THE UNIVERSITY OF CHICAGO

DISRUPTIONS TO THE COOPERATIVITY OF DNA BASE PAIRING AND HYBRIDIZATION PROBED WITH TEMPERATURE-JUMP INFRARED SPECTROSCOPY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

BRENNAN GEORGE ASHWOOD

CHICAGO, ILLINOIS

DECEMBER 2023

Contents

Chapter 2

Chapter 3

List of Figures

List of Tables

Acknowledgments

My journey to graduate school required a heavy dose of luck, and the same good fortune continued immediately upon joining the Tokmakoff group. I was welcomed by an extraordinary collection of senior graduate students and postdocs who maintained a thoughtful and rigorous scientific community and helped make science fun throughout my early years of graduate school. Andrei's scientific curiosity and calm demeanor were central to this community, and I would like to thank him for his wonderful mentorship over the years. From the beginning, Andrei has given me the freedom to pursue new directions, make plenty of mistakes, and learn from them to become a better scientist. Along the way, Andrei has always kept an open door for discussion and showed incredible patience in the development of my scientific, writing, and presentation skills. I am always amazed by how he strives to distill a complicated problem into a focused message, and this mentality has shaped my own efforts toward scientific presentations. Lastly, it is clear that Andrei is invested in the success of all group members after they leave the group, and I greatly appreciate his wisdom and honest feedback in helping me choose my next scientific steps.

Paul Sanstead was the best graduate-student mentor I could have asked for. Even as he was preparing his thesis, he always had time for questions and taught me much of what I know about nucleic acids and how to align the pulsed T-jump spectrometer. Whether DNA or insulin, Paul always obtained pristine data and performed careful analysis that set a high bar for all future big lab members. In addition to his skills as a scientist, Paul was an ideal group member and demonstrated the true patience of a saint during moments of adversity. He was always there to help someone in lab, provide feedback on a manuscript, or debate me on politics as we changed the

flashlamps. I will always look up to Paul for these qualities and thank him for his mentorship and collaboration during our overlap in the group.

Several other former and current group members shaped my development during graduate school. Among all, Nick Lewis likely had the largest impact. Nick was the first group member to teach me about 2D IR spectroscopy and later taught me the skills to build a spectrometer from scratch. He has continued to be an influential mentor and friend throughout my PhD, and I consistently admire the level of science he brings to the group. I always asked "What would Nick do?" when faced with a challenge in the lab, and I will continue to do so throughout my scientific career. Lukas Whaley-Mayda's beautiful BGMs, insightful questions, and in-depth discussions usually left me in awe or deeply confused, which always pushed me to improve my work. I also thank him for teaching me much of what I know about single-molecule spectroscopy, the intricacies of colormaps, and the good things about California. However, most impactful was Lukas' insistence for me to drink Rishi's Earl Grey Supreme tea during the first year, as 2+ cups per day have powered my last five years of research.

Yumin Lee joined the big lab as a postdoc at the start of my fourth year and has since been an invaluable colleague and one of my closest friends. Yumin's tremendous work ethic, high standard for research, and humble personality immediately promoted easy and productive collaboration between the two of us. Over this time, she has taught me a great deal about ultrafast spectroscopy and protein biophysics. I've also enjoyed learning about the unique qualities of Philly and Korea from her, as well as going on trips to try new food and tea throughout Chicago. I've recently been impressed by Yumin's relentless effort to express and purify isotope-labeled protein samples, especially given the absence of such knowledge in our group, and I am excited to see all that she accomplishes in her current work and future scientific endeavors.

Memo Carpenter was a constant source of energy and joy in the group and served as a terrific role model for the junior students. As my host student during the recruitment weekend, Memo's caring personality and clear passion for research amplified my interest in the Tokmakoff group, and I am grateful for our few years of overlap. Chi-Jui Feng taught me a lot about proteins, computational biophysics, and Mario tennis. He helped me through several research challenges early on and always provided encouragement and constructive feedback for other group members. Sam Penwell's talent and dedication to building new experiments were inspirational to all members of the group, and I appreciate the feedback he provided throughout my first two years. Although I overlapped with Joe Fournier for less than a year, he immediately welcomed me into the ball-toss circle and group. His 2017 BGM was the first group meeting I ever attended, and I remember being mesmerized by his remarkably crafted story, yet spooked by the hostility surrounding the structure of the aqueous proton. Tanya Shpigel was exceptionally efficient in handling just about any group matter and made life much easier during the first half of my PhD.

I had the pleasure of joining the group together with John Hack and Ram Itani. I admire John's careful approach to spectroscopy and have learned a lot from his discoveries and our discussions over the years. Ram was tasked with a difficult project for which our group had no background, and his perseverance and positive attitude through it all have been impressive and often a source of liveliness in the group. Abhirup Guha joined a year after me and has consistently demonstrated a careful and humble approach to research. He has taught me a lot about microscopy and FEIR, and I look forward to seeing the outcome of his current projects. Seung Yeon Lee has shown impressive talent and hard work in the lab, and I anticipate she will continue to push the FEIR lab to new heights. I've been lucky to have Melissa Bodine as a desk neighbor for the past three years. I could always count on a pointed joke to start the day (even though it usually went

over my head), and I thank Melissa for consistently being an uplifting spirit in the office. Jakob Schauss, Sam Knight, Liv Mumma, Rowan Simonet, and Ian Bongalonta all joined the group over the last couple of years, and I am excited to see where they take big lab and water lab in the near future.

I have been lucky to collaborate with multiple research groups at the University of Chicago. I've lost track of the number of projects Mike Jones, Prof. Andy Ferguson, and I have worked on together. Mike's efficiency, dedication, and flexibility opened many doors for our research, and along the way, he taught me much of what I know about computational biophysics. I thank Andy for always providing rapid, thoughtful feedback on our research and taking the time to help me decide my next steps after graduate school. Prof. Jack Szostak's group moved to the University of Chicago within the last two years of my PhD, opening several new pathways for nucleic acid research in the Tokmakoff group. Jack always found time to discuss with me and inspired our work on dinucleotide hybridization (Chapter 9). Even though Aleks Radakovic was busy setting up Jack's new labspace, he made time to teach me basic nucleic acid biochemistry and synthesized RNA samples for our work. I never directly collaborated with Marco Todisco, but that didn't stop us from having several thoughtful discussions on nucleic acid hybridization and non-enzymatic extension. I would like to thank Prof. Qing Dai and Prof. Chuan He for their collaboration with Paul and I on the modified cytosine work described in Chapter 8. I thank Joe Sachleben for teaching me about NMR spectroscopy of biomolecules and helping me perform the NMR experiments in this thesis. Although I never scientifically collaborated with Prof. Ka Yee Lee, I thank her for being part of my thesis committee and offering constructive feedback over the years.

I only made it to graduate school through the opportunities and trust provided by Marvin Pollum, Prof. Carlos Crespo-Hernández at Case Western Reserve University, and Prof. Geoff

Blake at the California Institute of Technology. Marvin was my graduate student mentor in the Crespo group and demonstrated an incredible level of patience with me. His willingness to teach me about transient absorption, photophysics, and scientific writing was the primary reason for my progress as an undergraduate and eventually the ability to attend graduate school. Carlos is an exceptional mentor and dedicated much of his time to improve my research and writing skills. His honest and constructive criticism prepared me well for graduate research. Geoff allowed me to work in his lab for a summer as an undergraduate, and this is where I first became aware of multi-dimensional spectroscopy and the Tokmakoff group. I thank his former students, particularly Ian Finneran, Brandon Carroll, Jacob Good, and Brett McGuire for taking time to discuss with me.

I thank my wife Rachel for being my best friend and a constant source of support, honesty, and fun. Whether a new board game, show, or speaking proper English, she always teaches me new things and helped push me in the right direction at my lowest points during graduate school. Rachel also turned me into a cat person. Nadine and Buddy's welcome home at night was always something to look forward to and their forceful wakeup in the morning kept me in line. Lastly, I would like to thank my parents and brother for supporting my scientific trajectory.

Funding

I am grateful to have received funding for my PhD through the National Science Foundation graduate research fellowship program (NSFGRFP) and the Seymour Goodman research fellowship award from the Department of Chemistry at the University of Chicago. I would also like to acknowledge the National Institute of General Medical Sciences of the National Institutes of Health (R01-GM118774) and the National Science Foundation (CHE-2155027) for funding the majority of the research described in this thesis.

Abstract

Hybridization of DNA strands and motions of the DNA duplex play fundamental roles in biology and the design of nanotechnology and biotechnology applications, and our understanding in these fields is rooted in the thermodynamic and kinetic models of hybridization developed throughout the twentieth century. Many of these technologies and functions of DNA act in concert with chemical modifications to nucleobases or the backbone that alter the physical and biochemical properties of DNA. The mechanisms by which modifications alter hybridization and interactions within the duplex remain poorly understood and are often not captured by the established thermodynamic and kinetic models of hybridization. Interpreting the effects of modifications instead requires a predictive understanding of the series of molecular events during hybridization and structural changes within the duplex state, for which there is little direct insight even for unmodified DNA. This thesis describes efforts to characterize the molecular dynamics of hybridization and the duplex ensemble in DNA oligonucleotides and the corresponding effects from modifications and lesions. Our approach relies on the development and application of time-resolved infrared (IR) spectroscopy in combination with molecular dynamics (MD) simulations to directly interrogate and visualize motion along the free-energy landscape of hybridization with previously unachieved molecular detail. Chapters 2 – 5 introduce the benefits and drawbacks of IR spectroscopy and MD simulations for studying these problems.

Investigation of oligonucleotides containing an abasic site (AP site) demonstrates how the loss of a single nucleobase disrupts the cooperativity to hybridization and ultimately destabilizes the duplex. Temperature-jump (T-jump) IR experiments reveal that an AP site creates a free-energy barrier for forming base pairs on each side of the modified site, effectively breaking

the duplex into two segments with free-energy barriers for nucleation even though the average structure of the duplex is nearly unperturbed relative to canonical DNA. Coarse-grained MD simulations reveal that an AP site constrains the trajectories of hybridization events (i.e. transition paths) to follow a mechanism of nucleating and zippering a stretch of base pairs on one side of the AP site before nucleating and zippering the other segment. The added free-energy penalty for nucleating the second segment destabilizes the duplex beyond what is expected from the well-established nearest-neighbor thermodynamic models of DNA. Shifting the position of an AP site within an oligonucleotide moves the location of the second nucleation barrier along the free-energy landscape, tuning the destabilization and dynamics of the duplex. T-jump IR experiments reveal that the barrier induces fraying of the short base-pair segment on nanosecond timescales when the AP site is close to the termini. The short segment gains binding stability as the AP site moves away from the termini and eventually overcomes the nucleation barrier at length scales of 2 – 5 base pairs depending on the sequence. The fully encompassed nucleation barrier maximally destabilizes the duplex and creates metastable configurations with one segment dissociated. Nuclear magnetic resonance (NMR) and two-dimensional IR spectroscopy (2D IR) measurements reveal that certain sequences can circumvent the nucleation barrier by base pairing out-of-register.

Our study of cytosine and 5-carboxylcytosine (caC) protonation in oligonucleotides reveals that small chemical modifications can also disrupt the cooperativity of base pairing and hybridization. While N3 protonation of cytosine is found to inhibit duplex formation, protonation of the exocyclic carboxyl group of caC weakens base pairing and promotes internal melting of the duplex at temperatures below the full dissociation transition of the duplex. T-jump IR measurements show that caC protonation reduces the free-energy barrier to dissociation and promotes stretched-exponential relaxation, a signature of conformational or dynamic

heterogeneity. Deprotonated caC exhibits a minor impact on duplex stability and kinetics, therefore exocyclic protonation acts as a switch between standard and weakened base-pairing that may be relevant for the recognition and excision of caC.

Chapter 1

Towards the dynamics of nucleic acid hybridization

1.1 Introduction

Short DNA and RNA strands (< 200 nucleotides), known as oligonucleotides, have become foundational molecules for applications in biotechnology, nanotechnology, and imaging applications biology. CRISP-Cas gene editing uses a 20-mer guide RNA segment to recognize sections of the genome,[1] DNA nanostructures are self-assembled from collections of short oligonucleotides,[2] dynamic functions of nanostructures can be programmed with oligonucleotide hybridization,[3] and oligonucleotide hybridization be may utilized for super-resolution imaging with DNA-PAINT.[4] Fundamental physical properties of oligonucleotides and their folding and assembly underpin the design of these methods. For example, the thermodynamic and kinetic penalties of forming base-pair mismatches are a primary factor in the fidelity of target specificity for CRISPR-Cas gene editing.[5-7] In DNA nanostructure assembly, tunable nucleic acid structures larger than 100 nm-scales form through the binding of hundreds to thousands of oligonucleotides with a degree of cooperativity that depends on the desired structure.[8-12] With DNA-PAINT, the rate of oligonucleotide hybridization and dehybridization determine the imaging speed, prompting optimization based on repetitive nucleobase sequence designs.[13-14]

Nucleic acid technologies hinge on an ability to predict how the thermodynamics, kinetics, and dynamics of oligonucleotide hybridization vary with factors such as nucleobase sequence, length, environmental conditions, chemical modifications, and non-canonical base pairing.

Empirical nearest-neighbor thermodynamic models predict the free-energy change of hybridization (ΔG_h°) as a function of nucleobase with quantitative accuracy,[15-18] and these models have emerged as the most widely applied tool to predict duplex stability[19-22] while methods to predict hybridization kinetics are still developing.[23-27] Calculation of ΔG_h° alone is sufficient for certain applications,[28] yet it is a poor predictor of the configurational ensemble adopted by duplexes and single strands as well as the rate and mechanism of duplex hybridization that are foundational aspects of technologies like dynamic DNA nanostructures and DNA-PAINT.[3, 14] Numerous research groups recognized this gap in the understanding of oligonucleotides, prompting experimental and computational investigations over the past fifteen years that have illuminated how nucleobase sequence tunes the rate and mechanism of duplex hybridization as well as the molecular properties of the duplex ensemble.[25, 29-36] There are still key questions that remain unanswered concerning the dynamics of hybridization, here defined as the sequence of molecular events that take place when oligonucleotides transition between the single-strand and duplex ensembles. Given that these properties remain ambiguous for canonical oligonucleotides, the influence of chemical modifications, lesions, non-canonical base pairs, and changes in environmental conditions on hybridization dynamics are almost entirely limited to thermodynamic and kinetic descriptions.[37-41] With these dynamic properties still elusive, nucleic acid technologies have rapidly developed using design principles based exclusively on thermodynamic and kinetic principles of hybridization. Therefore, improvement in our understanding of hybridization and duplex dynamics may similarly promote the development of new technologies based on controlled manipulation of hybridization mechanism and pathways.

Hybridization of oligonucleotides requires coordination amongst numerous structural degrees of freedom, such as those that describe base pairing, stacking, and backbone configurations, which is achieved through cooperative interactions between residues. Rooted in the statistical thermodynamic helix-coil models of the 1950s, we refer to cooperativity in the context of duplex hybridization as the favorability for forming a base pair adjacent to another relative to nucleating a base pair in isolation. This cooperativity is largely mediated by stacking interactions and backbone constraints between adjacent sites that act to pre-organize formation of the next base pair. Cooperativity has been a core concept of nucleic acid hybridization from its earliest thermodynamic investigations and is also fundamental to the dynamics of hybridization and the structural dynamics of the duplex. Full duplex hybridization between complementary oligonucleotides arises as each intact base pair within the continuous stretch of contacts pre-organizes the next through nearest-neighbor interactions. Oligonucleotides spend nearly all of their time as dissociated single-strands or an almost fully intact duplex because cooperative interactions hinder anything in between. However, this cooperativity is not infinite, as oligonucleotide duplexes have been shown to exhibit a variety of local transitions away from their canonical form,[42-46] and new studies demonstrate that these properties of the duplex are highly sensitive to nucleobase sequence, chemical modifications, and the presence of lesions.[47-50] In addition to tuning the duplex ensemble, these factors likely influence the dynamics of hybridization located within the rare and rapid crossing events between the duplex and dissociated states. Such crossing events are invisible to nearly all experimental measurements and remain difficult to characterize with computation. Accessing and understanding how to tune such dynamics requires new approaches to probe and model hybridization dynamics.

In this chapter, we aim to give an overview of the current understanding of oligonucleotide hybridization kinetics and dynamics. We first summarize work that has explored the physical properties of duplex and single-strand ensembles. As a foundation for investigating transitions between ensembles, the following section summarizes thermodynamic and kinetic principles of oligonucleotide hybridization established over the past sixty years. We then highlight recent breakthrough studies that have enabled the first experimental and computational visualization of hybridization dynamics and conclude by briefly discussing current gaps in our understanding of hybridization and an outlook for future investigations.

1.2 Duplexes, single strands, and hybridization from 60 years of thermodynamics and kinetics

Oligonucleotide hybridization is a multidimensional process that involves changes in many structural degrees of freedom, yet hybridization is often described along a single reaction coordinate. The optimal coordinate may report on combinations of the degree of base pairing between strands, stacking within strands, or hydration structure around the nucleic acid, yet most experiments directly probe only one of these parameters. Therefore, the hybridization coordinate in Fig. 1.1 depicts an ambiguous collective variable that separates the single-strand and duplex ensembles and captures the free-energy barrier in between. The barrier region can generally be broken down into three sections: an encounter complex where oligonucleotides rapidly shuffle around each other while forming and breaking base-pair contacts, a transition-state ensemble, and energetically downhill formation of the remaining base pairs.

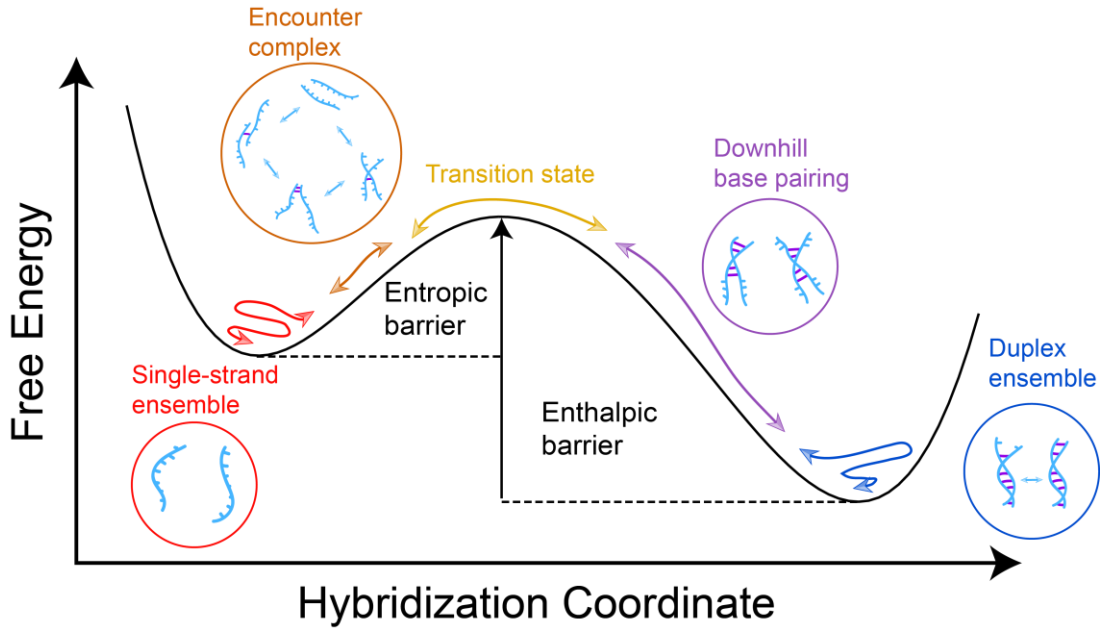

Figure 1.1 Overview of oligonucleotide hybridization on a two-state free-energy profile. Simple oligonucleotide systems may be separated into free-energy basins encompassing the single-strand and duplex ensembles. Unstructured single-strands must overcome an entropic barrier to hybridize that involves: (1) diffuse search between oligonucleotides, (2) an encounter complex where base-pair contacts are formed and broken until strands diffusive away or (3) until enough contacts are made to overcome the barrier and formation of the remaining base pairs is energetically downhill. In reverse, the free-energy barrier for dehybridization is dominated by the enthalpic penalty of breaking base-pair and stacking interactions yet slightly compensated by the favorable entropy change of dissociation.

1.2.1 Structural dynamics of the duplex

Although originally envisioned as a relatively static molecule, nucleic acid duplexes constantly undergo local and global structural motions that promote interactions with proteins and ligands. These motions span timescales from picoseconds to seconds and involve local fluctuations in base-pair orientation,[51-52] exchange of base-pair geometries,[42] and large-scale bending of the duplex.[53] We primarily focus on the dynamics of base-pairing interactions in this chapter. Advances in all-atom molecular dynamics (MD) simulations and experimental techniques over the

past twenty years have prompted several discoveries that reshaped our molecular interpretation of nucleic acid duplexes.[42, 53-57] As a result, the duplex ensemble is the most thoroughly characterized portion of the free-energy profile in Fig. 1.1.

The study of structural fluctuations in nucleic acid duplexes dates back more than 60 years to the concept of DNA "breathing" (Fig. 1.2a).[43, 58-59] Due to its structural ambiguity, "breathing" has taken on a multitude of meanings that generally refer to transient local disruptions of base pairs arising from thermal fluctuations.[43, 59-60] Evidence for "breathing" was first found in the 1960s by measuring the rates of hydrogen-tritium (H-T) exchange between duplex DNA and tritiated water (THO). The observation of complete exchange of labile nucleobase protons indicated that even the imino and amino hydrogens that participate in base pairs must become transiently accessible to exchange with tritium. Further evidence of transient base-pair openings came from experiments using formaldehyde, which hydroxymethylates amino and imino groups of nucleobases and disrupts base-pair interactions.[61-62] In the 1980s, the molecular detail accessible from exchange experiments greatly improved through combination with nuclear magnetic resonance (NMR) spectroscopy. NMR experiments monitoring hydrogen-deuterium (H-D) exchange at guanine and thymine imino sites enabled determination of exchange rates for individual base pairs in duplexes shorter than 30 nucleotides.[63-65] Two-state modeling of the exchange kinetics typically indicates opening on millisecond timescales and closing on nano-to-microsecond timescales in B-DNA and A-RNA duplexes, producing equilibrium constants for opening that ranged from 10^{-5} – 10^{-7} depending on the sequence context.[63, 66-67]

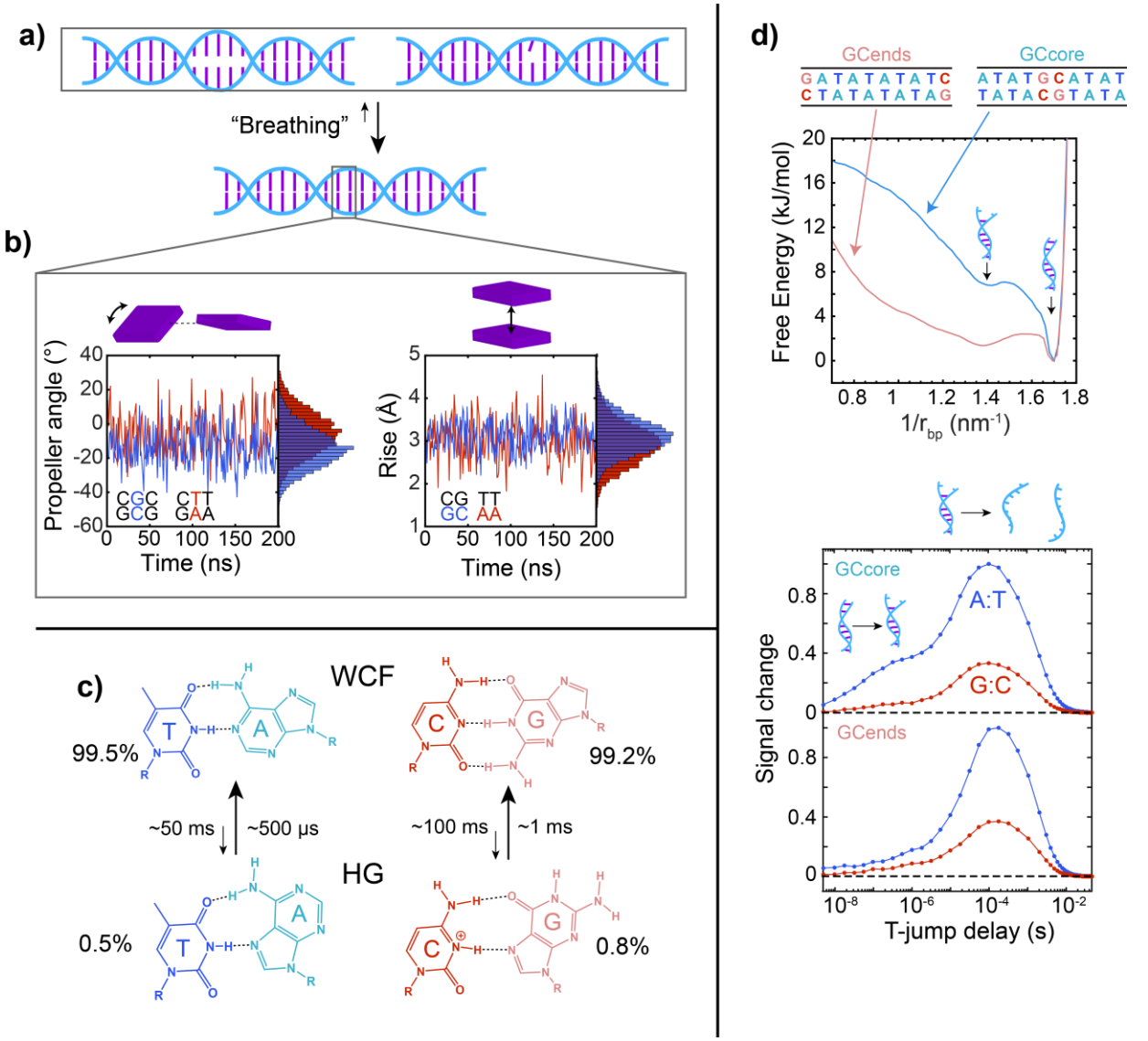

Figure 1.2 Structural motions in the duplex ensemble. (a) Schematic of base-pair "breathing" in a duplex where an individual or a region of base pairs transiently open to some degree. **(b)** Local fluctuations in base-pairing and stacking structure on pico-to-nanosecond timescales from all-atom MD simulations with the parmbsc1 force field. Trajectories for base-pair propeller twist angle and stacking rise are shown for G:C (blue) and A:T (red) sequence contexts. Histograms correspond to probability distributions from 1 µs of simulation time. **(c)** A:T and G:C base pairs in canonical DNA duplexes can exchange between Watson-Crick-Franklin (WCF) and lowly-populated Hoogsteen (HG) conformations on micro-to-millisecond timescales. Timescales and equilibrium populations are determined from NMR relaxation dispersion measurements.[36] **(d) (top)** Free-energy profiles for fraying of A:T (GCcore) and G:C (GCends) terminal base pairs calculated

Figure 1.2 Structural motions in the duplex ensemble, continued

from MD simulations with the 3SPN.2 coarse-grained model, highlighting the greater propensity for A:T fraying. Free energy is plotted as a function of the inverse average base-pair separation ($1/r_{bp}$) across the duplex where the free-energy minimum at ~1.7 nm^{-1} corresponds to a fully intact duplex while that at ~1.4 nm^{-1} corresponds to a duplex with one broken terminal base pair. **(bottom)** Experimental demonstration of high-amplitude A:T terminal fraying on 10 – 100 ns timescales with temperature-jump IR spectroscopy. The change in IR signal at adenine (blue) and guanine (red) vibrational bands is shown following a temperature jump across the duplex melting transition. GCcore exhibits kinetic components for A:T terminal fraying from 10 – 200 ns and full strand dissociation from 5 – 100 μs. GCends exhibits a low-amplitude kinetic component for G:C fraying from 0.2 – 1.0 μs and full-strand dissociation from 5 – 100 μs.

Even though imino H-D exchange measurements for individual base pairs are commonly interpreted to report on complete opening of Watson-Crick-Franklin (WCF) base pairs (or base flipping), the molecular characteristics of the "open" and "closed" states in these measurements are often ambiguous.[43, 59, 68-69] For example, all-atom MD simulations of duplexes indicate that "closed" base pairs constantly undergo fluctuations in base pair and stacking geometry on pico-to-nanosecond timescales that depend on sequence-context (Fig. 1.2b).[51-52, 54, 70] The "open" state corresponds to the ensemble of configurations that can undergo H-D exchange. But how open must a base pair be for H-D exchange to occur, and do multiple base pairs open cooperatively? Free-energy calculations of base flipping from all-atom MD simulations revealed that imino protons may undergo H-D exchange with base-pair opening angles as small as 30° and that the "open" state likely contains a large spread of opening angle values.[68-69, 71-72] Free energies for opening calculated using this cutoff show qualitative agreement with free energies for "opening" in H-D exchange measurements, supporting the assignment of "opening" in NMR experiments to flipping of one nucleobase by ≥30°.[45] The observation of distinct NMR H-D exchange rates for adjacent base pairs has led to suggestions that "breathing" occurs one base pair at a time rather than as a

cooperative unit at temperatures below the melting transition of the duplex.[59, 64, 73] However, H-D exchange timescales are an average over the ensemble of opening events at each site, and measurements that correlate opening events at adjacent sites are needed to evaluate cooperativity in base-pair opening. In contrast to NMR H-D exchange, measurements of duplex motions using single-molecule fluorescence spectroscopy reveal structural fluctuations on tens-to-hundreds of microsecond timescales that may correspond to transient opening of 2 – 10 base pairs (i.e. "bubbles"),[46, 74] and a full rationalization of these results with those from NMR H-D exchange has yet to be accomplished.

It was discovered roughly ten years ago that a portion of the "breathing" events observed in isotope-exchange and formaldehyde experiments correspond to transitions between WCF and Hoogsteen (HG)[75] base-pair geometries.[36, 42, 76] HG base pairs are created by rotating the purine base ~180° about the glycosidic bond to adopt a *syn* sugar-base orientation and form hydrogen bonds between 6 and 7 positions of the purine with the 4 and 3 positions of pyrimidine base, respectively (Fig. 1.2c). Using relaxation dispersion NMR spectroscopy (Chapter 2.3), Al-Hashimi and co-workers found that WCF base pairs in B-DNA spontaneously switch to HG conformations on tens of millisecond timescales via a free-energy barrier that requires flipping of the purine base into the major groove.[42, 77] The average lifetime of the HG conformation ranges from hundreds of microseconds to a couple of milliseconds depending on the sequence context.[36] The resulting HG equilibrium constants range from 10^{-2} to 10^{-1}, ~10^4-fold larger than the typical values obtained for base-pair opening, and sequence-dependent variation in the HG equilibrium constant is primarily set by the stability of the WCF base pair rather than the stability of the HG conformation.[36] Transient HG base pairs have not been observed in A-RNA experimentally and are estimated to

be disfavored by a free-energy penalty of 20 – 30 kJ/mol relative to B-DNA.[78] This energy penalty primarily arises from steric clashes of forming the *syn* orientation in the A-form duplex.

In addition to internal base-pair opening and HG formation, oligonucleotide duplexes may partially unwind from the terminal base pairs. These events are referred to as terminal fraying and are commonly observed in measurements of base-pair breaking and duplex dissociation.[79-82] Terminal fraying arises from the reduced stacking interactions at the terminus that give way to the entropic benefit of having a broken base pair with high configurational freedom. The lower enthalpic stability of base-pairing and stacking interactions in A:T base pairs makes them more likely to fray than G:C base pairs (Fig. 1.2d).[80, 83] Temperature-jump (T-jump) spectroscopy revealed that A:T base-pair fraying occurs on nanosecond timescales and is approximately 10-fold faster than fraying of G:C base pairs.[31-32, 80-81] However, the distribution of frayed configurations has yet to be determined directly from experiment and is instead estimated from modeling and MD simulations.[83-87] Statistical models are often used to calculate distributions of the number of frayed bases and predict that configurations with multiple frayed A:T base pairs can reach a population of 5 – 10% even at temperatures below the duplex melting transition.[83-84] All-atom MD simulations have been employed to probe the configurations of frayed states and pathways for fraying. An investigation of RNA oligonucleotides found that fraying preferentially initiates through unstacking of the 3′ nucleobase before the 5′ end.[86] These simulations, as well as others, also identify a population of noncanonical base-pairing geometries that play a significant role in the kinetics of fraying,[85-86] but it is unclear whether these configurations correspond to an artifact of current all-atom force fields for DNA and RNA.

