville von Neumann equation using a perturbative expansion of the Hamiltonian and inserting it back to Equation (2.4) gives the third-order polarization:

$$P^{(3)}(t) = \iiint \mathbf{R}^{(3)}(\tau_3, \tau_2, \tau_1) E_3(\mathbf{k}_3, t - \tau_3) \times E_2(\mathbf{k}_2, t - \tau_3 - \tau_2) \times E_1(\mathbf{k}_1, t - \tau_3 - \tau_2 - \tau_1) \, d\tau_1 d\tau_2 d\tau_3$$

(2.6)

$\mathbf{R}^{(3)}$ is the third-order nonlinear response function that is used to describe the majority of coherent nonlinear experiments such as pump-probe experiments, transient gratings, photon echoes, coherent anti-Stokes Raman spectroscopy (CARS), and degenerate four-wave mixing. This response function is a thrice-nested commutator of the transition dipole operators, $\mu$, evaluated at each light-matter interaction:

$$\mathbf{R}^{(3)}(\tau_3, \tau_2, \tau_1) = \left( \frac{-i}{\hbar} \right)^3 \theta(\tau_3) \theta(\tau_2) \theta(\tau_1) \times \langle [[[\mu(\tau_3 + \tau_2 + \tau_1), \mu(\tau_2 + \tau_1)], \mu(\tau_1)], \mu(0)] \rho_g \rangle$$

(2.7)
The time ordering of the three pulses are enforced by causality, which is represented by the Heaviside step function, \( \theta(\tau) \). We will assume the femtosecond pulses to be impulsive for simplicity, although infinite time resolution cannot be achieved in reality. In this semi-impulsive limit, the 100 fs width of the laser pulses is assumed to be short compared to the timescale of the system, but long compared to the oscillation period of the light field. Therefore, we can express the envelope of the pulses as the following:

\[
E_1(t) = |E_1| \delta(t) e^{i\omega_1 t + \Delta k_1 \cdot \mathbf{r}}
\]

\[
E_2(t) = |E_2| \delta(t - \tau_1) e^{i\omega_2 t + \Delta k_2 \cdot \mathbf{r}}
\]

\[
E_3(t) = |E_3| \delta(t - \tau_1 - \tau_2) e^{i\omega_3 t + \Delta k_3 \cdot \mathbf{r}}
\]  

We also have to make the rotating wave approximation, which drops the non-resonant term because the resonant term \( (\omega \approx \Delta \epsilon / \hbar) \) is much stronger. This is valid when the envelope of the electric field is slowly varying in time compared to the carrier frequency. The nonlinear polarization is then directly proportional to the nonlinear response function.

\[
P^{(3)}(t) \propto \mathcal{R}^{(3)}(\tau_3, \tau_2, \tau_1, t)|E_1| |E_2| |E_3|
\]  

\[2.2.2\] The third-order nonlinear response

The thrice-nested commutator in Equation (2.7) can be expanded into eight terms, however, only four are unique since the other four are the complex conjugates:

\[
\mathcal{R}^{(3)}(\tau_3, \tau_2, \tau_1) = \sum_{n=1}^{d} \mathcal{R}_n(\tau_3, \tau_2, \tau_1) - \mathcal{R}^{*}_n(\tau_3, \tau_2, \tau_1)
\]  

These four response function pairs differ by whether sequential operators act on the bra or ket side of \( \rho \) when enforcing the time-ordering. These can be pictorially-represented using double-sided Feynman diagrams shown in Figure 2.2, where \( a, b, c, d \) are the eigenstates sampled through in a Liouville pathway. The time axis runs from the bottom to the top. The right and left side of the Feynman diagram represent the bra and ket of the density matrix, respectively.

Now we will make further assumption to simplify the response functions. Assuming the vibrational and rotational degrees of freedom are separable, \[2.12,2.13\] then each of the tensorial responses in Equation (2.10) can be written as a product of an isotropic nonlinear response
Figure 2.2: Feynman diagrams for the four pairs of third-order nonlinear response functions. The indices $a$, $b$, $c$, $d$ represent the eigenstates of the system.

The exact functional form of $R_n^{a,b,c,d}$ depends on how the density matrix evolves in time under the Hamiltonian. For example, we will write out explicitly for $R_1$:

$$
R_1(\tau_3, \tau_2, \tau_1) = P_a \times \mu_{a,b} \mu_{d,e} \mu_{c,a} \mu_{b,a} \times 
\exp\left(-i\omega_{b,d}\tau_3 - i\omega_{b,e}\tau_2 - i\omega_{b,a}\tau_1\right) \times F_1^{a,b,c,d}(\tau_3, \tau_2, \tau_1)
$$
Here $P_a$ is the population in the ground state, where the system is originally prepared in. During the first time interval, $\tau_1$, the field acts on the *ket* side with $\mu^{a,b}$ to raise it from the ground state to the first excited state, $b$, thus creating a coherence $\rho_{ba}$ oscillating at frequency $\omega_{b,a}$. Based on this principle, we can complete the expression for the three interactions.

$F_{1}^{a,b,c,d}(\tau_3, \tau_2, \tau_1)$ is the dephasing function that is related to the exponential of a sum of different lineshape functions $g_{i,j}(t)$, which can be obtained from the integrals over a set of energy-gap, or frequency-frequency correlation functions $C_{i,j}$.

$$g_{i,j}(t) = \int d\tau_2 \int d\tau_1 C_{i,j}(\tau_2 - \tau_1)$$

$$C_{i,j}(t-t_0) = \langle \delta\omega_i(t)\delta\omega_j(t_0) \rangle$$

More detailed discussion on the lineshape function will come later when we present a model calculation.

### 2.2.3 Phase-matching and time-ordering

We now consider phase-matching condition for the resonant 2D IR experiments used in this thesis. For the three incoming pulses $E_a, E_b, E_r$, each with an incident wavevector $k_a, k_b, k_r$, respectively, the boxcar geometry allows the detection of the signal with the $k_{sig} = -k_a + k_b + k_r$ phase-matched direction (Figure 2.1c). Considering the time-ordering of $k_a$ and $k_b$, we can further separate the signal into two categories: $k_s = -k_1 + k_2 + k_3$ and $k_{II} = +k_1 - k_2 + k_3$. In the first case, since the phase acquired during $\tau_1$ and $\tau_3$ are opposite, we call it the rephasing signal. Vice versa, the second type of signal is called the non-rephasing signal. Going back to Figure 2.2, we can figure out which diagrams contribute to the rephasing and non-rephasing signals. In double-sided Feynman diagrams, positive and negative wavevectors are associated with an arrow pointing to the right and left, respectively. Arrows pointing towards and away from the diagram indicate absorption and emission, respectively. With the phase-matching conditions and the fact that there is no emission from the ground state, we find that diagrams $R_2, R_3$ and $R_1^*$ contribute to the rephasing signal; while $R_1, R_4$ and $R_2^*$ contribute to the non-rephasing signal, as shown.
Rephasing
\[ S_2 = -k_1 + k_2 + k_3 \]

Non-rephasing
\[ S_{11} = +k_1 - k_2 + k_3 \]

Figure 2.3: Feynman diagrams contributing to the rephasing and non-rephasing signals with the phase-matched condition \( k_{ag} = -k_a + k_{g} + k_s \) for an IR resonant vibrational system. The numberings 0, 1, and 2 represent the ground, singly-excited, and doubly-excited states.

in Figure 2.3. Here, 0, 1, and 2 represent the ground, singly-excited, and doubly-excited states of the system. In other words, for a system with two eigenstates \( a \) and \( s \), 1 can be either \( a \) or \( s \). We will show an example of two coupled oscillators in Section 2.3.

The rephasing and non-rephasing spectra are measured separately, and the sum of the time domain signals is then double-Fourier transformed to the frequency domain, resulting in purely absorptive lineshape, which we termed 2D IR correlation spectrum.

\[
S_r(\omega_2, \tau_2, \omega_1) \propto \Re \left[ \int_0^{\tau_2} \int_0^{\tau_3} R(\tau_1, \tau_2, \tau_3) e^{i\omega_1 \tau_1} e^{i\omega_2 \tau_2} d\tau_1 d\tau_3 \right] \tag{2.15}
\]

where \( \omega_1 \) and \( \omega_2 \) are Fourier transform pairs of the experimental time delays \( \tau_1 \) and \( \tau_3 \).

2.3 Six-level model system

In order to illustrate the information content provided by a 2D IR spectrum, we simulate a spectrum for a model system consisting of two anharmonically coupled carbonyl stretches (for example dicarboxylacetylacetonato rhodium or uridine shown in Chapter 4). We denote the two local C=O stretches as \( c_1 \) and \( c_2 \). If we consider up to total quantum number of 2, there exist six energy levels:

\[ |0, 0\rangle, |c_1, 0\rangle, |0, c_2\rangle, |2c_1, 0\rangle, |0, 2c_2\rangle, |c_1, c_2\rangle \]

where the Dirac notation \( |m, n\rangle \) is used to denote the state in which the system has \( m \) and \( n \) vibrational quanta in the \( c_1 \) and \( c_2 \) stretches, respectively. The two oscillators can be coupled
due to either electrostatic dipole-dipole interaction or mechanical through-chemical bond coupling. We introduce a bi-linear coupling $\beta$ with the Hamiltonian written as:

$$H = \hbar \Omega_i \left( a_i^e a_i + 1/2 \right) + \hbar \Omega_{i_2} \left( a_{i_2} a_{i_2} + 1/2 \right) + \beta \left( a_i^e a_i + a_{i_2} a_{i_2} \right)$$ \hfill (2.16)

where $\Omega_i$ is the excitation energy of the individual site, $a_i^e$ and $a_i$ are the raising and lowering operators of the local oscillators, respectively. Next, we introduce vibrational anharmonicity, $\delta$, which is the result of expanding the potential energy of the oscillators to the cubic term. We will later show that the presence of anharmonicity is important for observing 2D IR signals. The Hamiltonian in the six-state basis can now be expressed as:

$$\hat{H} = \begin{bmatrix}
0 & \Omega_{i_1} & \beta & 0 & \sqrt{2} \beta \\
\Omega_{i_1} & \beta & \Omega_{i_2} & 2 \Omega_{i_1} - \delta_i & 0 & \sqrt{2} \beta \\
\beta & \Omega_{i_2} & 2 \Omega_{i_2} - \delta_{i_2} & \sqrt{2} \beta & 0 & \sqrt{2} \beta \\
0 & 2 \Omega_{i_1} - \delta_i & \sqrt{2} \beta & \Omega_{i_1} + \Omega_{i_2}
\end{bmatrix}$$ \hfill (2.17)

Here the Hamiltonian is separated into blocks of the ground-state, the one-excitonic, and the two excitonic Hamiltonians. Diagonalizing the matrix will result in the symmetric and asymmetric C=O stretches and the six associated eigenstates: $|0,0\rangle, |a,0\rangle, |0,s\rangle, |2a,0\rangle, |0,2s\rangle, |a,s\rangle$. The energy levels of the eigenstates are plotted in Figure 2.4a, and the resulting FTIR and 2D IR spectra are shown in Figure 2.4b. The transformation from the local basis set into the eigenstate basis is defined by the matrix $T$,

$$\hat{H}' = T^{-1} \hat{H} T$$ \hfill (2.18)

where $\hat{H}'$ is the Hamiltonian in the new eigenstate basis set. Similarly, the transition dipole operators, which determine the peak intensities in the 2D IR spectra, can be transformed from the local basis to the eigenstate basis following:

$$M' = T^{-1} M T$$ \hfill (2.19)

where $M'$ is the transition dipole operator matrix in the new eigenstate basis set.
2.3.1 Peak position and sign

From Figure 2.4b, one can immediately see how much more information is provided by 2D IR spectroscopy compared to the conventional linear FTIR. The different peaks are the result of the different pathways, which can be obtained by extending the double-sided Feynman diagram shown in Figure 2.3 to include all six eigenstates. These Feynman diagrams are shown in Figure 2.5 and each diagram is labeled with the 2D IR peak that it contributes to. Peaks 1 and 2 have the same frequency as the ones observed in the FTIR both in $\omega_1$ and $\omega_3$, therefore, these peaks are called the diagonal peaks. Whereas peaks in the off-diagonal region are termed cross-peaks (3, 3', 4, and 4'). Here we will discuss how 2D IR spectra can be used to characterize the energies and transition dipoles of the system eigenstates.

The position of the peaks is determined by the consecutive interactions of the multilevel system with a sequence of three electric fields. The $\omega_1$ axis indicates the frequency of the vibrational coherence during $\tau_1$ after the first light-matter interaction. Therefore, all the peaks in this frequency dimension lie along the fundamental frequencies, $\omega_a$ and $\omega_s$. The $\omega_3$ axis represents the frequency of the coherence during $\tau_3$. Here the peaks do not necessarily lie along the fundamental frequencies because the eigenenergies of the two-quantum states are lowered by the anharmonicities. The diagonal peaks 1 and 2 originate from four

![Figure 2.4](image-url)

Figure 2.4: (a) The energy levels for the eigenstates of the two anharmonically coupled oscillators. (b) Simulated FTIR and 2D IR spectra for this six-level system.
transitions only between the ground and the first excited state, \(a\) and \(s\), respectively. The cross-peaks 3 and 4 are the result of change of coherence from one fundamental transition to another (in other words, pumping \(a\) and probing \(s\) or vice versa). The remaining peaks with negative intensity are due to transitions involving the two-quantum states. For instance, the pathway leading to peaks 1' involves coherence of \(|2a\rangle\langle a|\) during \(\tau_3\). Therefore, the peak is downshifted in \(\omega_3\) compared to peak 1 as \(\omega_{2a,a} - \omega_{a,0} = -\Delta_a\). Furthermore, when exciting the system into the two-quantum states, the combination band \(|as\rangle\) can be excited as well, which gives rise to peaks 3'-6' that are further red-shifted compared to peaks 1' and 2' (because \(\Delta_{as} > \Delta_a, \Delta_s\)). Peaks 5' and 6' are weaker transitions since there is only one non-phasing pathway for each peak. Moreover, when \(\Delta_{as} < \Delta_a, \Delta_s\), peaks 5' and 6' overlap with peaks 1' and 2', and therefore are not commonly observed experimentally.

The frequencies of these peaks characterize the anharmonic nuclear potential of the coupled symmetric and asymmetric carbonyl stretches. The peak splitting of a doublet peak (one positive and one negative) indicate the anharmonic shifts of the overtone \((\Delta_a, \Delta_s)\) and the combination band \((\Delta_{as})\). The sign of the peaks shows whether the system evolves in a superposition of the ground and fundamental state (positive, indicated by red) or fundamental and doubly-excited state (negative, indicated by blue) during \(\tau_3\). Therefore, it is now clear that for a harmonic system \((\Delta_a = \Delta_s = \Delta_{as} = 0)\) where there is no peak splitting, the positive and negative peaks of a doublet interfere destructively (for example, peak 3 and 3'), thus the 2D IR signal is very small or nonexistent. For two uncoupled anharmonic oscillators, the diagonal peaks can be observed while the cross peaks will disappear. To sum up, the selection rules for 2D IR spectrum include (1) anharmonicity in the ground-state potential, (2) nonlinear dependence of the transition dipole operator on vibrational coordinates, and (3) level-dependent dephasing dynamics (the dephasing of \(0 \rightarrow 1\) is different from \(1 \rightarrow 2\)).^15
Figure 2.5: Feynman diagrams for the six-level system that contribute to the rephasing (top) and non-rephasing (bottom) 2D IR signals. The associated 2D IR peaks are labeled for each diagram.
2.3.2 Lineshape

Spectroscopy is useful to describe phenomenon in the condensed matter because it provides information on the structure and dynamics. For a single vibration in a liquid or solid, at a given instant in time each member of an ensemble will have a different transition frequency due to different local environment. The lineshape of a peak in a 2D IR spectrum reflects this effect of the bath on the four interacting transition dipoles.

In general, the time-dependent correlated fluctuations between two transition frequencies (Equation (2.14)) can be described by a Gaussian stochastic model with exponentially decaying auto- \((i = j)\) and cross- \((i \neq j)\) correlation functions:

\[
C_y(t) = \rho_y \Delta_y \Delta_y \exp(-|t|/\tau_y)
\]  

(2.20)

where \(ij = a, s. \rho_y\) is the correlation coefficient with \(\rho_y = 1\). If the timescale of these frequency fluctuation is sufficiently small, or the correlation time is long, then the lineshape of the ensemble reflects the distribution of frequencies. In this limit the lineshape is asymmetric and called “inhomogeneously broadened”. The energy gap correlation function becomes \(C_y(t) = \rho_y \Delta_y \Delta_y\), and the lineshape function is \(g_y(t) = \rho_y \Delta_y \Delta_y t^2 / 2\). In this static limit, \(\Delta_y\) and \(\Delta_\nu\) are proportional to the magnitudes of the distribution of the transition frequencies of the system eigenstate around their central frequency. On the other hand, if the frequency fluctuation is sufficiently fast, or the correlation time is short, then the lineshape reflects both the frequency distribution and the time-dependent fluctuation. This limit is called the “homogeneous” limit where a symmetric peak is observed. The energy gap correlation function is a delta function in the homogeneous limit: \(C_y(t) = \delta(t)/T^y_2\), and the lineshape function is \(g_y(t) = t/T_2^y\), where \(\delta(t)\) is the Dirac-delta function the dephasing time \(T^y_2 = \rho_y \Delta_y \Delta_\nu \tau_y\). Experimentally, \(T_2\) can be obtained from the homogeneous linewidth \(\Gamma\):

\[
\Gamma = 1/\pi T_2
\]  

(2.21)

To include the intermediate regime, the frequency-frequency correlation function can be modeled using the generalized Kubo model, which will be used to fit the experimental spectra in this thesis (see Chapter 4),

\[
C(t) = \frac{\delta(t)}{T_2} + \Delta^2 e^{-\Delta^2} + \Delta_0^2
\]  

(2.22)

52
Figure 2.6: Simulated 2D IR spectrum consisting of homogeneous peak at 1600 cm\(^{-1}\) and an inhomogeneous peak at 1670 cm\(^{-1}\).

where \(\delta(t)/T_2\) describes the fast dynamics that determine the homogeneous linewidth, \(\tau\) and \(\Lambda\) describe the intermediate regime, and \(\Lambda_o\) determines the inhomogeneous linewidth. Inserting this form of correlation function into Equation (2.13), we obtain the lineshape function,

\[
g(t) = \frac{t}{T_2} + \Lambda^2 \tau^2 \left[ e^{-t/\tau} - 1 + \frac{t}{\tau} \right] + \Lambda_o^2 t^2 / 2
\]

Examples of homogeneous and inhomogeneous peaks are shown in Figure 2.6. The 1600 cm\(^{-1}\) diagonal peak is symmetric and homogeneous whereas the 1670 cm\(^{-1}\) diagonal peak is asymmetric and inhomogeneously broadened along the diagonal. This simulated spectrum is commonly seen for DNA bases and analogs as the high frequency mode represents the C=O stretch, which is broad due to the multiple configurations of hydrogen-bonding to the solvent, and the low frequency mode represents the sharp ring modes.

2.4 Polarization-selective 2D IR spectroscopy

We now return to the orientational nonlinear response function \((Y_n)_{\alpha,\beta,\epsilon,\delta}^{i,j,k,l}\) in Equation (2.11), which is an important factor that influence the 2D IR peak amplitudes. The orientational response depends on the mutual orientation of the four time-ordered transition dipole moments in the molecular fixed frame \((i,j,k,l) \in x,y,z\), and can be evaluated by transforming into the laboratory frame through an orientational averaging. The four \(Y_n\)
factors differ only in the sequence of dipole interactions and therefore we write out the expression for $Y_i$ only:

$$(Y_i)_{i,j,k,l}^{a,b,c,d} = \sum_{ijkl} \tilde{Y}_{ijkl} \left( \hat{\mu}^{a,b} \cdot \hat{i} \right) \left( \hat{\mu}^{a,c} \cdot \hat{j} \right) \left( \hat{\mu}^{c,a} \cdot \hat{k} \right) \left( \hat{\mu}^{b,a} \cdot \hat{l} \right)$$

(2.24)

The transformation from the molecular frame to the laboratory frame can be treated classically and discussed in detail in Refs. 15. For two orthogonal dipole moments, these $\tilde{Y}_{ijkl}$ tensor components are evaluated and listed in Table 2.1.

Table 2.1. Tensor components of the orientational response involving two orthogonal transition dipole moments oriented along $\hat{v}$ and $\hat{k}$. $C_i(t_1, t_2) = \exp[-2D(t_1 + t_2)]$ and $C_2(t_2) = \exp(-6Dt_2)$ where D is the orientational diffusion constant.

<table>
<thead>
<tr>
<th>Polarization</th>
<th>$Y^{vyy}$</th>
<th>$Y^{vxx}$</th>
<th>$Y^{vxx}$</th>
<th>$Y^{vxy}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZZZZ</td>
<td>$\frac{1}{9}C_1(1-\frac{4}{5}C_2)$</td>
<td>$\frac{1}{9}C_1(1-\frac{2}{5}C_2)$</td>
<td>$\frac{1}{15}C_1C_2$</td>
<td>$\frac{1}{15}C_1C_2$</td>
</tr>
<tr>
<td>ZZYX</td>
<td>$\frac{1}{9}C_1(1-\frac{2}{5}C_2)$</td>
<td>$\frac{1}{9}C_1(1+\frac{1}{5}C_2)$</td>
<td>$-\frac{1}{30}C_1C_2$</td>
<td>$-\frac{1}{30}C_1C_2$</td>
</tr>
<tr>
<td>ZYZZ</td>
<td>$\frac{1}{15}C_1C_2$</td>
<td>$-\frac{1}{30}C_1C_2$</td>
<td>$\frac{1}{4}C_1(1-\frac{1}{5}C_2)$</td>
<td>$\frac{1}{4}C_1(1+\frac{1}{5}C_2)$</td>
</tr>
<tr>
<td>ZYXZ</td>
<td>$\frac{1}{15}C_1C_2$</td>
<td>$-\frac{1}{30}C_1C_2$</td>
<td>$\frac{1}{4}C_1\left(1\frac{1}{5}C_2\right)$</td>
<td>$\frac{1}{4}C_1\left(1-\frac{1}{5}C_2\right)$</td>
</tr>
</tbody>
</table>

Because the 2D IR peak amplitudes depend on both the laser polarization and the angle between the two coupled transition dipoles, $\theta$, a judicious choice of incident field polarizations should in principle allow the determination of the relative dipole orientation. However, it is not straightforward to directly obtain $\theta$ from comparing the cross-peak intensities from polarization-selective 2D IR spectra, because some pathways involve four different transition dipole moments. For example, one of the diagrams leading to peak 3 involves the interaction of $\mu^{0,0}, \mu^{0,3} \mu^{0,0}, \mu^{0,0}, \mu^{0,0}$, and therefore includes three different angles: between $\mu^{0,0}$ and $\mu^{0,0}$, between $\mu^{0,0}$ and $\mu^{0,3}$, and between $\mu^{0,0}$ and $\mu^{0,0}$. If we assume that the angles between the transition dipole moments from the $0\rightarrow 1 \left(\mu^{1,0}\right)$ and $1\rightarrow 2 \left(\mu^{2,1}\right)$ quantum states are parallel, then the cross-peak amplitudes will be sensitive to only one angle,
between $\mu_1^{x,0}$ and $\mu_2^{x,0}$. A fit to the experimental 2D IR spectra of a rhodium dicarbonyl (RDC) compound showed that this is a valid assumption.\textsuperscript{14}

For two transition dipole moments oriented with an angle $\theta$ between them, the orientational factor $Y_{ijkl}^{\text{old}}$ can be expressed as the linear combination of the tensor components listed in Table 2.1 weighted by $\cos(\theta)$ or $\sin(\theta)$. A comprehensive list of the appropriate scaling factors are tabulated in Table 16S in Ref. 14. To show one example, the positive cross-peak 3 in Figure 2.4b has contributions from three different pathways (one of each from $R_3$, $R_3$, and $R_4$, see Figure 2.5), and we can obtain the expression for the orientational response for peak 3 by summing the three contributions:

\[
(Y_3)^{0,0,0,0}_{ijkl} + (Y_4)^{0,0,0,0}_{ijkl} + (Y_4)^{0,0,0,0}_{ijkl} = \\
\left[ Y_{ijkl}^{vvv} \cos^2(\theta) + Y_{ijkl}^{vxy} \sin^2(\theta) \right] + \left[ Y_{ijkl}^{vvv} \cos^2(\theta) + Y_{ijkl}^{vxy} \sin^2(\theta) \right] + \\
3Y_{ijkl}^{vvv} \cos^2(\theta) + \left[ 2Y_{ijkl}^{vxy} + Y_{ijkl}^{vxy} \right] \sin^2(\theta) \\
\text{(2.25)}
\]

Finally, the indices $I, J, K, L$ are replaced by the laser polarizations used in the experiment (such as $Z$ or $Y$) following Table 2.1. Since the peak amplitudes also depend on the transition dipole strengths ($\mu_1^2 \mu_2^2$ in this case), we remove this contribution by taking the ratio of the peak amplitudes measured with two different polarizations, such as perpendicular (ZZYY) and parallel (ZZZZ) polarizations used in this thesis. Inserting the tensor components listed in Table 2.1 into Equation (2.25), we obtain the relationship between the relative cross-peak amplitude and the angle between the two dipoles:

\[
R = \frac{(A3)^{ZZYY}_{ijkl}}{(A3)^{ZZZZ}_{ijkl}} = \frac{1}{5} \left( \frac{\cos^2(\theta) + \frac{7}{30} \sin^2(\theta)}{\cos^2(\theta) + \frac{1}{5} \sin^2(\theta)} \right) = \frac{7 - \cos^2(\theta)}{6 + 12 \cos^2(\theta)} \\
\text{(2.26)}
\]

where $(A3)_{ijkl}$ is the amplitude of the cross-peak 3 measured with $IJKL$ polarization. Therefore the angle is

\[
\theta = \cos^{-1} \left( \frac{7 - 6r}{\sqrt{1 + 12r}} \right) \\
\text{(2.27)}
\]
Equation (2.26) describes the relationship between the absorptive (positive signal) cross-peak centered at \((\omega_1, \omega_2) = (\omega_i, \omega_s)\) such as peak 3, but is also valid for the downhill cross-peak at \((\omega_1, \omega_2) = (\omega_i, \omega_s)\) — peak 4 in Figure 2.4. Similar analysis can be applied to the emissive (negative signal) cross-peaks at \((\omega_s, \omega_i - \Delta_0)\), the entire cross-peak (containing both absorptive and emissive signals), anisotropy, or signals measured with other polarizations. However, obtaining the angle from the emissive cross-peaks that involve transitions to the combination bands is not straightforward. For example, peak 3' has a pathway involving the interactions between \(\mu^{\alpha, \beta} \mu^{\gamma, \delta} \mu^{\sigma, \rho}\) and another pathway involving \(\mu^{\alpha, \beta} \mu^{\gamma, \delta} \mu^{\sigma, \rho}\), in order to separate the contribution of the transition dipole moments and the orientational factors, harmonic scaling of the transition dipole moments has to be assumed such that \(\mu^{\alpha, \beta} \approx \mu^{\gamma, \delta} \mu^{\sigma, \rho}\). This is not always valid as both electrical and nuclear anharmonicity result in a deviation of the harmonic scaling (both the amplitudes and collinearity). We will demonstrate the performance of these metrics later, but with the assumption of harmonic scaling, the dependence of the peak amplitudes as a function of \(\theta\) can be established and figure 2.7 shows a couple of examples. All of these relations involve the \(\cos^2(\theta)\) factor and therefore both \(\theta\) and \(\pi - \theta\) lead to the same value of the ratios. Resolving this degeneracy and ambiguity requires more insight of the system, such as knowledge about the molecular structure or comparison to computer simulations. For systems with more than two vibrations within the spectral region of interest, polarization-selective 2D IR experiments can simultaneously yield the relative orientations between all the transition dipoles and therefore provide more experimental constraints to resolve this issue (see Chapter 4).

Figure 2.7a plots the orientational factors of the absorptive cross-peak measured with the ZZZZ and ZZYY polarizations. As the angle between the two transition dipole moments changes from \(0^o\) to \(90^o\), the ZZYY amplitude \(A^{+\text{ZZYY}}\) (where the "+" sign indicates the positive part of the cross-peaks) increases gradually from 0.2 to 0.23 while the ZZZZ amplitudes \(A^{+\text{ZZZZ}}\) drops rapidly from 0.6 to 0.2. As a consequence, \(A^{+\text{ZZYY}}\) is smaller than \(A^{+\text{ZZZZ}}\) for most orientational configurations, but becomes bigger when the two transition dipole moments are close to being perpendicular. Taking the ratio of these two curves give the relation described in Equation (2.26) and plotted in Figure 2.7b. The ratio increases from 0.33 to 1.17 as the two dipole moments transition from parallel to perpendicular, and therefore this metrics is most sensitive when the two dipole moments are \(90^o\) apart. In fact, this is true also for the emissive cross-peak (Figure 2.7c) or the entire cross-peak (Figure 2.7d). In particular, the amplitude ratio of the entire cross-peak has a bigger range from 0.33 to 2, making it a more sensitive metrics. In addition, as will be shown later that this metrics is
more useful when the system has small off-diagonal anharmonicity such that the absorptive and emissive cross-peaks interfere significantly.

Figure 2.7e shows a metrics that involve a different polarization scheme, \( \Delta_{ZZZZ}/\Delta_{ZYY} \). This was recently used by the Fleming group to determine the relative angle between the electronic transition dipole moments of the chlorophyll in CP29 light harvesting complex.\(^{18}\) This amplitude ratio ranges from 1 to -0.25 as the angle changes from 0° to 90°. It
can be seen that this curve changes more drastically for angles near 0° compared to the previous three metrics, and therefore is more suited for angle determination in this range.

Finally, one could also determine the angle from measuring the cross-peak anisotropy $\alpha$ defined as:

$$\alpha = \frac{A_{ZZZ} - A_{ZZY}}{A_{ZZZ} + 2A_{ZZY}}$$  \hspace{1cm} (2.28)

We discuss anisotropy only for the absorptive cross-peak centered at $(\omega_i, \omega_f)$, but for three different representations of 2D IR spectra: rephasing, non-rephasing, and correlation spectra. Inserting the orientational factors into Equation (2.28) results in the curves plotted in Figure 2.7f. The anisotropy of the cross-peak is 0.4 for two parallel transition dipoles regardless of the type of 2D IR representation but becomes 0.1, -0.2, and -0.05 for rephasing, non-rephasing, and correlation spectra, respectively, when the two dipoles are perpendicular.

To demonstrate the applicability of these metrics, we simulated 2D IR spectra of RDC with harmonic scaling of the transition dipole moments (notice that this is not true in experimental spectra). The line shapes in this calculation has not been fitted to the experimental spectra. Figure 2.8a and b plot the absorptive correlation spectra with ZZZZ and ZZYY polarizations, respectively. Because the symmetric (2084 cm$^{-1}$) and asymmetric (2010 cm$^{-1}$) C=O stretches are perpendicular, the cross-peaks are enhanced in the ZZYY spectrum. However, we should point out that the cross-peak enhancement appeared in the normalized ZZYY spectrum is mostly due to the signal suppression of the diagonal peaks. Figure 2.8c compares the frequency slices at $\omega_i = 2010$ cm$^{-1}$ for the two polarization-selective spectra. Using the orientational factors, it can be derived that the amplitude ratio of the positive part of the diagonal peak changes from 0.33 to 0.26 when the two transition dipole moments vary from parallel to perpendicular. The ratio of the peak height at $\omega_3 = 2997$ cm$^{-1}$ shows a value of 0.28, which is consistent with our expectation. Nevertheless, since the dynamic range of this metrics is small (from 0.33 to 0.26), it is not a suitable metric to determine the relative dipole orientation. In addition, if the diagonal peak is coupled to more than one diagonal peak, its amplitude ratio will depend on all the angles between different pairs of transition dipole moments. On the other hand, the amplitude ratio of the cross-peaks is sensitive specifically to one angle between the two coupled transition dipoles (under the harmonic scaling assumption). As a result, it is more useful to analyze the cross-peaks for angle determination.

As pointed out earlier that the amplitude of the cross-peak measured with ZZYY polarization is normally smaller than the values measured with ZZZZ polarization, except
Figure 2.8: Simulated 2D IR spectra for RDC assuming harmonic scaling of the transition dipole moments. The lineshapes have not been fitted to the experimental spectra. Absorptive correlation spectrum with ZZZZ (a) and ZZYY (b) polarizations. The red box around the cross-peak represents the peak intensity integration box. (c) Frequency slices along $\omega_1 = 2015 \text{ cm}^{-1}$ from the ZZZZ (blue) and ZZYY (red) correlation spectra. Two-dimensional anisotropy map for the correlation (d), rephasing (e), and non-rephasing (f) surfaces. The anisotropy values expected for cross-peaks between two perpendicular transition dipoles are highlighted in red in the colorbar. In order to move noise originated from dividing numbers close to zero, anisotropy values less than 5% of the maximum value have been discarded.

for when the two dipoles are nearly orthogonal. Figure 2.8c shows a $\Lambda_{ZZYY}^+ / \Lambda_{ZZZZ}^+$ ratio of 1.14, which is very close to the theoretical value of 1.17 for two perpendicular dipoles (Figure 2.7b). Integrating the signal for the entire cross-peak doublet (both the absorptive and emissive signals, red box in Figure 2.8a,b) gives an amplitude ratio of 2, which corresponds to the theoretical value for $90^\circ$.

Next we examine the anisotropy of the correlation (Figure 2.8d), rephasing (Figure 2.8e), and non-rephasing (Figure 2.8f) 2D IR spectra, using Equation (2.28). As shown in Figure 2.7f, the anisotropy for two parallel transition dipoles should be 0.4, regardless of the
type of 2D IR spectra. This limit applies to the diagonal peaks (dark red indicates a value ~0.4 in Figure 2.8d-f) of a system containing only two coupled oscillators. However, the anisotropy values are different for various 2D IR representations when the two dipoles deviate from parallel. The values corresponding to 90° are highlighted in red next to the colorbar for the three representations. It is clear that all three anisotropy maps reveal orthogonal transition dipoles.

In theory and as shown by the RDC model calculation, these metrics provide a convenient way to measure the relative dipole orientation quantitatively (with the caveat of being either 0 or 180° - 0). As a consequence, it has become a common practice in the community to compare the frequency slices cutting through the cross-peak maximum from two different polarization spectra to obtain the angle.14,19,20 However, when the system under study has small anharmonicity that leads to significant overlap between the positive and negative peaks, the peak amplitude becomes an unreliable metric. We will use simulated 2D IR spectra for N-methylacetamide (NMA) in DMSO to illustrate the break-down of these metrics. Using the nonlinear response function formulism outlined previously, the experimental spectra of NMA in DMSO have been fitted (Figure 2.9) and the angle between the amide I (1665 cm⁻¹) and amide II (1555 cm⁻¹) transition dipoles was found to be 40°.21

![Figure 2.9: (left) Experimental and fitted 2D IR spectra with ZZZZ and ZZYY polarizations for NMA in DMSO-d₆. (right) The normal mode displacement of the amide I and amide II modes, and the relative orientation of these two transition dipoles. Figure adapted from ref. [21] with permission.](image-url)
Now we apply the different metrics to analyze the simulated spectra and see if we can reliably retrieve the input angle parameter of 40° from either comparing the integrated peak intensities or anisotropy. The peak intensity integration boxes are shown in Figure 2.10—blue, green, and red boxes are for the absorptive, emissive, and the entire cross-peak. It is usually better to examine the ratio of the uphill cross-peaks ($\omega_1 < \omega_2$) since the downhill cross-peaks often experience destructive interference from the $1 \rightarrow 2$ transition resonance of the diagonal peak at the same $\omega_1$ frequency. The results are summarized in Table 2.2.

The angles extracted from looking at either the positive or negative part of the cross-peak are very different from the input 40°, and do not vary significantly with the integration boxes. Therefore the discrepancy comes mainly from the fact that there is substantial interferences between the positive and negative cross-peaks. This is in sharp contrast to the results for the simulated RDC spectra where the off-diagonal anharmonicity is large compared to the linewidth. On the other hand, using the entire cross-peak slightly improves the determination ($49.5 \pm 1.5°$), even though it still does not reproduce exactly 40°. Finally, we also test the anisotropy method which calculates $(A_{zzzz} - A_{zzzy})/(A_{zzzz} + 2A_{zzzy})$, shown in Figure 2.10. Unfortunately this method results in the largest uncertainty as anisotropy ranging from ~0.1 to ~0.3 is observed in the cross-peak region. The average anisotropy value in the absorptive cross-peak region is 0.218 ± 0.078, which gives an angle of 45.2° ($+11.4°, -9.8°$). The anisotropy method provides the closest value to the input parameter but

![Figure 2.10: Simulated 2D IR spectra with ZZZZ (left) and ZZYY (middle) polarizations for NMA in DMSO-d$_6$. (right) Two-dimensional anisotropy calculated from the correlation spectra with ZZZZ and ZZYY polarizations: $(A_{zzzz} - A_{zzzy})/(A_{zzzz} + 2A_{zzzy})$. Anisotropy value of 0.26 is expected (highlighted in red next to the colorbar) for an angle of 40°. The blue, green, and red boxes in the 2D IR spectra show the integration region for the absorptive, emissive, and the entire cross-peak.](image-url)
also the largest ambiguity. These calculations show that even for the simulated spectra, it is challenging to retrieve the relative dipole orientations by directly applying these metrics.

In practice, experimental spectra inevitably contain noise due to various issues. The errors in determining dipole orientation using these methods will be worse. Furthermore, one should clearly specify which metrics is being used. Sometimes the equation for the entire cross-peak (Figure 2.7d) is mistakenly applied to the amplitude ratio of the positive part of the cross-peaks. Such inconsistency can result in enormous error, for example, two orthogonal transition dipole moments lead to a value of $A^{'zzz}/A^{'ZZZZ} = 1.17$, which based on the curve in Figure 2.7d will give an angle of $60^\circ$. Therefore, one should be careful in applying these methods, and the accuracy is highly dependent on the system being studied. In the end, we found that fitting the experimental spectra with the full nonlinear response function formulism seems to provide the best estimate of the angles because the interference effects are better captured in these simulations. In Chapter 4, we will use this approach to determine the relative dipole orientations of the various modes in DNA bases.

Table 2.2. Peak ratios between the ZZYY and ZZZZ 2D IR cross-peaks and the derived angle between the two transition dipole moments. The peak ratios were obtained from integrating the boxes shown in Figure 2.10. Ratios of the peak maximum values are close to the average values. The theoretical values for two dipoles oriented $40^\circ$ apart are also listed.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Theoretical value for $40^\circ$</th>
<th>Peak ratio</th>
<th>Calculated angle ($^\circ$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive cross-peak</td>
<td>0.49</td>
<td>0.626 ± 0.015</td>
<td>51.9 ± 1.1</td>
</tr>
<tr>
<td>Negative cross-peak</td>
<td>0.41</td>
<td>0.571 ± 0.007</td>
<td>63.4 ± 0.8</td>
</tr>
<tr>
<td>Entire cross-peak</td>
<td>0.65</td>
<td>0.855 ± 0.036</td>
<td>49.5 ± 1.5</td>
</tr>
<tr>
<td>Anisotropy</td>
<td>0.26</td>
<td>0.218 ± 0.078</td>
<td>45.2 (+11.4, -9.8)</td>
</tr>
</tbody>
</table>

2.5 Acknowledgments

I thank Mike Reppert for helpful discussion and careful reading over this chapter.
2.6 References


CHAPTER 3
Experimental Methods

3.1 Introduction

After presenting the theoretical background of 2D IR spectroscopy and demonstrating how we obtain specific information about the system such as eigenenergy, anharmonicity, vibrational coupling and relative dipole orientation, the experimental implementation is discussed in this Chapter. Details of the experimental methods have been published previously,\textsuperscript{1,2} so only a brief review is provided here. Equilibrium 2D IR experiments are introduced first, followed by transient experiments. In addition to the optical setup, data processing and basic interpretation for the 2D IR experimental data are given. Temperature-jump (T-jump) transient experiment offers a powerful tool to investigate fast chemical and biological process in the nanosecond to millisecond timescale. However, both the experimental implementation and data interpretation are more complicated than the equilibrium experiments due to the additional T-jump pulse. In fact, examples of transient 2D IR experiments (using either T-jump or UV excitation) are scarce, and to date, the Tokmakoff group is the only group that has successfully conducted T-jump 2D IR experiments. Here the spectral features and the time evolution of transient 2D IR signal are discussed.
3.2 Experimental setup for 2D IR spectroscopy

3.2.1 IR generation

Coherent 2D IR spectra are acquired using fs pulses with enough spectral bandwidth to excite all the vibrations of interest. Figure 3.1 shows the schematic of the front end of the 2D IR spectrometer. A Spectra-Physics Tsunami Ti:Sapphire oscillator pumped by a Spectra-Physics Millenia Nd:YAG laser (532 nm, CW, 6 W) generates 84 MHz of 790 nm light (FWHM=16 nm, ~600 mW). The 790 nm light is used as the seed for Spectral-Physics Spitfire regenerative amplifier, which is pumped by a Spectra-Physics Empower Nd: YLF laser (532 nm, 1 kHz, 10 W). The Spitfire generates a 1kHz train of 100 fs pulses which are sent into Spectra-Physics OPA-800 to create the IR pulses. OPA-800 is a two-stage optical parametric amplifier (OPA) that converts the 790 nm light into signal (1.4 μm, 90 μJ) and idler (1.84 μm, 40 μJ) after a Type II process in 1 mm β barium borate (BBO) crystal. The signal and idler are focused into 1 mm AgGaS₂ crystal for difference frequency generation (DFG). The spectrum is tunable from 4 to 6 μm with different DFG crystal angles. Typical pulses have energies of 3 to 4 μJ, and bandwidth of 160 cm⁻¹.

Figure 3.1: Schematic of the oscillator, amplifier, and OPA lay-out (not drawn to scale). A Ti:Sapphire oscillator pumped by a Nd:YAG laser produces 100 fs pulses at 790 nm. The 790 nm light is amplified by a regenerative amplifier and sent to an OPA. The signal (1.4 μm) and idler (1.84 μm) are used to generate the desired 4 ~ 6 μm light using difference frequency generation (DFG).
3.2.2 He-Ne overlap

Once the mid-IR light is generated, the downstream IR alignment in the interferometer and sample area is made convenient by first overlapping the IR with a visible HeNe beam. The HeNe overlap setup is shown in Figure 3.2. The HeNe beam size is adjusted to match that of the IR beam using a telescope (L1 and L2), and is combined with the IR beam at the germanium plate. The HeNe beam path is kept fixed and the IR overlap is done by maximizing the IR signal on the detector in the near field (Det after L3) and far field (Det after L4, path length ~experimental path length to the dual stripe MCT detect). The IR alignment in the near and far field is controlled by M4 and M5, respectively. After the IR and HeNe beams are spatially overlapped, they go through a periscope which rotates their polarization by 90° to S (vertical) polarization, and then through a pair of spherical mirrors which expands and collimates the beam size to 1 cm. The divergence of the IR beam is checked by mapping out the intensity profile at the far field detector as a function of distance.

Figure 3.2: Schematic of the HeNe overlap setup. M: mirror; SM: spherical mirror; fM: flip mirror; L: lens; I: iris; Ge: germanium; X: periscope.
3.2.3 2D IR interferometer

The 2D IR experiments are performed with a "boxcar" geometry where three pulses are focused into the sample from three corners of a 1 inch box and the nonlinear signal is emitted into the fourth corner (see inset in Figure 3.4). A five-beam interferometer (Figure 3.3) is used to split the IR pulse from the OPA (after HeNe overlap) into multiple pulses for the 2D IR experiments. The incoming IR beam is first split by a beam splitter (BS1) that is uncoated on the front and anti-reflection (AR) on the back. With an incident angle of 56°, 60% of the beam is transmitted while 40% is reflected. The transmitted beam is further split by BS2 (50%/AR front/back) to form beams 1 and 2, each with 30% power of the original IR beam.

The 40% reflected beam from BS1 is split by BS3 to form beam 3 and beam 4 (which is later split into tracer and local oscillator). Since in practice, the tracer beam is not used to generate nonlinear IR signal and that only a small local oscillator (LO) power is required for heterodyne detection, we use an AR/AR beam splitter to reflect only ~2% of the beam power for tracer/LO. Therefore, the transmitted beam 3 through BS3 has 40% x 98% = 39% of the original IR power, while the tracer/LO beam has 40% x 2% = 1% power. Beam 4 is further split by BS4 to give equal intensity tracer and LO (each with 0.5% power).

The intensities and polarizations of the beams are controlled by 2 mm thick MgF₂ half-wave plates, followed by 2 mm thick ZnSe wire-grid polarizers. Each of beams 1, 2, and 4 is independently delayed with respect to fixed beam 3 using a retroreflector mounted on an Aerotech stage (50 mm travel range, which corresponds to 333 ps delay). Beams 1, 2, 3, and tracer are aligned in a 1" boxcar geometry shown in the inset of Figure 3.3. Beam 2 is chopped by a 500 Hz chopper. The tracer beam, which propagates along nonlinear signal direction, is used for rough alignment and blocked during data acquisition. The LO is used for heterodyne detection and can also be utilized to take pump-probe spectra; with beam 2 and LO being the pump and probe, respectively.
Figure 3.3: Experimental layout for the five-beam interferometer, showing the three input pulses 1, 2, and 3 and the tracer and the LO pulse. The wheel is the 500 Hz chopper; BS: 4-mm-thick ZnSe beam-splitter; C: 4-mm-thick ZnSe compensation plate; WP: 2-mm-thick MgF₂ wave plates; P: wire-grid ZnSe polarizers. The intensity split ratio by the beam splitters is indicated, and the final powers of the five beams are also specified.

3.2.4 2D IR signal detection

Figure 3.4 illustrates the optical layout for signal detection. The IR pulses are focused to the sample using a 3-inch, 10 cm focal length, gold off-axis parabolic mirror (PM1). The boxcar geometry is done by aligning each of the beams (1, 2, 3, and tracer) through an iris (different iris for each beam) at the entry port of the detection box, using the horizontal and vertical control of the Aerotech stages. The four beams are then focused through a 75 μm pinhole at the sample position using the square mirrors after the retroreflectors. Generally, > 70 % throughput can be achieved. M11 is a removable square mirror that directs all beams to PM3 which focuses into a single-channel MCT detector used for alignment purposes.
Figure 3.4: Optical layout of the signal detection box. M: mirrors; TM: mirrors for the T-jump laser; PM, gold-coated parabolic mirrors; BS: beam splitter; S, sample; Mono, monochromator (not drawn to scale).

After PM2, beams 1-3 are blocked and the signal is sent to a 50%/AR beam splitter (BS1, shown in the inset of Figure 3.4). The reflected and transmitted beams are vertically separated (transmitted beam is displaced higher), and focused into the monochrometer, which diffracts to two vertically displaced stripes of 2x64-pixel MCT array. Balance detection is used for collecting the nonlinear signal to significantly increase the signal-to-noise ratio of the heterodyne measurement by eliminating the baseline noise in the DC component of the interferogram. The balanced heterodyne detection is performed by recombining the signal and LO at BS1. Because there is a 180° phase shift between the internal and external reflection, the upper and lower stripe of the MCT array collect equal intensity but oppositely signed cross-terms between the signal and LO:

\[
S_v = |E_S - E_{LO}|^2 = I_S + I_{LO} - 2\Re\left[ E_S E_{LO}^* \right] \\
S_l = |E_S + E_{LO}|^2 = I_S + I_{LO} + 2\Re\left[ E_S E_{LO}^* \right] \tag{3.1}
\]

Beam 2 is chopped at 500 Hz, so when it’s chopped, both stripes only measure \( I_{LO} \). Therefore, the differential signal of consecutive laser shots is free of \( I_{LO} \). Finally, the difference signal from both stripes subtracts the homodyne signal \( I_S \) and scattering from beam 2, and we end up with enhanced nonlinear signal.
3.2.5 Timing up pulses

In order to generate 3rd-order nonlinear signal, beams 1-3 have to be spatially and temporally overlapped at the sample position. To do so, one can either use intensity autocorrelation, or interferometric measurements as shown in Figure 3.5.

Figure 3.5: Optical setup for the three autocorrelation methods. (a) Intensity autocorrelation which is background-free because the two beams sent into the SHG crystal are not collinear and can be easily blocked. (b) Field autocorrelation of the scattered light. There is no nonlinear effect involved in this measurement. In our experiments, a 75 μm pinhole is used to scatter the IR pulses and measure the relative timings. (c) Interferometric autocorrelation. Two collinear and time delayed pulses are focused into a nonlinear crystal to generate SHG. The SHG signal generated in this scheme has more contributions than the signal measured in background-free intensity autocorrelation. See text for more details. The simulated signals for a 100 fs Gaussian pulse using the three methods are plotted on the right. For method (a), a Gaussian fit to the signal shows that the incoming pulse width is 141 fs/1.41 = 100 fs.
3.2.5.1 **Intensity autocorrelation**

The relative time delays between the beams can be determined by taking background-free intensity autocorrelation of each pulse pair in a type I AgGaS₂ crystal placed at the sample spot. The desired second harmonic generation (SHG) in the direction going to the detector is generated when the two pulses are spatially and temporally overlapped (timed-up at \( t_0 \)). The second polarization is proportional to the product of the two electric fields and the detector measures the intensity of the SHG as follows:

\[
I_{AC}(\tau) \propto \int_{-\infty}^{\infty} |E(t)E(t-\tau)|^2 \, dt
\]  

(3.2)

Because a cross-beam geometry is used in this measurement, the fundamental frequency light can be easily isolated, and hence the measurement is background-free. It is clear that the intensity autocorrelation does not contain full information about the electric field of the pulse. However, one advantage is that if the pulse shape is known, then the pulse width can be obtained from the FWHM of the intensity autocorrelation trace by dividing out a constant of 1.414 or 1.542 for Gaussian or Sech² pulses, respectively. As shown in Figure 3.5a, a Gaussian fit to the simulated intensity autocorrelation trace reveals the 100 fs input pulse width.

3.2.5.2 **Field autocorrelation**

A slightly simpler method to find temporal overlap is to use interferometric measurement of the two pulses by scattering them with a pinhole and stepping one of the beams (Figure 3.5b). In this case, the detector measures:

\[
I_{FAC}(\tau) \propto \int_{-\infty}^{\infty} |E(t) + E(t-\tau)|^2 \, dt
\]

\[
\propto \int_{-\infty}^{\infty} |E(t)|^2 \, dt + 2Re \left[ \int_{-\infty}^{\infty} |E(t)E(t-\tau)|^2 \, dt \right]
\]  

(3.3)

The first term represents the pulse energy and the second term is the field autocorrelation, of which the Fourier transform gives the spectrum (because the Fourier transformation of a convolution is the product of the Fourier transform of the two functions \( \hat{E}(\omega)\hat{E}^*(\omega) = I(\omega) \)). Therefore, measuring the interferogram is equivalent of measuring the pulse spectrum. However, this method does not give you phase information nor the pulse width. For our purposes, field autocorrelation provides a convenient way of obtaining the timing of two pulses, which is where the maximum signal occurs. Figure 3.6 illustrates how we find the timing in the lab by scattering two beams through a 75 μm pinhole and scanning...
the time delay of one beam relative to another. A 2D data array is measured as function of frequency and time delay (Figure 3.6a). At delays times away from \( t_0 \), the interferogram has many fringes (Figure 3.6c, e); and at \( t_0 \), it resembles the pulse spectrum (Figure 3.6d). Integrating the time traces (Figure 3.6b) from all 64 pixels results in the expected field autocorrelation signal (Figure 3.6f, see Figure 3.5b for comparison). The DC component is filtered out in the Fourier domain, and the absolute value square is taken to obtain a Gaussian profile, which could be fitted to get the \( t_0 \) value (Figure 3.6g).

Figure 3.6: Experimental data for field autocorrelation measurement using pinhole scattering. (a) Raw data as function of frequency (represented by the pixel number on the array) and time delay (fs). (b) The time trace for pixel \# 32. (c-e) The interferograms at negative time, time zero, and positive time. At \( \sim t_0 \), the interferogram resembles the pulse spectrum, whereas at times away from \( t_0 \), multiple fringes are observed. (f) Integrating over all the pixels results in the expected field autocorrelation signal, \( I_{FAAC}(\tau) \). (g) After the DC signal of \( I_{FAAC}(\tau) \) is filtered out in the Fourier domain, the absolute value squared of the filtered \( I_{FAAC}(\tau) \) results in a Gaussian profile that can be fitted to give the \( t_0 \).
3.2.5.3 Interferometric autocorrelation (IAC)

Interferometric autocorrelation (IAC) is done by sending two collinear and time-delayed pulses into a SHG crystal. Due to the collinear geometry, there are more contributions to the measured signal as shown below:

\[ I(\tau) \propto \left\{ \left| E(t) + E(t - \tau) \right|^2 \right\} dt \quad (3.4) \]

Equation (3.4) can be expressed as a sum of four terms:

\[ I(\tau) = I_{\text{back}} + I_{\text{Ac}}(\tau) + I_{\omega}(\tau) + I_{2\omega}(\tau) \quad (3.5) \]

where

\[ I_{\text{back}} = \int_{-\infty}^{\infty} \left( |E(t)|^4 + |E(t - \tau)|^4 \right) dt \]

\[ I_{\text{Ac}}(\tau) = 4 \int_{-\infty}^{\infty} |E(t)E(t - \tau)|^2 dt \]

\[ I_{\omega}(\tau) = 4 \int_{-\infty}^{\infty} \Re \left[ (I(t) + I(t - \tau)) E^*(t)E(t - \tau)e^{-i\omega \tau} \right] dt \]

\[ I_{2\omega}(\tau) = 2 \int_{-\infty}^{\infty} \Re \left[ (E(t)E^*(t - \tau))^2 e^{-2i\omega \tau} \right] dt \quad (3.6) \]

The first term \( I_{\text{back}} \) is independent of the time delay. Therefore, the interferometric autocorrelation result is sitting on a constant background which is usually normalized to 1. The second term \( I_{\text{Ac}}(\tau) \) is the signal obtained from the second harmonic intensity autocorrelation, and the last two terms \( I_{\omega}(\tau) \) and \( I_{2\omega}(\tau) \) are the cross-terms which oscillates at frequencies \( \omega \) and \( 2\omega \), respectively.

The interferometric autocorrelation signal is shown in Figure 3.5c. It has a constant background of 1 and a maximum value of 8 at \( t_0 \). IAC can also be used to check for dispersion (the baseline will not be flat for a chirped pulse). However, due to the different contributions as shown in Equation (3.6), the width of the envelope of IAC is sensitive to phase modulation (whereas intensity autocorrelation is not) and therefore the pulse width is not directly measured from the width of the IAC profile. However, isolating the \( I_{\text{Ac}}(\tau) \) component in the frequency domain will allow the width determination.
3.2.6 Data processing for equilibrium 2D IR spectra

To collect a 2D spectrum, the timing between the signal and LO \( \tau_1 \) is set to zero and the grating in the array detector effectively Fourier transforms the \( \tau_1 \) time period to generate the \( \omega_1 \) axis. The coherence time \( \tau_1 \) is step-scanned in a 4 fs interval from 0 to 2 ps to adequately resolve the 20 fs oscillation corresponding to the 6 \( \mu \)m light. For the rephasing experiment, \( E_1 \) enters the sample followed by \( E_2 \) and \( E_3 \); while for the non-rephasing experiment, the pulse sequence of \( E_1 \) and \( E_2 \) is reversed. Before entering the interferometer, the IR beams are polarized vertically (Z) in the laboratory frame. Polarization-sensitive 2D IR spectra are taken at two polarization geometries, all-parallel (ZZZZ) and perpendicular (ZZYY) by changing the polarizations of beams 1 and 2.

Figure 3.7 shows the raw experimental data of guanosine 5'-monophosphate (GMP) to illustrate how we process the data to get phased 2D IR spectra. The rephasing and non-rephasing surfaces are shown in panels a and b, respectively, and the time slices at \( \omega_1 = 1672 \) cm\(^{-1}\) are shown in panel c. The rephasing signal persists longer than the non-rephasing signal. Even though we have set the timings between the pulses to be near 0 fs using the methods described previously, in practice there is still a little bit of uncertainty in \( \tau_1 \) which results in a mixing of the real and imaginary parts of the rephasing and non-rephasing signals. In order to correct for the timing error, we use a phasing procedure based on the projection-slice theorem, \(^3\) which states that the \( \tau_1 = 0 \) fs slices of the rephasing and non-rephasing surfaces are equal to the independently collected two-beam pump-probe. The timing error \( \Delta \tau_1 \) can be fixed in the time domain by directly shifting the data along \( \tau_1 \). Therefore, in order to compensate for both negative and positive timing error, we scan the coherence time from -60 fs (however, we generally time up the pulses to within \( \pm 5 \) fs before taking data). Figure 3.7d shows the agreement between the pump-probe and the zero time slices of the rephasing and non-rephasing surfaces after phasing procedure. Further evidence of correct phasing comes from the agreement between the slices around time zero. As shown in Figure 3.7e, the 4 fs rephasing slice matches the -4 fs non-rephasing slice, and vice versa.
Figure 3.7: Experimental data for 20 mg/ml of guanosine 5'-monophosphate (GMP). (a, b) Rephasing and non-rephasing surfaces as a function of $\tau_1$ and $\omega_3$. (c) Time slices at $\omega_3 = 1672 \text{ cm}^{-1}$. Blue is rephasing and green is non-rephasing. (d) Frequency slices of rephasing and non-rephasing surfaces at 0 fs, and fit to the pump-probe signal by adjusting the phase. (e) Frequency slices at -4 and 4 fs.