1.2.2 Pre-melting of the duplex

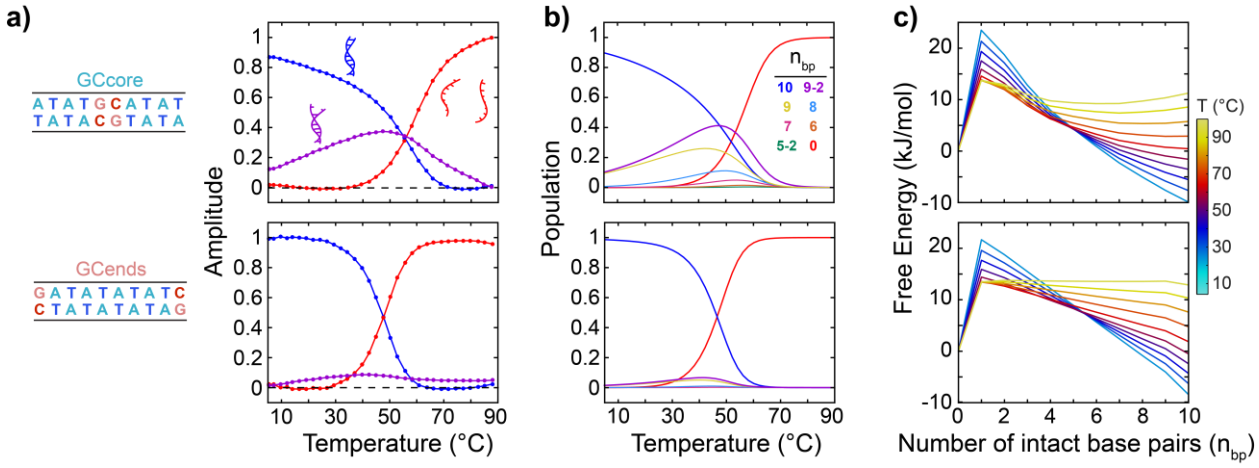

Figure 1.3 Pre-melting of duplexes through terminal fraying. (a) Three-component reconstructions of FTIR temperature series for GCcore and GCends, where the purple component corresponds to intermediate duplex configurations with frayed termini. The blue and red components correspond to the fully intact duplex and dissociated single strands, respectively. **(b)** Calculated populations of duplex species as a function of temperature using a lattice model. Different color curves correspond to the population of a species with a certain number of intact base pairs (n_{bp}), and the purple line is a sum over the population of all species containing frayed base pairs. **(c)** Free-energy profiles as a function of n_{bp} generated from the lattice model. Profiles are shown at temperatures ranging from 5 to 100 ˚C. Figure is adapted from Sanstead et al.[80]

The previous section summarized several internal and terminal base-pair structural motions exhibited by duplexes at temperatures well below their melting transition, and it is important to clarify the relationship and differences between these motions from "pre-melting" transitions that occur at temperatures closer to the duplex melting transition. The entropic benefit of base-pair breaking promotes a greater abundance of "breathing" or partial disruptions as temperature increases. Heterogeneity in base-pair and stacking stability throughout a duplex additionally enables local regions to dissociate with populations greater than 10% at temperatures below the

melting transition of the entire duplex,[80, 83] which is in significant contrast to the equilibrium constants of 10^{-2} to 10^{-7} for base-pair opening and HG formation discussed in the previous section.

The most abundant form of "pre-melting" in short canonical oligonucleotides (5 – 20 nucleotides) is terminal fraying. Due to its favorable entropy, the population of frayed configurations and average number of frayed base pairs increases as temperature increases and is particularly evident for A:T termini near a G:C-rich region of base pairs (Fig. 1.3).[80, 83, 87] In stark contrast to the sigmoidal melting transitions of duplexes, the relatively small enthalpy and entropy changes for fraying lead to a nearly linear increase in the frayed population with temperature until the duplex fully dissociates. Fraying manifests as a broadening of the duplex free-energy basin in hybridization free-energy profiles generated from statistical models and coarse-grained MD simulations (Fig. 1.3c).[83-84] Additionally, the free-energy barrier height for A:T fraying becomes nearly negligible close to the melting transition (Fig. 1.2d), evidenced by T-jump experiments where nanosecond temperature changes as small as 5 ˚C induce energetically downhill A:T fraying.[32] These experiments also suggest that fraying of multiple A:T base pairs is not a cooperative process and instead follows a diffusive random walk. On the other hand, G:C terminal fraying contains a significantly greater free-energy barrier height as temperature jumps as large as 20 ˚C are unable to promote downhill fraying.[31-32]

Simultaneous disruption of multiple base pairs internal to the duplex is severely unfavorable in 5 – 20 nucleotide oligonucleotides, but these bubble configurations become a significant form of "pre-melting" in intermediate and longer duplexes (>50 nucleotides).[60, 88-90] The favorability for bubble formation is greatest among a stretch of A:T base pairs that is

surrounded by G:C-rich regions. Using a combination of UV absorption temperature series and temperature-quenching experiments, Zocchi and co-workers experimentally determined the thermodynamic properties of A:T-rich bubble opening.[88, 91] Unlike A:T terminal fraying, bubble opening is proposed to be cooperative and requires an A:T-rich bubble region of at least 20 nucleotides in length to form. Coarse-grained MD simulations also suggest a cooperative mechanism for opening of large bubbles that operates through both disruption of base pairs and twisting of the loop.[92-93] The large free-energy barrier associated with this process leads to simulated opening and closing times of ~15 ms and 40 μs, respectively, at temperatures well below the melting transition of the duplex.[93] The calculated closing timescale is similar to the observed "breathing" timescales from single-molecular fluorescence spectroscopy noted in the previous section,[46, 74] supporting that these measurements probe cooperative bubble formation. These simulations and measurements are still in a temperature regime where bubble opening is unfavorable with equilibrium constants of 10^{-2} to 10^{-3}. It is unclear if the kinetics and dynamics of bubble formation differ at higher temperatures where they are formed with significant population as described in the work of Zocchi & co-workers.[88, 91]

1.2.3 Configurations of single-strand oligonucleotides

Potentially the most overlooked question in oligonucleotide hybridization is the structural nature of the single-strand ensemble. The energetics of this ensemble are significant to the free-energy change for hybridization, and the average structure and motions likely play a role in the rate and dynamics of hybridization. It is well-established that most oligonucleotide sequences can fold to form intramolecular base pairs, even if highly unstable, but what does an "unfolded"

oligonucleotide actually look like? The flexbility of the single strand is primarily dictated by a combination of excluded volume interactions, electrostatic repulsion between backbone units, and stacking between nucleobases. Therefore, the single-strand ensemble is highly sensitive to counterion concentration and valency as electrostatic shielding of the backbone from counterions promotes more compact single-strand conformations. These changes are most commonly expressed in terms of mechanical and global properties, such as the persistence length and end-to-end distance, extracted from force-pulling, small-angle x-ray scattering (SAXS), or single-molecule FRET experiments on intermediate size (~40 nucleotides) oligonucleotides.[94-98] The persistence length of ssDNA ranges from ~2.0 nm at sub-mM counterion concentrations, approximately the length of 6 nucleotides, to as low as 0.8 nm with >100 mM $MgCl_2$, and values are shifted about 0.3 nm higher in ssRNA.[96, 99] These properties are also expected to depend on nucleobase sequence through tuning the stacking interaction strengths in the single-strand, which is supported by the consistent observation that adenine-only oligonucleotides are more rigid than thymine-only oligonucleotides.[94, 98-99]

To understand the role of single-strand structure and dynamics in hybridization, it is important to go beyond a global description of the single strand and characterize the abundance and distribution of specific interactions such as stacking and non-canonical base pairing. Direct comparison between experiments and computations of single strands has begun to provide molecular insight into the ensemble of elongated and compact configurations adopted under different counterion conditions. Refinement of static ssDNA ensembles using SAXS data indicated that adenine-only and thymine-only oligonucleotides adopt similar size distributions even though the adenine-only strand contains 3-5-fold more stacked bases on average, suggesting that there is

little correlation between the mechanical properties and stacking interactions of the strand.[97] The ensembles indicated that stacking in the 30-mer is broken up into segments of 3 – 6 nucleotides separated by short unstacked regions, suggesting that stacking is only weakly cooperative. Extensions to probe the kinetics and dynamics of single-strand ensembles require comparison of experiments, specifically those sensitive to local structure like NMR and IR spectroscopy, with all-atom MD simulations, and studies to date have indicated sequence-dependent preferences for stacking and forming non-canonical intramolecular base pairs.[94, 100] These studies also show that oligonucleotides as short as 4 nucleotides still sample a wide ensemble of structures.[95, 100] However, the accuracy of all-atom force fields for describing single-strands is still under question,[94-95, 100-101] hindering progress towards an understanding of dynamics in single-strands.

1.2.4 Kinetics of transitions between duplex and single-strand ensembles

To understand how transitions between duplex and single-strand ensembles occur, the majority of research (including much of this thesis) has focused on how the timescale of hybridization and dehybridization change as a function of environmental variables and sequence composition.[25-26, 33, 37, 102-107] As these kinetic experiments measure the rate of population change in duplex and single-strand ensembles, rather than the crossings or "transition paths" in between, they lack direct insight into the series of molecular events during hybridization. However, the observed kinetics arise from the underlying dynamics, and kinetic characterization of hybridization puts constraints on its molecular dynamics. These constraints, as well as the fact that the most popular molecular pictures of hybridization are based upon kinetic models, encourage a discussion of the key findings on the kinetics of oligonucleotide hybridization.

As the proposed double-helical structure of dsDNA was being developed in the early 1950s,[108-109] researchers were already interested in the physical properties of duplex association and dissociation.[110-112] The structural complexities of polymeric DNA hindered the interpretation of early kinetic experiments,[103] but the advent of relaxation methods with sub-μs time resolution (particularly T-jump)[113] and short oligonucleotide synthesis enabled kinetic studies on sequences as short as 4 nucleotides.[102, 104, 114] Although these studies were limited to a small number of oligonucleotide sequences accessible at the time, they established foundational principles of hybridization kinetics. T-jump and mixing relaxation profiles for hybridization followed a single kinetic component, excluding terminal fraying and triplex dissociation, prompting treatment of the observed rate constant (k_{obs}) with a two-state kinetic model to extract rate constants for association (k_a) and dissociation (k_d) as shown in Fig. 1.4a and Chapter 3.6. The two-state behavior arises from the cooperativity of base pairing that produces a single free-energy barrier for hybridization and dehybridization. Temperature-dependent measurements of k_d and k_a modeled with transition-state theory revealed the energetic properties of each free-energy barrier and the transition-state ensemble for hybridization (Fig. 1.4b). k_d increases exponentially with temperature, reflecting a large enthalpic barrier ($\Delta H_d^{\ddagger}$ ~ 175 kJ/mol for 10-mer) and a favorable entropy change ($\Delta S_d^{\ddagger} > 0$).[102, 104] In contrast, k_a exhibits a weak dependence on temperature in the absence of hairpins and usually decreases slightly as temperature increases (depending on the temperature range),[33, 102, 104] indicating that the rate of hybridization is primarily set by an entropic barrier ($\Delta S_a^{\ddagger} < 0$) and a slightly favorable enthalpy change ($\Delta H_a^{\ddagger} < 0$).

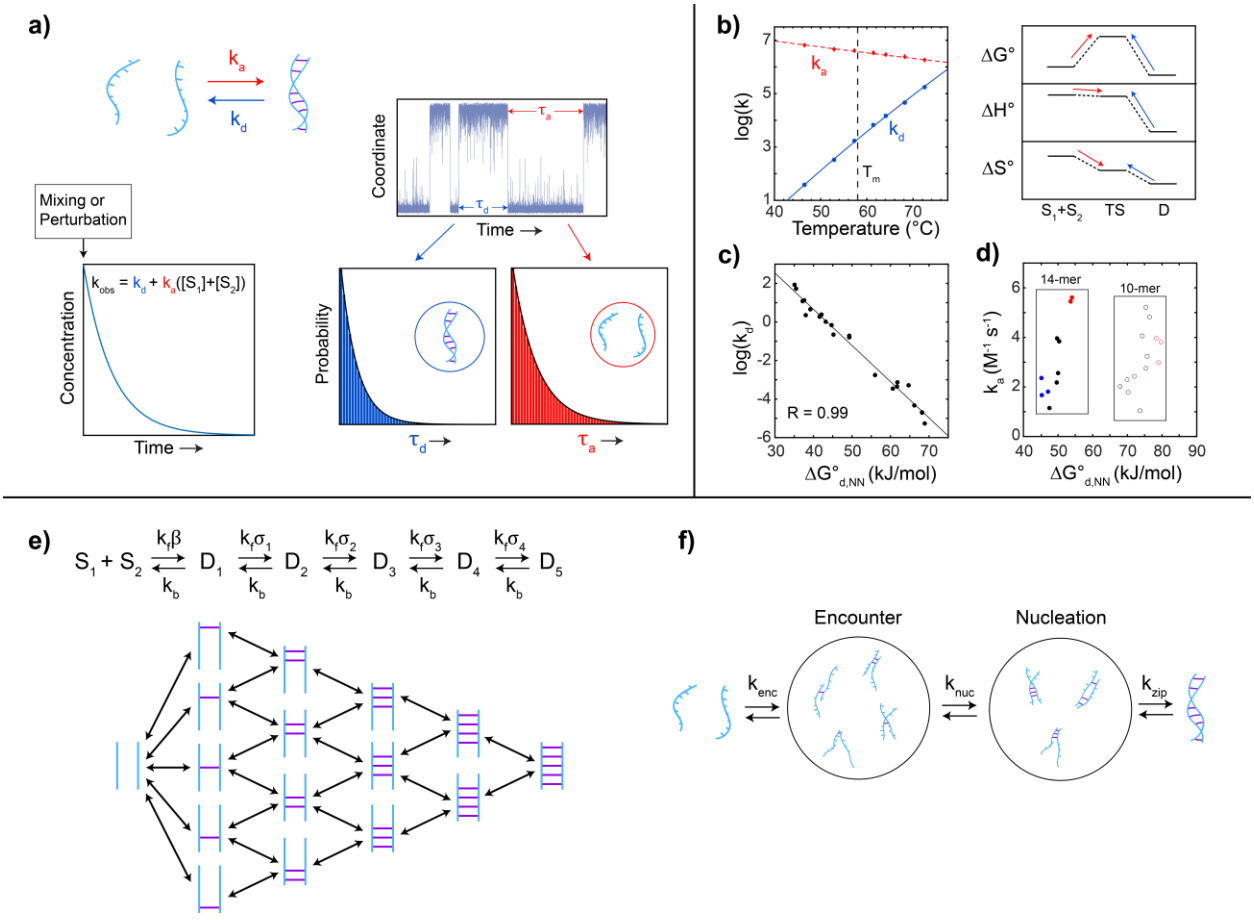

Figure 1.4 Kinetics of oligonucleotide hybridization. (a) Two-state kinetics commonly observed in ensemble and single-molecule measurements of hybridization. The observed rate of the slowest kinetic component (k_{obs}) from mixing or relaxation (usually temperature-jump) experiments is typically modeled as a sum of dissociation (k_d) and association (k_a) rate constants weighted by the monomer concentration at equilibrium. Single-molecule trajectories show jumps between duplex and single-strand states characterized by waiting times τ_d and τ_a, respectively, that follow exponential distributions. **(b)** (left) Temperature-trends in k_a and k_d for an 11-mer determined from T-jump experiments near the melting temperature (T_m), highlighting different energetic properties of the free-energy barriers for hybridization and dehybridization as shown schematically on the right. **(c)** Scatter plot showing the relationship between $\log(k_d)$ at 25 °C and the duplex dissociation free energy predicted from Santa Lucia's NN model ($\Delta G^\circ_{d,NN}$).[16] k_d data is taken from Rejali et al.[23] for oligonucleotides of 6 to 14 nucleotides in length. The solid line corresponds to a linear fit, and the data is well-correlated with a Pearson correlation coefficient (R) of 0.99. **(d)** Scatter plot of $\Delta G^\circ_{d,NN}$ vs. k_a using data on 10-mers and 14-mers from Hertel et al.[25] Open markers

Figure 1.4 Kinetics of oligonucleotide hybridization, continued

are 10-mers and closed markers are 14-mers, and blue, black, and red markers correspond to sequences with 42, 50, and 57% G:C base pair content. **(e)** Schematic of a nucleation-zipper kinetic model from Craig et al.[102] Each arrow involves the formation or breaking of an in-register base pair. Details are described in the text. **(f)** General schematic of recent kinetic models where hybridization is broken up into a diffusive encounter followed by nucleation and zippering.[25, 27] Unlike the nucleation-zipper model, these models account for out-of-register base pairing in the encounter complex and the possibility of multiple stretches of base pairs during the process of nucleation.

Early measurements of oligonucleotide hybridization were performed shortly after the development of statistical helix-coil models for nucleic acids.[90, 115-117] These statistical models were modified to create the "nucleation-zipper" kinetic model for oligonucleotide hybridization (Fig. 1.4e) that persists even today as the most common kinetic description of hybridization.[102, 104] The model is often expressed with minor modifications, but all versions assume that the rate-limiting step and free-energy barrier for hybridization corresponds to the formation of a small number of base pairs (known as nucleation) and that the remaining base pairs form in an energetically downhill process referred to as zippering. As for the thermodynamic zipper models, only configurations with a single stretch of base pairs are accounted for as illustrated in Fig. 1.4e. k_f and k_b are the rate constants for zippering or unzippering a single base pair, and their ratio gives the stability constant of a base pair from helix-coil models ($s = k_f/k_b$). Formation of the first few base pairs is slowed by factors β and σ_i, which are less than or equal to one and analogous to the cooperativity parameter, σ, in helix-coil models. β contains the translational diffusion for oligonucleotides to meet. However, measured k_a values are ~1000-fold smaller than the estimated diffusion-limited rate constant of short oligonucleotides,[37] suggesting that β also accounts for numerous failed attempts to form the first base pair or an encounter complex. The free-energy

barrier for hybridization may require the formation of additional base pairs that are energetically uphill. The formation of each base pair is described by their respective σ_i value, and $\sigma_i \sim 1$ for the formation of base pairs beyond the point of nucleation. These nucleation-zipper models have been successful in describing hybridization kinetics experiments, which is not surprising given that there are a large number of unknown parameters in the model and only a single kinetic observable (k_{obs}). However, the molecular characteristics of β and σ_i parameters are ambiguous and do not enable the prediction of hybridization rates for oligonucleotides of different sequences, lengths, and environments.

A resurgence of interest in hybridization kinetics over the past twenty years has reinforced aspects of the nucleation-zipper model and prompted modifications to maintain consistency with new experimental and computational results. Even though there have been numerous technological advances that enable new kinetic measurements of duplex conformational transitions or structural dynamics in proteins,[42, 118] the kinetic observables for nucleic acid hybridization are nearly the same as they were in the 1970s. T-jump spectroscopy and mixing are still two of the most common kinetic experiments,[25-26, 32, 119-120] but they have been improved to provide sensitivity to different structural coordinates, higher data quality and time resolution, and greater throughput. Single-molecule spectroscopy in principle provides access to the properties of transition paths between duplex and single-strand states, as demonstrated for protein association,[118] but insufficient time resolution typically limits these experiments to measuring distributions of k_a and k_d from which their average values are reported.[37, 105, 121] Using these methods, and helped by advances in oligonucleotide synthesis,[122] there have been multiple studies providing k_a and k_d values for over

40 oligonucleotide sequences.[23-26, 123] A common finding from these studies is that k_d is well-correlated with the predicted dissociation free-energy of the duplex from NN models ($\Delta G^{\circ}_{d,NN}$, Fig. 1.4c) and that the change in the free-energy barrier for dehybridization ($\Delta G^{\ddagger}_{d}$) exhibits a near 1-to-1 relationship with ΔG°_{d}.[23, 26, 119] In contrast, k_a usually varies only 2-to-5-fold with sequence and shows poor correlation with $\Delta G^{\circ}_{d,NN}$ (Fig. 1.4d).[23, 25-26] There is evidence that k_a is correlated with oligonucleotide length and G:C base-pair content, but these results vary across studies and require further experiments to establish empirical trends.[25-27, 37] These observations support previous notions that the transition-state ensemble of hybridization involves a small number of base-pair contacts and better resembles the single-strand ensemble relative to that of the duplex.

Sequence-dependent survey studies of hybridization kinetics have identified that k_a increases for oligonucleotides with increasingly repetitive nucleobase sequence due to the ability to base pair in out-of-register configurations.[25, 27] Binding out-of-register may be a kinetic trap in some cases, but increasing experimental and computational evidence suggests that it generally helps keep strands together longer in an encounter complex and provides more ways to nucleate (Fig. 1.4f).[25, 27, 29-31, 124] The combinatorics and stability of out-of-register base pairing are now being incorporated into kinetic models of hybridization with progress towards prediction of k_a as a function of sequence.[25, 27, 125]

1.3 Towards dynamics of hybridization

1.3.1 Encounter complex and nucleation

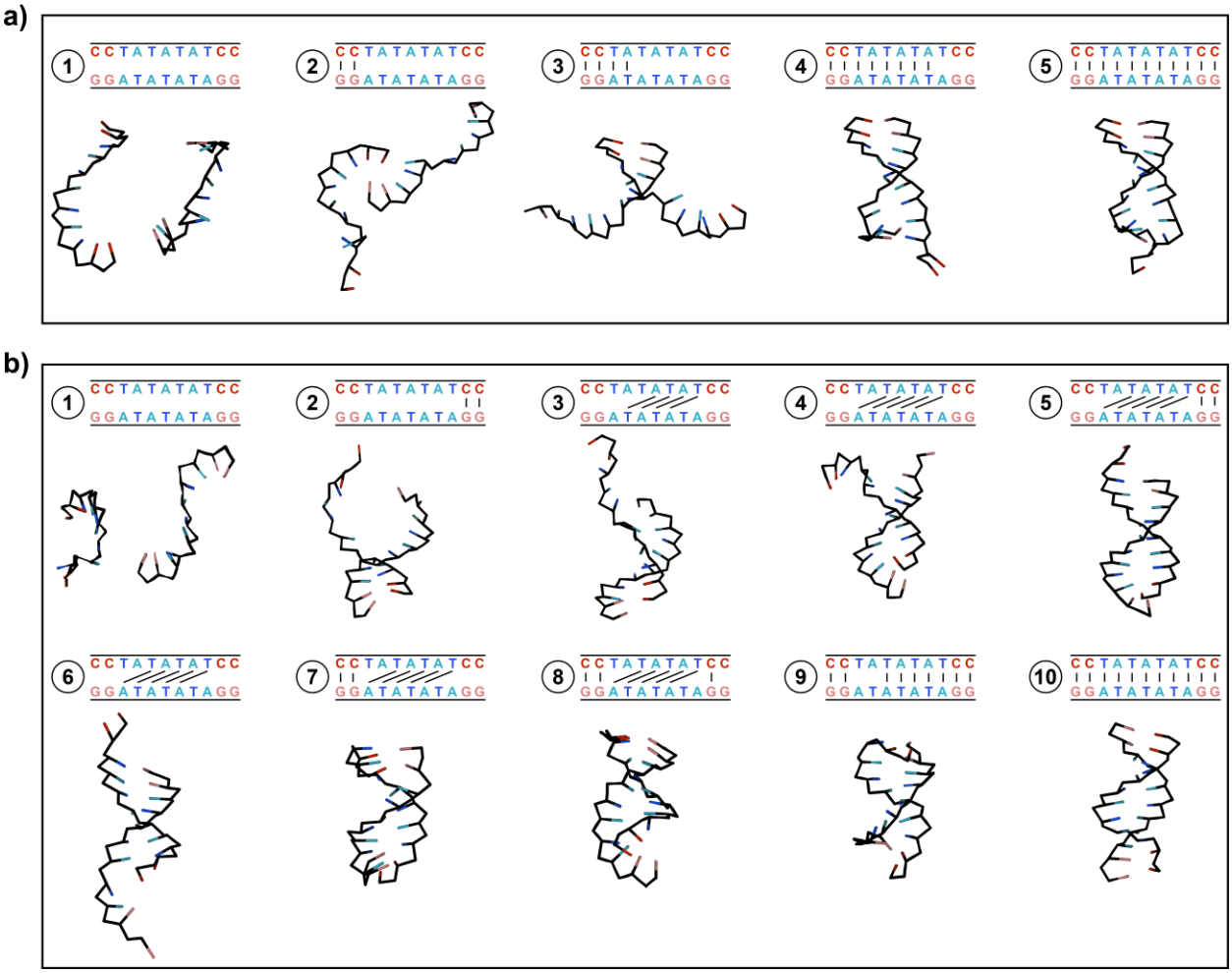

Figure 1.5 Multiple pathways of hybridization. (a) Snapshots from a single hybridization event following a classic nucleation-zipper mechanism in a MD simulation using the 3SPN.2 coarse-grained model. The simulation box contains 5′-CCTATATATCC-3′ and its complement with an implicit counterion ionic strength of 600 mM. **(b)** A second hybridization event that first involves binding with out-of-register base pairs followed by formation of in-register base pairs.

While kinetic models of hybridization are improving in their description of strand encounter and nucleation, they provide little insight into the molecular configurations that make

up the encounter complex and mechanism for nucleation. For example, what are the structural differences between the many failed nucleation attempts and those that actually succeed? Accessing these dynamics is challenging both with experiment and computation, and only recently has any direct insight been achieved with either.[29-31, 34, 126]

Inaccuracies in all-atom MD force fields for non-duplex structures and difficulties with sampling hybridization transitions pushed for the development of coarse-grained models to sample trajectories of hybridization events. Numerous coarse-grained models have been developed,[127-132] and oxDNA[127] has become the most popular for simulating oligonucleotide assembly in various contexts.[10, 29, 133-134] Regardless of the coarse-grained force field, a common observation is that oligonucleotides can hybridize via multiple pathways (Fig. 1.5). Canonical nucleation-zipper mechanisms, where strands bind together via a few in-register base pairs and subsequently zip up as a single-stretch of base pairs, are found in most sequences.[29-31, 135] Hybridization is also found to initiate via out-of-register base pairs.[29-31, 131] Out-of-register configurations are most abundant in fully repetitive sequences, but they can contribute to the hybridization transition paths of nearly any sequence. Upon binding in an out-of-register configuration, the partial duplex may dissociate and then quickly reform in-register base pairs and hybridize via a nucleation-zipper mechanism or the partial duplex may directly transition to a full duplex through configurations with both in- and out-of-register base pairs (Fig. 1.5b). Multiple types of such mechanisms have been observed in coarse-grained MD simulations, demonstrating the beneficial role of out-of-register contacts in providing more routes to hybridization.[29-30, 131]

Most experimental evidence for heterogeneity in hybridization pathways comes indirectly from the kinetic measurements discussed in Section 1.2.4,[25, 27] but advances in single-molecule spectroscopy may provide direct means to probe molecule configurations during hybridization events. Single-molecule FRET and optical tweezers have been applied to probe nucleic acid folding transition paths, but not yet bimolecular hybridization.[35, 136-139] However, a breakthrough study by Granick & co-workers used liquid-cell transmission electron microscopy (TEM) to directly image hybridization of 90-mer oligonucleotides in aqueous solution.[34] Interactions with the graphene sample cell reduce the translational diffusion coefficient of single strands to $\sim 2 \times 10^{-14}$ cm^2/s compared to $\sim 4 \times 10^{-7}$ cm^2/s in free solution, allowing visualization of oligonucleotides as they hybridize even with an imaging frame rate of ~ 3 s^{-1}. Even though only 18 hybridization events were captured, a variety of mechanisms were observed that include nucleation-zipper and those with apparent out-of-register base pairs. Multiple events were found to initiate via bubble configurations, which surprisingly promoted hybridization in certain sequences. Different mechanisms are also distinguished by time-dependent changes in two-dimensional orientation and the projected end-to-end distance of the bound oligonucleotides, providing the possibility for new quantitative measures of hybridization dynamics.

1.3.2 Base-pair zippering

It is generally assumed that the zippering phase of hybridization proceeds through rapid sequential formation of a continuous stretch of in-register base pairs, but such a picture had never been experimentally tested until recently. Single-molecule measurements of DNA hairpin folding cannot inform on the bimolecular encounter of oligonucleotides in hybridization, but they can

report on the dynamic properties of base-pair zippering. Using optical tweezers measurements at constant force, Woodside & co-workers have investigated the dynamic properties of zippering in DNA hairpins with 20 nucleotide stems and determined a number of quantitative properties of transition paths through measurements of thousands of folding and unfolding transitions.[35, 136-138, 140] They initially identified through statistical tests of transition paths and committor analysis (Chapter 5) that zippering is well-described as 1D diffusion along the end-to-end extension experimental coordinate,[140] similar to previous reports on protein folding.[141-142] Sequence-dependent variation in the average transition-path duration from 20 – 35 μs indicated that the diffusion coefficient along the extension coordinate increases roughly linearly with G:C base-pair content and is nearly 2-fold larger for a G:C base pair relative to A:T base pair.[35] The molecular origin for this sequence-dependence in the zippering diffusion coefficient is still unclear but was hypothesized to lower the free-energy barrier for rotation of the glycosidic bond in guanine relative to adenine. The different diffusion coefficients for A:T and G:C base pairing also imply that the diffusion coefficient for zippering of a mixed oligonucleotide is position-dependent along the reaction coordinate.

More recent investigations into the shape of transition paths for hairpin folding potentially reflect deviations from 1D diffusion along the end-to-end extension coordinate.[138] Transition path trajectories show pauses with durations of 1 – 20 μs that are defined as a region where the velocity is approximately zero. Pauses are observed across all regions of the transition paths, suggesting that they are ubiquitous during zippering and either correspond to small free-energy barriers for motion along the extension coordinate or structural motion along other coordinates to which extension is insensitive. Pause time distributions are multi-exponential regardless of the

nucleobase sequence, suggesting that either multiple free-energy barrier heights are encountered or that zippering proceeds through multiple types of multi-dimensional structural motions. In either scenario, the prevailing picture of 1D diffusive zippering is inconsistent with the results and instead a more complex set of dynamics must be at play.

1.3.3 Outlook for studying dynamics of hybridization

The collection of investigations on the energetics and time-dependence of hybridization spanning from the 1950s until today have established empirical properties of oligonucleotides that enabled the birth of numerous fields in nucleic acid technology. However, only within the last ten years have single-molecule experiments and molecular dynamics simulations of nucleic acids advanced far enough to bring us to a cusp in the direct investigation of structural dynamics in nucleic acid assembly and more generally in biomolecular recognition. Direct visualization of hybridization dynamics will provide a physical basis to explain the empirical trends observed to date and predictive power to design new experiments. As dynamic studies become more feasible, there are certain outstanding questions that may be most insightful to address: (1) structural differences between successful and failed hybridization encounters, (2) role of hydration and ion rearrangement during hybridization, and (3) influence of chemical modifications and lesions on dynamics of hybridization and base pairing within the duplex.

Tuning the propensity for failed vs. successful hybridization is potentially the most powerful design principle for optimizing hybridization kinetics. Some of the kinetic models and coarse-grained MD simulations summarized in this chapter suggest that having more possibilities for initial binding and rearrangement acts to speed up hybridization,[25, 27, 29-30] but there is still little

sense for which pathways are actually utilized in experiment. It is also important to clarify the meaning of a "failed" hybridization event. In one limit strands collide and quickly diffuse apart 1000 times on average before successful hybridization whereas in the other limit strands encounter and shuffle between native and non-native contacts until eventually nucleating as observed on graphene with liquid-cell TEM.[34] The reality of hybridization in free solution lies somewhere in between.

In addition to the formation of base-pairing and stacking interactions, hybridization requires disruption of hydration at the nucleobase interface, restriction of water molecules in the minor groove, and uptake in counterions associated with the phosphate backbone.[56, 143-145] While the thermodynamics of these changes are known to be significant factors in the stability of the duplex,[146-147] the time-dependence and dynamics are unknown. These dynamics are foundational to predicting how the properties of hybridization will change in different *in vitro* and *in vivo* environments.[107, 148] The significant dependence of k_a on counterion concentration yet minor dependence for k_d hints towards a mechanism where the uptake of counterions occurs in the encounter complex and nucleation phase of hybridization, but this is only speculative as counterion concentration will also influence the energetics of the single-strand ensemble. Computational and experimental advances over the past ten years have enabled more direct probing of the hydration and counterion structure around duplexes that may lend themselves to the study of rearrangement during hybridization.[144, 149-151]

1.4 Thesis Outline

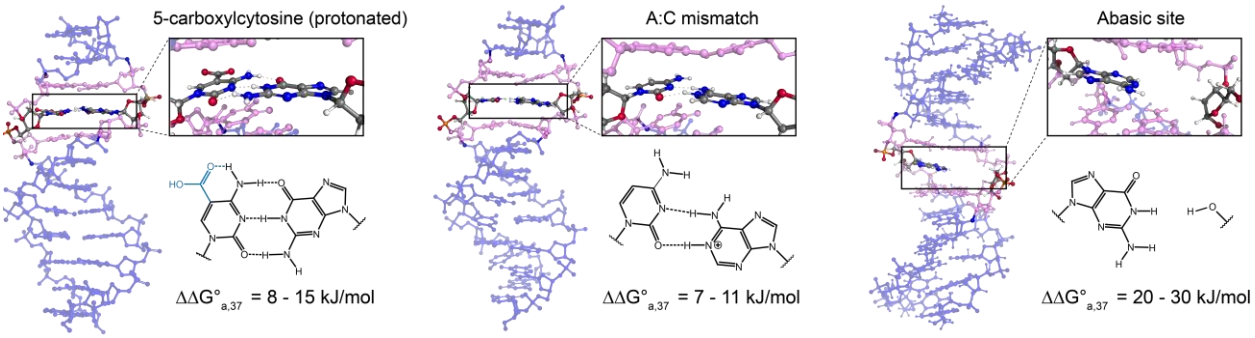

Figure 1.6 Duplex destabilization from non-canonical base pairs and modifications. Experimental structures and reported reduction in DNA duplex free energy ($\Delta\Delta G^{\circ}_{a,37}$)[38, 152-153] from (left) protonated 5-carboxylcytosine (PDB: 4PWM), (center) adenine:cytosine mismatch (PDB: 1D99), and (right) abasic site (PDB: 1G5D).

The body of work summarized in this chapter reveals several standing questions surrounding the molecular properties of nucleic acid hybridization between canonical oligonucleotides, and another layer of complexity arises from the multitude of non-canonical interactions that result in duplexes from chemical modifications, mismatches, and lesions. The stability of the duplex is extraordinarily sensitive to such modifications, often exhibiting free-energy penalties of 5 – 30 kJ/mol for the replacement or addition of a single chemical bond.[154-155] Such sensitivity in base-pair stability may play a role in the enzymatic recognition of modified and damaged sites in the genome and has been utilized in the development of nucleic acid technologies.[156-158] While some modifications induce large changes in the average structure of the duplex,[159-160] many appear to have a minor structural impact (Fig. 1.6).[50, 161] Instead, these sites may alter duplex stability through kinetic and dynamic effects and disruption of base-pairing cooperativity in the duplex.[47-49, 162]

The work in this thesis aims to improve our understanding of how chemical modifications and lesions alter DNA duplex and hybridization dynamics through the use of experiments, modeling, and simulations. We begin in Chapter 2 by summarizing the benefits and difficulties of using nonlinear infrared spectroscopy to study oligonucleotides with comparison to nuclear magnetic resonance and electronic spectroscopy. An emphasis is placed on the sensitivity of infrared spectroscopy to different types of base pairing. We refer readers to previous resources on 2D IR spectroscopy as its principles have been published repeatedly throughout the history of the Tokmakoff group.[163-164]

Chapters 3 and 4 contain a discussion of the principles and practical factors guiding the use of temperature-jump infrared spectroscopy for studying processes that occur on timescales ranging from nanoseconds to minutes. In addition to an outline of optical design, we offer a discussion of simple data processing, kinetic analysis, and characterization of potential artifacts.

Chapter 5 introduces concepts in canonical oligonucleotide hybridization dynamics through analysis of coarse-grained molecular dynamics simulations. We begin by assessing the quality of base-pairing collective variables for describing hybridization through statistical measures of transition paths. After establishing select collective variables, we move on to describe sequence-dependent trends in hybridization dynamics.

Chapters 6 and 7 provide an analysis of how removing a single nucleobase (i.e. formation of an abasic site) disrupts the stability, kinetics, and dynamics of the DNA duplex. A combination of experiments and simulations are used to directly show that an additional free-energy barrier arises for base pairing on each side of the abasic site, which disrupts the cooperativity of base

pairing observed in the dynamics, kinetics, and thermodynamics of hybridization. We then describe how the position of an abasic site in an oligonucleotide alters the stability and dynamics of the duplex through fraying, metastable states, and out-of-register base pairing.

In Chapter 8 we demonstrate how 5-carboxylcytosine, a nucleobase modification used in the active demethylation cycle of DNA, acts as a protonatable switch to reduce the cooperativity of G:C base pairing in the duplex. Direct comparison is made with N3 protonation of cytosine, illustrating distinct thermodynamic and kinetic influence from each type of protonation.

Lastly, in Chapter 9 we characterize the hybridization of dinucleotides onto oligonucleotide gap and overhang templates. The discussion primarily aims to introduce methods for accurately measuring temperature-dependent hybridization thermodynamics and kinetics in these systems, and along the way, we identify energetic properties distinct from hybridization between two canonical oligonucleotides.

1.5 Acknowledgements

I thank Yumin Lee and Jakob Schauss for helpful feedback on this chapter.

1.6 References

1. Jiang, F.; Doudna, J. A., CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **2017,** *46*, 505-529.

2. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mat.* **2017,** *3*, 1-23.

3. Song, J.; Li, Z.; Wang, P.; Meyer, T.; Mao, C.; Ke, Y., Reconfiguration of DNA molecular arrays driven by information relay. *Science* **2017,** *357*, eaan3377.

4. Jungmann, R.; Steinhauer, C.; Scheible, M.; Kuzyk, A.; Tinnefeld, P.; Simmel, F. C., Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **2010,** *10*, 4756-4761.

5. Klein, M.; Eslami-Mossallam, B.; Arroyo, D. G.; Depken, M., Hybridization kinetics explains CRISPR-Cas off-targeting rules. *Cell Rep.* **2018,** *22*, 1413-1423.

6. Boyle, E. A.; Andreasson, J. O.; Chircus, L. M.; Sternberg, S. H.; Wu, M. J.; Guegler, C. K.; Doudna, J. A.; Greenleaf, W. J., High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 5461-5466.

7. Strohkendl, I.; Saifuddin, F. A.; Rybarski, J. R.; Finkelstein, I. J.; Russell, R., Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Mol. Cell* **2018,** *71*, 816-824. e3.

8. Schneider, F.; Möritz, N.; Dietz, H., The sequence of events during folding of a DNA origami. *Sci. Adv.* **2019,** *5*, eaaw1412.

9. Wang, J.; Wei, Y.; Zhang, P.; Wang, Y.; Xia, Q.; Liu, X.; Luo, S.; Shi, J.; Hu, J.; Fan, C., Probing Heterogeneous Folding Pathways of DNA Origami Self-Assembly at the Molecular Level with Atomic Force Microscopy. *Nano Lett.* **2022,** *22*, 7173-7179.

10. Snodin, B. E.; Romano, F.; Rovigatti, L.; Ouldridge, T. E.; Louis, A. A.; Doye, J. P., Direct simulation of the self-assembly of a small DNA origami. *ACS Nano* **2016,** *10*, 1724-1737.

11. Cumberworth, A.; Frenkel, D.; Reinhardt, A., Simulations of DNA-origami self-assembly reveal design-dependent nucleation barriers. *Nano Lett.* **2022,** *22*, 6916-6922.

12. Jacobs, W. M.; Reinhardt, A.; Frenkel, D., Rational design of self-assembly pathways for complex multicomponent structures. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112*, 6313-6318.

13. Schueder, F.; Stein, J.; Stehr, F.; Auer, A.; Sperl, B.; Strauss, M. T.; Schwille, P.; Jungmann, R., An order of magnitude faster DNA-PAINT imaging by optimized sequence design and buffer conditions. *Nat. Methods* **2019,** *16*, 1101-1104.

14. Strauss, S.; Jungmann, R., Up to 100-fold speed-up and multiplexing in optimized DNA-PAINT. *Nat. Methods* **2020,** *17*, 789-791.

15. Gray, D. M.; Tinoco Jr, I., A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers* **1970,** *9*, 223-244.

16. SantaLucia Jr, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 1460-1465.

17. SantaLucia Jr, J.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-440.

18. Xia, T.; SantaLucia Jr, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H., Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson− Crick base pairs. *Biochem.* **1998,** *37*, 14719-14735.

19. Fornace, M. E.; Huang, J.; Newman, C. T.; Porubsky, N. J.; Pierce, M. B.; Pierce, N. A., NUPACK: analysis and design of nucleic acid structures, devices, and systems. *ChemRxiv* **2022**, DOI:10.26434/chemrxiv-2022-xv98l.

20. Owczarzy, R.; Tataurov, A. V.; Wu, Y.; Manthey, J. A.; McQuisten, K. A.; Almabrazi, H. G.; Pedersen, K. F.; Lin, Y.; Garretson, J.; McEntaggart, N. O., IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* **2008,** *36*, W163-W169.

21. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003,** *31*, 3406-3415.

22. Turner, D. H.; Mathews, D. H., NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* **2010,** *38*, D280-D282.

23. Rejali, N. A.; Ye, F. D.; Zuiter, A. M.; Keller, C. C.; Wittwer, C. T., Nearest-neighbour transition-state analysis for nucleic acid kinetics. *Nucleic Acids Res.* **2021,** *49*, 4574-4585.

24. Zhang, J. X.; Fang, J. Z.; Duan, W.; Wu, L. R.; Zhang, A. W.; Dalchau, N.; Yordanov, B.; Petersen, R.; Phillips, A.; Zhang, D. Y., Predicting DNA hybridization kinetics from sequence. *Nat. Chem.* **2018,** *10*, 91-98.

25. Hertel, S.; Spinney, R. E.; Xu, S. Y.; Ouldridge, T. E.; Morris, R. G.; Lee, L. K., The stability and number of nucleating interactions determine DNA hybridization rates in the absence of secondary structure. *Nucleic Acids Res.* **2022,** *50*, 7829-7841.

26. Todisco, M.; Szostak, J. W., Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Res.* **2022,** *50*, 9647-9662.

27. Phan, T. T.; Phan, T. M.; Schmit, J. D., Beneficial and detrimental effects of non-specific binding during DNA hybridization. *Biophys. J.* **2023,** *122*, 835-848.

28. Zhang, D. Y.; Winfree, E., Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **2009,** *131*, 17303-17314.

29. Ouldridge, T. E.; Šulc, P.; Romano, F.; Doye, J. P.; Louis, A. A., DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **2013,** *41*, 8886-8895.

30. Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J., Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* **2014,** *141*, 07B613_1.

31. Jones, M. S.; Ashwood, B.; Tokmakoff, A.; Ferguson, A. L., Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, markov state models, and infrared spectroscopy. *J. Am. Chem. Soc.* **2021,** *143*, 17395-17411.

32. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

33. Chen, C.; Wang, W.; Wang, Z.; Wei, F.; Zhao, X. S., Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* **2007,** *35*, 2875-2884.

34. Wang, H.; Li, B.; Kim, Y.-J.; Kwon, O.-H.; Granick, S., Intermediate states of molecular self-assembly from liquid-cell electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 1283-1292.

35. Neupane, K.; Wang, F.; Woodside, M., Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 1329-1334.

36. Alvey, H. S.; Gottardo, F. L.; Nikolova, E. N.; Al-Hashimi, H. M., Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. *Nat. Commun.* **2014,** *5*, 4786.

37. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.

38. Dubini, R. C.; Korytiaková, E.; Schinkel, T.; Heinrichs, P.; Carell, T.; Rovó, P., 1H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives. *ACS Phys. Chem. Au* **2022,** *2*, 237-246.

39. Ashwood, B.; Sanstead, P. J.; Dai, Q.; He, C.; Tokmakoff, A., 5-Carboxylcytosine and cytosine protonation distinctly alter the stability and dehybridization dynamics of the DNA duplex. *J. Phys. Chem. B* **2019,** *124*, 627-640.

40. Ghosh, S.; Takahashi, S.; Ohyama, T.; Endoh, T.; Tateishi-Karimata, H.; Sugimoto, N., Nearest-neighbor parameters for predicting DNA duplex stability in diverse molecular crowding conditions. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 14194-14201.

41. Shi, H.; Liu, B.; Nussbaumer, F.; Rangadurai, A.; Kreutz, C.; Al-Hashimi, H. M., NMR chemical exchange measurements reveal that N 6-methyladenosine slows RNA annealing. *J. Am. Chem. Soc.* **2019,** *141*, 19988-19993.

42. Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M., Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011,** *470*, 498-502.

43. von Hippel, P. H.; Johnson, N. P.; Marcus, A. H., Fifty years of DNA "breathing": Reflections on old and new approaches. *Biopolymers* **2013,** *99*, 923-954.

44. Kimsey, I. J.; Szymanski, E. S.; Zahurancik, W. J.; Shakya, A.; Xue, Y.; Chu, C.-C.; Sathyamoorthy, B.; Suo, Z.; Al-Hashimi, H. M., Dynamic basis for dG• dT misincorporation via tautomerization and ionization. *Nature* **2018,** *554*, 195-201.

45. Lindahl, V.; Villa, A.; Hess, B., Sequence dependency of canonical base pair opening in the DNA double helix. *PLOS Comput. Biol.* **2017,** *13*, e1005463.

46. Altan-Bonnet, G.; Libchaber, A.; Krichevsky, O., Bubble dynamics in double-stranded DNA. *Phys. Rev. Lett.* **2003,** *90*, 138101.

47. Shi, H.; Kimsey, I. J.; Gu, S.; Liu, H.-F.; Pham, U.; Schumacher, M. A.; Al-Hashimi, H. M., Revealing AT and GC Hoogsteen base pairs in stressed protein-bound duplex DNA. *Nucleic Acids Res.* **2021,** *49*, 12540-12555.

48. Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016,** *7*, 10813.

49. Yin, Y.; Yang, L.; Zheng, G.; Gu, C.; Yi, C.; He, C.; Gao, Y. Q.; Zhao, X. S., Dynamics of spontaneous flipping of a mismatched base in DNA duplex. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111*, 8043-8048.

50. Chen, J.; Dupradeau, F.-Y.; Case, D. A.; Turner, C. J.; Stubbe, J., DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* **2008,** *36*, 253-262.

51. López-Güell, K.; Battistini, F.; Orozco, M., Correlated motions in DNA: beyond base-pair step models of DNA flexibility. *Nucleic Acids Res.* **2023,** *51*, 2633-2640.

52. Dans, P. D.; Balaceanu, A.; Pasi, M.; Patelli, A. S.; Petkevičiūtė, D.; Walther, J.; Hospital, A.; Bayarri, G.; Lavery, R.; Maddocks, J. H., The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules. *Nucleic Acids Res.* **2019,** *47*, 11090-11102.

53. Vafabakhsh, R.; Ha, T., Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **2012,** *337*, 1097-1101.

54. Da Rosa, G.; Grille, L.; Calzada, V.; Ahmad, K.; Arcon, J. P.; Battistini, F.; Bayarri, G.; Bishop, T.; Carloni, P.; Cheatham III, T.; Collepardo-Guevara, R.; Czub, J.; Espinosa, J. R.;

Orozco, M.; Dans, P. D., Sequence-dependent structural properties of B-DNA: what have we learned in 40 years? *Biophysical Reviews* **2021**, 1-11.

55. Duboué-Dijon, E.; Fogarty, A. C.; Hynes, J. T.; Laage, D., Dynamical disorder in the DNA hydration shell. *J. Am. Chem. Soc.* **2016,** *138*, 7610-7620.

56. Kuchuk, K.; Sivan, U., Hydration structure of a single DNA molecule revealed by frequency-modulation atomic force microscopy. *Nano Lett.* **2018,** *18*, 2733-2737.

57. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D., Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006,** *34*, 564-574.

58. Printz, M. P.; von Hippel, P. H., Hydrogen exchange studies of DNA structure. *Proc. Natl. Acad. Sci. U.S.A.* **1965,** *53*, 363-370.

59. Frank-Kamenetskii, M. D.; Prakash, S., Fluctuations in the DNA double helix: A critical review. *Phys. Life Rev.* **2014,** *11*, 153-170.

60. Manghi, M.; Destainville, N., Physics of base-pairing dynamics in DNA. *Phys. Rep.* **2016,** *631*, 1-41.

61. McGhee, J. D.; Von Hippel, P. H., Formaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA. *Biochem.* **1977,** *16*, 3276-3293.

62. Lukashin, A. V.; Vologodskii, A. V.; Frank-Kamenetskii, M. D.; Lyubchenko, Y. L., Fluctuational opening of the double helix as revealed by theoretical and experimental study of DNA interaction with formaldehyde. *J. Mol. Biol.* **1976,** *108*, 665-682.

63. Guéron, M.; Leroy, J.-L., [16] Studies of base pair kinetics by NMR measurement of proton exchange. In *Methods in enzymology*, Elsevier: 1995; Vol. 261, pp 383-413.

64. Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Guéron, M., Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* **1988,** *200*, 223-238.

65. Guéron, M.; Kochoyan, M.; Leroy, J.-L., A single mode of DNA base-pair opening drives imino proton exchange. *Nature* **1987,** *328*, 89-92.

66. Krueger, A.; Protozanova, E.; Frank-Kamenetskii, M. D., Sequence-dependent basepair opening in DNA double helix. *Biophys. J.* **2006,** *90*, 3091-3099.

67. Wärmländer, S.; Sen, A.; Leijon, M., Imino proton exchange in DNA catalyzed by ammonia and trimethylamine: evidence for a secondary long-lived open state of the base pair. *Biochem.* **2000,** *39*, 607-615.

68. Várnai, P.; Canalia, M.; Leroy, J.-L., Opening Mechanism of G⊙ T/U Pairs in DNA and RNA Duplexes: A Combined Study of Imino Proton Exchange and Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2004,** *126*, 14659-14667.

69. Priyakumar, U. D.; MacKerell, A. D., Computational approaches for investigating base flipping in oligonucleotides. *Chem. Rev.* **2006,** *106*, 489-505.

70. Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham III, T.; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C., μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* **2014,** *42*, 12272-12283.

71. Giudice, E.; Várnai, P.; Lavery, R., Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* **2003,** *31*, 1434-1443.

72. Levintov, L.; Paul, S.; Vashisth, H., Reaction coordinate and thermodynamics of base flipping in RNA. *J. Chem. Theory Comput.* **2021,** *17*, 1914-1921.

73. Snoussi, K.; Leroy, J.-L., Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochem.* **2001,** *40*, 8898-8904.

74. Phelps, C.; Lee, W.; Jose, D.; Von Hippel, P. H.; Marcus, A. H., Single-molecule FRET and linear dichroism studies of DNA breathing and helicase binding at replication fork junctions. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 17320-17325.

75. Hoogsteen, K., The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1959,** *12*, 822-823.

76. Zhou, H.; Hintze, B. J.; Kimsey, I. J.; Sathyamoorthy, B.; Yang, S.; Richardson, J. S.; Al-Hashimi, H. M., New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic Acids Res.* **2015,** *43*, 3420-3433.

77. Ray, D.; Andricioaei, I., Free Energy Landscape and Conformational Kinetics of Hoogsteen Base Pairing in DNA vs. RNA. *Biophys. J.* **2020,** *119*, 1568-1579.

78. Rangadurai, A.; Zhou, H.; Merriman, D. K.; Meiser, N.; Liu, B.; Shi, H.; Szymanski, E. S.; Al-Hashimi, H. M., Why are Hoogsteen base pairs energetically disfavored in A-RNA compared to B-DNA? *Nucleic Acids Res.* **2018,** *46*, 11099-11114.

79. Nonin, S.; Leroy, J.-L.; Guéron, M., Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochem.* **1995,** *34*, 10652-10659.

80. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

81. Pörschke, D., A direct measurement of the unzippering rate of a nucleic acid double helix. *Biophys. Chem.* **1974,** *2*, 97-101.

82. Chen, X.; Zhou, Y.; Qu, P.; Zhao, X. S., Base-by-base dynamics in DNA hybridization probed by fluorescence correlation spectroscopy. *J. Am. Chem. Soc.* **2008,** *130*, 16947-16952.

83. Araque, J.; Robert, M., Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* **2016,** *144*, 125101.

84. Sanstead, P. J.; Tokmakoff, A., A lattice model for the interpretation of oligonucleotide hybridization experiments. *J. Chem. Phys.* **2019,** *150*, 185104.

85. Zgarbová, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P., Base pair fraying in molecular dynamics simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014,** *10*, 3177-3189.

86. Pinamonti, G.; Paul, F.; Noé, F.; Rodriguez, A.; Bussi, G., The mechanism of RNA base fraying: Molecular dynamics simulations analyzed with core-set Markov state models. *J. Chem. Phys.* **2019,** *150*, 154123.

87. Dabin, A.; Stirnemann, G., Toward a Molecular Mechanism of Complementary RNA Duplexes Denaturation. *J. Phys. Chem. B* **2023**.

88. Zeng, Y.; Montrichok, A.; Zocchi, G., Length and statistical weight of bubbles in DNA melting. *Phys. Rev. Lett.* **2003,** *91*, 148101.

89. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.

90. Poland, D.; Scheraga, H. A., *Theory of helix-coil transitions in biopolymers*. Academic Press.: 1970.

91. Zeng, Y.; Montrichok, A.; Zocchi, G., Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.* **2004,** *339*, 67-75.

92. Dasanna, A. K.; Destainville, N.; Palmeri, J.; Manghi, M., Slow closure of denaturation bubbles in DNA: Twist matters. *Phys. Rev. E* **2013,** *87*, 052703.

93. Sicard, F.; Destainville, N.; Manghi, M., DNA denaturation bubbles: Free-energy landscape and nucleation/closure rates. *J. Chem. Phys.* **2015,** *142*.

94. Grotz, K. K.; Nueesch, M. F.; Holmstrom, E. D.; Heinz, M.; Stelzl, L. S.; Schuler, B.; Hummer, G., Dispersion correction alleviates dye stacking of single-stranded DNA and RNA in simulations of single-molecule fluorescence experiments. *J. Phys. Chem. B* **2018,** *122*, 11626-11639.

95. Bergonzo, C.; Grishaev, A.; Bottaro, S., Conformational heterogeneity of UCAAUC RNA oligonucleotide from molecular dynamics simulations, SAXS, and NMR experiments. *RNA* **2022,** *28*, 937-946.

96. Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L., Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 799-804.

97. Plumridge, A.; Meisburger, S. P.; Andresen, K.; Pollack, L., The impact of base stacking on the conformations and electrostatics of single-stranded DNA. *Nucleic Acids Res.* **2017,** *45*, 3932-3943.

98. McIntosh, D. B.; Duggan, G.; Gouil, Q.; Saleh, O. A., Sequence-dependent elasticity and electrostatics of single-stranded DNA: signatures of base-stacking. *Biophys. J.* **2014,** *106*, 659-666.

99. Sim, A. Y.; Lipfert, J.; Herschlag, D.; Doniach, S., Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E* **2012,** *86*, 021901.

100. Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K., Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Sci. Adv.* **2018,** *4*, eaar8521.

101. Oweida, T. J.; Kim, H. S.; Donald, J. M.; Singh, A.; Yingling, Y. G., Assessment of AMBER force fields for simulations of ssDNA. *J. Chem. Theory Comput.* **2021,** *17*, 1208-1217.

102. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.

103. Wetmur, J. G.; Davidson, N., Kinetics of renaturation of DNA. *J. Mol. Biol.* **1968,** *31*, 349-370.

104. Pörschke, D.; Eigen, M., Co-operative non-enzymatic base recognition III. Kinetics of the helix—coil transition of the oligoribouridylic· oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **1971,** *62*, 361-381.

105. Schickinger, M.; Zacharias, M.; Dietz, H., Tethered multifluorophore motion reveals equilibrium transition kinetics of single DNA double helices. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115*, E7512-E7521.

106. Hart, D. J.; Jeong, J.; Gumbart, J. C.; Kim, H. D., Weak tension accelerates hybridization and dehybridization of short oligonucleotides. *Nucleic Acids Res.* **2023,** *51*, 3030-3040.

107. Chen, Y.-I.; Chang, Y.-J.; Nguyen, T. D.; Liu, C.; Phillion, S.; Kuo, Y.-A.; Vu, H. T.; Liu, A.; Liu, Y.-L.; Hong, S., Measuring DNA hybridization kinetics in live cells using a time-resolved 3D single-molecule tracking method. *J. Am. Chem. Soc.* **2019,** *141*, 15747-15750.

108. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **1953,** *171*, 737-738.

109. Franklin, R. E.; Gosling, R. G., Molecular configuration in sodium thymonucleate. *Nature* **1953,** *171*, 740-741.

110. Creeth, J.; Gulland, J. M.; Jordan, D., 214. Deoxypentose nucleic acids. Part III. Viscosity and streaming birefringence of solutions of the sodium salt of the deoxypentose nucleic acid of calf thymus. *J. Chem. Soc.* **1947**, 1141-1145.

111. Rice, S. A.; Doty, P., The thermal denaturation of desoxyribose nucleic acid. *J. Am. Chem. Soc.* **1957,** *79*, 3937-3947.

112. Marmur, J.; Doty, P., Thermal renaturation of deoxyribonucleic acids. *J. Mol. Biol.* **1961,** *3*, 585-594.

113. Eigen, M., Methods for investigation of ionic reactions in aqueous solutions with half-times as short as 10–9 sec. Application to neutralization and hydrolysis reactions. *Discuss. Faraday Soc.* **1954,** *17*, 194-205.

114. Pörschke, D.; Uhlenbeck, O. C.; Martin, F. H., Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs. *Biopolymers* **1973,** *12*, 1313-1335.

115. Applequist, J.; Damle, V., Theory of the Effects of Concentration and Chain Length on Helix—Coil Equilibria in Two-Stranded Nucleic Acids. *J. Chem. Phys.* **1963,** *39*, 2719-2721.

116. Crothers, D.; Kallenbach, N. R.; Zimm, B., The melting transition of low-molecular-weight DNA: Theory and experiment. *J. Mol. Biol.* **1965,** *11*, 802-820.

117. Gibbs, J.; DiMarzio, E., Statistical Mechanics of Helix-Coil Transitions in Biological Macromolecules. *J. Chem. Phys.* **1959,** *30*, 271-282.

118. Kim, J.-Y.; Chung, H. S., Disordered proteins follow diverse transition paths as they fold and bind to a partner. *Science* **2020,** *368*, 1253-1257.

119. Rauzan, B.; McMichael, E.; Cave, R.; Sevcik, L. R.; Ostrosky, K.; Whitman, E.; Stegemann, R.; Sinclair, A. L.; Serra, M. J.; Deckert, A. A., Kinetics and thermodynamics of DNA, RNA, and hybrid duplex formation. *Biochem.* **2013,** *52*, 765-772.

120. Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Minnes, L.; Towrie, M.; Parker, A. W.; Hunt, N. T., Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 6188-6199.

121. Ding, F.; Cocco, S.; Raj, S.; Manosas, M.; Nguyen, T. T. T.; Spiering, M. M.; Bensimon, D.; Allemand, J.-F.; Croquette, V., Displacement and dissociation of oligonucleotides during DNA hairpin closure under strain. *Nucleic Acids Res.* **2022,** *50*, 12082-12093.

122. Hoose, A.; Vellacott, R.; Storch, M.; Freemont, P. S.; Ryadnov, M. G., DNA synthesis technologies to close the gene writing gap. *Nat. Rev. Chem.* **2023,** *7*, 144-161.

123. Hata, H.; Kitajima, T.; Suyama, A., Influence of thermodynamically unfavorable secondary structures on DNA hybridization kinetics. *Nucleic Acids Res.* **2018,** *46*, 782-791.

124. Xiao, S.; Sharpe, D. J.; Chakraborty, D.; Wales, D. J., Energy landscapes and hybridization pathways for DNA hexamer duplexes. *J. Phys. Chem. Lett.* **2019,** *10*, 6771-6779.

125. Niranjani, G.; Murugan, R., Theory on the mechanism of DNA renaturation: stochastic nucleation and zipping. *PLOS One* **2016,** *11*, e0153172.

126. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. *Proc. Natl. Acad. Sci. U.S.A.* **2023,** *120*, e2219124120.

127. Ouldridge, T. E.; Louis, A. A.; Doye, J. P., Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011,** *134*, 02B627.

128. Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J., An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013,** *139*, 10B604_1.

129. Markegard, C. B.; Fu, I. W.; Reddy, K. A.; Nguyen, H. D., Coarse-grained simulation study of sequence effects on DNA hybridization in a concentrated environment. *J. Phys. Chem. B* **2015,** *119*, 1823-1834.

130. Chakraborty, D.; Hori, N.; Thirumalai, D., Sequence-dependent three interaction site model for single-and double-stranded DNA. *J. Chem. Theory Comput.* **2018,** *14*, 3763-3779.

131. Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A., DNA duplex formation with a coarse-grained model. *J. Chem. Theory Comput.* **2014,** *10*, 5020-5035.

132. Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J., Martini coarse-grained force field: extension to DNA. *J. Chem. Theory Comput.* **2015,** *11*, 3932-3945.

133. Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **2013,** *41*, 10641-10658.

134. Schreck, J. S.; Ouldridge, T. E.; Romano, F.; Šulc, P.; Shaw, L. P.; Louis, A. A.; Doye, J. P., DNA hairpins destabilize duplexes primarily by promoting melting rather than by inhibiting hybridization. *Nucleic Acids Res.* **2015,** *43*, 6181-6190.

135. Jin, R.; Maibaum, L., Mechanisms of DNA hybridization: Transition path analysis of a simulation-informed Markov model. *J. Chem. Phys.* **2019,** *150*.

136. Neupane, K.; Hoffer, N. Q.; Woodside, M., Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* **2018,** *121*, 018102.

137. Hoffer, N. Q.; Neupane, K.; Woodside, M. T., Measuring the average shape of transition paths during the folding of a single biological molecule. *Proc. Natl. Acad. Sci. U.S.A.* **2019,** *116*, 8125-8130.

138. Hoffer, N. Q.; Neupane, K.; Woodside, M. T., Observing the base-by-base search for native structure along transition paths during the folding of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U.S.A.* **2021,** *118*, e2101006118.

139. Truex, K.; Chung, H. S.; Louis, J. M.; Eaton, W. A., Testing landscape theory for biomolecular processes with single molecule fluorescence spectroscopy. *Phys. Rev. Lett.* **2015,** *115*, 018101.

140. Neupane, K.; Manuel, A. P.; Lambert, J.; Woodside, M. T., Transition-path probability as a test of reaction-coordinate quality reveals DNA hairpin folding is a one-dimensional diffusive process. *J. Phys. Chem. Lett.* **2015,** *6*, 1005-1010.

141. Neupane, K.; Manuel, A. P.; Woodside, M. T., Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat. Phys.* **2016,** *12*, 700-703.

142. Zheng, W.; Best, R. B., Reduction of all-atom protein folding dynamics to one-dimensional diffusion. *J. Phys. Chem. B* **2015,** *119*, 15247-15255.

143. Bastos, M.; Castro, V.; Mrevlishvili, G.; Teixeira, J., Hydration of ds-DNA and ss-DNA by neutron quasielastic scattering. *Biophys. J.* **2004,** *86*, 3822-3827.

144. McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B., DNA's chiral spine of hydration. *ACS Cent. Sci.* **2017,** *3*, 708-714.

145. Denisov, V. P.; Carlström, G.; Venu, K.; Halle, B., Kinetics of DNA hydration. *J. Mol. Biol.* **1997,** *268*, 118-136.

146. Son, I.; Shek, Y. L.; Dubins, D. N.; Chalikian, T. V., Hydration changes accompanying helix-to-coil DNA transitions. *J. Am. Chem. Soc.* **2014,** *136*, 4040-4047.

147. Nakano, M.; Tateishi-Karimata, H.; Tanaka, S.; Tama, F.; Miyashita, O.; Nakano, S.-i.; Sugimoto, N., Local thermodynamics of the water molecules around single-and double-stranded DNA studied by grid inhomogeneous solvation theory. *Chem. Phys. Lett.* **2016,** *660*, 250-255.

148. Schoen, I.; Krammer, H.; Braun, D., Hybridization kinetics is different inside cells. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 21649-21654.

149. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C., Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Cent. Sci.* **2022,** *8*, 1404-1414.

150. Giambasu, G. M.; Case, D. A.; York, D. M., Predicting site-binding modes of ions and water to nucleic acids using molecular solvation theory. *J. Am. Chem. Soc.* **2019,** *141*, 2435-2445.

151. Jang, S. S.; Dubnik, S.; Hon, J.; Hellenkamp, B.; Lynall, D. G.; Shepard, K. L.; Nuckolls, C.; Gonzalez Jr, R. L., Characterizing the conformational free-energy landscape of RNA stem-loops using single-molecule field-effect transistors. *J. Am. Chem. Soc.* **2022,** *145*, 402-412.

152. Allawi, H. T.; SantaLucia, J., Nearest-Neighbor Thermodynamics of Internal A⊙ C Mismatches in DNA: Sequence Dependence and pH Effects. *Biochem.* **1998,** *37*, 9435-9444.

153. Sági, J.; Guliaev, A. B.; Singer, B., 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines. *Biochem.* **2001,** *40*, 3859-3868.

154. Singh, S. K.; Szulik, M. W.; Ganguly, M.; Khutsishvili, I.; Stone, M. P.; Marky, L. A.; Gold, B., Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Res.* **2011,** *39*, 6789-6801.

155. Roost, C.; Lynch, S. R.; Batista, P. J.; Qu, K.; Chang, H. Y.; Kool, E. T., Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* **2015,** *137*, 2107-2115.

156. Hardwick, J. S.; Lane, A. N.; Brown, T., Epigenetic modifications of cytosine: biophysical properties, regulation, and function in mammalian DNA. *BioEssays* **2018,** *40*, 1700199.

157. Yoshida, T.; Morihiro, K.; Naito, Y.; Mikami, A.; Kasahara, Y.; Inoue, T.; Obika, S., Identification of nucleobase chemical modifications that reduce the hepatotoxicity of gapmer antisense oligonucleotides. *Nucleic Acids Res.* **2022,** *50*, 7224-7234.

158. E Wang, R.; Wu, H.; Niu, Y.; Cai, J., Improving the stability of aptamers by chemical modification. *Curr. Med. Chem.* **2011,** *18*, 4126-4138.

159. Kondo, J.; Yamada, T.; Hirose, C.; Okamoto, I.; Tanaka, Y.; Ono, A., Crystal Structure of Metallo DNA Duplex Containing Consecutive Watson–Crick-like T–HgII–T Base Pairs. *Angew. Chem. Int. Edit.* **2014,** *53*, 2385-2388.

160. Geacintov, N. E.; Broyde, S., Repair-resistant DNA lesions. *Chem. Res. Toxicol.* **2017,** *30*, 1517-1548.

161. Hardwick, J. S.; Ptchelkine, D.; El-Sagheer, A. H.; Tear, I.; Singleton, D.; Phillips, S. E.; Lane, A. N.; Brown, T., 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **2017,** *24*, 544-552.

162. Naiser, T.; Kayser, J.; Mai, T.; Michel, W.; Ott, A., Position dependent mismatch discrimination on DNA microarrays–experiments and model. *BMC Bioinformatics* **2008,** *9*, 1-12.

163. De Marco, L. The molecular dynamics of hydrogen-bonding explored with broadband two dimensional infrared spectroscopy. Massachusetts Institute of Technology, 2016.

164. Khalil, M.; Demirdöven, N.; Tokmakoff, A., Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution. *J. Phys. Chem. A* **2003,** *107*, 5258-5279.