Fourier transformation of the rephasing and non-rephasing $\omega_3-\tau_1$ surfaces results in two-dimensional spectra. These 2D spectra are complex values and can be represented as the real part, imaginary part, or absolute value shown in Figure 3.8. It is typical to show and analyze the absorptive 2D “correlation” spectrum, which is the real part of the sum of the rephasing and non-rephasing surfaces. The correlation spectrum provides the most spectral information, for example, the information on anharmonicity is lost in the power spectrum. The one-dimensional representations of the 2D data are shown on the left panel of Figure 3.8 and can be compared to the FTIR spectrum. The mathematical relations between the 2D data and the dispersed vibration echo (DVE), heterodyne dispersed vibration echo (HDVE), and dispersed pump-probe (DPP) are as follows:

$$\tilde{S}_{HDVE}(\omega_3) = \int_{-\infty}^{\infty} \tilde{S}_{2D}(\omega_1, \omega_3) d\omega_1 = \tilde{S}_{2D}(\tau_1 = 0, \omega_3)$$

$$S_{DVE}(\omega_3) = \left| \int_{-\infty}^{\infty} \tilde{S}_{2D}(\omega_1, \omega_3) d\omega_1 \right|^2 = |\tilde{S}_{HDVE}(\omega_3)|^2$$

$$S_{DPP}(\omega_3) = \int_{-\infty}^{\infty} S_{corr}(\omega_1, \omega_3) d\omega_1 = \Re \left[ \tilde{S}_{HDVE}(\omega_3) \right]$$

where $\tilde{S}_{2D}(\omega_1, \omega_3)$ is the sum of the complex rephasing and non-rephasing spectra.

76
Figure 3.8: Experimental spectra for GMP. (Left) FTIR and linear representations of the 2D data including DVE, HDVE, and DPP. These three spectra are of function of $\omega_3$ so the resolution is only 4 cm$^{-1}$. (Right) Different representations of the 2D IR spectra including the real and imaginary parts of the rephasing and non-rephasing spectra, as well as their power spectra. The absorptive correlation 2D spectrum is the real part of the sum of $S_R$ and $S_{NR}$. And the “absolute value” 2D spectrum is the power spectrum of the sum of $S_R$ and $S_{NR}$. 
3.3 Transient 2D IR experiments

The use of multiple pulses in 2D IR spectroscopy allows for time-dependent experiments by varying the waiting time $\tau_2$ between the pump and probe pulses. These experiments have proven to be useful in studying chemical exchange, water hydrogen-bond switching, protein conformational dynamics, and energy relaxation in DNA oligomers. However, these experiments are limited by the vibrational lifetime of the molecules, which is typically on the picosecond timescale in water. To overcome this problem in order to obtain fast kinetics in the nanosecond to microsecond timescale, such as thermal folding transition of biomolecules, one can perform transient experiments with a laser-induced temperature jump (T-jump), pH jump, or UV excitation to abruptly shift biochemical equilibria, and then monitor the relaxation processes. Transient 2D IR experiments following either UV–pump or T–jump have been utilized to study the conformational dynamics of peptides and proteins, the kinetics of metal-to-ligand charge transfer, electron injection in TiO$_2$ thin films, and vibrational relaxation dynamics of the photoproducts of metal carbonyls.

In our experiments, transient T-jump 2D IR spectroscopy was performed with the setup described in details (synchronization of the T-jump and 6 $\mu$m pulses, temperature characterization... etc) previously. Briefly, the 5 ns, temperature-jump is generated by a 20 Hz Nd:YAG pumped optical parametric oscillator (OPO). The output pulse of the OPO is tuned to 1.98 $\mu$m with 10 mJ energy, and focused into the sample. The T-jump beam path is shown in Figure 3.4 as the thick green line, and the geometries of the multiple laser beams at

Figure 3.9: (Left) Laser beam geometry at the sample. (Right) Temperature profile upon T-jump excitation. Green and blue pulses represent the T-jump and 6 $\mu$m pulses, respectively. The temperature is raised by $\sim 10^0$C in 5 ns, and stays relatively constant up to $\sim 100$ $\mu$s. Fast kinetics such as tautomerization or DNA/protein unfolding can be monitored by 2D IR experiments during this time period. Finally the temperature relaxes on the ms timescale.
the sample spot are shown in Figure 3.9. The T-jump pulse heats D₂O by exciting the O-D stretch overtone of the D₂O solvent, which then relaxes to lower frequency modes, raising the temperature of the sample. The T-jump pulse is focused to a 500 µm waist, which is about five times bigger than the IR beam size, therefore giving a uniform heating within the sample area being probed. About ~10°C T-jump can be generated with a single pass through the sample, and ~20°C can be created with double pass using TM4 in Figure 3.4.

The temperature profile is shown in Figure 3.9. The 10°C T-jump is achieved in about 5 ns, and the temperature stays relatively constant for approximately 100 µs. Fast kinetics such as tautomerization of aromatic heterocycles or DNA/protein unfolding can be monitored with transient 2D IR experiments during this time window. Fifty 1 kHz, 6 µm IR pulses (blue pulses) following the T-jump pulse (green pulse) are used to measure the transient IR spectra. Heterodyne-detected dispersed vibrational echo (HDVE) and 2D IR spectra are collected as a function of the electronically generated T-jump delay (τ), a time delay between the T-jump pulse and the first of the fifty IR beams. The time delay between the T-jump pulse and the jth fifty IR pulse is denoted as:

$$\tau_j = \tau + (j-1) \times 1 \text{ms}, \quad \text{for } j = 1, 2, ..., 50. \quad (3.8)$$

The T-jump re-equilibrates on the ms time scale following a stretched exponential form \( \exp[-(t/2.05\text{ms})^{0.67}] \) as the heat diffused to the sample cell CaF₂ windows (decayed to ~1% by 7 ms). Changing the path length or window materials can change the thermal profile. Ideally, we would like the temperature to stay constant over the entire time window before the next T-jump pulse comes in, and of course, we need to make sure the new path length and window material is suitable for measuring desired nonlinear IR signal.

### 3.3.1 Transient data collection

Both transient HDVE (t-HDVE) and transient 2D IR (t-2D IR) experiments are performed with processing the fifty 6 µm IR data sets independently after each T-jump. The double difference detection scheme is shown in Figure 3.10a. The 49th and 50th pulses are used as the reference spectra since the temperature has re-equilibrated back to the initial temperature. The 6 µm IR pulses (beam 2) are chopped at 500 Hz, so every other shot is blocked (open and close circle indicate unblocked and b, respectively). After the first data set
with 0° chopper phase is acquired, the chopper phase is flipped to 180° to measure the second data set. At each T-jump delay, the differential signal between the open and close pairs gives the transient spectrum. Finally the transient difference spectrum is obtained by subtracting the reference spectrum:

$$\Delta \tilde{S}(\tau_j) = \left[ I_0(\tau_j) - I_0^*(\tau_j) \right] - \left[ I_0(\tau_49) - I_0^*(\tau_49) \right] \quad \text{odd } j$$

$$= \left[ I_\pi(\tau_j) - I_\pi^*(\tau_j) \right] - \left[ I_\pi(\tau_50) - I_\pi^*(\tau_50) \right] \quad \text{even } j$$

(3.9)

### 3.3.1.1 Transient 2D IR

The data collection for t-2D IR is done in three loops as shown in Figure 3.10b: (1) chopper phase, (2) coherence time \(\tau_1\), and (3) T-jump delay \(\tau\). The T-jump delay and the coherence time are set to the first value. The chopper phase is set to 0°, and 7000 shots are acquired before the chopper phase is flipped to 180°, after which another 7000 shots are acquired. Then \(\tau_1\) delay is stepped in 8 fs step to next \(\tau_1\) value. We noticed that undersampling at 14 fs resulted in line-shape distortions in particular for the ring modes, most likely due to the coupling to lower frequency modes. The 2x7000 shots are collected and \(\tau_1\) is stepped again. The coherence time is scanned to 2 ps and 1 ps for rephasing and non-rephasing surfaces, respectively. The process is repeated until the entire rephasing and non-rephasing surfaces are mapped out, and then \(\tau\) is moved to the next T-jump delay. Spectra collected at the same \(\tau\) delay are averaged in the frequency domain. Because the ms spectra are acquired for each electronic delay \(\tau\) (ns to \(\mu s\)) and consider to be equivalent \((ms + (ns, \mu s) \sim ms)\), they have much
higher signal to noise ratio compared to the ns or μs transient spectra. In general, ≥ 5 averages of the complete set of t-2D IR are required to get good signal to noise ratio. Therefore, for a single T-jump delay 2D IR spectrum, it takes 5 avg*375 τ*2 chopper phase*7000 shots*1 ms ~ 7 hours. Since many T-jump delays are required for a full picture of the short-time dynamics, data acquisition time can exceed 24 hours for a single sample. Many technical issues can arise over this long period of heating time, for example solvent evaporation, sample degradation/aggregation, bubble formation, and laser instability.

3.3.1.2 Transient HDVE

Transient HDVE spectra are taken using the Fourier-transform spectral interferometry and phase-modulation spectral interferometry methods, in which the local oscillator (LO) is stepped from 0 to 25 fs in 5 fs steps from the nonlinear signal. Again, the data is collected in three loops (Figure 3.10b): (1) chopper phase, (2) LO timing, and (3) T-jump delay τ. 10,000 shots were collected for both chopper phase set to 0° and 180°, after which τ,LO is stepped to the next value. Finally, the T-jump delay τ is changed. Although t-2D IR spectra provide the most spectral information, t-HDVE can be acquired much faster and therefore we use it to get accurate kinetic traces. Normally, we take ≥ 8 averages of the complete t-HDVE data sets. This requires 8 avg*6 τ,LO*2 chopper phase*10,000 shots*1 ms = 16 min, for each T-jump delay. Therefore, a kinetic trace containing 30 T-jump delays takes 8 hours.

3.3.1.3 Correction for linear absorption

One technical difficulty encountered in T-jump experiments is that the linear absorption of the D₂O solvent is also temperature-dependent, which leads to peak distortions in the transient 2D IR spectrum. The peak distortion and the appearance of some spurious peaks can overshadow the spectral features that encode the desired chemical dynamics. Therefore, a method to correct for the linear absorption is necessary and is briefly outlined below, following Ref. [2]. The 6 μm IR pulses propagate through the sample and experience the complex refractive index n(ω) = n(ω) + iκ(ω), where n(ω) and τ(ω) are the real and imaginary part of the effective index that has been shaped by the T-jump. The electric field of these pulses after passing through the sample can be expressed as:

$$\tilde{E}_\omega(\omega) = \epsilon_\omega(\omega)e^{i\phi_\omega(\omega)}e^{i\tilde{n}(\omega)\omega d/c}$$  \hspace{1cm} (3.10)
where $e_\alpha$ is the amplitude, $\phi_\alpha$ is the phase, $l$ is the path length, and $c$ is the speed of light. The last term describes the propagation effect on the refractive index. The combined absorption of beams 1 and 2 can be approximated by a depletion term along the $\omega_1$ axis:

$$K(\omega) = \frac{1 - \exp\left[-2\kappa(\omega)\omega_1/c\right]}{2\kappa(\omega)\omega_1/c} \tag{3.11}$$

The combined absorption effect of beam 3 and the signal is $\exp[-\kappa(\omega_3)\omega_3/c]$. The LO also accumulates the same absorption term $\exp[-\kappa(\omega_3)\omega_3/c]$. As a consequence, the combined effect of distortion along $\omega_1$ by the depletion of beams 1 and 2, and along $\omega_3$ by the absorption of beams 3, signal, and LO leads to distortions in the 2D IR spectrum:

$$\tilde{S}_{obs,1,3}(\omega_1, \omega_3) \approx \tilde{S}_{true}(\omega_1, \omega_3) \times K(\omega_1) \times e^{-2\kappa(\omega_3)\omega_3/c} \tag{3.12}$$

The subscripts 1 and 3 indicate that the observed signal is distorted in both the $\omega_1$ and $\omega_3$ axis. So our goal is to correct for the last two terms in Equation (3.12). The signals detected by the dual array MCT detector in the chopped mode can be expressed as:

$$I^{*}_{U/L}(\omega) = \left|\tilde{E}_{LO}(\omega_3)\right|^2 = e^{2\kappa(\omega_3)\omega_3/c}$$

$$I^{*}_{U/L}(\omega) = \left|\tilde{E}_{Lo,1,3}(\omega_3) + \tilde{E}_{Lo}(\omega_3)e^{-i\omega_3t_{LO}}\right|^2$$

$$= e^{-2\kappa(\omega_3)\omega_3/c} \left\{ e^{2\kappa(\omega_3)\omega_3/c} + e^{2\kappa(\omega_3)\omega_3/c} + 2e_{LO}(\omega_3)e_{s,1}(\omega_3)\cos[\Delta\phi(\omega_3) + \tau_{LO}\omega_3] \right\} \tag{3.13}$$

where the close and open circles in the superscript represent blocked and unblocked, respectively, and $U$ and $L$ represent upper and lower stripes of the MCT array, respectively. $\Delta\phi$ and $\tau_{LO}$ are the phase difference and time delay between the signal and LO, respectively.

Using the balance detection scheme to remove the homodyne signal followed by shot normalization, we get:

$$\frac{I^*_U - I^*_L}{I^*_U + I^*_L} = \frac{e^{-2\kappa(\omega)\omega_1/c} \times 4e_{s,1}(\omega_3)e_{s,1}(\omega_3)\cos[\Delta\phi(\omega_3) + \tau_{LO}\omega_3]}{e^{-2\kappa(\omega)\omega_1/c} \times 2 \left[ e^{2\kappa(\omega_3)\omega_3/c} + e^{2\kappa(\omega_3)\omega_3/c} \right]} \tag{3.14}$$

$$\approx 2 e_{LO}(\omega_3)e_{s,1}(\omega_3)\cos[\Delta\phi(\omega_3) + \tau_{LO}\omega_3] / e^{2\kappa(\omega_3)\omega_3/c}$$
where the triangular bracket indicates averages over the data sets taken at the same T-jump delay. In the second step, we use the fact that \( \varepsilon_{\omega_i}^2 \) is small compared to \( \varepsilon_{\omega_i}^2 \) and therefore dropped. In the lab, the data is actually collected as follows:

\[
\tilde{S}_{\text{measured}}(\omega_3, \tau_1) = \left[ \frac{I_{l1} - I_{l2}}{I_{l1} + I_{l2}} \right] - \left[ \frac{I_{r1} - I_{r2}}{I_{r1} + I_{r2}} \right]
\]

The term in the second bracket is theoretically zero, but in practice it is useful to remove scatter, electrical noise, and residual intensity mismatch between the two stripes.

What we measure in the lab is proportional to the desired signal whose \( \omega_3 \) distortion is corrected, by a factor of \( \varepsilon_{\omega_3}^2 \) since:

\[
\tilde{S}_i(\omega_3, \tau_1) = 2\varepsilon_{\omega_3}^2(\omega_3)\varepsilon_{\omega_3}(\omega_3)\cos[\Delta\phi(\omega_3) + \tau_{\omega_i, \omega_3}] \times K(\omega_3)
\]

As a consequence, to correct for the \( \varepsilon_{\omega_3}^2(\omega_3) \) factor and absorption along \( \omega_3 \), we post-process the data using Matlab. \( \varepsilon_{\omega_3}^2(\omega_3) \) is obtained with the measured LO spectrum that goes through the sample, \( \tilde{E}_{\omega_3}(\omega_3) = \varepsilon_{\omega_3}^2(\omega_3)e^{-\omega_3/\omega_3} \), and an independently measured FTIR at the final temperature used in the T-jump experiment,

\[
\tilde{S}_{\text{FTIR}} = -\log T = -\log \left( e^{-2\kappa\omega_3/c} \right) = \frac{2\kappa\omega_3}{c\ln10}
\]

Finally \( K(\omega_3) \) can also be calculated from the FTIR spectrum and divided out along \( \omega_3 \) axis.

### 3.3.2 Interpretation of the transient data

#### 3.3.2.1 Spectral features in t-2D IR

Throughout the entire thesis, equilibrium 2D IR spectra are plotted with positive peaks in yellow/red and negative peaks in blue. Since we measure the difference spectra in transient experiments, we use a different color gradient to plot t-2D IR spectra. As illustrated by cartoon spectra in Figure 3.11a, a loss of population appears in t-2D IR difference spectra as a negative (blue) peak on the top with a positive (red) peak on the bottom. Vice versa, a gain of population has a positive (red) peak on the top and a negative (blue) peak on the bottom.
Depending on the thermal response of the sample being studied, t-2D IR can exhibit complicated peak patterns that make interpretation challenging. This is also why it is important to correct for the linear absorption which can result in artifact peaks to further complicate the data interpretation (normally we get very small transient signal on the order of < 10% of the equilibrium signal).

Figure 3.11: (a) Cartoon showing the different peaks observed in equilibrium and transient experiments. (b) Simulated 2D IR spectra to illustrate the spectral features seen in t-2D IR spectrum. In this simulation, both peaks are blueshifted upon “T-jump” as shown by the diagonal slices plotted on the top.

Figure 3.11b shows one simulation example to illustrate some t-2D IR spectral features. In this simulation, the initial state at low temperature has two peaks: one at 1550 cm\(^{-1}\) which is inhomogeneously broadened along the diagonal and one at 1640 cm\(^{-1}\) which is more homogeneous. Upon raising temperature, both peaks blueshift (more easily seen from the diagonal slices), which shows up in t-2D IR as one gain feature at high frequency and one loss feature at low frequency. Similarly, the cross-peaks between the two modes show spectral shifts. The uphill cross-peak (\(\omega_1 < \omega_3\)) appears to shift to the blue along the \(\omega_1\) axis, while the downhill cross-peak (\(\omega_1 > \omega_3\)) appears to blueshift along the \(\omega_3\) axis. This is due to the larger frequency shift of the low frequency mode (otherwise the cross-peaks should also be shifting along the diagonal direction). Further discussion of t-2D IR spectral features can be found in the work of Cervetto et al.\(^{23}\)
3.3.2.2. Time evolution of transient data

The fast kinetics such as tautomeration or folding dynamics can be obtained by tracking the time evolution of specific spectral features in t-2D IR or t-HDVE. The observed time-dependence of the transient signal $S$ is the convolution of the sample response $R$ and the temperature profile $\Delta T$. That is:

$$S(\tau) = R(\tau) \otimes \Delta T(\tau) \quad (3.18)$$

Figure 3.12a shows one example of simulated T-jump response. The response of the sample to the T-jump takes the form $R(\tau) = \exp[-\tau / \lambda]$ with $\lambda = 150 \mu s$, which is a reasonable timescale for the unfolding of small DNA double helix. This sample response is convoluted with the solvent temperature which relaxes on the ms timescale to generate the observed transient signal. The transient signal rises initially due to the DNA's response (unfolding) to the T-jump, and then it falls due to the temperature re-equilibration.

It is a common practice to directly fit the rise of the observed transient signal to some functional form, for instance a single exponential, and obtain the time constant of the sample response. Figure 3.12b shows such fit which gives a time constant of 113 $\mu s$. The discrepancy comes from the fact that the timescale of the sample response and the temperature profile are too similar, and the subsequent error in fitting. Therefore, in order to get the actual response of the sample, one has to perform deconvolution:
\[ R(\tau) = \mathcal{FT}^{-1} \left\{ \frac{\mathcal{FT}[S(\tau)]}{\mathcal{FT}[\Delta T(\tau)]} \right\} \]  

(3.19)

In theory, deconvolution based on Fourier transformation should accurately recover the response from the sample (in the above example, the correct time of 150 \( \mu \)s is recovered); however, in practice, noisy data can be problematic in the division step. Furthermore, the time spacing needs to be constant throughout the many decades in time.

Fortunately, the empirical relation between the observed and correct time constants has been previously determined by the Tokmakoff group.\(^{21}\) It was found that when the sample response is fast (\( \lambda_{\text{obs}} < 50 \mu \)s), the measured rise time matches the input time constant. Between 50 \( \mu \)s to 3 ms, the measure rate is faster than expected (113 \( \mu \)s versus 150 \( \mu \)s in our case). At slower rates (\( \lambda_{\text{obs}} > 3 \) ms) where the temperature drops before the sample’s response, the rise time reflects the temperature profile while the fall time depends on the sample’s response. The dynamics studied in this thesis fall into the first two categories: (1) tautomerization (\( \lambda_{\text{obs}} < 50 \mu \)s) and (2) DNA duplex unfolding (50 \( \mu \)s < \( \lambda_{\text{obs}} < 3 \) ms). For the intermediate region, a polynomial fit was used to approximately map the observed time constant to the actual time constant:\(^{21}\)

\[ \lambda_{\text{wpul}} = \left( 2.166 \times 10^6 \right) \lambda_{\text{obs}}^3 + \left( -297.2 \right) \lambda_{\text{obs}}^2 + \left( 1.134 \right) \lambda_{\text{obs}} - 1.848 \times 10^{-6} \]  

(3.20)

3.4 Acknowledgments

I thank Kevin Jones for the improvements he has made on the T-jump 2D IR spectrometer and the procedure he has developed for correcting the linear absorption in the transient data. I thank Paul Stevenson and Ann Fitzpatrick for their careful reading and helpful comments on this chapter.
3.5 References


CHAPTER 4
Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy

The work presented in this chapter has been published in the following paper:


4.1 Introduction

Vibrational spectroscopy is a valuable technique to study nucleic acids because of its sensitivity to base pairing, sugar conformation, and glycosidic torsion angles. The vibrational IR and Raman spectra of nucleic acids and the general assignment of resonances have been known for decades and have been used to structurally characterize single-stranded, double-stranded DNA, and various RNA structural elements. One particularly useful IR spectral region is the fingerprint in-plane base vibrations in the 1500 -1800 cm⁻¹ region, such as C=O stretch, C=N stretch, and ND₂ bending modes. However, even with the extensive body of work that exists, a description of the vibrational motions assigned to different resonances, and an understanding of how these are influenced by interactions in secondary structures remain elusive. Traditional spectral assignments are based on simple local modes such as C=O, C=C, and C=N double bond stretches, whereas many computational studies suggest highly delocalized DNA vibrations. Progress towards the study of DNA/RNA conformational dynamics is limited by the ability to accurately assign the different vibrational bands. This is because most experiments are limited to linear FTIR or Raman spectroscopy where the details
of vibrational couplings are not readily accessible. On the other hand, two-dimensional infrared spectroscopy (2D IR) possesses the capability to characterize potential energy surfaces by obtaining the diagonal and off-diagonal anharmonicities and vibrational couplings through the analysis of cross-peaks. 2D IR experiments open an avenue to directly measure the vibrational couplings without the aid of isotope labeling, which can potentially alter the delocalized nature of these vibrations. Recent theoretical work on DNA bases and Watson-Crick base pairs showed that it is important to include anharmonic corrections for normal mode frequency calculations in order to reach reasonable agreement with the gas-phase IR measurement. Additionally, the ability to perform polarization selective experiments allows one to obtain the relative orientations of the transition dipoles, which is a sensitive measure of the vibrational mixing underlying the anharmonic modes of the system.

As a step toward understanding the spectral signatures of nucleotide interactions, we characterized the DNA and RNA vibrations of individual bases by acquiring polarization dependent 2D IR spectra of the purine and pyrimidine base vibrations of five nucleotide monophosphates (NMPs, structures shown in Figure 4.1). We focus on the 1500 –1700 cm\(^{-1}\) spectral region that contains in-plane base vibrations. The distinctive cross-peaks between the vibrational modes of NMPs, such as ring vibrations and C=O stretches, indicate that these vibrational modes are highly coupled anharmonic bond stretching vibrations. Through the analysis and modeling of the experimental 2D IR spectra containing the IR active in-plane vibrations, we have characterized the transition energies, vibrational anharmonicities and couplings, and the magnitude and relative orientations of the transition dipoles. To help interpret the molecular origins of these vibrational modes, we also performed density functional theory (DFT) calculations. The consistency between experiment and calculations indicates that multiple ring vibrations are delocalized over the purine or pyrimidine rings.
AMP GMP
R
UMP TMP CMP
 ND 2 0

\[ \text{ND} \]
\[ \text{IN} \]
\[ \text{IN} \]
\[ \text{IN} \]
\[ \text{IN} \]

Figure 4.1: Structure and numbering convention of the deuterated five nucleotide 5'-monophosphates investigated. R represents D-ribose-5-phosphate.

4.2 Results

The FTIR spectra of the five NMPs in D\textsubscript{2}O in the 1500–1700 cm\textsuperscript{-1} region are presented in Figure 4.2. In this spectral region, these bands mainly originate from in-plane double bond vibrations, such as C=O, C=N, and C=C stretch, and the ND\textsubscript{2} bend. The vibrational frequencies, intensities, and linewidths are sensitive to solvation and hydrogen to deuterium exchange.\textsuperscript{15} Because an assignment convention for these modes has not been established, in this paper we will label the peaks under study with the first letter of the nucleotides, associated with a numbering scheme that goes from higher frequency to lower frequency (Figure 4.2).

Nucleotides having carbonyl groups have IR peaks in the 1650–1700 cm\textsuperscript{-1} region. These peaks have been traditionally assigned to local C=O stretches: 1663 cm\textsuperscript{-1} for C\textsuperscript{6}=O in GMP, 1651 cm\textsuperscript{-1} for C\textsuperscript{2}=O in CMP, 1691 cm\textsuperscript{-1} for C\textsuperscript{2}=O and 1654 cm\textsuperscript{-1} for C\textsuperscript{4}=O in UMP, and 1690 cm\textsuperscript{-1} for C\textsuperscript{2}=O and 1661 cm\textsuperscript{-1} for C\textsuperscript{4}=O in TMP. These peaks in general are more intense due to the strong dipole moment of carbonyl stretches. Furthermore, they have broader linewidths (11-13 cm\textsuperscript{-1} FWHM) compared to the other lower frequency peaks, due to water hydrogen-bonding to the carbonyls that induce frequency shifts of these oscillators. Lower frequency modes have been assigned to be the ring vibrations consisting of C=C and C=N double bonds.\textsuperscript{4}

Although there are similar chemical structures, the FTIR spectra of the NMPs in the spectral region of the ring vibrations are in fact quite distinct. No simple pattern exists that determines the frequency, intensity and linewidth of a particular type of vibration, leading to
a number of questions. For example, within the pyrimidines, why is there a drastic intensity variation for the two similar local C=O stretches in UMP and TMP, and why do these differ dramatically in the number and position of peaks compared to CMP? Why do the purines, AMP and GMP, have such dramatic differences in the number of resonances, their frequency, and their lineshapes?

Figure 4.2: FTIR spectra of the five NMPs in PBS buffer at 20°C and pH 6.9. The D₂O background has been subtracted. The peak labels used in this paper are indicated for the various base vibrations, and their peak frequencies can be found in Table 4.1.
4.2.1 2D IR Spectra of Purines

4.2.1.A AMP

2D IR spectra of AMP are shown in Figure 4.3a and b. The infrared spectra of AMP display one intense peak at 1625 cm\(^{-1}\) and a weak peak at 1578 cm\(^{-1}\), denoted as A1 and A2 respectively, which are assigned to pyrimidine ring modes.\(^{13,21}\) 2D IR resonances display positive and negative peaks that arise from the 0-1 and 1-2 transitions, respectively, so that the splitting of the doublet along the \(\omega_3\) axis gives a measure of the diagonal anharmonicity. The presence of cross-peaks indicates two anharmonically coupled vibrational modes, as would be expected for two delocalized modes of a conjugated ring system. The overall eight-peaks in the spectrum follows a classic pattern that emerges for transitions between the ground singly and doubly excited eigenstates of a pair of coupled vibrations.\(^{22}\) The variation in cross-peak amplitude between parallel (ZZZZ) and perpendicular (ZZYY) polarized spectra depends on the projection angle between the two transition dipole moments. A significant enhancement of the cross-peak amplitude relative to the diagonal peaks is observed for AMP, indicating that the projection angle is close or equal to 90 degrees.\(^{23}\) Both A1 and A2 and their cross-peaks appear homogeneously broadened, as seen from the diamond-like lineshapes, with diagonal linewidths (FWHM) of 11 and 9.4 cm\(^{-1}\), respectively. The comparison of diagonal linewidth (\(\sigma\)), anti-diagonal linewidth (\(\Gamma\)), and ellipticity,\((\sigma^2 - \Gamma^2) / (\sigma^2 + \Gamma^2)\), for the five NMPs is shown in the Figure 4.5.
Figure 4.3: Experimental ZZZZ (top) and ZZYY (bottom) 2D IR spectra of the five NMPs (a,b: AMP; c,d: GMP; e,f: UMP; g,h: TMP; i,j: CMP) at 20mg/ml in PBS buffer at 20°C and pH 6.9. The experimental FTIR spectra have been plotted on top of the corresponding 2D IR spectra for comparison. All contours are plotted in 7.5% intervals.
Figure 4.4: Simulated ZZZZ (top) and ZZYY (bottom) 2D IR spectra of the five NMPs (a,b: AMP; c,d: GMP; e,f: UMP; g,h: TMP; i,j: CMP) from 2D spectral fitting. The fitting parameters are summarized in Table 1, Table 2, and SI. The simulated FTIR spectra using the lineshape function obtained from the 2D spectral fitting have been plotted on top of the corresponding 2D IR spectra for comparison. All contours are plotted in 7.5% intervals.
To obtain information on the anharmonic nuclear potential and projection angles of the transition dipole moments in these systems, the 2D IR ZZZZ and ZZYY spectra were fit simultaneously with a least-squared fitting algorithm using a nonlinear response function formalism described in Chapter 2. Calculated 2D IR spectra using these parameters are shown in Figure 4.4. The important parameters in describing the potential surfaces such as transition frequencies, diagonal and off-diagonal anharmonicities, as well as the angles between transition dipoles are tabulated in Table 4.1 and Table 4.2 for all nucleotides. A full list of the fitting parameters including the lineshape parameters and dipole strengths for higher order transitions is included in Appendix 4.A. The extracted diagonal anharmonicities for AMP are $\Delta \omega_{d1} = 8.0$ cm$^{-1}$ and $\Delta \omega_{d2} = 9.8$ cm$^{-1}$. The experimental results can be compared with the anharmonicities obtained from DFT calculations in the gas phase by Cho and coworkers: $\Delta \omega_{d1} = 7.0$ cm$^{-1}$ and $\Delta \omega_{d2} = 0.8$ cm$^{-1}$. The off-diagonal anharmonicity ($\Delta \omega_{d1,d2} = 3.7$ cm$^{-1}$) is significantly less, suggesting that the system can be modeled as two weakly coupled normal modes. The coupling strength between the two oscillators can therefore be calculated perturbatively to be 15 cm$^{-1}$, compared to the energy splitting of 47 cm$^{-1}$. From fitting the ZZZZ and ZZYY spectra simultaneously, the projection angle between $\mu_{d1}$ and $\mu_{d2}$ was determined to be $90^\circ \pm 2^\circ$.

DFT calculations on deuterated adenosine in the gas phase observed two normal modes in the spectral region of interest at 1588 cm$^{-1}$ and 1565 cm$^{-1}$. FTIR stick spectra for all five NMPs simulated from the calculated IR intensities are given in Figure 4.6. The 1588 cm$^{-1}$ band (A1) is a pyrimidine ring mode consisting mostly of the $\text{C}^4=\text{C}^5$, $\text{C}^5-\text{C}^6$ out-of-phase stretch. More complete description of the vibration's composition can be found in Table 4.1. The 1565 cm$^{-1}$ band (A2) contains the ring vibrations of both the imidazole and pyrimidine rings. Specifically, this highly delocalized ring deformation mode involves the $\text{C}^4=\text{C}^5$, $\text{C}^5-\text{C}^6$ in-phase stretch. The angle between the two transition dipole moment was found to be 101° (shown in Figure 4.7), in good agreement with the experiment. Therefore, the qualitative agreement of IR intensities, symmetry, and transition dipole angle with the experimental results assure us of the interpretation of these vibrational modes.
Table 4.1. Characterization for the five deuterated NMPs in D2O in experiments and DFT calculations. Fundamental transition energies, $\omega_i$, diagonal anharmonicities ($\Delta \omega_{ij}$, cm$^{-1}$) and the 0-1 transition dipole moments $\mu_{1,0}$.

<table>
<thead>
<tr>
<th>Nucleotide Peak</th>
<th>$\omega_i$ (cm$^{-1}$)</th>
<th>$\Delta \omega_{ij}$ (cm$^{-1}$)</th>
<th>$\mu_{1,0}$ (cm$^{-1}$)</th>
<th>$\omega_j$ (cm$^{-1}$)</th>
<th>Vibrations $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Fit)</td>
<td>(Fit)</td>
<td>(Fit)$^a$</td>
<td>(Calc'd)$^b$</td>
<td></td>
</tr>
<tr>
<td><strong>AMP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>1625</td>
<td>8.0</td>
<td>1.0</td>
<td>1588</td>
<td>$\nu (C^1=C^3, C^5-C^6$ out-of-phase), $\delta (C^2-H)$, $\delta (N^4D_2)$, $P_y$</td>
</tr>
<tr>
<td>A2</td>
<td>1578</td>
<td>9.8</td>
<td>0.41</td>
<td>1565</td>
<td>$\nu (C^1=C^3, C^5-C^6$ in phase), $\nu (N^1-C^6)$, $\nu (N^3-C^4)$, $\nu (N^7-C^8)$, $\delta (C^6-H)$, $P_y + Im$</td>
</tr>
<tr>
<td><strong>GMP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>1665</td>
<td>11</td>
<td>1.0</td>
<td>1692</td>
<td>$\nu (C^6-O)$, $\delta (N^1-D)$, $P_y$</td>
</tr>
<tr>
<td>G2</td>
<td>1579</td>
<td>4.0</td>
<td>0.63</td>
<td>1556</td>
<td>$\nu (C^2=N^4)$, $\nu (C^5-O)$, $\delta (N^1-D)$, $\delta (N^2D_2)$, $P_y$</td>
</tr>
<tr>
<td>G3</td>
<td>1565</td>
<td>7.8</td>
<td>0.62</td>
<td>1533</td>
<td>$\nu (C^2=N^1C^4=C^5)$, $\delta (C^6-H)$, $P_y + Im$</td>
</tr>
<tr>
<td>G4</td>
<td>1539</td>
<td>18</td>
<td>0.29</td>
<td>1511</td>
<td>$\nu (C^2=C^5)$, $\nu (N^7=C^9)$, $\delta (C^4-H)$, $P_y + Im$</td>
</tr>
<tr>
<td><strong>UMP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>1693</td>
<td>16</td>
<td>0.76</td>
<td>1681</td>
<td>$\nu (\text{sym } C=O)$, $\nu (C^1=C^5)$, $\delta (N^1-D)$, $\delta (C^3-H)$</td>
</tr>
<tr>
<td>U2</td>
<td>1655</td>
<td>18</td>
<td>1.0</td>
<td>1670</td>
<td>$\nu (\text{asym } C=O)$, $\nu (C^2=C^6)$, $\delta (C^3-H)$, $\delta (C^6-H)$, $\delta (N^1-D)$</td>
</tr>
<tr>
<td>U3</td>
<td>1617</td>
<td>25</td>
<td>0.46</td>
<td>1615</td>
<td>$\nu (C^3=C^6)$, $\nu (\text{asym } C=O)$, $\delta (C^3-H)$, $\delta (C^6-H)$, $\delta (N^1-D)$</td>
</tr>
<tr>
<td><strong>TMP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1690</td>
<td>11</td>
<td>0.74</td>
<td>1677</td>
<td>$\nu (C^2=O)$, $\nu (C^1=O)$, $\delta (N^1-D)$</td>
</tr>
<tr>
<td>T2</td>
<td>1663</td>
<td>5.0</td>
<td>1.0</td>
<td>1657</td>
<td>$\nu (C^2=O)$, $\nu (C^3=C^9)$, $\delta (N^1-D)$, $\delta (C^3H_3)$, $\delta (C^6-H)$</td>
</tr>
<tr>
<td>T3</td>
<td>1629</td>
<td>10</td>
<td>0.94</td>
<td>1627</td>
<td>$\nu (C^3=C^5)$, $\nu (C^1=O)$, $\delta (C^3H_3)$, $\delta (C^6-H)$</td>
</tr>
<tr>
<td><strong>CMP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>1651</td>
<td>9.2</td>
<td>1.0</td>
<td>1690</td>
<td>$\nu (C^2=O)$, $\nu (N^1=C^9)$, $\delta (C^6-H)$</td>
</tr>
<tr>
<td>C2</td>
<td>1614</td>
<td>15</td>
<td>0.55</td>
<td>1624</td>
<td>$\nu (N^3=C^1-C^3=C^6)$, $\nu (C^2=O)$, $\delta (C^3-H)$, $\delta (C^6-H)$</td>
</tr>
<tr>
<td>C3</td>
<td>1583</td>
<td>17</td>
<td>0.21</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>1524</td>
<td>6.1</td>
<td>0.57</td>
<td>1503</td>
<td>$\nu (C^1=C^5)$, $\nu (N^1-C^9)$, $\delta (C^3-H)$, $\delta (C^6-H)$, $\delta (N^4D_2)$</td>
</tr>
<tr>
<td>C5</td>
<td>1504</td>
<td>8.9</td>
<td>0.94</td>
<td>1483</td>
<td>$\nu (N^1=C^1-N^3)$, $\nu (C^3=C^9)$, $\delta (C^3-H)$, $\delta (C^6-H)$, $\delta (N^4D_2)$</td>
</tr>
</tbody>
</table>

$^a$ The dipole strengths reported here are from fitting the FTIR spectra using lineshape parameters from the 2D fitting. Peak amplitudes directly from 2D fitting are reported in Appendix 4.A.

$^b$ These vibrational frequencies obtained from DFT calculations have been scaled by 0.9614.

$^c$ Listed here are the decomposed vibrations with larger amplitudes obtained from the DFT calculations. $\nu$: stretching; $\delta$: bending; $P_y$: pyrimidine ring vibration; $Im$: imidazole ring vibration.
Table 4.2. Couplings between base vibrations for the five deuterated NMPs in D$_2$O. Off-diagonal anharmonicities ($\Delta \omega_{i,j}$, cm$^{-1}$). $\Theta$'s are projection angles obtained from fitting the experimental 2D IR spectra, and $\theta$'s are the angles obtained from DFT calculations.

<table>
<thead>
<tr>
<th>Nucleotides</th>
<th>Coupled Modes</th>
<th>$\Delta \omega_{i,j}$ (cm$^{-1}$)</th>
<th>$\Theta$ (deg)</th>
<th>$\theta$ (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMP</td>
<td>A1 &amp; A2</td>
<td>3.7</td>
<td>90 (±2)</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>G1 &amp; G2</td>
<td>1.3</td>
<td>89 (+1/-10)</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>G1 &amp; G3</td>
<td>4.6</td>
<td>111 (+11/-12)</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>G1 &amp; G4</td>
<td>3.3</td>
<td>37 (+16/-37)</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>G2 &amp; G3</td>
<td>7.5</td>
<td>21 (±1)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>G2 &amp; G4</td>
<td>7.0</td>
<td>121 (+8/-7)</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>G3 &amp; G4</td>
<td>6.8</td>
<td>134 (+11/-8)</td>
<td>134</td>
</tr>
<tr>
<td>GMP</td>
<td>U1 &amp; U2</td>
<td>8.0</td>
<td>90 (±7)</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>U1 &amp; U3</td>
<td>25</td>
<td>35 (+5/-6)</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>U2 &amp; U3</td>
<td>15</td>
<td>100 (+1/-2)</td>
<td>101</td>
</tr>
<tr>
<td>UMP</td>
<td>T1 &amp; T2</td>
<td>1.7</td>
<td>117 (±2)</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>T1 &amp; T3</td>
<td>4.9</td>
<td>34 (±3)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>T2 &amp; T3</td>
<td>6.2</td>
<td>144 (±1)</td>
<td>135</td>
</tr>
<tr>
<td>TMP</td>
<td>C1 &amp; C2</td>
<td>7.2</td>
<td>121 (±1)</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>C1 &amp; C3</td>
<td>7.4</td>
<td>108 (±2)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>C1 &amp; C4</td>
<td>1.6</td>
<td>93 (+4/-3)</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>C1 &amp; C5</td>
<td>5.2</td>
<td>16 (+6/-16)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>C2 &amp; C3</td>
<td>12</td>
<td>23 (+7/-13)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>C2 &amp; C4</td>
<td>14</td>
<td>134 (±1)</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>C2 &amp; C5</td>
<td>7.6</td>
<td>144 (±2)$^*$</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>C3 &amp; C4</td>
<td>13</td>
<td>152 (+5/-4)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>C3 &amp; C5</td>
<td>8.2</td>
<td>124 (±2)$^*$</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>C4 &amp; C5</td>
<td>1.2</td>
<td>71 (±1)</td>
<td>84</td>
</tr>
</tbody>
</table>

$^*$ The error bars for these cross-angles were determined from the negative cross-peaks because their positive cross-peaks are interfered too heavily by nearby peaks to be clearly identified. The equation relating the relative cross-peak amplitude to the dipole angle is shown in Appendix 4.A.
Figure 4.5: Lineshape analysis of the experimental ZZZZ 2D IR spectra of the five NMP's shown in Figure 4.3. (a) The diagonal FWHM (σ) are obtained from fitting the diagonal slices of the 2D spectra to Gaussian functions. (b) The anti-diagonal FWHM (Γ) are determined directly from the anti-diagonal slices that cut through the peak maxima of the diagonal peaks. (c) The ellipticities are calculated from the diagonal and anti-diagonal widths, \((σ^2 - Γ^2) / (σ^2 + Γ^2)\).

4.2.1.B GMP

Infrared spectra of GMP between 1500–1800 cm\(^{-1}\) display four vibrational resonances. The intense peak at 1663 cm\(^{-1}\) (G1) is widely accepted as a local C=O stretch. Its 2D lineshape is elongated along the diagonal with an unusually broad inhomogeneous linewidth of 33 cm\(^{-1}\). The broadening of G1 is not surprising, given our understanding of how hydrogen bonding to carbonyls influences frequency shifts and line broadening. Notably, the similar broadening of the amide I band of N-methylacetamide in D\(_2\)O has been explained as containing both singly and doubly hydrogen-bonded configurations of water to the carbonyl. Therefore, given that multiple water sites have been observed from crystal structures of DNA, the inhomogeneity is most likely the result of various hydrogen-bonding configurations around the carbonyl.
FTIR shows a second peak G2 at 1578 cm\(^{-1}\), along with a shoulder peak G3 at 1567 cm\(^{-1}\). The presence of this shoulder peak is clearly identified by the pronounced cross-peaks between G2 and G3. The positive downhill cross-peak destructively interferes with the overtone of the G2 diagonal peak, resulting in a special peak pattern (contrasted to the eight-peak pattern for AMP) for two coupled oscillators that are separated by only a small energy gap. Additionally, the lack of relative cross-peak enhancement from the polarization-selective 2D IR spectra suggests a pair of parallel transition dipoles. On the other hand, the cross-peaks of these two modes to G1 are greatly intensified in the ZZYY spectrum, indicating a perpendicular geometry between G1 and these two vibrations. Finally, a fourth peak G4 at 1539 cm\(^{-1}\) that is barely seen in the FTIR, is detected by its cross-peaks to the higher frequency modes in the 2D spectrum. However, due to the low intensity of peak G4, the cross-peak enhancement is difficult to determine and its relative dipole orientation can only be extracted from fitting the spectra. The lineshapes of G2-G4 are homogeneous with diagonal linewidths of \(\sim 13\) cm\(^{-1}\), as in the case of the ring modes in AMP.

Figure 4.6: Calculated FTIR spectra using DFT calculated normal mode frequencies and transition dipole strengths of the five deuterated nucleosides: adenosine (a), guanosine (b), uridine (c), thymidine (d), and cytidine (e). The vibrational frequencies have been scaled by 0.9614. All peak lineshapes are generated by convoluting stick spectra with a Lorenztian function with linewidth \(\Gamma = 2\) cm\(^{-1}\).
Previous assignments for these three peaks have been inconsistent: Tsuboi and Mondragon-Sanchez \textit{et al.}\textsuperscript{7} assigned them to be local C\textsuperscript{4}=C\textsuperscript{5} stretch and different C=N stretches; Toyama \textit{et al.}\textsuperscript{21} performed isotope labeling experiments to assign G2 and G3 to be pyrimidine ring vibrations localized to N\textsuperscript{3} and C\textsuperscript{2}, respectively, and G4 to be a delocalized vibration over the entire purine ring; while DFT calculations showed delocalized and coupled vibrations across the purine ring for these modes.\textsuperscript{31} By fitting the GMP ZZZZ and ZZYY 2D IR spectra, the diagonal anharmonicities were obtained as: $\Delta \omega_{G1} = 11\ \text{cm}^{-1}$, $\Delta \omega_{G2} = 4.0\ \text{cm}^{-1}$, $\Delta \omega_{G3} = 7.8\ \text{cm}^{-1}$, and $\Delta \omega_{G4} = 18\ \text{cm}^{-1}$. Krummel and Zanni measured $\Delta \omega_{G1}$ to be 14 cm\textsuperscript{-1} by fitting the pump-probe spectrum of dGMP.\textsuperscript{31} The experimental values are in reasonable agreement with DFT predicted values. Wang and coworkers\textsuperscript{20} obtained $\Delta \omega_{G4} = 19\ \text{cm}^{-1}$ for guanine, and Lee and Cho\textsuperscript{17} calculated diagonal anharmonicities for C\textsuperscript{6}=O stretch (16.8 cm\textsuperscript{-1}) and ring vibration (13.0 cm\textsuperscript{-1}) for 9-methylated guanine-d3. The off-diagonal anharmonicities were found to be of nearly equal magnitude to the diagonal anharmonicities.

The relative angles between the four transition dipole moments obtained from fitting, along with comparison to DFT results, are presented in Table 4.2. The constraint that all four transition dipoles are in-plane was applied to match the six fitted angles into a simple geometry, remembering that solutions to the equation: \[ \cos \Theta = \sqrt{(7-6\alpha)(1+2\alpha)} \] can adopt the value of either \( \Theta \) or \( 180^\circ - \Theta \). It was found that $\mu_{G2}$ and $\mu_{G3}$ are close to being perpendicular with $\Theta_{G1,G2} = 89^\circ$, $\Theta_{G1,G3} = 111^\circ$, while $\mu_{G4}$ is 37° away from $\mu_{G1}$. The uncertainties in these angles are $\sim 10^\circ$.
Figure 4.7: Transition dipole orientations (orange arrows) of the normal modes from DFT calculations on the five deuterated nucleosides. Shown here are the optimized structures with different number of explicit D$_2$O molecules making hydrogen-bonds to the carbonyl or ND groups on the bases. $\theta$'s define the calculated angles between different transition dipole moments. Only the angles between the nearest transition dipoles are shown, see Table 4.2 for a complete list of angles. The ribose is made less opaque for clearer viewing of the bases. Mode C3 is not seen in our DFT calculation and therefore its dipole orientation (dotted arrow) is deduced from its relative angles to other modes ($\Theta_{23}, \Theta_{13}$) based on 2D spectral fitting (Table 4.2).

DFT calculations on deuterated guanosine were performed to get a picture of the vibrations and help interpret these fitted angles. It was found that the number and location of the added water molecules in DFT calculations alters the directions of these transition dipoles, although the bond stretches involved did not change dramatically (shown in Figure 4.8 and Figure 4.9). Therefore, care is needed when interpreting the 2D IR results using DFT calculations. When at least one water molecule is hydrogen-bonded to the carbonyl, the
agreement between the experimentally determined angles and DFT calculated results is good, even for the projection-angles involving the low intensity peak G4. Furthermore, from the calculation it was observed that by solvating guanosine with more D$_2$O molecules, the frequency of G1 redshifts while the frequencies of G2, G3, and G4 blueshift. This anti-correlated frequency shift can be rationalized through the stabilization of resonance structures which lengthen the carbonyl bond while shortening the CN bond, analogous to protein Amide I' and Amide II' modes.$^{32,33}$

The calculated frequencies and intensities can be easily matched to the experimental results, except for the fact that the G3 mode is weaker than the G4 mode. G1 was found to contain mainly C$^\beta$=O stretch and its transition dipole aligns roughly with the CO bond, indicating a more local vibration character. G2 has larger amplitude of delocalized pyrimidine ring vibration. G3 mode is a more delocalized ring vibration that extends to the imidazole ring. G4 is also a delocalized vibration across the entire purine ring, with strong C$^4$=C$^5$ and N$^7$=C$^8$ stretches.

4.2.2 2D IR Spectroscopy of Pyrimidines

4.2.2.A UMP

UMP has a more congested spectrum compared to those of the purines with three modes close in energy: one intense peak at 1654 cm$^{-1}$ (U2), a medium peak at 1691 cm$^{-1}$ (U1) and a weak peak at 1617 cm$^{-1}$ (U3). Previously, U3 was attributed to the C$^5$=C$^6$ stretch where U1 and U2 have been assigned to the local C$^2$=O and C$^4$=O, respectively.$^{14,15}$ However, 2D IR spectra of UMP show couplings between all three modes, revealing some degree of delocalization. In addition, the distinct intensities and lineshapes of U1 and U2 suggest a more complicated interpretation than a pure local mode scheme. Their anti-diagonal linewidths are 16 cm$^{-1}$ and 13 cm$^{-1}$ for U1 and U2, respectively, which fall in the typical range (10-15 cm$^{-1}$) for homogeneous linewidths of carbonyls in aqueous solution. On the other hand, the diagonal linewidths are 32 cm$^{-1}$ and 17 cm$^{-1}$ for U1 and U2 respectively. As a consequence, U1 is an inhomogeneous peak whereas U2 appears homogeneous. Spectral fitting resulted in a $90^\circ \pm 7^\circ$ angle between the two vibrations, which suggests the two modes are close to being the symmetric and asymmetric C=O stretches. In fact, this observation parallels with our DFT calculation on uracil.

DFT calculations showed that U1 is the symmetric C=O stretch mixed with N$^1$-D bend. U2 was the asymmetric C=O stretch mixed with some C$^3$=C$^6$. Finally, U3 was C$^5$=C$^6$ stretching in-phase with C$^2$=O and out-of-phase with C$^4$=O. The mode composition of these
Figure 4.8: Vibrational normal modes G1 – G4 of guanosine as a function of the number of solvating D₂O molecules: (a) no D₂O; (b) one D₂O; (c) three D₂O; and (d) seven D₂O molecules. The frequencies have been scaled by a factor of 0.9614 and the normalized IR intensities to the most intense mode are given in parentheses.
vibrations does not change dramatically with the number of solvating water molecules; however, the vibrational frequencies and the relative dipole orientations do. Specifically, three water molecules were required to be present in the proximity of the C=O’s and ND group in order to obtain an agreement with the experimentally measured relative dipole orientations.

We determined the degree of correlation between the spectral broadening of the resonances for UMP. The correlated shift of the transition frequencies \(\delta\omega_i\), and \(\delta\omega_j\) between two distributions is quantified by the correlation coefficient \(\rho_{ij} = (\delta\omega_i, \delta\omega_j)/(\sigma_i, \sigma_j)\) where \(\sigma_i\) and \(\sigma_j\) are the corresponding distribution widths.\(^{23}\) Although the correlations can be expressly addressed in Kubo models of the cross-peak lineshape,\(^{23,34}\) the correlation coefficient can also be more simply related to the tilt angle of the cross peak nodal line, \(\Psi\), through \(\rho_{ij} = \tan(\Psi) = (A_R - A_{NR})/(A_R + A_{NR})\), where \(A_R\) and \(A_{NR}\) are the cross-peak amplitudes in rephasing and non-rephasing spectra, respectively.\(^{34,35}\) High degree of correlation and anti-correlation correspond to \(\rho_{ij}\) of +1 and -1, respectively, where no correlation results in \(\rho_{ij} = 0\).\(^{36,37}\)

Based on the results shown in Table 4.3, U3 is well correlated with the other vibrations, whereas U1 and U2 are not. Highly correlated broadenings indicate that the vibrational couplings are large compared to the disorder,\(^{34}\) as illustrated by the cross-peak between U2 and U3 \((\omega_i = 1617 \text{ cm}^{-1}, \omega_j = 1654 \text{ cm}^{-1})\), where the tilt of the nodal line is nearly parallel to the diagonal (39°). This is an indication that both modes have character of the conjugated \(C^4=O\) and \(C^5=C^6\) stretches, where weakening/strengthening of one bond is
correlated to the other in this conjugated polyene system. This observation again matches the vibrational modes seen in our DFT calculations and other normal mode analysis on uracil.\textsuperscript{5,38} However, it contrasts with some \textit{ab initio} studies which argued that U3 is the C\textsuperscript{5}=C\textsuperscript{6} stretch while U1 and U2 are coupled C=O stretches with different participation ratio of the two carbylons.\textsuperscript{15,19}

Table 4.3. Correlation coefficients, $\rho$, and slope of the cross-peak nodal line, $\Psi$, for UMP and TMP. $\rho$ varies from 1 for correlated to -1 for anticorrelated broadening, whereas $\Psi$ varies from 45° to -45° over the same range.

<table>
<thead>
<tr>
<th></th>
<th>UMP</th>
<th>$\rho$</th>
<th>$\Psi$ (deg)</th>
<th>TMP</th>
<th>$\rho$</th>
<th>$\Psi$ (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1 &amp; U2</td>
<td>0.20</td>
<td>11</td>
<td></td>
<td>T1 &amp; T2</td>
<td>0.29</td>
<td>16</td>
</tr>
<tr>
<td>U1 &amp; U3</td>
<td>0.65</td>
<td>33</td>
<td></td>
<td>T1 &amp; T3</td>
<td>0.28</td>
<td>15</td>
</tr>
<tr>
<td>U2 &amp; U3</td>
<td>0.81</td>
<td>39</td>
<td></td>
<td>T2 &amp; T3</td>
<td>0.98</td>
<td>44</td>
</tr>
</tbody>
</table>

4.2.2.B TMP

TMP spectra have three peaks with similar positions to those of UMP; however, the relative intensities change dramatically upon the methyl insertion, indicating that the TMP modes are significantly different from the UMP modes. In particular, the 1628 cm\textsuperscript{-1} peak (T3) is much more intense than U3, indicating a lower symmetry of the vibrational mode involving C\textsuperscript{5}=C\textsuperscript{6} stretch. Similar to UMP, prior studies used the local mode picture and assigned the 1690 cm\textsuperscript{-1} peak (T1) to $C^2$=O stretch, 1661 cm\textsuperscript{-1} peak (T2) to $C^4$=O, and T3 to $C^5$=C\textsuperscript{6} stretch.\textsuperscript{1,5,13,15} However, Zhang \textit{et al.} suggested that these peaks may be due to mixed vibrations.\textsuperscript{38}

2D IR spectra of TMP reveal couplings between all three modes ($\leq$6 cm\textsuperscript{-1}) that are less than the diagonal anharmonicities, which is indicative of relatively weak couplings between the modes. Spectral fitting reveals that the transition dipole moments between the carbonyl transitions T1 and T2 are 117° (+8/-7) away from each other, which is also suggestive of a local mode assignment for the carbonyl bands. This can be compared to an angle of 133° obtained from our DFT calculation on thymidine.

Interestingly, the TMP lineshapes are very different from those of UMP, most notably the broad overtone peak of T3 compared to U3. Although T1 is still inhomogeneous and T2 is homogeneous, their anti-diagonal linewidths are larger than those of U1 and U2 (Figure 4.5). The correlation coefficients of the three pairs of TMP cross-peaks were calculated (Table 4.3).
The result is similar to UMP in that T1 and T2 are uncorrelated while T2 and T3 are highly correlated. However, U1 and U3 are more correlated compared to T1 and T3. This effect may be explained by the insertion of the methyl group that results in the loss of C5-H bend observed in U1, U2 and U3 modes.

Just as in our Dft calculations on uridine, three water molecules hydrogen bonded to thymidine’s C=O's and ND group were needed to reach a reasonable agreement with 2D IR results. The calculations reveal that the methyl substitution decreases the degree of coupling of the two carbonyl stretches. T1 was found to be mostly a C2=O stretch, accompanied by a small magnitude of in-phase vibration of C4=O, with its transition dipole aligned nearly parallel to the C2=O bond. T2 has contributions from in-phase stretching of C4=O and C5=C6. Finally, T3 is mostly C5=C6 stretch mixed with a little C=O stretch. The normal modes and angles from the DFT calculations match the experimental results reasonably well, and suggest that the coupled vibrations in TMP are less delocalized than the ones in UMP, due to the methyl substitution.

4.2.2. C CMP

The spectra of CMP display five peaks in the spectral region of interest. Structurally, CMP has one C=O bond, and only one high frequency vibrational mode attributable to a CO stretch is observed at 1651 cm⁻¹ (C1). From the 2D IR spectra, one can see that C1 is homogeneous, in sharp contrast to the inhomogeneous lineshapes of U1 and T1. There are two other homogeneous peaks that are close in energy at 1614 cm⁻¹ (C2) and 1583 cm⁻¹ (C3) with much lower intensities. Purrello and coworkers⁴⁹ assigned C2 in their UVRR spectrum of dCMP in D2O to the C5=C6 stretch. Miles⁴⁰ observed that C2 and C3 disappear upon methylation on the N3 position and dimethylation on the N4 position, respectively. This study indicates that C2 is mixed with N3=C4 while C3 involves motions of the ND2 group. The more intense cross-peaks between modes C1/C2 than C1/C3 in our 2D spectra support this peak assignment because C2=O, N3=C4, and C5=C6 form a conjugated polyene system. Based on the enhancement of the relative amplitudes of these cross-peaks between C1, C2, and C3, the transition dipole of C1 should deviate from those of the nearly parallel C2 and C3.

Additionally, there are two more modes at 1524 cm⁻¹ (C4) and 1504 cm⁻¹ (C5) which also have been assigned to in-plane ring vibrations.¹ These two peaks are much more intense than C2 and C3, which can be explained by the involvement of the exocyclic ND2 that breaks the symmetry of these ring vibrations. Interestingly, 2D IR spectra reveal that these two ring modes are coupled to the higher frequency modes C1-C3, which explain why the observed frequencies are much lower than expected for a standard C=N. In order to reduce the number
of free parameters for fitting the CMP 2D spectra, the diagonal peaks C4 and C5, as well as their influence on associated cross-peaks, were set to be purely homogeneous. The anharmonicity of C1 is relatively small compared to those of C2 and C3, but comparable to those of C4 and C5. A value of 9.2 cm⁻¹ for Δω₁ is consistent with the previously reported result of 9 cm⁻¹.³¹

From our DFT calculations of harmonic vibrational frequencies on deuterated cytidine, only four vibrations were observed in the spectral window of interest, rather than the five seen experimentally. Previous DFT calculations on single-stranded poly-C also reported only one medium intensity peak in the 1616-1624 cm⁻¹ region, which was assigned to an in-plane ring vibration.¹ Since the presence of C1-C5 is experimentally established, we believe that one of the experimental peaks may be a combination band or Fermi resonance. A DFT calculated transition at 1690 cm⁻¹, corresponding to the C1 mode, has dominant C²=O character. A second peak at 1624 cm⁻¹ involves the enimine bond N³=C⁴=C⁵=C⁶ stretches and C-H bend. Since these contributions match our description of the C2 mode based on experimental results, we assign this peak to C2 and conclude that the DFT calculation does not find the C3 mode which requires significant ND₂ motions. Gavira and coworkers ⁴¹ observed a medium band at 1616 cm⁻¹ and a very weak peak at 1582 cm⁻¹ in the FTIR spectrum of cytidine 3'-monophosphate (3'-CMP) and assigned them to be the ribose C-H vibrations and in-plane base vibrations, respectively. However, this interpretation does not explain the FTIR of cytosine (data not shown) where both peaks are present. The fact that the C3 band is barely visible in the FTIR is also suggestive of a combination band.

It was found that when cytidine makes a hydrogen bond to a water molecule through the carbonyl, the relative transition dipole orientations match well with the results obtained from the 2D spectral fitting. Although the C3 mode is not described by the harmonic mode DFT calculation, it was determined from the fitting that the transition dipoles of C2 and C3 are 23° apart and its direction is further constrained by its angles relative to other modes in a planar geometry (Figure 4.7).

4.3 Discussion

4.3.1 Vibrational couplings and mode character

In our characterization of the IR active in-plane double bond stretching vibrations of nucleotides, we see that the 2D IR spectra and corresponding vibrational anharmonic modes
are distinct, even with the structural similarities between the purine and pyrimidine bases. Except perhaps for the case of carbonyls, there are no simple or intuitive structural correlations to the spectral features, for instance with respect to conjugation, resonance structures, and ring placement of exocyclic groups. The number of resonances between 1650 cm\(^{-1}\)–1700 cm\(^{-1}\) matches the number of carbonyls in the base, and these are inhomogeneously broadened, reflecting the sensitivity of the carbonyl frequency to hydrogen bonding to the oxygen. Comparison of the experimental results to the harmonic modes derived from DFT calculations of explicitly hydrated nucleosides is generally favorable. For the most part, significant anharmonic couplings within the bases lead to vibrational modes that are delocalized over large portions of the base.

Comparing the two purines only the G1 peak that is largely due to the C\(^{\equiv}\)=O vibration is readily assigned. Due to the large energy splitting between G1 and the ring modes (> 43 cm\(^{-1}\)) and the small off-diagonal anharmonicities (< 5 cm\(^{-1}\)), it is reasonable to describe G1 as a local mode C=O stretch. The frequencies and amplitudes for the ring vibrations are difficult to predict and depend on where the exocyclic ND\(_2\) group resides. These lower frequency (1500 cm\(^{-1}\) – 1600 cm\(^{-1}\)) ring modes in AMP, GMP, and CMP originate from the one- or two-member ring deformation coupled to the small amplitude C-ND\(_2\) bend vibrations. It should be noted that H/D exchange redshifts the NH\(_2\) scissorings from the 1500–1600 cm\(^{-1}\) region to ~1100 cm\(^{-1}\).

The assignment of the carbonyl double bond stretches in UMP and TMP is more complicated as the two C=O groups are more strongly coupled. Traditionally, C=O vibrations are assigned as local modes with C\(^{\equiv}\)=O corresponding to U1/T1 and C\(^{\equiv}\)=O to U2/T2. However, the 2D IR spectroscopy indicates that observed couplings between the carbonyl stretches are not negligible, especially for UMP. UMP has diagonal anharmonicities greater than 15 cm\(^{-1}\), and the off-diagonal anharmonicities vary from 8.0 cm\(^{-1}\) (U1/U2) to 25 cm\(^{-1}\) (U1/U3). The coupling constant between carbonyls in the local mode picture can be calculated perturbatively\(^{25}\) to be \(2V = 26 \text{ cm}^{-1}\). This coupling strength is comparable to the frequency splitting of \(\Delta E = 38 \text{ cm}^{-1}\). Therefore, the system is not weakly coupled, and in fact intermediate between the strong (\(|2V| \gg |\Delta E|\)) and weak coupling (\(|2V| \ll |\Delta E|\)) regimes. This finding along with the projection angle of 90\(^{\circ}\) ± 7\(^{\circ}\) from the 2D spectral fitting match the results from the DFT calculations where symmetric and asymmetric C=O stretches are found. Further, the disorder (diagonal linewidth) for the carbonyls is ~30 cm\(^{-1}\), indicating that significant variation in the mode mixing and projection angle may exist within the ensemble.
On the other hand, TMP has much smaller anharmonicities relative to UMP. With a value of \( 2V = 12 \text{ cm}^{-1} \) between \( T1 \) and \( T2 \), compared to the \( 27 \text{ cm}^{-1} \) energy splitting, a weak coupling description of the two carbonyl stretches is more appropriate. This is in accordance with an angle of \( 117^\circ \pm 2^\circ \) between \( T1 \) and \( T2 \), indicating a more localized \( C=O \) stretch for \( T1 \). However, \( 2V \approx 34 \text{ cm}^{-1} \) was obtained between \( T2 \) and \( T3 \) (\( \Delta E = 38 \text{ cm}^{-1} \)), showing significant mixing for the \( C=O \) stretch. A similar effect of couplings between the carbonyl and ring modes was also observed for CMP vibrations, indicated by the pronounced cross-peaks and calculations based on perturbation theory (\( 2V \approx 28 \text{ cm}^{-1} ; \Delta E = 37 \text{ cm}^{-1} \)). Furthermore, the order of the frequencies of the ring vibrations involving \( C=C \) in these three pyrimidines \( (\omega_{C2} < \omega_{C1} < \omega_{C3}) \) is likely to be correlated with the degree of conjugation of the \( C=C \) bond in the pyrimidine ring. Therefore we conclude that the underlying in-plane base motions in UMP and TMP cannot be attributed to simple local \( C=O \) and \( C=C \) vibrations.

### 4.3.2 Experimental and theoretical limitations

By independently controlling the excitation and detection polarizations in 2D IR, one can measure the cross-peak anisotropy from which the relative dipole orientations can be obtained.\(^{23,42,43}\) This is the most sensitive probe of the vibrational potential, since the relative angle between the transition dipole moments is unusually sensitive to the directionality of the bonds in the molecule, the couplings and delocalization of the vibration, and the solvent environment. In the simple case where the peaks are well-separated, the relative peak amplitude of peaks in parallel and perpendicular spectra can be obtained directly from peak heights or integrated peak volume. However, in most cases, including the spectra presented here, spectral overlap between resonances leads to interferences that greatly impair the ability to directly extract angles from spectral amplitudes. Although 2D IR spectra are linearly additive in the components, the sum over positive and negative overlapping features means that quantitative analysis of amplitudes, and therefore angles, is unreliable without knowledge of the spectral components. In the present case, it was found that angle determination based on peak volume analysis was only reliable when applied to well-isolated cross-peaks such as the pairs \( A1/A2, G1/G2, \) and \( G1/G3 \). In the case of overlapping vibrational bands, it was found that the accurate dipole projection angles obtained from spectral fitting were as much as \( 40^\circ \) off the angles obtained from volume integration. Strategies that may help simplify such problems include the polarization-angle-scanning (PAS) 2D IR spectroscopy developed by Lee and coworkers.\(^{44}\)
Our measured transition dipole orientations compare favorably with DFT calculations on the five nucleosides provided that we include explicit water molecules to hydrogen bonding sites on the base, particularly to carbonyls. Hydrogen-bonding interactions between the water and solute strongly influence the vibrational modes, and thereby limit the ability to assign vibrations and relate them to molecular structure based solely on unsolvated DFT calculations. We find that in nucleotides, the hydrogen bonding to the water molecules can lead to vibrational frequency shifts as much as $-60\,\text{cm}^{-1}$ for carbonyl stretches, but only $+10\,\text{cm}^{-1}$ for the ring modes. However, the relative dipole orientations of the ring modes are more sensitive (dipole moments can rotate as much as $70^\circ$) to the solvent environment compared to the carbonyl stretches that have more local characters such as in the case of GMP and CMP. These issues take on new meaning when considering the dynamics and disorder of the system, which can lead to sharp variations in transition dipole direction. Although ab initio DFT calculations are an important component of the present work, it is the combination of calculations with the 2D IR experimental constraints that lead to accurate spectral assignments and interpretation.

4.3.3 Toward quantitative models of nucleic acid secondary structure.

Although the IR spectroscopy of DNA oligomers in the region of base vibrations is sensitive to the secondary structures of DNA, the underlying physics required to link the molecular structure and the IR spectral features is not well understood. Current use of FTIR for investigating DNA structure is largely based on empirical findings on marker bands. For instance, it is found that G1 mode observed for free GMP at $1663\,\text{cm}^{-1}$ is blue-shifted to $1689\,\text{cm}^{-1}$ in Watson-Crick (WC) G-C base-pairs or $1672\,\text{cm}^{-1}$ when GMP is assembled into G-quadruplexes stabilized via Hoogsteen G-G interactions. Also, upon WC duplex formation, it is also observed that G2 and G3 modes decrease in intensity. A more detailed understanding of how inter-molecular interactions influence the base vibrations is needed to move beyond qualitative analysis of IR spectra.

The characterization of the individual base vibrations forms a starting point for developing structure-based spectroscopic models of nucleic acid secondary structure, and for interpreting changes in frequency, intensity, and lineshape of the base vibrations in these structures. Studies of the protein backbone amide I vibration using 2D IR spectroscopy have offered increased structural resolution and have opened up an avenue for detailed probing protein dynamics largely as a result of excitonic models that describe the coupling between local mode peptide carbonyls. In the case of DNA and RNA, structure based models will likely take a different form, given the delocalized nature of the base vibrations and the variety
of base-base couplings available. However, intermolecular interactions can be modeled as perturbations to the anharmonic vibrational modes of single bases. In the case of the CO vibrations of GC oligonucleotides, Krummel and Zanni modeled the 2D IR spectra found that C=O vibrations of double helical GC pairs interact through electrostatics and base stacking, leading to delocalized modes across the base-pairs and along the helices. They also recognized that the coupling between the carbonyl and the ring modes are important for determining the inter-base coupling and the transition dipole directions.

4.4 Concluding Remarks

2D IR spectroscopy provides clear evidence that the in-plane ring stretching vibrations of nucleic acid bases between 1500 -1700 cm$^{-1}$ are significantly delocalized over the purine and pyrimidine rings. Although this has been recognized in prior $ab$ initio calculations, the traditional assignment and interpretation of DNA vibrations was typically based on a local mode description of particular double bond stretches. Such assignments remain reasonable for C=O vibrations of G and C. Modeling these spectra also provides a quantitative understanding of the anharmonic vibrational potential coupling the ring modes, and the relative orientation of transition dipole moments. Such work forms the basis for constructing models for the vibrational spectroscopy of DNA and RNA secondary structures and protein-DNA interactions, in which inter-base interactions modify the frequency, intensity, mode composition, and dipole orientations of the single-base vibrations.

4.5 Acknowledgments

I thank Professor Kimberly Hamad-Schifferli for providing us with the thrombin binding aptamer problem that motivated our first steps into the area of DNA vibrational spectroscopy. I thank Ziad Ganim for helpful discussion on 2D spectral fitting, and Andrew Horning, Carlos Baiz and Rebecca Nicodemus for assistance in DFT calculations.
4.6 References


Appendix 4.A: 2D IR spectra fitting model and parameters

The following tables list all the parameters used to fit the experimental 2D IR spectra for the five NMPs. $\omega_i$ is the transition frequency; $\Delta \omega_{i,j}$ and $\Delta \omega_{i,j}$ are the diagonal and off-diagonal anharmonicities, respectively; $\mu_i$ is the transition dipole strength for the zero- to one-quantum states; $\mu_{2i,j}$ is the transition dipole strength for the one- to two-quantum states; $\mu_{i,j}$ is the transition dipole strengths from the one-quantum state to a combination band; $T_2$, $\Delta$, $\tau$, and $\Delta_o$ are the lineshape parameters. It should be emphasized again that these lineshape parameters are only for the purpose of fitting the peak amplitudes, and they do not imply the real dynamics. In addition, the fundamental transition dipole strengths are influenced by the spectrum used in the 2D experiments, therefore the more accurate determination of the dipole strengths were obtained from fitting the FTIR using the lineshape parameters. However, the scaling of the dipole strengths for higher transitions remains accurate in an anisotropy measurement.

a) AMP

<table>
<thead>
<tr>
<th>$\omega_{A1}$</th>
<th>$\omega_{A2}$</th>
<th>$\mu_{A1}$</th>
<th>$\mu_{A2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1625</td>
<td>1578</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>$\Delta \omega_{A1}$</td>
<td>$\Delta \omega_{A2}$</td>
<td>$\Delta \omega_{A1,A2}$</td>
<td>$\Delta \omega_{A2,A1}$</td>
</tr>
<tr>
<td>8.00</td>
<td>9.80</td>
<td>3.65</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T_2$</th>
<th>$\Delta$</th>
<th>$\tau$</th>
<th>$\Delta_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal peaks A1</td>
<td>1909.6</td>
<td>1.7</td>
<td>95.0</td>
</tr>
<tr>
<td>A2</td>
<td>1995.0</td>
<td>35.8</td>
<td>800.3</td>
</tr>
<tr>
<td>Cross-peaks</td>
<td>A1/A2</td>
<td>2000.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>
### b) GMP

<table>
<thead>
<tr>
<th>$\omega_{G1}$</th>
<th>$\omega_{G2}$</th>
<th>$\omega_{G3}$</th>
<th>$\omega_{G4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1665</td>
<td>1579</td>
<td>1565</td>
<td>1539</td>
</tr>
<tr>
<td>$\Delta\omega_{G1}$</td>
<td>$\Delta\omega_{G2}$</td>
<td>$\Delta\omega_{G3}$</td>
<td>$\Delta\omega_{G4}$</td>
</tr>
<tr>
<td>10.94</td>
<td>4.00</td>
<td>7.76</td>
<td>17.98</td>
</tr>
<tr>
<td>$\Delta\omega_{G1/G2}$</td>
<td>$\Delta\omega_{G1/G3}$</td>
<td>$\Delta\omega_{G1/G4}$</td>
<td>$\Delta\omega_{G2/G3}$</td>
</tr>
<tr>
<td>1.34</td>
<td>4.57</td>
<td>3.27</td>
<td>7.51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{G1}$</th>
<th>$\mu_{G2}$</th>
<th>$\mu_{G3}$</th>
<th>$\mu_{G4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.87</td>
<td>0.76</td>
<td>0.21</td>
</tr>
<tr>
<td>$\mu_{G1/G1}$</td>
<td>$\mu_{G1/G2}$</td>
<td>$\mu_{G1/G3,G1}$</td>
<td>$\mu_{G1/G4,G1}$</td>
</tr>
<tr>
<td>1.40</td>
<td>1.18</td>
<td>1.04</td>
<td>0.49*</td>
</tr>
<tr>
<td>$\mu_{G1/G2,G1}$</td>
<td>$\mu_{G1/G3,G1}$</td>
<td>$\mu_{G1/G4,G1}$</td>
<td>$\mu_{G1/G4,G1}$</td>
</tr>
<tr>
<td>0.84</td>
<td>0.76</td>
<td>0.36</td>
<td>0.77</td>
</tr>
<tr>
<td>$\mu_{G1/G2,G2}$</td>
<td>$\mu_{G1/G3,G3}$</td>
<td>$\mu_{G1/G4,G4}$</td>
<td>$\mu_{G2/G3,G3}$</td>
</tr>
<tr>
<td>1.00</td>
<td>1.01</td>
<td>1.08</td>
<td>0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T$_2$</th>
<th>$\Delta$</th>
<th>$\tau$</th>
<th>$\Delta_\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal peaks</td>
<td>G1</td>
<td>566.1</td>
<td>49.7</td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>623.2</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>598.8</td>
<td>59.9</td>
</tr>
<tr>
<td></td>
<td>G4</td>
<td>1645.7</td>
<td>122.8</td>
</tr>
<tr>
<td>Cross-peaks</td>
<td>G1/G2</td>
<td>4544.3</td>
<td>42.7</td>
</tr>
<tr>
<td></td>
<td>G1/G3</td>
<td>4363.7</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>G1/G4</td>
<td>1866.0</td>
<td>118.9</td>
</tr>
<tr>
<td></td>
<td>G2/G3</td>
<td>5111.3</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>G2/G4</td>
<td>1025.5</td>
<td>34.0</td>
</tr>
<tr>
<td></td>
<td>G3/G4</td>
<td>1901.0</td>
<td>9.9</td>
</tr>
</tbody>
</table>

* A deviation from the harmonic scaling in the transition dipole strength was observed for mode G4 by comparing the amplitudes of the diagonal absorptive and emissive peaks. A relative strength of $\mu_{G4,G4} / \mu_{G4} = 2.3$ suggests significant electrical anharmonicity, however, this observation can also be caused by the low peak intensity of G4 that led to the insensitivity of amplitude fitting.
### c) UMP

<table>
<thead>
<tr>
<th>$\omega_{U1}$</th>
<th>$\omega_{U2}$</th>
<th>$\omega_{U3}$</th>
<th>$\mu_{U1}$</th>
<th>$\mu_{U2}$</th>
<th>$\mu_{U3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1693</td>
<td>1655</td>
<td>1617</td>
<td>0.76</td>
<td>1.01</td>
<td>0.46</td>
</tr>
<tr>
<td>$\Delta \omega_{U1}$</td>
<td>$\Delta \omega_{U2}$</td>
<td>$\Delta \omega_{U3}$</td>
<td>$\mu_{2U1/U1}$</td>
<td>$\mu_{2U2/U2}$</td>
<td>$\mu_{2U3/U3}$</td>
</tr>
<tr>
<td>15.96</td>
<td>18.32</td>
<td>25.00</td>
<td>0.95</td>
<td>1.35</td>
<td>0.67</td>
</tr>
<tr>
<td>$\Delta \omega_{U1/U2}$</td>
<td>$\Delta \omega_{U1/U3}$</td>
<td>$\Delta \omega_{U2/U3}$</td>
<td>$\mu_{U1/U2/U1}$</td>
<td>$\mu_{U1/U3/U1}$</td>
<td>$\mu_{U2/U3/U2}$</td>
</tr>
<tr>
<td>7.96</td>
<td>24.98</td>
<td>15.40</td>
<td>0.9</td>
<td>0.55</td>
<td>0.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T_2$</th>
<th>$\Delta$</th>
<th>$\tau$</th>
<th>$\Delta_\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal peaks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>1291.7</td>
<td>98.0</td>
<td>199.6</td>
</tr>
<tr>
<td>U2</td>
<td>637.7</td>
<td>100.3</td>
<td>17.4</td>
</tr>
<tr>
<td>U3</td>
<td>1773.9</td>
<td>93.2</td>
<td>643.3</td>
</tr>
<tr>
<td>Cross-peaks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1/U2</td>
<td>397.3</td>
<td>160.2</td>
<td>141.8</td>
</tr>
<tr>
<td>U1/U3</td>
<td>862.6</td>
<td>102.4</td>
<td>338.8</td>
</tr>
<tr>
<td>U2/U3</td>
<td>1977.9</td>
<td>31.5</td>
<td>518.8</td>
</tr>
</tbody>
</table>

### d) TMP

<table>
<thead>
<tr>
<th>$\omega_{T1}$</th>
<th>$\omega_{T2}$</th>
<th>$\omega_{T3}$</th>
<th>$\mu_{T1}$</th>
<th>$\mu_{T2}$</th>
<th>$\mu_{T3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1690</td>
<td>1663</td>
<td>1629</td>
<td>0.71</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>$\Delta \omega_{T1}$</td>
<td>$\Delta \omega_{T2}$</td>
<td>$\Delta \omega_{T3}$</td>
<td>$\mu_{2T1/T1}$</td>
<td>$\mu_{2T2/T2}$</td>
<td>$\mu_{2T3/T3}$</td>
</tr>
<tr>
<td>10.71</td>
<td>5.00</td>
<td>10.00</td>
<td>1.2</td>
<td>1.42</td>
<td>1.37</td>
</tr>
<tr>
<td>$\Delta \omega_{T1/T2}$</td>
<td>$\Delta \omega_{T1/T3}$</td>
<td>$\Delta \omega_{T2/T3}$</td>
<td>$\mu_{T1/T2/T1}$</td>
<td>$\mu_{T1/T3/T1}$</td>
<td>$\mu_{T2/T3/T2}$</td>
</tr>
<tr>
<td>1.73</td>
<td>4.92</td>
<td>6.15</td>
<td>0.80</td>
<td>1.07</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T_2$</th>
<th>$\Delta$</th>
<th>$\tau$</th>
<th>$\Delta_\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal peaks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>8873.7</td>
<td>163.43</td>
<td>139.88</td>
</tr>
<tr>
<td>T2</td>
<td>5510.29</td>
<td>334.96</td>
<td>16.73</td>
</tr>
<tr>
<td>T3</td>
<td>10904.81</td>
<td>175.05</td>
<td>97.86</td>
</tr>
<tr>
<td>Cross-peaks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1/T2</td>
<td>5920.99</td>
<td>31.38</td>
<td>20.12</td>
</tr>
<tr>
<td>T1/T3</td>
<td>6050.92</td>
<td>30.01</td>
<td>20.14</td>
</tr>
<tr>
<td>T2/T3</td>
<td>6227.86</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>
c) CMP

<table>
<thead>
<tr>
<th>$\omega_{C1}$</th>
<th>$\omega_{C2}$</th>
<th>$\omega_{C3}$</th>
<th>$\omega_{C4}$</th>
<th>$\omega_{C5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1651</td>
<td>1614</td>
<td>1583</td>
<td>1524</td>
<td>1504</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Delta \omega_{C1}$</th>
<th>$\Delta \omega_{C2}$</th>
<th>$\Delta \omega_{C3}$</th>
<th>$\Delta \omega_{C4}$</th>
<th>$\Delta \omega_{C5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.19</td>
<td>15.40</td>
<td>16.70</td>
<td>6.12</td>
<td>8.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Delta \omega_{C1/C2}$</th>
<th>$\Delta \omega_{C1/C3}$</th>
<th>$\Delta \omega_{C1/C4}$</th>
<th>$\Delta \omega_{C1/C5}$</th>
<th>$\Delta \omega_{C2/C3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.15</td>
<td>7.38</td>
<td>1.64</td>
<td>5.19</td>
<td>11.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Delta \omega_{C2/C4}$</th>
<th>$\Delta \omega_{C2/C5}$</th>
<th>$\Delta \omega_{G3/G4}$</th>
<th>$\Delta \omega_{G3/G5}$</th>
<th>$\Delta \omega_{C3/C4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.38</td>
<td>7.64</td>
<td>12.76</td>
<td>8.24</td>
<td>1.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{C1}$</th>
<th>$\mu_{C2}$</th>
<th>$\mu_{C3}$</th>
<th>$\mu_{C4}$</th>
<th>$\mu_{C5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>0.62</td>
<td>0.19</td>
<td>0.63</td>
<td>0.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{2C1C1}$</th>
<th>$\mu_{2C2C2}$</th>
<th>$\mu_{2C3C3}$</th>
<th>$\mu_{2C4C4}$</th>
<th>$\mu_{2C5C5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>0.98</td>
<td>0.55</td>
<td>0.72</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{C1C2}$</th>
<th>$\mu_{C1C3}$</th>
<th>$\mu_{C1C4}$</th>
<th>$\mu_{C1C5}$</th>
<th>$\mu_{C2C3}$</th>
<th>$\mu_{C2C4}$</th>
<th>$\mu_{C2C5}$</th>
<th>$\mu_{C3C4}$</th>
<th>$\mu_{C3C5}$</th>
<th>$\mu_{C4C5}$</th>
<th>$\mu_{C4C5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.61</td>
<td>0.24</td>
<td>0.64</td>
<td>0.59</td>
<td>0.40</td>
<td>0.62</td>
<td>0.40</td>
<td>0.30</td>
<td>0.70</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{C1C2}$</th>
<th>$\mu_{C1C3}$</th>
<th>$\mu_{C1C4}$</th>
<th>$\mu_{C1C5}$</th>
<th>$\mu_{C2C3}$</th>
<th>$\mu_{C2C4}$</th>
<th>$\mu_{C2C5}$</th>
<th>$\mu_{C3C4}$</th>
<th>$\mu_{C3C5}$</th>
<th>$\mu_{C4C5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.81</td>
<td>0.87</td>
<td>0.81</td>
<td>0.81</td>
<td>0.51</td>
<td>0.78</td>
<td>0.66</td>
<td>0.39</td>
<td>0.28</td>
<td>0.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T_2$</th>
<th>$\Delta$</th>
<th>$\tau$</th>
<th>$\Delta_\varphi$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Diagonal peaks</th>
<th>C1</th>
<th>1584</th>
<th>109</th>
<th>204</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>2500</td>
<td>78</td>
<td>600</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>2500</td>
<td>17</td>
<td>45</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>651</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>583</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

| C1/C2         | 2326| 104  | 499 | 0   |
| C1/C3         | 862 | 150  | 143 | 0   |
| C1/C4         | 577 | --   | --  | --  |
| C1/C5         | 1844| --   | --  | --  |
| C2/C3         | 1285| 81   | 200 | 0   |
| C2/C4         | 2000| --   | --  | --  |
| C2/C5         | 1698| --   | --  | --  |
| C3/C4         | 1559| --   | --  | --  |
| C3/C5         | 509 | --   | --  | --  |
| C4/C5         | 1961| --   | --  | --  |
CHAPTER 5
Identification of Lactam–Lactim Tautomers of Aromatic Heterocycles Using 2D IR Spectroscopy

The work presented in this chapter has been published in the following papers:


5.1 Introduction

Understanding the tautomerism of aromatic heterocycles is of great importance in many areas of chemistry and biology. Histidine residues are often present in enzyme active sites because tautomerism of their imidazole rings provides chemical versatility and facilitates proton transfer in the catalytic steps. Since the canonical keto-amino forms of DNA nucleobases define the hydrogen-bonding structure for Watson-Crick base-pairing, tautomerization can lead to spontaneous mutagenesis when minor tautomers mispair during replication.
To characterize the tautomeric equilibria, researchers have used techniques such as UV absorption, circular dichroism, X-ray crystallography, NMR, Raman and IR absorption spectroscopy. However, these methods face several challenges to characterizing thermodynamic and kinetic data for tautomeric equilibria and exchange processes. For example, electronic spectra are broad and featureless which complicates spectral interpretation, particularly for systems with multiple tautomers. On the other hand, NMR provides excellent structural resolution, but it can only directly monitor chemical exchange in real time on millisecond and longer time scales, rather than the picosecond to nanosecond time scales expected for proton transfer under physiological conditions. From the theoretical point of view, quantum mechanical calculations have been performed extensively to characterize the relative stability of tautomeric systems, their activation barriers, and the reaction mechanisms; nevertheless, experimental validation of these predictions is scarce. In order to provide a characterization under ambient aqueous conditions, we need a technique with both the structural sensitivity to unambiguously identify various tautomers, and the time resolution to probe the rapid proton transfer dynamics.

Two-dimensional infrared spectroscopy (2D IR) is an emerging molecular vibrational spectroscopy which can be used to characterize tautomeric equilibria. These experiments use sequences of ultrafast IR pulses to characterize vibrational couplings, which appear as cross-peaks in 2D IR spectra. The distinct cross-peaks originating from different tautomers in 2D IR spectra enable unambiguous peak assignments that are often not possible in the congested FTIR or Raman spectra. Furthermore, the intrinsic picosecond time-resolution means that these measurements can potentially characterize the time-scale of tautomer exchange processes.

In this chapter, we first demonstrate the capability of 2D IR spectroscopy to characterize tautomeric equilibria using a study of keto-lactam to enol-lactim tautomerization in a model system, 2-pyridone derivatives. 2-pyridone has been regarded as the prototype for the keto-enol tautomeration in heterocycles. We focus on 2-pyridone and 6-chloro-2-pyridone to discuss the spectral signatures of the lactam and lactim tautomers. We chose these two model compounds to benchmark with existing experimental and computational results, and also to examine the agreement between the experimental IR spectra and DFT calculations. We study 4-pyrimidinone as an application to systems with multiple lactam tautomers. Finally, we apply the methodology developed in the first part of this chapter to study the tautomerism of oxythiamine, the oxidized form of thiamine, and discuss the implications on its recognition by the thiamine pyrophosphate riboswitch.
5.2 Tautomerism of pyridone derivatives in aqueous solution

5.2.1 Equilibrium spectra and peak assignments

The chemical structures of the possible tautomers of the molecules studied, along with their FTIR and 2D IR spectra at room temperature in D$_2$O, are shown in Figure 5.1. Previous 2D IR work on nucleic acid bases in the in-plane base vibration region has shown that clear cross-peaks can be observed between all of the diagonal peaks due to the delocalized nature of these in-plane C=O and ring vibrations in the aromatic system. The infrared spectra of 2-pyridone from 1500 cm$^{-1}$ – 1700 cm$^{-1}$ display three well-separated peaks at 1643 cm$^{-1}$ ($\Lambda_1$), 1560 cm$^{-1}$ ($\Lambda_2$), and 1541 cm$^{-1}$ ($\Lambda_3$). In general, vibrational modes with high intensity and frequency $> 1640$ cm$^{-1}$ can be assigned to C=O stretching; therefore $\Lambda_1$ is expected to be the carbonyl stretch of the lactam tautomer 1b. Additionally, since prominent cross-peaks exist between all three modes in the 2D IR spectrum, these vibrational modes are attributed to the same chemical species, the lactam tautomer 1b.

The presence of a monomeric species is confirmed with concentration dependence study shown in Figure 5.2, as well as studies in varying organic solvents (see section 5.4). The normalized FTIR spectra of 2-pyridone from 1 mg/ml to 20 mg/ml are almost identical, except for the slight offset seen in the 1 mg/ml sample due to imperfect background subtraction. The concentration dependence indicates that 2-pyridone is in the monomer state within the concentration range in our experiments. The monomer/dimer equilibrium of 2-pyridone is dependent on solvent polarity, with monomers and dimers predominating in polar and non-polar solvents, respectively.
Figure 5.1: Chemical structures of possible tautomers (top), FTIR (middle) and 2D IR spectra (bottom) for: (a) 2-pyridone; (b) 6-chloro-2-pyridone; (c) 4-pyrimidinone. The IR spectra were taken at 23°C and in D₂O (phosphate buffer at pH* = 7.4). The 2D IR spectra, acquired with all-parallel polarization (ZZZZ), were normalized to the maximum of absolute value, and 25 equally spaced contours from -1 to 1 were plotted. 2D IR spectrum of 4-pyrimidinone have been scaled by x2.5 for frequencies < 1605 cm⁻¹ to allow better visualization of the cross-peaks.

Figure 5.2: Concentration-dependent FTIR of 2-pyridone in phosphate buffer.
To assign the three observed vibrational resonances in Figure 1a to 1b, DFT calculations of deuterated 1b and 1a (2-hydroxypyridine) were performed using B3LYP/6-31G(d,p). Figure 5.5 shows the calculated IR spectra of 1b with zero, one, and two explicit D$_2$O molecules making hydrogen-bonds (H-bonds) to the oxygen and nitrogen atoms (calculated frequencies are tabulated in Table 5.1). Interestingly, as the number of explicit water molecules increase, the C=O and ring modes redshift due to the H-bond interactions with the solvent molecules. The intensity of the A2 mode also changes dramatically. The calculated frequencies and intensities for 1b with two explicit D$_2$O molecules match experiment very well. This allows us to conclude that 1b is the major species for 2-pyridone under the experimental conditions. This is consistent with previous findings that the lactam-lactim equilibrium shifts to lactam in polar solvents such as water.$^{19,20}$

The importance of including explicit water molecules for describing the vibrational spectra, especially the two higher frequency modes A1 and A2, can be illustrated by looking at the normal mode composition in these calculations. We find that all three modes are highly delocalized with C=O stretch strongly mixed with the ring vibrations. Although A1 is predominately C=O stretch, as the number of D$_2$O molecules hydrogen bonded to the oxygen increases from 0 to 2, the A1 frequency red-shifts from 1713 cm$^{-1}$ to 1646 cm$^{-1}$, and its ring breathing mode character increases. Dramatic red-shifting and intensification were also observed for A2. Frequency shift and intensity variation upon hydration have also been observed in DFT calculations of DNA base-pairs.$^{21}$ On the other hand, A3, a ring

![Figure 5.3: DFT calculated FTIR spectra of 2-pyridone with different number of explicit D$_2$O molecules: 0 (blue), 1 (green) and 2 (red). All peak line-shapes were generated by convoluting stick spectra with a Gaussian profile with $\sigma = 5$ cm$^{-1}$.](image)
deformation mode consisting of symmetric C=C stretches, is less affected by solvation. Interestingly, the vibrational modes of 1a are only marginally affected by the presence of solvent molecules (Figure 5.4). These results not only show the influence of water on vibrational frequencies, but also provide evidence for the presence of a bridging water molecule that facilitates the proton transfer induced exchange between 1a and 1b.

Table 5.1. Vibrational frequencies of experiment v.s. DFT calculations. The frequencies are in units of cm⁻¹, and the calculated results have been scaled by 0.9614. All calculations were with explicit D₂O molecules.

### 2-pyridone

<table>
<thead>
<tr>
<th>Peak</th>
<th>Experiment</th>
<th>DFT calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lactam</td>
</tr>
<tr>
<td>A1</td>
<td>1643</td>
<td>1647</td>
</tr>
<tr>
<td>A2</td>
<td>1560</td>
<td>1575</td>
</tr>
<tr>
<td>A3</td>
<td>1541</td>
<td>1528</td>
</tr>
</tbody>
</table>

### 6-chloro-2-pyridone

<table>
<thead>
<tr>
<th>Peak</th>
<th>Experiment</th>
<th>DFT calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lactam</td>
</tr>
<tr>
<td>B1</td>
<td>1637</td>
<td>1642</td>
</tr>
<tr>
<td>B2</td>
<td>1591</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>1550</td>
<td>1568</td>
</tr>
<tr>
<td>B4</td>
<td>1528</td>
<td>1507</td>
</tr>
</tbody>
</table>

### 4-pyrimidinone

<table>
<thead>
<tr>
<th>Peak</th>
<th>Experiment</th>
<th>DFT calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N1H lactam</td>
</tr>
<tr>
<td>C1</td>
<td>1654</td>
<td>1657</td>
</tr>
<tr>
<td>C2</td>
<td>1646</td>
<td>1654</td>
</tr>
<tr>
<td>C3</td>
<td>1593</td>
<td>1577</td>
</tr>
<tr>
<td>C4</td>
<td>1585</td>
<td>---³</td>
</tr>
<tr>
<td>C5</td>
<td>1538</td>
<td>1529</td>
</tr>
<tr>
<td>C6</td>
<td>1503</td>
<td>1439</td>
</tr>
</tbody>
</table>

³ The C4 mode is shown to have cross-peaks to C2, C5, and C6 in the 2D IR spectrum, therefore we assign it to be one of the N1H lactam vibration. However, this mode is not observed in the DFT calculations.
Substituents on aromatic heterocycles are known to shift the tautomeric equilibrium, and the addition of chloride at the 6 position was found to increase the lactim population.\textsuperscript{19,22} We therefore investigated the IR spectral signatures of 6-chloro-2-pyridone (2b) and its lactim tautomer, 6-chloro-2-hydroxypyridine (2a). The FTIR spectrum of 6-chloro-2-pyridone shows 4 bands of which the B1, B3, and B4 peak frequencies and 2D IR lineshapes are close to those of the parent 2-pyridone, suggesting that these peaks may arise from the lactam tautomer. Interestingly, the 2D IR spectrum shows that the 1591 cm\(^{-1}\) peak (B2) has no cross-peaks to the three peaks mentioned above, indicating that this vibrational mode originates from a separate species. IR harmonic frequencies and amplitudes of 2a and 2b with two solvating D\(_2\)O molecules were computed and shown in Figure 5.5 (spectra with zero and one water shown in Figure 5.6). There is a good correspondence with the experimental spectra where the three measured peaks at 1637 cm\(^{-1}\) (B1), 1550 cm\(^{-1}\) (B3), and 1528 cm\(^{-1}\) (B4) match well with the calculated modes for 2b at 1642 cm\(^{-1}\), 1568 cm\(^{-1}\), and 1506 cm\(^{-1}\). Likewise, the experimental 1591 cm\(^{-1}\) (B2) peak can be assigned to the calculated 1593 cm\(^{-1}\) peak for the lactim tautomer 2a. The second calculated band for 2a was not observed experimentally, most likely due to the lower oscillator strength and the low population of 2a. In addition, this peak overlaps with B3 and B4 peaks in 2b in a congested spectral region that would also make it difficult to be detected. The 2D IR spectrum of 6-chloro-2-pyridone clearly reveals the existence of two chemical species, and by comparing...
the experimental spectra to DFT calculations, we assigned the two species to be the lactam and lactim tautomers hydrogen-bonded with D$_2$O. IR spectroscopy has been used in the gas phase to identify the lactim tautomer by detecting the O-H stretching vibration. This approach is not feasible in aqueous systems since the water O-H signal dominates. The identification of the distinct lactim vibration in the fingerprint region therefore provides an alternative means of characterizing the lactim tautomer.

Figure 5.5: DFT calculated FTIR spectra of the lactam (blue) and lactim (red) tautomers of 6-chloro-2-pyridone with two explicit D$_2$O molecules.

Figure 5.6: DFT calculated IR spectra of 6-chloro-2-pyridone-d$_1$ (left) and 6-chloro-2-hydroxypyridine-d$_1$ (right) with different number of D$_2$O making HB’s to O and N atoms.
As a comparison, we have also measured UV spectra for the first \( \pi \rightarrow \pi^* \) absorption band of 2-pyridone and 6-chloro-2-pyridone in H\(_2\)O and THF, shown in Figure 5.7. Previous calculations\(^{23}\) have shown that for a given pyridone derivative, the wavelength of the absorption maximum follows a general trend: lactim monomer < lactam monomer. The spectrum of 2-pyridone in H\(_2\)O has a maximum at 294 nm, which has been attributed to the lactam tautomer.\(^{24}\) In general, the lactim tautomers are more populated in non-polar solvents as seen by the spectrum of 6-chloro-2-pyridone in THF, where the absorption maximum is at shorter wavelength 278 nm. However, for 50 µg/ml 2-pyridone in THF, the spectrum displays a maximum at 307 nm (with vibrational bands at 335 nm, 320 nm, 307 nm, and 286 nm) indicating that the lactam tautomer is the dominating species. As the concentration of 2-pyridone in THF increases, 2-pyridone dimerizes and lead to a shift of the band structures.\(^{14}\) For 6-chloro-2-pyridone in H\(_2\)O, the spectrum has a maximum at 300 nm, which would indicate that the lactam form is the predominant species, and a 5 \% of lactim was reported by Katritzky and coworkers.\(^{24}\) Nevertheless, the IR spectra reported in this paper reveal a significant amount of lactim population, demonstrating that IR spectroscopy is more sensitive to the lactam-lactim tautomerism when the two tautomer populations are close to each other, leading to an overlap in the UV absorption.

Figure 5.7: UV absorption spectra of 2-pyridone (blue) and 6-chloro-2-pyridone (red) in H\(_2\)O (solid) and THF (dashed). The samples in H\(_2\)O were buffered with phosphate buffer at pH = 7.4. The sample concentrations were 50 µg/ml.
5.2.2 DFT calculation: relative stabilities of the lactam-lactim tautomers

Numerous studies have been done to predict the tautomeric equilibria using computational methods. Here we address some of the issues and challenges in this approach.

First of all, it should be kept in mind that the calculated stabilities of these tautomers are sensitive to not only the solvent environment but also the computational methods used. The relative energies can be differed by as much as 5 kcal/mol between different levels of theory, and sometimes even the sign can be reversed.\(^{15,25,26}\) However, a 0.6 kcal/mol energy difference would result in a drastic 36% difference in relative population. It has not been established that a specific computational method works the best when compared to the experimental results. The relative stabilities predicted by these calculations in general agree with the experiments, but the absolute numbers are still under debate.

Secondly, it has become a common practice for experimentalists to run simple DFT calculations in the gas phase to compare to the experimental results. However, tautomerism in biologically relevant processes happen in aqueous solutions, therefore, it is not appropriate to compare the energy difference between tautomers from DFT calculations to experimental results. A proper selection of the solvent model is crucial for even attempting to compare the experimental results with calculations. Take 2-pyridone for example, gas phase experiments have shown that the lactim form is more stable by about -0.58 kcal/mol.\(^{27}\) On the other hand, the equilibrium is shifted towards the lactam tautomers by about 1 kcal/mol in solution, which has been explained by the larger dipole moment of 2-pyridone and its ability to form hydrogen-bonds with water.\(^{28}\) Additionally, the activation barrier for the tautomerization of 2-pyridone to 2-hydroxypyridine was calculated to be 38.45 kcal/mol.\(^{26}\) This value is reduced by about 15 kcal/mol when a single water molecule is included in the calculation and can be further reduced with two hydrating water.\(^{28}\) A better solvation model is clearly needed to bridge the gap between the experiments and calculations.
Table 5.2. Relative energies of the tautomers. Zero-point energies have been included. E: electronic energy; H: enthalpy; G: Gibbs free energy. Energies are expressed in units of kcal/mol.

<table>
<thead>
<tr>
<th></th>
<th>PD</th>
<th>HP1</th>
<th>HP2</th>
<th></th>
<th>PD</th>
<th>HP1</th>
<th>HP2</th>
<th></th>
<th>PD</th>
<th>HP1</th>
<th>HP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-pyridone</td>
<td>E</td>
<td>0</td>
<td>0.036</td>
<td>5.365</td>
<td>0</td>
<td>0.859</td>
<td>12.7</td>
<td>0</td>
<td>3.145</td>
<td>10.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E+H</td>
<td>0.003</td>
<td>0</td>
<td>5.409</td>
<td>0</td>
<td>0.787</td>
<td>13.43</td>
<td>0</td>
<td>3.195</td>
<td>11.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E+G</td>
<td>0</td>
<td>0.104</td>
<td>5.348</td>
<td>0</td>
<td>0.939</td>
<td>10.94</td>
<td>0</td>
<td>2.959</td>
<td>8.679</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PD</th>
<th>HP1</th>
<th>HP2</th>
<th></th>
<th>PD</th>
<th>HP1</th>
<th>HP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-chloro-2-pyridone</td>
<td>E</td>
<td>3.096</td>
<td>0</td>
<td>5.192</td>
<td>2.699</td>
<td>0</td>
<td>12.97</td>
</tr>
<tr>
<td></td>
<td>E+H</td>
<td>3.156</td>
<td>0</td>
<td>5.257</td>
<td>2.719</td>
<td>0</td>
<td>13.84</td>
</tr>
<tr>
<td></td>
<td>E+G</td>
<td>3.001</td>
<td>0</td>
<td>5.127</td>
<td>2.696</td>
<td>0</td>
<td>10.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N1H</th>
<th>N3H</th>
<th>Lactim</th>
<th></th>
<th>N1H</th>
<th>N3H</th>
<th>Lactim</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-pyrimidinone</td>
<td>E</td>
<td>9.826</td>
<td>0</td>
<td>1.276</td>
<td>10.9</td>
<td>0</td>
<td>3.656</td>
</tr>
<tr>
<td></td>
<td>E+H</td>
<td>9.954</td>
<td>0</td>
<td>1.237</td>
<td>11.81</td>
<td>0</td>
<td>3.697</td>
</tr>
<tr>
<td></td>
<td>E+G</td>
<td>9.622</td>
<td>0</td>
<td>1.344</td>
<td>8.052</td>
<td>0</td>
<td>3.409</td>
</tr>
</tbody>
</table>

Since the calculations show that HP2 is much higher in energy compared to HP1, we only compared the experimental results to the calculation results for HP1. We can notice that lactim becomes more stable than the lactam tautomers for 6-chloro-2-pyridone compared to 2-pyridone. This observation is consistent with our IR experimental results. However, the experiments show that the lactam population is greater than the lactim population for 6-chloro-2-pyridone ($K_{eq} = 2.1$ and $\Delta H = -3.3$ kcal/mol), which is contradictory to the calculations ($\Delta H = +0.35$ kcal/mol). This discrepancy demonstrates that current DFT calculations cannot predict accurately the relative stabilities between tautomers, yet predicting a general trend going from one molecule to another is plausible.

![Figure 5.8: Structures of the two forms of 2-hydroxypyridone.](image)
5.2.3 Temperature dependence of 6-chloro-2-pyridone tautomerism

Temperature-dependent FTIR spectra of 6-chloro-2-pyridone (Figure 5.9a) show gain of the B2 peak and loss of the other three peaks, providing evidence that the ratio of lactim to lactam population increases with temperature. For the temperature-dependent 2D IR spectra (Figure 5.10), we performed singular value decomposition (SVD), which linearly decomposes the spectra and encapsulates the major spectral changes in response to temperature in the second spectral component (Figure 5.9b). Similar observations were found with a gain of the 2a peak (red dotted ellipse) and loss of 2b diagonal and cross-peaks. Integrating the FTIR peak intensities under bands B1 and B2 allows a thermodynamic analysis of the tautomerism within a simple two-state picture (Appendix 5.A). The resulting temperature dependent tautomer populations are shown in Figure 5.9c. At room temperature, the equilibrium constant $K_{eq} = [2b]/[2a] = [\text{lactam}]/[\text{lactim}]$ was found to be 2.1 and in agreement with the value reported by Forlani and coworkers.\textsuperscript{19} Fitting the Arrhenius equation resulted in $\Delta H = -3.3$ kcal/mol and $\Delta S = -9.8$ cal/mol·K for the lactim to lactam conversion.

![Figure 5.9: Temperature dependence of 6-chloro-2-pyridone. (a) FTIR from 10°C (blue) to 90°C (red). (b) The second component spectrum calculated from the SVD analysis of the temperature dependent 2D IR spectra. (c) Populations of the lactim (2a, red circle) and lactam (2b, blue square) tautomers determined from the temperature-dependent FTIR. (d) Arrhenius plot.](image-url)
Figure 5.10: Temperature-dependent 2D IR spectra of 6-chloro-2-pyridone demonstrating the increase of lactim population at the expense of lactam tautomers as the temperature increases. The lactam tautomers also shift to less hydrogen-bonded states.

In addition to the tautomer equilibria, changes of the lactam hydration environment were also captured in 2D IR spectra. On the diagonal, Figure 5.9b shows that both B1 and B3 peaks are lost with temperature, but that this loss originates from the red-side of the lineshape, and is accompanied by some gain on the blue side (purple arrows). This spectral feature corresponds to a blue-shift of the peak that is consistent with the weakening and disruption of the H-bonds around the molecule. A blue-shift of 12 cm\(^{-1}\) of the C=O peak can be compared to the amide carbonyl red-shifts of \(~16\) cm\(^{-1}\) with each addition of a H-bond to oxygen.\(^{29,30}\) Similar to 2-pyridone, DFT calculations show that the two higher frequency modes of the lactam tautomer 2b red-shift with increasing number of H-bonds to the carbonyl, whereas the modes of the lactim tautomer 2a are less affected (Figure 5.6).

5.2.4 Tautomerism of 4-pyrimidinone

2D IR spectrum of 4-pyrimidinone (Figure 5.1c) demonstrates how different lactam tautomers can be distinguished. 4-pyrimidinone is structurally similar to 2-pyridone with the replacement of C by N at the para-position of the carbonyl. Therefore, besides the lactim tautomer (3a), 4-pyrimidinone can exist in two different lactam forms depending on which N atom is protonated: the N1H lactam (3b) and N3H lactam (3c) tautomers. The FTIR
The spectrum of 4-pyrimidinone is distinct from those of 2-pyridone and 6-chloro-2-pyridone, most noticeably by the two carbonyl peaks at 1654 cm\(^{-1}\) (C1) and 1646 cm\(^{-1}\) (C2). This spectral feature is also reflected by the well-separated diagonal peaks in the 2D IR spectrum. As demonstrated earlier, the diagonal line-width reports the degree of solvent exposure since different surrounding environment leads to frequency distribution that results in peaks elongating along the diagonal. As a consequence, the C=O peak splitting can potentially be due to different solvent configurations around the carbonyl group which differ by one hydrogen bond to the C=O on average. However, the distinct cross-peak patterns of these two diagonal peaks to the 1503 cm\(^{-1}\) (C6) mode rule out this possibility, and indicate that they are two separate chemical species.

The DFT calculations of the three tautomers 3a-3c with explicit solvent molecules predict that the carbonyl stretching mode of 3b is 5 cm\(^{-1}\) lower than that of 3c (Figure 5.11). Furthermore, the calculations show that, besides both lactam tautomers sharing a peak at ~1520 cm\(^{-1}\), 3b has a distinct mode at 1439 cm\(^{-1}\) whereas 3c has a peak at 1578 cm\(^{-1}\). These results are consistent with the cross-peak patterns seen in the experimental 2D IR spectrum, where cross-peaks are observed between the C5 peak and both C=O peaks, in addition to the distinctive cross-peaks at \((\omega_1, \omega_2) = (1503 \text{ cm}^{-1}, 1646 \text{ cm}^{-1})\) and \((1593 \text{ cm}^{-1}, 1654 \text{ cm}^{-1})\). Therefore, we assign C1 and C2 to 3c and 3b, respectively.

![Graphical representation of DFT calculations and IR spectra](image.png)

Figure 5.11: DFT calculated IR spectra of N1H lactam (top), N3H lactam (middle), and lactim (bottom) tautomers of 4-pyrimidinone with 3 D\(_2\)O molecules.
Comparing the experimental spectra and DFT calculations, we did not detect significant lactim tautomer. This is in agreement with previous IR results of 4-pyrimidinone in CCl₄ where an equilibrium constant of 0.012 was measured using [OH]/[NH]². However, despite the general agreement between the DFT calculations and experimental spectra, we observed a larger intensity mismatch at 1590 cm⁻¹ compared to the cases of 2-pyridone and 6-chloro-2-pyridone. The additional intensity is likely contributed by some amount of deprotonated 4-pyrimidinone, as shown by the pH⁺ = 13 spectrum in Figure 5.12.

![Figure 5.12: Experimental pH-dependent FTIR of 4-pyrimidinone (left) and DFT calculated FTIR (right) for different protonation states.](image)

In summary, we have demonstrated the capability of 2D IR spectroscopy in identifying the lactam and lactim tautomers in aromatic heterocycles using the distinctive cross-peak patterns. The observation of a ring vibrational marker for the lactim tautomer in the fingerprint region offers a way to probe the lactim tautomer in aqueous solution where water signal dominates in the O-H or O-D stretching region. We concluded that 2-pyridone exists mainly in the lactam form. 6-chloro substitution of 2-pyridone shifts the tautomeric equilibrium towards the lactim tautomer, and the fraction of this form increases with temperature. We have also observed that the N1H and N3H lactam tautomers of 4-pyrimidinone lead to frequency shifts of the carbonyl group and different cross-peak patterns. We should note that in principle, the equilibrium distribution can be obtained solely with FTIR once the assignment has been determined with 2D IR spectra and if the peaks are well-separated. The analysis presented here thereby provide a starting point to further develop 2D IR spectroscopy into a complementary technique to add to the existing toolbox.
in probing tautomerism in complex biological systems. Relevant biological processes take place in aqueous solutions, and water facilitates proton transfer by forming water bridges. With the intrinsic ultrafast time resolution and the capability of fast triggering technique such as photo-initiation or temperature-jump, transient 2D IR spectroscopy will offer insights into the dynamics of proton transfer during tautomerization. Such studies will help to explain phenomena such as preferential ligand recognition, spontaneous mutation, and enzyme catalytic functions.

5.3 Observation of oxythiamine tautomers

5.3.1 Background of oxythiamine and its recognition by the TPP riboswitch

After demonstrating the capability of 2D IR spectroscopy in identifying the various tautomers in model heterocycles, we wish to apply this method to study molecules of biological significance. Riboswitches are structured RNA regulatory elements found in the non-coding region of many bacterial mRNAs. They bind to small ligands with high affinity and specificity to regulate the expression of downstream encoded genes. Tautomerism has been predicted to influence the ligand recognition by RNA aptamers. For example, the purine riboswitch, which regulates genes involved in guanine metabolism, has been suggested to bind to the enol tautomer of xanthine whereas the thiamine pyrophosphate (TPP) riboswitch has been proposed to recognize oxythiamine pyrophosphate (OxyTPP), a model ligand, in its enol form instead of the canonical keto tautomer.

Thiamine pyrophosphate (TPP, Figure 5.13a) is an important cofactor in the catabolism of sugars and amides by catalyzing the transfer of aldehyde or acyl groups. It is one of the phosphate derivatives of thiamine, commonly known as vitamin B1, which is composed of thiazolium and 4'-aminopyrimidine rings. The TPP riboswitch (TPP, Figure 5.13b) binds to TPP and negatively regulates the expression of genes involved in the biosynthesis. Structural and biochemical studies on TPP riboswitches with the bound ligand have revealed the importance of the hydrogen-bond (HB) interactions between guanine and the 4'-aminopyrimidine ring of TPP. These HB interactions are highly sensitive to the tautomerism of the bound molecules as the proton can be transferred between different nitrogen and oxygen atoms on the ring. Furthermore, it was suggested that the binding pockets of the RNA riboswitches may be formed via a ligand-mediated induced-fit mechanism following the molecular recognition of the binding ligands.
The X-ray crystal structure of the riboswitch with the TPP ligand shows that the amino group at the 4'-position of TPP act as a hydrogen bond donor to the N3 position of G28 (Figure 5.14a). Similar structural studies of the TPP riboswitch and the oxidized form of TPP, oxythiamine pyrophosphate (OxyTPP), indicated that its hydrogen bonding interactions to G28 are almost identical to those of TPP. Since the 4'-position of OxyTPP can only act as a hydrogen bond donor in the enol form (Figure 5.14b, c, d), it was proposed that OxyTPP exists as an enol tautomer in the bound form. Nevertheless, given that the resolution of the crystal structure was not enough (2.6Å) to determine the precise proton position, the exact tautomer bound to TPP is still under debate.