Chapter 2

Spectroscopy of Nucleic Acid Structure and Dynamics

2.1 Tools for probing structural dynamics

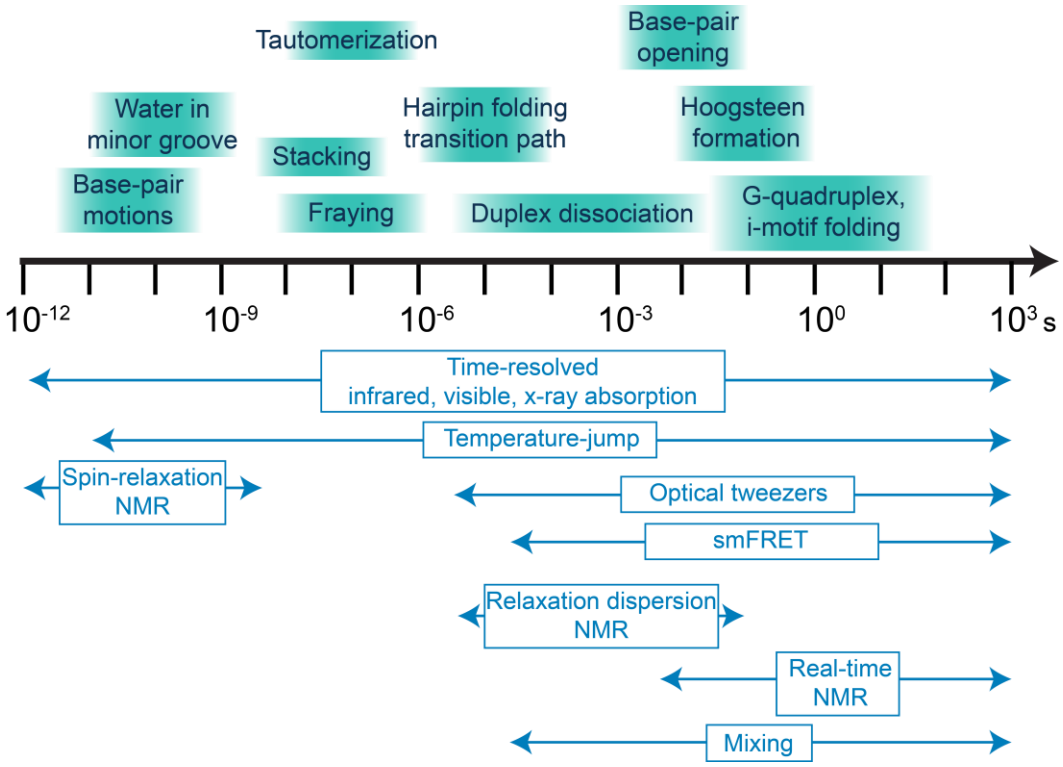

Figure 2.1 Timescales of nucleic acid motions and accessibility by select time-resolved experiments. Optical spectroscopy may be coupled with perturbations such as photoexcitation and temperature-jumps to probe structural motions occurring from picoseconds to hours, but these methods suffer from limited structural information and cannot directly probe dynamics of barrier crossing. Single-molecule methods can visualize trajectories of barrier crossing, but are often limited by time resolution ($\geq$ 10 μs).

As introduced in Chapter 1, nucleic acids and proteins exhibit a hierarchy of structural dynamics spanning from local motions on ultrafast (fs – ps) timescales to large-scale rearrangement and assembly that can take up to hours. Unfortunately, no single experiment can

probe all of these dynamics at once, and high-quality biophysics studies instead resort to combining a series of experimental tools to characterize biomolecular dynamics. The most common reason for employing multiple probes is the experimental tradeoff between structure resolution and time resolution (Fig. 2.1). Methods such as cryogenic electron microscopy (cryo-EM),[1] x-ray crystallography,[2] and nuclear magnetic resonance (NMR)[3] offer unmatched structural information but are usually limited to motions occurring on millisecond timescales or slower.[4-6] Time-resolved X-ray scattering and diffraction methods are showing promise for nanosecond-to-millisecond structural dynamics, but are still in development.[7-8] In contrast, methods capable of accessing ultrafast timescales such as UV absorption, circular dichroism, vibrational spectroscopy, and fluorescence provide relatively limited structural information.

Coverage of a wide range of timescales does not ensure accessibility to the true dynamics of a structural process. As stated in Chapter 1, our goal is to visualize how a molecule transforms from one state to another through what is called a transition path (the actual barrier-crossing event). This can only be achieved by monitoring time trajectories of single molecules. Single-molecule experiments have directly revealed insight into transition paths of protein folding, protein-protein association, and DNA hairpin folding.[9-10] However, these methods suffer from limited time resolution ($\geq$ 10 μs), lack of local structural information, and require potentially perturbative immobilization and chemical labeling of the sample.[11-13] Although single-molecule methods are powerful for probing biomolecular dynamics, they were not employed for the work in this thesis and their discussion will be limited.

Ensemble experiments, which make up the entirety of this thesis, measure a signal originating from 10^6 – 10^{12} molecules that averages out the trajectories and dynamic information of individual molecules. These experiments offer accessibility to the fastest molecular motions, but provide indirect insight into the underlying structural dynamics relative to single-molecule methods. In almost all cases, the dynamics of interest must be synchronized by applying an external perturbation such as a photoexcitation or temperature change. For a transition between two states separated by a large free-energy barrier – a common scenario for nucleic acids and proteins – only the change in the population of states is observed as a function of time. The spectroscopic probe then reports how long molecules remain in a given state rather than watching the movement of molecules from one state to the other. If stable or metastable configurations exist in between basins or off-pathway, they can often be resolved as well. Ensemble methods can most directly access dynamics of energetically downhill conformational motion, in which the movement of population across the free-energy landscape may be visible. Perturbation-induced downhill dynamics is most easily achieved for photophysical and photochemical phenomena where relaxation from an initially excited high-energy configuration may proceed partially or fully downhill.[14] It is more difficult to promote downhill structural dynamics in the electronic ground state (e.g. structural motions of nucleic acids),[15-17] and the approach for doing so may bias the observed dynamics from its native mechanism.[18]

This chapter summarizes the advantages and disadvantages of employing infrared and NMR spectroscopy to investigate structural motions in nucleic acids. Many details of nonlinear infrared spectroscopy and their application to nucleic acids have previously been described within

our group,[19-22] and comparison with modern time-resolved NMR methods[23-24] should assist in identifying which method is most appropriate for the problem at hand.

2.2 Infrared spectroscopy of nucleic acids

2.2.1 FTIR spectroscopy of nucleobases and secondary structure

Infrared (IR) spectroscopy probes molecular vibrations that offer sensitivity to local nucleic acid structure as well as ultrafast intrinsic time resolution for monitoring the time-dependence of structural changes. As a result, IR spectroscopy resolves rapidly exchanging species (tautomers, oligonucleotide strands, etc.) as separate sets of absorption peaks weighted by their population and can be implemented in time-resolved approaches to track real-time kinetics down to hundreds of femtosecond timescales.[25-28] Time-resolved IR spectroscopy is discussed in Chapters 3 and 4, and here we summarize the structural sensitivity of IR spectroscopy to nucleic acid structural changes. Nucleic acids contain numerous types of chemical groups that lead to absorption across much of the mid-IR spectral region (1000 – 4000 cm^{-1} or 10 – 2.5 μm).[29-30] Each of these groups reports on different aspects of nucleic acid structure. For instance the asymmetric phosphate stretch at 1200 – 1250 cm^{-1} is sensitive to the structure and electrostatic environment of the phosphate backbone.[31] Base-ribose and low-frequency ring vibrations from 1300 to 1450 cm^{-1} are sensitive to ribose configuration and glycosidic bond orientation and can distinguish A, B, and Z duplex conformations, C2′-endo vs. C3′-endo puckering conformations, and anti vs. syn sugar-base configurations.[22, 29-30, 32] Higher frequency N-H and C-H stretch vibrations are also sensitive to nucleic acid structural changes but are more difficult to study.[33-34]

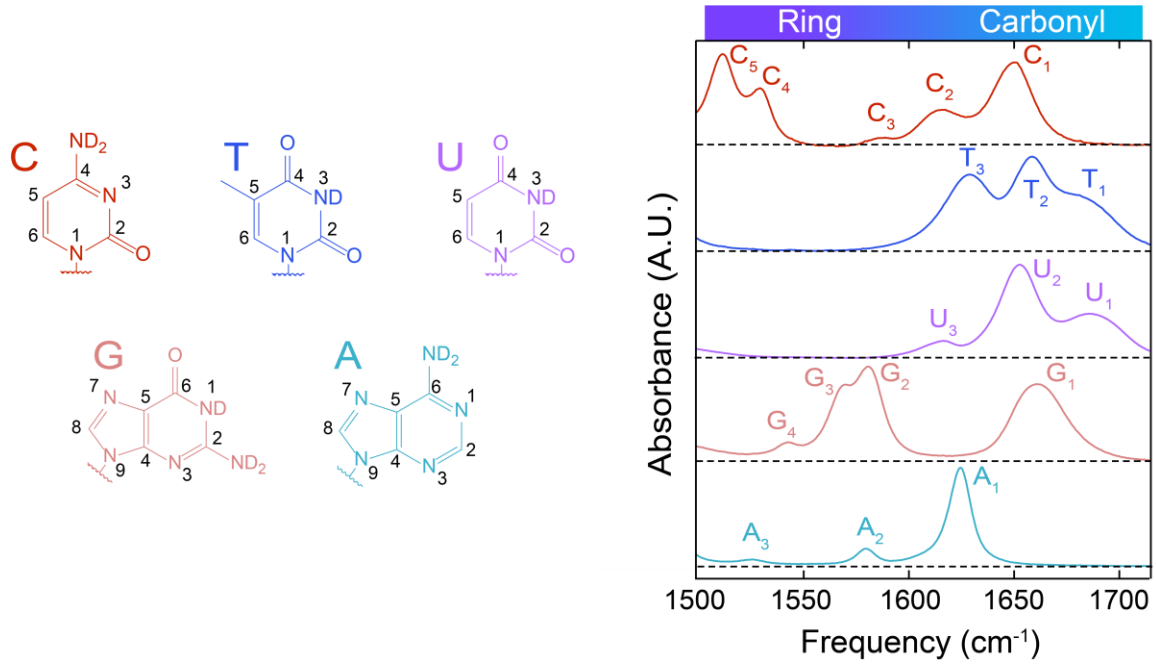

Figure 2.2 Free nucleotide FTIR spectra. (Left) Chemical structures of nucleobases. (Right) FTIR spectra of cytidine (C), thymidine (T), uridine (U), guanosine (G), adenosine (A) 5′-monophosphates in deuterated pH 7 PBS. FTIR bands are labeled X_n for each nucleobase. Adapted from Peng et al.[35]

The work in this thesis exclusively utilizes the in-plane ring and carbonyl stretching vibrations of the nucleobases (1450 to 1750 cm^{-1}). These are among the brightest DNA vibrational bands (molar absorption coefficient ε = 500 – 800 M^{-1} cm^{-1}) and are highly sensitive to changes in base pairing and stacking as well as chemical modifications to the nucleobase. Figure 2.2 shows FTIR spectra of each nucleotide in deuterated aqueous solution. There are three to five vibrational bands in the 1500 – 1750 cm^{-1} spectral region depending on the nucleobase substituents and ring structure. In general, the vibrations are highly delocalized with a combination of C=O, C=N, C=C stretching and N-H bending, and there is no clean separation of carbonyl-only and ring-only vibrations.[35] A possible exception is the guanosine carbonyl stretch (G1), which is relatively

localized to the C^6=O carbonyl group. The spectra instead transition from carbonyl-dominant (1640 – 1750 cm^{-1}) to ring-dominant (1500 – 1640 cm^{-1}) as indicated in Table 2.1.

Table 2.1 IR vibrational mode assignments for nucleotides. Experimental FTIR peak frequency and contributions to vibrational modes determined from DFT frequency calculations of solvated nucleotides. ν indicates stretching and δ indicates bending. Py and Im correspond to vibrations on the pyrimidine and imidazole ring, respectively. Adapted from Peng et al.[35]

Base	Peak	Freq. (cm^{-1})	Contributions
C	C1	1651	$\nu(C^2{=}O)$, $\nu(N^1{=}C^6)$, $\delta(C^6{-}H)$
	C2	1614	$\nu(N^3{=}C^4{-}C^5{=}C^6)$, $\nu(C^2{=}O)$, $\delta(C^5{-}H)$, $\delta(C^6{-}H)$
	C3	1583	-
	C4	1524	$\nu(C^4{-}C^5)$, $\nu(N^1{-}C^6)$, $\delta(C^5{-}H)$, $\delta(C^6{-}H)$, $\delta(N^4D_2)$
	C5	1504	$\nu(N^3{=}C^4{-}N^4)$, $\nu(C^5{=}C^6)$, $\delta(C^5{-}H)$, $\delta(C^6{-}H)$, $\delta(N^4D_2)$
T	T1	1690	$\nu(C^2{=}O)$, $\nu(C^4{=}O)$, $\delta(N^3{-}D)$
	T2	1663	$\nu(C^4{=}O)$, $\nu(C^5{=}C^6)$, $\delta(N^3{-}D)$, $\delta(C^5H_3)$, $\delta(C^6{-}H)$
	T3	1629	$\nu(C^5{=}C^6)$, $\nu(C^4{=}O)$, $\delta(C^5H_3)$, $\delta(C^6$-H)
U	U1	1693	ν(sym C=O), $\nu(C^5{=}C^6)$, $\delta(N^3{-}D)$, $\delta(C^5{-}H)$
	U2	1655	ν(asym C=O), $\nu(C^5{=}C^6)$, $\delta(C^5{-}H)$, $\delta(C^6{-}H)$, $\delta(N^3{-}D)$
	U3	1617	$\nu(C^5{=}C^6)$,ν(asym C=O), $\delta(C^5{-}H)$, $\delta(C^6{-}H)$, $\delta(N^3{-}D)$
G	G1	1665	$\nu(C^6{=}O)$, $\delta(N^1$-D), Py
	G2	1579	$\nu(C^2{=}N^3)$, $\nu(C^6{=}O)$, $\delta(N^1{-}D)$, $\delta(N^2D_2)$ Py
	G3	1565	$\nu(C^2{=}N^3{-}C^4{=}C^5)$, $\delta(C^8{-}H)$, Py + Im
	G4	1539	$\nu(C^4{=}C^5)$, $\nu(N^7{=}C^8)$, $\delta(C^8{-}H)$, Py + Im
A	A1	1625	$\nu(C^4{=}C^5$, $C^5{-}C^6$ out-of-phase), $\delta(C^2$-H), $\delta(N^6D_2)$, Py
	A2	1578	$\nu(C^4{=}C^5$, $C^5{-}C^6$ in-phase), $\nu(N^1{-}C^6)$, $\nu(N^3{-}C^4)$, $\nu(N^7{=}C^8)$, $\delta(C^8{-}H)$, Py + Im
	A3	1525	-

The distinct mid-IR fingerprint of each nucleobase enables IR spectroscopy to provide base-specific information in the study of oligonucleotides. In the duplex, guanine and cytosine vibrations, and to a lesser extent adenine and thymine/uracil vibrations, are coupled through base pairing.[26, 36-38] Figure 2.3 shows FTIR spectra of self-complementary 5′-ATATATATAT-3′, $(AT)_5$, and 5′-GCGCGCGC-3′, $(GC)_4$, oligonucleotides where each sample is fully duplexed at 1 °C and dissociated into single strands at 90 °C. Numerous IR bands are sensitive to changes in base pairing and stacking. Even though the character of the vibrations are different in the duplex

than in free nucleotides, the nucleotide notation in Fig. 2.2 is used for simplicity. A:T duplex formation primarily induces intensity changes in the 1600 – 1750 cm^{-1} region, most significantly the suppression of A_1 and T_3. Intensity changes also dominate the ring mode region ($G_{2\text{-}4}$ and $C_{4\text{-}5}$) during G:C duplex formation, yet the most notable change is the blueshift of G_1 by ~20 cm^{-1} (G_{1*}) that may result from coupling to intrabase ring and cytosine vibrations.[39] The intensity changes of the $C_{4\text{-}5}/G_{2\text{-}3}$ and A_1/T_3 ring modes are well isolated, enabling use as markers for A:T and G:C base pairing, respectively, even in mixed duplexes (Fig. 2.3b). It is worth noting that while these changes are easy to isolate experimentally, they are more difficult to interpret and reproduce computationally than the shifts in the carbonyl region.[36, 40] As a result, direct connections between experimental and computational measurements of nucleic acid structural dynamics using vibrational spectroscopic maps have yet to be achieved. It is also important to note that ^{13}C, ^{15}N, and ^{18}O isotope-labeling may be employed to measure nucleobase- and bond-specific IR spectra, which may be more suitable to accurate spectral simulation.[36, 41-42]

The in-plane ring and carbonyl vibrations depend on the nature of hydrogen bonding and stacking interactions and therefore show sensitivity to different types of secondary and tertiary structure. First, we may consider the comparison of DNA and RNA duplexes. As an example, FTIR spectra of G:C-rich gap complexes are shown in Fig. 2.3c. Small differences between DNA and RNA may arise in the 1620 – 1720 cm^{-1} region from replacement of two thymines in DNA to uracils in RNA, but the most notable spectral differences occur in the G/C ring and carbonyl bands. The spectral differences between DNA and RNA most likely stem from their different duplex conformation. DNA is typically B-form (rise per base pair ~ 3.4 Å) while RNA is the more compact A-form (rise ~ 2.9 Å), which has been shown to alter the intrastrand coupling between nucleobase

vibrations.[43] Indeed, spectral simulations of G:C-rich A-form DNA show a blueshift and broadening of G_{1*} relative to B-form DNA as shown in Fig. 2.3c, but these simulations do not reproduce the enhanced intensity of G_3 observed in RNA.[43] The FTIR spectra of single-stranded DNA and RNA at 90 °C are nearly the same, suggesting similar stacking structure.

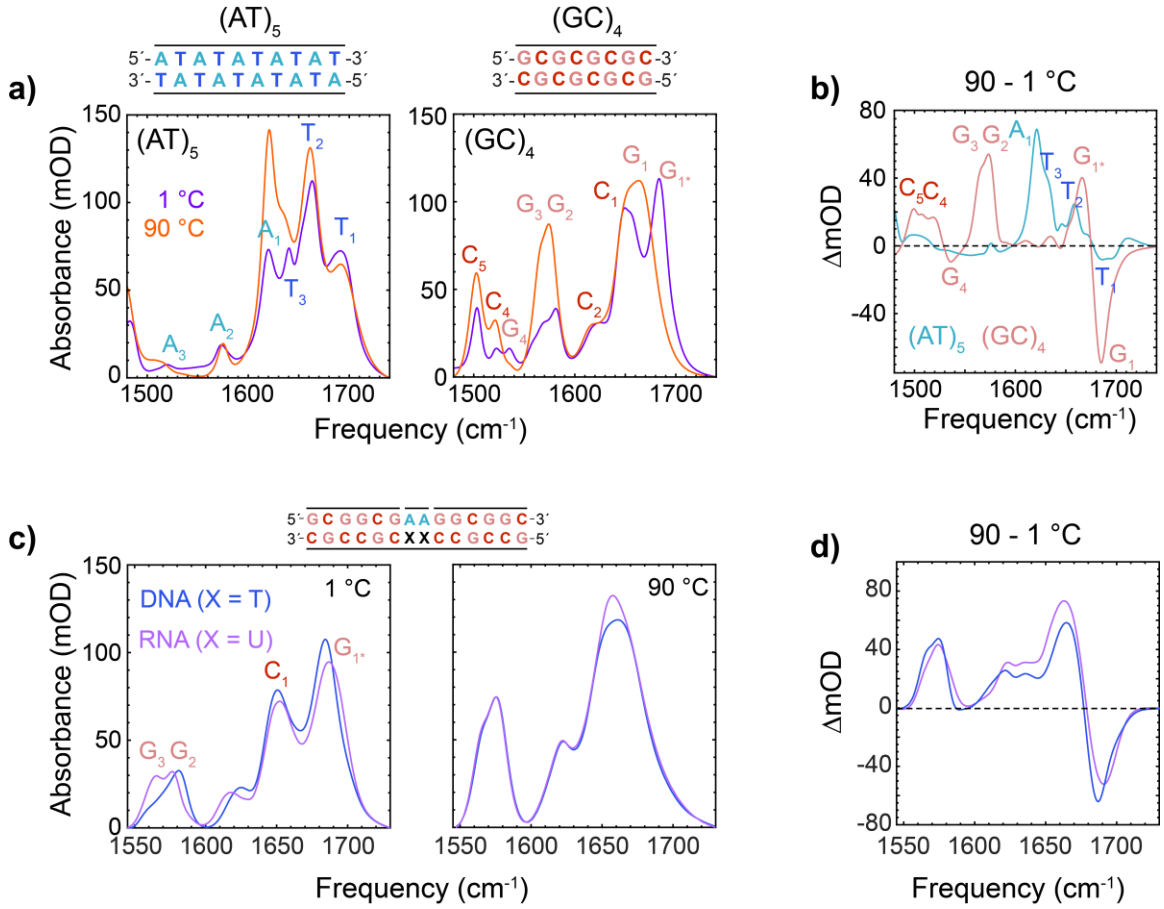

Figure 2.3 FTIR spectral changes of DNA and RNA duplex melting. (a) FTIR spectra of 5′-ATATATATAT-3′, $(AT)_5$, and 5′-GCGCGCGC-3′, $(GC)_4$, at 1 °C (duplex) and 90 °C (single-strand). Nucleobase vibrations from Fig. 2.2 are indicated. Oligonucleotides were prepared at 4 mM in pH* 6.8 40 mM sodium phosphate buffer (SPB). **(b)** Difference spectra between 90 and 1 °C for each sequence. **(c)** 1 °C (left) and 90 °C (right) spectra of DNA and RNA gap duplex complex. **(d)** Difference spectra between 90 and 1 °C.

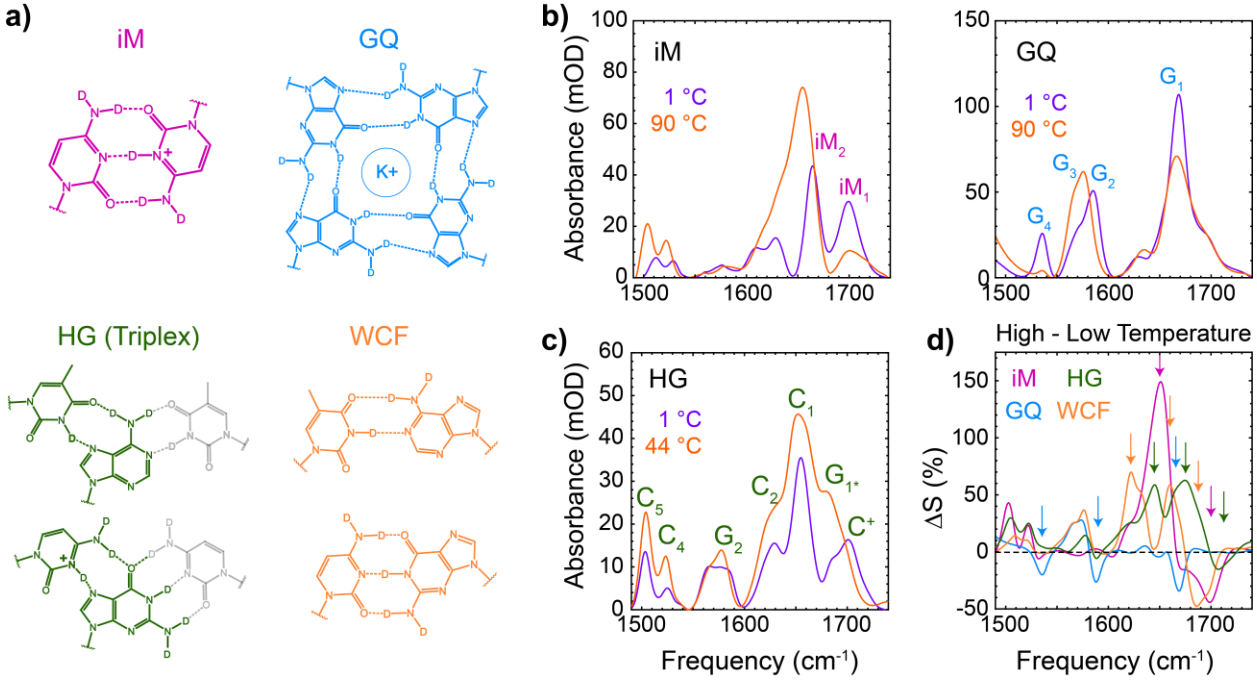

Figure 2.4 FTIR spectroscopy of different types of base pairs. (a) Chemical structures of C:C base pairing in an i-motif (iM), guanine-quartet base pairing in a G-quadruplex (GQ), Hoogsteen (HG) base pairing in a triplex, and Watson-Crick-Franklin (WCF) base pairing in a B-DNA duplex. **(b)** FTIR spectra of folded (1 °C) and unfolded (90 °C) iM and GQ strands. The iM sequence is 5′-CCCTTTCCCTTTCCCTTTCCC-3′ and the GQ sequence is 5′-GGTTGGTGTGGTTGG-3′. **(c)** Spectra of a G:C-rich intramolecular DNA triplex with the HG segment folded (1 °C) and unfolded (44 °C). The HG base pairing segment is CCTCCCCTCC/GGAGGGGAGG. **(d)** Difference spectra between high and low temperatures for the strands in **(b, c)** and a DNA duplex of 5′-CCTATATATCC-3′ + complement (WCF). Data are plotted in percent change relative to the 1 °C spectrum, $\Delta S = [S(90) - S(1)]/\max(S(1))$. Color-coded arrows indicate key difference features for distinguishing species in a mixture of structures.

Figure 2.4 illustrates how the in-plane ring and carbonyl vibrations are distinctly sensitive to base pairing geometry using the examples of C:C base pairing in an i-motif (iM), guanine-quartet in G-quadruplex, and G:C Hoogsteen (HG) base pairing in triplex DNA to compare with Watson-Crick-Franklin (WCF) base pairing in duplex DNA. There are other types of base pairing that may exhibit distinct behavior, but these examples are chosen to illustrate the

sensitivity of nucleobase vibrations to common secondary structures. In the iM, base pairing between cytosines is mediated by N3 protonation such that three hydrogen bonds are formed in the base pair, and the C:C base pair energy is larger than that of WCF G:C and A:T.[44] N3 protonation partially breaks up the delocalized nature of vibrations in cytosine, producing a vibrational band at ~1710 cm^{-1} that is dominated by C^2=O stretching motion (iM_1) and altering the character of the other cytosine ring modes. Although C:C base pair formation influences the width and frequency of the C^2=O stretch, the high-frequency band generally arises from N3 protonation in free nucleotides and G:C HG base pairing (C_*, Fig. 2.4c).[45] There is also a large reduction in the intensity of the band near 1650 cm^{-1}, which is a distinct feature of C:C base pairing. Guanine-quartet formation does not produce any new vibrational bands but does alter the guanine vibrations distinctly from WCF base pairing. Guanines in the quartet participate in hydrogen bonding through both the N1 and N7 faces while coordinated around a monovalent cation of the correct size (usually K^+). This leads to large changes in the ring vibrations due to varying degrees of delocalization across the quartet. Interestingly, G_1 does not blueshift as observed in WCF base pairing, and previous analysis shows it corresponds to asymmetric stretching of C^6=O in diagonal monomers.[22] Potentially the most distinct feature of GQs is the enhanced intensity of G_4 relative to WCF base pairing, which comes from increased coupling to the C^6=O stretch on the same monomer. As illustrated in the melting difference spectra in Fig. 2.4d, the distinct spectral changes associated with each of these structures allows IR spectroscopy to disentangle complicated mixtures of secondary and tertiary structure, which is necessary for studying biologically relevant systems with competition between structures.[46]

2.2.2 Nonlinear IR spectroscopy of DNA duplex melting

Nonlinear vibrational spectroscopy is capable of providing enhanced structural information of nucleic acids relative to FTIR spectroscopy. Such methods include sum-frequency generation for probing interfacial water and nucleic acids at interfaces and two-dimensional IR (2D IR) for probing structural changes in solution.[26, 47-49] This thesis utilizes 2D IR and its 1D projection, pump-probe (PP) spectroscopy, as a structural probe for steady-state and transient methods of nucleic acid folding and hybridization. Details of 2D IR spectroscopy are well documented and its implementation is discussed in Chapter 4,[19-20] so this section will focus on the application to nucleic acids and the basic concepts needed for understanding the spectra.

2D IR is a well-known tool for studying ultrafast (fs – ps) solvent dynamics, chemical exchange, and vibrational energy transfer in solution.[50-51] 2D IR presents several advantages over FTIR. First of all, it spreads the vibrational spectrum over two frequency axes, excitation (ω_1) and detection (ω_3), creating a correlation map between vibrational transitions. The presence or absence of cross-peaks report on coupling and energy flow between vibrations and can determine whether a set of peaks arises from one or multiple chemical species.[22, 25] Each peak consists of a doublet with a positive peak that corresponds to the 0-1 vibrational transition (ground-state bleach and stimulated emission; abbreviated GSB for simplicity) and a negative peak due to the 1-2 vibrational excited-state absorption (ESA). In the nucleic acid spectra, the ESA always appears shifted to lower ω_3 frequencies than the GSB due to normal anharmonicity of the vibrational potentials. Another advantage comes from the proportionality of the 2D IR and PP signals to the fourth order of the transition dipole moment for a given oscillator ($|\mu|^4$) that leads to negligible

2D IR and PP background signal from D_2O in the 1450 – 1750 cm^{-1} region. The linear IR absorption bands are proportional to $|\mu|^2$, so the high concentration and low $|\mu|$ of D_2O leads to an absorbance that is typically ~10 times greater than that of nucleic acids (at ~2 mM). Extracting the nucleic acid linear absorption spectrum requires careful subtraction of the D_2O background, which ultimately limits signal detection at low concentrations. 2D IR spectroscopy may also be employed to monitor the kinetics of ultrafast processes through changes in peak amplitude, lineshape, and anisotropy as a function of waiting time (τ_2),[49-50] but the congested nature of nucleic acid spectra limits the use of lineshape and anisotropy analysis methods.

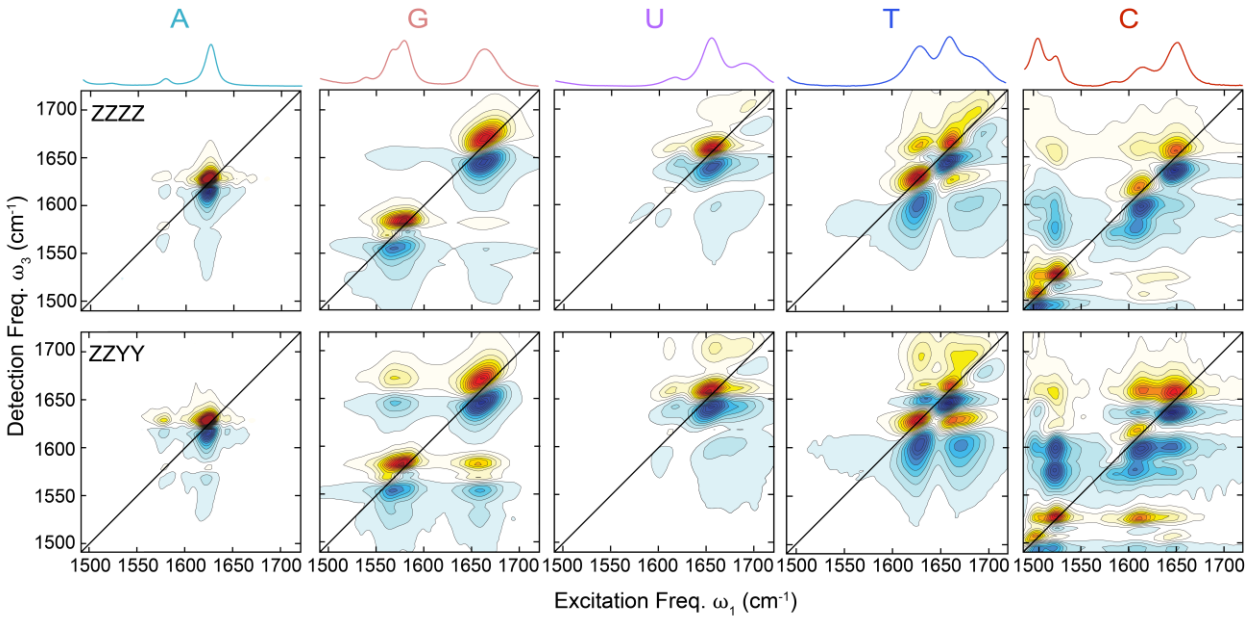

Figure 2.5 2D IR spectra of nucleobases. 2D IR spectra of nucleotide monophosphates at a concentration 20 mg/ml at 20 °C and pH 6.9. Spectra were acquired with parallel (ZZZZ, top) and perpendicular (ZZYY, bottom) pulse polarization. FTIR spectra from Fig. 2.2 are shown atop for each nucleotide. Adapted from Peng et al.[53]

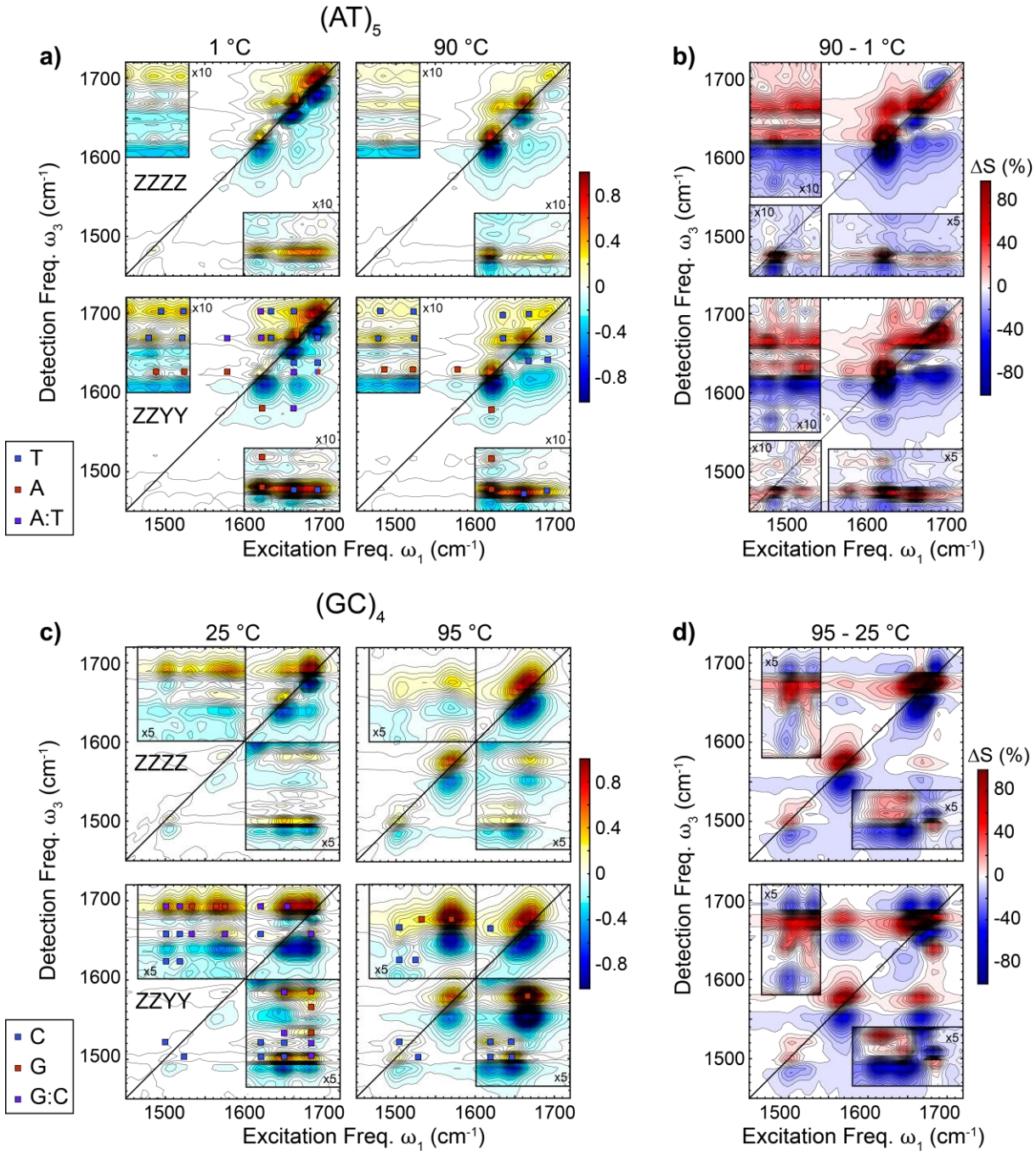

Figure 2.6 2D IR spectroscopy of DNA duplex melting. (a) 2D IR spectra of $(AT)_5$ at 1 °C (duplex) and 90 °C (single-strand) using ZZZZ and ZZYY pulse polarization. Data are collected at a waiting time (τ_2) of 150 fs. Markers in ZZYY spectra indicate intrabase thymine (blue), intrabase adenine (teal), and interbase A:T (purple) cross peaks. **(b)** Difference spectra between 90 and 1 °C plotted in percent change, $\Delta S = [S(90) - S(1)]/\max[S(1)]$. **(c-d)** Plots for $(GC)_4$.