The thiamine pyrophosphate riboswitch bound to OxyTPP provides a good model system for comprehensively studying tautomerism in nucleic acid complexes because structural studies have predicted that the riboswitch binds to the minor tautomeric form of OxyTPP. OxyTPP can theoretically exist in three tautomeric forms depending on the position of the active proton (Figure 5.14b, c, d); two 4'-keto forms that are protonated either at the N1'-position (4'-keto-N1'H-OxyT) or at the N3'-position (4'-keto-N3'H-OxyT) and a 4'-enol (4'-enol-Oxy T) form. In this study we characterize the various tautomeric forms of oxythiamine (not the diphosphorylated form) in solution using 2D IR spectroscopy.
Figure 5.14: (a) Comparison of hydrogen bonding interactions between TPP and G28 (guanine at the 28th position) of the TPP riboswitch (b, c, d) and various tautomers of OxyTPP and G28. R stands for pyrophosphate and R' stands for riboswitch backbone.

5.3.2 Oxythiamine tautomers in the unbound form

5.3.2.1 Characterization using infrared spectroscopy

Following the approach outlined in the pyridone work, we performed variable temperature FTIR and 2D IR experiments to investigate the tautomerism of oxythiamine in TPP buffer (1M HEPES buffer pH 7.5, 100 mM KCl and 15 mM MgCl₂). The variable temperature FTIR spectra of OxyT and the 2D IR spectrum at 10°C in the region of in-plane double bond vibrations for aromatic heterocycles are shown in Figure 5.15a and Figure 5.15b, respectively. Vibrational bands in this region, which we number from high to low frequency, are expected to have a distinct pattern for each tautomer, in particular when the C=O character is altered significantly. The X1 mode is composed of X1a at 1658 cm⁻¹ with a shoulder (X1b) at 1646 cm⁻¹. The X1 mode was assigned to the carbonyl stretch in the keto form based on the characteristic vibrational frequency, high IR intensity, and the broad line-shape. Since a single tautomer of aromatic heterocycles should exhibit intense cross-peaks between all of the diagonal peaks, the contrast between the pronounced X1a/X4 cross-peaks at (ω₁, ω₄) = (1658 cm⁻¹, 1525 cm⁻¹) and the absence of X1b/X4 cross-peaks clearly indicate that X1a and X1b originate from two different species.
The DFT frequency calculations predicted that the C=O frequency of the 4'-keto-N3'H-OxyT tautomer is slightly lower than that of the 4'-keto-N1'H-OxyT, suggesting that the X1b shoulder is from the C=O stretch of the 4'-keto-N3'H-OxyT (Figure 5.15c, Figure 5.16). The calculations also show that both keto tautomers have a peak at ~ 1545 cm\(^{-1}\) corresponding to the pyrimidine ring vibrations with strong C=N character. This peak matches well with the experimental peak at 1558 cm\(^{-1}\) (X3), which has cross-peaks to both C=O peaks. Furthermore, the calculations showed that the 4'-keto-N1'H-OxyT tautomer has noticeable methyl group vibrations that contribute to the 1450 cm\(^{-1}\) band, whereas the 4'-keto-N3'H-OxyT tautomer has a much weaker peak at 1450 cm\(^{-1}\). The intensity difference in the methyl vibration manifested itself in the 2D IR spectrum where cross-peaks were observed between X1a/X4 but not between X1b/X4. Finally, an unusually broad peak (X2) was seen at 1598 cm\(^{-1}\), which also has cross-peaks to X4. Again comparing it to the DFT calculations, X2 was assigned to be the collective ring vibration of the enol tautomer that involves the C=N stretches as well as the O-D stretch coupled to the water vibrations (Figure 5.16c). The 4'-enol-OxyT tautomer had an even more intense methyl vibration than the 4'-keto-N1'H-OxyT tautomer, which resulted in the cross-peaks between X2/X4.

Figure 5.15: (a) FTIR spectra of OxyT in TPP buffer taken at various temperatures (5°C – 90°C). (b) 2D IR spectrum of OxyT under the same solvent condition at 10°C. (c) DFT calculated IR spectra for the three OxyT tautomers using QChem.
distinct cross-peak patterns in the 2D IR spectrum enabled us to identify and distinguish among all three tautomers of OxyT under relatively physiological conditions.

Figure 5.16: DFT calculated normal modes of (a) 4'-keto-N1'H-OxyT; (b) 4'-keto-N3'H-OxyT; and (c) 4'-enol-OxyT.
Since the three tautomers have distinct peaks at 1658 cm\(^{-1}\), 1646 cm\(^{-1}\), and 1598 cm\(^{-1}\) for the 4'-keto-N1'H-OxyT, 4'-keto-N3'H-OxyT, and 4'-enol-OxyT, respectively, their relative populations can be extracted out by simultaneously fitting the FTIR and 2D IR spectra in the 1580 – 1700 cm\(^{-1}\) frequency range.\(^{41,42}\) To calculate and fit the experimental FTIR and 2D IR spectra, a nonlinear response function was used and described in detail previously.\(^{42,43}\) Since the three tautomers of OxyT have distinct vibrations in the 1580–1680 cm\(^{-1}\), only the experimental spectra in this frequency range were fit. Before fitting the FTIR spectra, we applied a linear correction from 1580 to 1700 cm\(^{-1}\) to the temperature dependent OxyT FTIR spectra after the temperature dependent buffer spectra have been subtracted. Three uncoupled anharmonic oscillator with Kubo line-shape were used to fit both the FTIR and 2D IR spectra. The Kubo line-shape is calculated from an energy-gap correlation function which contains a single exponential decay \(\Delta \exp(-|t|/\tau)\), and a \(\delta(t)/T_2\) term to produce the homogeneous linewidth in the 2D IR spectra.\(^{44}\) A simultaneous fit of both the FTIR and 2D IR spectra well reproduced the experimental data (Figure 5.17) with the fitting parameters listed in Table 5.3.

In this model since the three anharmonic oscillators are uncoupled, no cross-peaks should be present in the calculated spectrum. The off-diagonal features seen in the experimental spectrum are from the tails of the Lorentzian peaks extending to the X1/X3 cross-peaks (see Figure 5.15) which do not exist in the calculation. In addition, the DFT calculations show that both the keto tautomers may have contributions to the X2 mode as

![Figure 5.17](image)

Figure 5.17: (a) Experimental FTIR of OxyT in TPP buffer taken at room temperature (blue curve), and calculated FTIR (red curve) using the fitting parameters listed in Table 5.3. (b) Experimental (left) and fitted (right) 2D IR spectra.
vibrations corresponding to the thiazole ring vibrations (Figure 5.16) are observed at 1560 cm\(^{-1}\). We have used a model that includes the additional mode at 1600 cm\(^{-1}\) for the N1'H tautomer (Figure 5.18), and found that its contribution is very weak and can be neglected. The errors in determining the tautomer populations from neglecting these contributions are small and within the error bars reported in this study.

Table 5.3. Fitting parameters from simultaneous fitting of experimental FTIR and 2D IR at room temperature. The model includes three independent oscillators representing the three tautomers. \(\omega\) is the transition frequency; \(\Delta\omega\) is the diagonal anharmonicity; \(\mu_{0}\) is the transition dipole strength for the 0-1 quantum states; \(\mu_{21}\) is the transition dipole strength for the 1-2 quantum states; \(T_2\), \(\Delta\), and \(\tau\) are the lineshape parameters in the Kubo model:

\[
g(t) = t / T_2 + \Delta^2 \tau^2 \left[ \exp(-t / \tau) - 1 + t / \tau \right].
\]

<table>
<thead>
<tr>
<th></th>
<th>4'-keto-N1'H-OxyT</th>
<th>4'-keto-N3'H-OxyT</th>
<th>4'-enol-OxyT</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>0.5</td>
<td>0.42</td>
<td>0.08</td>
</tr>
<tr>
<td>(\omega)</td>
<td>1658</td>
<td>1646</td>
<td>1597</td>
</tr>
<tr>
<td>(\Delta\omega)</td>
<td>8.5</td>
<td>12.9</td>
<td>19</td>
</tr>
<tr>
<td>(\mu_{0})</td>
<td>0.7</td>
<td>0.68</td>
<td>0.74</td>
</tr>
<tr>
<td>(\mu_{21})</td>
<td>1.12</td>
<td>0.93</td>
<td>1.29</td>
</tr>
<tr>
<td>(T_2)</td>
<td>1515</td>
<td>600</td>
<td>2066</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>45.6</td>
<td>40.1</td>
<td>52.8</td>
</tr>
<tr>
<td>(\tau)</td>
<td>617</td>
<td>80.7</td>
<td>593.6</td>
</tr>
</tbody>
</table>

In order to determine the error bars on the tautomer populations, we tested the robustness of the two keto populations in reproducing the FTIR spectrum. We fit the FTIR spectrum by fixing \(\mu_{0}\) from the simultaneous fit and either one of the keto populations, and then allowed the other parameters (another keto population and line-shape parameters) to float. We stepped the population of 4'-keto-N1'H-OxyT from 25% to 75%, and calculated the residual relative to the experimental spectrum. We then defined the acceptance for a good fit to be 20% deviation of the relative residual from the best fit. We found that the experiment can be well reproduced with the population of 4'-keto-N1'H-OxyT varying between 50% to 60%. Similar procedure was done for the population of 4'-keto-N3'H-OxyT and it was found to be between 30% to 44%. The population of the enol tautomer was found to vary between 7% to 9%. The result shows about 14% error in determining the keto populations as opposed to the 2% error in enol population. The overlap between the two keto peaks leads to the larger uncertainty. Other sources of error include baseline

142
Figure 5.18: Experimental and simulated FTIR and 2D IR spectra, following the same formulism as the simulation shown in Figure 5.17. However, one additional peak at 1600 cm$^{-1}$ has been included in this model as a weak vibration of the N1'H tautomer to account for the weak cross-peak at (1600 cm$^{-1}$, 1658 cm$^{-1}$). Hypobolic sine contours are plotted to better show the weak cross-peaks.

uncertainty in the FTIR spectra, lineshapes, and the fact that we may be neglecting keto contributions to the X2 mode. However, self-consistent fitting the 2D IR spectrum in addition to the FTIR enhances our confidence limits.

A self-consistent fitting of the FTIR and 2D IR spectra at room temperature resulted in 8% ± 1% 4'-enol-OxyT tautomer, 55% ± 5% 4'-keto-N1'H-OxyT and 37% ± 7% 4'-keto-N3'H-OxyT tautomers (Figure 5.17). The variable temperature FTIR spectra show that X1a intensifies at the expense of the X1b shoulder as the temperature is raised, whereas X2 has weak temperature dependence. Fitting the variable temperature FTIR spectra using the spectroscopic parameters from the room temperature spectra showed that the 4'-keto-N1'H-OxyT population increased to ~ 70% at 90 °C, the 4'-keto-N3'H-OxyT population decreased to ~ 20%, whereas the 4'-enol-OxyT tautomer was relatively constant within the error bars (Figure 5.19 and Figure 5.20).
Figure 5.19: Variable temperature FTIR spectra from 5 °C to 90 °C: (a) experiment; (b) fit. The fitting procedure is the same as described for fitting the room temperature spectrum. The transition dipole strengths were assumed to be independent of temperature. Transition frequencies were allowed to float within ± 3 cm⁻¹ from the values determined at room temperature to account for the temperature induced frequency shifts.

Figure 5.20: Temperature dependence of populations of tautomers of OxyT. The populations were obtained from fitting the variable temperature FTIR. Representative error bars are shown at 90 °C and are roughly independent of temperature.

5.3.2.2 Characterization using NMR spectroscopy

Because tautomerization of aromatic heterocycle are expected to occur on nanosecond time scales under physiological conditions, traditional NMR approaches could not resolve tautomeric forms of OxyT. Nevertheless, NMR spectroscopy can still provide important and complementary structural information in certain solvents and temperature range where the tautomer exchange is slow. Figure 5.21a displays the variable temperature
1H NMR spectra in DMF and at low temperatures. At -60 °C, three distinct proton resonances between 12.5 to 17.0 ppm were observed that were attributable to protons bound either to N1' or N3' or to the 4'-oxygen on the pyrimidine ring. Considering that each tautomer can only have the proton at one of these three positions, the presence of three proton resonances indicated the presence of three OxyT tautomers in DMF solvent at -60 °C.

Comparing to the literature, the proton resonance at 16.6ppm was assigned to the N3' amide proton in the 4'-keto-N3'H-OxyT tautomer, 15.5ppm to the 4'-enol-OxyT proton, and 13.3ppm to the N1' amino proton in the 4'-keto-N1'H-OxyT tautomer. Integrated peak intensities allowed determination of the relative ratios of the three tautomeric species that exist in DMF at low temperature (4'-keto-N3'H-OxyT: 4'-enol-OxyT: 4'-keto-N1'H-OxyT = 0.34 : 0.38 : 0.28). Additionally, the separation of chemical shifts for three tautomers provides estimate of the timescale of tautomer exchange: the first appearance of three distinguishable peaks of the tautomeric proton is -50°C (the coalescence temperature for the tautomer exchange). The exchange rate between peak 16.6 ppm and 15.5 ppm can be calculated using the equation $\sqrt{2 / (\pi \Delta \nu)} = \sqrt{2 / (\pi \times 1.1 \text{ppm} \times 500 \text{Hz/ppm})}$, which is 0.82 ms. The other two exchange rates were calculated accordingly: the exchange rate between 15.5 ppm and 13.3 ppm is 0.41 ms and the exchange rate between 16.6 ppm and 13.3 ppm is 0.27 ms. It is clear that at higher temperatures, the exchange becomes too fast for NMR to resolve individual peaks. In contrast to OxyT, variable temperature NMR data for thiamine indicates that it is present in solution as a single 4'-amino tautomer (Figure 5.21b).

![Diagram](image_url)

**Figure 5.21:** Variable temperature 1H NMR spectra of (a) oxythiamine and (b) thiamine in DMF-d7 (20°C to -60°C).
5.3.3 Tautomeric form of OxyTPP bound to the TPP riboswitch

The spectroscopic approaches described above allowed us to characterize the tautomeric forms of OxyT in the unbound form. Given the complexity of the TPP aptamer, these approaches could not be directly applied to determine the tautomeric form of OxyTPP bound to the riboswitch. Instead, we studied the tautomerism of bound OxyTPP using binding isotope effects (BIEs), which characterize the increase or decrease in binding strength upon substitution of an atom with its heavier isotope. \(^4\) BIEs are useful as they are sensitive to change in bond order between the bound and the unbound states of a ligand. In general, if a bond order of the isotope labeled atom decreases upon binding, a BIE of greater than 1.0 is observed. Vice versa, an inverse BIE (BIE < 1.0) is indicative of an increase in bond order upon binding. The magnitude of the effect is correlated with the change in bond order. BIEs have been used to study enzyme-substrate interactions in enzymes such as dehydrogenases and to detect subtle changes in bonding in glucose when bound to the hexokinase enzyme.\(^{45,49}\)

The \(^18\text{O} \text{BIE}\) at the 4'-position of OxyTPP, \(^4\text{O} - 18\text{O} \text{BIE}\), along with DFT calculations were used to establish the tautomeric form of OxyTPP bound to the TPP riboswitch. An inverse \(^4\text{O} - 18\text{O} \text{BIE}\) value of 0.984 was measured which suggested that the binding of OxyTPP to the TPP riboswitch resulted in increased bonding to the 4'-oxygen of OxyTPP. This observation is inconsistent with the proposal of an enol tautomer in the bound form given the fact that the N1'H keto tautomer is the predominant form of the unbound OxyTPP. To first order approximation, transition from the 4'-carbonyl group to the hydroxyl group results in decrease in bond order. As a consequence, the BIE results suggest that the bound form of OxyTPP should be one of the two keto tautomers (distinction between the two cannot be derived from this result).

Additionally, we measured the dissociation constant of OxyTPP bound to the TPP riboswitch to be 1.67 \(\mu\text{M}\), which is an approximately 10X to 200X increase over the natural ligand TPP (~ 9-200 nM depending on the Mg\(^{2+}\) concentration).\(^{50,51}\) The crystal structure of the TPP riboswitch with its ligand predicts that the pyrimidine ring of TPP forms two hydrogen bonds with G28, between the exocyclic 4'-NH\(_2\) group of TPP and N3 of G28 and N3 of TPP and 2-NH\(_2\) of G28 (Figure 5.14a).\(^{36}\) The 4'-OxyTPP in the keto form is expected to bind unfavorably compared to TPP due to the potential loss of hydrogen bonding with N3 of G28 (Figure 5.14c, d). Furthermore, the N3'H keto tautomer cannot form favorable hydrogen-bonds with G28 in the geometry solved from the X-ray crystal structure due to significant geometric constraints (clash between the N3'H and 2-NH\(_2\) of G28). Therefore it is suggested that the bound form of OxyTPP is 4'-keto-N1'H-OxyTPP.
Figure 5.22: (a) Proposed model for the enol tautomer of OxyTPP bound to G28 of the TPP riboswitch. (b) Proposed models based on BIE showing interactions for the 4'-keto-N1'H-OxyTPP tautomer with protonated G28. (c) Proposed models based on BIE showing interactions for the 4'-keto-N1'H-OxyTPP tautomer with G28 in the imino form. All three models would be consistent with the crystal structure of the riboswitch with OxyTPP.

Taken together, the spectroscopic data show that 4'-keto-N1'H-OxyT is the predominant tautomeric state in the unbound form followed by 4'-keto-N3'H-OxyT and then 4'-enol-OxyT. In the bound form, the BIE data is consistent with the presence of 4'-keto-OxyTPP tautomer. The structural data indicates that the bound OxyTPP is unlikely to be the N3'H keto tautomer. As a consequence, it is suggested that bound form is the 4'-keto-N1'H-OxyTPP tautomer, which is also the predominant tautomer in the unbound state. This result, nevertheless, does not fit perfectly with the observation from the crystal studies which showed that the binding interaction and geometry are almost identical for both TPP and OxyTPP. Identical hydrogen-bond interactions can be achieved with the 4'-enol-OxyTPP tautomer (Figure 5.22a) but not with the 4'-keto-N1'H-OxyTPP tautomer due to the loss of hydrogen-bond to the N3 atom (Figure 5.14c), unless G28 is protonated (Figure 5.22b) or in the imino form (Figure 5.22c). Presence of altered guanine charge states or minor tautomeric forms has been predicted in other RNA systems such as small self-cleaving ribozyme and riboswitches for instance Hammerhead, Varkud Satellite, Hairpin and GlmS ribozymes.34,35,37,52,53 Future structural studies are required to obtain more detailed information on the ionic or tautomeric state of G28.
5.4 Dimerization of pyridone in organic solvents

In the previous sections, we have shown that 2D IR spectroscopy opens a new avenue for studying the tautomerism of aromatic heterocycles in aqueous solutions based on the distinct cross-peaks arising from the various tautomers. However, we wish to point out that once the peak assignments are made with the help of 2D IR spectra, more quantitative measurements can be achieved by analyzing the FTIR spectra because of the higher signal-to-noise ratio and the linear dependence on the concentration. That is to say, if peak assignments are straightforward, FTIR spectra alone provide a convenient way for us to extract out information on tautomeric systems. This may be achieved when studying tautomerism in organic solvents whose vibrations do not overlap significantly with important functional groups that are used to identify certain tautomers, such as the OH stretch in enol tautomers. Even though the tautomerism of heterocycles changes upon the solvent environment, it would still be informative to study the solvent effect. Nevertheless, a different technical difficulty other than background contamination is often encountered when studying tautomerism of heterocycles in organic solvents—dimerization. To illustrate the structural differences due to the solvent, we present the results of pyridone derivatives in organic solvents in this section.

5.4.1 Lactam dimers of 2-pyridone

FTIR spectra of 2-pyridone (Figure 5.23) in different solvents show dramatic changes in peak positions, intensity and lineshapes, especially when compared to the spectrum taken in D$_2$O where 2-pyridone was found to be primarily in the lactam tautomer. The differences in the FTIR spectra demonstrate the significant effect of solvents on the tautomeric equilibria between the keto, enol, and different dimers. In the 6 μm region (Figure 5.23a), peaks above 1650 cm$^{-1}$ are assigned to the carbonyl stretches, and their frequencies are sensitive to the HB environment to the C=O. In D$_2$O, the carbonyl group of 2-pyridone can accept up to two HBs from D$_2$O molecules (Figure 5.3), which leads to a big redshift. In organic solvents, the C=O stretch was found to be at higher frequencies and sometimes split into multiple peaks, suggesting that the HB interaction due to the solvent is weaker and that dimers may be present which give rise to the symmetric and asymmetric coupled C=O stretches. The peaks below 1650 cm$^{-1}$ are typically ring modes consisting C=N and C=C double bond stretches.
2-pyridone is known to form cyclic dimers in non-aqueous solvents. The presence of the cyclic dimer can be identified from the 3 μm NH stretching region (Figure 5.23b). When the cyclic dimers form, the NH groups are strongly hydrogen-bonded to the C=O (Figure 5.24), thus red-shifting the NH stretching band. The FTIR spectra of 2-pyridone in CD₂Cl₂ and CCl₄ show a broad band centered around 2900 cm⁻¹ due to the strongly hydrogen-bonded NH stretches in the cyclic dimer, as well as C-H stretching modes. The fine structure on top of the broad peak is due to the Fermi resonances to the combination or overtone bands of fingerprint modes, which is commonly seen in strongly hydrogen-bonded systems. On the other hand, a narrower peak at around 3400 cm⁻¹ is observed in the FTIR of 2-pyridone in MeCN-d₃ and DMSO-d₆, indicating the presence of monomers. Comparing the four spectra, we conclude that 2-pyridone has the largest dimer fraction in CCl₄ and the smallest dimer fraction in DMSO-d₆. We will later present the 2D IR spectra in the 6 μm region to confirm this observation.

2-pyridone can in principle undergo lactam-lactim tautomerization to form 2-hydroxypyridine. Since the lactam tautomer has larger dipole moment, it is believed that the lactam form is more stable in polar solvents whereas the lactim form is favored in nonpolar solvents. However, no trace of free OH stretching was observed in the 3 μm region of CCl₄ spectrum, we thus conclude the absence of lactim tautomers even in nonpolar solvent at the concentrations used in our experiments. UV-vis spectra of 2-pyridone in H₂O, MeCN, and CCl₄ were taken (Figure 5.25), and showed no evidence of the lactim tautomer at 260 nm.
Figure 5.24: In non-aqueous solvents, 2-pyridone can exist as 2-pyridone (PD) cyclic dimer or PD chain where only some of the carbonyl or NH groups are hydrogen-bonded. Similarly, 6-chloro-2-hydroxypyridine (HP) can form cyclic dimer, or a chain of lactim tautomers.

Figure 5.25: UV-vis of 2-pyridone in various solvents.

The temperature-dependent FTIR spectra of 2-pyridone in MeCN-d₃ are shown in Figure 5.26. As the temperature increases, the peak at 1660 cm⁻¹ decreases dramatically, suggesting that it is the dimer C=O peak since dimer dissociation is sensitive to temperature. On the other hand, the 1676 cm⁻¹ peak with relatively constant intensity is likely to be from the free carbonyl stretch. The 2D IR spectra of 2-pyridone in various solvents are shown in Figure 5.27. In the CCl₄ spectrum, intense cross-peaks between all the diagonal peaks are observed, indicating that these modes arise from the same chemical species. This observation is in agreement with the assumption that 2-pyridone exists primarily in the dimer state in CCl₄. The 1683 cm⁻¹ and 1660 cm⁻¹ peaks are likely to be from the dimer C=O peaks. The 1621 cm⁻¹ and 1546 cm⁻¹ peaks correspond to the ring modes.
Figure 5.26: Temperature-dependent FTIR spectra of 2-pyridone (5mg/ml) in MeCN-d$_3$.

Figure 5.27: 2D IR spectra of 2-pyridone in different solvents: (a) MeCN-d$_3$; (b) DMSO-d$_6$; (c) CD$_2$Cl$_2$; (d) CCl$_4$. All spectra were acquired with ZZZZ polarization.

Spectral features reflecting the melting of 2-pyridone dimer can be clearly seen in Figure 5.28 which shows spectra of 2-pyridone in MeCN-d$_3$ at various temperatures. Dimer melting leads to the intensity drop of the 1660 cm$^{-1}$ peak and the disappearance of the cross-
peak (purple arrow) at \((\omega_1, \omega_3) = (1684 \text{ cm}^{-1}, 1660 \text{ cm}^{-1})\). The higher frequency diagonal peak is not apparent in the MeCN-d₃ spectrum due to the presence of the free monomer C=O peak at 1676 cm⁻¹; however, it is clearly seen in the CCl₄ spectrum. The 1684 cm⁻¹ peak has weak intensity and can be assigned to the Fermi resonance of some combination or overtone bands of the lower frequency modes. Complicated infrared spectra consisting of extra peaks in the carbonyl stretching region have been reported before. Alternatively, the 1684 cm⁻¹ peak can be assigned to the other C=O dimer peak. From DFT calculation, the symmetric C=O stretch in a dimer has zero intensity due to the planarity of the dimer and high symmetry. However, this band is expected to gain some intensity when the symmetry is broken. If this is the case, from the C=O frequencies of the dimer peaks and the free monomer, the coupling constant and the frequency shift due to dimer HB can be calculated to be 12 cm⁻¹ and 4 cm⁻¹, respectively.

The 2D IR spectrum of 2-pyridone in CD₂Cl₂ is similar to the CCl₄ spectrum at first glance. However, temperature dependent 2D IR spectra (Figure 5.29) show the presence of the 2⁻ species. As the temperature increases, the diagonal peaks at 1683 cm⁻¹, 1658 cm⁻¹, and 1619 cm⁻¹ (blue grids) decrease as well as the cross-peaks between them. These peaks correspond well with the peaks seen in the CCl₄ spectrum, thus representing the melting of the dimer. Interestingly, another set of peaks (peaks with cross-peaks between them are considered as a set, red grids) grow in at 1675 cm⁻¹, and 1611 cm⁻¹. This observation is consistent with previous experiments done by Szyc and coworkers. The authors assigned
these peaks to the C=O stretch and ring mode of 2-pyridone hydrogen-bonded to two CD$_2$Cl$_2$ molecules.

The 2D IR spectrum of 2-pyridone in DMSO-d$_6$ also shows cross-peaks between all the diagonal peaks, indicating that they originate from the same species. Since the intensity ratio of the two C=O peaks is very different from that in the CCl$_4$ spectrum, and the fact that little amount of the broad band at ~2900 cm$^{-1}$ was observed, we ruled out the possibility that 2-pyridone exists in the dimer state in DMSO-d$_6$. As a consequence, the DMSO-d$_6$ spectrum may very well represent the free monomer spectrum. These peaks at 1670 cm$^{-1}$, 1651 cm$^{-1}$, and 1611 cm$^{-1}$ match well with the 2nd species seen in the CD$_2$Cl$_2$ spectrum, therefore, we assign them to the same species. However, this assignment rules out the possibility that this species is the 2-pyridone monomer since two C=O stretches are observed at 1670 cm$^{-1}$ and 1655 cm$^{-1}$. The presence of the 1655 cm$^{-1}$ peak in CD$_2$Cl$_2$ was suggested by Szyc and coworkers, and assigned to the C=O that is hydrogen-bonded in a 2PD-chain (see Figure 5.24 for structure). This peak was not observed in our spectrum most likely due to the overlap with the dimer C=O peak at 1660 cm$^{-1}$. Additionally, the concentration of 2-pyridone used in our experiment is four-fold smaller than theirs (50 mM versus 200 mM), therefore the dimer and 2PD-chain fractions are likely different.

The 2D IR spectrum of 2-pyridone in MeCN-d$_3$ has the most spectral features. The 1st set of peaks at 1684 cm$^{-1}$, 1660 cm$^{-1}$, and 1619 cm$^{-1}$ has cross-peaks between all three modes, showing the presence of cyclic dimers. The 2nd set of peaks at 1676 cm$^{-1}$, 1655 cm$^{-1}$, and 1613 cm$^{-1}$ correspond to the species seen in the DMSO-d$_6$ spectrum. Furthermore, as the temperature increases, two new species were identified (indicated by the purple and
green grids in Figure 5.30). The 3rd set of peaks at 1664 cm\(^{-1}\) and 1596 cm\(^{-1}\), and the 4th set of peaks at 1643 cm\(^{-1}\) and 1563 cm\(^{-1}\) were observed. Since these peaks only appear at high temperatures, there is a bigger chance of them being the monomers as opposed to 2PD-chains suggested by Szyc and coworkers.\(^{15}\) Szyc and coworkers’ DFT calculations show that the C=O frequencies span a wide range depending on the structure of 2-pyridone and the HB environment around the carbonyl.\(^{55}\) Comparing different sets of data, we assign the 1664 cm\(^{-1}\) peak to the monomer C=O hydrogen-bonded to one solvent molecule, and the 1643 peak to the monomer C=O hydrogen-bonded to two solvent molecules, which can be compared to the 1643 peak in D\(_2\)O. The 1596 cm\(^{-1}\) and 1563 cm\(^{-1}\) peaks are assigned to the ring modes corresponding to the two abovementioned species. Furthermore, the 2nd species that is also seen in CD\(_2\)Cl\(_2\) and DMSO-d\(_6\) is attributed to the PD-chain where the 1676 cm\(^{-1}\) peak is due to the free C=O and the 1655 cm\(^{-1}\) peak is from the hydrogen-bonded C=O. It is not surprising to see couplings between the two carbonyls which lead to the presence of cross-peaks, as couplings were observed in G-C carbonyl stretches,\(^{59,60}\) which are in similar geometry as the 2-pyridone chain-like structure.

![Figure 5.30: 2D IR spectrum of 2-pyridone (5 mg/ml) in MeCN-d\(_3\) at 50°C with ZZZZ polarization. This figure shows the appearance of 2-pyridone in the monomeric forms at high temperature. Purple grids indicate the positions of the monomer hydrogen-bonded to one solvent molecule: C=O at 1664 cm\(^{-1}\) and the ring mode at 1596 cm\(^{-1}\). Green grids show the positions of the monomer hydrogen-bonded to two solvent molecules: C=O at 1643 cm\(^{-1}\) and the ring mode at 1563 cm\(^{-1}\). The highest contours of the cross-peaks are highlighted with thicker black lines.](image)
These 2D IR spectra of 2-pyridone in non-aqueous solvents show that 2-pyridone exists in complicated equilibrium between the cyclic dimers, 2-pyridone-solvent complex, and chain-like dimers. On contrary, 2D IR spectrum of 2-pyridone in D$_2$O (Figure 5.1a) shows that only one species is present, and was assigned to the lactam monomer. No evidence of the lactim tautomer was observed in these any of the solvents.

To provide further evidence for the spectral assignment made based on the experimental spectra, we performed DFT calculations of 2-pyridone in various structural forms shown in Figure 5.31, and the calculated frequencies are tabulated in Table 5.4.

![Figure 5.31: DFT [B3LYP/6-31G (d,p)] optimized structures of 2-pyridone in monomer or dimer states. The explicit solvent molecules included were MeCN-d$_3$ (a, b) and D$_2$O (c).](image)

155
Table 5.4. Experimentally observed peaks for 2-pyridone in various solvents and the DFT (B3LYP/6-31G (d,p)) calculated frequencies for different species in MeCN-d3. The frequencies have been scaled by 0.9614. The optimized structures are shown in Figure 5.31.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>DFT</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeCN-d3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMSO-d6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD2Cl2-d2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCl4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer +MeCNd3</td>
<td>1684</td>
<td>Fermi resonance or dimer sym C=O</td>
</tr>
<tr>
<td>Monomer +2MeCNd3</td>
<td>1676</td>
<td>2PD chain C=O (free)</td>
</tr>
<tr>
<td>Chain</td>
<td>1664</td>
<td>Monomer C=O HB to 1 solvent</td>
</tr>
<tr>
<td>Monomer +MeCNd3</td>
<td>1660</td>
<td>Dimer C=O, asym</td>
</tr>
<tr>
<td>Cyclic dimer</td>
<td>1667</td>
<td>2PD chain C=O (HB)</td>
</tr>
<tr>
<td>Dimer</td>
<td>1655</td>
<td>Monomer C=O HB to 2 solvent</td>
</tr>
<tr>
<td>2PD</td>
<td>1643</td>
<td>C=O HB to 2 water</td>
</tr>
<tr>
<td>Monomer</td>
<td>1619</td>
<td>Dimer ring mode 1</td>
</tr>
<tr>
<td>solvent, ring mode 1</td>
<td>1613</td>
<td>Monomer ring mode 1</td>
</tr>
<tr>
<td>Monomer</td>
<td>1596</td>
<td>Monomer: 2 solvent, ring mode 1</td>
</tr>
<tr>
<td>2solvent, ring mode 1</td>
<td>1563</td>
<td>Monomer: 2 solvent, ring mode 1</td>
</tr>
<tr>
<td>Monomer in water ring mode 1</td>
<td>1560</td>
<td>Monomer in water ring mode 1</td>
</tr>
<tr>
<td>Monomer, dimer ring mode 2</td>
<td>1542</td>
<td>Monomer, dimer ring mode 2</td>
</tr>
</tbody>
</table>
The assignments made in Table 5.4 are tentative assignments based on the
eperimental spectra. It is clear that some peaks do not match well with the DFT calculated
frequencies. For example, the symmetric C=O stretch of the cyclic dimer was calculated to
be at 1632 cm\(^{-1}\) with zero intensity. Therefore, if the calculation reflects the reality, the
experimental peak at 1683 cm\(^{-1}\) could only be explained as the Fermi resonance. Moreover,
there is significant mismatch in the frequencies for the two species observed in the high
temperature spectrum which we assigned to the monomer hydrogen-bonded to either one or
two solvent molecules. The monomer assignment was made based on the fact that they only
appear at high temperatures. A better calculation model would be required to confirm this
assignment. Alternatively, the recently developed broadband 2D IR spectroscopy\(^{51,62}\) can be
used to study the coupling between the carbonyl and HN groups that will provide more
information on the HB interactions and help differentiate between monomer, chain-like,
dimer or even higher-order aggregate species.

5.4.2 Lactim dimers of 6-chloro-2-pyridone

As discussed in Section 5.2, substituents on aromatic heterocycles can shift the
tautomeric equilibrium, and the addition of chloride at the 6 position was found to increase
the lactim population.\(^{19,22}\) We have found that 32% of 6-chloro-2-pyridone exists in the
lactim form in aqueous solution at room temperature. From the UV absorption spectra
shown in Figure 5.32, we can see that the lactim population is bigger in non-aqueous
solution such as THF (lactim peak is blueshifted compared to lactam). Similar conclusions
can be drawn from the infrared spectra of 6-chloro-2-pyridone in various solvents (FTIR
shown in Figure 5.33 and 2D IR shown in Figure 5.34). The FTIR spectra show dramatic
decrease of C=O stretch intensity and a much more complicated peak pattern in the 1550
\(\text{cm}^{-1}\) to 1600 \(\text{cm}^{-1}\) region. The 2D IR spectra clearly indicate the presence of at least two
distinct species in all solvents. The peaks enclosed by the blue box include the C=O

![Figure 5.32: UV-vis absorption of 50 \(\mu\text{g/ml}\) 6-chloro-2-pyridone in THF and H\(_2\)O.](image-url)
stretching mode and thus correspond to the lactam tautomer vibrations. On the other hand, the peaks highlighted by the magenta box arise from the lactim tautomer. We will start with the assignments of the lactam peaks followed by the assignment on the lactim peaks since the lactam spectra have great resemblance to those of 2-pyridone.

Figure 5.33: FTIR spectra of 5mg/ml 6-chloro-2-pyridone in various solvents.

Figure 5.34: 2D IR spectra (ZZZZ) of 6-chloro-2-pyridone in different solvents: (a) MeCN-d₃; (b) DMSO-d₆; (c) CD₂Cl₂; (d) CCI₄; (e) THF. The blue and magenta boxes are intended to highlight the vibrational modes from the lactam and lactim species, respectively.
Even though the C=O stretching region may seem similar in the five solvents at first glance, a closer examination reveals that varied number of C=O stretching modes are observed which indicate that different species are present in these solvents. First of all, temperature-dependent FTIR spectra in CCl₄ (Figure 5.35a) do not show significant decrease of the 1669 cm⁻¹ peak, arguing against the assignment of lactam dimer whose formation should be sensitive to temperature as seen in the 2-pyridone case. Therefore, we can assign the 1669 cm⁻¹ peak to be from the free C=O of a chain-like structure, and the 1651 cm⁻¹ peak to be the hydrogen-bonded C=O of the same species.

In MeCN-d₃ and THF, a peak at 1679 cm⁻¹ was observed with high intensity. The higher frequency and strong intensity suggest that it is due to the lactam monomer C=O stretch (instead of the Fermi resonance seen in the 2-pyridone dimer). Additionally, the temperature-dependent FTIR spectra in MeCN-d₃ (Figure 5.35b) show that upon heating, the 1650 cm⁻¹ peak drops to zero and the 1679 cm⁻¹ peak increases slightly, suggesting that the 1650 cm⁻¹ peak is due to higher order aggregate such as chain-like structures or dimers, and that the 1679 cm⁻¹ peak is from the C=O stretch of the monomers.

In DMSO-d₆, a carbonyl peak was observed at 1668 cm⁻¹, which coincides with the frequency of the free C=O in chain-like structure seen in CD₂Cl₂-d₂ and CCl₄. However, a second C=O peak that would correspond to the hydrogen-bonded C=O in the chain was not detected. This observation suggests the 1668 cm⁻¹ peak is due to monomer whose C=O interacts with the solvent and therefore redshifts its frequency from 1679 cm⁻¹. However, temperature or concentration dependence is warranted to confirm this assignment.

The 1550 cm⁻¹ to 1600 cm⁻¹ spectral region exhibits complicated peak patterns. Due to the lack of cross-peaks to the C=O stretch, these peaks are assigned to the lactim

![Figure 5.35: Temperature-dependent FTIR spectra of 6-chloro-2-pyridone (5mg/ml) in (a) CCl₄ and (b) MeCN-d₃.](image-url)
tautomer. The lactim monomer in water only has one peak in this spectral region (perhaps another peak which overlaps with the lactam ring modes, see discussion Section 5.2), but up to four peaks associated with the lactim tautomer are observed in spectra taken in organic solvents. The 2D IR spectra show checkboard-like peak pattern between the diagonal and cross-peaks, indicating that these peaks originate from the same species. Since the formation of lactim dimer (Figure 5.24, cyclic dimer or chain-like) can give rise to many low frequency modes such as out-of-plane C-H bend, and N···O-H bend...etc, it is reasonable to suspect that extra peaks show up due to the Fermi resonance of those low frequency modes.

The temperature-dependent FTIR (Figure 5.35) and 2D IR (Figure 5.36) in this region show small drop in signal (in particular at ~ 1600 cm\(^{-1}\) and 1575 cm\(^{-1}\)) upon heating. More interestingly, the thermal difference spectrum in MeCN-d\(_3\) reveals a gain feature at 1590 cm\(^{-1}\). A plausible explanation for the temperature-dependent spectra is that upon heating, the lactim dimer melts into monomers, and the 1590 cm\(^{-1}\) could be assigned to the lactim monomer (this frequency also coincides with the value observed in D\(_2\)O). Nevertheless, due to significant spectral overlap and small degree of dimer melting, it is difficult to extract more information from this data set.

Finally, we note that the assignments made here (summarized in Table 5.5) are based on the limited available experimental results and therefore should be regarded as speculative. The preliminary data are presented to serve as a starting point for future studies. Further experiments including temperature and concentration dependence are required to confirm these assignments. It would also be interesting to investigate the origin of the Fermi resonances that give rise to the complicated peak pattern seen in the lactim dimer spectra.

![Figure 5.36: 2D IR spectra of 6-chloro-2-pyridone in MeCN-d\(_3\) at 10\(^\circ\)C and 40\(^\circ\)C, and the thermal difference spectrum.](image-url)
Table 5.5. Experimentally observed peaks for 6-chloro-2-pyridone in various solvents. The assignments were made based solely on the experimental spectra, without DFT calculations.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>MeCN-\text{d}_3</th>
<th>DMSO-\text{d}_6</th>
<th>\text{CD}_2\text{Cl}_2-\text{d}_2</th>
<th>\text{CCl}_4</th>
<th>THF</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1679</td>
<td></td>
<td></td>
<td></td>
<td>1681</td>
<td></td>
<td>Lactam monomer C=O</td>
</tr>
<tr>
<td>1650</td>
<td>1668$^a$</td>
<td></td>
<td>1669</td>
<td></td>
<td></td>
<td>Lactam chain free C=O</td>
</tr>
<tr>
<td>1605</td>
<td></td>
<td></td>
<td>1649</td>
<td>1651</td>
<td>1607</td>
<td>Lactam dimer C=O (HB)</td>
</tr>
<tr>
<td>1603</td>
<td>1602</td>
<td></td>
<td>1602</td>
<td>1603</td>
<td>1604</td>
<td>Lactam dimer</td>
</tr>
<tr>
<td>1590</td>
<td></td>
<td></td>
<td>1600$^a$</td>
<td>1592</td>
<td>1595</td>
<td>Lactam monomer ring mode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lactam dimer ring mode</td>
</tr>
<tr>
<td>1574</td>
<td></td>
<td></td>
<td>1578</td>
<td>1575</td>
<td>1574</td>
<td>Lactam dimer</td>
</tr>
<tr>
<td>1561</td>
<td>1564</td>
<td></td>
<td>1560</td>
<td>1561</td>
<td>1561</td>
<td>Lactam dimer</td>
</tr>
<tr>
<td>1547</td>
<td></td>
<td>1546</td>
<td>1543</td>
<td>1539</td>
<td>1549</td>
<td>Lactam dimer</td>
</tr>
</tbody>
</table>

$^a$ More descriptive assignment could not be made due to the lack of temperature-or concentration dependent data.

5.5 Acknowledgments

I thank Vipender Singh, Deyu Li and Professor John Essigmann for starting the collaboration on studying tautomerism which led to multiple exciting projects presented here and in the following chapters. The NMR experiments of oxythiamine were performed by Deyu Li and Katherine Silvestre. The binding isotope effects experiments of oxythiamine were carried out by Vipender Singh and Koyel Mitra. I also thank Mike Reppert for his stimulating discussions, Carlos Baiz for his regular assistance in the lab, and Luigi De Marco for his careful reading of the pyridone dimer part of this chapter.
5.6 References


Appendix 5.A: Determination of equilibrium constant for 6-chloro-2-pyridone

For the tautomeric equilibrium of 6-chloro-2-pyridone:

\[ \text{la} \xrightarrow{} \text{lb} \]  
(5.A.1)

Where 1a and 1b represent the lactim and lactam tautomers, respectively. The equilibrium constant as a function of temperature therefore is written as the ratio of the populations of two tautomers:

\[ K_{eq}(T) = \frac{P_{lb}(T)}{P_{la}(T)} \]  
(5.A.2)

We express the tautomer populations using the Boltzmann factors and the relative Gibbs free energy \( \Delta G \):

\[ P_{la}(T) = \frac{1}{1 + \exp\left(-\Delta G / kT\right)} \]  
(5.A.3)

\[ P_{lb}(T) = \frac{\exp\left(-\Delta G / kT\right)}{1 + \exp\left(-\Delta G / kT\right)} \]

From the FTIR spectrum of 6-chloro-2-pyridone, we identify the two separate peaks for the lactim and lactam tautomers at 1591 cm\(^{-1}\) and 1637 cm\(^{-1}\), respectively. We can express the integrated intensities of these two peaks using Beer’s law:

\[ \overline{A}_{la}(T) = A_{la}(T) / bC = \varepsilon_{la} P_{la}(T) \]  
(5.A.4)

\[ \overline{A}_{lb}(T) = A_{lb}(T) / bC = \varepsilon_{lb} P_{lb}(T) \]

Where \( b \) is the path-length, and \( C \) is the total concentration of the sample which include both tautomers. If we take the ratio of the two integrated intensities:

\[ \frac{\overline{A}_{lb}(T)}{\overline{A}_{la}(T)} = \frac{P_{lb}(T)}{P_{la}(T)} \times \frac{\varepsilon_{lb}}{\varepsilon_{la}} = K_{eq}(T) \frac{\varepsilon_{lb}}{\varepsilon_{la}} = e^{-\Delta G / kT} \times \frac{\varepsilon_{lb}}{\varepsilon_{la}} \]  
(5.A.5)

\[ \Rightarrow \ln \left[ \frac{\overline{A}_{lb}(T)}{\overline{A}_{la}(T)} \right] = \frac{-\Delta H}{kT} + \frac{\Delta S}{k} + \ln \left( \frac{\varepsilon_{lb}}{\varepsilon_{la}} \right) \]  
(5.A.6)

167
Here we have made the assumptions that the changes in enthalpy and entropy do not depend on temperature. This should be valid given that the temperature range in the experiment is small (from 10°C to 90°C). Fitting this curve as a function of $1/T$ results in:

$$\frac{-\Delta H}{k} = 1665 \pm 33$$
$$\frac{\Delta S}{k} + \ln \left( \frac{\varepsilon_{lb}}{\varepsilon_{ua}} \right) = -4.678 \pm 0.103$$

(5.A.7)

Once we have obtained $\Delta H$, we can then fit the integrated intensities to get $\varepsilon_1$, $\varepsilon_2$, and $\Delta S$:

$$\bar{A}_{1a}(T) = \frac{1}{1 + e^{-\Delta H/kT} e^{-\Delta S/k} \varepsilon_{1a}}$$
$$\bar{A}_{1b}(T) = \frac{e^{-\Delta H/kT} e^{-\Delta S/k} \varepsilon_{1b}}{1 + e^{-\Delta H/kT} e^{-\Delta S/k} \varepsilon_{1b}}$$

(5.A.8)

The results are listed and shown in Figure 5.A.1:

$$\varepsilon_{1a} = 0.437 \pm 0.016$$
$$\varepsilon_{1b} = 0.563 \pm 0.010$$
$$\frac{\Delta S}{k} = -4.931 \pm 0.074$$

(5.A.9)

Figure 5.A.1: Analysis of the temperature-dependent FTIR spectra of 6-chloro-2-pyridone. (left) Fit to Equation 5.A.6. (right) Fit to Equation 5.A.8.
CHAPTER 6
Direct Observation of Ground State Lactam–Lactim Tautomerization Using Temperature-Jump Transient 2D IR Spectroscopy

The work presented in this chapter has been published in the following paper:


6.1 Introduction

Tautomerism of aromatic heterocycles has been extensively studied due to its importance in biochemical processes such as enzyme catalysis, ligand binding, and spontaneous mutagenesis. In Chapter 5, we have shown that the unique cross-peak patterns in 2D IR spectra can be used to separate different tautomers under equilibrium. However, to gain further insight into the mechanism of tautomerism, it is also required to be able to monitor the tautomer exchange in real time. To this end, we demonstrate the capability of 2D IR spectroscopy coupled with a nanosecond temperature-jump (T-jump) laser to reveal the non-equilibrium lactam–lactim tautomerization kinetics of pyridone derivatives.

Time-resolved studies of tautomerism have a long and storied history. In the 1960’s, Manfred Eigen pioneered the T-jump relaxation method using rapid electrical heating in electrolyte solutions to trigger fast chemical reactions, such as the recombination of the enolate ion of barbituric acid with H⁺ to form either a keto or enol tautomer. Others have since used a similar approach, using a T-jump monitored by UV absorption, to study tautomerism of
small molecules on the microsecond time scale.\textsuperscript{5,7} These groundbreaking experiments provided the early insights into proton transfer and tautomerization kinetics. However, they still suffer from certain limitations including the difficulty of interpreting variations in electronic spectra, the inclusion of pH indicators, and questions about reactive species that may be created by the capacitive discharge. The more recent studies with time-resolution to capture proton transfer dynamics have used ultrafast photo-excitation to induce tautomerization on the excited electronic state.\textsuperscript{8-13} This has proven a fruitful test bed for dynamics, but experiments in room temperature aqueous solution would more accurately report on physiological tautomerism.

Two-dimensional IR spectra, analogous to 2D NMR, exhibit cross-peak patterns between vibrational modes that provide distinct signatures of molecular structures, although with sub-picosecond time resolution.\textsuperscript{14} Previously, we have shown that intramolecular couplings between the vibrational modes of aromatic heterocycles result in pronounced 2D IR cross-peaks,\textsuperscript{15} and therefore the lactam and lactim tautomers of 2-pyridone derivatives can be differentiated based on the distinct cross-peak patterns from individual species.\textsuperscript{16} Our interest is to make use of 2D IR to explore the time scale and mechanism for tautomer exchange. Chemical exchange processes at equilibrium have been characterized through waiting-time 2D IR experiments where the delay between the excitation and detection dimensions is varied.\textsuperscript{17} Nevertheless, the time window for these experiments is limited to the vibrational lifetime of the systems, typically a few picoseconds. To circumvent this limitation, one can perform transient 2D IR (t-2D IR) experiments that follow re-equilibration after an abrupt perturbation.\textsuperscript{18-23} Our strategy to monitor ground state reactivity is to perform t-2D IR triggered by a nanosecond T-jump and observe the tautomeric interconversion in real time during the re-equilibration. In the following, we demonstrate the nanosecond lactam–lactim tautomerization kinetics of pyridone derivatives in D\textsubscript{2}O, and find that tautomerization of pyridone proceed through a concerted intramolecular proton transfer facilitated by bridging water molecules. This approach opens up an opportunity to study tautomerization under physiological conditions, and provides a platform for experimental validation of electronic structure models that offer an atomistic picture of the barrier-crossing processes.

6.2 Equilibrium measurements

Because of its simple structure and similarity to DNA bases, 2-pyridone has drawn considerable interest and is regarded as the prototype molecule to study lactam–lactim tautomerization. It is known that in aqueous solution, 2-pyridone exists primarily in the lactam
form, and the addition of chlorine at the 6-position significantly shifts the tautomeric equilibrium towards the lactim form.\textsuperscript{7} The structure, FTIR and 2D IR spectrum of 6-chloro-2-pyridone at 40°C in D\textsubscript{2}O is shown in Figure 6.1. In an absorptive 2D IR spectrum, peaks appear as a doublet with a positive peak (red) originating from the ground to first excited vibrational state (0–1) transition, and a negative peak (blue) due to the 1–2 transition. The diagonal features, labeled A\textsubscript{1}–A\textsubscript{4}, can be mapped to the corresponding peaks in the linear absorption spectrum, whereas the cross-peaks correlate the vibrational excitation ($\omega_{ij}$) and detection ($\omega_{3}$) frequencies, which allow the separation of different chemical species in a mixture, even with significant spectral congestion.

Peak A\textsubscript{1} is the carbonyl stretch of the lactam form (\textit{1a}), and peaks A\textsubscript{3} and A\textsubscript{4} are ring vibrations of the same species, indicated by their pronounced cross-peaks to A\textsubscript{1}.\textsuperscript{16} On the other hand, peak A\textsubscript{2}, which does not have cross-peaks with A\textsubscript{1}, A\textsubscript{3}, and A\textsubscript{4}, arises from the ring vibration of the lactim tautomer (\textit{1b}), coupled to D\textsubscript{2}O bending motions. These peak assignments are further validated by comparing to the 2D IR spectrum of N-methyl-2-pyridone (2, Figure 6.1b), whose labile proton has been replaced by a methyl group to block tautomerization. The 2D IR spectrum of 2 closely resembles the spectral features of \textit{1a} (indicated by the dash purple lines), except for the extra ring mode and the absolute peak frequencies. Previous 2D IR studies of the lactam–lactim tautomerization of 1 were used to determine the equilibrium constant $K_{eq} = [\textit{1b}] / [\textit{1a}] = 0.53 \pm 0.02$ at 25°C and found that this value increases with temperature.\textsuperscript{16}

![Figure 6.1: Chemical structures (a,b), equilibrium FTIR (c,d), and 2D IR spectra (e,f) with ZZZZ polarization for 6-chloro-2-pyridone at 40°C (left), and N-methyl-2-pyridone at 25°C (right). The dash grids show the coupled vibrations from the lactam tautomers.](image-url)
Figure 6.2: Transient difference 2D IR spectra at T-jump delays of 10 ns (a), 40 ns (b), and 100 ns (c) for I following a T-jump from 40 to 50°C. These spectra are plotted between ±6% difference signal relative to the equilibrium spectrum. (d) Normalized second component spectrum calculated from the SVD analysis of the temperature-dependent 2D IR spectra from 10°C to 80°C. The frequency axes are plotted differently from Figure 6.1 to focus on the growth of the lactim peak.

6.3 Transient 2D IR spectroscopy

We performed T-jump infrared experiments using the methods described in Refs. [31-33] to probe time-dependent changes to the tautomer populations. We recorded the difference signal relative to a reference spectrum, $\Delta S(\tau, T) = S(\tau, T) - S(T_i)$, where $\tau$ is the time delay between the T-jump and 2D IR pulses and $S(T_i)$ is the equilibrium spectrum measured prior to the T-jump. Tautomeric interconversion was monitored with transient 2D IR (t-2D IR) and transient heterodyne-detected dispersed vibrational echo (t-HDVE). Transient HDVE were performed with fixing the excitation at 0 fs and the waiting time at 150 fs. The complex value t-HDVE signal was measured by Fourier-transform spectral interferometry, and its information content is equivalent to the projection of the t-2D IR spectrum onto the detection frequency axis. While t-2D IR spectra provide the most information on the system, the measurement requires roughly 100 times longer sampling time than t-HDVE. Therefore, we assigned the transient spectral features using full t-2D IR spectra at selected time delays, and obtained kinetics from the more finely-sampled one-dimensional t-HDVE spectra.

Transient 2D IR and the corresponding t-HDVE spectra of I in phosphate buffered D$_2$O at pD 7.8 starting from an initial temperature $T_i = 40°C$ are displayed in Figure 6.2 and Figure 6.4b, respectively. We use different color maps for the equilibrium and transient spectra in order to emphasize that t-2D IR spectra are time-dependent difference spectra with respect to the equilibrium spectra at $T_i$. A loss of population appears in t-2D IR difference spectra as
a negative (blue) peak on the top with a positive (red) peak on the bottom. At 10 ns following the T-jump, a bleach of the C=O stretch of 1a at 1634 cm⁻¹ is observed (box A1 in Figure 6.2a). Moreover, a gain feature (red above and blue below the diagonal) is identified for peak A2, which is a ring vibrational mode of 1b. These spectral changes indicate the growth of lactim tautomer 1b at the expense of the lactam tautomer 1a. Here a 10° temperature change corresponds to an equilibrium population drop of 1a from 59% to 55%. Both the A1 loss and the A2 gain increase with T-jump delay, as illustrated in Figure 6.4a by the integrated peak intensities obtained from the t-2D IR absolute value spectra (Figure 6.3). We fit both time traces to mono-exponential decays with relaxation time constants of 56 ns and 60 ns for 1a and 1b, respectively. The non-stretched exponential behavior and the common time constant provide experimental evidence for two-state tautomerization kinetics, and a concerted proton transfer mechanism without a long-lived intermediate state.

Figure 6.3: Equilibrium and transient 2D IR absolute value spectra for 6-chloro-2-pyridone. The t-2D IR spectra were collected following a T-jump from 40°C to 50°C, at selected T-jump delays 10, 40, and 100 ns. The magenta and green boxes indicate the spectral regions where the peak intensities for the lactam and lactim tautomers were integrated to obtain the kinetics of the tautomeric interconversion, shown in Figure 6.4a and on the right. The error bars (ranging from 2 to 20%) were obtained from the standard deviation of 4 t-2D IR measurements.

The loss of the lactam C=O stretch is also observed in the t-HDVE spectra as a negative difference signal marked by the purple line in Figure 6.4b. A single exponential decay with a time constant of 56±9 ns was measured for this T-jump response at T = 40°C → 50°C (Figure 6.4c), which is in excellent agreement with the kinetics obtained from t-2D IR. Considering that t-HDVE can be collected much faster than t-2D IR to allow more accurate kinetic measurements, in the following discussion, we used the A1 response from t-HDVE to represent the kinetics of both the loss of 1a and gain of 1b. We note however, due to significant spectral congestion in this frequency region, the gain of the lactim tautomer cannot be readily identified in t-HDVE. In contrast to the clear gain feature at ~ 1600 cm⁻¹ observed in the t-
2D IR spectra, t-HDVE exhibits only a small positive peak on top of the broad loss features from A1 and A3. The ability to resolve both tautomeric species in t-2D IR spectra demonstrates the increased spectral information content available in 2D IR spectra compared to the equivalent one-dimensional representations.

To verify that the observed response corresponds to tautomerization, and rule out any T-jump artifacts such as cavitation or thermally induced change in dielectric constants, we carried out a negative control experiment on 2, whose N1 tautomerization site has been blocked. The t-HDVE time trace of the C=O stretch of 2 is plotted as orange crosses in Figure 6.4a (spectra shown in Figure 6.5). Only a T-jump pulse-width limited baseline shift that results from increased solvent transmission was observed, confirming that the ns exponential relaxation is due to the lactam–lactim tautomerization.