Figure 2.5 shows 2D IR spectra for each nucleotide monophosphate in aqueous solution. Each spectrum is characterized by cross-peaks from vibrational coupling between ring and carbonyl vibrations, providing direct evidence of the delocalized nature of the nucleobase vibrations.[35] The spectra of the nucleobases help interpret those of self-complementary A:T and G:C oligonucleotides shown in Fig. 2.6. The low-temperature duplex spectrum of $(AT)_5$ is characterized by cross-peaks from intrabase vibrational coupling within adenine and thymine nucleobases. Cross-peaks exist at both above ($\omega_3 > \omega_1$) and below ($\omega_3 < \omega_1$) the diagonal, but those below the diagonal are more difficult to observe due to interference with ESAs. There are a few interbase cross-peaks between adenine and thymine due to base pairing that are lost upon thermal dissociation of the duplex, as described previously.[26, 49] Intrabase cross-peak amplitudes change with the amplitude of their diagonal peaks. The low-temperature spectra of $(GC)_4$ exhibit interbase cross peaks between guanine and cytosine modes in the carbonyl region as well as directly between carbonyl and ring vibrations. The cross peaks between G_{1*} and C_4/C_5 are particularly well-isolated from intrabase vibrations and cross peaks in A:T base pairs, allowing robust measurement of G:C base pairing in mixed-sequence duplexes.[26] The employed pulse-polarization scheme modifies the relative amplitude between cross peaks and diagonal peaks and can be used to better isolate peaks of interest in the 2D IR spectra.[19, 39, 52] Figures 2.5 and 2.6 show spectra measured with two of the most common polarization schemes where the pump and probe pulses have either parallel (ZZZZ) and perpendicular polarization (ZZYY). Many cross peaks show an increase in amplitude relative to the diagonal peaks upon switching from ZZZZ to ZZYY, arising from projection angles of 50 – 130° between transition dipoles of coupled nucleobase vibrational modes.[35, 39] Isolation and characterization of cross peaks using ZZYY or other

polarization schemes can provide enhanced insight into the secondary and tertiary structure of biomolecules.[26, 53-54]

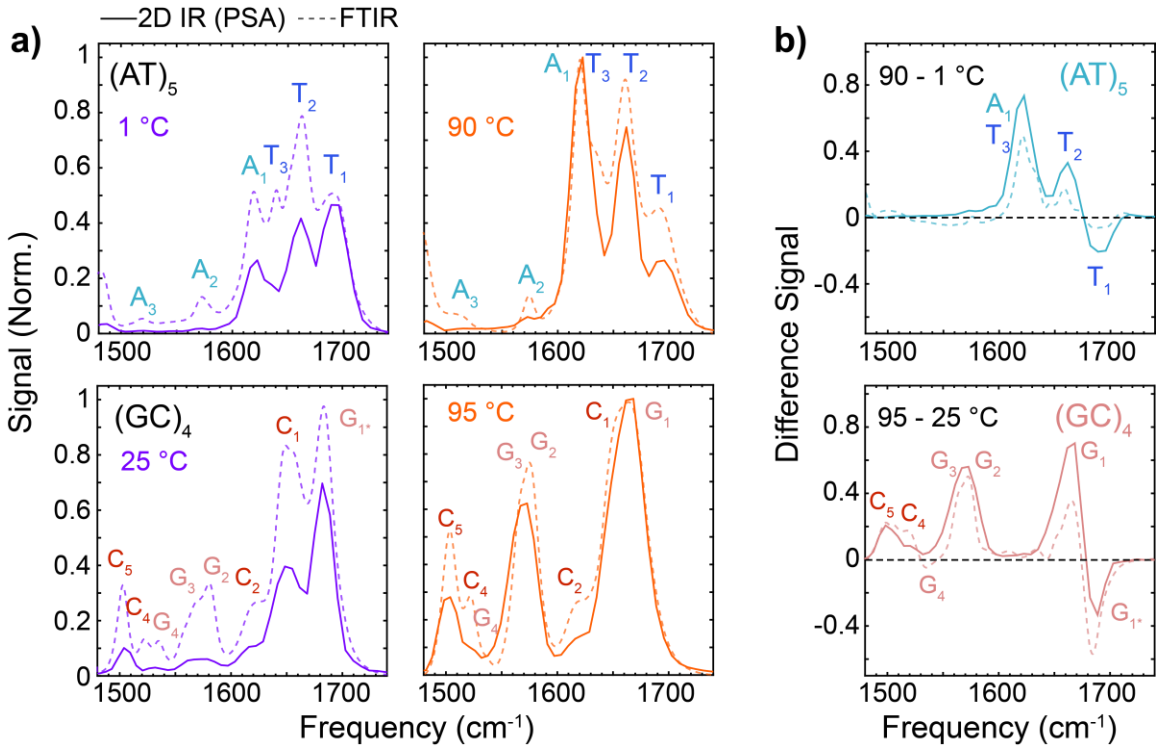

Figure 2.7 Comparison of 2D IR and FTIR spectral changes during DNA duplex melting. (a) 2D IR diagonal slices extracted using the pump-slice amplitude method (PSA)[55] overlaid with FTIR spectra at the same temperature for $(AT)_5$ and $(GC)_4$. Data at each temperature are normalized to the maximum signal of the high-temperature spectrum. **(b)** Difference spectra between the high and low-temperature normalized spectra in **(a)**.

Figure 2.7 illustrates that the spectral changes along the diagonal are distinct from those in FTIR for duplex dissociation. The higher-order proportionality to $|\mu|$ in the 2D IR signal leads to different relative amplitudes among the peaks for both the duplex and single-strand as well as different relative signal change upon duplex melting compared to FTIR. For $(AT)_5$ the signal changes in $T_{1\text{-}3}$ and A_1 are all amplified relative to FTIR, and the behavior of $(GC)_4$ is more

complicated. In 2D IR, there is nearly half the change in G_{1*} and double the change in G_1 relative to FTIR. The normalized increase in $G_{2\text{-}3}$ and $C_{4\text{-}5}$ is similar for each method, but the percent change is greater in 2D IR. Overall, these differences give 2D IR greater contrast between melting and non-melting temperature-dependent changes in the IR spectrum, which provides higher sensitivity to partial duplex melting transition as shown in Chapters 7 and 9.

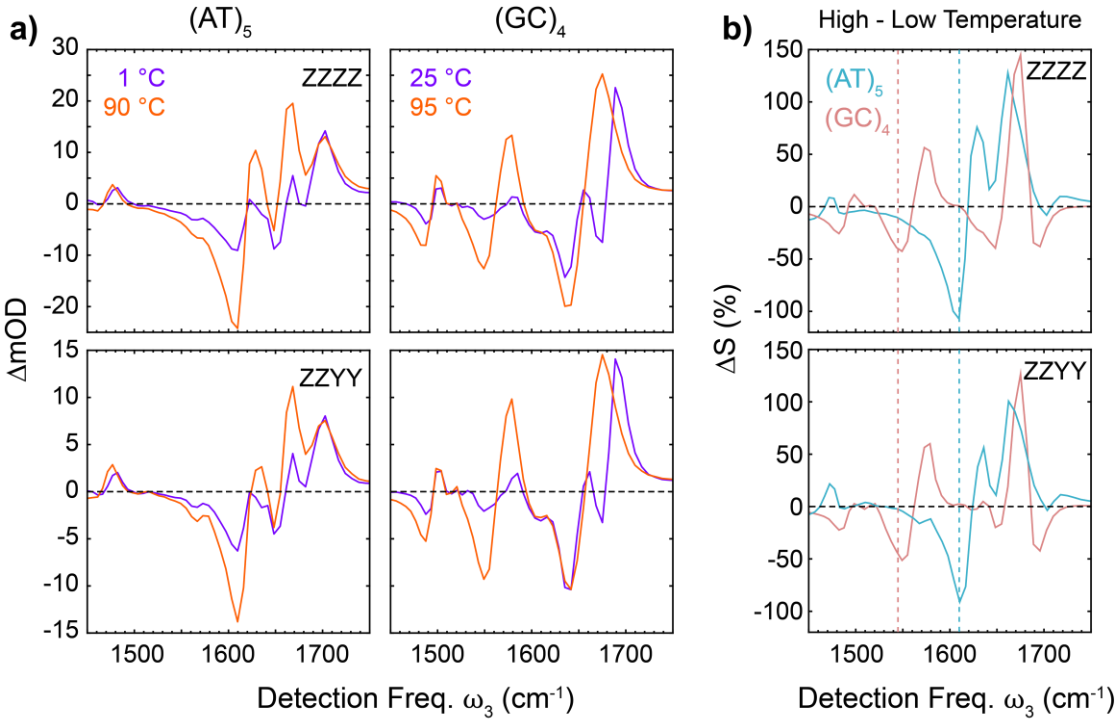

Figure 2.8 Pump-probe spectroscopy of DNA duplex melting. (a) Low and high temperature pump-probe spectra of $(AT)_5$ and $(GC)_4$ collected with parallel (ZZZZ, top) and perpendicular (ZZYY, bottom) pulse polarization at τ_2 = 150 fs. **(b)** Percent change difference spectra for both sequences and polarizations. Dashed lines indicate adenine and guanine ring ESA bands that are used for probing G:C and A:T base pairing, respectively, in mixed-sequence duplexes.

When employing nonlinear IR spectroscopy as a probe for structural dynamics, PP spectroscopy is often employed instead of 2D IR due to its faster acquisition speed.[27, 56-57] The PP spectrum is equivalent to integrating the 2D IR spectrum across ω_1.[58] Therefore, the PP spectrum

still contains the advantages of fourth-order proportionality to the transition dipole moment but does suffer a loss of information on cross-peaks and lineshape relative to 2D IR. Certain polarization schemes may still allow to extract some cross-peak information.[52, 57] Another drawback of PP is the partial overlap of GSB and ESA peaks, which can hide peaks of interest. This is particularly troublesome for nucleic acids and proteins due to the large number of vibrational transitions in the 1500 – 1750 cm^{-1} region. Figure 2.8 shows DNA duplex and single-strand PP spectra collected for $(AT)_5$ and $(GC)_4$, and differences between ZZZZ and ZZYY spectra are subtle relative to the 2D IR spectra (Fig. 2.6). The highest frequency carbonyl GSB contains the least overlap with other bands and is by far the largest positive signal for each duplex. All GSB and ESA peaks from 1550 to 1660 cm^{-1} interfere with others, leading to complicated PP spectra. However, similar to the fundamental transitions in the FTIR spectra, melting-induced signal increase in the G_2/G_3 ESAs are separated from A:T spectral changes, and A_1/T_3 ESAs are separated from G:C spectral changes (dashed lines in Fig. 2.8b). These bands are used as reporters for G:C and A:T base pairing in transient measurements of DNA structural changes.[15, 59]

In Chapter 4, we will discuss how 2D IR may be combined with a synchronization event, such as a temperature change, to probe structural changes that occur on picosecond to minute timescales. Potentially due to technical difficulties, only a few research groups employ such transient 2D IR methods for probing structural processes in biomolecules.[27, 60-61]

2.2.3 Comparison with UV absorption and circular dichroism spectroscopy

UV absorption and circular dichroism (CD) are the most commonly employed spectroscopic probes of nucleic acid conformational changes dating back to the 1950s.[62-63] These

early experiments laid the foundation for nucleic acid biophysics, and the methods still offer some advantages over other methods. First of all, nucleic acid bases have relatively large absorption coefficients from 250 – 280 nm (ε = 5 – 10 × 10^3 $M^{-1}cm^{-1}$) that allow for high-quality measurements with micromolar nucleotide concentrations. Additionally, easy-to-use commercial UV/Vis and CD spectrometers with automated temperature control and commercial data analysis software are available in nearly all research institutions, making the method accessible to researchers from a wide breadth of backgrounds.

Nucleobases contain a combination of ππ* and dark nπ* electronic transitions from 200 – 300 nm. The absorption band(s) from 240 – 300 nm, which are most commonly used for melting studies, correspond to the lowest energy ππ* transitions. Overall, the spectra of each nucleobase have significant spectral overlap, but minor differences arise in pyrimidines versus purines (Fig. 2.9a).[64] The lowest-energy ππ* transition of pyrimidines is well-separated from other bright transitions, leading to a single absorption band from 250 – 300 nm. Purines contain two similar low-energy ππ* transitions, denoted L_a and L_b. The guanine spectrum is distinguished by a two-peak structure arising from significant oscillator strength in each transition whereas only the L_a transition is bright in adenine. Stacking between nucleobases couples transitions on neighboring nucleobases, generating exciton transitions with different energies and oscillator strengths than in the free nucleotides.[65] As a result, the UV spectrum of oligonucleotides reshapes upon duplex melting (Fig. 2.9a,b). For both $(AT)_5$ and $(GC)_4$, melting leads to an increase in absorption with difference features centered at 260 and 280 nm, respectively. Although the A:T and G:C difference

features are not as separated as in FTIR spectroscopy, they still offer the capability to resolve each type of melting in mixed-sequence duplexes.

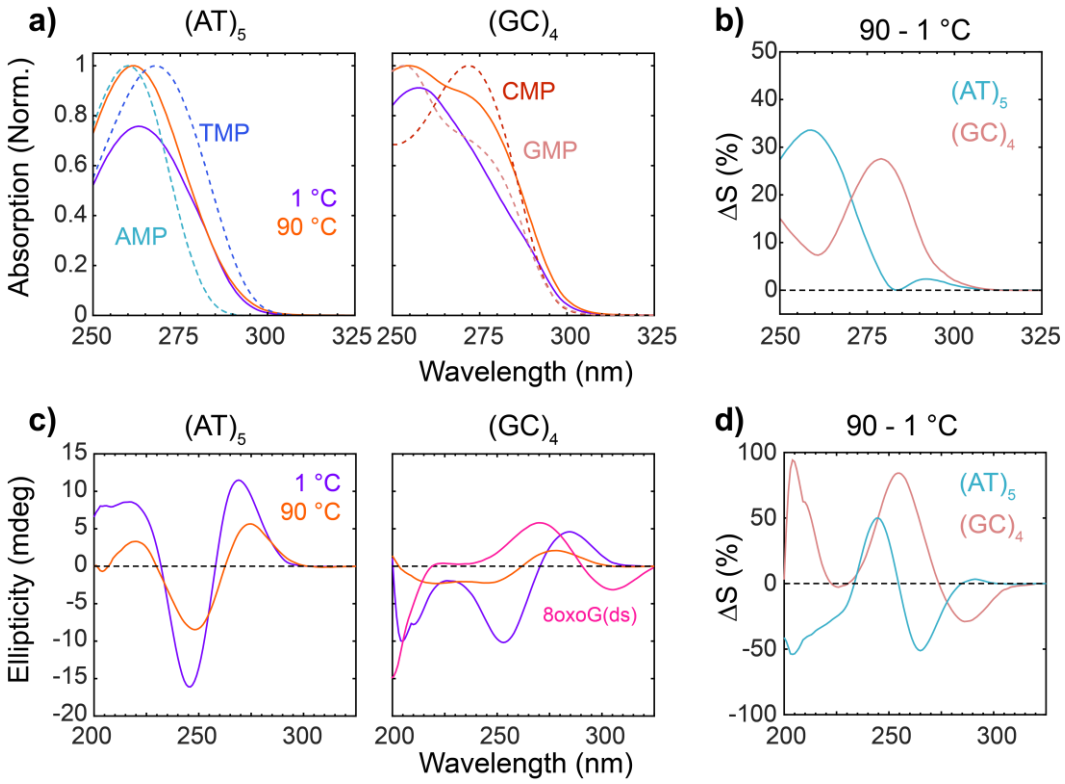

Figure 2.9 Signatures of DNA duplex melting from UV absorption and CD spectroscopy. (a) UV absorption spectra of $(AT)_5$ and $(GC)_4$ at 1 and 90 °C from 250 to 325 nm. Oligonucleotides were prepared at 50 μM in pH 7 40 mM SPB. Dashed lines correspond to absorption spectra of respective nucleotide monophosphates. **(b)** Percent change difference spectra between 90 and 1 °C for each sequence. **(c)** UV circular dichroism (CD) spectra of $(AT)_5$ and $(GC)_4$ at 1 and 90 °C. The room temperature spectrum of a modified self-complementary Z-form duplex, 5′-TACXCXCXTA-3′ where X is 8-oxoguanosine, is overlaid with $(GC)_4$. **(d)** Percent change difference spectra for melting of $(AT)_5$ and $(GC)_4$.

UV CD spectroscopy is now more commonly applied than UV absorption to monitor nucleic acid structural changes. CD reports on the chirality of the whole DNA molecule, and is therefore best for detecting associated and folded structures with different orientations. Significant spectral differences are found between different duplex conformations (A, B, Z) and between

parallel and anti-parallel G-quadruplex orientations,[66] whereas these conformations may be harder to detect with FTIR.[22, 30] As shown in Fig. 2.9c, the greater ordering of duplexed nucleic acids enhance their CD signal relative to the single-strands such that melting leads to an overall loss of signal. Similar to the UV absorption, the CD spectral changes of $(GC)_4$ in the 225 – 300 nm window are redshifted 25 – 30 nm relative to $(AT)_5$. However, larger sequence differences are observed from 200 – 225 nm. As an example of CD's sensitivity to helical conformation, a left-handed Z-form DNA duplex modified with 8-oxoguanosine shows a near flip of the CD spectrum in the 225 – 300 nm window relative to the B-form duplex of $(GC)_4$. A drawback of CD is that its interpretation is largely empirical as the conformation of a measured sample is assigned based on its similarity to established reference samples. As a result, inaccurate structural assignments may be made when CD spectra deviate from well-established spectra.[67]

2.3 NMR for structural dynamics

Nuclear magnetic resonance (NMR) spectroscopy is well-established as one of the most useful approaches for studying protein and nucleic acid structure and kinetics. The high sensitivity of nuclei to their local environment allows for residue-specific probing of oligonucleotides and small proteins with ^{15}N or ^{13}C isotope labeling,[68-69] and nucleobase-specific resolution may still be achieved for oligonucleotides containing less than ~15 nucleotides using ^{1}H NMR spectroscopy.[70-71] Additionally, numerous NMR active nuclei provide distinct information on nucleic acid structure and dynamics (Table 2.2). Much of the structural insight is learned through homonuclear and heteronuclear multidimensional NMR methods that probe correlations between the same type (homonuclear) and different types of nuclei (heteronuclear) that nowadays can be performed with

commercial spectrometers and software, making NMR a widely used approach in biophysics among non-specialists. While NMR was commonly viewed as a structural tool for many decades, recent advances have enabled studies of kinetics on micro-to-millisecond timescales with sensitivity to minor configurations unmatched by other spectroscopic techniques.[23-24, 72]

Table 2.2 Common NMR-active nuclei for studies of nucleic acids and proteins

Nuclei	Natural Abundance	Gyromagnetic ratio γ/2π (MHz/T)	Applications for nucleic acids
^{1}H	99.990	42.576	H-D exchange, structure determination, large conformational changes
^{13}C	1.110	10.708	Relaxation dispersion, base pair and sugar conformation
^{15}N	0.370	-4.316	Relaxation dispersion, base pair conformation
^{31}P	100.000	17.235	Backbone conformation[73]
^{19}F	100.000	40.078	Ligand-binding,[74] site-specific dynamics, in cell labeling
^{23}Na	100.000	11.262	Cation-nucleic acid interactions[75]
^{7}Li	92.580	16.546	Cation-nucleic acid interactions[76]
^{17}O	0.037	-5.772	Water-nucleic acid interactions[77]

2.3.1 NMR spectra of nucleic acids

NMR obtains its structural sensitivity and resolution from the numerous possible active nuclei in nucleic acids. The precession frequency of each nucleus is quite sensitive to the local environment, leading to frequencies on each nucleobase that are often separated enough to be resolved by 2D NMR measurements in short oligonucleotides. As the system and number of nuclei grow, the ability to resolve individual nuclei becomes more difficult. This challenge of resolving nuclei depends on the type of nuclei being measured, as each exhibits a different range of resonant frequencies (ω). These frequencies are typically expressed in chemical shift (δ) relative to a reference resonance (ω_{ref}) where ω_0 is the Larmor frequency set by the spectrometer's static magnetic field (B_0).

$$\delta = \frac{\omega - \omega_0}{\omega_0} - \frac{\omega_{ref} - \omega_0}{\omega_0} \tag{2.1}$$

The shift in ω relative to ω_0 arises from shielding of the nuclei by a local magnetic field that opposes B_0 and is produced by the movement of the electron cloud around the nuclei. The resulting value of δ for a given nucleus will depend on the local electron density, and δ is expressed in parts per million (ppm).

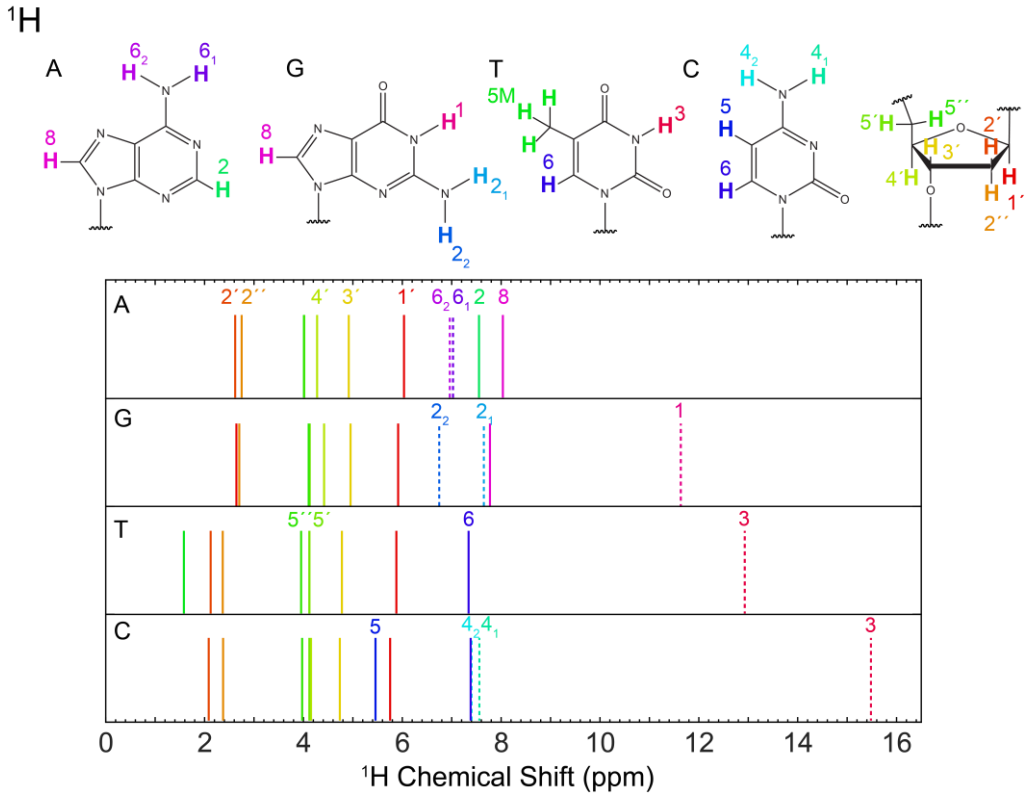

Figure 2.10 ^{1}H chemical shifts of nucleobase and deoxyribose nuclei in DNA. ^{1}H chemical shifts for each nucleobase and deoxyribose color-coded with the chemical structure. Dashed lines correspond to exchangeable nuclei. Chemical shifts correspond to average values across all structures deposited in the Biological Magnetic Resonance Bank (BMRB).[3] The average amino and imino chemical shifts for thymine and guanine come from duplex DNA structures, but the chemical shift for cytosine H3 comes from triplex and i-motif structures.

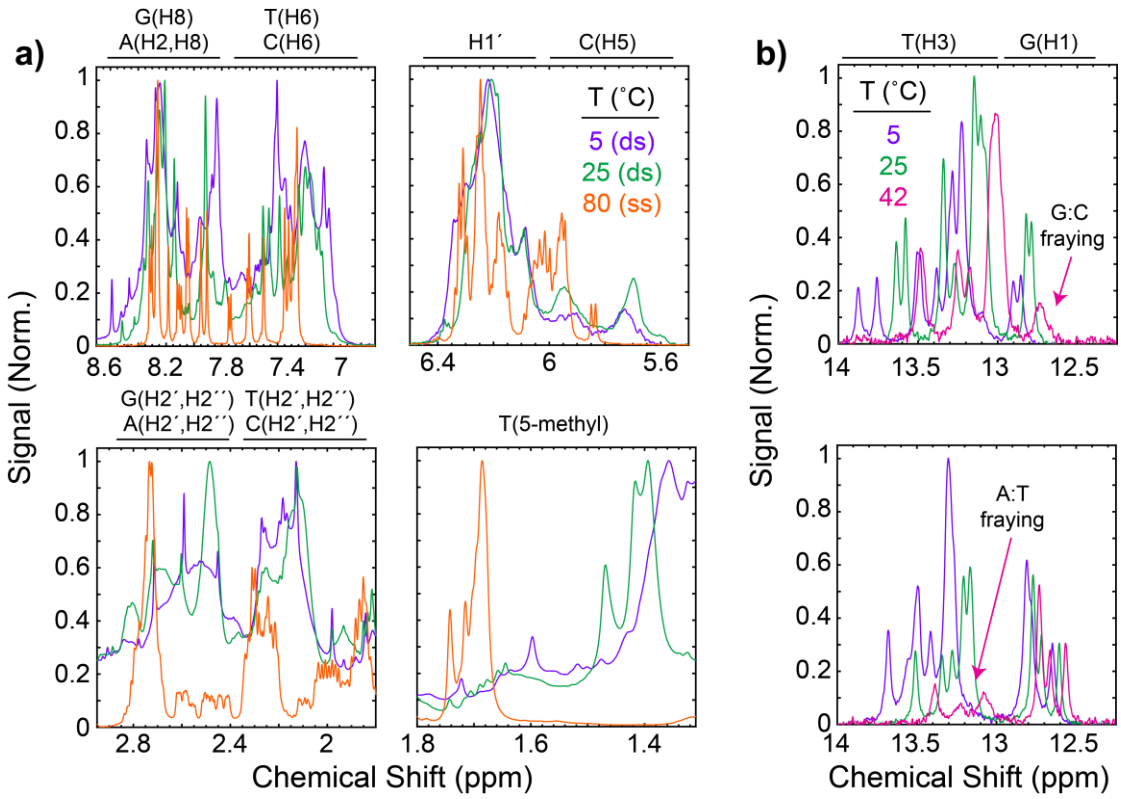

Figure 2.11 ^{1}H NMR spectra of DNA oligonucleotides. (a) Spectra of non-exchangeable protons in D_2O for 5′-CGCATATATAT-3′ + complement. Duplex (ds) spectra are measured at 5 and 25 °C and single-strand (ss) spectra are measured at 80 °C. Approximate chemical shifts regions for certain nuclei are labelled above each plot. **(b)** Imino proton spectra of (top) 5′-CGATATATTATATACG-3′ + complement and (bottom) 5′-ATATATCGCGTATATA-3′ + complement in 95% H_2O/5% D_2O. Samples are highly duplexed at each temperature, and fraying of terminal base pairs leads to faster exchange of imino protons with the solvent and broadening of peaks.

^{1}H NMR spectroscopy is commonly applied to nucleic acids due to its high sensitivity to global and local structural changes as well as simple implementation relative to ^{13}C and ^{15}N NMR.[70, 78-79] However, as illustrated in Figs. 2.10 and 2.11, ^{1}H resonance frequencies are only spread out across a ~15 ppm range, and most are contained within 1 – 8.5 ppm. The region from 7 – 8.6 ppm contains all aromatic protons except H5 of cytosine, which is typically at 5.5 – 6 ppm.

The chemical shifts of these aromatic protons and the thymine 5-methyl group (1.2 – 2.0 ppm) are highly sensitive to base pairing and stacking,[79-80] but these frequency regions suffer from spectral congestion that hinders interpretation even for the short duplexes shown in Fig. 2.11. The spectral overlap is greatest at low temperatures due to broadening of the peaks from faster transverse spin relaxation (Section 2.3.3). The amino and imino protons, on the other hand, exchange with water molecules, and their chemical shift and peak width report on the solvent accessibility of a given base pair. The amino peaks overlap with the aromatic protons, but the imino protons are shifted well upfield from the other resonances. In duplex DNA, guanine H1 resonances are typically located from 12 – 13 ppm and thymine H3 resonances occur from 13 – 14 ppm, but imino resonances may shift by more than 1 ppm in other structural contexts.[81] In general, sharp imino resonances indicate that the proton exchanges slowly with water. Increased solvent accessibility leads to broadening and eventually complete loss of the peak. In duplexes, this is most commonly observed from fraying of the terminal base pairs. Figure 2.11b shows how imino resonances of nucleobases closer to the termini broaden and diminish at lower temperatures than those internal to the duplex as a result of increased accessibility to water.

^{1}H-^{1}H 2D NMR measurements provide spectra that correlate the frequency of two ^{1}H nuclei and are critical for peak assignments. Nuclear Overhauser effect correlation spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) are the most commonly applied ^{1}H-^{1}H 2D measurements for macromolecules.[82-83] NOESY cross peaks arise from dipole-dipole (or dipolar) couplings mediated by nuclei that are close to each other in space. In contrast, TOCSY cross peaks come from J-coupling (or scalar coupling) mediated through chemical bonds and indicate whether

two ^{1}H nuclei are connected through a network of carbon-carbon bonds. NOESY utilizes a series of radiofrequency (RF) pulses separated by sequential time delays τ_1, the mixing time (τ_{mix}), and τ_2. Data are collected as a function of τ_1 and τ_2 and Fourier transformed along each time axis to produce a 2D correlation map with frequency axes ω_1 and ω_2. The τ_{mix} delay allows for spin transfer between nuclei and is typically set to 100 – 400 ms for macromolecules to maximize spin transfer between nearby nuclei ($\leq$ 6 nm) yet minimize subsequent transfer to nuclei at further distances (spin diffusion). Acquisition of TOCSY spectra is similar to NOESY except an additional set of pulses is applied after τ_1 to remove contributions from dipolar coupling, commonly referred to as an isotropic mixing sequence.[84]

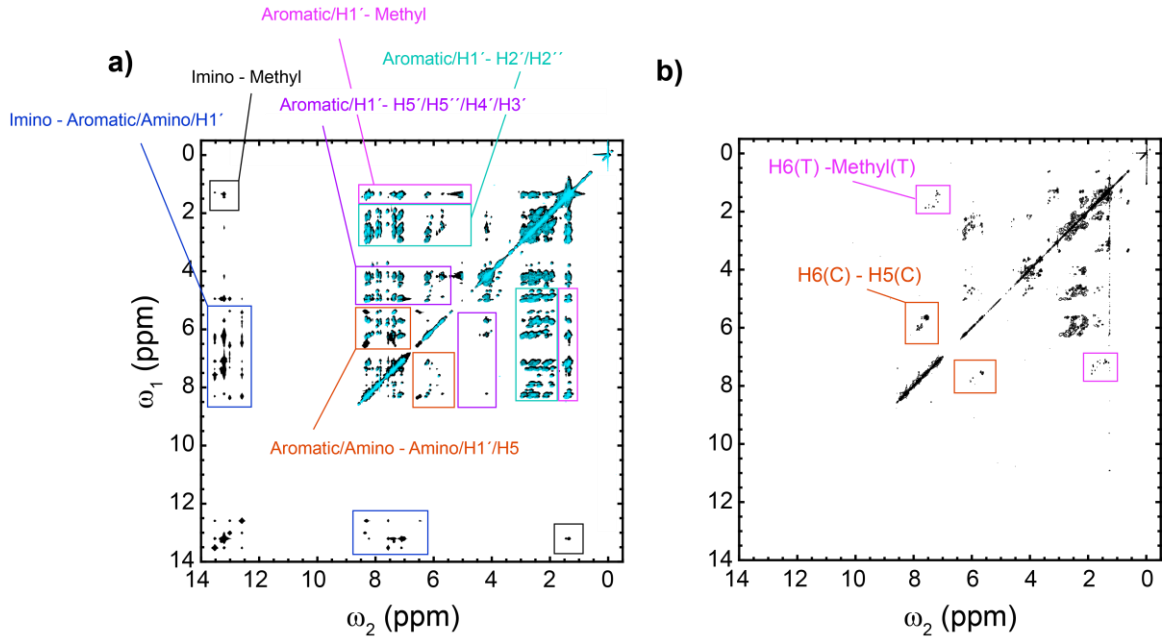

Figure 2.12 NOESY and TOCSY spectra of an 11-mer DNA duplex. (a) NOESY spectrum of 5′ -CGCATATATAT-3′ + complement in pH7 400 mM SPB at 15 °C using a mixing time of 200 ms. Spectra in 95% H_2O/5% D_2O (black) and 100% D_2O (cyan) are overlaid. Selected cross-peak regions are indicated. **(b)** TOCSY spectrum of 5′-CCTATATATCC-3′ + complement using a mixing time of 60 ms. The regions containing cross peaks between H5 and H6 nuclei of cytosine and between H6 and methyl nuclei of thymine are indicated.

Figure 2.12 shows example NOESY and TOCSY spectra for an 11-mer DNA duplex. Even for a short duplex, the NOESY spectrum is congested with various interbase, intrabase, base-sugar, and sugar-sugar cross peaks and quickly becomes difficult to interpret for larger duplexes. When there is sufficient separation, networks of cross-peaks among aromatic (except adenine H2), methyl, H1′, and H2′/H2″ nuclei may be used to identify the nucleobase associated with each resonance. H1′ protons are oriented in the 3′ direction, therefore aromatic resonances will only show strong cross peaks to the H1′ nuclei on their nucleotide and their 5′ neighbor. As a result, the network of cross peaks between aromatic and H1′ nuclei indicates the order of nucleobases in a strand, and this ordering can be supported through cross peaks between aromatic and H2′/H2″ nuclei and between H1′ and H2′/H2″. Cross peaks between methyl or aromatic protons with imino protons can greatly simplify peak assignments when available. Cross peaks to imino and amino resonances are easily identified by comparing spectra measured in 95% H_2O/5% D_2O and 100% D_2O as the peaks are lost due to deuteration in the latter (Fig. 2.12a).

In contrast to NOESY, TOCSY spectra of nucleic acids provide relatively select information. The through-bond spin networks of nucleobases are interrupted by many nitrogen atoms such that TOCSY cross peaks only arise from H6 to H5 couplings in cytosine/uracil and H6 to 5-methyl in thymine and 5-methylcytosine (Fig. 2.12b). The H6-H5 cross peaks distinguish H6 cytosine peaks from those of thymine and as well as H5 from H1′. The H6-5-methyl cross peak in turn isolates thymine H6 peaks in the aromatic region. Additional cross peaks arise between sugar protons on the same ring, which help assign H1′ and H2′/H2″ nuclei. In general, a combination of NOESY and TOCSY measurements is necessary for peak assignment and is routinely performed in practice.

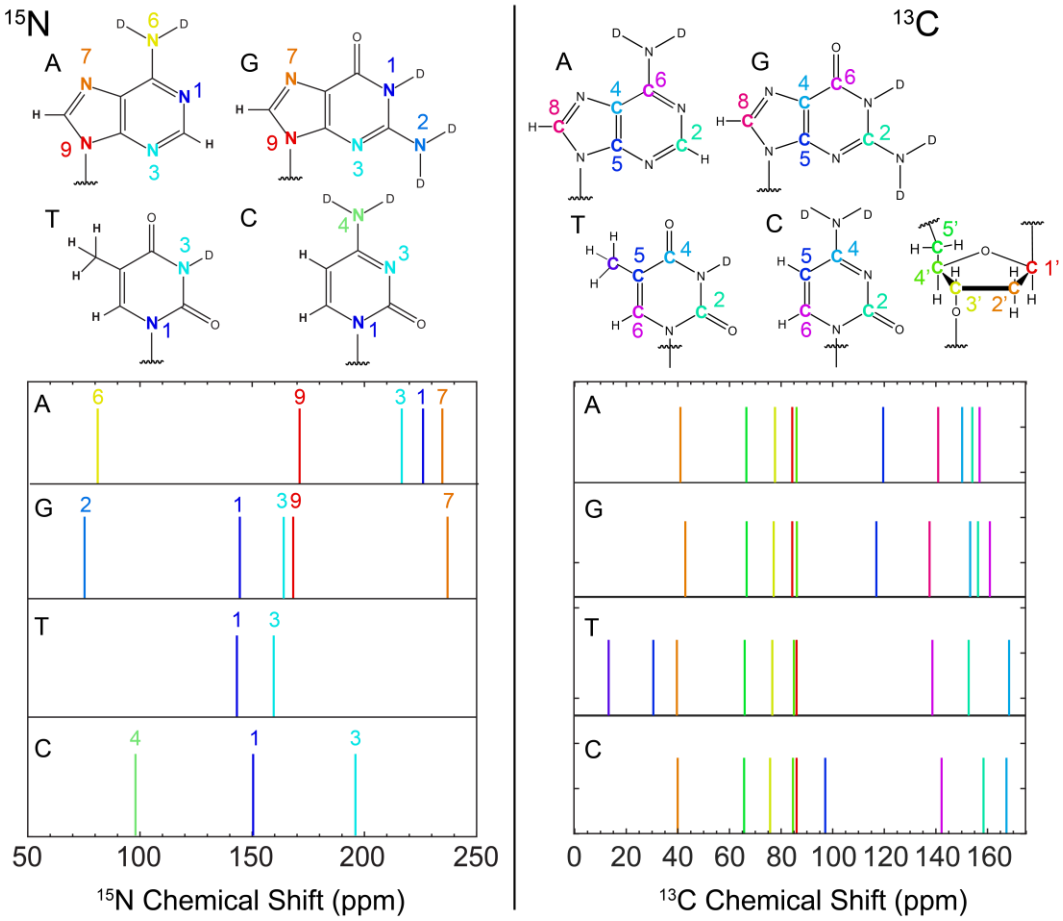

Figure 2.13 ^{15}N and ^{13}C chemical shifts of DNA nucleobase and deoxyribose nuclei. Chemical shifts for each nucleobase and deoxyribose color-coded with the chemical structure. Chemical shifts correspond to average values across all structures deposited in the Biological Magnetic Resonance Bank (BMRB).[3]

Due to the spectral congestion of ^{1}H NMR spectroscopy, nucleic acids are now more commonly studied with ^{13}C and ^{15}N NMR. These methods require isotopic labeling for which robust methods were first developed for RNA and are now being developed for DNA. For RNA, isotope labeling may be accomplished through chemical synthesis of labeled nucleobases, *in vitro* transcription, or *in vivo* expression. The toolbox for isotopic labeling of DNA is less developed and typically relies on chemical and enzymatic synthesis.[85] The chemical shifts of ^{13}C and

particularly ^{15}N nuclei are spread over a wider range relative to ^{1}H (Fig. 2.13), allowing better isolation of peaks arising from each nucleobase. ^{13}C spectroscopy is highly sensitive to base protonation, large conformational changes, and sugar conformational changes whereas ^{15}N is particularly useful for tracking changes in base pairing.[24, 86-87]

2.3.2 NMR time resolution

Those of us interested in studying structural dynamics occurring on sub-millisecond timescales often justify our use of optical spectroscopy with the notion that NMR is a "slow" technique. However, a wide suite of NMR experiments has been developed for studying biomolecular dynamics, and the time resolution of NMR depends on the information that is desired. Indeed, real-time NMR measurements are conventionally limited to probing timescales slower than 1 s, yet recent perturbation and trap-freezing methods have pushed real-time measurements to milliseconds.[6, 88] The intrinsic timescale of NMR is commonly defined in terms of its ability to spectrally distinguish exchanging species. In a standard 1D NMR spectrum of a system undergoing two-state exchange between species A and B, a given nucleus will only show two separate peaks if the rate of exchange between A and B is much slower than the frequency difference between the nuclei in states A and B ($k_{ex} \ll |\Delta\omega_{AB}|$) whereas a single peak centered at the population-weighted frequency of A and B will be observed when $k_{ex} \gg |\Delta\omega_{AB}|$. While k_{ex} is intrinsic to the system under study, $\Delta\omega_{AB}$ depends on the nucleus of interest and strength of the static magnetic field, B_0. For example, the chemical shift difference for a duplexed vs. single-strand nucleobase aromatic proton is ~0.5 ppm, which with B_0 = 500 MHz corresponds to $|\Delta\omega_{AB}|$ = 250 Hz. NMR measurements typically require millimolar oligonucleotide concentrations for

reasonable signal-to-noise, leading to $k_{ex} \geq 500$ Hz for hybridization. Therefore, duplex and single-strand resonances are averaged as a single set of peaks in 1D NMR measurements. In contrast, ^{13}C and ^{15}N nuclei of nucleobases often shift by many ppm during hybridization such that faster exchange may be resolved.[89-90]

While real-time NMR is unable to monitor sub-millisecond structural dynamics, numerous spin relaxation methods have been developed to detect such processes with varying levels of structural and kinetic information. Carr-Purcell-Meiboom-Gill (CPMG), rotation frame spin relaxation ($R_{1\rho}$), and chemical exchange saturation transfer (CEST) are the most common NMR methods for probing chemical exchange on second to microsecond timescales and enable extraction of the exchange rates and chemical shifts for multi-state exchange. Relative to other time-resolved spectroscopy techniques, the most advantageous aspect of these NMR methods is their sensitivity to scarcely populated states (even less than 0.1% population).[87, 91] Of the three experiments, $R_{1\rho}$ can typically access the shortest timescales of 1-10 μs whereas CPMG and CEST are usually limited to 100 μs or slower timescales.[24] The basis for these different time windows is explored in the following sections.

NMR spin relaxation is also sensitive to structural motions occurring on pico-to-nanosecond timescales but cannot offer the same level of kinetic and structural detail achieved in ms-to-μs relaxation dispersion measurements. Spin relaxation data is analyzed using the "model-free" method of Liparit and Szabo[92] or through spectral density mapping[93] to extract the overall rotational correlation time of a molecule (τ_c), the effective internal rotational correlation time of a given bond vector (τ_e), and a general order parameter (S) that is a measure of

the distribution of orientations for a bond vector. These parameters do not directly inform on the kinetics or structural changes associated with the motions, but instead are primarily used to identify which regions of a biomolecule are more dynamic or disordered than others.[94] Molecular detail into pico-to-nanosecond motions in proteins and nucleic acids can be extracted through refinement of molecular dynamics (MD) simulations using residue-specific S.[95]

2.3.3 NMR relaxation for detecting chemical exchange

T1 and T2 Relaxation

One of the oldest and most common methods of extracting chemical exchange timescales is through measurements of the timescales for either restoring magnetization parallel to B_0 (longitudinal, T1) or the loss of net magnetization in the xy plane (transversal, T2) following a 90° pulse along xy. T1 relaxation primarily occurs through dipole-dipole interactions with nearby nuclei and the fluctuation of dipolar interactions that largely result from rotation of molecules in solution. T1 is shortest when the rotational correlation time (τ_c) is near the inverse of the Larmor frequency (ω_0). τ_c for oligonucleotide nuclei are typically 5-10 ns and approximately increase with molecular size by the Stokes-Einstein-Debye equation where η is the solvent viscosity and r is the spherical radius of the molecule.

$$1/\tau_c = \frac{k_B T}{8\pi\eta r^3} \tag{2.2}$$

For common NMR spectrometers with $1/\omega_0$ = 1-3 ns, T1 will always increase as the length of the nucleic acid oligonucleotide increases, and the protons of DNA duplexes with 6-10 base pairs typically exhibit T1 of 1-2 s.[96-97] T2 relaxation results from field inhomogeneity and interactions

between spin-1/2 nuclei, and T2 monotonically decreases as τ_c increases because larger molecules have a greater number of spin-spin interactions. As a result, T2 and T1 are nearly identical for small molecules in solution whereas T2 < T1 for macromolecules in solution. T2 values of 10 – 50 ms have been reported for short DNA duplexes.[96-97] It is important to note that T1 and T2 will depend on temperature due to their relationship with τ_c through eq. 2.2. T2 increases as temperature increases, and T1 of macromolecules decreases as temperature increases.

The timescale for spin relaxation is also sensitive to chemical exchange of nucleic acids between different conformations where nuclei precess at different frequencies. Converting between states with different precession frequency leads to dephasing of the bulk magnetization generated in the xy-plane, and its corresponding effect on spin relaxation depends on the timescale of exchange. We consider the case of two-state chemical exchange between species A and B with rate constants k_A and k_B,[98-100] yet more complicated kinetic schemes have been treated under various assumptions.[93]

$$A \underset{k_B}{\overset{k_A}{\rightleftharpoons}} B \tag{2.3}$$

The time-dependence of nuclear magnetization (M) along x, y, z for A and B, accounting for longitudinal and transversal relaxation as well as chemical exchange, is described by the Bloch-McConnell equations.[98]

$$\frac{d}{dt}\begin{pmatrix} M_{x,A} \\ M_{y,A} \\ M_{z,A} \end{pmatrix} = \begin{pmatrix} -\left(R_{2,A}+k_A\right) & -\Delta\omega_A & 0 \\ \Delta\omega_A & -\left(R_{2,A}+k_A\right) & 0 \\ 0 & 0 & -\left(R_{1,A}+k_A\right) \end{pmatrix}\begin{pmatrix} M_{x,A} \\ M_{y,A} \\ M_{z,A} \end{pmatrix} + \begin{pmatrix} k_B M_{x,B} \\ k_B M_{y,B} \\ k_B M_{z,B} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ R_{1,A} M_{0,A} \end{pmatrix} \tag{2.4a}$$

$$\frac{d}{dt}\begin{pmatrix} M_{x,B} \\ M_{y,B} \\ M_{z,B} \end{pmatrix} = \begin{pmatrix} -\left(R_{2,B}+k_B\right) & -\Delta\omega_B & 0 \\ \Delta\omega_B & -\left(R_{2,B}+k_B\right) & 0 \\ 0 & 0 & -\left(R_{1,B}+k_B\right) \end{pmatrix}\begin{pmatrix} M_{x,B} \\ M_{y,B} \\ M_{z,B} \end{pmatrix} + \begin{pmatrix} k_A M_{x,A} \\ k_A M_{y,A} \\ k_A M_{z,A} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ R_{1,B} M_{0,B} \end{pmatrix} \quad (2.4b)$$

Here, $R_{1,x}$ and $R_{2,x}$ correspond to longitudinal ($1/T_{1,x}$) and transversal ($1/T_{2,x}$) relaxation rates, respectively, for species X. $\Delta\omega_x$ is the frequency offset for each species, and $\Delta M_{0,x}$ is the equilibrium magnetization along z. General solutions for the observed longitudinal ($R_{1,obs}$) and transversal ($R_{2,obs}$) relaxation rates have been determined from solving eqs. 2.4:[99]

$$R_{1,obs} = A_1 \pm \sqrt{A_1^2 - \left(R_{1A,eff} R_{1B,eff} - k_A k_B\right)} \quad (2.5a)$$

where $A_1 = \left(R_{1A,eff} + R_{1B,eff}\right)/2$, $R_{1A,eff} = R_{1,A} + k_A$, and $R_{1B,eff} = R_{1,B} + k_B$

$$R_{2,obs} = A_2 \pm \sqrt{\left[B + \left(B^2 + C^2\right)^{1/2}\right]/2} \quad (2.5b)$$

where $A_2 = \left(R_{2A,eff} + R_{2B,eff}\right)/2$, $R_{2A,eff} = R_{2,A} + k_A$, and $R_{2B,eff} = R_{2,B} + k_B$

$$B = \frac{\left(R_{2A,eff} - R_{2B,eff}\right)}{4} + k_A k_B - \frac{\Delta\omega_{AB}^2}{4},\ C = \frac{\left(R_{2A,eff} - R_{2B,eff}\right)}{2}\Delta\omega_{AB}$$

$\Delta\omega_{AB}$ is the resonance frequency difference between nuclei in the A and B states. The observed resonance frequency (ω_{obs}) occurs in between the precession frequency in each state.

$$\omega_{obs} = A_\omega \pm \sqrt{\left[-B + \left(B^2 + C^2\right)^{1/2}\right]/2}, \text{ where } A_\omega = \left(\omega_A + \omega_B\right)/2 \quad (2.5c)$$

Figure 2.14 shows an example of how T1 relaxation is sensitive to H-D exchange in duplex DNA. Longitudinal magnetization of the imino protons is inverted using a 180° pulse, and T1 relaxation is monitored as the recovery of positive signal. $R_{1,obs}$ varies across the imino resonances due to different rates of H-D exchange with the solvent. Terminal base-pair fraying increases the solvent exposure of nucleobases near the termini, leading to faster H-D exchange and consequently T1 relaxation. As introduced in Chapter 1.2, similar H-D exchange measurements are widely applied to study base-pair kinetics and stability in nucleic acids.[71, 101-102]

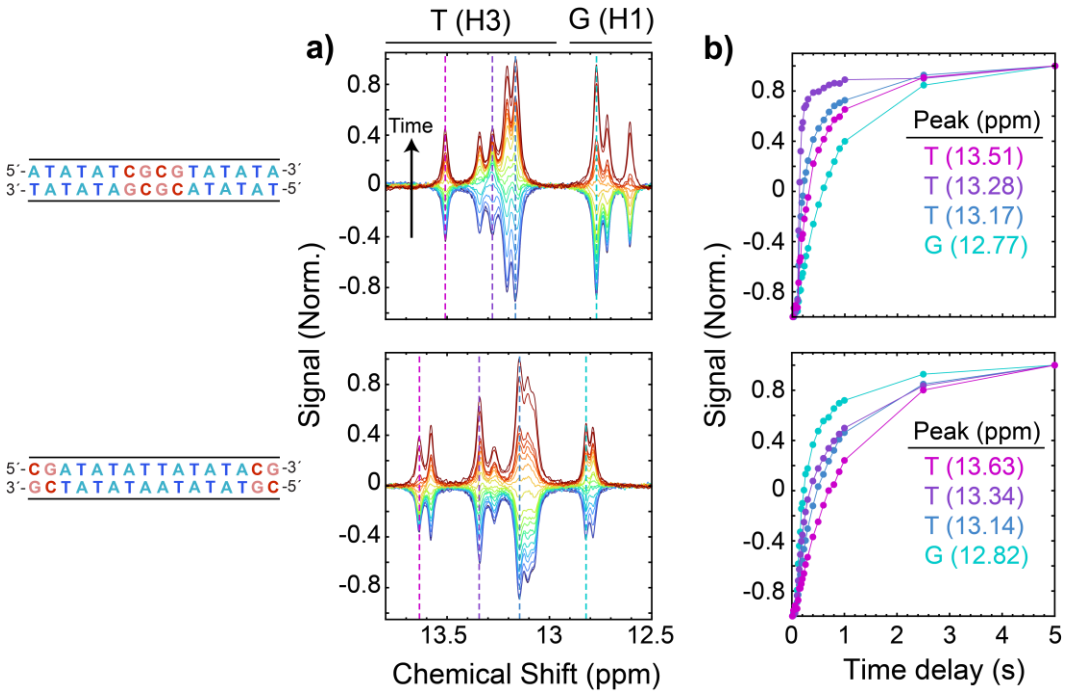

Figure 2.14 Influence of H-D exchange on T1 relaxation in a DNA duplex. (a) Imino proton spectra of DNA duplexes with (top) A:T termini and (bottom) G:C termini at time delays (blue to red) following inversion of the magnetization parallel to B_0 using a 180° pulse. Thymine H3 and guanine H1 resonances occur from 13 – 13.7 and 12.5 – 12.9 ppm, respectively. **(b)** Time traces at frequencies marked in **(a)** showing different rates of T1 relaxation depending on the rate of H-D exchange with water. Nucleobases closer to the termini exhibit faster H-D exchange.

The majority of structural transitions investigated in this thesis occur with timescales ranging from nanoseconds to milliseconds, therefore we are particularly interested in assessing the sensitivity of NMR relaxation to sub-millisecond chemical exchange. In such cases, the Bloch-McConnell equations may be solved under the conditions of fast exchange ($k_{ex} = (k_A + k_B) \gg |R_{2.A} - R_{2,B}|$) and exchange narrowing ($k_{ex} \gg \Delta\omega_{AB}$). A single nuclei resonance is observed at a ω_{obs} weighted by the population of each state (p_A, p_B).

$$\omega_{obs} = p_A\omega_A + p_B\omega_B \tag{2.6}$$

In this limit, $R_{1,obs}$ is also the population-weighted sum of $R_{1,A}$ and $R_{1,B}$.

$$R_{1,obs} = p_A R_{1,A} + p_B R_{1,B} \tag{2.7}$$

When $R_{1,A}$ and $R_{1,B}$ are significantly different, $R_{1,obs}$ reports on p_A and p_B but not k_{ex}. In contrast to $R_{1,obs}$, $R_{2,obs}$ directly depends on k_{ex}, $\Delta\omega_{AB}$, and the population-weighted sum of $R_{2,A}$ and $R_{2,B}$ (denoted $\overline{R}_{2,0}$).

$$R_{2,obs} = \bar{R}_{2,0} + \frac{p_A p_B \Delta\omega_{AB}^2}{k_{ex}} \tag{2.8}$$

$$\text{where, } \bar{R}_{2,0} = p_A R_{2,A} + p_B R_{2,B}$$

Due to its direct relationship with k_{ex}, $R_{2,obs}$ is typically used to report on fast chemical exchange. $R_{2,obs}$ increases with the magnitude of $\Delta\omega_{AB}$ due to more efficient dephasing of the bulk magnetization. However, $R_{2,obs}$ decreases as k_{ex} increases in the fast exchange limit because the

respective lifetimes in states A and B become too short to induce dephasing of the bulk magnetization at the motionally averaged resonance frequency. Figure 2.15a shows how $R_{2,obs}$ varies with k_{ex} in the fast exchange limit. k_{ex} is resolved through the deviation of $R_{2,obs}$ from $\overline{R}_{2,0}$, which is greatest at relatively low k_{ex} and/or large $\Delta\omega_{AB}$. The values of k_{ex} where $R_{2,obs}$ becomes indistinguishable from $\overline{R}_{2,0}$ sets the time resolution for measuring chemical exchange, which is ~10 μs in the model data.

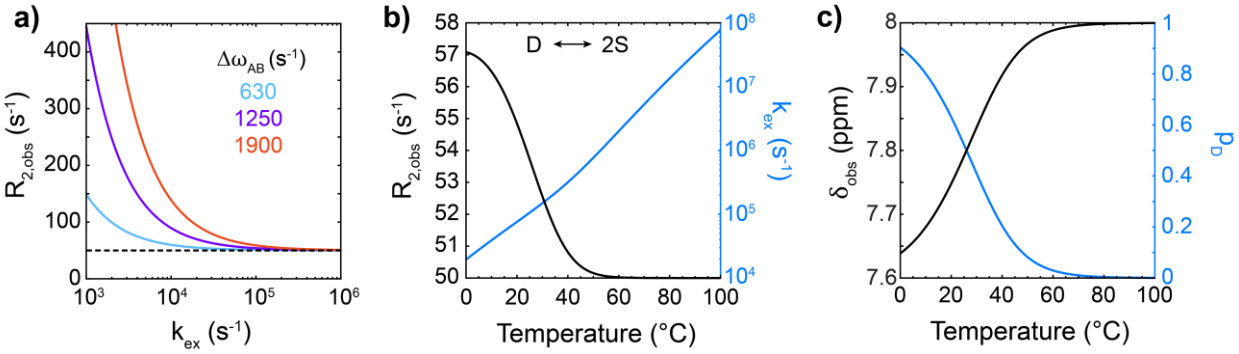

Figure 2.15 Behavior of transverse relaxation rate with two-state chemical exchange. (a) Observed transversal relaxation rate ($R_{2.obs}$) for a motionally averaged resonance as a function of two-state chemical exchange rate (k_{ex}). Data are shown for variable $\Delta\omega_{AB}$, and $\overline{R}_{2,0}$ = 50 s^{-1} (denoted by dashed line). Values are chosen to be reasonable hybridization of short oligonucleotides. **(b)** Example of temperature-dependence of $R_{2.obs}$ for oligonucleotide hybridization using k_{ex} determined from experimental data (blue line) in Chapter 9. The temperature-dependence of $\overline{R}_{2,0}$ is neglected and $R_{2.A} = R_{2,B}$. **(c)** Observed resonance frequency (ω_{obs}) plotted in terms of chemical shift (δ) as a function of temperature for the same system in **(b)**. The chemical shift of the proton in the duplex state is 7.6 ppm and that in the single-strand state 8.0 ppm. The blue line shows the duplex population (p_D).

An example in Figs. 2.15b,c shows how $R_{2,obs}$ may vary with temperature-dependent hybridization kinetics using data for a DNA dinucleotide from Chapter 8. As the temperature increases, the observed exchange rate, $k_{ex} = k_d + k_a([S_1] + [S_2])$, increases, where k_d and k_a are the dissociation and association rate constants, respectively. $R_{2,obs}$ approaches $\overline{R}_{2,0}$ as k_{ex}

increases, and exchange becomes undetectable. The loss and gain of duplex (p_D) and single-strand (p_S) population, respectively, as well as other temperature effects will cause $\overline{R}_{2,0}$ to decrease as temperature increases, but these effects are not included in Fig. 2.15. It is also worth noting that the observed motionally-averaged resonance frequency shifts from the duplex to single-strand frequency in accordance with the population change as expressed in eq. 2.6.

NMR for microsecond chemical exchange

Nowadays, micro-to-millisecond exchange is most commonly characterized using relaxation dispersion measurements such as CPMG, $R_{1\rho}$, and CEST.[103] CPMG is one of the oldest NMR methods for probing chemical exchange and is still used today to measure sub-millisecond processes. During the experiment, $R_{2,obs}$ is modulated via a train of 180° pulses (spin-echo sequence) with a repetition rate of ν_{CPMG}. Larger values of ν_{CPMG} are required to modulate, and therefore detect, faster exchange. For two-state chemical exchange in the motionally averaged limit, $R_{2,obs}$ may be expressed in terms of ν_{CPMG} as well as the factors that contribute to general $R_{2,obs}$ (eq. 2.8).

$$R_{2,obs}\left(\nu_{CPMG}\right) = \bar{R}_{2,0} + \frac{p_A p_B \Delta\omega_{AB}^2}{k_{ex}}\left(1 - \frac{2\nu_{CPMG}\tanh\left[k_{ex}/2\nu_{CPMG}\right]}{k_{ex}}\right) \tag{2.9}$$

Figure 2.16a,b shows calculated $R_{2,obs}$ vs. ν_{CPMG} profiles for sub-millisecond two-state exchange. At ν_{CPMG} = 0 (i.e. no modulating pulse train), $R_{2,obs}$ is identical to the transverse relaxation rate in eq. 2.8. Therefore, in the fast exchange limit, $R_{2,obs}$ becomes less sensitive to ν_{CPMG} as k_{ex} increases. For the example in Fig. 2.16a, the $R_{2,obs}$ vs. ν_{CPMG} profile becomes nearly flat at k_{ex} >

10,000 Hz. In this limit, the offset between $R_{2,obs}$ and $\overline{R}_{2,0}$ may be used to determine a lower bound for the exchange rate but k_{ex} itself cannot be determined. In practice, ν_{CPMG} is limited to values of 1 – 2 kHz as higher values can damage NMR probes and heat the sample,[104] which limits the exchange rates accessible with CPMG.

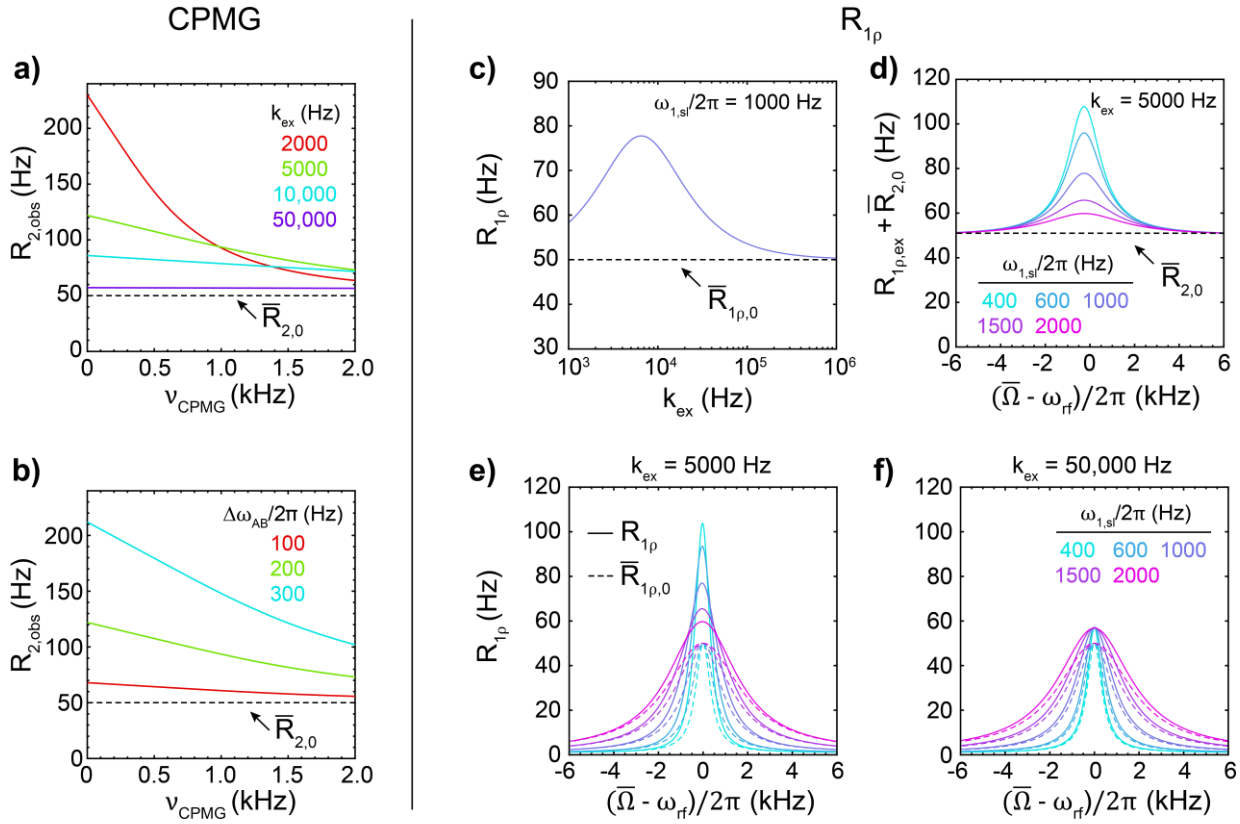

Figure 2.16 Manifestation of fast chemical exchange in CPMG and $R_{1\rho}$ NMR experiments. **(a)** Observed dependence of transverse relaxation rate ($R_{2,obs}$) on CPMG repetition rate (ν_{CPMG}) in the presence of fast two-state exchange calculated by eq. 2.9. $\Delta\omega_{AB}$ = 2000 Hz, p_A = 0.9, and p_B = 0.1. **(b)** $R_{2,obs}$ vs. ν_{CPMG} for various values of $\Delta\omega_{AB}$ with k_{ex} = 5000 Hz. **(c)** $R_{1\rho}$ relaxation rate calculated as a function of k_{ex} using eq. 2.12 with $\overline{\Omega}$ = 0, $\omega_{1,sl}/2\pi$ = 1000 Hz, $\Delta\omega_{AB}$ = 2000 Hz, p_A = 0.9, and p_B = 0.1. The dashed line corresponds to the non-exchange contributions to $R_{1\rho}$ ($\overline{R}_{1\rho,0}$, eq. 2.10). **(d)** $R_{1\rho,ex} + \overline{R}_{2,0}$ profiles vs. $\overline{\Omega}$ for multiple values of $\omega_{1,sl}$ with k_{ex} = 5000 Hz. The black dashed line corresponds to $\overline{R}_{2,0}$. $R_{1\rho}$ and $\overline{R}_{1\rho,0}$ vs. $\overline{\Omega}$ are shown for multiple values of $\omega_{1,sl}$ with **(e)** k_{ex} = 5000 Hz and **(f)** k_{ex} = 50,000 Hz.