---

Figure 6.4: (a) Integrated peak intensities of the t-2D IR absolute value spectra for 1 at T = 40°C. The integration regions for lactim and lactam are shown by the green and magenta boxes in Figure 6.3, at ~1600 cm⁻¹ and ~1640 cm⁻¹, respectively. The t-2D IR amplitudes have been scaled to be better compared with the t-HDVE traces (the error bars are smaller than the marker size, see Figure 6.3). The solid lines are the single exponential fits. The orange crosses display the t-HDVE time trace for 2 at ω₀ = 1649 cm⁻¹ at T = 25°C. (b) Transient difference spectra derived from the absolute value of the complex value t-HDVE for 1 from 40 to 50°C at selected T-jump delays. (c) Representative t-HDVE time traces at selected T, for the C=O stretch of 1 at ω₀ = 1634 cm⁻¹. (d) The Arrhenius plot.
Figure 6.5: (Left) t-HDVE spectra of N-methyl-2-pyridone at 25°C at selected T-jump delays. (Right) The time dependence of the t-HDVE spectra at 1568 cm\(^{-1}\) and 1649 cm\(^{-1}\) shows a T-jump pulse-width-limited increase in solvent transmission, indicating that spectral changes are all complete within \(\ll 1\) ns.

In addition to reporting on the tautomeric interconversion, t-2D IR provides spectral signatures of the solvent environment around the carbonyl group. In the 10 ns spectrum, positive features (violet arrows in Figure 6.2a) appear on the blue side of the A1 and A3 modes—a sign of a blue-shift of the peak, which suggests weakening or loss of hydrogen-bonds (HBs) to \(1a\).\(^{16}\) The more complicated line-shapes of the A1/A3 cross-peaks (green arrows) are interpreted as a convolution between the blue-shifts and the population decrease of \(1a\) (see Figure 6.6 for more details). Cervetto and coworkers\(^{25}\) have also discussed different t-2D IR spectral features in details. At longer T-jump delays, most of the spectral changes are solely due to the tautomeric interconversion, indicating that the solvent HB equilibration is finished by \(\sim 10\) ns. Overall, these changes in the t-2D IR spectra due to both the tautomerization and solvent fluctuation are encapsulated in the second spectral component from the singular value decomposition (SVD) analysis of the temperature-dependent equilibrium 2D IR spectra from 10°C to 80°C (Figure 6.4d).

### 6.3.1 Temperature dependence

To determine the thermodynamic properties of tautomerization, we monitored the time evolution of A1 from the t-HDVE spectra at different initial temperatures (Figure 6.4c). With increasing temperature, the relaxation rate speeds up and the amplitude of the difference signal increases. First-order kinetics for the two-state exchange leads to an observed relaxation rate of \(k_{obs} = 1/\tau_{obs} = k_f + k_r\), where \(\tau_{obs}\) is the observed relaxation time constant and \(k_f\) and \(k_r\) are the forward and reverse rate constants. Using the equilibrium constant
$K_{eq} = [b]/[a] = k_f/k_r$ \cite{16}, $k_f$ and $k_r$ can be determined at each temperature. The temperature dependence of the forward and reverse rates (Figure 6.4d) was used to extract the activation barrier $E_a$ and pre-exponential factors $A$ for these reactions (Table 6.1). We obtained an activation barrier of $7.4 \pm 0.5$ kcal/mol ($\sim 12 k_BT$) for the lactam to lactim tautomerization. This can be compared to the calculated results for the parent molecule 2-pyridone, using DFT and MP2 electronic-structure methods by Sonnenberg and coworkers.\cite{26} The energy barrier was found to be $\sim 40$ kcal/mol in isolation, and further reduced to $\sim 15$ kcal/mol with the addition of one explicit water molecule to mediate the proton transfer. The microsolvation configuration of pyridone used in their calculation is similar to the structure (Figure 6.7) that gives reasonable agreement between the experimental and calculated IR absorption spectra.\cite{16}

It is not surprising that the calculated $E_a$ for 2-pyridone is higher than the experimental value for 6-chloro-2-pyridone since the chlorine substitution significantly lowers the relative energy of the lactim tautomer. Additionally, it has not been established that a specific computational method can accurately predict the energetics of a tautomeric system. Most importantly, the significant reduction in the activation energy by hydrating water molecules argues strongly for a solvent mediated tautomerization mechanism in our experiments.

![Figure 6.6: Simulated difference spectrum to demonstrate the origins of some spectral features observed in the t-2D IR spectra shown in Figure 6.2. We only included peaks A1 and A3, and blue-shifted both A1 (1634 $\rightarrow$ 1637 cm$^{-1}$) and A3 (1550 $\rightarrow$ 1558 cm$^{-1}$) to represent the high temperature spectrum. The two cross-peaks marked by green arrows respond differently to blue-shifts: the uphill ($\omega_1 < \omega_3$) and downhill ($\omega_1 > \omega_3$) cross-peaks blue-shift along the $\omega_1$ and the $\omega_3$ axes, respectively. We note that the line-shapes are not fully reproduced due to incorrect shifts, and inhomogeneous to homogeneous transition. Fitting to the full nonlinear response function including line-shape parameters will be required to account for the discrepancies.](image-url)
Table 6.1. Thermodynamic and kinetic fit parameters obtained for 6-chloro-2-pyridone and 2-chloro-4-pyridone at temperatures of 25°C and 30°C, respectively.

<table>
<thead>
<tr>
<th>Tautomer</th>
<th>k, $10^6$ s$^{-1}$</th>
<th>$A_s$, s$^{-1}$</th>
<th>$E_a$, kcal/mol</th>
<th>$\Delta H^\ddagger$, kcal/mol</th>
<th>$\Delta S^\ddagger$, cal/mol·K</th>
<th>$\Delta H$, kcal/mol</th>
<th>$\Delta S$, cal/mol·K</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-chloro-2-pyridone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lactam→Lactim</td>
<td>3.1 ± 0.1</td>
<td>8.2 ± 0.3 × 10$^{11}$</td>
<td>7.4 ± 0.5</td>
<td>6.7 ± 0.5</td>
<td>-6.4 ± 0.2</td>
<td>3.3 ± 0.1</td>
<td>9.8 ± 0.2</td>
</tr>
<tr>
<td>Lactim</td>
<td>5.9 ± 0.2</td>
<td>6.0 ± 0.2 × 10$^3$</td>
<td>4.1 ± 0.5</td>
<td>3.4 ± 0.5</td>
<td>-16.2 ± 0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-chloro-4-pyridone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lactam→Lactim</td>
<td>10.2 ± 2.5</td>
<td>1.8 ± 0.5 × 10$^3$</td>
<td>8.7 ± 0.6</td>
<td>8.1 ± 0.6</td>
<td>0.1 ± 0.0</td>
<td>3.9 ± 0.1</td>
<td>15.3 ± 0.4</td>
</tr>
<tr>
<td>Lactim</td>
<td>3.2 ± 0.8</td>
<td>8.6 ± 2.2 × 10$^3$</td>
<td>4.8 ± 0.6</td>
<td>4.1 ± 0.6</td>
<td>-15.2 ± 0.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.7: Chemical structures of the lactim and lactam tautomers of 6-chloro-2-pyridone hydrogen-bonded to two water molecules. QChem was used to perform geometry optimization and frequency calculations with the B3LYP hybrid functional and 6-31G (d,p) basis set. The labile protons have been replaced with deuterium atoms.

We stress again that our experiments probe the rare tautomerization events on the ground electronic state, in contrast to excited-state proton transfer (ESPT). Electronic excitation reduces pKa, reshapes the energy surface, and lowers the activation barrier for tautomerization, thus leading to much faster proton transfer rates than in the ground state. For example, photo-initiated ESPT in 7-azaindole/alcohol complexes occurs within ~ 100 ps with negligible barrier (< 1 kcal/mol), and the double proton transfer in 7-azaindole dimer proceeds with a 1.1 ps time constant.

We also calculated the activation enthalpy and entropy (Table 6.1) using linear fits to the Eyring-Polanyi equation $\ln (k / T) = -\Delta H^\ddagger / RT + \Delta S^\ddagger / R + \ln (k_B / h)$. Since the internal entropies of the two tautomeric forms are not expected to differ significantly, the negative $\Delta S^\ddagger$ observed suggests that the transition state involves a well-structured solvation complex, for instance a water wire or ring structure between the solute and the solvating water molecules. Both the small activation barrier and negative activation entropy suggest the presence of bridging water molecules during tautomerization.
6.3.2 Solvent dependence

In order to explore the involvement of the solvent, we measured the relaxation rates in different EtOD/D$_2$O solvent mixtures at $T_i = 25^\circ$C (Figure 6.8). We found that the rate dropped by a factor of 1.8 when the D$_2$O volume percentage decreased from 100% to 40%, even though the viscosity of ethanol/water mixtures exhibits a maximum at 0.25 mole fraction of ethanol.\textsuperscript{28} This observation suggests that the dense hydrogen-bonding network of water may play an important role in facilitating tautomerization. Even though the exact nature of the solvent effect on tautomerization requires further investigation, our present study is not consistent with tautomerism via intramolecular proton transfer without solvent mediation, which has been observed for protonated uracil in the gas phase using IR photo-dissociation spectroscopy\textsuperscript{29} and suggested for glycine in water using NMR relaxation methods.\textsuperscript{30}

6.3.3 2-Chloro-4-pyridone

Several mechanisms have been proposed for proton transfers in tautomerization. For example, intermolecular double proton tunneling has been suggested for the spontaneous formation of enol tautomers in Watson-Crick base pairs,\textsuperscript{31} and has been observed for imidazole and pyrazole in ether/THF or ether/acetone mixed solvents at -100$^\circ$C.\textsuperscript{32} We precluded the possibility of dimer formation in water because both the FTIR spectra\textsuperscript{10} and T-jump relaxation rates (Figure 6.9) were found to be independent of concentration. Alternatively, intramolecular or bifunctional proton transfer with water participation can proceed via a concerted (Figure 6.10a) or a stepwise (Figure 6.10b) mechanism. The stepwise

![Figure 6.8](image)

Figure 6.8: Lactam-lactim tautomerization rate as a function of the percentage of D$_2$O in a D$_2$O/EtOD mixture at $T_i = 34^\circ$C. An exponential curve $8.8 \times 10^{-6} \exp(0.0656 \times \text{pct}) + 7.34 \times 10^{-3}$ is plotted to show the trend of the solvent effect.
mechanism, also referred to as "dissociative" mechanism, involves an ionic intermediate in which the nitrogen proton becomes solvated. The excess proton can then diffuse or undergo an undirected Grothuss transport to the proton acceptor site. On the other hand, proton transfer by concerted tautomerization mostly likely proceeds as a Grothuss transfer with the assistance of a water bridge.

Figure 6.9: Concentration dependence of the relaxation rate constant for 6-chloro-2-pyridone at $T_f = 34^\circ C$ and pH = 7.8. The concentration independence rules out the tautomerization mechanism via dimerization (right). The error bars are reported as the standard deviation from measurements of more than five times. The error bars are larger for lower concentration measurements due to the lower signal-to-noise level.

Diffusive proton transfer or direct intramolecular tunneling should be strongly influenced by changes in proximity between proton acceptor and donor sites. We tested this by measuring the T-jump relaxation rates for 2-chloro-4-pyridone (3a, structure shown in Figure 6.11b), in which the carbonyl and N-D groups are in the para position—as opposed to ortho position in 1a. Temperature-dependent FTIR of 3 at pD = 7.8 (Figure 6.11a) shows that the C=O intensity at 1626 cm$^{-1}$ decreases while the 1581 cm$^{-1}$ peak increases with temperature. The equilibrium 2D IR spectrum of 3 (Figure 6.11c) reveals that these two peaks are from two separate species, which we attribute to the lactam and lactim tautomers based on DFT calculations and the similarities to the spectra of 1. We obtained the T-jump relaxation rates through the loss of the lactam C=O peak measured with the t-HDVE spectra (Figure 6.12). The relaxation time constant measured under the same conditions (pD = 7.8, $T_i = 40^\circ C$) for 3 was 49 ± 13 ns, which is slightly faster than that of 1. We calculated the activation thermodynamic parameters using the temperature dependence of the relaxation rates (Figure 6.11d and Table 6.1), and found that they are within 2 kcal/mol to those of 1. The similar kinetics of 1 and 3 exclude the possibility of intramolecular tunneling since the distance between proton donor and acceptors sites of 3 is too large for efficient tunneling.
Figure 6.10: Possible mechanisms for the tautomerization of 1. (a) Concerted mechanism with water-bridge facilitating the proton transfer. Longer bridging networks are possible. (b) Stepwise tautomerization through ionic intermediates, catalyzed by water molecules.

Figure 6.11: Experimental data of 2-chloro-4-pyridone, whose tautomeric structures are shown in (b). (a) Temperature-dependent FTIR. (c) Equilibrium 2D IR spectrum at 25°C. (d) Arrhenius plot for the T-jump relaxation rate.
The similarity of proton transfer kinetics between 1 and 3 indicates that aqueous proton migration does not limit the efficiency of the tautomer exchange. We note that using the proton diffusion coefficient in water of $8 \times 10^{-9}$ m$^2$ s$^{-1}$, the time for a proton to sample an rms distance equal to the proton donor-acceptor site distance for 1 and 3 is 1.2 ps and 5.4 ps, respectively. This indicates that a released proton has sufficient time to diffuse to the proton acceptor site in either case, and that the ion pair is a short-lived metastable intermediate. The degree to which diffusion and proton delocalization along a water wire are involved remains to be determined, but our results show that the rate limiting step is the release of the proton from the substrate to water. The slightly faster tautomerization rate for 3 is the result of its smaller pKa = 7.26 compared to 7.91 for 1a (Figure 6.14). Dissociation of covalently bound protons into a fluctuating and strongly hydrogen bonded aqueous environment can be rationalized in terms of pre-organization of the hydrogen-bonded water to accept the proton. But in physical terms, transfer to solvent likely involves rare polarization fluctuations that align the electric fields of the surrounding water so that it distorts the proton potential into a dissociative state, suggesting that the entropic contribution from the solvent may significantly modulate the reaction barrier.
6.3.4 pD dependence

In order to gain insights into the acid- and base-catalyzed tautomerization mechanism, we examined the pD dependence of the T-jump relaxation rate of 1 at \( T_i = 25^\circ C \), illustrated in Figure 6.14. We describe the variation of T-jump relaxation rate using the following expression:

\[
\tau_{\text{obs}}^{-1} = k_0 + k_D[D^+] \tag{6.1}
\]

where \([D^+]\) is the concentration of deuterium ion and \(k\)'s are the rate constants. The fit shown in the dash curve results in \( k_0 = 1.23 \times 10^7 \text{ s}^{-1} \), and \( k_D = 5.03 \times 10^{10} \text{ M}^{-1}\text{s}^{-1} \). Clearly, near neutral pD, \( k_0 \) dominates over \( k_D[D^+] \) by more than three orders of magnitude. In general, tautomerization via an "intramolecular dissociative" mechanism (Figure 6.10b) near neutral pH is catalyzed by water molecules to form cationic or anionic intermediates. In such cases, the pH-independent rate constant \( k_0 \) has been estimated to be \( \sim 10^2-10^3 \text{ s}^{-1} \).33 However, \( k_0 \) determined here greatly exceeds such a typical value and therefore further supports a concerted or "non-dissociative" mechanism shown in Figure 6.10a. The large value is not entirely surprising as Chang and Grunwald have reported a value of \( k_0 = 1.1 \times 10^8 \text{ s}^{-1} \) for the uracil N1H and N3H proton transfer using dynamic NMR,34 and concluded that the mechanism must be non-dissociative. The concerted mechanism supported by the weak pD-dependence is consistent with the observation that the lactam and lactim peaks in the t-2D IR spectra follow the same single exponential kinetics (Figure 6.4a). We should note that Figure

![Figure 6.13: Intensity of the C=O stretch (\( \omega = 1637 \text{ cm}^{-1} \)) of 10mg/ml 6-chloro-2-pyridone as a function of pH* (pH reading in D_2O) at 25°C. The data was fitted to the Henderson-Hasselbalch equation: \( A = a[1+10^{\text{pKa-pH*}}]^{-1} + b \), where a, b, and pKa are the fitting parameters. The pKa obtained was 7.91.](image-url)
6.10a represents the simplistic picture of the concerted mechanism with one water molecule. A short-lived transition state with a negative charge delocalized across the N-C-O bonds, although possible, cannot be detected with our experiments. Moreover, weak dependence on proton concentration also argues against physical diffusion of the proton from donor to acceptor site, and in favor of a Grothuss proton hopping mechanism facilitated by water wires, i.e. some degree of hydrogen bonding connectivity between sites.

Under very acidic conditions, the reaction rate is governed by the \( k_{D^+} [D^+] \) term, and the increased rate is explained by the widely studied acid-catalyzed mechanism by \( D^+ \), which involves the protonation of the substrate. Although few data points were measured beyond neutrality to ensure the fitting results, the obtained recombination rate constant with \( D^+ \) is consistent with previously reported values \( k_{D^+} \sim 10^{10} - 10^{11} \text{ M}^{-1} \text{s}^{-1} \).^33,35,36

![Figure 6.14: pD dependence of the T-jump relaxation rate of 1 at T = 25°C. The dash curve shows the fit to Equation (6.1). In general, the relaxation rate described by Equation (6.1) can also depend on the concentrations of deuterium hydroxide, cationic and anionic substrate. Since the pKa's for proton gain and loss are -0.18 and 7.91 respectively, the cationic and anionic substrates are negligible within the pD range under study and hence excluded from the analysis. At pD \( \gg 7.91 \), 1 is deprotonated to form enolate ion, and the resulting loss of the C=O stretch intensity prevents measurements of the T-jump relaxation rate.](image)
6.4 Comparison with prior results

Finally, we wish to address the pioneering work by Dubois and coworkers, who used a Joule-heating T-jump apparatus with a UV spectrometer to study tautomerization kinetics and mechanism over 35 years ago.\textsuperscript{5,7,37-40} Our experiments showed tautomerization kinetics that are three orders of magnitude faster than those reported by the Dubois group. They reported that the relaxation time for 1 is about 10-20 µs at T\textsubscript{f} = 10°C, and that the tautomerization of 3 is six times slower.\textsuperscript{7,37} Nonetheless, they also observed a large \( k_0 \) contribution near neutral pH, and concluded that tautomerization occurs via a non-dissociative mechanism. Given the striking differences between our experimental observations, we performed the control experiments to validate our conclusions.

The Dubois group studied tautomerism using UV absorption spectroscopy coupled to a T-jump induced by capacitor discharge.\textsuperscript{5,7,37-40} In most of their studies, they observed a large amplitude, < 5 µs response, followed by a slower response on the time scale of tens to hundreds of µs. They assigned the fast response to temperature-dependent variation in solvation environment and the slower process to tautomerism. They argued that since the sub-µs (faster than their time resolution) response was observed for both N-methyl-2-pyridone and 6-chloro-2-pyridone, this response cannot originate from tautomerism. In order to further understand their results and compare to our infrared experiments, we measured the UV absorption spectra of 6-chloro-2-pyridone and N-methyl-2-pyridone in ethanol and H\textsubscript{2}O, and at different temperatures shown in Figure 6.15.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
 & EtOH – H\textsubscript{2}O & & High T – Low T & \\
\hline
Max (nm) & 6-chloro-2-pyridone & N-methyl-2-pyridone & 6-chloro-2-pyridone & N-methyl-2-pyridone \\
\hline
 274 & 322 & 274/322 & 322 \\
\hline
Min (nm) & 300 & 288 & 300 & 288 \\
\hline
\end{tabular}
\caption{UV absorption maximum and minimum of 6-chloro-2-pyridone and N-methyl-2-pyridone under different solvent and temperature conditions.}
\end{table}
Figure 6.15: Variation of UV absorption spectra of 6-chloro-2-pyridone and N-methyl-2-pyridone with solvents (a) and temperatures (b,c). Water and ethanol were used in the solvent studies. The bottom panel shows the solvent difference spectra (EtOH-H₂O) and temperature difference spectra (high T - low T). The black dash lines are the solvent difference spectra to be compared with the thermal difference spectra.

The tautomeric equilibrium of 6-chloro-2-pyrdione shifts from lactam to lactim going from H₂O to ethanol, indicated by the peak maximum shift from 300 nm to 274 nm. Based on this solvent difference UV absorption spectrum, Bensaude and coworkers assigned the observed T-jump relaxation spectrum to be the result of tautomerization. However, we also observed solvent induced variation in the UV absorption spectra of N-methyl-2-pyridone, a model compound that does not tautomerize. Therefore, we recorded the equilibrium temperature difference absorption spectra of both compounds to compare to the solvent difference spectra. Interestingly, the thermal difference spectra share similar features with the solvent difference spectra. This observation has been noted by Johnson et al. and attributed the variations to the difference between two hydrogen-bonded solvent-solute complexes. UV absorption spectra for both chemicals change with temperature, resulting in difference spectra with one positive and one negative peak. The difference spectrum for N-methyl-2-pyridone is solely from the temperature-dependent change in dielectric constant while the response for 6-chloro-2-pyridone has contribution from change in dielectric constant as well as tautomerism. As a consequence, it is possible that the sub-μs response of 6-chloro-
2-pyridone contains a contribution from tautomerism, but is below their time resolution to be further studied. This result again demonstrates that IR spectroscopy allows clearer assignments and provides better structural resolution. Furthermore, we noticed that the T-jump amplitude spectrum reported in their publication (Fig. 2b in ref. [7]) has the opposite sign to the ones we measured. Their spectrum would suggest a growth of the lactam tautomer and loss of the lactim tautomer subject to T-jump, in contradiction to the equilibrium studies. As the time-dependent T-jump amplitude spectra and the data on N-methyl-2-pyridone were not provided in their publications, we cannot make more definitive comments on their results. We can only suspect that the T-jump UV absorption results are more complicated to interpret (distinction between solvent effects and tautomerism) than the authors concluded.

Given that our results are internally self-consistent, and yet so different from those of Dubois and coworkers, our only remaining explanation is that experimental differences may play a role in the findings. We note that in the early experiments the time resolution was limited to 5 μs, and that a significant sub-microsecond response was observed preceding the reported microsecond kinetics. Additionally, we point out that the T-jump triggering methods used in these two studies were different. It is possible that a T-jump method based on capacitor discharge autoionizes water and creates a reactive species which interfere with the ability to characterize the tautomerism of the substrate unambiguously. Clearly such questions can only be answered definitively by performing experiments with varying T-jump triggering method. We end by noting that our optical T-jump method creates a small perturbation of the system compared with a capacitive discharge and that 2D IR spectroscopy offers substantive structural information and direct probing of the molecules of interest.

6.5 Conclusion

In conclusion, we have used the T-jump transient 2D IR spectroscopy to provide an in-depth characterization of the ground state lactam–lactim tautomerization of pyridone derivatives in aqueous solution under ambient conditions. Using the distinctive cross-peak patterns, we can clearly distinguish the tautomeric forms. We track the time evolution of the lactam carbonyl peak following the T-jump to obtain the kinetics and calculate the activation thermodynamic variables. The concentration independence of the T-jump relaxation rate rules out the tautomerization mechanism via dimerization. The time constants of the single exponential t-2D IR signals of the lactam loss and lactim gain are nearly identical, indicating that the tautomerization proceeds through a two-state, concerted mechanism without a long-lived intermediate species. The comparison between the two pyridone derivatives, the solvent
dependence, and the pH dependence preclude intramolecular tunneling and strongly support proton transfer in water wires following a Grotthuss mechanism. We concluded that tautomerization of pyridone derivatives in water proceeds via the two-state concerted, and non-dissociative mechanism, and that the rate-limiting step is the release of proton into the solvating water environment.

6.6 Acknowledgments

I thank Carlos R. Baiz for his assistance in setting up T-jump experiments. I thank Mike Reppert and Kevin C. Jones for their stimulating discussions.
6.7 References


CHAPTER 7
Monitoring Rapidly Interconverting Tautomers of an Anti-HIV Agent KP1212

The work presented in this chapter is adapted from the following manuscripts:


7.1 Introduction

Many viruses utilize a high replication error rate to evolve rapidly and thereby adapt to environmental pressures. In a process called lethal mutagenesis to combat invading viruses, the innate immune system uses nucleobase deaminases such as APOBEC and ADAR\textsuperscript{1,2} to boost the already high viral mutation rate toward the viability threshold of the virus, sometimes leading to population collapse.\textsuperscript{3} The pharmaceutical analog of lethal mutagenesis uses mutagenic nucleoside analogs to chemically elevate the viral mutation rate and force the viral population to exceed its error catastrophe limit.\textsuperscript{4,7} Examples of this type of antiviral agents include ribavirin against hepatitis C virus\textsuperscript{8} and T705 against influenza.\textsuperscript{9} Additionally, deoxycytidine (dC) analogs have proven to be promising HIV drug candidates. 5-Hydroxydeoxycytidine (5-OH-dC) was the first identified molecule to display a lethal mutagenic property against HIV.\textsuperscript{16} 5,6-Dihydro-5-aza-2'-deoxycytidine (KP1212, Figure 7.1a),
the subject of this study, is the first compound intentionally designed to employ lethal mutagenesis against HIV.\textsuperscript{10-13} KP1212 has been shown to cause A-to-G and G-to-A mutations in biochemical studies\textsuperscript{10-12} and in a Phase IIa clinical trial.\textsuperscript{13}

Figure 7.1: (a) Structure of 2'-deoxycytidine and KP1212 in the keto-amino form. (b) Structure of the base portion of five possible neutral KP1212 tautomers. Here, cis-imino tautomers are shown, but they can also be present in the trans form. (c) pH-dependent FTIR spectra of KP1212 from pH\textsuperscript{*} = 1.6 to 13.9 at 25 °C. (d) The populations of the protonated, neutral, and deprotonated KP1212 as a function of pH\textsuperscript{*} obtained from the first three components of SVD analysis. The solid lines show the results from simultaneous fitting of the Henderson-Hasselbalch equation with pKa\textsubscript{1} and pKa\textsubscript{2}. (e-g) SVD reconstructed spectra representative for the protonated, neutral and deprotonated KP1212 spectra. The grey curves in (e,f) are the experimental FTIR spectra of CMP at pH\textsuperscript{*} 1.6 and 7.4, respectively.

Despite its promising antiviral activity, the molecular mechanism by which KP1212 promotes mutagenesis has been elusive. The leading proposal is the rare tautomer hypothesis, which states that KP1212 exists in multiple tautomeric forms (Figure 7.1b) and each tautomer displays a distinct base-pairing preference during replication which eventually results in the observed mutations.\textsuperscript{12-14} In the case of 5-OH-dC (also a G-to-A mutagen),\textsuperscript{15,16} Raman spectroscopic analysis identified a < 1% population of anionic imino-keto tautomer,\textsuperscript{17} which was proposed to base-pair with A. In this study we employ two-dimensional infrared spectroscopy (2D IR) to show the presence of multiple KP1212 tautomers in aqueous solution at physiological temperatures, and address key unknowns regarding the mechanism by which
KP1212 induces mutagenesis, such as the temperature and pH dependence of KP1212 tautomer populations as well as their interconversion rates.

Studying tautomerization of heterocycles under physiological conditions is difficult and requires a technique with both high structural sensitivity and with sufficient time-resolution to distinguish rapidly interconverting tautomers. Vibrational infrared and Raman spectroscopy are promising techniques in this regard because they are sensitive to the changes in functional groups, and they have intrinsic picosecond time resolution. However, it remains challenging to separate multiple species with these methods due to peak overlap and uncertainty in peak assignments. 2D IR was recently shown to offer a unique way for unambiguously separating the tautomers of pyridone derivatives, and for measuring the tautomerization kinetics in temperature-jump (T-jump) experiments. In this work, infrared spectroscopy was applied to characterize the tautomerization of KP1212, using Fourier transform infrared (FTIR) to measure the ionic distributions, 2D IR to distinguish among various tautomers, and density functional theory (DFT) calculations to assist in spectral assignments. We found that the neutral KP1212 exists primarily in the unexpected enol-imino form and interconverts with the keto form on the 20 nanosecond timescale. The enol tautomers can in principle base-pair with adenine, which offers strong support for the tautomerization hypothesis for lethal mutagenesis by KP1212. Moreover, the data reveal that KP1212 has unique shape shifting properties that afford rich opportunities to probe the impact and roles of tautomeric equilibria in biological systems.

7.2 Results

7.2.1 Extracting the pKa’s from the pH-dependent FTIR spectra

Because tautomerization is closely related to the molecule’s protonation state, we first investigated the pH dependence of KP1212 between pH* 1.6 and 13.9, where pH* refers to standard glass electrode readings for deuterated water solutions. The pH-dependent FTIR spectra (Figure 7.1c) display intricate behavior in the in-plane base vibration region from 1500 to 1750 cm⁻¹. This region contains the carbonyl stretching window 1630-1750 cm⁻¹ and the combinations of C=C and C=N stretches from 1500-1650 cm⁻¹. At low pH*, an intense peak at 1662 cm⁻¹ and a broader 1703 cm⁻¹ peak are observed. As the pH* increases, these two peaks decrease in intensity concomitantly with the appearance of three lower frequency modes. At high pH*, the spectrum is dominated by the appearance of a broad peak at 1569 cm⁻¹.
To analyze the pH dependence, we performed singular value decomposition (SVD), which linearly decomposes these spectra into basis spectra that share common pH dependence. Each spectrum at a specific pH has contributions from the various tautomers of KP1212 at different protonation states. Our goal is to identify these “basis spectra” or the difference spectra between each titratable species in the mixture. We introduce a matrix $D$ whose columns consist of these difference spectra and one reference spectrum. A transformation matrix $F$ specifies the pH dependence of each difference spectrum and the mixing of the columns of $D$ to form $A$, the pH-dependent spectra. The columns of $F$ contain the Henderson-Hasselbalch equations for the individual transitions and the last column contains ones for the constant reference spectrum. The transformation follows:

$$DF^T = A$$  \hfill (7.1)

The raw data, which can be written as the sum of matrix $A$ and the experimental noise $E$, contains information in the frequency space (rows) and pH space (columns). SVD is used to separate them into individual matrices, denoted as $\bar{U}$ (frequency) and $\bar{V}$ (pH), following:

$$A + E = \bar{U}S\bar{V}^T$$  \hfill (7.2)

The $S$ matrix generated from the SVD analysis is square and diagonal. The diagonal values of $S$ are called the singular values and indicate the weights for the corresponding columns of $\bar{U}$ and $\bar{V}$. The singular values start from a high value and decrease to a low value plateau which indicates the noise level. All of the essential information about the system should be contained in the first few component spectra with high singular values. Figure 7.2a and b show the first five components from the $\bar{U}$ and $\bar{V}$ matrices, along with the corresponding normalized singular values. We found three significant spectral components, and therefore the data set can be described with a reduced 3x3 matrix:

$$DF^T = A = USV^T$$  \hfill (7.3)

where $U$, $S$, and $V$ represent the reduced SVD components. The goal is to solve for $D$ which contains the true difference spectra between the individual tautomer spectra. Multiplying both sides of the equation by the pseudoinverse of $F^T$ gives

$$D = USV^T \left(F^T \right)^+ = USH$$  \hfill (7.4)
Now we fit the experimentally determined V components to find the F matrix, whose first two columns consist of the two Henderson-Hasselbalch equations with pKa, and pKa2, respectively, and the third column is a column of ones to represent the background spectrum. The Henderson-Hasselbalch terms are expressed as functions of pH,

\[
C_i [1 + 10^{(pK_a - pH)}]^{-1}
\]

(7.5)

where \(C_i\) is a constant fitting parameter. We fit the V components (weighted by the appropriate square of the singular value) simultaneously by minimizing the sum of the squares of the residuals (Figure 7.2c), and the resulting pKa, = 7.0 and pKa2 = 13.4.

After the best F is found, the D matrix can be reconstructed as follows,

\[
D = A (F')^T
\]

(7.6)
These three D components displayed in Figure 7.2d represent one reference spectrum, and two difference spectra. It is clear that D₃ should be the reference spectrum due to the absence of negative peaks. The blue-shifted peak at 1704 cm⁻¹ suggests that it arises from the C=O stretch of the protonated keto tautomer, therefore D₃ is assigned to the “protonated” spectrum. Further inspection informs us that D₁ represents the difference spectrum between the “protonated” and “neutral” spectra, and that D₂ represents the difference spectrum between the “neutral” and “deprotonated” spectra.

Finally, to obtain the spectra for protonated, neutral, and deprotonated KP1212, we apply a transformation matrix W on D and F.

\[ A = DF^T = (DW)(W^{-1}F^T) = BY^T \]  

(7.7)

where the columns of B (Figure 7.1e-g) contain the “protonated”, “neutral”, and “deprotonated” spectra, and Y (Figure 7.1d) shows their associated populations.

The first pKₐ (pKₐ₁ = 7.0) reports on the protonation at N3, and is significantly higher than the value of 4.3 for dC. Protonated KP1212 is in the keto form as indicated by the presence of the high frequency C=O vibration at 1703 cm⁻¹, which can be compared to the C=O frequency of cytidine 5'-monophosphate (CMP) that shifts from 1651 to 1709 cm⁻¹ when protonated (spectra at pH 1.6 and 7.4 are shown in Figure 7.1e and f, respectively). Much like the guanidinium cation, the presence of N5 and the sp³ center at C6 in KP1212 stabilizes the extra positive charge by resonance structures, leading to an increase of pKₐ. The second pKₐ (pKₐ₂ = 13.4) corresponds to the deprotonation of KP1212, and the formation of enolate ions is revealed by the complete loss of the C=O stretch.

The interpretation of the neutral spectrum is more challenging. In principle, neutral KP1212 can interconvert between the five tautomers shown in Figure 7.1b by keto–enol or amino–imino tautomerization. Also, cis and trans imine isomers may be present. Although the canonical form of dC is the keto-amino tautomer, the SVD reconstructed spectrum for neutral KP1212 (Figure 7.1f) shows only one peak above 1650 cm⁻¹ with very low intensity when compared with the strong C=O peak in CMP, indicating a strongly reduced keto population. This observation appears at odds with the proposed role of the keto-imino (KI) tautomer in the mutagenicity of KP1212.

### 7.2.2 Separating tautomers using 2D IR spectra

In order to gain more insight into the spectra, we acquired 2D IR spectra at different pH* and temperatures (Figure 7.3). The 2D IR spectrum of KP1212 at 37 °C and pH* 7.9
(Figure 7.3d) shows a pronounced grid of cross-peaks between the three low frequency modes at 1529, 1572, and 1612 cm⁻¹, indicating that these vibrations originate from the same molecular species. However, the 1662 cm⁻¹ C=O peak does not have cross-peaks to the other modes, suggesting that this vibration is due to a chemically distinct species. This is in sharp contrast to the 2D IR spectrum of CMP (Figure 7.3b), in which cross-peaks between all of the diagonal peaks occur. Because of the characteristic carbonyl stretch frequency, we assign the 1662 cm⁻¹ mode of KP1212 to a keto C=O vibration, and the three lower frequency modes to the enol tautomers.

Further evidence for this assignment comes from low pH* spectra. The pH*=1 2D IR spectrum of KP1212 resembles that of CMP (Figure 7.4), with three diagonal peaks and cross-peaks between all of them. We conclude that both protonated bases are in the keto-amino form based on the blue-shifted C=O peak (~1700 cm⁻¹). Because of its pKₐ at 7.0, KP1212 exists in a mixture of protonated and neutral forms in the physiological pH range of 6–8. The IR spectra at pH* 6.6 (Figure 7.3e, f) show that compared to pH* 7.9, the keto population is elevated while the enol population is reduced. The protonated keto-amino tautomer manifests itself in the 2D IR spectrum as the peaks highlighted by the pink grids.
The remaining neutral KP1212 population is distributed between the keto and enol tautomers, as shown by the presence of the modes with frequencies <1640 cm\(^{-1}\).

![Figure 7.4: FTIR (top) and 2D IR (bottom) spectra of CMP (left) and KP1212 (right) at pH* 1.0 at 25 °C.](image)

The temperature dependence of KP1212 (Figure 7.3c, e) also provides strong evidence for the existence of multiple tautomers. Unlike the variable temperature FTIR of CMP (Figure 7.3a) which show marginal changes, the KP1212 spectra exhibit dramatic intensity variation for multiple peaks and exhibit clear isosbestic points. Opposite temperature trends were observed at these two pH* values: upon heating, the keto tautomers increase at pH* 7.9 while the enol tautomers grow at pH* 6.6. An analysis of the temperature-dependent spectra allows us to characterize the thermodynamic parameters governing their equilibrium. Although 2D IR peak heights are proportional to concentration and can help constrain the fitting, interference effects between peaks complicate the quantification of weakly populated tautomer. We therefore analyzed the FTIR spectra with a two-state model that considers only the broader categories of keto and enol species, i.e. \( K \rightleftharpoons E \). It was found that at 37 °C, KP1212 exists in 66% of enol form at pH* 7.9 and 75% of protonated keto form at pH* 6.6.

7.2.2.1 pH* 7.9 data set

For the pH* 7.9 temperature-dependent FTIR spectra, the SVD analysis (Figure 7.5c) reveal two significant components. A two-state model using the first two SVD components was applied to reconstruct two physically-relevant spectra that can represent the two species. Since the temperature profiles obtained from the SVD analysis (Figure 7.5d) are not well-defined functional forms, as with the pH titration profiles extracted from the pH-dependent
FTIRs (Figure 7.2c), a direct fitting of these temperature profiles to Boltzmann distribution does not give robust results. This is because Boltzmann distribution for a two-state system can be approximated as a linear line within a small temperature range.

Figure 7.5: (a) Temperature-dependent FTIR of KP1212 at pH* 7.9 (same as shown in Figure 7.3c). (b) Thermal difference spectra by subtracting the 10 °C spectrum. (c) The first four SVD spectral components with the associated singular values listed on top of each component. (d) The corresponding thermal profile of the four SVD spectral components.

Alternatively, we chose to fit the spectral components by using a matrix transformation to link the mathematically calculated spectra to physically meaningful spectra. However, because in reality we are not able to separate out pure tautomers and measure their spectra, we will not be able to fit the reconstructed spectra to the spectra of individual tautomers. As a result, we introduced constraints during the spectral fitting. The first constraint is the non-negativity of these spectra, which is a loose constraint. The second constraint is that certain frequency range of the individual spectrum has to be zero, and this constraint comes from the inspection of the experimental FTIR and 2D IR spectra, as well as comparison to the DFT calculations. This reconstruction procedure is subjective in a sense—the peak assignment to a particular tautomer can influence the reconstruction result. The reconstructed spectra are shown in Figure 7.6a and b, and their corresponding populations are shown in Figure 7.6c. The blue spectrum accounts for the dominant species at this pH, and can be assigned mostly to the enol tautomer because of its characteristic triplet peak structure. The red spectrum can represent the sum of the different keto tautomers based on the broad C=O peak observed ~1650 cm⁻¹. The features seen at ~1550 cm⁻¹ could be attributed to peak shifting, broadening or contributions from other enol tautomers that grow in with temperature (Figure 7.11 shows
that different enol tautomers can have slightly different ring mode frequencies). With the populations of the two species obtained, a van't Hoff plot (Figure 7.6d) can be used to extract the thermodynamic variables listed in Table 7.1.

![Figure 7.6: The results from the SVD reconstruction using a two-state model: (a) the reconstructed spectrum representing the enol tautomer; (b) the reconstructed spectrum representing the keto tautomer, and possibly other minor enol tautomers. (c) The populations of the two species shown in (a) and (b) as a function of temperature. (d) The van't Hoff plot.](image)

**Table 7.1.** Thermodynamic variables obtained from the SVD two-state analysis of the temperature-dependent FTIR spectra. The values are shown for a general keto to enol tautomerization.

<table>
<thead>
<tr>
<th></th>
<th>pH* 7.9 (keto $\rightarrow$ enol)</th>
<th>pH* 6.6 (protonated keto $\rightarrow$ enol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta H$ [kcal/mol]</td>
<td>-1.3</td>
<td>3.9</td>
</tr>
<tr>
<td>$\Delta S$ [cal/mol/K]</td>
<td>-2.9</td>
<td>16.8</td>
</tr>
<tr>
<td>$\Delta G^0$ (37 °C)</td>
<td>-0.4</td>
<td>0.69</td>
</tr>
<tr>
<td>$K_{eq}$ (37 °C)</td>
<td>1.78</td>
<td>0.33</td>
</tr>
</tbody>
</table>

In the present analysis, we did not account for the contribution of thermally-induced frequency shift, which can also result in spectral features of peak gain and loss. However, this contribution is believed to be small as small frequency shifts are observed in the temperature-dependent FTIR spectra of CMP. We have also tried three-state and four-species models in an attempt to further separate the red spectrum into more distinct species. Unfortunately, we were not able to reconstruct physically meaningful spectra that are all positive across the spectral ranges. Therefore, the two-state analysis seems to be the most reasonable given the small populations of the minor tautomers which cannot be accurately quantified from our experiments. From the two-state analysis, we obtained 66% of enol at 37 °C. We have tested
with simulations that if the real system consists of four tautomers, using a two-state analysis can result in as much as 10% discrepancy in the calculated populations (Appendix 7.A).

In order to provide further support the populations obtained from the two-state SVD analysis, we compared the FTIR spectra of KP1212 to that of 5-aza-2'-deoxycytidine (5-aza-dC, structure shown in Figure 7.7b). 5-aza-dC is the precursor of KP1212 in the keto-amino form, we therefore assume that the keto C=O stretch of 5-aza-dC and KP1212 have the same dipole strength. The keto population of KP1212 can then be estimated by comparing the integrated C=O peak intensity (shown as blue dash curves). It was found to be ~30% at 25 °C, which is close to 35% estimated from the two-state SVD analysis. This result suggests the validity of the SVD analysis. The 10% protonated keto tautomer at this pH also contributes to the intensity measured in the C=O stretch region as shown by the small peak at 1660 cm⁻¹.

![Figure 7.7: FTIR spectra of 20 mg/ml of KP1212 (a) and 5-aza-dC (b). The experimental spectra (black circles) were fitted to the sum of Voigt profiles shown in red (individual peak in dash blue curves).](image)

7.2.2 pH* 6.6 data set

Similar analysis was applied to the pH* 6.6 data set, and the results are shown in Figure 7.8 and Figure 7.9. The reconstructed spectrum plotted in blue in Figure 7.9a represents the spectrum for protonated keto tautomer, and can be compared to the low pH FTIR spectrum. On the other hand, the spectrum in red (Figure 7.9b) represents the EI spectrum. It is likely that this spectrum contains contributions from the other tautomers since the spectrum does not match perfectly with the EI tautomer spectrum obtained from the SVD analysis of the pH* 7.6 data set, in particular ~1550 cm⁻¹ region.
Figure 7.8: (a) Temperature-dependent FTIR of KP1212 at pH* 6.6 (same as shown in Figure 7.3e). (b) Thermal difference spectra by subtracting the 10 °C spectrum. (c) The first four SVD spectral components with the associated singular values listed on top of each component. (d) The corresponding thermal profile of the four SVD spectral components shown in (c).

Figure 7.9: The results from the SVD reconstruction of the pH* 6.6 spectra using a two-state model: (a) the reconstructed spectrum representing the protonated keto tautomer; (b) the reconstructed spectrum representing the enol tautomer, and possibly other minor enol tautomers. (c) The populations of the two species shown in (a) and (b) as a function of temperature. (d) The van't Hoff plot.

7.2.3 Peak assignments based on DFT calculations

To assign vibrations to specific tautomers, we calculated the IR absorption spectra for the five tautomers shown in Figure 7.1b using DFT calculations of harmonic vibrations at the
B3LYP/6-31G(d,p) level. Similar to previous observations, we found that the number and configuration of explicit water molecules included in these calculations significantly influence the basic qualitative spectral features (Figure 7.10). Calculated spectra in vacuo (Figure 7.10a) show only two significant peaks in the spectral region of interests for each of the five tautomers, which clearly cannot explain the pronounced cross-peaks between the three low frequency modes observed in the 2D IR spectra. As more explicit water molecules were placed near the hydrogen bond donor and acceptor sites to solvate KP1212, the calculated spectra change dramatically (Figure 7.10b), especially for the enol tautomers. Figure 7.10c shows the progression of the calculated spectra for the cis-enol-imino (cEI) tautomer as the number of explicit increases. Upon solvation, the coupling of the O-D stretch of D$_2$O and the KP1212 ring modes gives rise to three significant peaks in the 1500—1600 cm$^{-1}$ spectral region that resemble the triplets seen in the experimental spectra. The positions of the explicit water molecules greatly influence the resulting IR absorption spectra due to different degrees of couplings between the solvent and the solute. In principle, we can use polarization dependent 2D IR spectra to measure the transition dipole orientations to help improve the simulated spectra. This finding should also invoke further theoretical studies to investigate the physical origins of these spectral variations, such as intermolecular coupling or electrostatic interactions.

Figure 7.11 displays the calculated IR spectra with five explicit water molecules. At this level of hydration we feel that vibrational frequencies within ±30 cm$^{-1}$ can be used for assignment of tautomers, recognizing that anharmonic effects such as Fermi resonances in experimental spectra would not be captured by the calculations.
Figure 7.10: Calculated IR absorption spectra for the five possible KP1212 tautomers without (a) and with (b) five explicit water solvent molecules. The \textit{cis} isomers of the imino tautomers are shown. (c) Calculated spectra for the \textit{cis}-enol-imino tautomer with different microsolvation environments. Cyclopentane was used to replace the ribose moiety.

Qualitative similarities between calculated and experimental spectra support the separation of spectra into keto and enol species. In particular, the triplet pattern of the \textit{cis}-enol-imino (cEI) spectrum at 1502, 1537, and 1591 cm\(^{-1}\) matches qualitatively with the peak positions in the 1500-1620 cm\(^{-1}\) region of the experimental spectra (blue grids in Figure 7.3d). Although neither the \textit{cis} nor \textit{trans} isomer calculations predict the experimentally observed intensity pattern, we find that the calculated intensities are more easily influenced by solvent structural details than frequencies. Keto-amino-\textit{N}^3 (KA-N\textit{N}^3) and keto-imino (KI) calculations both predict two transitions >1600 cm\(^{-1}\). Keto-amino-\textit{N}^3 (KA-N\textit{N}^3) is the only keto tautomer predicted to have a strong transition near 1550 cm\(^{-1}\), which should result in a strong cross-peak from \(~1550\) cm\(^{-1}\) to the C=O stretch. Such a cross-peak is not observed in any of the experimental spectra containing only the uncharged KP1212, meaning that if any KA-N\textit{N}^3 is present, its population fraction is estimated to be <5%.

204
Figure 7.11: Experimental (left) and DFT calculated (right) IR absorption spectra for the different KP1212 species with five explicit water molecules. The frequencies have been scaled by 0.9614 and $\sigma = 5$ cm$^{-1}$ Gaussian broadening of the lines has been applied. Spectra for the cis and trans imino tautomers are plotted with solid and dash lines, respectively.

Table 7.2. DFT calculated vibrational frequencies of the various KP1212 tautomers. The frequencies have been scaled by 0.9614.

<table>
<thead>
<tr>
<th>$[\text{KA}^+]$</th>
<th>$\text{KA-N}_3$</th>
<th>$\text{KA-N}_3$</th>
<th>cKI</th>
<th>tKI</th>
<th>EA</th>
<th>cEI</th>
<th>tEI</th>
<th>$[\text{EA}]^-$</th>
<th>$[\text{cEI}]^-$</th>
<th>$[\text{tEI}]^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1708</td>
<td>1634</td>
<td>1657</td>
<td>1678</td>
<td>1712</td>
<td>1631</td>
<td>1591</td>
<td>1563</td>
<td>1585</td>
<td>1607</td>
<td>1626</td>
</tr>
<tr>
<td>1635</td>
<td>1544</td>
<td>1626</td>
<td>1589</td>
<td>1538</td>
<td>1537</td>
<td>1518</td>
<td>1545</td>
<td>1548</td>
<td>1529</td>
<td></td>
</tr>
<tr>
<td>1530</td>
<td>1512</td>
<td></td>
<td>1498</td>
<td>1502</td>
<td>1498</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

205
In addition to the results for neutral KP1212, we also found very good agreements between the experimental and calculated spectra for protonated and deprotonated species. Protonated KP1212 exists in the keto-amino (KA) form whereas deprotonated KP1212 are in the enol-amino (EA) and enol-imino (EI) forms. The calculation shows that protonated KA has a C=O stretch peak blue-shifted compared to the neutral KP1212, which is consistent with the experimental spectra. The peak positions and relative intensities of the two ring modes also match the experimental peaks. In the case of deprotonated KP1212, the combination of the doublet peak of EA and the intense peak of EI can explain the broad peak in the experimental spectrum. The agreement between the experimental and DFT calculations for the protonated and deprotonated cases provides further support for the assignment on the neutral KP1212.

7.2.3.1 Minor keto tautomer

Assignment of minor tautomers requires careful analysis of 2D lineshape contours, which we performed at pH* 8.9 to eliminate contributions from the protonated keto tautomer. We present zoom-ins to the keto and enol regions in Figure 7.3g and h. Figure 7.3g shows that the C=O peak at 80 °C is about two times broader than that of CMP (Figure 7.3b). The peak (FWHM = 46 cm⁻¹) exceeds typical broadening of C=O transitions due to solvent hydrogen-bonding,²² implying the presence of two overlapped keto transitions from separate tautomers.¹⁹ The thermal difference FTIR spectra show an additional peak at 1675 cm⁻¹ (Figure 7.5b), which cannot be attributed to a simple blueshift of C=O stretch because a corresponding bleach on the red side is not observed. It has been reported¹⁷,²³ that the C=O frequency of N3-methylcytidine, which adopts the keto-imino form, is blue-shifted to ~1670 cm⁻¹, suggesting that the 1675 cm⁻¹ peak in the KP1212 spectrum could be the KI tautomer.

A cross-peak from the C=O stretch to a weak peak between 1620-1640 cm⁻¹ is suggested by a ridge extending along ω₁ at ω₃ = 1675 cm⁻¹. To confirm the presence of this cross-peak, we fit the experimental spectrum with the sum of tilted 2D Gaussian peaks. The experimental spectrum used in the fit (Figure 7.12a) was taken at 1 ps waiting time in order to enhance the cross-peak structure. We fixed a single peak at 1612 cm⁻¹ (Figure 7.12f) to represent the ring mode from the EI tautomer, which does not have any cross-peaks in this spectral region. In the first model, we assumed that there is only one keto tautomer with one C=O stretching vibration. The best fit using two uncoupled peaks (Figure 7.12b) shows that the ridge along ω₃ = 1675 cm⁻¹ could not be captured.
Figure 7.12: (a) Carbonyl stretch region of the experimental 2D IR spectrum of KP1212 at pH* 8.9 and 80 °C. Waiting time was set to 1 ps in order to enhance the weak cross-peak features. Spectrum with 150 fs waiting time is shown in the SI. (b, c) Fits to the experimental spectrum using 2D Gaussian profiles with the absence (b) and presence (c) of cross-peaks. In both fits, a peak at 1612 cm⁻¹ is fixed for the EI vibration shown in (e). In (b), only one peak is included to simulate the broad C=O lineshape. In (c), two different keto tautomers each with two vibrations are considered, as shown in (d) and (e).

In the second model which is guided by the DFT calculations, we considered both the KA-N³ and KI tautomers. The KA-N³ tautomer has a C=O stretch at 1664 cm⁻¹ and a ring mode at 1645 cm⁻¹; however, as expected, the two modes and their cross-peaks overlap to give rise a broad peak shown in Figure 7.12d. The KI tautomer (Figure 7.12e) has a C=O stretch at 1675 cm⁻¹ and a ring mode at 1630 cm⁻¹. The best fit shown in Figure 7.12c is the superposition of Figure 7.12d-f, with a 7:1 KA-N³ to KI ratio. A better qualitative agreement between the experimental and calculated lineshapes was achieved with this model. The small KI population explains why the cross-peak appears as a ridge instead of a distinct peak. Its breadth and the diffuseness of the cross-peak may reflect a mixture of the cis and trans isomers. Although we cannot guarantee the uniqueness of the spectral decomposition, the fit presented is consistent with the DFT calculations and offers a reasonable explanation to the experimental spectrum.
7.2.3.2 Minor enol tautomer

Whereas a cursory analysis reveals that the EI tautomer is the dominant species, a closer look at the enol region of the 2D IR spectrum indicates that multiple species are present. In Figure 7.3d, the position of negative and positive peaks does not align on the same vertical excitation frequency (for instance $\omega_1 \approx 1530$ cm$^{-1}$) indicating multiple species in the sample. Since the imino tautomers can be either in the trans or cis form, and the variations in frequency of their vibrational transitions are on the order of the vibrational linewidths, the observed broad peaks could possibly be explained by a superposition of these isomers (e.g. feature 1 in Figure 7.3h). Similarly, it is possible that the enol-amino (EA) tautomer is present at low population. The calculation predicts that EA also has a three peak pattern like EI, but with two modes of different intensity that overlap with the EI modes at 1498 and 1538 cm$^{-1}$, and a blue-shifted ring mode at 1631 cm$^{-1}$. The 2D IR spectrum of KP1212 at 4 °C and pH* 8.9, a condition under which the enol tautomers are favored, shows an additional diagonal peak with weak intensity at 1638 cm$^{-1}$ (feature 2 in Figure 7.3h). This feature could be due to the inhomogenous broadening of the 1621 cm$^{-1}$ ring mode; however, when we increased the waiting time $\tau_2$ to allow spectral diffusion to symmetrize the peaks, feature 1 persists and exhibits spectral diffusion as well. This spectral signature indicates that it is a distinct peak.

![Image of 2D IR spectra at selected waiting times](image)

Figure 7.13: 2D IR spectra at selected waiting times of KP1212 at pH* 8.9 and 4 °C. All-parallel polarization was used to enhance the diagonal features.

Moreover, fitting a frequency slice along $\omega_1$ at $\omega_1 = 1529$ cm$^{-1}$ (Figure 7.14) suggests a cross-peak feature at (1638 cm$^{-1}$, 1529 cm$^{-1}$). The vibrational band pattern of the new species at low temperature resembles the calculated EA spectrum.
In order to further support our assignment, we simulated 2D IR spectra to illustrate how the weakly populated \textbf{EA} tautomer may manifest itself as the spectral features described above. 2D IR spectra for \textbf{EA} and \textbf{EI} were generated using 2D Gaussian functions and shown in Figure 7.15. The relative peak positions were guided by the DFT calculations and the exact frequencies were set to match the experiment. Both \textbf{EA} and \textbf{EI} have two ring vibrations at 1529 cm\textsuperscript{-1} and 1572 cm\textsuperscript{-1}, and \textbf{EA} has a ring mode at 1638 cm\textsuperscript{-1} that is blueshifted compared to the \textbf{EI} ring mode at 1612 cm\textsuperscript{-1}. The peak intensities and linewidths are arbitrary chosen to roughly match the experiment. The \textbf{EA} and \textbf{EI} spectra were added with a 1:9 ratio, and the resulting spectrum displays similar spectral features identified in Figure 7.14 (the corresponding features are marked). The $\omega_3 = 1529$ cm\textsuperscript{-1} slice is plotted on top of the spectrum and it shows a weak bump at 1638 cm\textsuperscript{-1}.

Figure 7.15: The simulated 2D IR spectra using tilted 2D Gaussian functions for the \textbf{EA} (left) and \textbf{EI} (middle) tautomers, and the superposition of the two spectra (right). The $\omega_3 = 1529$ cm\textsuperscript{-1} slice is plotted on top of the spectrum.
7.2.4 Tautomeric distribution probed by NMR spectroscopy

Taken together, the combination of the experimental constraints from 2D IR spectra and the DFT calculations provides consistent evidence that multiple tautomers co-exist for KP1212 and that the enol-imino tautomer is the major species under physiological pH and temperature. The presence of multiple tautomer under physiological condition derived from 2D IR spectroscopy is consistent with our variable temperature NMR spectra taken in DMF-d7 (Figure 7.16b). At 20 °C, the exchangeable protons of KP1212 give rise to three broad peaks denoted i', ii' and iii' in Figure 7.16b. When the temperature is decreased to -50 °C, the three broad peaks resolve into six distinct proton resonances (denoted as i-vi). Because each KP1212 tautomer only has a total of three exchangeable protons on the base moiety, the observation of six proton resonances indicates the presence of a minimum of two tautomeric forms. Detailed analysis allowed the assignment of these peaks to imino (colored blue in Figure 7.16a), amido (purple), enolic (red) or amino (green) protons.

**Figure 7.16:** (a) Structures of the five possible tautomeric forms of KP1212. The active protons (a-o) on the nucleobase portion of the molecule are designated with different colors to indicate their chemical environment (type): blue (imino), purple (amido), red (enol), and green (amino). (b) Variable temperature ¹H NMR spectrum of KP1212 in DMF-d7. (c) Schematic of the deconvolution process of the ¹H NMR spectrum of KP1212 at -50 °C.
With the peak assignment, deconvolution of the NMR spectrum at -50 °C allows quantification of the tautomeric species (Figure 7.16c). We should note that the trans and cis imino protons have the same chemical shift, and therefore are treated together. Even though some of the peaks overlap, there is enough experimental constraint for us to obtain a unique solution. The main advantage of deconvoluting NMR spectra is that all the protons have signals of the same intensity, a luxury feature that does not hold true in IR spectra (i.e. different vibrations have different oscillator strengths). Nevertheless, due to the rapid tautomerization, different tautomers cannot be resolved in aqueous solution under ambient temperatures. The deconvolution at -50 °C shows the presence of all five tautomers and that 51% of KP1212 exists as the enol-amino form and 38% in the enol-imino form. It is not unexpected that the tautomeric distribution from the two experimental conditions (D₂O, 37 °C v.s. DMF-d₇, -50 °C) are different. Nevertheless, combining results from both IR and NMR spectroscopy, we obtained compelling evidence of the presence of multiple KP1212 tautomers, with the unexpected enolic form dominating under both experimental conditions.