$R_{1\rho}$ experiments employ larger effective field strengths than CPMG that enable resolution of chemical exchange timescales down to 3 – 10 microseconds.[23-24] Therefore, $R_{1\rho}$ is the most commonly applied NMR experiment to probe sub-millisecond nucleic acid dynamics and is particularly useful for detecting scarcely populated states.[24, 91] In these experiments, a continuous magnetic field is applied in the x or y direction (B_1) to spin-lock nuclei. Depending on the spin-lock field strength ($\omega_{1,sl}$) and nuclei resonance frequency of interest ($\overline{\Omega}$), the nuclei spin vector is tilted away from the z plane by $\theta = \tan^{-1}\left(\omega_{1,\mathrm{sl}} / \overline{\Omega}\right)$ and spin-locked along an average effective magnetic field $\omega_e = \sqrt{\overline{\Omega}^2 + \omega_{1,\mathrm{sl}}^2}$. For a motionally narrowed resonance, $\overline{\Omega}$ is the population-weighted average of Ω for each contributing state. In contrast to standard R_2, R_1, and CPMG measurements, $R_{1\rho}$ corresponds to the rate of relaxation for magnetization spin-locked along ω_e. Therefore, by changing ω_e, the final equilibrium position of the experiment is altered. To extract the most information on chemical exchange, $R_{1\rho}$ experiments are typically performed as a function of $\omega_{1,\mathrm{sl}}$ as well as by scanning off-resonance by varying the frequency of the spin-lock field (ω_{rf}).

$R_{1\rho}$ is weighted by both transverse ($\overline{R}_{2,0}$) and longitudinal relaxation ($R_{1,obs}$) depending on θ. In the absence of chemical exchange, $R_{1\rho}$ is denoted as $\overline{R}_{1\rho,0}$:

$$\bar{R}_{1\rho,0} = R_{1,obs}\cos^2\theta + \bar{R}_{2,0}\sin^2\theta \tag{2.10}$$

In macromolecules, $\overline{R}_{2,0}$ is usually much greater than $R_{1,obs}$ and dominates $\overline{R}_{1\rho,0}$. Since θ depends on $\overline{\Omega}$ and ω_1, $R_{1\rho}$ varies with $\overline{\Omega}$ and $\omega_{1,sl}$ even in the absence of chemical exchange as shown in

Fig. 2.16e,f. In general, $R_{1\rho}$ will contain contributions from both chemical exchange ($R_{1\rho.ex}$) and $\overline{R}_{1\rho,0}$.

$$R_{1\rho} = \bar{R}_{1\rho,0} + \sin^2\theta R_{1\rho,\mathrm{ex}} \tag{2.11}$$

Obtaining analytical solutions to $R_{1\rho}$ that account for chemical exchange is more difficult than for $R_{2,obs}$, but solutions have been obtained in the limit of asymmetric populations where $p_A >> p_B$ for a two-state system of states A and B, which has been used to study Hoogsteen base-pair formation, tautomerization, and low-temperature dehybridization.[89, 91, 105] For asymmetric populations and fast two-state chemical exchange:[23]

$$R_{1\rho} = \bar{R}_{1\rho,0} + \sin^2\theta \left[\frac{p_A p_B \Delta\omega_{AB}^2 k_{ex}}{\frac{\omega_A^2\omega_B^2}{\omega_e^2} + k_{ex}^2 - \sin^2\theta\, p_A p_B \Delta\omega_{AB}^2 \left(1 + \frac{2k_{ex}^2\left(p_A\omega_A^2 + p_B\omega_B^2\right)}{\omega_A^2\omega_B^2 + \omega_e^2 k_{ex}^2} \right)} \right] \tag{2.12}$$

where, $\omega_A^2 = \left(\Omega_A - \omega_{rf}\right)^2 + \omega_{1,\mathrm{sl}}^2$, $\omega_B^2 = \left(\Omega_B - \omega_{rf}\right)^2 + \omega_{1,\mathrm{sl}}^2$, and $\omega_e^2 = \left(\bar{\Omega} - \omega_{rf}\right)^2 + \omega_{1,\mathrm{sl}}^2$

Here, Ω_A and Ω_B correspond to the resonance frequencies of states A and B, respectively. Figure 2.16c shows $R_{1\rho}$ calculated as a function of k_{ex} for $(\overline{\Omega} - \omega_{rf}) = 0$ and fixed $\omega_{1,sl}$, illustrating that the fastest exchange timescales detectable with $R_{1\rho}$ are within approximately 1 – 10 μs. The reduction in $R_{1\rho}$ for k_{ex} >> 10^4 s^{-1} is analogous to that of $R_{2,obs}$ in the presence of fast chemical exchange (Fig. 2.15a). Profiles of $R_{1\rho}$ calculated as a function of $\overline{\Omega} - \omega_{rf}$ and at various $\omega_{1,sl}$ are shown in Fig. 2.16d,e,f. For exchange on hundreds of microsecond timescales, $R_{1\rho,ex}$ is significant

and is well resolved from the intrinsic $\overline{R}_{1\rho,0}$ background. For k_{ex} = 50,000 Hz, the $R_{1\rho,ex}$ enhancement of $R_{1\rho}$ is much smaller.

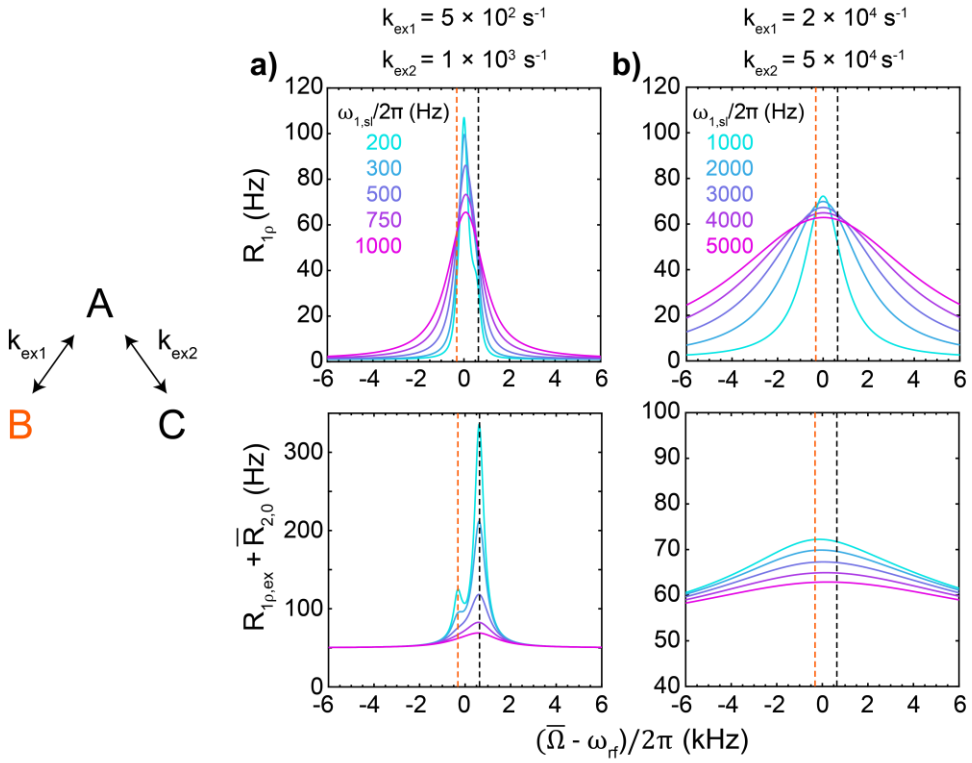

Figure 2.17 Comparison of $R_{1\rho}$ experiments for slow and fast 3-site exchange. (a) (top) $R_{1\rho}$ vs. $(\overline{\Omega} - \omega_{rf})$ and (bottom) $R_{1\rho,ex} + \overline{R}_{2,0}$ vs. $(\overline{\Omega} - \omega_{rf})$ profiles for slow exchange with k_{ex1} = 500 Hz and k_{ex2} = 1000 Hz as defined in the schematic on the left. p_A = 0.9, p_B = 0.05, p_C = 0.05, $\Delta\omega_{AB}$ = 2000 Hz, and $\Delta\omega_{AC}$ = -4000 Hz. Vertical dashed lines are placed at $\left(\overline{\Omega} - \omega_{rf}\right) = -\Delta\omega_{AB}$ (orange) and $\left(\overline{\Omega} - \omega_{rf}\right) = -\Delta\omega_{AC}$ (black). **(b)** Same plots for fast exchange with k_{ex1} = 20,000 Hz and k_{ex2} = 50,000 Hz.

In addition to variable $R_{1\rho}$ enhancement, the rate of chemical exchange also determines what type of information is attainable by $R_{1\rho}$. When exchange is fast, which effectively occurs when $k_{ex} \gg |\Delta\omega_{AB}|$ or $\omega_1 \gg |\Delta\omega_{AB}|$, then only k_{ex} and the product term $p_A p_B \Delta\omega_{AB}^2$ can be determined from fitting the $R_{1\rho}$ data to eq. 2.12. Independent measurements of the exchange

thermodynamics (i.e. p_A and p_B) allow for determination of the magnitude of $\Delta\omega_{AB}$ but not the sign. In contrast, much more structural information is attainable for slowly exchanging species. In particular, the sign and magnitude of the frequency difference between any number of well-separated states i and j ($\Delta\omega_{ij}$) can be determined. The difference between $R_{1\rho}$ in fast and slow exchange limits is illustrated for the case of 3-site exchange in Fig. 2.17 using previously published solutions.[23-24] When each exchange timescale is 1 – 2 ms, the magnitude and sign of the frequency difference associated with each exchange process are resolved as separate peaks in the $R_{1\rho,ex} + \overline{R}_{2,0}$ vs. ($\overline{\Omega} - \omega_{rf}$) profiles. In contrast, only a single broad maximum is observed for 20 – 50 μs exchange timescales (Fig. 2.17b), limiting the ability to directly resolve multiple exchanging states.

2.3.4 Optical and NMR spectroscopy for nucleic acid structural dynamics

Our goal is to monitor transitions occurring on timescales from picoseconds to minutes with as much structural information as possible, and this chapter demonstrates that no single ensemble method can satisfactorily address this goal. Tracking a process in real-time is most direct as the results are less dependent on processing and modeling. In this sense, direct kinetics from NMR is limited to events occurring over milliseconds and longer. While not a real-time technique, relaxation dispersion is sensitive to structural transitions as fast as ~5 μs and some that are unattainable in any other ensemble method. Most notable are (1) transitions to species with far less than 1% of the ensemble population and (2) transitions that cannot be synchronized via a perturbation. For these scenarios, relaxation dispersion NMR is the method of choice and has led to major discoveries in nucleic acids.[87, 91, 105]

Although large structural rearrangements occur on microsecond to hour timescales, the pico-to-microsecond window contains local dynamics that are best accessed by other ensemble methods like IR, UV, and X-ray methods.[15, 106] Most of these methods offer coarse yet distinct structural information. For example, CD reports on global helical structure, UV absorption is most sensitive to stacking, and IR to base-pairing structure. All of these methods are amenable to site-specific labeling that drastically enhances structural insight, yet with large costs and time-consuming synthesis. Direct comparison between data obtained with these methods may provide distinct and complementary kinetic information that can be self-consistently modeled to better describe the underlying structural dynamics.[107-108] In Chapters 3 and 4, we specifically discuss how optical spectroscopic probes can be incorporated into time-resolved methods for probing real-time kinetics on nanosecond to minute-long timescales.

2.4 Acknowledgements

I thank Jakob Schauss for careful reading and feedback on the chapter. I thank Joseph Sachleben for helpful discussions regarding NMR spectroscopy.

2.5 References

1. Lawson, C. L.; Patwardhan, A.; Baker, M. L.; Hryc, C.; Garcia, E. S.; Hudson, B. P.; Lagerstedt, I.; Ludtke, S. J.; Pintilie, G.; Sala, R., EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **2016,** *44*, D396-D403.

2. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic Acids Res.* **2000,** *28*, 235-242.

3. Hoch, J. C.; Baskaran, K.; Burr, H.; Chin, J.; Eghbalnia, H. R.; Fujiwara, T.; Gryk, M. R.; Iwata, T.; Kojima, C.; Kurisu, G., Biological magnetic resonance data bank. *Nucleic Acids Res.* **2023,** *51*, D368-D376.

4. Kaledhonkar, S.; Fu, Z.; Caban, K.; Li, W.; Chen, B.; Sun, M.; Gonzalez Jr, R. L.; Frank, J., Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. *Nature* **2019,** *570*, 400-404.

5. Chim, N.; Meza, R. A.; Trinh, A. M.; Yang, K.; Chaput, J. C., Following replicative DNA synthesis by time-resolved X-ray crystallography. *Nat. Commun.* **2021,** *12*, 2641.

6. Jeon, J.; Wilson, C. B.; Yau, W.-M.; Thurber, K. R.; Tycko, R., Time-resolved solid state NMR of biomolecular processes with millisecond time resolution. *J. Magn. Reson.* **2022,** *342*, 107285.

7. Thompson, M. C.; Barad, B. A.; Wolff, A. M.; Sun Cho, H.; Schotte, F.; Schwarz, D. M.; Anfinrud, P.; Fraser, J. S., Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme. *Nat. Chem.* **2019,** *11*, 1058-1066.

8. Wolff, A. M.; Nango, E.; Young, I. D.; Brewster, A. S.; Kubo, M.; Nomura, T.; Sugahara, M.; Owada, S.; Barad, B. A.; Ito, K., Mapping Protein Dynamics at High-Resolution with Temperature-Jump X-ray Crystallography. *bioRxiv* **2022**, 2022.06. 10.495662.

9. Kim, J.-Y.; Chung, H. S., Disordered proteins follow diverse transition paths as they fold and bind to a partner. *Science* **2020,** *368*, 1253-1257.

10. Hoffer, N. Q.; Woodside, M. T., Probing microscopic conformational dynamics in folding reactions by measuring transition paths. *Curr. Opin. Chem. Biol.* **2019,** *53*, 68-74.

11. Peterson, E. M.; Reece, E. J.; Li, W.; Harris, J. M., Super-Resolution Imaging of Competitive Unlabeled DNA Hybridization Reveals the Influence of Fluorescent Labels on Duplex Formation and Dissociation Kinetics. *J. Phys. Chem. B* **2019,** *123*, 10746-10756.

12. Chan, V.; Graves, D. J.; McKenzie, S. E., The biophysics of DNA hybridization with immobilized oligonucleotide probes. *Biophys. J.* **1995,** *69*, 2243-2255.

13. Treasurer, E.; Levicky, R., How Surfaces Affect Hybridization Kinetics. *J. Phys. Chem. B* **2021,** *125*, 2976-2986.

14. Kumpulainen, T.; Lang, B.; Rosspeintner, A.; Vauthey, E., Ultrafast elementary photochemical processes of organic molecules in liquid solution. *Chem. Rev.* **2017,** *117*, 10826-10939.

15. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

16. Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B., Protein folding and unfolding on a complex energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97*, 2527-2532.

17. Liu, F.; Gruebele, M., Downhill dynamics and the molecular rate of protein folding. *Chem. Phys. Lett.* **2008,** *461*, 1-8.

18. Bozovic, O.; Jankovic, B.; Hamm, P., Using azobenzene photocontrol to set proteins in motion. *Nat. Rev. Chem.* **2022,** *6*, 112-124.

19. Khalil, M.; Demirdöven, N.; Tokmakoff, A., Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution. *J. Phys. Chem. A* **2003,** *107*, 5258-5279.

20. De Marco, L. The molecular dynamics of hydrogen-bonding explored with broadband two dimensional infrared spectroscopy. Massachusetts Institute of Technology, 2016.

21. Sanstead, P. J. C., *Investigation of DNA Dehybridization through Steady-State and Transient Temperature-Jump Nonlinear Infrared Spectroscopy*. The University of Chicago: 2018.

22. Peng, C. S. Two-dimensional infrared spectroscopy of nucleic acids: application to tautomerism and DNA aptamer unfolding dynamics. Massachusetts Institute of Technology, 2014.

23. Palmer, A. G.; Massi, F., Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **2006,** *106*, 1700-1719.

24. Rangadurai, A.; Szymaski, E. S.; Kimsey, I. J.; Shi, H.; Al-Hashimi, H. M., Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance R1ρ relaxation dispersion. *Prog. Nucl. Mag. Res. Spectrosc.* **2019,** *112*, 55-102.

25. Peng, C. S.; Baiz, C. R.; Tokmakoff, A., Direct observation of ground-state lactam–lactim tautomerization using temperature-jump transient 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 9243-9248.

26. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

27. Bozovic, O.; Zanobini, C.; Gulzar, A.; Jankovic, B.; Buhrke, D.; Post, M.; Wolf, S.; Stock, G.; Hamm, P., Real-time observation of ligand-induced allosteric transitions in a PDZ domain. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 26031-26039.

28. Zhang, Y.; Dood, J.; Beckstead, A. A.; Li, X.-B.; Nguyen, K. V.; Burrows, C. J.; Improta, R.; Kohler, B., Efficient UV-induced charge separation and recombination in an 8-oxoguanine-containing dinucleotide. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111*, 11612-11617.

29. Banyay, M.; Sarkar, M.; Gräslund, A., A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **2003,** *104*, 477-488.

30. Taillandier, E.; Liquier, J., [16] Infrared spectroscopy of DNA. *Method. Enzymol.* **1992,** *211*, 307-335.

31. Elsaesser, T.; Schauss, J.; Kundu, A.; Fingerhut, B. P., Phosphate vibrations probe electric fields in hydrated biomolecules: spectroscopy, dynamics, and interactions. *J. Phys. Chem. B* **2021,** *125*, 3899-3908.

32. Stelling, A. L.; Xu, Y.; Zhou, H.; Choi, S. H.; Clay, M. C.; Merriman, D. K.; Al-Hashimi, H. M., Robust IR-based detection of stable and fractionally populated G-C+ and A-T Hoogsteen base pairs in duplex DNA. *FEBS Lett.* **2017,** *591*, 1770-1784.

33. Prescott, B.; Steinmetz, W.; Thomas Jr, G., Characterization of DNA structures by laser Raman spectroscopy. *Biopolymers* **1984,** *23*, 235-256.

34. Greve, C.; Elsaesser, T., Ultrafast Two-Dimensional Infrared Spectroscopy of Guanine–Cytosine Base Pairs in DNA Oligomers. *J. Phys. Chem. B* **2013,** *117*, 14009-14017.

35. Peng, C. S.; Jones, K. C.; Tokmakoff, A., Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 15650-15660.

36. Lee, C.; Cho, M., Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **2006,** *125*, 114509.

37. Howard, F.; Frazier, J.; Miles, H. T., Interbase vibrational coupling in G: C polynucleotide helices. *Proc. Natl. Acad. Sci. U.S.A.* **1969,** *64*, 451-458.

38. Jiang, Y.; Wang, L., Modeling the vibrational couplings of nucleobases. *J. Chem. Phys.* **2020,** *152*, 084114.

39. Krummel, A. T.; Zanni, M. T., DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: Insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* **2006,** *110*, 13991-14000.

40. Jiang, Y.; Wang, L., Development of vibrational frequency maps for nucleobases. *J. Phys. Chem. B* **2019,** *123*, 5791-5804.

41. Meng, W.; Peng, H.-C.; Liu, Y.; Stelling, A.; Wang, L., Modeling the Infrared Spectroscopy of Oligonucleotides with 13C Isotope Labels. *J. Phys. Chem. B* **2023,** *127*, 2351-2361.

42. Price, D. A.; Kartje, Z. J.; Hughes, J. A.; Hill, T. D.; Loth, T. M.; Watts, J. K.; Gagnon, K. T.; Moran, S. D., Infrared spectroscopy reveals the preferred motif size and local disorder in parallel stranded DNA G-quadruplexes. *ChemBioChem* **2020,** *21*, 2792-2804.

43. Lee, C.; Cho, M., Vibrational dynamics of DNA: IV. Vibrational spectroscopic characteristics of A-, B-, and Z-form DNA's. *J. Chem. Phys.* **2007,** *126*, 04B607.

44. Yang, B.; Rodgers, M., Base-pairing energies of proton-bound heterodimers of cytosine and modified cytosines: implications for the stability of DNA i-motif conformations. *J. Am. Chem. Soc.* **2014,** *136*, 282-290.

45. Dai, Q.; Sanstead, P. J.; Peng, C. S.; Han, D.; He, C.; Tokmakoff, A., Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chem. Biol.* **2016,** *11*, 470-477.

46. Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D., I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018,** *10*, 631-637.

47. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C., Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Cent. Sci.* **2022,** *8*, 1404-1414.

48. McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B., DNA's chiral spine of hydration. *ACS Cent. Sci.* **2017,** *3*, 708-714.

49. Hithell, G.; González-Jiménez, M.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Wynne, K.; Hunt, N. T., Ultrafast 2D-IR and optical Kerr effect spectroscopy reveal the impact of duplex melting on the structural dynamics of DNA. *Phys. Chem. Chem. Phys.* **2017,** *19*, 10333-10342.

50. Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, A. W.; Hunt, N. T., Long-range vibrational dynamics are directed by Watson–Crick base pairing in duplex DNA. *J. Phys. Chem. B* **2016,** *120*, 4009-4018.

51. Fayer, M., Dynamics of liquids, molecules, and proteins measured with ultrafast 2D IR vibrational echo chemical exchange spectroscopy. *Annu. Rev. Phys. Chem.* **2009,** *60*, 21-38.

52. Farrell, K. M.; Yang, N.; Zanni, M. T., A polarization scheme that resolves cross-peaks with transient absorption and eliminates diagonal peaks in 2D spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2022,** *119*, e2117398119.

53. Price, D. A.; Wedamulla, P.; Hill, T. D.; Loth, T. M.; Moran, S. D., The polarization dependence of 2D IR cross-peaks distinguishes parallel-stranded and antiparallel-stranded DNA G-quadruplexes. *Spectrochim. Acta A* **2022,** *267*, 120596.

54. DeFlores, L. P.; Ganim, Z.; Nicodemus, R. A.; Tokmakoff, A., Amide I′− II′ 2D IR spectroscopy provides enhanced protein secondary structural sensitivity. *J. Am. Chem. Soc.* **2009,** *131*, 3385-3391.

55. Valentine, M. L.; Al-Mualem, Z. A.; Baiz, C. R., Pump slice amplitudes: a simple and robust method for connecting two-dimensional infrared and Fourier transform infrared spectra. *J. Phys. Chem. A* **2021,** *125*, 6498-6504.

56. Meuzelaar, H.; Marino, K. A.; Huerta-Viga, A.; Panman, M. R.; Smeenk, L. E.; Kettelarij, A. J.; van Maarseveen, J. H.; Timmerman, P.; Bolhuis, P. G.; Woutersen, S., Folding dynamics of the Trp-cage miniprotein: evidence for a native-like intermediate from combined time-resolved vibrational spectroscopy and molecular dynamics simulations. *J. Phys. Chem. B* **2013,** *117*, 11490-11501.

57. Meuzelaar, H.; Panman, M. R.; van Dijk, C. N.; Woutersen, S., Folding of a zinc-finger ββα-motif investigated using two-dimensional and time-resolved vibrational spectroscopy. *J. Phys. Chem. B* **2016,** *120*, 11151-11158.

58. Gallagher Faeder, S. M.; Jonas, D. M., Two-dimensional electronic correlation and relaxation spectra: Theory and model calculations. *J. Phys. Chem. A* **1999,** *103*, 10489-10505.

59. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. *Proc. Natl. Acad. Sci. U.S.A.* **2023,** *120*, e2219124120.

60. Jones, K. C.; Peng, C. S.; Tokmakoff, A., Folding of a heterogeneous β-hairpin peptide from temperature-jump 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 2828-2833.

61. Tucker, M. J.; Abdo, M.; Courter, J. R.; Chen, J.; Brown, S. P.; Smith III, A. B.; Hochstrasser, R. M., Nonequilibrium dynamics of helix reorganization observed by transient 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 17314-17319.

62. Doty, P.; Marmur, J.; Eigner, J.; Schildkraut, C., Strand separation and specific recombination in deoxyribonucleic acids: physical chemical studies. *Proc. Natl. Acad. Sci. U.S.A.* **1960,** *46*, 461-476.

63. Brahms, J.; Mommaerts, W., A study of conformation of nucleic acids in solution by means of circular dichroism. *J. Mol. Biol.* **1964,** *10*, 73-88.

64. Sapunar, M.; Domcke, W.; Došlić, N., UV absorption spectra of DNA bases in the 350–190 nm range: Assignment and state specific analysis of solvation effects. *Phys. Chem. Chem. Phys.* **2019,** *21*, 22782-22793.

65. Plasser, F.; Aquino, A. J.; Hase, W. L.; Lischka, H., UV absorption spectrum of alternating DNA duplexes. Analysis of excitonic and charge transfer interactions. *J. Phys. Chem. A* **2012,** *116*, 11151-11160.

66. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M., Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **2009,** *37*, 1713-1725.

67. Hardwick, J. S.; Ptchelkine, D.; El-Sagheer, A. H.; Tear, I.; Singleton, D.; Phillips, S. E.; Lane, A. N.; Brown, T., 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **2017,** *24*, 544-552.

68. Hansen, A. L.; Al-Hashimi, H. M., Dynamics of large elongated RNA by NMR carbon relaxation. *J. Am. Chem. Soc.* **2007,** *129*, 16072-16082.

69. Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H., Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angew. Chem. Int. Edit.* **2012,** *51*, 250-253.

70. Braunlin, W. H.; Bloomfield, V. A., Proton NMR study of the base-pairing reactions of d (GGAATTCC): salt effects on the equilibria and kinetics of strand association. *Biochem.* **1991,** *30*, 754-758.

71. Steinert, H. S.; Rinnenthal, J.; Schwalbe, H., Individual basepair stability of DNA and RNA studied by NMR-detected solvent exchange. *Biophys. J.* **2012,** *102*, 2564-2574.

72. Korzhnev, D. M.; Kay, L. E., Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **2008,** *41*, 442-451.

73. Gorenstein, D. G., [14] 31P NMR of DNA. *Method. Enzymol.* **1992,** *211*, 254-286.

74. Scott, L. G.; Hennig, M., 19F-site-specific-labeled nucleotides for nucleic acid structural analysis by NMR. *Method. Enzymol.* **2016,** *566*, 59-87.

75. Yu, B.; Bien, K. G.; Pletka, C. C.; Iwahara, J., Dynamics of cations around DNA and protein as revealed by 23Na diffusion NMR spectroscopy. *Anal. Chem.* **2022,** *94*, 2444-2452.

76. Yu, B.; Bien, K. G.; Wang, T.; Iwahara, J., Diffusion NMR-based comparison of electrostatic influences of DNA on various monovalent cations. *Biophys. J.* **2022,** *121*, 3562-3570.

77. Denisov, V. P.; Halle, B., Protein hydration dynamics in aqueous solution: a comparison of bovine pancreatic trypsin inhibitor and ubiquitin by oxygen-17 spin relaxation dispersion. *J. Mol. Biol.* **1995,** *245*, 682-697.

78. Nonin, S.; Leroy, J.-L.; Guéron, M., Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochem.* **1995,** *34*, 10652-10659.

79. Patel, D. J.; Pardi, A.; Itakura, K., DNA conformation, dynamics, and interactions in solution. *Science* **1982,** *216*, 581-590.

80. Patel, D. J.; Kozlowski, S. A.; Marky, L. A.; Broka, C.; Rice, J. A.; Itakura, K.; Breslauer, K. J., Premelting and melting transitions in the d (CGCGAATTCGCG) self-complementary duplex in solution. *Biochem.* **1982,** *21*, 428-436.

81. Phan, A. T.; Mergny, J. L., Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. *Nucleic Acids Res.* **2002,** *30*, 4618-4625.

82. Han, X.; Leroy, J.-L.; Guéron, M., An intramolecular i-motif: the solution structure and base-pair opening kinetics of d (5mCCT3CCT3ACCT3CC). *J. Mol. Biol.* **1998,** *278*, 949-965.

83. Chen, J.; Dupradeau, F.-Y.; Case, D. A.; Turner, C. J.; Stubbe, J., DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* **2008,** *36*, 253-262.

84. Rucker, S. P.; Shaka, A., Broadband homonuclear cross polarization in 2D NMR using DIPSI-2. *Mol. Phys.* **1989,** *68*, 509-517.

85. Nelissen, F. H.; Tessari, M.; Wijmenga, S. S.; Heus, H. A., Stable isotope labeling methods for DNA. *Prog. Nucl. Mag. Res. Spectrosc.* **2016,** *96*, 89-108.

86. Al-Hashimi, H. M., NMR studies of nucleic acid dynamics. *J. Magn. Reson.* **2013,** *237*, 191-204.

87. Kimsey, I. J.; Szymanski, E. S.; Zahurancik, W. J.; Shakya, A.; Xue, Y.; Chu, C.-C.; Sathyamoorthy, B.; Suo, Z.; Al-Hashimi, H. M., Dynamic basis for dG• dT misincorporation via tautomerization and ionization. *Nature* **2018,** *554*, 195-201.

88. Pintér, G.; Hohmann, K. F.; Grün, J. T.; Wirmer-Bartoschek, J.; Glaubitz, C.; Fürtig, B.; Schwalbe, H., Real-time nuclear magnetic resonance spectroscopy in the study of biomolecular kinetics and dynamics. *Magn. Reson.* **2021,** *2*, 291-320.

89. Shi, H.; Liu, B.; Nussbaumer, F.; Rangadurai, A.; Kreutz, C.; Al-Hashimi, H. M., NMR chemical exchange measurements reveal that N 6-methyladenosine slows RNA annealing. *J. Am. Chem. Soc.* **2019,** *141*, 19988-19993.

90. Gao, X.; Jones, R., Nitrogen-15-labeled oligodeoxynucleotides. Characterization by nitrogen-15 NMR of d [CGTACG] containing 15N6 or 15N1 labeled deoxyadenosine. *J. Am. Chem. Soc.* **1987,** *109*, 3169-3171.

91. Kimsey, I. J.; Petzold, K.; Sathyamoorthy, B.; Stein, Z. W.; Al-Hashimi, H. M., Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. *Nature* **2015,** *519*, 315-320.

92. Lipari, G.; Szabo, A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982,** *104*, 4546-4559.

93. Palmer III, A. G., NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **2004,** *104*, 3623-3640.

94. Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wöhnert, J.; Stock, G.; Schwalbe, H., NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* **2008,** *36*, 1928-1940.

95. Stenström, O.; Champion, C.; Lehner, M.; Bouvignies, G.; Riniker, S.; Ferrage, F., How does it really move? Recent progress in the investigation of protein nanosecond dynamics by NMR and simulation. *Curr. Opin. Struc. Biol.* **2022,** *77*, 102459.

96. Hogan, M. E.; Jardetzky, O., Internal motions in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1979,** *76*, 6341-6345.

97. Mirau, P. A.; Behling, R. W.; Kearns, D. R., Internal motions in B-and Z-form poly (dG-dC). cntdot. poly (dG-dC): proton NMR relaxation studies. *Biochem.* **1985,** *24*, 6200-6211.

98. McConnell, H. M., Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **1958,** *28*, 430-431.

99. Leigh Jr, J., Relaxation times in systems with chemical exchange: Some exact solutions. *J. Magn. Reson.* **1971,** *4*, 308-311.

100. McLaughlin, A.; Leigh Jr, J., Relaxation times in systems with chemical exchange: approximate solutions for the nondilute case. *J. Magn. Reson.* **1973,** *9*, 296-304.

101. Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Guéron, M., Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* **1988,** *200*, 223-238.

102. Guéron, M.; Leroy, J.-L., [16] Studies of base pair kinetics by NMR measurement of proton exchange. In *Methods in enzymology*, Elsevier: 1995; Vol. 261, pp 383-413.

103. Marušič, M.; Schlagnitweit, J.; Petzold, K., RNA dynamics by NMR spectroscopy. *ChemBioChem* **2019,** *20*, 2685-2710.

104. Vallurupalli, P.; Bouvignies, G.; Kay, L. E., Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **2012,** *134*, 8148-8161.

105. Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M., Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011,** *470*, 498-502.

106. Eaton, W. A., Modern kinetics and mechanism of protein folding: A retrospective. *J. Phys. Chem. B* **2021,** *125*, 3452-3467.

107. Ma, H.; Gruebele, M., Kinetics are probe-dependent during downhill folding of an engineered λ6–85 protein. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 2283-2287.

108. Bartlett, A. I.; Radford, S. E., An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* **2009,** *16*, 582-588.

109. Pazdernik, N.; Speicher, N. Oligo synthesis: Why IDT leads the oligo industry.

Appendix 2.A Preparation of DNA samples for IR and NMR spectroscopy

The majority of oligonucleotides used in this thesis were synthesized by Integrated DNA Technologies (IDT) using a modern solid-state synthesis protocol.[109] Synthesis is carried out with phosphoramidite monomers that contain numerous protection groups that must be cleaved from the oligonucleotide and removed from the solution to avoid spectral artifacts and unwanted interactions with the nucleic acid. Canonical DNA oligonucleotides synthesized by IDT are typically fully deprotected, but cleaved protecting groups often remain in the sample even after commercial desalting. Some of the most common leftover protecting groups are shown in Fig. 1. A dimethoxytrityl (DMT) group is used to protect the 5´ hydroxyl of the deoxyribose group and leaves the oligonucleotide as dimethoxytritylchloride (DMTCl) when treated with trichloroacetic acid. DMTCl is weakly soluble in water and is typically observed as a white solid when preparing millimolar concentrations of oligonucleotide from IDT. Therefore, DMTCl can easily be separated. Benzoyl chloride and isobutyryl chloride are formed from protecting groups for the exocyclic amine groups on cytosine/adenine and guanine, respectively.