### 7.2.5 Measuring the tautomerization kinetics with T-jump experiments

With a characterization of the KP1212 tautomers present in solution, we turn to measure tautomer exchange kinetics, which crucially influence tautomer variation during the steps of nucleobase diffusion, recognition and addition in DNA replication. Tautomer interconversion has proven difficult to measure due to the fast time-scales for the process. For this purpose, we employed T-jump transient 2D IR (t-2D IR)²⁴ to characterize time-dependent tautomer populations.

Figure 7.17a displays the t-2D IR spectrum at 15 ns after the T-jump from an initial temperature of 27 °C. As illustrated by the cartoon spectra in Figure 7.17b, a loss of population appears in t-2D IR difference spectra as a doublet with a negative (blue) peak on the top and with a positive (red) peak on the bottom. For example, boxes A and B mark the loss of the vibrational modes of the protonated keto tautomer. On the other hand, boxes C-E show the gain features of the enol peaks. These spectral signatures indicate that upon heating, the enol population increases at the expense of the protonated keto tautomer. Although we present the T-jump results obtained at pH* 6.6 which show clear signature of the loss of protonated keto species, similar kinetics were observed for the neutral keto species at pH* 7.9 with smaller magnitude (Figure 7.18), indicating that the conversion to both the protonated and neutral keto species takes place on similar timescales and could both be occurring at pH* 6.6.
Figure 7.17: Transient 2D IR results at pH* 6.6 and T-jump from 27 °C. (a) Transient 2D IR spectrum (ZZZZ) at 15 ns T-jump delay. (b) Cartoon 2D IR spectra to illustrate how gain and loss features are seen in a transient spectrum. (c) Transient HDVE difference spectra at selected T-jump delays. (d) Time evolution of the 1520 and 1654 cm⁻¹ peaks (dotted lines in panel c), representing the enol and keto species, respectively. (e) The Arrhenius plot.

Figure 7.18: (a) t-HDVE spectra of KP1212 at pH* 8.1 and T-jump from 27 °C. (b) Time evolution of the signal at ω₁ = 1659 cm⁻¹ and 1685 cm⁻¹. Single-exponential fits give time constants of 21.3±6.1 ns and 25.3±16.7 ns for 1659 cm⁻¹ and 1685 cm⁻¹, respectively.

The t-HDVE spectra (Figure 7.17c) resemble the thermal difference FTIR spectra (Figure 7.8b), with the lower and higher frequency ranges reporting on the gain of enol tautomers and the loss of keto tautomers, respectively. The keto-enol tautomer exchange kinetics can be extracted from the time evolutions of the 1654 and 1520 cm⁻¹ peaks, which were fitted to single exponential functions (Figure 7.17d). From the relaxation time constants of 20 ns and 19 ns for keto and enol, respectively, we find that the loss of keto and gain of enol are well correlated, suggesting that the kinetics can be described by a two-state exchange between keto and enol tautomers without significant metastable intermediates. T-jump relaxation rates were found to increase with temperature as shown in Figure 7.17e. Applying
a two-state kinetic analysis for keto-to-enol exchange gives the activation barriers as $\Delta E_{K\rightarrow E} = 10.7 \text{ kcal/mol}$ and $\Delta E_{E\rightarrow K} = 4.8 \text{ kcal/mol}$.

### 7.3 Discussion

#### 7.3.1 Proposed mechanism for the lethal mutagenesis of KP1212

It has been well established that nucleobases are stable in their canonical keto-amino form in aqueous solutions. Nevertheless, Watson and Crick suggested that if nucleobases occasionally take on the non-canonical enol or imino tautomeric forms, the hydrogen-bond (H-bond) interactions can be altered, leading to formation of base-pair mismatches and spontaneous mutations. Direct structural evidence for the mismatch has been challenging to obtain due to the low frequency of minor tautomers, and therefore the rare tautomer hypothesis remains as a subject of intense speculation. Recently, a crystal structure of a DNA polymerase with a C-A mismatch at the base incorporation site was reported. Though the exact proton positions could not be determined, the Watson-Crick (WC) geometry adopted by the mismatch suggested the presence of imino tautomer of either C or A. The present study provides concrete evidence of the existence of multiple tautomers for the mutagenic nucleoside analog KP1212, and shows that the normally rare enol-imino tautomer dominates under physiological conditions. The findings support the proposal that the observed mutations result from tautomerism and also demonstrate that the tautomeric equilibrium of nucleobases can be engineered with chemical modifications.

Based on the presence of multiple tautomers, we outline a number of possible base-pairing scenarios that could influence the outcome of a replication or transcription step involving KP1212. Since KP1212 is a dC analog which should pair with G and that the observed mutations are G-to-A or A-to-G, we examine the possibility of mispair between A and the newly identified KP1212 tautomers. Several factors influence the base-pairing stability, notably H-bond, base-pair geometry, and base-stacking. We consider only how KP1212 may doubly or triply H-bond with the canonical DNA bases. Previously, KP1212 was believed to be structurally similar to C in the keto-amino form, which makes favorable H-bonds with G either in a WC or a wobble geometry depending on whether the N5 or N3 atom is protonated (Figure 7.19a, b). The G-to-A mutation has been proposed to be the result of base-pairing between the KP1212 keto-imino tautomer and A (Figure 7.19d). However, our studies show that the prevalent KP1212 species under physiological conditions is the normally rare enol-imino tautomer. When positioned in a wobble position, the EI tautomer can form two.
H-bonds with A (Figure 7.19f). Similarly, the enol-amino tautomer could potentially pair with A with the same H-bond configuration (Figure 7.19e). These observations offer additional explanations for the role of tautomerism in KP1212's promiscuous base-pairing ability and mutagenicity.

![Proposed base-pairing geometries between the different KP1212 tautomers and either G or A that may explain the observed lethal mutagenesis.](image)

Figure 7.19: Proposed base-pairing geometries between the different KP1212 tautomers and either G or A that may explain the observed lethal mutagenesis. Scheme (a) resembles the canonical GC base-pair, and scheme (d) is the previously proposed explanation for KP1212 mutagenicity. Scheme (c) is for the protonated KP1212 whereas the rest are for neutral KP1212 tautomers. These schemes are examples illustrating that KP1212 tautomers can pair with either G or A, but other base-pair geometries are also possible (see Figure 7.20).

Furthermore, due to its pKa of 7.0, a substantial portion of KP1212 is positively charged in the physiological pH range, and such protonated keto-amino form is expected to pair with G in a wobble configuration (Figure 7.19c). The significantly elevated pKa of KP1212 results in a broader distribution of structural forms, which should be an influential factor in its high mutation rate as slight changes of environmental pH will alter the ionization state and tautomeric distribution, and subsequently the base-pairing promiscuity.

We note that other base-pairing geometries are also possible for these tautomers involving cis/trans isomerization of the imine bond, anti/syn rotamerism of the base, or rotation of the enol hydroxyl group (Figure 7.20). Overall, the canonical keto-amino forms are proposed to pair exclusively with G, whereas the KI, EA and EI tautomers can potentially pair with either A or G. Therefore, the presence of multiple tautomers opens up avenues for KP1212 to select different base-pairing partners during replication.
Figure 7.20: Possible base-pairing geometries between the various KP1212 tautomers with either G or A. These base-pairs can adopt Watson-Crick, wobble geometries, or involve the syn-conformer of KP1212. The top panel shows the protonated keto-amino KP1212 and the rest show the tautomers at neutral state. The glycosidic conformation is the more common anti-form unless otherwise specified. For the anti-enol tautomers (bottom two panels), the hydroxyl proton can point towards or away from the opposing base, leading to different base-pairs. However, there may be steric hindrance with the hydroxyl proton pointing towards the deoxyribose. The protonated KP1212 and the two keto-amino (KA) tautomers can only base-pair with G, whereas the keto-imino (KI), enol-amino (EA), and enol-imino (EI) tautomers can base-pair with either A or G.
Even though the schemes proposed in Figure 7.19 do not pair enol tautomers with A in a canonical WC geometry, the wobble base-pair exhibits comparable thermodynamic stability and is a common structural motif in RNA. Crystal structures of a wobble G*T mismatch within a polymerase have been reported, which show stabilizing interactions from the polymerase and the solvating water molecules. Base-stacking of KP1212 is not optimal intrinsically because of the lack of aromaticity and planarity. The energy cost due to unfavorable base stacking and perturbed overall structure likely will result in partial discrimination against KP1212 during DNA synthesis, which is consistent with the lower incorporation efficiency seen in the in vitro experiments.

Our proposed model based on the base-pairing of KP1212 tautomers provides a direct explanation of the recent measurement of KP1212 mutagenicity both in vitro and in vivo. In clinical studies where KP1212 was introduced into the dNTP pool, the observed mutations could be argued by indirect effects such as the reduction in the intracellular dCTP concentrations. In contrast, we incorporated a single KP1212 base into a plasmid and sequenced the resulting DNA after several rounds of DNA replication in E. coli at pH 7.0. A 10% G-to-A mutation occurred exclusively at the site opposite of KP1212, demonstrating that the mutation in living cells is an intrinsic property of KP1212. Taken together, these findings suggest that the high fraction of non-canonical tautomers in solution correlates with the ability of DNA polymerase to incorporate A across from KP1212, and provides a compelling structural evidence for the rare tautomer mechanism of lethal mutagenesis.

7.3.2 Connection between the tautomer populations and mutagenicity

Our experiments revealed that neutral KP1212 is present as ~66% EI plus trace amounts of KI and EA, which are presumed base-pairing partners for A. Although this population differs from the observed 10% mutation rate, there are several additional factors beyond the solution thermodynamics that will determine the final incorporation rate of A. It stands to reason that different KP1212 tautomers will have different base-pairing ability depending on the type and registry of H-bonds involved, and other short range interactions such as base-stacking. KI, EI and EA are all capable of base-pairing with either G or A as shown in Figure 7.20. Additionally, a significant fraction of KP1212 is protonated at physiological pH, and the protonated KA cannot pair with A. Furthermore, the environment at the base incorporation site in a DNA polymerase is drastically different than in bulk water, with the presence of several amino acids and metal ions, as well as a certain degree of deosolvatoin. These factors can markedly influence the pKa, the tautomeric distribution, and
the exchange rate of KP1212. FTIR spectrum of KP1212 in the aprotic solvent dimethyl sulfoxide (Figure 7.21) shows increased keto population. Clearly further studies are needed to establish the tautomeric equilibrium of KP1212 in the presence of a polymerase, and the mechanism of base incorporation in order to better estimate the mutation frequencies.

![Normalized FTIR spectra of KP1212 in D$_2$O/phosphate buffer at pH* 7.9 (red) and DMSO (blue). The increased peak intensity above 1650 cm$^{-1}$ indicates the predominance of the keto tautomers in DMSO.](image)

Figure 7.21: Normalized FTIR spectra of KP1212 in D$_2$O/phosphate buffer at pH* 7.9 (red) and DMSO (blue). The increased peak intensity above 1650 cm$^{-1}$ indicates the predominance of the keto tautomers in DMSO.

### 7.3.3 Tautomerization kinetics

From the T-jump experiments, we find that the tautomer exchange for KP1212 in free solution is on the ns timescale. But how does this timescale relate to the dynamics of DNA polymerization by HIV reverse transcriptase (RT)? Initially, the incoming nucleotide diffuses from the nucleotide pool to the catalytic site and makes H-bonds with the opposing template nucleobases.$^{34}$ Using a diffusion constant of $5 \times 10^3$ μM$^{-1}$s$^{-1}$ $^{35}$ and an dNTP concentration of 20 μM in human cells, $^{36}$ the approximate diffusion timescale is tens of microseconds. Assuming that the free solution measurements are representative, this is sufficient time for KP1212 to sample several tautomeric forms before forming a template-primer complex with the right tautomer. Without forming covalent bonds, the nucleotide is transiently stabilized by base-stacking and coordination of the triphosphate moiety with neighboring metal ions and side-chains.$^{34}$ Tautomerization at this stage would disrupt the H-bond interaction and cause nucleotide dissociation. Finally, conformational changes to the RT is induced and the phosphodiester bond is formed. If tautomerization happens after the nucleotide is incorporated, the base-pairing will be destroyed and the unpaired 3'-OH of the primer will block further elongation. Our measurement of KP1212 toxicity in vivo $^{31}$ did not show
significant replication inhibition, which seems to suggest that a particular KP1212 tautomer can be “frozen” in the context of double stranded DNA.

It is interesting to note that all of the clearly evident tautomers in our data belong to a set which can interconvert via proton transfer involving a single bridging water molecule (Figure 7.22a). Our studies of pyridone derivatives indicate that tautomerization is mediated by proton transfer along hydrogen bonding water wires. In the present case, simple proton rearrangement steps can be expected for the four species involved in the equilibria $\text{KA-N}^3 \rightleftharpoons \text{cKI} \rightleftharpoons \text{cEI} \rightleftharpoons \text{EA}$. Similarly, rapid proton transfer is possible for $\text{KA-N}^5 \rightleftharpoons \text{tKI} \rightleftharpoons \text{tEI}$ (Figure 7.22b). However, proton transfer between the two sub-groups requires a more extended water wire network, which would be less probable. This observation and the unknown timescale for cis-trans imino isomerization, serve to demonstrate that, although our experiments clearly indicate that the 20 ns tautomer exchange kinetics involve population changes between keto and enol, the mechanism and transition states for this process are not readily deduced.

Figure 7.22: Two subgroups in which the tautomers can interconvert by one bridging water molecule. Amino-imino and keto-enol tautomerization is assisted by the water molecule in red and blue, respectively. The interconversion between $\text{KA-N}^5$ and $\text{tEI}$ requires two water molecules.
7.4 Outlook

Our data provide clear evidence for multiple tautomers of KP1212 in an aqueous environment, which offer strong support of the tautomer hypothesis for lethal mutagenesis. Nonetheless, this mechanism should still be regarded as a proposal. Definitive evidence supporting this hypothesis would be provided by observing the base-pair between A and the KP1212 enol or KI tautomer, which could be only stably accommodated at the polymerase active site. Even if it is correct, much remains to be investigated in order to fully understand the complete process, such as the effect of DNA backbone and polymerases.

Despite these uncertainties, the observed A-to-G and G-to-A mutation patterns from in vitro, in vivo, and clinical studies fit logically with the tautomer model proposed here. Our experiments confirm the presence of rapidly interconverting tautomeric structures of KP1212, as well as measure their exchange rates. The data presented are expected to form the basis of future experiments with DNA oligomers containing KP1212 in order to gain more insights into the mechanism of lesion incorporation. The methods established here are also generally applicable to studying other antiviral drugs utilizing lethal mutagenesis.

7.5 Acknowledgments

I thank Vipender Singh, Deyu Li Bogdan Fedeles and Professor John Essigmann for their initiation of this collaboration on KP1212. The constant discussion and the perspectives from different angles to the project have been tremendously helpful to further our understanding of the molecular mechanism of lethal mutagenesis by KP1212. I thank Vipender Singh and Tiffany Amariuta for their help in the initial pH dependence experiment. The NMR experiments were performed by Deyu Li and Katherine Silvestre. I thank Ann Fitzpatrick for her careful reading and comments on this chapter.
7.6 References


Appendix 7.A: SVD analysis on simulated FTIR

In order to extract thermodynamic parameters from the temperature-dependent FTIR spectra, we performed singular value decomposition (SVD) analysis. Here we show some simulations to demonstrate the power of SVD analysis in a perfect world where the temperature-dependent FTIR only report on the population changes of the various tautomers, i.e., without thermally induced frequency shifts or peak broadening. Even though the model calculations do not fully reflect the reality, they still serve as a good practice to illustrate the basic principles of this method. In all the simulations, we tried to mimic the experimentally measured KP1212 spectra.

7.A.1 Three-state model

We start with a three-state model that includes KA-N, KI, and EI tautomers, and we made up the tautomer spectra based on comparison between the experimental and DFT calculated spectra. The basis spectra and the tautomer populations are shown in Figure 7.A.1, based on the thermodynamic parameters (also made up) listed in Table 7.A.7.A.1.

Figure 7.A.1: (a) temperature-dependent FTIR spectra (top) and difference spectra (bottom). (b) Simulated FTIR spectra for the EI, KA-N, and KI tautomers (from top to bottom). (c) The tautomer populations as a function of temperature.
Table 7.A.1: Thermodynamic parameters used in the three-state simulation.

<table>
<thead>
<tr>
<th></th>
<th>EI</th>
<th>KA-N^5</th>
<th>KI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH [kcal-mol⁻¹]</td>
<td>0</td>
<td>3.3</td>
<td>5</td>
</tr>
<tr>
<td>ΔS [cal-K⁻¹-mol⁻¹]</td>
<td>0</td>
<td>2.8</td>
<td>9</td>
</tr>
</tbody>
</table>

Using SVD analysis as described in 7.2.2.1, we can apply a matrix transformation to retrieve the basis spectra from the first SVD spectral components. The reconstruction follows two constraints: (1) nonnegativity of these spectra; (2) certain frequency range of the individual spectrum has to be zero, which is based on the inspection of the experimental FTIR and 2D IR spectra. This reconstruction procedure is subjective in a sense—the peak assignment to a particular tautomer will drastically influence the reconstruction result. Special care should be taken to ensure the validity of the results.

The mathematical basis of this method follows: we first obtain the objective SVD components $U$, $S$ and $V$ as done previously. The components with small singular values are considered as noise and discarded. The reduced-rank SVD components are now represented by $U$, $S$, and $V$ matrices. We now introduce a transformation matrix $W$ which satisfies $WW^{-1} = I$, and the reconstructed spectra $B$ can be calculated through

$$A = USV^T = US \left( WW^{-1} \right) V^T$$

$$= \left( USW \right) \left( W^{-1} V^T \right)$$

$$= BY^T \quad (7.A.1)$$

where $Y$ contains the populations of the tautomers as a function of temperature. The population curves will accurately report the thermodynamic variables of the system if we can perfectly reconstruct the basis spectra for the tautomers (the $B$ matrix). In practice, it is easier to break down the $W$ matrix into subsequent matrices $W_a$, $W_b$, $W_c$, ... etc and reconstruct one tautomer spectrum at a time:

$$A = USV^T = US \left( WW^{-1} \right) V^T = US \left( W_a W_b W_c W^{-1} W_a^{-1} W_c^{-1} W_b^{-1} W_{a,b} \right) V^T$$

$$= \left( USW_a \right) W_b W_c W_{a,b}^{-1} W_{a,b}^{-1} V^T$$

$$= \left( U_a W_b W_c \right) W_{a,b}^{-1} W_{a,b}^{-1} V^T$$

$$= \left( U_{a,b} W_c \right) W_{a,b}^{-1} V_{a,b}^T$$

$$= BY^T \quad (7.A.2)$$
where $U_a$ indicates that the spectrum of tautomer $a$ has been reconstructed using $W_a$ matrix; similarly, $U_{a,b}$ indicates that the spectra of both tautomers $a$ and $b$ have been recovered. Each column of $W_j$ contains coefficients that will take the linear combination of the three $U$ spectra to give the best spectrum for tautomer $j$ based on the input constraints. The transformation matrix $W$ takes the form of (using $W_a$ as an example)

$$W_a = \begin{pmatrix} 1 & 0 & 0 \\ \beta & 1 & 0 \\ \gamma & 0 & 1 \end{pmatrix} \quad (7.4.3)$$

Therefore, the reconstructed spectrum for tautomer $a$ is expressed as

$$u_a = u_1 + \beta u_2 + \gamma u_3 \quad (7.4.4)$$

The SVD reconstruction procedure is illustrated in Figure 7.A.2. Inspection of the first SVD spectral component (Figure 7.A.2 top), we can notice that it resembles our assumed $E_1$ spectrum which has three low frequency modes at 1529, 1572, and 1612 cm$^{-1}$. In order to transform this SVD component into the $E_1$ spectrum, we set the constraint that the spectral region above 1650 cm$^{-1}$ should be zero by adding contributions from the second and third SVD components. The resulting $U_a$ and $V_a$ are plotted in the middle panel. Similarly, we can then apply $W_b$ and $W_c$ transformation matrices to reconstruct the remaining two spectra for tautomers $K_1$ and $K_2$. However, because the amplitudes in SVD analysis are arbitrary, we have to use a final constraint that the sum of the three tautomer populations has to be conserved in order to obtain meaningful population curves. By doing so, we can fully recover the input spectra and populations (bottom panel in Figure 7.A.2).
Figure 7.A.2: The steps involved in the SVD reconstruction procedure. (Top) The first four SVD components. (Middle) The intermediate $U$ and $V$ matrices during the reconstruction. (Bottom) The reconstructed tautomer spectra and associated populations, which match the input parameters.
7.A.2 Four-state model

Since the simulated temperature-dependent FTIR spectra for the three-state model (Figure 7.A.1) still fail to capture some features observed in the experimental spectra, such as the broad gain feature at 1538 cm\(^{-1}\) and from 1620 cm\(^{-1}\) to 1660 cm\(^{-1}\). We hypothesize that these are the contributions from the EA tautomer. We can easily extend our previous analysis to a four-species model. Figure 7.A.3b, c plot the input spectra and populations (open circle) for simulating the temperature-dependent spectra shown in Figure 7.A.3a. We can see now the simulated spectra resemble the experimental spectra even better. And most importantly, using the transformation matrices described in the previous section, we can almost fully retrieve the input parameters (shown as red lines in Figure 7.A.3b and c). The reconstructed spectra match perfectly with the input spectra while the populations are off by a couple of percent.

![Figure 7.A.3: (a) Simulated temperature-dependent FTIR for the four-species system. (b,c) SVD analysis. The open circles show the input spectra (b) and populations (c) for the four tautomers: KA-N\(^5\), K1, EA, and EI (from top to bottom). The red solid lines show the reconstructed spectra and populations.](image)

Table 7.A.2: Thermodynamic parameters used in the four-state simulation.

<table>
<thead>
<tr>
<th></th>
<th>KA-N(^5)</th>
<th>K1</th>
<th>EA</th>
<th>EI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta H) [kcal·mol(^{-1})]</td>
<td>3.3</td>
<td>5.0</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>(\Delta S) [cal·K(^{-1})·mol(^{-1})]</td>
<td>2.8</td>
<td>7.5</td>
<td>1.0</td>
<td>0</td>
</tr>
</tbody>
</table>
However, when we applied this method to analyze the experimental spectra, we were not able to reconstruct four spectra that are physically relevant (for instance, some spectra always have negative peaks). Additionally, the inclusion of the third and fourth SVD components introduces a lot of noise into the reconstructed spectra, which is not desirable. This observation illustrates the importance of considering noise and thermally-induced frequency shifts that are measured by our experiments.

So let’s consider another case: assuming that we did not know a priori that the data set was the result of a four-state system, we decided that there are only two significant components since the third singular value is less than 1%. We then applied transformation matrices Wa and Wb as it was a two-state system to reconstruct the spectra for the two tautomers. The results are shown in Figure 7.A.4. We can see that the first reconstructed spectrum resembles that of the EI tautomer; however, it also contains some C=O stretch vibrations, indicating that the spectrum is not completely separated from those of the keto tautomers. The second reconstructed spectrum looks like a spectrum made from a linear combination of the KA-N₃, KI, and EA spectra, weighted by populations. Additionally, the calculated populations (solid lines) are ~ 15% off from the input populations (open circles).

Figure 7.A.4: Reconstructed spectra and populations for the four-species system using a two-state model. The blue circles represent the input population for the EI tautomer, whereas the red circles show the rest of the population.
CHAPTER 8
Vibrational Signatures of Base-pairing and KP1212 Oligonucleotides

8.1 Introduction

Even though we have shown the presence of multiple tautomers for KP12212 which provides strong support for the rare tautomer hypothesis for lethal mutagenesis, direct evidence of base-pairing between KP1212 and adenine has not been found. In this chapter, preliminary attempts to study base-pairing of KP1212 are presented. Signatures of canonical Watson-Crick GC and AT base-pairing in the context of DNA oligomer are described first to lay the foundation for interpreting KP1212 oligomer data.

8.2 Vibrational signatures of Watson-Crick base-pairs

8.2.1 AT base-pairs

Figure 8.1 shows the FTIR spectra of AMP, TMP, and d(AT)$_{11}$. Although in principle we should be comparing to the spectra of dA$_{11}$ and dT$_{11}$, we have not measured their experimental 2D IR spectra. Slight frequency shifts are expected between single-stranded DNA and NMPs, but the general spectral features should be the same. Significant spectral changes can be observed when AT base-pairs are formed. In the free TMP spectrum,
the T1 (1690 cm\(^{-1}\), mostly C2=O) and T2 (1663 cm\(^{-1}\), mostly C4=O) modes have quite distinct oscillator strengths; however, their intensities become roughly the same (without much frequency shift) upon base-pairing. The T3 thymine ring mode blue shifts from 1629 cm\(^{-1}\) to 1641 cm\(^{-1}\) while the A1 adenine ring mode red shifts from 1625 cm\(^{-1}\) to 1621 cm\(^{-1}\). The four sharp peaks in the d(AT)\(_{11}\) spectrum all have similar intensities.

Figure 8.1: FTIR and 2D IR ZZYY spectra of AMP, TMP, and d(AT)\(_{11}\). The DNA concentrations were 50 mM of AMP and TMP, 2 mM of d(AT)\(_{11}\). The buffer used for d(AT)\(_{11}\) was 50 mM phosphate buffer with 240 mM NaCl and 18 mM MgCl\(_2\).

Vibrational couplings between A and T can be clearly seen in the 2D IR spectra as pronounced cross-peaks exist between the A and T modes. In the individual AMP and TMP spectra, cross-peaks exist due to intramolecular vibrational couplings between the carbonyl stretch and ring breathing modes. In the d(AT)\(_{11}\) spectrum, new cross-peaks appear between the A1 mode and the three thymine modes marked by the pink boxes. In addition, the lineshapes also change dramatically: for example the thymine cross-peak at (1629 cm\(^{-1}\), 1690 cm\(^{-1}\)) is broad and diffuse, but it becomes narrow and sharp in the duplex spectrum. Lee and Cho\(^1\) showed that the vibrational couplings are indeed sizable, e.g., \(J(A1T2) = -10.7\) cm\(^{-1}\).
However, they observed that the two C=O stretches are mostly localized on the carbonyls even when base-paired. It was also found that microsolvation was necessary in the DFT calculations to get the intensities roughly right. With the addition of five D$_2$O molecules the most significant change, except for the expected frequency shifts, was the enhancement of the thymine ring mode that leads to reasonable agreement with the experimental spectra.

8.2.1.1 Thermal denaturation

These spectral changes can be used to monitor the thermal denaturation of AT base-pairs (Figure 8.2). With increasing temperature, we can clearly see the progressive modification of the spectra. At high temperatures, the spectra appear to be the superposition of the free base spectra, indicating the melting of the double-helix. Using SVD analysis, two significant components were found as shown in Figure 8.2c-f.

![Figure 8.2](image)

Figure 8.2: (a) Temperature-dependent (10-85°C) FTIR spectra of d(AT)$_{11}$, experimental conditions same as in Figure 8.1. (b) Difference spectra by subtracting the first spectrum. (c, d) The 1st and 2nd SVD spectral components. (e, f) The 1st and 2nd SVD thermal components.

We can determine the duplex melting temperature and thermodynamics. The reaction scheme for dissociating the DNA double-helix (D) into dA$_{11}$ (M$_1$) and dT$_{11}$ (M$_2$) is

$$D \rightleftharpoons M_1 + M_2$$  \hspace{1cm} (8.1)
\[ K_d = \frac{[M_1][M_2]}{[D]} = \exp\left(-\frac{\Delta G^o}{RT}\right) \]  
(8.2)

We define the temperature-dependent duplex fraction \( f_D \) as

\[
f_D = \frac{\text{Total number of DNA oligos in dsDNA}}{\text{Total number of DNA oligos}} \]
(8.3)

\[
= \frac{2[D]}{C_{\text{total}}} = \frac{2[D]}{2[D] + [M_1] + [M_2]}
\]

The melting temperature \( T_m \) is defined as the temperature when \( f_D = 0.5 \). For equal total concentration of \( dA_1 \) and \( dT_1 \), as we have used here, this refers to the case where half of the oligos present are in duplex form and half in single stranded form. Therefore, the dissociation constant becomes

\[
K_d = \frac{[M_1][M_2]}{[D]} = \frac{[M]^2}{[M]} = \frac{C_{\text{total}}}{4}
\]  
(8.4)

The melting temperature can be found by assuming that the second SVD component for the FTIR spectra is proportional to duplex fraction, and fitting to the following model:

\[
\Delta G^o = \Delta H^o + \Delta C_p \left[T - T_m - T \ln\left(\frac{T}{T_m}\right)\right] - T \Delta S^o
\]  
(8.5)

Here we see that at \( T_m \), unlike the two-state system for protein folding/unfolding, \( \Delta G^o \) may not be necessarily zero, and it actually depends on the total DNA concentration. At the melting temperature, the following relation is satisfied:

\[
\Delta G^o = \Delta H^o - T_m \Delta S^o = -RT_m \ln\left[K_d(T_m)\right]
\]

\[
\rightarrow \Delta S^o = \frac{\Delta H^o + RT_m \ln\left(C_{\text{total}}/4\right)}{T_m}
\]  
(8.6)

In practice, the melting curve measured from the second SVD component or a single frequency trace often exhibits sloping baselines which prevents a direct fitting to the dimer fraction. This can be corrected by fitting to

\[
S(T) = f_D(T)\left[S_D(T) - S_M(T)\right] + S_M(T)
\]  
(8.7)
where $S_D(T)$ and $S_M(T)$ are sloping baselines for the duplex (low $T$) and monomer (high $T$) states, respectively.

Figure 8.3a shows the fit which results in $T_m = 49$ °C, $\Delta H_{\text{diss}} = 69$ kcal/mol, and $\Delta C_p = 0$ kcal/mol·K. Fitting temperature traces for a single frequency such as 1695, 1662, and 1624 cm$^{-1}$ where significant spectral changes are observed gives $T_m$ values within ±2 °C. Therefore fitting the 2$^{nd}$ SVD component was chosen to represent changes within the whole spectral region. A melting curve expressed as the fraction of folded states can be calculated based on these parameters and is displayed in Figure 8.3b. It is well known that the melting temperature increases with DNA concentration, salt concentration, and GC content.

![Figure 8.3: (a) 2$^{nd}$ SVD thermal component (black circles) fit to two-state van’t Hoff expression (red). Sloping baselines are used (dotted lines). (b) The calculated folded double-helix fraction based on the fit in (a).](image)

Melting signatures can be observed in the 2D spectra as well (Figure 8.4). As the temperature increases, the sharp cross-peaks between the A and T modes disappear. The A ring mode at 1621 cm$^{-1}$ broadens and recovers the lineshape of free A (Figure 8.1). Similarly, the cross-peak between T1 and T3 at (1635 cm$^{-1}$, 1700 cm$^{-1}$) transitions from a narrow sharp peak to a diffuse peak as observed in the free T spectrum. The thermal difference 2D IR spectrum is messy as multiple gain and loss features interfere with one another. However, significant features reflecting the duplex melting are still clear—the gain of A ring mode and the loss of T1 on the diagonal. The diagonal slice of the difference spectrum resembles the thermal difference FTIR spectra (Figure 8.2). The 2$^{nd}$ SVD component (Figure 8.4c) of the 2D spectra yields a melting curve from which $T_m$ and $\Delta H$ can be obtained, which are similar to the values from FTIR spectra even though we have significantly fewer data points.
8.2.1.2 Secondary structure

Liquier and Taillandier discussed the experimental spectral features that can be used as marker bands to distinguish between the A, B and Z double-helical conformations (some are listed in Table 8.1); however, they focused mostly on the phosphate and deoxyribose vibrations, instead of in-plane base vibrations. The advantage of using the in-plane base vibration is the much higher oscillator strength. Figure 8.5 shows the FTIR of d(AT)$_{11}$ in the 1300–1500 cm$^{-1}$ region. The marker bands at 1329, 1345, and 1430 cm$^{-1}$ which correspond to the dT, dA, and deoxyribose peaks indicate a B form double-helix according to Table 8.1. However, it is clear that these peaks are very weak on the mOD scale, and can therefore easily be overwhelmed by experimental noise.

Figure 8.4: (a) Temperature-dependent 2D IR spectra of d(AT)$_{11}$. (b) Difference spectrum ($70$ °C-$10$ °C), and the diagonal slice. (c) 2$^{nd}$ SVD component fit to a two-state model.
Table 8.1. IR marker bands for A, B, and Z forms of double-helix. d: deoxyribose; P: phosphate.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1418 d</td>
<td>1425 d</td>
<td>1408 d</td>
<td></td>
</tr>
<tr>
<td>1375 dGdA</td>
<td>1375 dGdA</td>
<td>1355 dGdA</td>
<td></td>
</tr>
<tr>
<td>1335 dA</td>
<td>1344 dA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1335 dT</td>
<td>1328 dT</td>
<td>1320 dG</td>
<td></td>
</tr>
<tr>
<td>1240 P</td>
<td>1225 P</td>
<td>1215 P</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8.5: FTIR spectrum of d(AT)$_{11}$ in the 1300—1500 cm$^{-1}$ region.

In order to explore the potential of using the in-plane base vibrations to distinguish between different forms of double-helices, Lee and Cho performed DFT calculations with Hessian reconstruction methods to simulate FTIR and 2D IR spectra (shown in Figure 8.6). They showed that the higher frequency C2=O mode at 1700 cm$^{-1}$ does not change frequency based on the DNA form. Nevertheless, the rest of the in-plane base vibrations change dramatically with DNA structure. For example, the thymine and adenine ring modes appear as two distinct peaks in the calculated B form spectrum, while they merge into one broad band in the A and Z spectra. Furthermore, the peak positions and intensities of the cross-peaks in the calculated 2D IR spectra also vary significantly with the double-helical conformations, demonstrating the potential sensitivity of 2D IR spectroscopy to secondary structure. When compared to the experimental d(AT)$_{11}$ spectrum, the calculated B form spectrum bears the greatest resemblance, in particular capturing the four-peak structure. Nonetheless, the relative ring mode and cross-peak intensities do not match quantitatively, illustrating that a more accurate theoretical model is needed. Furthermore, these calculations were plotted as absolute value spectra, therefore reducing the spectral information. More experimental constraints can be offered by the absorptive correlation spectra that can help refine the simulation model.
8.2.2 GC base-pairs

Figure 8.7 shows the FTIR and 2D IR spectra of the self-complimentary d(GC)$_8$ strand. Upon double-helix formation at low temperature, the guanine C6=O stretch blue shifts significantly from 1662 cm$^{-1}$ to 1689 cm$^{-1}$. However, the cytosine C2=O stretch stays at 1651 cm$^{-1}$. There is guanine ring-mode intensity suppression at 1581 cm$^{-1}$, which is evident from its intensity growth at higher temperature when the duplex melts. When the temperature increases, the duplex melts and the G C=O red shifts back to $\sim$1665 cm$^{-1}$ which is the frequency of the free G nucleotide in D$_2$O. The frequency shifts of the carbonyl groups upon duplex formation are due to vibrational coupling between the two C=O stretches. This was first shown in isotope labeling experiments where both the G and C modes shift upon $^{18}$O labeling of the G C6=O, and the absolute value of the coupling constant was calculated to be 15.3 cm$^{-1}$. Almost four decades later, Krummel and Zanni measured the first 2D IR spectrum of GC base-pairs and reported $-7.4$ cm$^{-1}$ for the coupling strength by fitting the absolute value 2D IR spectrum. DFT calculations by Lee and Cho obtained a coupling constant of $-10.5$ cm$^{-1}$. 
Figure 8.7: FTIR and 2D IR ZZYY spectrum of d(GC)$_8$. Sample was at 10 mg/mL in 10 mM potassium phosphate buffer with 100 mM NaCl. 2D spectrum was taken at room temperature.

Figure 8.7c plots the 2D IR spectrum of d(GC)$_8$ at room temperature, where the duplex is favored. Similar to what we observed in the 2D IR spectra of AT base-pair, cross-peaks between all the vibrational modes of G and C are found when GC base-pairs. Cross-peak at (1651 cm$^{-1}$, 1690 cm$^{-1}$) reveals the couplings between the C=O stretches of G and C. The overtone of that cross-peak interferes with the fundamental transition of C C=O and therefore is not observed clearly. Similarly, the positive cross-peak at (1690 cm$^{-1}$, 1650 cm$^{-1}$) overlaps with the overtone of the G C=O peak. Interestingly, the G ring mode at 1575 cm$^{-1}$ shows cross-peaks to both C=O transitions (although weak). The couplings between the ring modes and the C=O stretches have not been studied, but are essential for the development of quantitative modeling of 2D IR spectra for nucleic acids.

The understanding of the ring mode behavior is less clear. Maevsky and Sukhorukov reported the FTIR spectra of cytidine at high concentration in D$_2$O and concluded that the reduced intensity of the 1524 cm$^{-1}$ ring mode is due to base-stacking. While this explanation is highly possible, it is unclear whether cytidine forms planar dimers or higher order aggregates in D$_2$O as shown by studies in CDCl$_3$. The assumption that DNA bases aggregate in solution by vertical base-stacking is mostly based on studies of caffeine which does not form planar dimers due to the lack of hydrogen bond donors. As a consequence, the physical picture behind the ring mode reduction remains elusive.
Future studies are also necessary to elucidate the different effects of hydrogen-bonds and base stacking on the ring modes. For instance, base stacking is believed to blue shift the C=O stretch while H-bonds are known to red shift the C=O stretch, and the combined contribution may be responsible for the little affected cytosine C2=O in GC base-pairs.

8.2.2.1 Secondary structures

In terms of distinguishing between the different double-helical geometries, unlike the AT base-pair case, the calculated FTIR spectra for the A and B form of the d(GC)$_{2n}$ double-helix suggest similar lineshape in the in-plane base vibration region. The calculations show a slight frequency difference between the guanine C=O stretch in the A and B forms, however, this could be due to different solvation environment in the calculation, which has not been tested against experimental spectra. The intensities of the cross-peaks between the guanine ring mode and C=O stretches in the calculated spectra for these two forms appear to be distinct which suggests that 2D IR spectra may be more sensitive to the conformational variation. In contrast, the calculated spectrum for the Z form d(GC)$_{2n}$ double-helix exhibits a distinct spectral feature: the cytosine ring mode at 1622 cm$^{-1}$ becomes more intense than the cytosine C=O stretch, which is also manifested in the 2D IR cross-peak intensities.

The spectral signatures predicted from the DFT calculations have yet to be tested against experimental data, especially for 2D IR spectra. The simulated 2D IR spectra have only been qualitatively compared to the experimental spectra for the GC base-pair. The ring mode vibrations have not been studied experimentally. In order to further develop the infrared spectroscopy of nucleic acids, particularly non-canonical and more complex conformations such as triplexes or guanine-quadruplexes, systematic empirical studies aimed at elucidating the varying effects of hydrogen bonding, base-stacking, and helix conformation would prove invaluable.

8.3 KP1212 oligomers

Although the exact structure-spectrum relationship has not been established, model studies of GC and AT base-pairs demonstrate that 2D IR cross-peaks, as well as frequency shifts and intensity variations, can be used to reveal double-helix formation. In the following sections, preliminary results are presented to show the base-pairing ability of KP1212 with the canonical nucleobases.
8.3.1 K·G base-pair

Our first goal was to detect base-pairing between the KP1212 keto-amino tautomer with G, which happens ~ 90% of the time in *in vitro* and *in vivo* experiments. The X-ray crystal structure of the KG base-pair has been solved (private communication with Li and Singh), although the structural resolution was not high enough (~ 1 Å) to determine which tautomer binds to G, the near WC base-pair geometry suggests the KA-N\(^5\) tautomer.

The intuitive oligomer design to monitor KG base-pairing is to measure poly(dG) mixed with poly(KP1212). However, there are a couple of technical difficulties: (1) G-rich sequences tend to form G-quadruplexes by forming Hoogsteen H-bond network (described in more details in Chapter 9), at the mM DNA concentration required by 2D IR experiments. (2) Since KP1212 is not an aromatic molecule due to the sp\(^3\) carbon at the 6 position, base-stacking is disfavored and hence poly(KP1212) is unstable and has not been synthesized. The lack of optimal base-stacking ability, unfortunately, also tremendously limits our design for IR experiments. In principle, we would like to maximize the signal of KP1212 by increasing its content in an oligomer, yet the high KP1212 concentration destabilizes the oligomer and prevents our ability to see base-pairing.

As a result, the initial approach to detect KG base-pairing using IR spectroscopy was to use oligomers consisting of KP1212 but without C in order to avoid spectral contamination from GC base-pair signals. The oligomer used has a sequence of 5'-TKTKTKTKTKT-3' (abbreviated as d(TK)\(^5\)T), and the complementary strand is 5'-AGAGAGAGAGA-3'. Obviously the carbonyl of T will also interfere with that of KP1212, but certain compromises have to be made, as the inclusion of AT base-pairs stabilize the double-helix.

Figure 8.8 shows the FTIR spectrum of d(TK)\(^5\)T, along with the spectra of the KP1212 and TMP bases. It is clear that the oligo spectrum is the superposition of the two (except for small frequency shifts due to the backbone conformation). The low frequency modes at 1612, 1568, and 1527 cm\(^{-1}\) indicate the presence of the EI tautomer of KP1212 even in the context of a DNA oligomer. However, fitting the oligo spectrum with the unit spectra of KP1212 and TMP resulted in a ratio of TMP/KP1212 = 4.4, which is drastically different from the base composition of the oligo (TMP/KP1212 = 1.2). This result suggests that in the context of the oligomer, either the oscillator strength of individual bases change or that the tautomeric equilibrium of KP1212 shifts towards the keto form. To test the first hypothesis, one would need to compare the spectra of dNTP and poly(dN). It is likely that the ring modes of KP1212 are suppressed upon base-stacking, as seen for the ring modes of
C and G. To test the second possibility, the straightforward approach is to measure the spectrum of poly(KP1212), which is not possible as just described. Therefore, comparing spectra of oligomers containing varying amounts of KP1212 may resolve this issue.

![FTIR spectrum comparison](image)

Figure 8.8: FTIR of KP1212 (black), TMP (blue dotted), and d(TK)_5T.

Next we examined the spectrum of mixing equal molar d(TK)_5T and d(AT)_5T, shown as red open circles in Figure 8.9. It is clear that the spectrum is different from the simple linear combination of the two spectra of d(TK)_5T and d(AG)_5A shown in blue. The most striking difference is the peak intensification at 1696 cm\(^{-1}\) which is the spectral signature of the thymine C=O stretch when base-paired (see Figure 8.2). The appearance of a peak at 1635 cm\(^{-1}\) also suggests base-pairing of thymine, although it is not blue shifted to 1642 cm\(^{-1}\) as observed in d(AT)_11. Base-pairing of KP1212 may have an effect on the frequency shift of this peak. The observation of AT base-pairing strongly suggests that K and G are also paired because it has been found that 7 contiguous base-pairs are needed to maintain the stability of the double-helix (rule of seven). The alternating T and K sequence in our oligomer ensures that AT base pair only when GK base pair.

Further evidence of KP1212 base-pairing comes from the frequency shift of the C=O stretch. The mixture spectrum has a red-shifted C=O at 1659 cm\(^{-1}\). Both G and T C=O blue shift upon base-pairing (Figure 8.2 and Figure 8.3), therefore the 1659 cm\(^{-1}\) peak is likely the red-shifted C=O stretch of KP1212 (from 1662 cm\(^{-1}\)) due to vibrational coupling in base-pairs. More importantly, the lower frequency ring modes of the EI tautomer of KP1212 are significantly reduced in the mixture spectrum. This is suggestive of a shift in the tautomeric equilibrium, i.e. the favorable base-pairing between G and the keto tautomer of KP1212 drives the equilibrium away from the enol species.
Temperature-dependent FTIR spectra of the mixture (Figure 8.10) show the melting of the duplex. As the duplex melts, the T C=O peak intensity decreases, the C=O stretch at 1655 cm\(^{-1}\) blue shifts to 1662 cm\(^{-1}\) and intensifies. The A ring mode intensifies at 1621 cm\(^{-1}\) and the T ring mode weakens at 1634 cm\(^{-1}\). The G ring mode intensifies at 1575 cm\(^{-1}\). All of these are signatures of melting base-paired A, T and G. Moreover, the peak intensification at 1530 and 1565 cm\(^{-1}\) at high temperatures indicate the appearance of KP1212 EI tautomer. This observation shows that as the base-pairs between KP1212 and G are destroyed, single-stranded KP1212 takes the form of the EI tautomer, as in the case of free-bases. The T\(_{m}\) was determined from the 2\(^{nd}\) SVD component to be 28 °C, which is ~20 °C lower than that of d(AT)\(_{11}\). In principle, G base-pairs with the KP1212 keto tautomer with three H-bonds (as if it were C) and the increased “GC” content should increase T\(_{m}\), the significant reduction in T\(_{m}\) reports on the destabilization of incorporating KP1212 (not optimal for base stacking) into the duplex. Without the addition of NaCl and MgCl\(_{2}\), base-pairing was also observed but the signature is weaker and the T\(_{m}\) was below 20°C.

Our hope of directly observing base-pair between G and KP1212 lies in the 2D spectra. Figure 8.11 displays the temperature-dependent 2D IR spectra of an equimolar mixture of d(TK)\(_{5}\)T and d(AG)\(_{5}\)A. At 10°C, many features correspond to the spectral signatures of AT base-pairing. The set of cross-peaks along \(\omega_{1} = 1621\) cm\(^{-1}\) are not as sharp and clear as in the d(AT)\(_{11}\) spectrum, probably due to the half reduction of AT base-pairs in the sample, as well as the lower signal-to-noise ratio. The low temperature spectrum shows complicated behavior in the C=O stretch region from 1650 to 1700 cm\(^{-1}\). The 1698 cm\(^{-1}\)
diagonal peak is from H-bonded T C=O. Another T C=O mode is at 1662 cm\(^{-1}\). Interestingly, there appears to be a shoulder for the 1698 cm\(^{-1}\) peak at 1689 cm\(^{-1}\) (upper-right corner of the pink box), which disappears at high temperature. This peak frequency coincides with that of H-bonded G C=O (Figure 8.7). The lower-left of the pink box marks the presence of another C=O peak at 1653 cm\(^{-1}\), which is red shifted compared to the C=O peaks of any of the free bases in this sample (T, G, and KP1212). Since the C=O peaks of T and G have been identified, the logical assignment for the 1653 cm\(^{-1}\) peak is KP1212, which would suggest that KP1212 is also H-bonded. Evidence of base-pairing between KP1212 and G would come from the cross-peaks around (1653 cm\(^{-1}\), 1689 cm\(^{-1}\)). There may have been traces of cross-peak (certainly not present in the high T spectrum), but not definitive enough to make a strong conclusion; spectra taken at higher signal-to-noise ratio should elucidate this problem. However, keep in mind that the cross-peaks between the G and C C=O peaks are also weak (Figure 8.7).

Overall, the IR results indicate that KP1212 base-pair with G in the keto-amino form as predicted in Chapter 7. The near WC base-pairing geometry observed in the crystal structure suggests the K\text{A-N}^5 form. 2D IR spectra of at higher signal-to-noise can confirm this suggestion by identifying cross-peaks between the ring modes of G and K\text{A-N}^5, which are predicted to be \(\sim 1530\) and 1570 cm\(^{-1}\) (see DFT calculations in Chapter 7).

![Figure 8.10: Temperature-dependent FTIR of an equal mixture of d(TK)_3T and d(AG)_3A, and the SVD analysis. Same buffer as in Figure 8.9.](image)
Figure 8.11: Temperature-dependent 2D IR spectra of equal mixture of d(TK)_5 T and d(AG)_5 A. The pink box marks the place where spectral signatures of G-K are expected (1653 and 1689 cm⁻¹).

8.3.2 K-A base-pairs

8.3.2.1 Combination 1

In order to detect base-pairing between KP1212 and A, several different DNA sequences were studied and listed in Table 8.2. Our first attempt was to use the same d(TK)_5 T strand and put it against d(A)_11. Figure 8.12a shows that, unfortunately, the spectrum of the equimolar mixture is identical to the sum of the two component spectra, indicating the absence of duplex formation. Further evidence comes from the little spectral change during a temperature ramp (Figure 8.12b), and the lack of cross-peaks between the A and KP1212 vibrational modes (Figure 8.12c). The nonexistence of duplex DNA is probably due to the high content of “K•A mutation” within this oligomer. For short DNA duplexes consisting of canonical base-pairs, each mutation can decrease the Tₘ by 5-10°C. KP1212 was proposed to base-pair with A in the wobble position which disrupts the backbone integrity much more than a normal mismatch. The sub-optimal base stacking of KP1212 is also expected to lower the stability of the duplex.
Figure 8.12: (a) FTIR spectra of d(TK)₅T and d(A)₁₁, and their equimolar mixture. The black line is the sum of the two individual component spectra. The mixture has 2 mM of each strand, dissolved in 50 mM phosphate buffer at pH 8, without additional salt. (b) Temperature-dependent FTIR spectra of the mixture. (c) 2D IR ZZYY spectrum of the mixture. The black, magenta, and purple boxes mark the vibrations from KP1212, A, and T, respectively. Except for the temperature dependence, all other spectra were taken at 24 °C.

Table 8.2. Sequences of the DNA oligomers used in the experiments to find K·A base-pair.

<table>
<thead>
<tr>
<th>Combination</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5'-TKTKTKTKTKT-3'</td>
</tr>
<tr>
<td></td>
<td>3'-AAAAAAAAAAA-5'</td>
</tr>
<tr>
<td>2</td>
<td>5'-TKTKTKTKTKT-3'</td>
</tr>
<tr>
<td></td>
<td>3'-AGAGAGAAAAA-5'</td>
</tr>
<tr>
<td>3</td>
<td>5'-TKTKTKTKTKT-3'</td>
</tr>
<tr>
<td></td>
<td>3'-AGAGAGAAAGA-5'</td>
</tr>
<tr>
<td>4</td>
<td>5'-TTTTTTTTTT-3'</td>
</tr>
<tr>
<td></td>
<td>3'-AAAAAAAAAAAAA-5'</td>
</tr>
</tbody>
</table>

8.3.2.2 Combination 2

With an effort to minimize the content of "K·A mutation", we replaced d(A)₁₁ with 3'-AGAGAGAAAAA-5' so there are only two K·A base-pairs ifduplexation with d(TK)₅T occurs (see Table 8.2). Figure 8.13a shows the temperature-dependent FTIR spectra of 4 mM mixture, where the spectra bear great resemblance to the spectra in Figure 8.10a.
The spectral changes in the range of 1600–1700 cm⁻¹ suggests at least the presence of AT base-pairs. However, it is clear that even at the lowest temperature, the AT base-pair is not very strong (the C=O peak at 1695 cm⁻¹ is not as sharp). Analysis on the 2nd SVD component estimated a Tₘ of 16 °C (the error on this could be big due to the lack of lower baseline). Since AT base-pairs are observed in this sample, it is reasonable to assume that KG base-pairs are also formed (a C=O stretch at 1658 cm⁻¹ was seen at low T). The Tₘ obtained here is ~12 °C lower than that of d(TK)₅T + d(AG)₅A, which may seem too big of a difference if Combination #2 also makes a full 11mer duplex. Using a Tₘ calculator provided by Integrated DNA Technology (IDT) under similar solvent conditions, it was found that removing the last four base-pairs of the duplex formed by 5'-TCTCTCTCTCT-3' and its complementary strand reduces Tₘ by 12 °C. Therefore, it is highly possible that our sample has fraying ends as illustrated in Figure 8.13g. The 2D IR spectra for this sample, unfortunately were noisy (due to laser instability) and could not provide more definitive information. No sign of base-pairing to the enol tautomer of KP1212 was observed.

Figure 8.13: Temperature-dependent FTIR of 4 mM mixture of d(TK)₅T and 3'-AGAGAGAAAAA-5', and the SVD analysis. Sample was dissolved in 50mM phosphate buffer, with 240 mM NaCl and 18 mM MgCl₂.

8.3.2.3 Combination 3

In order to overcome the problem of a fraying tail, we substituted the second to last A in the complementary strand with G and hoped that the end of the duplex can be stabilized by 5'-TKT-3' and 3'-AGA-5' (see Table 8.2, Combination # 3). As a result, there is
only one "K•A mutation" in this sequence combination. Figure 8.14 shows the temperature-dependent FTIR spectra for this combination. Even though there are signs of duplex formation, strangely; these signatures seem to be weaker than shown in Figure 8.13. The higher \( T_m \) (28.9°C) measured from the 2\(^{nd}\) SVD component suggests that this combination is more stable than combination # 2, which could be due to better base-pairing at the tail. However, due to the weak signatures of duplex formation, K•A base-pairing was not expected and 2D IR spectra were not measured for this combination.

Figure 8.14: Temperature-dependent FTIR of 2 mM mixture of d(TK)\(_5\)T and 3'-AGAGAGAAAGA-5', and the SVD analysis. Sample was dissolved in 50 mM phosphate buffer, with 240 mM NaCl and 18 mM MgCl\(_2\).

8.3.2.4 Combination 4

Finally, in order to eliminate the instability that may be caused by the unfavorable base-pairing of KP1212, we decided to compromise the signal level and incorporate just a single KP1212 in the 11mer. Using 5'-TTTKTTTTTTT-3' against d(A)\(_{11}\), we hoped that the ten AT base-pairs would be sufficient to hold together the duplex. Figure 8.15 shows the temperature-dependent FTIR spectra of a 2mM mixture of the two oligomers. Signatures of AT base-pairing are clear in the 1600-1700 cm\(^{-1}\) region. The nonzero intensity from 1500 cm\(^{-1}\) to \(~1570\) cm\(^{-1}\) indicates the presence of KP1212 EI tautomer. However, the EI vibrations do not seem to change with temperature, suggesting a lack of base-pairing. The melting temperature of the duplex was measured to be 40 °C, which is \(~9\) °C lower than the
d(AT)$_{11}$ oligomer under similar buffer conditions (Figure 8.3). Again, this could be due to the fraying of the tail or single mismatch (K•A).

Figure 8.15: Temperature-dependent FTIR of 2 mM mixture of 5'-T$_3$K$_T$-3' and d(A)$_{11}$, and the SVD analysis. Sample was dissolved in 50 mM phosphate buffer, with 240 mM NaCl and 18 mM MgCl$_2$.

Furthermore, there is also the possibility of forming four-stranded Holliday junctions as illustrated in Figure 8.16. In this structure, AT base-pairs are still formed but the mismatch between KP1212 and A is avoided. However, studies have also shown that Holliday junctions are stable when each arm has more than 12 base-pairs, which is not the case here. Running gel electrophoresis should be able to test this possibility.

Figure 8.16: Possible formation of Holliday junction by 5'-T$_3$K$_T$-3' and d(A)$_{11}$.
2D IR spectra were measured to provide more information on base-pairing. Figure 8.17a shows the 2D IR spectrum of a 4 mM mixture of 5'-T₃KT₇-3' and d(A)₁₁ in phosphate buffer with Mg²⁺. Because the peak intensity of the KP1212 EI tautomer in the 1500–1600 cm⁻¹ region is weak, noise and scattering greatly hinder our ability to get a clean spectrum. On the diagonal, a weak peak at 1524 cm⁻¹ is from the EI tautomer; and the broader peak at ~1570 cm⁻¹ has contributions from both A and KP1212; the strong peak at 1625 cm⁻¹ is from A. As a result, the cross-peaks at (1575 cm⁻¹, 1625 cm⁻¹) is due to the intramolecular coupling of A modes. We also measured 2D IR spectra of the same mixture but with Mn²⁺ replacing Mg²⁺ (Figure 8.17b) since it has been reported that adding Mn²⁺ can induce base mismatching.¹⁸ This spectrum is slightly cleaner; however, the diagonal feature at 1524 cm⁻¹ is barely visible. Interestingly, a cross-peak at (1625 cm⁻¹, 1524 cm⁻¹) marked by red circle shows up that may suggest the coupling between A and the EI tautomer of KP1212. Nevertheless, a comparison to the 2D IR spectrum of pure d(A)₁₁ (Figure 8.17d) at low contours reveals that the cross-peak is from A itself.

Even though we did not observe spectral features for K•A base-pairing, this cross-peak is expected to be strong if the K•A base-pair does occur. The position of this cross-peak is also less contaminated by other peaks that could be observed in DNA oligomers. Frequency slices along ω₁ at ω₁ = ~ 1625 cm⁻¹ (shown on top of the 2D spectra) may also be helpful to identify the presence of cross-peaks at (1524 cm⁻¹, 1625 cm⁻¹). Unfortunately, the currently available data are too noisy to make conclusive statements.

![Figure 8.17: 2D IR spectra of (a) 4 mM mixture of 5'-T₃KT₇-3' and d(A)₁₁ with 18 mM MgCl₂ (b) 4 mM mixture of 5'-T₃KT₇-3' and d(A)₁₁ with 55 mM MnCl₂ (c) 4 mM mixture of 5'-T₃KT₇-3' (d) 4 mM of d(A)₁₁. Sample was dissolved in 50 mM phosphate buffer, with 240 mM NaCl. Frequency slices (pink line) are plotted on the top.](image-url)
So far, the experimental results have not shown evidence of K•A base-pairing, indicating that if it does happen, it is a rare event. It is also possible that DNA polymerase is required to stabilize this mismatch through interactions which certain crucial amino acids. Perhaps even just the transition state of the K•A base-pair is stabilized, and once the polymerase moves onto the next few base-pairs, KP1212 flips out of the duplex and leaves an abasic site. 2D IR experiments involving the polymerase will be challenging as the amide I vibrations will interfere severely with the DNA base vibrations. Isotope labeling of either the protein or DNA will be necessary to resolve the spectral overlap.

### 8.4 Base-pairs in nonpolar solvents

In Section 8.3 we showed that it is technically difficult to detect possible K•A base-pairing signal when we are restricted to incorporating only one KP1212 in a ~10mer duplex. Accurate difference spectra are required to eliminate any artifact peaks that may be due to laser intensity drift, phasing error, or difference in sample concentration. As a result, it would be tremendously helpful to prepare a single base-pair in which the signal is not contaminated by other bases. This is not straightforward to achieve in aqueous solution as the strong dipole moment of water keeps the nucleobases in the monomer state even at high concentrations; however, it is possible in nonpolar solvents where aromatic heterocycles tend to dimerize (as discussed in Chapter 5 for pyridone derivatives).

Nucleosides are barely soluble in weakly polar solvents. In order to increase their solubility, the hydroxyl groups on the ribose can be replaced with tertbutyldimethylsilyl (TBDMS) groups. 2', 3', 5'-TBDMS protected adenosine, uridine, guanosine, and cytidine were synthesized for our initial experiments to establish the spectral features of base-pairing between these modified bases (see Figure 8.18 for structures). This strategy of using TBDMS protecting groups to probe the ultrafast dynamics of DNA bases and base-pairs has been used extensively by the Elsaesser group to study, for example GC base-pair,\(^{19}\) AT base-pair,\(^{20}\) and adenosine monomer.\(^{21}\)

![Figure 8.18: Structure of (a) 2', 3', 5'-TBDMS protected nucleoside, and (b) tertbutyldimethylsilyl (TBDMS) group.](image)
8.4.1 Modified AU base-pairs

Figure 8.19 shows the FTIR spectra of modified A, U, and an equimolar mixture of the two in CDCl₃. The 3525 cm⁻¹ and 3413 cm⁻¹ peaks for A are assigned to the asymmetric and symmetric NH₂ stretches of the monomer.²¹ Aromatic heterocycles tend to dimerize in nonpolar solvents, for example, signatures of dimer formation can be found at 3482, 3310, 3167, and 3126 cm⁻¹, which are assigned to the H-bonded and red-shifted NH stretches.²¹ The in-plane base vibration region is similar to the spectrum in D₂O with peaks at 1630 and 1586 cm⁻¹. Presumably there is also a peak at 1604 cm⁻¹ that can be assigned to the NH₂ bending.²² The spectrum of modified U in the 3 μm region is similar to that of T, with a NH stretching mode at 3396 cm⁻¹.²⁰ Dimer formation can also occur with higher concentration that will lead to the appearance of downshifted peaks ~3180 cm⁻¹ (not obvious in this spectrum as bubbles were formed during the measurement).²¹ The 6 μm region also resembles the spectrum taken in D₂O, with two C=O stretches at 1716 and 1688 cm⁻¹ (blue-shifted because of the lack of H-bonds to water), and a low intensity peak at 1634 cm⁻¹. When equimolar modified A and U are mixed together at 10 mM, the resulting spectrum is the linear combination of the two component spectra, indicating the lack of base-pairing. However, when the concentration is increased to 100 mM, several new peaks appear that indicate the formation of base-pairs. For instance, 3485, 3324, 3260, 3212, 3186 cm⁻¹ peaks in the 3 μm region; and 1641, 1600 cm⁻¹ in the 6 μm region. These spectral features decrease in intensity as the base-pairs are melted away (Figure 8.20). The results show that IR spectroscopy is sensitive to the formation of base-pairs, however, determining the exact base-pair geometry requires further information such as 2D IR spectra. It has been reported that these modified AT base-pairs adopt WC, reverse WC, Hoogsteen, and reverse Hoogsteen configurations.²⁰
Figure 8.19: FTIR spectra of modified (a, b) A, (c, d) U, and (e-h) equimolar mixture of A and U in CDCl₃. The red curves in (e) and (g) are the spectra of free A and U. Notations used in (b) and (d) are: M: monomer; D: dimer; s: symmetric; a: asymmetric.

Figure 8.20: FTIR and difference spectra of 100 mM mixture of modified A and U.
8.4.2 Modified GC base-pairs

Figure 8.21 presents the FTIR spectra of modified G, C, and an equimolar mixture of GC. For G, peaks at 3521 and 3411 cm\(^{-1}\) are assigned to the asymmetric and symmetric NH\(_2\) stretches.\(^8\) At 10 mM, there is already a significant amount of G dimers as indicated by peaks at 3490, and the broad feature spanning 3100-3250 cm\(^{-1}\). The higher tendency to dimerize is in contrast to A and U. Similarly, for C, the 3534, and 3417 cm\(^{-1}\) peaks are assigned to the asymmetric and symmetric NH\(_2\) stretches,\(^9\) while dimer formation was also observed. It has been reported that trimers or tetramers may also exist for high concentration of C in CDCl\(_3\).\(^8\) The 6 \(\mu\)m region of the spectra of G and C are quite different from their spectra taken in D\(_2\)O due to the appearance of the NH\(_2\) bend (ND\(_2\) is shifted to \(\sim1100\) cm\(^{-1}\)) and dimer formation. When equimolar G and C are mixed at 50 mM, significant spectral changes are observed in the 3 \(\mu\)m region which shows three main peaks that are assigned to the free NH stretches in the amino group (3490 cm\(^{-1}\)), the bound NH stretch of the G amino group (3306 cm\(^{-1}\)) and the bound NH of C and G (3143 cm\(^{-1}\)).\(^9\) Spectral changes in the 6 \(\mu\)m region are less clear. The G C=O blue shifts to 1690 cm\(^{-1}\), and the C C=O narrows at 1650 cm\(^{-1}\). It is expected that 2D IR spectra will provide more definitive evidence of base-pairing in this region.

Figure 8.21: FTIR spectra of modified (a, b) G, (c, d) C, and (e, f) equimolar mixture of G and C in CDCl\(_3\). The dotted curves in (e) are the spectra of free G and C.
8.5 Conclusion

2D IR spectra of oligonucleotides consisting of Watson-Crick GC and AT base-pairs demonstrate that 2D IR spectra are sensitive to the H-bond network involved in the base-pairs. These H-bonds between the carbonyl and NH groups cause the in-plane base vibrations to delocalize across different bases, resulting in pronounce cross-peaks, in addition to peak shifts and intensity variations observed in the linear FTIR spectra. As a consequence, together with its ability to resolve rapidly interconverting tautomers, 2D IR spectroscopy holds great promise to investigate the base-pairing ability of different KP1212 tautomers. The experiments can be performed in the context of DNA oligomers to mimic the physiological condition, or as single base-pair in organic solvents with the aid of TBDMS protecting groups in order to enhance the KP1212 signal. Future experiments involving DNA polymerase may be required if the enzyme is essential for stabilizing the K•A mismatch at the active site. Isotope labeling of either the protein or DNA will be essential to separate the overlapping vibrations. Such experiments will provide direct evidence to prove (or disprove) the rare tautomer hypothesis proposed for the lethal mutagenic property of KP1212.

8.6 Acknowledgments

I thank Vipender Singh, Deyu Li, Bogdan Fedeles, and John Essigmann for invaluable discussions and suggestions regarding the design of KP1212 oligomers experiments. I thank Paul Sanstead and Ann Fitzpatrick for setting up the “Compact 2D IR” spectrometer for measuring the 2D IR spectrum of d(GC)₈. I thank Paul Sanstead for measuring the temperature-dependent FTIR and 2D IR spectra of d(GC)₈, and his careful reading of this chapter. I thank Qing Dai for synthesizing the TBDMS protected nucleobases.
8.7 References


the mutagenic properties of the anti-HIV nucleoside analog 5-aza-5,6-dihydro-2'-deoxycytidine. (in preparation).


Appendix 8.A: DNA oligomer sample preparation

It was found that DNA oligomers synthesized either by Sigma-Alrich, Integrated DNA Technology (IDT), or Keck Laboratory at Yale University all contain impurities whose vibrations interfere with the DNA base vibrations. In order to remove these impurities, the samples were filtered using Amicon Ultra 0.5 mL centrifugal filters (3 kDa). The FTIR spectra of the filtered wastes are shown in Figure 8.A.1, and they all show, regardless of the company or purification method, absorption bands in the 6 μm region which overlaps with DNA base vibrations. According to Sigma, possible leftover small molecules that may contribute to these absorptions are benzyol chloride, isobutyryl chloride, dimethoxytrityl chloride, or monomethoxytrityl chloride (Figure 8.A.2). There might also be some depurination product of A,T, C, G bases.

Figure 8.A.1: FTIR spectra of filtered waste from several commercially synthesized DNA oligomers. Purification was done with either HPLC or standard desalting.

Figure 8.A.2: Structures for the possible impurities.
CHAPTER 9
Vibrational Signatures of G-quadruplex Revealed by 2D IR Spectroscopy

9.1 Introduction to G-quadruplex

Thrombin-binding aptamer (TBA) is a single-stranded DNA 15mer (5'-GGTTGGTGTGGTTGG-3') that binds to the protein thrombin, inhibiting its bioactivity of cleaving fibrinogen to make fibrin. TBA is a promising new anticoagulant because its complementary DNA strand can serve as an antidote, which is lacking for most currently used anticoagulants. The basic structural motif of TBA has been shown by both X-ray crystallography\(^1\)\(^3\) and solution NMR spectroscopy\(^4\)\(^6\) to be an anti-parallel chair-like structure (Figure 9.1a,b) when bound to thrombin. This chair structure is stabilized through guanine-quadruplex (G-quadruplex) that is formed by Hoogsteen base-pairing between four guanine bases (referred to as a G-quartet, Figure 9.1c). A G-quartet is a planar arrangement of four guanine bases associated with eight cyclic Hoogsteen hydrogen-bonds where each guanine donates and accepts two H-bonds. Considering only the guanine base without the backbone, the G-quartet structure has \(C_{4h}\) symmetry (\(D_4\) for two layers of G-quartet) with each guanine base rotated 90° with respect to its neighboring base, and the H-bonds are formed between \(C^6=O\) and \(N^7=C^8\) of one base to \(N^1-H\) and \(N^2-H\) of the next base. G-quadruplexes are further stabilized by a presence of metal ion in the center which coordinates with the carbonyl groups. The two G-quartets are linked by two TT loops at one end that mimics the legs of a chair, and another GTG loop at another end that mimics the
chair back. Other variants of TBA has been found, for example, Tasset et al. a reported 29-nucleotide single-stranded DNA aptamer, which possesses similar G-quadruplex core structure as the TBA sequence found by Bock et al., but has a flanking duplex tail formed by base-pairing between the 5′ and 3′ ends.

In order to understand the folding of TBA, we first examine the infrared spectra of various G-quadruplexes. G-rich oligonucleotides and high concentration of (deoxy)guanosine monophosphate (dGMP or GMP) can form a variety of multiple-stranded G-quadruplexes. For G-rich oligonucleotides, G-quadruplexes can be formed either by folding of a single-stranded oligomer into an intramolecular quadruplex, or assembly of four single-stranded oligomers into an intermolecular quadruplex. In the case of dGMP or GMP, a helical stack of tetramer units leads to the formation of G-quadruplex.

Structural studies of G-quadruplex have attracted great interest because they participate in various important cellular processes. For example, G-quadruplexes and homopurine repeats are commonly present in telomeres and non-coding regions of the genome. G-quadruplexes are believed to act as promoters for transcription, and they are known to be present in regulatory regions for several important oncogenes, making them potential therapeutic targets. G-quartet formation has also found applications in supramolecular chemistry and nanotechnology. They have also been used as gelling agents, biosensors, nanowires, nanotubes, ion channels, and artificial synapses.

![Figure 9.1](image_path)

**Figure 9.1:** (a) X-ray crystal structure of TBA bound to thrombin (PDB code: 1HAO). The blue and magenta bases in TBA are guanine and thymine, respectively. (b) Chair form of thrombin-binding aptamer. The yellow rectangles represent the base moieties of guanine. (c) Structure of the G-quartet that is stabilized by Hoogsteen H-bond network and a potassium ion placed in the center.
The formation of guanine self-association was first identified in 1962 when Gellert et al. reported that both 5'-GMP and 3'-GMP form stacks of H-bonded tetramers. Since then, the basis of G-quartets have been studied with a variety of techniques such as dynamic light scattering, X-ray crystallography, NMR spectroscopy, circular dichroism, IR spectroscopy, and Raman spectroscopy. Each method is only sensitive to certain structural features, in particular, vibrational spectroscopy has proven to be sensitive to H-bond interactions, and therefore suitable for studying the self-assembly of G-quartets. Despite this, the molecular origins of the infrared spectral features of molecular structures formed by G-rich oligomers have not been well-established. Many empirical relations have been reported, but a robust theoretical model is still lacking, hence hindering the efforts to investigate more detailed structural dynamics.

In this chapter, we employ multiple techniques to characterize G-quadruplexes with an emphasis on establishing the physical origins leading to their IR spectral features. We first carry out the basic characterization using circular dichroism and PAGE gels on multiple G-quadruplex models. We then present their FTIR and 2D IR spectra and discuss the important features associated with the formation of G-quadruplexes, with the aid of DFT calculations and phenomenological modeling.

9.2 Basic characterization of G-quadruplex

9.2.1 Parallel and anti-parallel G-quadruplexes

G-quadruplexes can adopt various different conformations depending on the number of oligonucleotide strands, and orientations. For example, the X-ray crystal structure of d(TG₄T) shows the association of four strands in parallel orientation, and the guanine bases in the same plane forms G-quartet, therefore creating the so called “parallel G-quadruplex” (Figure 9.2a). Figure 9.2b illustrates another type of parallel G-quadruplex adopted by d(G₄TG₄). In this case, the guanine bases from the 5'-end of one strand make H-bonds with the guanine bases from the 3'-end of another strand. With this structural motif, a continuous supramolecule is formed and is called the “G-wire”.

Besides parallel G-quadruplex, anti-parallel G-quadruplex can be formed as well, as shown in Figure 9.2c and d. These two DNA strands represent the short repeats found in the telomeric DNA of Oxytricha. The telomeres of most eukaryotes contain a repeating G-rich sequence that can form G-quadruplexes, which have been observed in vivo and

259
found to possess important biological functions.\textsuperscript{42} Here we show the solution NMR structures of d\((G_4T_4G_4)\) (Oxyl.5, Figure 9.2c) and d\((G_4T_4G_4T_4G_4)\) (Oxy3.5, Figure 9.2d). Both structures contain anti-parallel G-rich strands. However, there is still debate over their exact strand polarities.\textsuperscript{6,31,43,44} For example, the X-ray crystal structure of Oxyl.5 is a dimer of two hairpins with the loops along the edges,\textsuperscript{43} whereas the solution NMR structure has loops across the diagonal.\textsuperscript{31} In the NMR structure, each guanine base H-bonds with another guanine base from another DNA strand. Similarly, different structures have been reported for Oxy3.5. In the first model, the adjacent strands are anti-parallel and the loops are along the edges of the G-quartet.\textsuperscript{6} In a second model, a diagonal loop was observed (as illustrated in Figure 9.2d).\textsuperscript{44}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure9.2.png}
\caption{Structures and circular dichroism (CD) spectra of four different G-quadruplexes at 25 °C and 95 °C. (a) d\((TG_4T)\). The cartoon structure shows a dimer of the parallel G-quadruplex.\textsuperscript{45} (b) d\((G_4TG_4)\), which forms G-wires with parallel G-rich strands.\textsuperscript{8} (c) d\((G_4T_4G_4)\), also known as Oxyl.5, which forms an anti-parallel G-quadruplex with a dimer of hairpins.\textsuperscript{44} (d) d\([(G_4T_4)G_4]\), also known as Oxy3.5, which forms an anti-parallel G-quadruplex with a diagonal loop.\textsuperscript{44} Samples were dissolved in 150 mM potassium phosphate buffer, except for Oxyl.5 which was dissolved in 30 mM NaCl in D\textsubscript{2}O. All sample concentrations were kept at 5 mg/mL. The cartoon structures were taken from the respective refs with permission.}
\end{figure}
To distinguish between the parallel and anti-parallel G-quadruplexes, simple empirical relationship has been developed using circular dichroism (CD) spectroscopy. For parallel G-quadruplex, a positive signal is observed at ~260 nm and a negative signal is observed at ~240 nm; whereas for anti-parallel G-quadruplex, a positive signal is seen at ~290 nm and a negative signal is seen at 260 nm. The CD spectra of the four aforementioned G-quadruplexes are shown in Figure 9.2 and are consistent with the characteristic features for parallel and anti-parallel G-quadruplexes. As the temperature increases, these signals decrease. Depending on the thermal stability, some residual structures can be seen even at 95 °C. Recent theoretical work by Masiero et al has shown that these spectral features can be reproduced using a simple exciton coupling model. They concluded that the differences in the CD signal is due to the different base-stacking orientations. Figure 9.3 illustrates that the guanine bases stack in the “head-to-tail” (six member ring on top of the five member ring) and “head-to-head” (five member ring on top of the five member ring) configurations in the parallel and anti-parallel G-quadruplexes, respectively.

![Figure 9.3: Structures of the stacked guanine bases in parallel and anti-parallel G-quadruplexes. In parallel configuration, the bases are stacked in the head-to-tail orientation, and the glycosidic bond configuration is always anti. For anti-parallel G-quadruplex, the bases are stacked in the head-to-head orientation, and the glycosidic configuration alternates between anti and syn.](image)

### 9.2.2 Anti and syn glycosidic configurations

We have presented the CD spectral signatures which distinguishes between the parallel and anti-parallel G-quadruplexes, and now we examine the IR spectra of G-quadruplexes which offer additional structural information. In addition to the base-
stacking orientations, the two types of G-quadruplexes also differ in the glycosidic-bond configurations. Figure 9.3 shows that in the parallel form, all the guanine bases are in the anti configuration, whereas in the anti-parallel form, the bases take on alternating anti and syn configurations. The 1250-1500 cm⁻¹ IR spectral region is sensitive to the base-sugar vibrations.⁴⁹ A comparison between the FTIR spectra of d(TG₄T) and Oxy1.5 in the spectral region of 1300-1400 cm⁻¹ (Figure 9.4) shows that in addition to the 1352 cm⁻¹ feature which exists in both spectra, the spectrum of Oxy1.5 has another peak at 1326 cm⁻¹. This observation suggests that the 1352 cm⁻¹ and 1326 cm⁻¹ peaks correspond to the anti and syn configurations, respectively. This assignment is consistent with the observed red shift (~1375 cm⁻¹ - ~1355 cm⁻¹) going from anti to syn configurations measured in H₂O.⁶⁹ Similar spectral features discerning the different glycosidic conformations have also been reported in Raman spectroscopy studies.³⁹,⁴¹

![Figure 9.4: FTIR spectra of d(TG₄T) and Oxy1.5 in the 1300-1400 cm⁻¹ region.](image)

**9.2.3 Basic characterization of thrombin-binding aptamer**

With the basic characterization of G-quadruplexes in hand, we can now apply these methods to study thrombin-binding aptamer (TBA). The structure of TBA has been solved using X-ray crystallography¹ and NMR spectroscopy⁴, and was found to be an anti-parallel G-quadruplex that resembles a chair form (Figure 9.1b). Other techniques have been used to characterize the stability of TBA G-quadruplex formation, such as circular dichroism (CD) spectroscopy, temperature-dependent UV absorption spectroscopy, differential scanning calorimetry (DSC), and fluorescence.⁶,³³,⁴⁸,⁵⁰,⁵⁴

The anti-parallel G-quadruplex that TBA adopts can be confirmed by the characteristic CD signal at 292 nm.⁴⁸ Figure 9.5a plots the temperature-dependent CD
spectra of 10 µM TBA in H₂O with 150 mM KCl. These spectra resemble those of Oxy1.5 and Oxy3.5 with a maximum peak at 292 nm. As the temperature increases, the overall signal decrease, indicating the melting of the chair form. The signal at 292 nm as a function of temperature (Figure 9.5c) provides a melting curve whose melting temperature was found to be 53 °C using a fit to a two-state model. The black curve in Figure 9.5a shows the spectrum for 10 µM TBA, but without the addition of potassium ions. The lack of CD signal indicates that 10 µM of TBA does not fold into the native structure without the stabilizing interactions provided by the potassium ions. To test the thermal reversibility of the chair formation, CD spectra were taken during three thermal cycles. As shown in Figure 9.5b, it is clear that the signal representing the chair form is fully reversible.

![Figure 9.5: Circular dichroism (CD) spectra of TBA: (a) 10 µM TBA in H₂O with 150 mM KCl at selected temperatures. 10 µM TBA without the presence of K⁺ (offset by -1.5 mdeg) indicates the lack of chair formation. (b) CD spectra of 10 µM TBA in H₂O with 150 mM KCl taken during three thermal cycles to demonstrate the thermal reversibility of the chair formation. (c) Signal at 292 nm as a function of temperature, and the two-state model fit shown as the red curve.](image)

Typically, CD spectra are collected with micromolar concentrations, however, millimolar concentrations are required to provide the required signal-to-noise ratio for IR experiments. Therefore, in order to check that there are no higher-order aggregates at high TBA concentrations, we measured the CD spectra at 1 and 2 mM TBA concentrations, as shown in Figure 9.6. Figure 9.6a shows the CD spectra of 1 mM TBA with increasing KCl concentration. Surprisingly, even at 0 mM KCl, the spectrum exhibits the characteristic signal of the chair structure, demonstrating that the chair form can be stable at high TBA concentrations even without the presence of stabilizing metal ions. Since the chair formation is unimolecular reaction and should not depend on the TBA concentration, crowding effect may contribute to the formation of G-quadruplex at high concentration. A previous Raman
The study also showed that the TBA chair form can exist at 10 μM TBA, 4 °C, and pure H₂O. As the potassium ion concentration increases, the CD spectra remain constant until [K⁺] reaches 2 M. At 2M [K⁺], there is a significant drop in the 292 nm signal, along with a frequency shift of the 245 nm peak, which can be explained as a peak growing in at 260 nm that would correspond to the formation of parallel G-quadruplex, as shown in Figure 9.2. The melting curves of 1 mM TBA with various [K⁺] are plotted in Figure 9.6b. In the absence of stabilizing K⁺, Tₑ is about 30 °C. The addition of a small amount of KCl (50 mM) is found to increase Tₑ by 20 °C. At 2 M KCl, the melting curve becomes less sharp, suggesting a more complicated unfolding process that may involve intermediate states during the melting of parallel G-quadruplex. Figure 9.6c also shows that parallel G-quadruplexes can be formed with 2 mM TBA and higher concentration of potassium phosphates.

Figure 9.6: CD spectra of TBA: (a) 1 mM TBA in H₂O with various KCl concentrations as indicated in the plot. (c) CD spectra with different TBA and K⁺ concentrations. (b,d) Melting curves corresponding to the spectra shown in (a,c).
CD spectroscopy provides a convenient way of distinguishing between the parallel and anti-parallel G-quadruplexes; however, it does not necessarily inform on the intermolecular or intramolecular nature of the G-quadruplex. Even though all the intermolecular four-stranded G-quadruplex have parallel strand orientations, dimeric G-quadruplex can have anti-parallel strands such as the structure of Oxy1.5 (Figure 9.2c). To ensure that purely monomeric TBA in the chair structure is present in solution, we use PAGE gels to characterize the structure of TBA under different conditions. The gel (Figure 9.7) shows that 10 μM TBA (with or without K⁺) and 1 mM TBA without K⁺ are purely in the monomer state. However, at 1 mM TBA, as the concentration of the KCl or potassium phosphate increases, dimers and trimers begin to appear, indicated by the weak bands at 30 bp and 50 bp, respectively. The formation of intermolecular parallel G-quadruplex is consistent with the proposal by Miura and Thomas Jr. Since the signal from the higher-order aggregates is significantly weaker compared to the monomer band (estimated to be <1%), our IR experiments are performed at TBA concentrations under 1 mM in a 50 mM potassium phosphate buffer.

Figure 9.7: PAGE gel of TBA under various conditions. The DNA ladder is shown on the left for reference. TBA is a 15mer so 30 bp and 50 bp would correspond to dimers and trimers. Experiment was performed with non-denaturing 20% TBE Gel with 0.5X TBE running buffer, at 180 V and ~17mA. The whole gel box was placed in ice to maintain low T for TBA folding. For 10 μM TBA samples, 2 μL of sample was added with 1 μL of loading dye and 2 μL of buffer. For 1 mM TBA samples, 0.4 μL of sample was added with 0.1 μL of loading dye. After sufficient separation, the gel was stained with 10ul SybrGold in 100ml 0.5X TBE buffer for 30min on shaker.
The conclusion from the characterization of TBA using CD spectroscopy and PAGE gel is that the chair form is stable under high salt concentration and at low temperatures. However, excess salt and oligonucleotide concentration will induce the formation of intermolecular parallel G-quadruplexes.

9.3 Equilibrium IR spectroscopy of G-quadruplexes

9.3.1 Infrared absorption spectroscopy

Figure 9.8 shows the FTIR spectra of the four G-quadruplexes illustrated in Figure 9.2. The spectra were taken at 25 °C and in the presence of K⁺ ions. Since these four oligonucleotides are composed of just G and T, the FTIR spectra of GMP and TMP are plotted for comparison. The peak frequencies are summarized in Table 9.1. CD spectra of these four oligonucleotides indicate the formation of G-quadruplexes under these experimental conditions. Significant spectral changes can be observed upon G-quadruplex formation. Firstly, the carbonyl stretch is sharper and has a higher frequency (~1674 cm⁻¹) compared to that of the free GMP in D₂O at 1664 cm⁻¹. The sharpening is most likely due to the formation of the highly symmetric G-quartet (Figure 9.1b). Secondly, upon the formation of G-quartet, the guanine C=O group loses the H-bond interactions with the surrounding solvent, therefore reducing its frequency distribution. Sharpening of the C=O peak has also been observed in other base-pair systems (see Chapter 8). The C=O stretching frequencies of free GMP and G-quadruplex provide us important information on the strength of H-bonds from water and the NH groups in G-quartet. Thirdly, the G C=O stretch also blue shifts (relative to the free GMP) to 1682 cm⁻¹ in GC Watson-Crick base-pairs. For d(G₄TG₄), two C=O peaks were observed at 1675 cm⁻¹ and 1682 cm⁻¹. This is perhaps due to the fact that this oligonucleotide forms G-wires (Figure 9.2c), which are expected to have more structural disorder that breaks the degeneracy of the C=O stretching frequency within the G-quartet. See discussion below for additional details.
Three thymine vibrations are expected in the 1600-1750 cm$^{-1}$ region: two carbonyl stretches at 1690 and 1663 cm$^{-1}$, and a ring mode at 1629 cm$^{-1}$. The frequencies of the two C=O peaks are not greatly influenced by H-bonds in base-pairing (Figure 8.1), and since the thymines in these systems do not participate in the Hoogsteen H-bond network, the T C=O's should not experience significant frequency shifts in G-quadruplexes either. These T C=O peaks appear as shoulders on the two sides of the narrowed G C=O. The T ring mode appears to decrease its intensity in the G-quadruplexes, especially in the parallel G-quadruplexes. The peak at $\sim$1613 cm$^{-1}$ seen in d(TG,T) and d(G,TG) is not expected to correspond to the T ring mode since a 16 cm$^{-1}$ shift is unlikely.
Table 9.1. FTIR absorption bands in the in-plane base vibration region for various G-quadruplexes at 25 °C and D₂O. Na⁺ ion was present in the Oxy1.5 sample and K⁺ ion was present in all other samples.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>G4</th>
<th>G3</th>
<th>G2</th>
<th>---</th>
<th>G1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMP</td>
<td>1538</td>
<td>1567</td>
<td>1579</td>
<td>---</td>
<td>1664</td>
</tr>
<tr>
<td>TMP</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>1629</td>
<td>---</td>
</tr>
<tr>
<td>d(TG₄T)</td>
<td>1535</td>
<td>1566</td>
<td>1586</td>
<td>1613b</td>
<td>1676</td>
</tr>
<tr>
<td>d(G₄TG₄)</td>
<td>1535</td>
<td>---</td>
<td>1583</td>
<td>1615b</td>
<td>1675,1682</td>
</tr>
<tr>
<td>Oxy1.5</td>
<td>1535</td>
<td>1567</td>
<td>1585</td>
<td>1626</td>
<td>1673</td>
</tr>
<tr>
<td>Oxy3.5</td>
<td>1535</td>
<td>1568</td>
<td>1584</td>
<td>1629</td>
<td>1675</td>
</tr>
<tr>
<td>TBA</td>
<td>1535</td>
<td>1567</td>
<td>1584</td>
<td>1632</td>
<td>1668</td>
</tr>
<tr>
<td>(GMP)₄a</td>
<td>1534</td>
<td>1567</td>
<td>1590</td>
<td></td>
<td>1671</td>
</tr>
<tr>
<td>polyG</td>
<td>1534</td>
<td>1569</td>
<td>1589</td>
<td></td>
<td>1675</td>
</tr>
</tbody>
</table>

---

a 200 mg/ml of GMP in the presence of K⁺.

b the bands ~1613 cm⁻¹ in d(TG₄T) and d(G₄TG₄) may be assigned to the guanine ring modes in G-quadruplex. See modeling below section for additional details.

The ring mode suppression is more profound for the G ring modes G2 and G3 at 1579 and 1567 cm⁻¹, respectively. The G ring modes are stronger than the C=O stretch in free GMP; however, their intensities almost vanish in the G-quadruplexes.

Finally, another G ring deformation mode at 1538 cm⁻¹ is red shifted to 1535 cm⁻¹ upon G-quadruplex formation. Such a peak is absent in GC Watson-Crick base-pairing.⁵⁶ In the free GMP case, this G4 mode is a delocalized ring deformation mode across both the pyrimidine and imidazole rings, with strong contribution from the N⁷–C⁸ bond.⁵⁷ The N⁷–C⁸ bond directly participates in the Hoogsteen H-bond network by forming a strong H-bond to the amino group of the next guanine base. Therefore, this band is expected to be sensitive to the formation of G-quadruplexes.

Further experimental support for the above interpretation comes from the temperature dependence of IR spectra. Figure 9.9 shows the FTIR spectra of GMP, TMP, and TBA between 10 °C to 90 °C. As shown by Figure 9.9a and Figure 9.9b, marginal spectral features are observed for free GMP and TMP. The small blue shift of the C=O stretches upon heating can be rationalized as the weakening of H-bonds from the C=O to D₂O solvent molecules. The spectral changes for the G ring modes ~1570 cm⁻¹ with temperature are less well understood. Figure 9.9d plots the variable temperature spectra of TBA in the presence of K⁺, and three distinct spectral features corresponding to the formation of G-quadruplex are identified by the green arrows. At low temperature where the
chair form is stabilized, the G C=O sharpens and narrows, the G ring mode is suppressed, and the 1535 cm⁻¹ peak is enhanced. As the temperature increases and the chair melts, these spectral features disappear. A linear combination of the temperature-dependent spectra of GMP and TMP, weighted by the base composition of TBA, is calculated and presented in Figure 9.9c. It is clear that the low temperature spectra in Figure 9.9c and Figure 9.9d are quite distinct while the high temperature spectra are similar. It is interesting to note that the ring mode suppression for TBA is not as pronounced as in the other G-quadruplexes shown in Figure 9.8. This observation is consistent with the assumption that G-quartet stacking reduces the G ring mode intensity since the four G-quadruplexes shown in Figure 9.8 consist of more layers of G-quartet compared to TBA, and therefore greater intensity reduction.

![Figure 9.9](image)

Figure 9.9: Temperature-dependent FTIR spectra of 20 mg/ml of GMP (a), 20 mg/ml of TMP (b) and 1 mM of TBA in 50 mM PBS (d). (c) Linear combination of the spectra of GMP and TMP, weighted by the base composition of TBA (G:T = 3:2). The temperature goes from 10 °C (blue) to 90 °C (red). The green arrows in (d) mark the significant spectral features associated with the chair formation.
CD spectroscopy shows that 1 mM TBA can adopt the chair structure even in the absence of K\(^+\) at low temperature (Figure 9.6), and this observation is confirmed by the spectral changes seen in the temperature-dependent FTIR plotted in Figure 9.10b. Similar to the TBA spectra taken with K\(^+\) (Figure 9.10a), the 1535 cm\(^{-1}\) mode enhances and the ring modes at ~1570 cm\(^{-1}\) decreases upon chair formation at low temperature. However, the C=O peak is not sharpened in the pure D\(_2\)O spectra. This suggests that the chair structure is less rigid and symmetric without the stabilization from K\(^+\).

G-quadruplexes can also be formed at high concentration of GMP in the presence of K\(^+\).\(^{11-16}\) The FTIR spectra of 200 mg/ml of GMP with K\(^+\) (Figure 9.10c) exhibit the same spectral features as those for the different G-quadruplex models. Temperature-dependent gel formation was observed for 25 mg/ml of GMP at pH 5, and the structure was proposed to be helical stacking of the G-quartet based on its X-ray diffraction pattern.\(^{12}\) Such gel formation was not seen for GMP at neutral pH, though the similarities in IR spectra suggested G-quadruplex formation in both cases.\(^{14}\) However, Miles and Frazier measured the NMR linewidth of the H8 proton and found that it was sharper than what would be expected for a long rigid rod, concluding that the degree of GMP aggregation was small.\(^{14}\)

![Figure 9.10: Temperature-dependent FTIR spectra of 1 mM of TBA in 50 mM PBS (a), 1 mM of TBA in D\(_2\)O (b), and 200 mg/ml of GMP in 50 mM PBS. The high GMP concentration spectra were taken with 6 \(\mu\)m spacer.](image)
9.3.1.1 Intermolecular couplings between the C=O vibrations

In order to understand the physical origins of these spectral features observed in the IR spectra of G-quadruplexes, we will first consider the couplings between the carbonyl groups within a G-quartet. Considering the guanine tetramer with a C₄ᵥ symmetry as shown in Figure 9.1b, the four coupled C=O stretches should give rise to two degenerate IR active modes of Eᵥ symmetry, and two Raman active modes with Aᵥ and Bᵥ symmetry. It has been reported that 30 mM of guanosine in 100 mM of KCl forms a gel and its Raman spectrum shows two C=O peaks at 1659 and 1702 cm⁻¹.⁵⁸ Furthermore, the IR spectrum of 1:1 mixture of ¹⁶O:¹⁸O guanosine shows multiple C=O peaks, demonstrating that the four carbonyls in the tetramer are indeed coupled.⁵⁸

With the frequencies of the Raman C=O peaks at 1702 and 1659 cm⁻¹, and IR C=O mode at 1672 cm⁻¹, we can calculate the couplings between the four carbonyl groups within the G-quartet. We began by testing the transition dipole coupling (TDC) model, a model which expresses the coupling constants βᵢ,ⱼ based solely on transition dipole interactions:

\[
\beta_{i,j} = \frac{1}{4\pi\varepsilon_0} \left[ \frac{\mathbf{\mu}_i \cdot \mathbf{\mu}_j - 3 \left( \mathbf{\mu}_i \cdot \mathbf{r}_{i,j} \right) \left( \mathbf{\mu}_j \cdot \mathbf{r}_{i,j} \right)}{r_{i,j}^3} \right]
\]

(9.1)

where \( \mathbf{\mu} \) is the carbonyl transition dipole, and \( r \) is the distance between the center of the C=O bonds. If we use the units of \([\mu] = \text{D} \text{Å}^{-1} \text{amu}^{-1/2} \) and \([r] = \text{Å} \), then the coupling constant in mdyn/(Å amu) is given by,

\[
\beta_{i,j} \left[ \frac{\text{mdyn}}{\text{Å} \cdot \text{amu}} \right] = 9.79 \times 10^{-2} \left[ \frac{\mathbf{\mu}_i \cdot \mathbf{\mu}_j - 3 \left( \mathbf{\mu}_i \cdot \mathbf{r}_{i,j} \right) \left( \mathbf{\mu}_j \cdot \mathbf{r}_{i,j} \right)}{r_{i,j}^3} \right]
\]

(9.2)

which is further converted to the more commonly used units in spectroscopy, cm⁻¹, by using a constant of \( 8.48 \times 10^5 / a_0 \text{[cm}^{-1}] \).⁵⁹

To perform the calculation we first extract the position of a single G-quartet from the NMR structure of TBA (PDB: 148D). We then optimize its geometry via density functional theory (DFT), with a B3LYP/6-31G(d,p) functional and basis, in the presence of Na⁺. As shown in Figure 9.11, the initial structure of the G-quartet is twisted, but becomes planar after geometry optimization with the Na⁺ ion placed in the center.
Figure 9.11: (a) Structure of the upper G-quartet of TBA from the solution NMR results (PDB: 148D).\(^5\) (b) Structure of the G-quartet with a sodium ion in the center after geometry optimization using DFT calculation with B3LYP/6-31G(d,p). The ribose and phosphate groups have been replaced by methyl groups in order to reduce computational cost. The side views are presented to show the planarity of the structures. The transition dipoles of the C=O stretches are shown as orange arrows. (c) Illustration of the couplings between different guanine monomers.

We used the value of 3.9 D Å\(^{-1}\) amu\(^{-1/2}\) to approximate the C=O stretch of GMP,\(^{55}\) and placed the transition dipoles according to their relative orientations to the carbonyl bond in the monomer (Figure 4.7), shown in Figure 9.11b. The nearest-neighbor and cross-strand coupling constants (Figure 9.11c) were calculated to be \(\beta_{1,2} = 6.6\ \text{cm}^{-1}\), and \(\beta_{1,3} = 4.7\ \text{cm}^{-1}\). Since the carbonyls do not interact with water molecules in a G-quartet, we used the C=O frequency of monomeric GMP in CDCl\(_3\), 1685 cm\(^{-1}\), to approximate the site energy, \(\epsilon\). Due to the symmetry of the G-quartet, we can construct a symmetric 4-by-4 element Hamiltonian:

\[
\hat{H} = 
\begin{bmatrix}
\epsilon & \beta_{1,2} & \beta_{1,3} & \beta_{1,2} \\
\beta_{1,2} & \epsilon & \beta_{1,2} & \beta_{1,3} \\
\beta_{1,3} & \beta_{1,2} & \epsilon & \beta_{1,2} \\
\beta_{1,2} & \beta_{1,3} & \beta_{1,2} & \epsilon 
\end{bmatrix}
\]

(9.3)
By diagonalizing the matrix, we obtained a pair of degenerate IR active modes at 1680 cm\(^{-1}\), and two other transitions at 1677 and 1703 cm\(^{-1}\) with very weak intensities. Obviously, these calculated frequencies do not match the experimental values, and it appears that the TDC model underestimates the coupling strengths, leading to smaller frequency splitting. Further, we have not accounted for couplings to other modes.

In order to gain more insight into how the coupling constants affect the IR spectrum, we varied the coupling constants in this simple model and summarized some general rules. Figure 9.12 shows the results from the calculation. We individually tested the effect of the magnitude and sign of the coupling constants \(\beta_{1,2}\) and \(\beta_{1,3}\). The symmetric coupling constant matrix is shown in the left column. The middle column shows the eigenvectors of the 4-by-4 element Hamiltonian (indicates the phase and amplitude of the wavevector on each monomer), and the right column plots the calculated FTIR spectra. Some noticeable observations are as follows: (1) A positive value of \(\beta_{1,3}\) results in a red shift of the asymmetric C=O stretch (IR active) and blue shift of the symmetric C=O stretch, leading to an overall red shift in the observed C=O frequency (Figure 9.12a). Conversely, negative \(\beta_{1,3}\) leads to an overall blue shift in the C=O frequency (Figure 9.12b). (2) The inclusion of only the nearest neighbor coupling \(\beta_{1,2}\) results in two IR active modes with the same frequency as the site energy, and two IR inactive modes which are on the blue and red sides of the IR active modes respectively. The magnitude of \(\beta_{1,2}\) determines the frequency shifts of the IR inactive modes (Figure 9.12c,e), and the sign of \(\beta_{1,2}\) determines the frequency ordering of the two IR inactive modes (Figure 9.12c,d). (3) Deviation from perfect planarity increases the intensity of the IR inactive modes.
Figure 9.12: Calculated C=O frequencies and mode decompositions for G-quartet. The left column shows the coupling constants used in the simulation. The middle column shows the mode decomposition from the four C=O sites. The number listed on top of each eigenvector indicates the corresponding energy and normalized intensity (in parenthesis). The right column shows the simulated FTIR using a Lorentzian linewidth of 3 cm⁻¹. The vertical black line marks the site energy of 1685 cm⁻¹.
Finally, we fit the experimental C=O frequencies (IR mode at 1671 cm$^{-1}$, and Raman modes at 1659 and 1702 cm$^{-1}$) by floating the site energy and the two coupling constants, while imposing symmetry onto the coupling constants. The fit result gives a site energy of 1675.8 cm$^{-1}$, and $\beta_{1,2} = 10.75$ cm$^{-1}$, and $\beta_{1,3} = 4.25$ cm$^{-1}$. The fitted coupling constants show the same trend as the TDC model in which the nearest-neighbor coupling $\beta_{1,2}$ is stronger than the cross coupling $\beta_{1,3}$. The magnitude and sign of $\beta_{1,3}$ is similar to the value from TDC calculation which contributes to the red shifting of the main IR band from the site energy. However, the magnitude of $\beta_{1,2}$ is much bigger than the TDC value. The positive sign of $\beta_{1,2}$ would give rise to a red shifted $B_g$ mode with alternating phase between the four C=O groups ($+-+$, see Figure 9.12c). Similar underestimation of the C=O coupling constant using TDC model was observed in Watson-Crick GC base-pairs (TDC gives a -5.5 cm$^{-1}$ shift where experiment shows a $\sim$-9.6 cm$^{-1}$ shift). Furthermore, this mode carries a weak IR intensity compared to the other IR inactive mode (the all-symmetric $A_g$ mode) which is blue shifted. This result may account for the slight asymmetry (slightly stronger on the red side) in the peak intensity of GMP at high concentration. Figure 9.13 shows the results from the fit and the comparison to the experimental spectrum.

![Figure 9.13](image)

Figure 9.13: Fitting results to the experimental peak frequencies of 200 mg/ml GMP in K*'. The calculated FTIR was simulated with a Lorentzian linewidth of 8 cm$^{-1}$.

We should point out that even though the calculated FTIR of the C=O stretch seems to reproduce the experimental peak relatively well, the contributions from the guanine ring modes was neglected in the model. The 2D IR spectrum of GMP shows significant
coupling between the C=O and ring modes, and therefore the reductive model presented here represents only explains the FTIR spectrum of G-quadruplexes. The importance of the coupling between the C=O and ring modes will become apparent from the 2D IR spectra of G-quadruplexes shown below.

9.3.1.2 Spectral changes of the ring modes

The guanine ring modes in the 1500-1600 cm\(^{-1}\) also experience dramatic spectral changes upon G-quadruplex formation. First of all, the G2 mode shifts from 1579 cm\(^{-1}\) to 1585 cm\(^{-1}\) and its intensity drops significantly. Secondly, the G4 mode shifts from 1538 cm\(^{-1}\) to 1535 cm\(^{-1}\) and grows in intensity.

DFT calculations using B3LYP/6-31G** by Setnicka et al. showed that ring-mode suppression can be achieved just with the formation of a single layer of G-quartet, regardless of the orientations of the ribose moieties from the four guanine bases. The authors further compared the calculations for a G-quartet with and without the presence of a sodium ion in the center of the G-quartet. The optimized geometries of these two G-quartet are twisted and near-planar in the absence and presence of a Na\(^{+}\) ion, respectively, similar to the structures found here (Figure 9.11). Better agreement is found between calculations and experimental IR absorption and vibrational circular dichroism (VCD) of the G-quadruplex when Na\(^{+}\) is included. Additionally, this calculation exhibits greater suppression of the ring-mode intensity. However, the enhanced peak intensity at 1535 cm\(^{-1}\) was reproduced in the calculation without Na\(^{+}\), but not with Na\(^{+}\). The authors attributed the discrepancy to the fact that their calculation was on a single G-quartet whereas in reality it should be a helical stack of G-quartets, which would give rise to a chiral signal (this peak is strong in the VCD spectrum).

More recently, with the aid of the Cartesian coordinate tensor transfer technique which can extend quantum calculations to larger systems, Andrushchenko and coworkers simulated the IR and VCD spectra of d(G), G-quadruplex and found close agreement with the experimental spectra. The calculations for four different H-bonded guanine dimers, a single G-quartet, single-stranded, duplex, and quadruplex structures of d(G)\(_8\) were performed, and only the single G-quartet and d(G)\(_8\) quadruplex showed ring-mode suppression, with the G-quadruplex having a more pronounced effect. This result is consistent with the findings reported by Setnicka et al. which demonstrates that the formation of G-quartet leads to the ring-mode suppression.

In addition, Andrushchenko and coworkers’ work also shows that stacking of G-quartets induces further suppression. Nevertheless, the 1535 cm\(^{-1}\) ring mode was predicted
in the G-quartet spectrum by Andrushchenko et al., in contradiction to Setnicka et al.
Therefore, the exact nature of these ring modes still require further studies.

The intensity reduction of the G ring mode has also been observed in other systems
such as Watson-Crick GC base-pairs. The temperature-dependent FTIR spectra of d(GC)n
plotted in Figure 8.7 show a similar ring mode suppression when the duplex is formed, and
the intensity is regained when the duplex is melted. Simulated spectra of two different
stacked GC base-pairs (self-complementary 5'-d(GC)-3' and 5'-d(CG)-3') by Andrushchenko
and coworkers in a separate publication showed similar ring mode suppression.

To separate out the contribution of H-bonds and base-stacking to ring mode
suppression, we measured the FTIR spectrum of a stacked guanine base, but not H-bonded.
The experiment was performed based on a strategy for measuring the base-stacking ability
introduced by Kool in which the base of interest was placed at the 5'-end of a short self-
complementary strand. The dangling base provides π-π base-stacking stabilizing interaction
but not H-bonds, and was found to increase the melting temperature of the duplex. Figure
9.14 plots the FTIR spectra of self-complementary 5'-G(AT)-3' and 5'-(AT)-3'. The AT
repeats in both sequence form base-pairs as evident from the spectral features (see Chapter 8
for peak assignments), while the 5'-end guanine base should stack onto the second base
adenine. Figure 9.14b shows the comparison between the GMP spectrum and the difference
spectrum of the two oligonucleotides. The peaks at 1680, 1640, 1620 cm⁻¹ indicate improper
subtraction of the A and T modes presumably due to minor differences in the content of
duplex formation, or the effect of guanine base-stacking on the C=O modes. Nevertheless,
the relative intensities of the 1665 cm⁻¹ C=O stretch, and the 1570 cm⁻¹ ring modes are
similar in both cases, demonstrating that guanine stacking alone does not cause ring mode
suppression (if it occurs, we would see much lower ring mode intensity compared to the
C=O mode). As a consequence, it is suggested that ring mode suppression can happen for a
stack of H-bonded base-pairs or Hoogsteen H-bonded guanine tetramer.
5'-GATATATATAT
TATATATATAG-5'

Figure 9.14: (a) 8 mM of d[G(AT)₅] and d(AT)₅ in 100 mM PBS, 250 mM NaCl, and 20 mM MgCl₂ at 6 °C. (b) The red curve is the scaled difference spectrum between the two spectra plotted in (a). The sharp features at 1685, 1640 cm⁻¹ are the residual thymine peaks due to improper subtraction. The black dash line is the normalized GMP spectrum.

9.3.2 2D IR spectroscopy

With the goal of obtaining more vibrational spectral features to constrain a theoretical model to predict the IR spectrum of G-quadruplexes, we measured the 2D IR spectra of several G-quadruplex models. Figure 9.15 shows the comparison between the temperature-dependent 2D IR spectra of 0.9 mM TBA in D₂O with 50 mM KCl, a mixture of 60% GMP and 40% TMP, pure GMP, and pure TMP. Several significant spectral features indicate that the low temperature spectra of TBA are different from those of the free nucleotides. In contrast, the 2D IR spectra of the free nucleotides are basically identical across different temperatures. In Figure 9.15, cross-peaks between the C=O stretch at 1670 cm⁻¹ and the G ring mode at 1535 cm⁻¹ appear (highlighted by the pink ellipses). These cross-peaks have very weak intensities in the pure GMP spectra, but are clearly present in the TBA spectra. Upon heating, these cross-peaks disappear and exhibit intensity similar to what is observed in the free GMP spectra. The G2 mode (marked by green arrow) is sharpened and shifted to 1586 cm⁻¹ and its intensity is suppressed at low temperature, as is the G3 mode at 1566 cm⁻¹. Upon heating, both ring modes recover their intensity and the lineshapes resemble those of the free GMP. The thymine ring mode also experiences slight intensity drop at low temperature. These 2D IR signatures for G-quadruplex formation can also be clearly seen in the spectrum of poly(G), plotted in Figure 9.16.
Figure 9.15: Temperature-dependent 2D IR spectra (ZZZZ) of 0.9 mM TBA in D$_2$O with 50 mM KCl, a mixture of 60% GMP and 40% TMP, pure GMP, and pure TMP. PBS buffer was used for the nucleotides. Note: these 2D spectra were taken with samples treated without membrane filters, so they contain a small peak at ~1670 cm$^{-1}$ which overlaps with the C=O stretch. Hyperbolic sine contours are used in order to enhance the cross-peaks marked by pink ellipses.
Based on X-ray diffraction of fibers, Zimmerman et al. proposed that poly(G) forms a four-stranded helical G-quadruplex. A later IR study suggested that poly(G) undergoes a slow transition (~5 days) from a metastable to stable form. This transition was observed with three notable IR spectral features: (1) C=O stretch shifts from 1692 cm\(^{-1}\) to 1674 cm\(^{-1}\); (2) a ring mode shifts from 1585 cm\(^{-1}\) to 1591 cm\(^{-1}\) and intensifies; (3) a weak band at 1610 cm\(^{-1}\) disappears within an hour. Using electron micrographs, Lesnik et al. concluded that the transition takes place via untwisting and elongation of a globular state to a stable linear form, and that both structures are four-stranded. Because of its polymorphism and the long transition timescale, experiments carried out under different experimental conditions showed inconsistent results. Howard et al. suggested that heating the sample facilitates the transition from metastable to stable state, while Petrovic and Polavarapu concluded that heating or lowering the pH induces unfolding of the G-quadruplex into a duplex.

In our experiments, the poly(G) sample was received from Sigma-Aldrich as lyophilized powder, which is in the metastable form according to the IR work by Howard et al. The sample was lyophilized once in D\(_2\)O for H/D exchange, and stored at -20 °C for 5 days. No heating was applied before the measurement to prevent potential duplex formation. The vibrations observed at 1675 cm\(^{-1}\) and 1589 cm\(^{-1}\) indicate the formation of G-quadruplex instead of the metastable form. The 2D IR spectra of poly(G) exhibits many spectral features similar to the TBA spectra, such as enhancement of the (G1,G4) cross-peaks, and the sharpening and blue shift of the G2 ring mode.

It is interesting to point out that Guzman et al. reported that the G C=O stretch of parallel and anti-parallel G-quadruplexes are at 1693 cm\(^{-1}\) and 1682 cm\(^{-1}\), respectively. However, of the G-quadruplexes we have studied so far, only d(GTGG), a parallel G-quadruplex, has a peak at 1682 cm\(^{-1}\) (others are mostly at ~1675 cm\(^{-1}\)). The higher C=O frequency suggests that those samples may have been prepared in the metastable form, similar to the metastable poly(G).

Besides the spectral features discussed above, G-quadruplex formation also manifests itself in the polarization-dependent 2D IR spectra. Figure 9.16 compares the ZZZZ and ZZYY 2D IR spectra. For GMP, the cross-peaks between the C=O stretch (G1) and the two ring modes at 1579 cm\(^{-1}\) (G2) and 1565 cm\(^{-1}\) (G3) are significantly enhanced in the ZZYY spectrum, indicating that the transition dipole moments of G2 and G3 are nearly parallel to G1. Indeed, in Chapter 4, we have shown that by fitting the ZZZZ and ZZYY 2D IR spectra, the relative dipole angles are 89° and 111° for G1/G2 and G1/G3, respectively. However, for TBA and poly(G) ring modes, not only do their intensity drop upon G-quadruplex formation, their cross-peaks to the C=O stretch become less sensitive.
to laser polarizations. These cross-peaks exhibit only slight increase in their relative intensity between $ZZZZ$ and $ZZYY$ polarizations. Additionally, these cross-peaks are much sharper compared to those in free GMP. Since the modes in TBA and poly(G) are the result of strong intra- and intermolecular coupling between modes from individual monomers and therefore heavily delocalized across the tetramer, it is more challenging and ambiguous to define the relative dipole orientations. Instead, we first point out the differences in the $ZZYY/ZZZZ$ relative cross-peak intensity for $(G1,G1)$, $(G1,G3)$, and $(G1,G4)$ in Table 9.2. It can be seen that the relative $(G1,G2)$ cross-peak intensity for TBA and poly(G) are quite similar, but drastically lower than GMP. However, the relative $(G1,G4)$ cross-peak intensity for all three molecules are relatively low. The results suggest that the vibrational modes of the G-quadruplex is the remixing of the basis modes in the free GMP, therefore changing the relative dipole orientations.

Figure 9.16: ZZZZ and ZZYY 2D IR spectra of 20 mg/ml GMP in PBS buffer, 0.9 mM TBA in D$_2$O, 30 mg/ml poly(G) in 10 mM sodium phosphate buffer.
Table 9.2. Cross-peak intensity ratio ZZYY/ZZZZ for GMP, TBA, and poly(G). Intensities were obtained by integrating the positive uphill (ω₁ < ω₂) cross-peak.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>G1,G4</th>
<th>G1,G3</th>
<th>G1,G2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMP</td>
<td>0.65±0.01</td>
<td>1.35±0.15</td>
<td>1.65±0.09</td>
</tr>
<tr>
<td>TBA</td>
<td>0.45±0.02</td>
<td>0.52±0.01</td>
<td>0.49±0.04</td>
</tr>
<tr>
<td>Poly(G)</td>
<td>0.79±0.14</td>
<td>1.38±0.27</td>
<td>0.54±0.01</td>
</tr>
</tbody>
</table>

To gain more insights into the vibrational modes in G-quadruplex, we modelled the tetramer of methylated guanine using DFT, as shown in Figure 9.11b. Since the G-quartet forms stable Hoogsteen H-bond network between most of the labile protons and H-bond acceptor oxygen and nitrogen atoms (except for N³ atom and a free N-D bond in N²-D₂), it is expected that the influence of solvent molecule on the vibrational modes will be less significant than in the free GMP case, and therefore no solvent model was included in the calculation. Figure 9.17a-c show the calculated vibrational modes that would correspond to the G1 and G2 modes in guanosine, which are illustrated in Figure 9.17d and e, respectively. In guanosine, G1 is mostly C=O stretch plus N¹-D bend and slight pyrimidine ring vibration. G2 is a stronger pyrimidine vibration that consists of stretching motions of C²=N³ and slight C=O, and bending of N¹-D and N²-D². These two transition dipoles were determined to be 89° apart by fitting the polarization selective 2D IR spectra.57

In the tetramer G1 mode (denoted as GT₁, Figure 9.17a), the M₁ monomer (lower left corner, see Figure 9.17f for notation) primarily exhibits the same vibrational motions as the free monomer (Figure 9.17d), while the M₃ monomer shows the asymmetric stretches. This is consistent with the model calculations of the coupled C=O stretches shown earlier: there are two degenerate C=O modes, one of which has asymmetric C=O stretches of the cross-strand carboxyls (M₁, M₃), and another mode has the same motions from M₂ and M₄. Here only one of the two degenerate GT₁ modes is shown. We will use this mode as a reference to discuss how its coupling to the ring vibrations may influence the cross-peak strengths. It can be seen that the vibrational motions are not exclusively localized onto these two monomers. The N¹-D stretch of the M₂ monomer is coupled to the C=O stretch of M₁ through H-bond interaction. Overall the transition dipole points in a direction that is nearly parallel to the carbonyl bonds of M₁ and M₃.

Figure 9.17b and c show the two degenerate tetramer GT₂ and GT₂' modes. The GT₂ modes are more delocalized across all four bases than GT₁. Looking at the GT₂ mode (Figure 9.17b), the vibrations of the free guanosine G2 mode (Figure 9.17e) such as N¹-D and N²-D₂ bending are more preserved in M₁ and M₃ compared to M₂ and M₄. The transition
The dipole of $G^{T2}$ was calculated to be $70^\circ$ rotated from that of $G^{T1}$ (see Figure 9.17f). However, the motions of $M_2$ in the $G^{T2}$ mode does not seem to have strong inter-base coupling to the vibrations of $M_1$ in the $G^{T1}$ mode.

Figure 9.17: (a-c) DFT calculated vibrational modes for mG4 that represent the $G^{T1}$ and $G^{T2}$ modes. The force vector and transition dipole are indicated by the green and orange arrows, respectively. The $G^{T2}$ and $G^{T2'}$ modes are degenerate. Another degenerate $G^{T1}$ mode which consists of primary C=O stretch of the other two guanines is not shown. The center ion is Na$^+$. (d,e) The calculated $G1$ and $G2$ modes of guanosine with one D$_2$O molecule. (f) Summary of the relative transition dipole moments for mG4.

In the $G^{T2'}$ mode, which can be obtained simply by rotating $G^{T2}$ by $90^\circ$. The $M_2$ and $M_4$ monomers carry most of the N$^1$-D and N$^2$-D$_2$ bending vibrations. These two motions are important in the Hoogsteen H-bond network and couple to the C=O stretch of a neighboring base. Consequently, the inter-base coupling between $G^{T1}$ and $G^{T2'}$ is likely to contribute more significantly to the ($G1,G2$) cross-peaks. The transition dipoles of these two
modes are only 20° apart, which is consistent with the lower sensitivity of the cross-peak to the laser field polarization (Table 9.2). It should be emphasized that we are not ruling out the contribution to the cross-peaks by the intra-base coupling (between G\textsuperscript{T}1 and G\textsuperscript{T}2), which is clearly present from the GMP 2D IR spectra. So far, we have not accurately determine the relative dipole orientations from the experimental spectra to be able to quantify the inter- and intra-base coupling. Nonetheless, the remixing of the vibrational modes and the appearance of the G\textsuperscript{2'} 2 mode in a G-quartet offers a reasonable explanation that is consistent with the experimental spectra.

Similar analysis can be applied to the tetramer G\textsuperscript{T}4 mode (Figure 9.18). The G4 mode in free guanosine (Figure 9.18e) is assigned to a delocalized ring mode across both the pyrimidine and imidazole rings, with significant motions from the stretch of N\textsuperscript{7}=C\textsuperscript{8}, and C\textsuperscript{4}=C\textsuperscript{5}, as well as bending of C\textsuperscript{8}-H.\textsuperscript{37} In G-quartet, the G\textsuperscript{T}4 modes (Figure 9.18b,c) are not as delocalized as the G\textsuperscript{T}2 modes. The G\textsuperscript{T}4 mode has participation from M\textsubscript{1} and M\textsubscript{3} (Figure 9.18b), while the G\textsuperscript{T}4' mode has contribution from M\textsubscript{2} and M\textsubscript{4} (Figure 9.18c). Even though both the G\textsuperscript{T}4 modes in G-quartet or guanosine have significant N\textsuperscript{7}=C\textsuperscript{8} stretch, the G\textsuperscript{T}4 modes exhibit additional contribution from the C=O stretch. This observation may explain the enhanced (G1,G4) cross-peaks. It also suggests that the main coupling between G1 and G4 modes is intramolecular (which makes sense because the N\textsuperscript{7}=C\textsuperscript{8} bond from a neighboring M\textsubscript{2} guanine is actually far away from the C=O of M\textsubscript{1}), and therefore the two transition dipoles are nearly parallel (Figure 9.18a and b). The calculated angle between these two transition dipoles is 9° (compared to 37° in guanosine), which is consistent with the small ratio reported in Table 9.2. On the other hand, if the coupling was intermolecular between the neighboring guanines, for example between the two modes shown in Figure 9.18a and c, the angle would become 99°, which would lead to significant enhancement of the cross-peak in the ZZYY spectrum.

Furthermore, larger degree of participation of C=O in the G\textsuperscript{T}4 mode is also consistent with the correlated spectral broadening of the G1 and G4 modes. The correlations can be related to the tile angle of the cross-peak nodal line with 45° and -45° being highly correlated or anti-correlated, while 0 being not correlated. The (G1,G4) cross-peak in the ZZYY spectrum of poly(G) shows a nodal line of close to 45°, indicating the high correlation coefficient.
Interestingly, the relative \((G1,G3)\) cross-peak intensity of TBA is lower than that of poly(G), which is similar to that of GMP. The observation suggests that the distinction between parallel (poly(G)) and anti-parallel (TBA) G-quadruplexes may be found in the subtle differences of the ring modes, even though the general 2D IR features for the G-quadruplex formation is similar for TBA and poly(G).

Following this idea, we focus our analysis on the ring modes. Figure 9.19 shows the guanine \(C=N\) and \(C=C\) stretch region (1520-1610 cm\(^{-1}\)) for GMP, TBA, and poly(G). The lineshapes of the ring modes in these three samples seem to differ quite dramatically. Here the spectra are normalized to the maximum within this spectral region and therefore the ring-mode intensity suppression is not observed. GMP shows two dominant transitions for the free nucleotide state; TBA shows three strongly coupled bands for the anti-parallel G-quadruplex; and poly(G) shows four strongly-coupled peaks for the parallel G-quadruplex. Both TBA and poly(G) have an enhanced 1535 cm\(^{-1}\) peak and a blue-shifted ring mode.

Figure 9.18: (a-c) DFT calculated vibrational modes for mG4 that represent the \(G^T1\) and \(G^T4\) modes. The force vector and transition dipole are indicated by the green and orange arrows, respectively. The \(G^T4\) and \(G^T4'\) modes are degenerate. Another degenerate \(G^T1\) mode which consists of primary \(C=O\) stretch of the other two guanines is not shown. The center ion is \(Na^+\). (d,e) The calculated \(G1\) and \(G4\) modes of guanosine with one \(D_2O\) molecule. (f) Summary of the relative transition dipole moments for mG4.
~1580 cm$^{-1}$ (with poly(G) further shifted). The G3 mode of TBA at 1567 cm$^{-1}$ is much weaker than the corresponding GMP peak, whereas this mode in poly(G) is quite strong. The much weaker G3 mode in TBA may be due to the fact that anti-parallel guanine bases stack, leading to cancellation of the transition dipoles that are not perfectly in-plane. These dramatic changes cannot be simply ascribed to changes in H-bonding, and are very likely an indication of excitonic vibrational couplings of the G-quartet. It is clear from the spectra that different conformations will be separable and our next goal is to find a definitive assignment and the underlying physics explaining these spectral features. The polarization selective 2D IR spectra, and the distinct cross-peaks will provide structural constraints (dipole orientations) in modeling G-quartet and TBA folding.

![GMP, TBA, polyG spectra](image)

Figure 9.19: Ring mode spectral region zoom area of the spectra shown in Figure 9.16 collected in the ZZYY geometry. The contours were renormalized to the maximum amplitude within this spectral region.

9.4 Acknowledgements

I thank Kimberly Hamad-Schifferli and Helena de Puig Guixe for helpful discussions, in particular on the background of TBA. I thank Joyce L. Yang for performing the PAGE gel experiments, which provided important information on the molecular aggregate of TBA at high concentrations. I thank Carlos Baiz and Luigi De Marco for their careful reading of this chapter and much needed comments.
9.5 References


CHAPTER 10
Dynamics of TBA Unfolding and TBA-antidote Dissociation

10.1 Introduction

It is well known that most proteins and nucleic acids adopt specific three-dimensional structures in order to perform their biological function. With the increasing discoveries of the diverse role of RNA, especially as catalytic enzymes, there is a growing appreciation of the folding dynamics of RNA. Similarly, enzymatic reactions carried out by proteins in well-defined structures and neurodegenerative disease due to protein misfolding has raised the importance of studying protein folding. Thermodynamically, the folding process is a delicate balance between the loss of configurational entropy and gain of favorable enthalpy by making native contacts. However, the exact driving forces for the folding of nucleic acids and proteins can be quite different. The constituent building blocks of nucleic acids and proteins vary both in number (five vs. twenty), and in their chemical properties (nucleic acid bases are mostly hydrophobic while amino acid side chains can be polar, hydrophobic, or ionic). Nucleic acids are also drastically different from proteins in that their backbone is highly negatively charged, so their folding is sensitive to stabilizing ions. Numerous experimental and theoretical work has been done in both fields in order to describe the interplay between different interactions that ultimately leads to the formation of secondary and tertiary structures.
Although the detail nature of the driving forces for folding of nucleic acids and proteins are distinct, for example base-stacking (nucleic acids) versus hydrophobic collapse (proteins), multiple competing interactions involved in both folding problems result in “rugged” energy surfaces. An important question in the field of biomolecular folding is whether the process is heterogeneous, meaning if the reaction has multiple pathways and if there are kinetically trapped states in the local minima. Two-state model consisting of the folded and unfolded states is the most simplistic view of biomolecular folding. However, recently numerous studies have shown multi-exponential, stretched-exponential, and non-exponential folding kinetics, suggesting a heterogeneous folding picture.\textsuperscript{2,9,12}

Small proteins and nucleic acids have been found to fold in the sub-millisecond timescale, and the maximum rate of protein folding was estimated to be the on the order of $\mu s$.\textsuperscript{1,45} In order to observe these fast reactions, it is necessary to use rapid initiation methods which introduces sudden change to the solvent environment, shifting the equilibrium constant of [folded]/[unfolded]. The change in population can then be monitored with structural sensitive probes. The stability of folded nucleic acids is highly dependent on temperature and ionic strength. In our experiments, taking advantage of the steepness of the melting curves of these biomolecules, we use laser induced temperature-jump (T-jump) to abruptly unfold the biomolecules with an abrupt 10-20 °C increase in temperature. Although different initial states are prepared in T-jump unfolding and folding experiments, microscopic reversibility requires that the contributing rate equations be the same.

Investigation of the folding kinetics of nucleic acids using T-jump methods began in the early seventies when people reported single-exponential kinetics and proposed a two-state mechanism, with the folded and unfolded states separated by a significant energy barrier.\textsuperscript{13,15} More recently, techniques with improved time resolution and structural sensitivity have started to reveal the complex nature of nucleic acid folding. Non-Arrhenius behavior and stretched exponential kinetics have been observed for hairpin unfolding.\textsuperscript{2,12} Multiphase kinetics for RNA hairpin unfolding have been reported at different temperatures.\textsuperscript{10} These studies raise the question of whether nucleic acid hairpins fold in a two-state cooperative fashion and suggest the presence of partially folded or misfolded intermediate states.

Here we study the early unfolding events of thrombin-binding aptamer (TBA) which is a single-stranded DNA 15 mer that folds into a chair-like G-quadruplex.\textsuperscript{16,17} TBA folding has attracted much attention because when they fold into the chair form, they can bind to the protein thrombin and inhibit its activity of generating blood clots, serving as an alternative anticoagulant.\textsuperscript{18} Additionally, the complementary strand of TBA (denoted as cTBA) can serve as an antidote of the anticoagulant by releasing TBA from thrombin and...
restoring its bioactivity. The dynamics and the interplay between TBA folding, binding to thrombin, and association with cTBA remain unclear even though the chair structure has been solved for years. Therefore, as a first step towards understanding the process, IR in this chapter we investigate the unfolding of TBA and the dissociation of TBA-cTBA using T-jump 2D IR spectroscopy.

Among the variety of possible scenarios, previous investigations have proposed two unfolding mechanisms for TBA. Using NMR spectroscopy, Mao and Gmeiner measured the difference in hydrogen exchange rate of the guanine imino protons, and proposed an unfolding model in which the $G_1-G_{15}$, $G_2-G_{14}$, and $G_5-G_{11}$ base-pairs uncouple first, followed by the opening of the TGT loop and ending with the opening of the TT loops (Figure 10.1a). Similar to the proposed model based on experiments, Jayapal et al. performed a 2 ns MD simulation under equilibrium condition and observed an unfolding event which was initiated by the breaking of $G_1-G_{15}$ base-pair, followed by the uncoupling of the $G_2-G_{14}$ pair and loss of π-stacking between the two G-tetrads. The model from these two studies suggests a folding intermediate that consists of a 3+1 (three guanine bases are H-bonded) top tetrad and a 2+2 bottom tetrad.

![Proposed molecular mechanisms of TBA unfolding from the literature. Figures in panel (a) and (b) are adapted from refs. [1] and [3], respectively, with permissions.](image)

However, work from the Parrinello group proposed a different unfolding scheme. Their 80 ns metadynamics simulation of TBA exhibits three main energy minima (Figure 295).
The first energy minimum represents the native folded chair-form. In the second minimum, the G₁₅ base at the 3'-end is solvent exposed so that both the G₁-G₁₅ and G₁₀-G₁₅ base-pairs are broken. In the third minimum, the H-bonds to G₁₄ are also lost such that the entire 3'-end is frayed. The remaining structure forms a stable folding intermediate state, a G-triplex. The authors have also experimentally proved the existence of this triplex structure by solving the NMR structure of a modified TBA whose last four bases were removed.²²

To test these proposed unfolding mechanisms, which we term "pacman" (Figure 10.1a) and “fraying” (Figure 10.1b), and to detect these folding intermediate states, we performed T-jump 2D IR experiments, in which a laser T-jump perturbs the chair form rapidly within 10 ns, and the relaxation is monitored.

10.2 Equilibrium temperature-dependent FTIR spectra of TBA and tTBA

The details of the in-plane base vibrations of G-quadruplex have been discussed in Chapter 9. Figure 10.2a and b show the temperature-dependent FTIR spectra of TBA with and without the presence of K⁺, respectively. Briefly, when the TBA chair structure melts with increasing temperature, the sharp C=O stretch at ~1670 cm⁻¹ broadens, the ring modes at ~1580 cm⁻¹ gain intensity and red shift, and the ring mode at 1535 cm⁻¹ drops in intensity.

These spectral changes can be compared to those of the constituent free nucleotides, GMP and TMP, shown in Figure 10.3. Upon increasing temperature the H-bonds between the carbonyls of the free bases and the water molecules are disrupted and the C=O bond increases in strength and shifts to higher frequency, leading to a gain feature on the blue side and a loss feature on the red. This is in sharp contrast to the C=O stretch region of the TBA thermal difference spectra (Figure 10.2e and f), which shows a loss feature at 1670 cm⁻¹ accompanied with gain features on both sides. The difference between the spectra of TBA and the free bases clearly indicates unfolding of TBA and not simply thermal effects. Upon chair formation in the presence of K⁺, the eight carbonyls are electrostatically coordinated by the metal ion, causing a transition from the structurally disordered ensemble to a more rigid and symmetric one. The resulting ordered structure and vibrational coupling within the complex results in sharpening of the C=O peak.²⁵ Without the presence of coordinating ions, the C=O peak is not narrowed even with the chair formation.
Figure 10.2: (a-d) Temperature-dependent FTIR spectra for 1mM TBA in 50 mM PBS (a), TBA in D$_2$O (b), tTBA in 50 mM PBS (c), and tTBA in D$_2$O (d). (e-h) Thermal difference spectra. The dash lines are shown to guide the eyes. (i-l) The normalized 2nd SVD component for the guanine 1535 cm$^{-1}$ ring mode spectral region (1510-1550 cm$^{-1}$).

The second spectral feature associated with the chair formation is the ring-mode suppression in the range of 1560-1580 cm$^{-1}$. In the case of GMP (Figure 10.3d), the loss at 1575 cm$^{-1}$ reflects a red shift of the ring modes upon weakening of solvent H-bonds to GMP. In contrast, a much smaller loss feature is seen in the TBA spectra at 1580 cm$^{-1}$, indicating a real peak intensity growth at 1565 cm$^{-1}$, accompanied with a small red shift. Finally, a clear loss feature of the 1535 cm$^{-1}$ mode is observed in all TBA spectra, which is absent in the GMP spectra. This ring mode which exhibits pronounced 2D IR cross-peaks to the C=O stretch at 1670 cm$^{-1}$ (Chapter 9) is a clear indicator of G-quadruplex formation.$^{24,25}$
To characterize the spectral features associated with the proposed folding intermediate, we investigated the IR spectroscopy of the 11-mer oligomer 5'-GGTTGGTGTGG-3' corresponding to the 1-11 sequence of TBA. This oligo was shown by NMR to fold into the triplex structure shown in Figure 10.4, and is referred to as triplex TBA (tTBA). Besides the missing 3'-end, the triplex structure is similar to the chair structure by retaining the Hoogsteen H-bond network among the remaining three guanine bases in each layer of G-triads (Figure 10.4b,c). The guanine bases in the G-triplex remain at the four corners of a square and do not form an equidistant triangular triplex. In Figure 10.2, the spectra of tTBA show a striking resemblance to those of the unmodified TBA. Comparing the difference FTIR of TBA and tTBA with K+ in Figure 10.2e and g, the tTBA spectra reproduce the primary spectral features identified with G-quadruplex formation, although slight frequency shifts are observed in the 1682 and 1670 cm⁻¹ modes and decreased intensity.
in the 1535 cm⁻¹ mode. The spectral similarity can be rationalized with the similarity in the structural motifs between G-quadruplex and G-triplex. The CD spectra of TBA and tTBA reported by the Parrinello group are also similar, with slight frequency shifts. The small frequency shifts observed in the IR spectra may indicate the difference in the couplings between the C=O modes in a G-quadruplex or G-triplex.

Figure 10.4: NMR structure of triplex TBA. Figures are adapted from ref. [3] with permissions.

Melting curves were obtained by applying singular value decomposition (SVD) to the G-quadruplex sensitive 1535 cm⁻¹ ring mode region (Figure 10.2i-l). This results in sigmoidal melting curves, whereas the analysis on the free bases over the same range produces a straight line. This difference indicates structural changes and allows for the determination of the melting temperatures. For TBA in K⁺, a two-state analysis reveals a Tₘ value of 50.8 °C, which agrees well with the value of 50 °C obtained from CD spectra (Figure 9.6). Removing the last four bases of TBA destabilizes the secondary structure by ~15 °C. Without K⁺, the stability of G-quadruplex decreases dramatically. Although it is clear that we did not reach 100% folded fraction state for TBA and tTBA in D₂O, an estimate of Tₘ based on the mid-point of the melting curve gives a value of <23 °C for TBA, which is in close agreement with the 20 °C value obtained from CD spectroscopy (Figure 9.6). The Tₘ for tTBA in D₂O is estimated to be <21 °C.
10.3 T-jump transient 2D IR of TBA unfolding

10.3.1 T-jump experiments of TBA

We performed T-jump experiments on TBA in D$_2$O with and without excess K$^+$. The T-jump was initiated 10 °C below the melting temperature, and jumped from 40 to 50 °C and 15 to 25 °C for TBA with and without K$^+$, respectively. Our experiments measure the T-jump induced changes to the equilibrium dispersed pump-probe signal or 2D IR spectrum, which we refer to as transient dispersed pump-probe (t-DPP) and transient 2D IR (t-2DIR) experiments, respectively.

Figure 10.5 shows the time-dependent kinetics from the T-jump t-DPP spectra measured at the peak of the G ring mode at 1580 cm$^{-1}$, where the time-axis has been partitioned to illustrate relaxation behavior on the nanosecond and microsecond time scales. This preliminary analysis illustrates the difference in the kinetics of TBA with and without K$^+$, and compares both to the response of the free nucleotide GMP. For both TBA with K$^+$ and free GMP, only a 5% increase in the t-DPP signal was observed, which rose with a T-jump pulse-width limited time-scale (<5 ns) and remained constant throughout the 5 ns to 100 μs time window probed by our experiment. In contrast, TBA in D$_2$O shows time-dependent absorption changes in addition to the instantaneous response. Additional increases in t-DPP signal are observed on the ~50 ns timescale and the few-μs timescale. The abrupt increase in signal is common to all T-jump experiments we have performed, and originates in the picosecond temperature dependent changes in hydrogen bonding configurations of the water about the solute. The lack of other signals for TBA with K$^+$ suggest that the global unfolding response must occur on longer timescales than our instrument accesses. Indeed, this is consistent with previous stopped-flow experiments which reported that the Pb$^{2+}$-induced folding of TBA chair structure follows a single-exponential behavior with a time constant of 1.05 sec$^{26}$ The additional time-dependent spectral changes for TBA in D$_2$O on the ns and μs timescale are therefore of a different origin.
Figure 10.5: The T-jump kinetic traces measured from the t-DPP spectra at 1584 cm\(^{-1}\), for TBA in D\(_2\)O at T-jump of 15-25 °C (red); TBA in the presence of K\(^+\) at T-jump of 40-50 °C (black); and GMP in 50 mM potassium phosphate buffer at T-jump of 25-25 °C (purple). The nanosecond and microsecond responses were measured from different sets of data under the same experimental conditions. The μs response of TBA in D\(_2\)O was scaled by 0.625 to match the amplitude of the ns response. The lines are only meant to guide the eye, but were obtained from fitting. This μs response was fit to an exponential function with a time constant \(\sim\)1 μs. In the ns regime, the solid lines represent a convolution of a Gaussian pulse (with a \(\sim\)5 ns FWHM) that represents the T-jump pulse, with a double-exponential function that describes the sample response. Both the responses from TBA with K\(^+\) and GMP contain only the pulse-width-limited step response, whereas TBA in D\(_2\)O shows an additional kinetics phase at \(\sim\)50-200 ns.

Figure 10.6c plots the t-DPP spectra for TBA in D\(_2\)O as a function of both the \(\omega_3\) detection axis and T-jump delays showing both the signal increases (red) and decreases (blue) on a log-time axis with time increasing vertically. The t-DPP spectra at selected T-jump delays are shown in Figure 10.6b. For comparison, we also show the thermally-induced spectral changes to the equilibrium DPP spectra, which we denote as \(\Delta DPP_{eq}\). Figure 10.6a presents the \(\Delta DPP_{eq}\) of GMP, TMP and TBA in D\(_2\)O, which were derived from the second SVD components of projected 2D IR spectra. In contrast to FTIR spectra, each peak in DPP spectra shows up as a doublet with opposite signs, corresponding to the ground state bleach and excited state absorption. As a consequence, the positive and negative features in
the difference DPP spectra can be either a gain or a loss, and a comparison to the equilibrium DPP spectra is important for proper interpretation (Appendix 10.A).

Immediately following the 5 ns T-jump, the t-DPP spectra are dominated by the positive response at 1580 cm\(^{-1}\). The additional loss feature at 1590 cm\(^{-1}\) indicates that these features arise from a red shift of the to the G ring mode, similar to what is observed in the ADPP\(_{eq}\) spectra of GMP and TBA (Figure 10.6a). The lack of a positive peak at 1531 cm\(^{-1}\) in the t-DPP spectrum of TBA indicates that this response is not due to the melting of G-quadruplex. At 5 ns, the C=O spectral region in the t-DPP spectrum is drastically different from the ADPP\(_{eq}\) spectrum of TBA, which shows a feature corresponding to the broadening of the C=O peak. The observed broad positive peak at 1664 cm\(^{-1}\) accompanied with a negative dip at 1649 cm\(^{-1}\) in the t-DPP spectrum actually resembles the ADPP\(_{eq}\) spectrum of TMP.

Similar features are present in the transient 2D IR (t-2D IR) spectra shown in Figure 10.7. For comparison, the equilibrium spectrum and the 2\(^{nd}\) SVD spectral component of the temperature-dependent 2D IR spectra of TBA in D\(_2\)O are shown in panel (c) and (d), respectively. The 56 ns t-2D IR spectrum exhibits a red shift of the 1580 cm\(^{-1}\) ring mode (marked by a yellow box), and the lineshape remains sharp and narrow. The loss of cross-peaks between the 1535 and 1670 cm\(^{-1}\) diagonal peaks (pink ellipses in Figure 10.7d) is not observed. Given the signal-to-noise level in this cross peak region, we conclude that the guanine core stays intact. A zoomed-in spectrum of the 1600-1700 cm\(^{-1}\) region (Figure 10.8a) shows a loss of the thymine ring mode intensity at 1630 cm\(^{-1}\) (green ellipse), which can be compared to the 2\(^{nd}\) SVD component of the temperature-dependent 2D IR spectra of TMP (Figure 10.8c). The C=O region is more complicated to interpret. The positive feature that elongates along \(\omega_3 = 1660\) cm\(^{-1}\) resembles the feature seen in the TMP spectrum. However, the negative feature marked by the purple arrow matches neither the TBA nor the TMP spectrum.

Taken together, the data indicate that the ns response does not originate from the melting of the chair form due to the absence of the three spectral features: (1) loss of the 1535 cm\(^{-1}\) peak; (2) the response from the guanine C=O; (3) and the loss of cross-peaks between the 1535 and 1670 cm\(^{-1}\) modes. While these spectral changes suggest that the G-quadruplex core remains mostly intact, the slight increase of the guanine ring-mode intensity may indicate increased conformational fluctuations and the weakening of base-stacking. Similarly, the C=O response comes mainly from the thymine bases which could be due to breaking of H-bonds between thymine bases and the solvent molecules.
Figure 10.6: (a) Thermal spectral changes of the equilibrium dispersed pump-probe (DPP) spectra, represented by the second SVD components of the temperature-dependent spectra. Equilibrium DPP spectra were derived from the 2D IR spectra. The samples are TMP (blue) and GMP (red) in 50 mM phosphate buffer, and 1 mM TBA in D₂O (green). (b) Transient difference DPP spectra for 1 mM TBA in D₂O with T-jump from 15-25 °C. Spectra at selected T-jump delays are shown. (c) 2D plot of the t-DPP spectra as a function of both the detection frequency axis and T-jump delay. The spectra have been normalized to the maximum value. (d) The time traces at four different wavelengths (marked in (b)). Decreasing signals are plotted with increasing trend for a more convenient comparison. These time traces are fit to double-exponentials.
Figure 10.7: (Left) T-jump transient 2D IR spectra of TBA in D$_2$O from 15-25 °C, at T-jump delays of 56 ns (a) and 3.2 µs (b), in the 1520-1725 cm$^{-1}$ spectral region. (Right) The equilibrium spectrum (c), and the 2$^{nd}$ SVD component of the temperature-dependent 2D IR spectra to represent thermal changes to the spectrum.
Figure 10.8: T-jump transient 2D IR spectra of TBA in D₂O from 15-25 °C, at T-jump delays of 56 ns (a) and 3.2 μs (b), in the 1590-1725 cm⁻¹ spectral region. (c) The second SVD components of the temperature-dependent 2D IR spectra of TMP to represent the thermal difference spectrum.

With increasing T-jump delay, the primary change observed is an increase in the peak amplitudes, with minimal frequency shifts or lineshape changes. One noticeable new feature is a loss of 1699 cm⁻¹ in the t-DPP spectra after 100 ns (Figure 10.6b). This feature is not observed in the difference DPP spectra of GMP, TMP, or TBA. The negative ring mode peak shifts from 1546 cm⁻¹ to 1558 cm⁻¹. Moreover, the ring mode intensity at ~1570 cm⁻¹ increases in the 3.2 μs t-2D IR spectrum (Figure 10.7b).

To monitor the time evolution of these different spectral features, the peak amplitudes of different frequency slices from the t-DPP spectra are plotted in Figure 10.6d. These time traces clearly show deviation from mono-exponential behavior, with a fast sub-μs and a slower few-μs phase. It is not immediately clear whether these traces follow multi-exponential, stretched-exponential, or power laws. Reasonable fits are obtained with a double-exponential function, shown as a black line in the figure; however, due to the large error bars associated with the time constants, we cannot draw conclusions on the frequency-dependent variation in the timescales. Moreover, the timescale obtained from a single frequency slice can be erroneous since different peaks may overlap in pump-probe spectra. Therefore, we applied SVD analysis to describe the global time dependence of a broader spectral region, as well as remove experimental noise.
10.3.2 SVD analysis

Figure 10.9 shows the first SVD spectral and temporal components when SVD analysis is applied to different regions of the t-DPP spectra. The first SVD spectral component of the difference spectra represents the spectral change with increasing T-jump delay. We first analyzed the 1520-1600 cm⁻¹ spectral region (Figure 10.9a) which contains the ring mode of both guanine and thymine. The 1570-1600 cm⁻¹ region exhibits the red shift and intensity increase of the guanine ring mode, while the 1520-1560 cm⁻¹ region is mostly due to the broad induced absorption of the thymine ring mode (see Figure 10.6a for comparison). Figure 10.9c shows the associated time trace and is well-fit to a double-exponential function (38±12 ns, 0.92±0.36 μs). The time constants obtained from this fit are summarized in Table 10.1.

Figure 10.9: The first SVD spectral (a,b) and temporal (c,d) components of different spectral regions of the t-DPP spectra for TBA in D₂O at T-jump of 15-25 °C.
Table 10.1. Fitting parameters for the first SVD component of the t-DPP spectra for TBA in D₂O.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>1520-1600 cm⁻¹</th>
<th>1600-1720 cm⁻¹</th>
<th>1520-1600 cm⁻¹</th>
<th>1600-1720 cm⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_1 ) [ns]</td>
<td>41±11</td>
<td>40±20</td>
<td>31±22</td>
<td>39±22</td>
</tr>
<tr>
<td>( \tau_2 ) [μs]</td>
<td>0.97±0.37</td>
<td>1.94±0.42</td>
<td>0.66±0.35</td>
<td>1.31±0.41</td>
</tr>
</tbody>
</table>

Similar analysis was performed on the spectral region that reports on the C=O stretch region (1640-1720 cm⁻¹, Figure 10.9b). The kinetic trace of the C=O stretch region can also be described with a double-exponential function with relaxation time constants of 40±20 ns and 1.84±0.42 μs. The faster response from C=O is the same (within experimental error) as the ring mode response, while the slower phase of C=O is roughly a factor of two slower than that of the ring modes. It is not unreasonable to observe variations in the time constants, particularly since different spectral regions report on the kinetics of different bases, and the unfolding mechanism of TBA has been proposed to involve multiple intermediate states with a rugged energy landscape. Nonetheless, the small separation of these time constants does not allow us to conclusively distinguish between simple two- or three-state kinetics or heterogeneity in the underlying dynamics.

As the initial temperature of the T-jump experiment is raised above \( T_m \) to 25 °C, the μs kinetics become slightly faster while the ns kinetics remain the same. In general, the spectral features observed with a T-jump of 25-35 °C (Appendix 10.B) are similar to those observed with T-jump experiments jumping from 15-25 °C.

### 10.3.3 T-jump experiments of TBA triplex

In order to test the origin of the T-jump response and test the “fraying” model for TBA unfolding, we repeated the T-jump experiments on \( t \)TBA in D₂O. The transient t-DPP spectra for a T-jump from 25 to 35 °C are presented in Figure 10.10b. These show a striking resemblance to the t-DPP spectra of the TBA G-quadruplex structure. A closer examination reveals the absence of the 1699 cm⁻¹ loss feature together with a narrowing of the C=O peak. The time dependence of the response is also drastically different. The transient spectra are invariant after 100 ns. The kinetic traces at 1584 cm⁻¹ for both TBA and \( t \)TBA are plotted in Figure 10.11, which clearly shows that a ns response is shared in both samples while a μs response is only present for TBA in D₂O.

307
Figure 10.10: Transient t-DPP spectra for TBA in D₂O at 15-25 °C (a) and tTBA in D₂O at 25-35 °C (b).

Figure 10.11: The T-jump kinetic traces measured from the t-DPP spectra at 1584 cm⁻¹, for TBA in D₂O at T-jump of 15-25 °C (red); and tTBA in D₂O at T-jump of 25-35 °C (blue). Same representation as Figure 10.5.
10.4 Discussion

Previous studies have suggested two different unfolding mechanisms. Based on the hydrogen-deuterium exchange rate of the imino protons, the "pacman" model was proposed in which the TBA chair form is split into half consisting of two Hoogsteen hairpins. The "3'-fraying" model was proposed on the basis of MD simulation and the experimental prove for the existence of the stable TBA triplex lacking the last four bases of TBA. Figure 10.12a illustrates these and other possible unfolding mechanisms, showing with blue arrows the connectivity of conformations with partial secondary structure stabilized by Hoogsteen hydrogen bonding. In principle, because of the symmetry of the chair structure, fraying at the 5' should be equally likely. However, both the crystal and NMR structures show that G8 in the TGT loop stacks onto G1 (Figure 10.13a), causing the 5'-end to be more stable than the 3'-tail. A transition from these proposed unfolding intermediates (frayed and pacman) to the completely unfolded random coil is expected to undergo an energy barrier crossing which may involve additional hairpin intermediates (Figure 10.12a). With these different possible unfolding scenarios, we discuss the most likely pathway that is consistent with our T-jump results.

10.4.1 Nanosecond response

A pulse-width-limited <5ns response is observed in all the samples including the free GMP nucleotides (Figure 10.5). This is attributed to the heating of the solvent that weakens the water H-bond network which modifies the interactions with the bases and generally leads to changes in water IR absorption. The t-DPP spectra of GMP (Appendix 10.B) agrees well with the equilibrium thermal difference spectra (Figure 10.6a), providing evidence to support this assignment. This phenomenon is generally observed in T-jump experiments on small aqueous solutes.27,28

After the response from local heating, a ~40 ns spectral response is observed for TBA and tTBA in D2O, but not for TBA in the presence of K+. The t-2D IR spectra (Figure 10.7) do not show a loss of the (ωL=1535 cm⁻¹, ωS=1670 cm⁻¹) cross-peak even after 50 milliseconds, indicating that the overall structure remains intact as a G-quadruplex or triplex. The guanine ring mode red shifts while preserving its sharp and narrow lineshape in the 56 ns t-2D IR spectrum, suggesting that the ~40 ns response corresponds to weakening of the base-pair H-bonds for the guanine. Loosening of the base-stacking is expected to accompany the local fluctuation of the H-bond network. However, the lack of significant increase in ring-mode increase indicates only a small perturbation to the base-stacking. These
interactions are stabilized by the potassium ion in the center of TBA, and therefore this response was not observed in the spectra for TBA with K⁺.

Figure 10.12: (a) Possible unfolding pathways for TBA. Red arrow indicates the most likely pathway. (b) Proposed free energy surfaces for TBA unfolding based on our T-jump results.

The lack of apparent temperature dependence of the rate constant also suggests that there is no large-scale disruption of the structure on this timescale, which is consistent with the assignment of local base-stacking loosening, base-paring fluctuations, or unsuccessful attempts at fraying the 3'-end (fast tail opening and closing without being stabilized). The 40 ns timescale is also consistent with the relaxation time constants observed for base-stacking of single stranded oligonucleotides (τ = 100-300 ns).

In the case of single-stranded DNA, which lacks backbone stabilization, base destacking leads to fast helix-to-coil transition. In the case of TBA, the Hoogsteen H-bond network stabilizes the overall structure even in the absence of K⁺, allowing for local fluctuations only. Moreover, the timescale of adding a base-pair onto the helix end has been determined to be also ~100 ns, which suggests that transient 3'-tail opening and closing can occur during this kinetic phase.
10.4.2 Microsecond response

Since the relaxation time constants for single stranded oligonucleotide unstacking were found to be sub-\(\mu s\),\textsuperscript{29,31} the slower few-\(\mu s\) response seen in our data has to involve larger conformational changes. The unfolding of peptide hairpins has been observed in this time window.\textsuperscript{5,27,32} For example, loop rearrangement whose reaction rate could be slowed down by the steric constraints imposed by the loop structure. The \(\mu s\) response vanishes for tTBA, indicating the important role of the 3\textquotesingle-tail of TBA in this response. We propose that the \(\mu s\) response corresponds to the formation of the fully solvent-exposed 3\textquotesingle-tail and the structural reorganization of the remaining chair to form a stable triplex. The structural rearrangement can involve both the loops and the guanine core which may become more compact in the triplex form.

Figure 10.13: TBA structures from PDB: 148D. (a) Top view showing the base-stacking of G8 (grey) and G1 (red). The top layer G-quartet is also shown. (b) Side view showing the thymine bases.

In t-DPP spectra of TBA in D₂O, the loss feature at 1546 cm\(^{-1}\) shifts to 1558 cm\(^{-1}\) on the \(\mu s\) timescale, matching the loss feature of the free GMP spectra. It is suggested that the 1546 cm\(^{-1}\) peak corresponds to weakening the base-stacking of the two G-quartets, while the 1558 cm\(^{-1}\) peak is attributed to the response of the free guanine bases G14 and G15 after the 3\textquotesingle-tail is extended into the solvent. In the t-DPP spectra of TBA with K\(^+\), only a loss feature at 1546 cm\(^{-1}\) is observed, which supports its assignment to the weakening of the G-quartet stacking. The 3.2 \(\mu s\) t-2D IR spectrum (Figure 10.7b) shows significant enhancement of the
guanine ring mode intensity, suggesting base unstacking. However, the ring mode lineshapes remain sharp, indicating that the structure is still ordered. TBA only has one major base stack between the two layers of G-quartet, and therefore the observed response must involve certain degree of unstacking between these two layers.

The C=O stretch region of the transient spectra is more intriguing. Since the equilibrium spectra of tTBA is strikingly similar to those of TBA, it is expected that the guanine C=O does not contribute significantly to the observed signal even if the 3'-tail frays. Nevertheless, the loss of eight H-bonds during fraying should broaden the guanine C=O peak. In fact, as discussed earlier, the C=O response comes mainly from the thymine bases, which indicates the changes to the solvation around thymine carbonyls. A closer examination of the solution NMR structures of TBA shows that some thymine bases are solvent exposed, such as T3 and T7 (Figure 10.13b). Significant fluctuations are expected for these bases, especially when the 3'-tail frays. These structural changes are expected to result in complicated lineshape changes and frequency shifts to the thymine C=O.

We propose a free energy landscape for TBA unfolding that is consistent with our data, shown in Figure 10.12b. Before the T-jump at $T < T_m$, the most stable conformation of TBA is the chair-like G-quadruplex, which is separated by energy barriers from the triplex and random coil states. The abrupt T-jump shifts the energy barriers toward a less stable native chair form and a more stable coil form. The increasing temperature at $T > T_m$ increases the configurational flexibility. The distribution of structures in the native well exhibits a ~40 ns relaxation with weakening of the base-stacking and Hoogsteen H-bonds. Since the energy barrier is now lowered, the sub-ensemble at or near the transition state (the deformed G-quadruplexes) undergoes a few-μs fraying of the 3'-tail. G15 unstacks from G14 and the Hoogsteen H-bonds to G14 and G15 are broken, causing the 3'-tail to fray into the solvent. The remaining chair structure undergoes a large motion of rearrangement such as loop deformation and core tightening to stabilize the triplex structure.

Our T-jump experiments probe the early unfolding transition to the triplex structure, which involves the deformed G-quadruplex and 3'-frayed triplex. We expect that the actual breaking of the G-quadruplex core would be observed on timescales from ms to seconds, as reported for the unfolding of various G-quadruplexes.\textsuperscript{26,33-35} Currently the 1-100 milliseconds window is not accessible by our instruments.
10.4.3 Isotope labeled TBA

Experiments on isotope labeled TBA can further confirm the assignment of the T-jump response and obtain more detailed insight into the mechanism. Figure 10.14a shows the FTIR spectra of unlabeled GMP and uniform $^{13}$C$_{10}$ labeled GTP (Sigma-Aldrich). The C=O peak is red shifted from 1662 cm$^{-1}$ to 1616 cm$^{-1}$, the ring mode G2 is red shifted from 1579 cm$^{-1}$ to 1538 cm$^{-1}$, and the ring mode G4 is red shifted from 1538 cm$^{-1}$ to 1499 cm$^{-1}$.

Since $^{13}$C$_{10}$/$^{15}$N$_5$ isotope labeled guanosine completely shifts the guanine ring modes away, isotope labeling the guanine bases at the 3'-end, G14 and G15, should allow the assignment of the two distinct kinetic phases observed in the T-jump experiments. For example, if the μs response corresponds to the fraying of the tail and the reorganization of the triplex structure, the enhancement of the 1499 cm$^{-1}$ peak and its cross-peaks to the C=O stretch should disappear on the μs timescale as the G14 and G15 bases become solvent-exposed. Figure 10.14b shows the FTIR spectra of unlabeled TBA and $^{13}$C$_{10}$/$^{15}$N$_5$ isotope labeled TBA on G14 base. Significant spectral changes occur for the ring modes; however, the redshifted ring mode at 1515 cm$^{-1}$, for example, does not align with the 1499 cm$^{-1}$ seen in the labeled GTP. This indicates that the vibrational coupling within the G-tetrad is perturbed. Further experiments are needed to properly assign the spectra.

Figure 10.14: (a) FTIR of unlabeled GMP and $^{13}$C$_{10}$ labeled GTP. (b) FTIR of unlabeled TBA and G14 labeled TBA. The isotope label is $^{13}$C$_{10}$/$^{15}$N$_5$. 

313
10.5 Melting of TBA-antidote duplex

Having studied the early unfolding events of TBA, we moved on to investigate the binding interaction of TBA with its complementary strand (abbreviated as cTBA, 5'-CCAACCACACCAACC-3'). The binding between TBA and cTBA leads to the formation of a DNA double helix with 15 base-pairs that include both AT and CG base-pairs. The TBA-cTBA interaction is stronger than that between TBA and thrombin, so that the presence of cTBA acts to separate TBA from thrombin. cTBA acts as an antidote to TBAs inhibition of thrombin activity. We seek to understand the interplay between events of TBA unfolding, release from thrombin, and duplex formation requires knowing the timescales of these events. As a step in this direction, we applied T-jump 2D IR spectroscopy to study the TBA-cTBA association and dissociation kinetics.

10.5.1 Equilibrium temperature dependence

Figure 10.15a shows the temperature-dependent FTIR spectra of equimolar mixture of TBA with its complementary strand, cTBA, in 50 mM potassium phosphate buffer. The low temperature spectrum exhibits spectral features associated with both AT and GC base-pairs (see Chapter 8). Briefly, AT base-pairs result in a blue shift and intensification of the T C=O mode to 1695 cm⁻¹, a blue shift and narrowing of the T ring mode to 1641 cm⁻¹, and a red shift and intensity loss of the A ring mode at 1621 cm⁻¹. GC base-pairs cause significant blue shift of the G C=O from 1662 to 1689 cm⁻¹, while the C C=O remain at 1651 cm⁻¹. Additionally, the G ring modes ~1560–1580 cm⁻¹ experience intensity suppression, but not as much as for G-quadruplex formation. As the temperature is raised, these spectral features disappear and the high temperature spectrum resembles the free nucleotide spectrum.
At low temperature, the lack of the 1535 cm$^{-1}$ mode in the FTIR spectrum and the lack of 2D IR cross-peaks at (1535 cm$^{-1}$, 1670 cm$^{-1}$) indicate that in the presence of the complementary strand, the TBA-cTBA duplex formation is more energetically favorable than G-quadruplex, in agreement with previous study using FRET.$^{36}$ This result also validates the applicability of using the complementary strand as an antidote for TBA. The thermal difference FTIR spectra (Figure 10.15b) show spectral changes that correspond to the melting of the WC duplex. The increased absorbance at 1621 and 1575 cm$^{-1}$ are due to the thymine and guanine ring modes, respectively; while the 1521 and 1503 cm$^{-1}$ modes are the cytosine ring modes. The assignment of the enhanced signal at 1659 cm$^{-1}$ is ambiguous, as the C=O vibrations of free G, T, and C all exhibit peaks in this spectral region. The loss feature around 1680-1700 cm$^{-1}$ is attributed to breaking of base-paired G and T C=O modes.

By monitoring the three main gain features at 1659, 1621, and 1575 cm$^{-1}$, we can determine the duplex melting temperature and thermodynamics, following Equations (8.1)-(8.7). Briefly, the reaction scheme for dissociating the DNA double-helix (D) into TBA (M$_1$) and cTBA (M$_2$) is
\[
D \rightleftharpoons M_1 + M_2
\]  \hspace{1cm} (10.1)

\[
K_d = \frac{[M_1][M_2]}{[D]} = \exp\left(-\frac{\Delta G^o}{RT}\right)
\]  \hspace{1cm} (10.2)

The melting temperature \( T_m \) is defined as the temperature when the duplex fraction \( f_D \) reaches 0.5. The dimer fraction is assumed to be proportional to the second SVD component for the temperature-dependent FTIR spectra, and can be obtained by fitting to the following model:

\[
\Delta G^o = \Delta H^o + \Delta C_p \left[ T - T_m - T \ln \left( \frac{T}{T_m} \right) \right] - T \Delta S^o
\]  \hspace{1cm} (10.3)

If \( f_D = 0.5 \) at \( T_m \), the following relation must be satisfied,

\[
\Delta S^o = \frac{\Delta H^o + RT_m \ln \left( C_{\text{total}}/4 \right)}{T_m}
\]  \hspace{1cm} (10.4)

where \( C_{\text{total}} \) is the total DNA monomer concentration, in this case 2 mM. It should be pointed out that \( \Delta G^o \) may not be necessarily zero at \( T_m \), and it actually depends on \( C_{\text{total}} \). A self-consistent analysis of the thermodynamic parameters describing the thermodynamics and melting temperature, shown as the red curve in Figure 10.15c, leads to \( T_m = 66 \) °C, \( \Delta H_{\text{diss}} = 154 \) kcal/mol, and \( \Delta C_p = 0 \) kcal/mol·K.

The 2D IR spectra for the TBA-cTBA mixture at 20 °C is shown in Figure 10.16b. As discussed in Chapter 8, the 2D IR cross-peaks between vibrations of different bases provide clear evidence for the base-pairing between oligomers. For example, cross-peaks can be observed between the C=O stretches of G and C (purple arrow), between the G ring mode at 1575 cm\(^{-1}\) and C C=O stretch at 1651 cm\(^{-1}\) (green arrow), and between the A ring mode at 1621 cm\(^{-1}\) and the T C=O at 1661 cm\(^{-1}\) (black arrow). Upon increasing temperature, the peaks broaden and shift to where the free base vibrations are (Figure 10.16d).
10.5.2 T-jump transient 2D IR spectroscopy of duplex melting

T-jump transient HDVE experiments to probe the melting kinetics of the TBA-antidote duplex are presented in Figure 10.17. Figure 10.17a shows the t-HDVE spectra for the duplex following a T-jump from 62-72 °C. In this case the t-HDVE signal-to-noise ratio is sufficient to analyze the t-HDVE data without the aid of t-DPP. Following the pulse-width-limited response that is attributed to solvent heating, the earliest T-jump delay at which significant changes are observed is approximately 1 μs. At this delay, gain features are observed from the C=O stretch at 1659 cm⁻¹ and the ring modes of both G and A between 1550-1600 cm⁻¹. A loss feature is seen at 1647 cm⁻¹, which can be assigned to the loss of H-bonded T ring mode. Overall, the t-HDVE spectra resemble the thermal difference spectra shown in Figure 10.15b. As the T-jump delay increases, these spectral features become enhanced. Since only amplitude changes are observed, it is suggested that the population of one species increases at the expense of another. Consequently we can assign the T-jump response to breaking of the base-pairs in the TBA-antidote duplex. However, to what extent the duplex is unzipped is not immediately clear from the spectra.

The kinetic traces at four different wavelengths are plotted in Figure 10.17b in the range of 1 μs to 1 ms. Within this range one kinetic phase is observed, with the signal rising on the hundreds of μs timescale. These μs time traces can be fit well with a single-exponential decay (Table 10.2). At a T-jump of 62-72 °C, the time evolution of transient
signal measured at these four wavelengths are invariant within the experimental noise with a time constant of ~250 µs. A single relaxation time is consistent with two-state kinetics.

We varied the T-jump temperature ranges and in all cases we observed single-exponential behavior in the µs time window (Figure 10.18a and Table 10.2). The transient spectra are also consistent across all temperature ranges studied, indicating that the same melting behavior is being monitored. At initial temperatures lower than the $f_0 = 0.5$ transition temperature ($T_m = 66.4°C$), there seems to be a separation in timescales for the different probes, which is suggestive of heterogeneous unzipping, although the error bars for this measurement are large. In general the guanine ring mode at $1562 \text{ cm}^{-1}$ has faster response than the higher frequency modes. This may suggest that the melting of the DNA duplex occurs with destacking of the bases before breaking of the H-bonds. Similar trends have been observed in T-jump FTIR experiments of small RNA hairpins. At temperature higher than $T_m$, no wavelength dependence is observed for the time constants.

According to the cooperative zipping model, nucleic acid double helix formation consists of two distinct steps: (1) two single stranded DNA diffusing in solution find each other and form enough base pairs to nucleate helix formation, and (2) the helix grows with sequential addition of base-pairs. The first process, nucleation of the initial contacts, has been described through a transition state consisting of a few base-pairs (three or four for oligomers shorter than 20 bp) can grow as fast as it dissociates. Pöschke and Craig et al.

![Figure 10.17](image1.png)

Figure 10.17: (a) Transient HDVE spectra for the equimolar mixture of TBA and antidote, at selected T-jump delays. The initial temperature was 62 °C. (b) The time dependent traces at four wavelengths which are marked by black dash line in (a). The black solid lines represent single exponential fit.
used UV absorption spectroscopy coupled to T-jump to study this process in oligo(A)-oligo(U) and reported a relaxation time constant varying from 100 μs to 1 sec depending on conditions. This relaxation time constant depended on both the DNA concentration (shorter with higher concentration) and temperature (shorter with higher temperature).

Table 10.2. Summary of relaxation kinetics for 1 mM mixture of TBA and antidote. The error bars were obtained from the 95% confidence level of the single exponential fit.

<table>
<thead>
<tr>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_1 (°C)</td>
</tr>
<tr>
<td>56</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>62</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>74</td>
</tr>
</tbody>
</table>

The second step in duplex formation is the addition of single base-pair next to the helix end. This is a relatively fast reaction and can be experimentally probed when the chain length is long, or when the temperatures is below the melting temperature such only a small population is in the partially melted state. A T-jump experiment performed in this range will break a few base-pairs at the helix end, which is called an unzipping reaction. Signal contributed by the “all or none” helix-coil transition involving the entire duplex should be small in this case. Pörschke\textsuperscript{14} reported that the relaxation time constant of unzipping for oligo(A)-oligo(U) with chain lengths of 14 and 18 base-pairs is in the range of 100-300 ns. The rate of adding a base-pair was calculated to be 8x10^6 sec\textsuperscript{-1}, with an activation energy of 4 kcal/mol. This rate constant does not depend on the oligomer concentration.

319
On the basis of these studies, we believe the hundreds of \( \mu \)s kinetic phase observed in our experiment corresponds to the “all or none” helix-coil transition. The timescale falls in the range of 100 \( \mu \)s to 1 sec, and is much slower than the sub-\( \mu \)s timescale for unzipping. Moreover, significant fraction of duplex is in the dissociated state at the final temperatures in our experiments, which would not favor the unzipping reaction. Since the reciprocal of the relaxation time constant for the helix-coil transition is linearly proportional to the total DNA concentration, a concentration dependence in the future would help elucidate the assignment.

The reaction scheme for dissociating the DNA double-helix (D) into TBA (M₁) and cTBA (M₂) is:

\[
D \xrightleftharpoons[k_u]{k_y} M_1 + M_2
\]  

(10.5)
Solving the kinetic scheme for this reaction, the observed T-jump relaxation time constant is expressed as:

\[
\frac{1}{\tau_{\text{obs}}} = k_d + k_a ([M_1] + [M_2]) \tag{10.6}
\]

Since in our experiments, the final temperatures after the T-jump are higher than \(T_m\), the equilibrium is shifted towards the dissociated form. Therefore, the dissociation rate \(k_d\) dominates the observed relaxation kinetics. This means that \(k_d\) will share similar temperature trend as \(\tau_{\text{obs}}^{-1}\): increasing temperature results in faster dissociation, indicating a smaller activation barrier for dissociation. This can be interpreted as the destabilization of both the transition state and the duplex state due to larger configurational entropy cost at higher temperature. Using an Arrhenius fit to the last four temperature data points, an activation energy of \(\sim38\) kcal/mol was obtained, which is consistent with the literature values.\(^{15}\)

### 10.6 Conclusion and outlook

In summary, we have performed T-jump 2D IR experiments on TBA to measure the kinetics of its early unfolding event. With TBA in D\(_2\)O, we observed two kinetic phases on \(\sim40\) ns and \(\sim1\) \(\mu\)s timescales. The faster kinetic phase vanishes in the presence of stabilizing metal ions, and the slower kinetic phase was not observed in tTBA nor TBA in the presence of K\(^+\). From transient 2D IR spectra, it is suggested that the \(\sim40\) ns phase corresponds to local weakening of both the H-bond network and the base-stacking of the G-tetrads, and that the \(\sim1\) \(\mu\)s phase report on the fraying of the 3'-tail and the structural rearrangement to stabilize the triplex conformation. We have also studied the melting kinetics of TBA-antidote duplex and found the relaxation time constants to be hundreds of \(\mu\)s.

Although the kinetics of short DNA double helix formation and unzipping have been studied with T-jump UV absorption four decades ago, a detailed picture of conformational changes during the reaction has not been obtained due to the lack of structural sensitivity of the methods applied. The experiments presented in this chapter are the first T-jump transient 2D IR experiments on DNA oligomers that establish a framework upon which numerous future studies can be conducted. They have shown great promise of tracking the structural changes by monitoring the different base vibrations. In addition, these experiments have revealed that the DNA folding and binding are not simple two-state systems, and a rough energy surface is required to describe the processes. With the
advancement of theoretical modeling of the IR spectra of nucleic acids, we believe that transient 2D IR will become a powerful tool to study fast conformational changes of nucleic acids, especially for RNA, which exhibit faster and more fluctuating dynamics. A more ambitious goal of studying the dynamics of protein-nucleic acid interactions may even be possible with strategic incorporation of isotope labels.

10.7 Acknowledgments

I thank Mike Reppert for performing the MD simulations of TBA which provide valuable information on the frayed TBA structure.
10.8 References


Appendix 10.A: Linear projections of the 2D IR spectra

Figure 10.A.1: (Left) 1st (black dash) and 2nd (red solid) SVD components of the 2D IR derived squared root of DVE spectra. (Right) 1st (black dash) and 2nd (red solid) SVD components of the 2D IR derived pump-probe spectra. These spectra are scaled for better comparison.
Appendix 10.B: T-jump transient spectra

Figure 10.B.1: Transient t-HDVE and t-DPP spectra for TBA, tTBA, and GMP.
Chunte (Sam) Peng

929 E. 57th Street, GCIS E001A
Chicago, IL 60637
(773)702-4969
E-mail: sampeng@mit.edu

EDUCATION

2008 - 2014 Massachusetts Institute of Technology
PhD, Physical Chemistry (GPA: 5.00/5.00)

2004 - 2008 University of California at Berkeley
BS, Chemistry (Highest Honors, GPA: 3.99/4.00)

RESEARCH EXPERIENCE

2013/09 – present • Department of Chemistry, University of Chicago

2008/11- 2013/08 • Department of Chemistry, Massachusetts Institute of Technology
• Professor Andrei Tokmakoff
• Two-dimensional infrared spectroscopy of nucleic acids: application to tautomerism and DNA aptamer unfolding dynamics

2006/07- 2008/06 • Chemical Science Division, Lawrence Berkeley National Laboratory
• Department of Chemistry, University of California at Berkeley
• Professor Stephen R. Leone
• High harmonic generation and ultrafast dynamics in Helium nanodroplets

TEACHING EXPERIENCE

• Principle of Chemical Science, HHMI TA Fellow, MIT, Fall 2008
• Thermodynamics and Kinetics, Teaching Assistant, MIT, Fall 2009
• Physical Chemistry Laboratory, Teaching Assistant, MIT, Spring 2009
• Biophysical Chemistry, Grader, MIT, Spring 2011

HONORS & AWARDS

2012 Morse Travel Grant, MIT
2010 Department of Chemistry Awards for Outstanding Teaching, MIT
2008-2009 Presidential Fellowship, MIT
2008 Erich O. & Elly M. Saegbarth Prize in Chemistry, UCB
2004-2007 Honor Student of the Department of Chemistry
2007  Melvin J. Heger-Horst Scholarship
2007  Bruce Howard Memorial Scholarship
2006  College of Chemistry Undergraduate Scholarship Fund
2005  The Edward Frank Kraft Scholarship Prize
2005  Jack and Avis Olsen Undergraduate College of Chemistry Scholarship
2005  Laskey, N. Scholarship

PUBLICATIONS


TALKS
1. 2nd MIT Biophysics Retreat, September 2012; North Falmouth, MA.
3. Seminar at Research Center for Applied Sciences at Academia Sinica, June 2013; Taipei, Taiwan.
4. Seminar at National Sun Yat-sen University, June 2013; Kaosiung, Taiwan.
5. 58th Biophysical Society Meeting, February 2014; San Francisco, CA.

POSTERS

PROFESSIONAL ACTIVITIES
- Biophysical Society
- American Physical Society
- Alpha-Chi-Sigma: Professional Chemistry Fraternity

COMMUNITY SERVICES
- Vice President, MIT Taiwanese Student Association (ROCSA), 2010-2011
- Co-Chair, MIT ROCSA Connect Program, 2010-2011