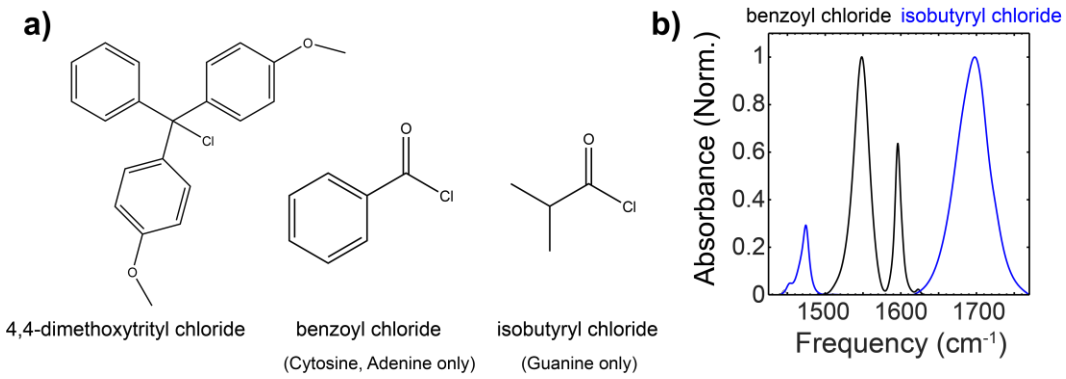

Figure 2.A1 Synthetic oligonucleotide impurities. (a) Chemical structures of some common small molecule impurities leftover from solid-state oligonucleotide synthesis. **(b)** FTIR spectra of benzoyl chloride and isobutyryl chloride mixed in deuterated pH* 6.8 buffer.

In addition to protecting groups, failed oligonucleotide products shorter than the desired may remain in the sample. The yield of desired oligonucleotide depends on the number of nucleobases (N) in the full oligonucleotide and coupling efficiency (C) for each nucleobase addition.

$$Y = C^{N-1} \tag{2.A1}$$

IDT quotes an average C of 99.4%, which leads to full-length product yields greater than 90% for N = 10-15 oligonucleotides. The typical oligonucleotides used in this thesis range from N = 2 to 20, so failed oligonucleotide products are not expected to significantly contribute to our results. However, the final deprotection step requires heating of oligonucleotide solution that may create significant hydrolyzed product from RNA oligonucleotides.

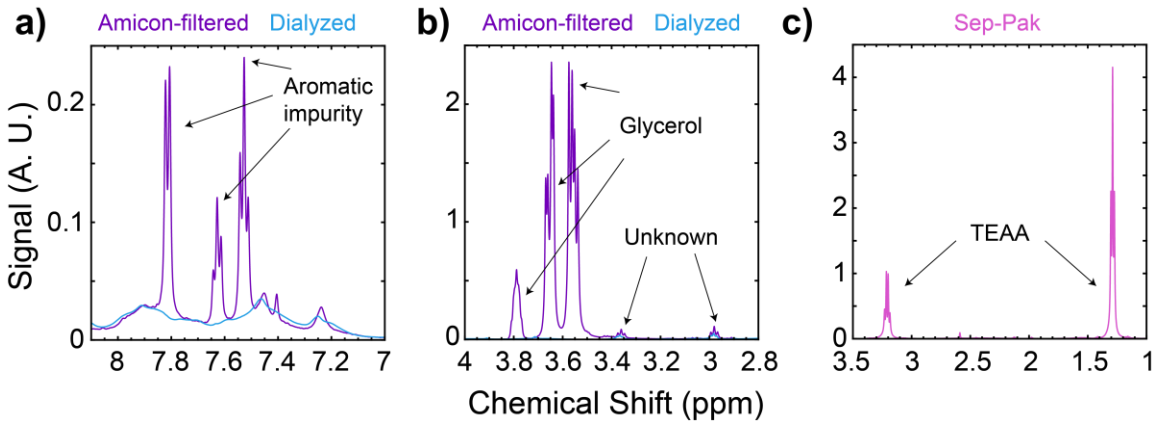

Figure 2.A2 Assessment of small molecule impurities in DNA samples. (a) ^{1}H NMR spectrum in the aromatic region for a DNA duplex sample (5′ -GCGGCGAAGGCGGC-3′ + complement) desalted using Amicon centrifugal filters (0.5 mL, 3 kDa cutoff) or by dialysis in ultrapure water using Slide-a-Lyzer cassettes (3 mL, 3 kDa cutoff). Spectra are measured at 1 °C. **(b)** CH and CH_2 region for the same samples showing the introduction of glycerol by centrifugal filters. **(c)** C18 Sep-Pak desalted sample showing residual triethylammonium acetate (TEAA).

Since we are not too concerned with failed oligonucleotide products, our primary concern is desalting. DNA oligonucleotides from IDT with MW > 2 kDa were desalted using Amicon centrifugal spin filters or through dialysis against ultrapure water using Slide-A-Lyzer cassettes (Thermo). Each method may appear to adequately remove aromatic impurities due to the low sensitivity of IR spectroscopy to aromatic impurities. However, ^{1}H NMR measurements reveal that samples desalted with Amicon filters still retain an aromatic impurity at concentrations similar to the oligonucleotide and also introduce glycerol (Fig. 2.A2). Based on the ^{1}H NMR spectrum, this impurity may be benzoyl chloride. In contrast, dialysis of the oligonucleotides in ultrapure water for 24 hours (refreshing water every 6 hours) removes most of the impurity. An additional alkyl impurity is present and not fully removed by either method. DNA oligonucleotides with MW < 2 kDa were desalted using C18 Sep-Pak cartridges (Waters, 55 – 105 μm, 360 mg sorbent, 125

Å pore size). This method is also effective for removing small aromatic impurities, but typically leads to leftover triethylammonium acetate (TEAA) in the eluted DNA sample (Fig. 2.A2c).

Once the sample is purified to the desired extent, it must be prepped for IR or NMR spectroscopy. Neither of these methods require additional labels or "magic cocktails" as in other methods, so sample preparation is relatively straightforward. Oligonucleotides are first lyophilized to dryness and then re-dissolved in a deuterated buffer at millimolar strand concentrations. Strand concentration is verified using the UV absorption of the nucleobases with a Nanodrop spectrometer (Thermo). To minimize HOD content in the sample, samples are prepared at the desired concentration and pH and then lyophilized again prior to measurement. Samples are always heated to 90 °C and annealed under room temperature conditions before an IR or NMR measurement, which may introduce a small amount of HOD into the sample. Sample preparation for NMR spectroscopy is similar to IR spectroscopy except a frequency reference must be added at a concentration similar to the oligonucleotide, (3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid was used in this thesis) and measurements are performed in fully deuterated solution or 95% H_2O/5% D_2O. In this thesis, all samples were prepared in sodium phosphate buffer at pH* 6.8 to prevent overlap with peaks from the buffer in IR and NMR spectral regions of interest.

Chapter 3

T-jump IR spectroscopy for biomolecular reactions

3.1 Why T-jump for probing biomolecular reactions?

The principle advantage of optical spectroscopic probes like IR, UV absorption, and CD over NMR and single-molecule techniques is that they are capable of probing real-time kinetics on timescales ranging from femtoseconds to hours when incorporated into time-resolved methods. These spectroscopic probes report on the ensemble of molecules, therefore their exchange kinetics are only observed when the system is synchronized by an external perturbation that changes the sample environment and shifts the relative population of states. If the perturbation is faster than the kinetics of interest, then relaxation of the ensemble to the new equilibrium can be tracked as a function of time. Biomolecular conformational states often exhibit small free energy differences arising from counteracting enthalpic and entropic factors, leading to extraordinary sensitivity in the relative populations of conformational states to variables such as temperature, pressure, pH, salt concentration, chemical denaturant, and others.[1-5] Utilizing this sensitivity, many types of triggers have been employed to synchronize biomolecular conformational changes.[6-11]

Rapid changes in temperature, or temperature-jumps (T-jumps), are the relaxation method used throughout this thesis. While one of the oldest relaxation methods,[12] T-jump spectroscopy is also one of the most versatile approaches for time-resolved measurements of biomolecules. T-jumps can be incorporated with nearly any spectroscopic probe,[13-20] and laser excitation enables T-jumps within a few picoseconds such that thermally-induced kinetics occurring within ~10 ps

or slower may be monitored.[21-23] Additionally, the properties of a T-jump (magnitude, duration, etc.) are easily tunable and allow for well-controlled measurements of non-equilibrium dynamics.[24-26]

T-jump spectroscopy is limited in the types of processes that may be studied. A temperature increase usually promotes unfolding or dissociation in biomolecular systems, and therefore the observed signals most directly report on dynamics within the folded or associated ensemble.[9, 24, 27-28] Exceptions are cold denaturation in proteins and phase separation with lower critical solution temperature (LCST) transitions, in which folding and association are directly monitored with T-jump spectroscopy.[29-30] In principle, the temperature-drop (T-drop) back to the initial temperature may be used to study folding or association, but T-drops can at best be achieved on microsecond timescales and are usually limited to milliseconds.[20, 26, 31] Another drawback of T-jump spectroscopy is that some processes simply cannot be synchronized with a rapid temperature change, a typical issue for competitive equilibria between different nucleic acid structures.[32-33] Once the system is equilibrated at a given temperature, T-jumps usually promote the unfolding or dissociation of certain components rather than an interchange between different complexes. Experiments that monitor assembly from a completely dissociated starting point or unsynchronized methods like NMR relaxation dispersion or single-molecule spectroscopy are best suited to characterize these equilibria.

3.2 Using near-IR laser excitation to create T-jumps

T-jump relaxation experiments were pioneered in the 1950s by combining Joule heating (via capacitive discharge) and UV/Vis absorption detection.[12, 34] These apparatuses produced

temperature changes up to ~10 °C with microsecond rise times and enabled the first observations of sub-millisecond reaction kinetics. Capacitive discharge methods were fast enough to resolve DNA duplex dissociation and informed much of the current understanding of DNA duplex hybridization (Chapter 1.2).[35-38] However, the high electrolyte concentration required for Joule heating is undesirable for many studies, and numerous chemical reactions and biomolecular dynamics occur on sub-microsecond timescales inaccessible to Joule heating apparatuses of the time. The advent of Q-switched Nd:YAG lasers and Raman shifting techniques enabled the generation of 1.4 μm pulses that could produce multi-degree optical T-jumps in aqueous solution on <50 ns timescales.[39-41] Optical T-jump methods provide faster temperature changes than capacitive discharge and are suitable with a wider range of sample conditions, leading to their tenure as a mainstay of T-jump spectroscopy and the field of biophysics.

Typical optical heating approaches use near-infrared (NIR) lasers to either excite a dye molecule in solution that quickly transfers energy to the solvent environment or to directly excite high-energy molecular vibrations of H_2O or D_2O.[42-43] The former method can achieve sub-nanosecond changes in temperature of greater than 100 °C but requires high dye concentrations that are likely perturbative to biomolecules. For biomolecular systems, it is instead more common to directly excite water using the first overtone of the O-H and O-D stretch vibrations at 1.45 and 1.95 μm, respectively, or the combination band between the O-H stretch and H_2O bending vibrations at ~1.95 μm. A temperature rise occurs as the excitation of these high-energy vibrations relaxes on ultrafast timescales to populate low-frequency phonon modes of the liquid. Direct excitation of the O-H or O-D stretch fundamental transition produces greater temperature changes than excitation of the overtone but suffers from the generation of large temperature gradients in

the sample that lead to stretched-exponential relaxation kinetics.[44] The low absorption cross-sections of the overtone O-H and O-D stretch vibrations (20 – 30 cm^{-1}) minimize attenuation of the heating beam and the resulting temperature gradient along the direction of beam propagation. T-jump experiments incorporated with IR spectroscopy use short enough sample path lengths ($\leq$ 50 μm) such that 20% or less of the T-jump excitation light is absorbed across the sample. T-jump measurements probed using UV absorption or fluorescence use pathlengths of 100 – 500 μm that absorb most of the T-jump light and create significant temperature gradients across the sample, and counter-propagating T-jump excitation is often implemented to reduce the temperature gradient in these experiments.[16, 43] Our group has developed two T-jump setups based on NIR excitation of H_2O and D_2O using pulsed and continuous wave (CW) laser excitation. Pulsed excitation is necessary to achieve heating on 10 – 1000 picosecond timescales while CW excitation is ideal for probing longer time windows. Each approach has different mechanisms for heating and suffers from distinct limitations and artifacts that are discussed throughout this chapter and Chapter 4.

3.3 Pulsed near-IR T-jump spectroscopy

3.3.1 Sample heating

Pulsed heating is typically achieved through NIR excitation with laser pulses of picosecond or nanosecond duration. For pulses longer than ~20 picoseconds, sample heating is assumed to occur within the pulse duration.[21-22] The pulse duration is typically much shorter than the timescale for thermal diffusion out of the sample cell, which depends on the sample pathlength and thermal properties of the solution and sample windows. Therefore, the temperature change in a pulsed

heating experiment occurs within the pulse duration and its magnitude (ΔT) is determined by the energy initially deposited (*Q*) into the solvent by the laser pulse.

$$Q = \alpha(\omega, T) l E_{pulse} \tag{3.1}$$

In eq. 3.1, α is the absorption coefficient of the sample at a given frequency (ω) and temperature (T), l is the sample path length, and E_{pulse} is the laser pulse energy. The deposited energy creates a ΔT set by the specific heat (C_v) and density (ρ) of the sample as well as the heating volume.

$$\Delta T(T) = \frac{Q}{\rho(T) C_v(T) A l} = \frac{\alpha(\omega, T) E_{pulse}}{\rho(T) C_v(T) A} \tag{3.2}$$

We approximate the heating volume as a cylinder with area *A* determined by the $1/e^2$ diameter of the heating beam. T-jump experiments described or referred to in this thesis are typically performed on dilute aqueous solutions of biomolecules, and their thermal properties are assumed to be equivalent to pure H_2O or D_2O (Table 3.1). In reality, the initial ΔT distribution is not a uniform cylinder. ΔT will match the heating pulse spatial profile (approximately Gaussian) in the direction perpendicular to beam propagation (*r*). Parallel to beam propagation (z), ΔT is attenuated due to absorption of the pulse by the sample.

Table 3.1 Thermal properties of D_2O and common window materials for optical spectroscopy. Density (ρ), specific heat (C_v), thermal conductivity (κ), and thermal diffusivity (D) of D_2O and various sample cell window materials. D is calculated as $D = \kappa/\rho C_v$.

Property	D_2O	CaF_2	BaF_2	MgF_2	FEP	Quartz	CVD Diamond
ρ (g/cm^3)	1.09 (60°C)[45]	3.18 (25 °C)	4.89 (25 °C)	3.18 (25 °C)	2.15 (25 °C)[46]	2.65	3.51
C_v (J/gK)	4.00 (60°C)[47]	0.85 (0 °C)	0.41 (25 °C)	1.00 (25 °C)	1.10 (25 °C)[46]	0.71	0.50 (27 °C)[48]
κ (W/cmK)	6.3×10^{-3} (60°C)[49]	1.0×10^{-1} (25 °C)[50]	1.1×10^{-1} (25 °C)[50]	3.4×10^{-1} (27 °C)[51]	2.0×10^{-3} (25 °C)[46]	6.2×10^{-2} (50 °C)[46]	22 (25 °C)[52]
D (cm^2/s)	1.4×10^{-3}	3.4×10^{-2}	5.5×10^{-2}	1.1×10^{-1}	8.5×10^{-4}	3.3×10^{-2}	12

Given fixed values of ρ and C_v for the solvent, the properties and delivery of the heating laser may be chosen to optimize ΔT. In practice, α, A, and E_{pulse} are limited by factors unique to the spectroscopic technique used to probe the sample. Here we focus on considerations for T-jump spectroscopy employing linear or nonlinear mid-IR spectroscopy as the probe. D_2O is typically used as the solvent for mid-IR spectroscopy of nucleobase vibrations in DNA and RNA or Amide I/II vibrations in proteins. Therefore, T-jumps in this scenario will utilize excitation of the O-D stretch overtone of D_2O. Figure 3.1 shows FTIR spectra of the O-D stretch overtone band from 1 to 96 °C. The frequency of the band maximum shifts from 5030 to 5175 cm^{-1} over this temperature range, leading to changes in the absorbance at single frequencies used for T-jump excitation. For example, the temperature-dependence of α at typical excitation frequencies used in our pulsed (5075 cm^{-1}) and CW T-jump (5150 cm^{-1}) spectrometers (Chapter 4) follow quite different trends.[53] These trends, along with the temperature-dependence of D_2O density and heat capacity,[45, 47] cause

ΔT to increase approximately linearly as a function of the sample temperature for a fixed T-jump excitation energy (Fig. 3.1c).

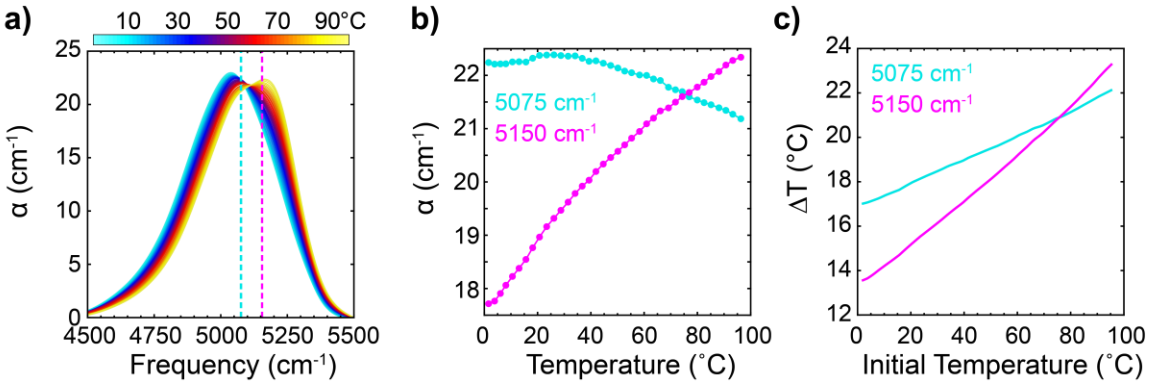

Figure 3.1 Temperature-dependence of the O-D stretch overtone band of D_2O. (a) FTIR temperature series of the O-D stretch overtone band of D_2O from 1 to 96 °C. Spectra are plotted in terms of absorption coefficient (α) using previously reported data.[53] **(b)** Temperature-dependent α at 5075 cm^{-1} (1.97 μm) and 5150 cm^{-1} (1.94 μm), which are the typical T-jump excitation frequencies used in our pulsed and CW T-jump setups, respectively (See Chapter 4). **(c)** Estimation of ΔT from eq. 3.2 as a function of initial temperature using 5075 cm^{-1} and 5150 cm^{-1} excitation frequencies. E_{pulse} = 5 mJ and A = 1.3 x 10^{-3} cm^2. ρ and C_v of D_2O were used from previous studies.[45,47]

3.3.2 Thermal relaxation following a T-jump

Window materials used for mid-IR spectroscopy generally have thermal diffusion coefficients (D) much greater than that of D_2O or H_2O. As a result, the time-dependent temperature profile of a sample following a T-jump is dictated by the thermal properties of the windows and sample pathlength. CaF_2 and BaF_2 are the two most commonly used window materials for mid-IR spectroscopy and have D values of 3.4 × 10^{-2} and 5.5 × 10^{-2} cm^2/s, respectively (Table 3.1). Each of these values are >10-fold larger than that of D_2O such that the ΔT profile induced by the T-jump in D_2O decays by thermal diffusion through the windows.

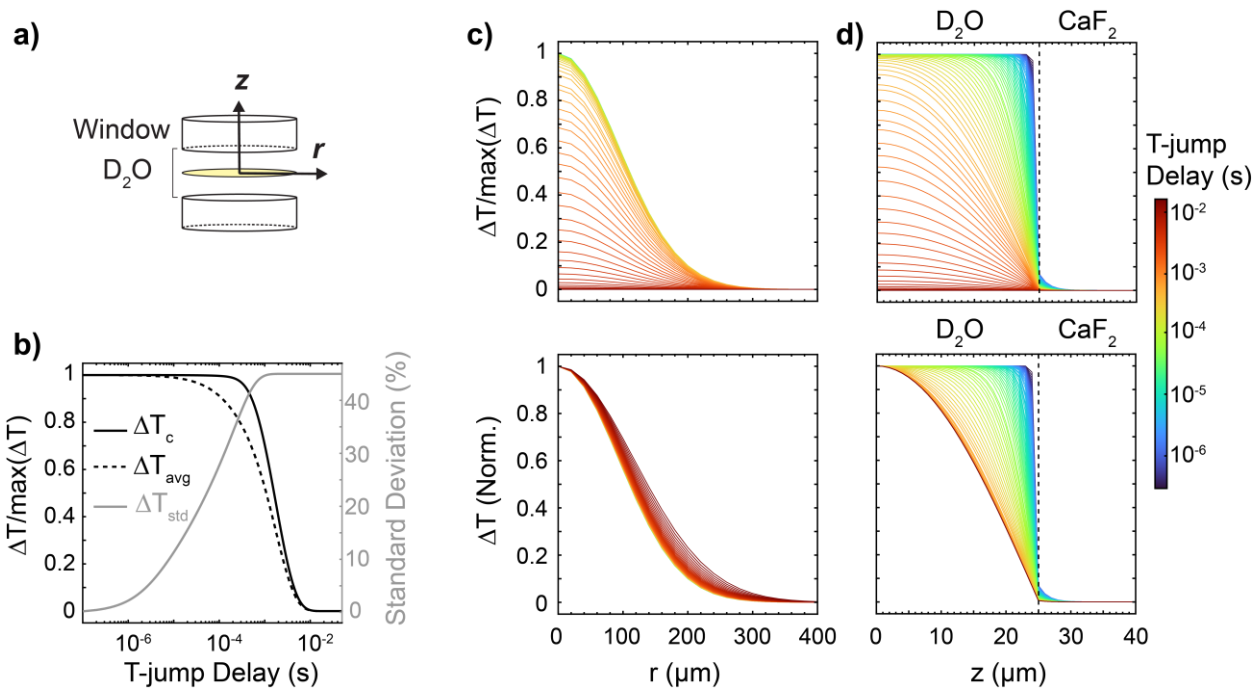

Figure 3.2 Simulation of thermal relaxation following an instantaneous temperature-jump. **(a)** Schematic of sample cell used for simulations containing a cylindrical volume of D_2O between CaF_2 windows. The sample path length (along *z*) is set to 50 μm and the radius of the sample (along *r*) is 500 μm. The thickness of CaF_2 along *z* is 1 mm. **(b)** Simulation of temperature change (ΔT) as a function of time following an instantaneous T-jump out to 50 ms. The 2D heat-diffusion equation (eq. 3.3) was propagated in 100 ns steps using the Crank-Nicolson method.[55] Step sizes of 20 μm and 1 μm were used along *r* and *z*, respectively. The initial ΔT distribution along *r* is a Gaussian with $1/e^2$ diameter of 400 μm. Attenuation of the T-jump along *z* is neglected. The solid black line indicates ΔT at the center of the heated volume (ΔT_c) whereas the dashed line is the average along *r* and *z* (ΔT_{avg}). The gray line indicates the standard deviation of ΔT across *r* and *z* (ΔT_{std}). Time-dependent ΔT distribution along **(c)** *r* and **(d)** *z*. The bottom panels show distributions where ΔT is normalized at each time point. Dashed lines in **(d)** indicate the D_2O- CaF_2 boundary.

To gain insight into the evolution of ΔT following a T-jump, we simulate thermal relaxation through numerical propagation of the 2D heat-diffusion equation in cylindrical coordinates as introduced previously for T-jump experiments.[13, 54]

$$\frac{\partial \Delta T(r,z,t)}{\partial t} = D\left[\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\right)\Delta T(r,z,t) + \frac{\partial^2}{\partial z^2}\Delta T(r,z,t)\right] \tag{3.3}$$

The 2D heat-diffusion equation was propagated using the Crank-Nicolson method[55] with a time step of 100 ns out to 50 ms. Figure 3.2 shows the simulated thermal relaxation for a standard sample design used in our group where a cylindrical volume of D_2O with a path length of 50 μm (along *z*) and radius of 500 μm (along *r*) is placed between 1 mm thick CaF_2 windows. The T-jump is instantaneous relative to the timescale of thermal relaxation. ΔT starts with a Gaussian profile along *r* that matches the spatial mode of the heating beam, and the profile along *z* is uniform (attenuation from absorption is neglected). There is essentially no change in the ΔT distribution up until 1 μs. After 1 μs, heat begins to transfer through the CaF_2 windows, leading to a reduction in the average ΔT value across the sample (ΔT_{avg}). ΔT at the center of the sample (ΔT_c) and ΔT_{avg} follow different decay profiles (Fig. 3.2b). ΔT_c is best described by a compressed exponential, $\Delta T_c(t) = \exp\left[-(t/\tau_D)^{\beta}\right]$, with τ_D = 2.2 ms and β = 1.38 whereas ΔT_{avg} is stretched with τ_D = 1.6 ms and β = 0.90. These differences arise from the creation of a ΔT gradient along *z* due to faster thermal relaxation of solvent near the windows relative to the center whereas diffusion along *r* is negligible on <2 ms timescales (Fig. 3.2c,d). The gradient is quantified by an increase in the standard deviation ΔT along *z* (ΔT_{std}) that follows a stretched exponential with τ_{std} = 0.20 ms and β = 0.87. Therefore, the stretched exponential form of the ΔT_{avg} decay arises from a distribution of decay times along *z*. In typical T-jump IR experiments, the mid-IR probe volume extends across most of the sample path length, and ΔT_{avg} represents a reasonable approximation of the average ΔT within the probe volume.

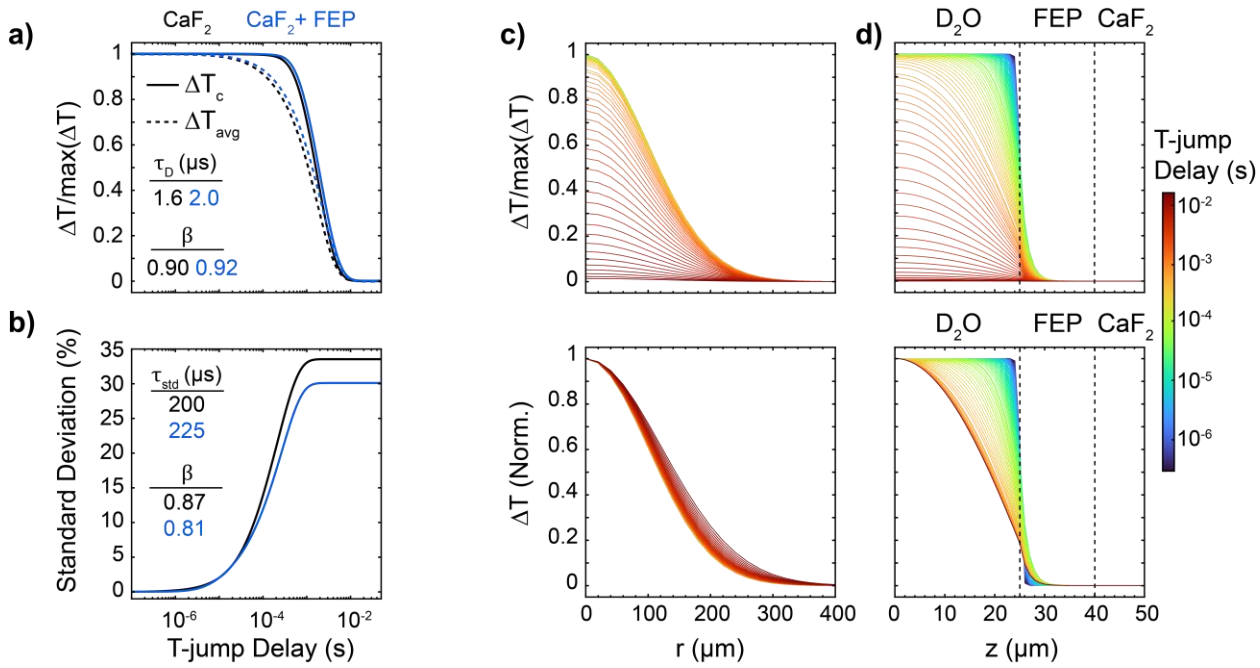

Figure 3.3 Simulation of thermal relaxation following a pulsed T-jump with FEP-coated CaF_2 windows. (a) Simulated time-dependence of ΔT_c (solid lines) and ΔT_{avg} (dashed lines) following an instantaneous T-jump using the same procedure as in Fig. 3.2 except that a 15 μm layer of FEP is added between the D_2O and CaF_2. Profiles for bare CaF_2 are shown in black. **(b)** Time-dependence of ΔT_{std}. Time-dependent ΔT distribution along **(c)** *r* and **(d)** *z*. The bottom panels show distributions where ΔT is normalized at each time point. Dashed lines in **(d)** indicate the D_2O-FEP and FEP-CaF_2 boundaries.

Fast thermal relaxation and temperature gradients along *z* limit the accessible time window and interpretability of pulsed T-jump experiments. While more insulating mid-IR window materials are not commercially available, it is possible to slow thermal relaxation and minimize thermal gradients by coating CaF_2 windows with an insulating film. For this purpose, we have used fluorinated ethylene propylene (FEP, DuPont), which has properties similar to PTFE (i.e. Teflon). FEP has a *D* value >100-fold smaller than CaF_2/BaF_2 (Table 3.1), is transparent in the mid-IR, and can easily be melted onto CaF_2. Figure 3.3 shows a simulation of thermal relaxation where a 15 μm layer of FEP is added between D_2O and CaF_2. The FEP layer slightly extends τ_D

of ΔT_{avg} and reduces the stretched nature of the cooling profile (β = 0.92). The plateau temperature gradient along *z* is reduced from ΔT_{std} = 34 to 30 % (Fig. 3.3b). Experimentally, the addition of a 12.7 µm FEP layer to CaF2 extends τ_D 4-fold relative to bare CaF_2 and greatly reduces the stretch factor (Fig. 3.4). Melted FEP is much more insulating than predicted by the heat diffusion simulations, but the reason for this inaccuracy in the simulation is unclear.

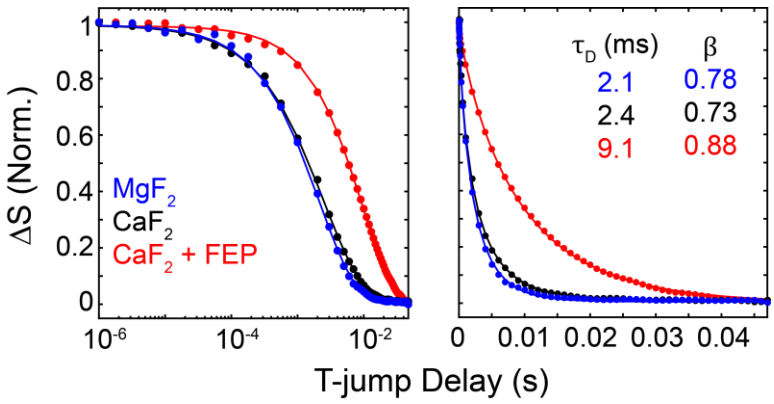

Figure 3.4 Experimental measurement of thermal relaxation following a pulsed T-jump. Time-dependence of mid-IR transmission through 5 mg/ml diglycine in D_2O measured with the pulsed T-jump setup described in Chapter 4.1. Samples with a 50 µm pathlength were placed between 1 mm MgF_2 (blue), 1 mm CaF_2 (black), or 1 mm CaF_2 coated with a 12.7 µm FEP sheet (red). Traces were fit to a stretched exponential $\Delta S(t) = \exp\left(-\left(t/\tau_D\right)^{\beta}\right)$.

The timescale of temperature relaxation is dependent on the distance for thermal diffusion (i.e. sample pathlength for optical spectroscopy). Since temperature relaxation is dominated by thermal diffusion along *z*, we may expect τ_D to follow the quadratic length-dependence for 1D thermal diffusion.

$$\tau_D = \frac{z_{path}^2 C_v \rho}{2\kappa} \tag{3.4}$$

In eq. 3.5, ρ, C_v, and κ correspond to values for D_2O and z_{path} is the distance from the sample center along *z* to the window. Numerical simulations for z_{path} from 4 – 25 μm show a power law relationship between z_{path} and τ_D, $\tau_D = Az_{path}^{\nu}$, with ν = 1.88, which is slightly smaller than predicted for 1D thermal diffusion (ν = 2.0). Over the range of path lengths used in IR spectroscopy, the numerical τ_D value is smaller than the 1D diffusion limit and the two overlap at $2z_{path}$~ 1.5 μm. The normalized ΔT distributions along *r* and *z* are essentially independent of z_{path}. Experimental measurements of τ_D were performed with 50 (z_{path} = 25 μm), 16 (z_{path} = 8 μm), and 6 μm (z_{path} = 3 μm) Teflon spacers as well as without a spacer (z_{path} ~ 1 μm). The change in HDVE signal of diglycine relative to the initial temperature spectrum was used to monitor thermal relaxation (ΔS, Fig. 3.5f). ΔT appears to be independent of sample pathlength as the maximum signal change is nearly the same for each. This observation is consistent with eq. 3.2, which predicts that ΔT is independent of pathlength when heating is much faster than thermal relaxation because both the energy deposited by the laser pulse and heating volume are linearly proportional to pathlength. Experimental τ_D values tend to be slower than those from numerical simulation or eq. 3.4 but do change by similar orders of magnitude in going from z_{path} = 25 to 1 μm. Although a more thorough measurement of the pathlength-dependence of thermal relaxation is needed for quantitative assessment, it is clear that sample pathlength is an important factor to take into account when designing T-jump experiments.

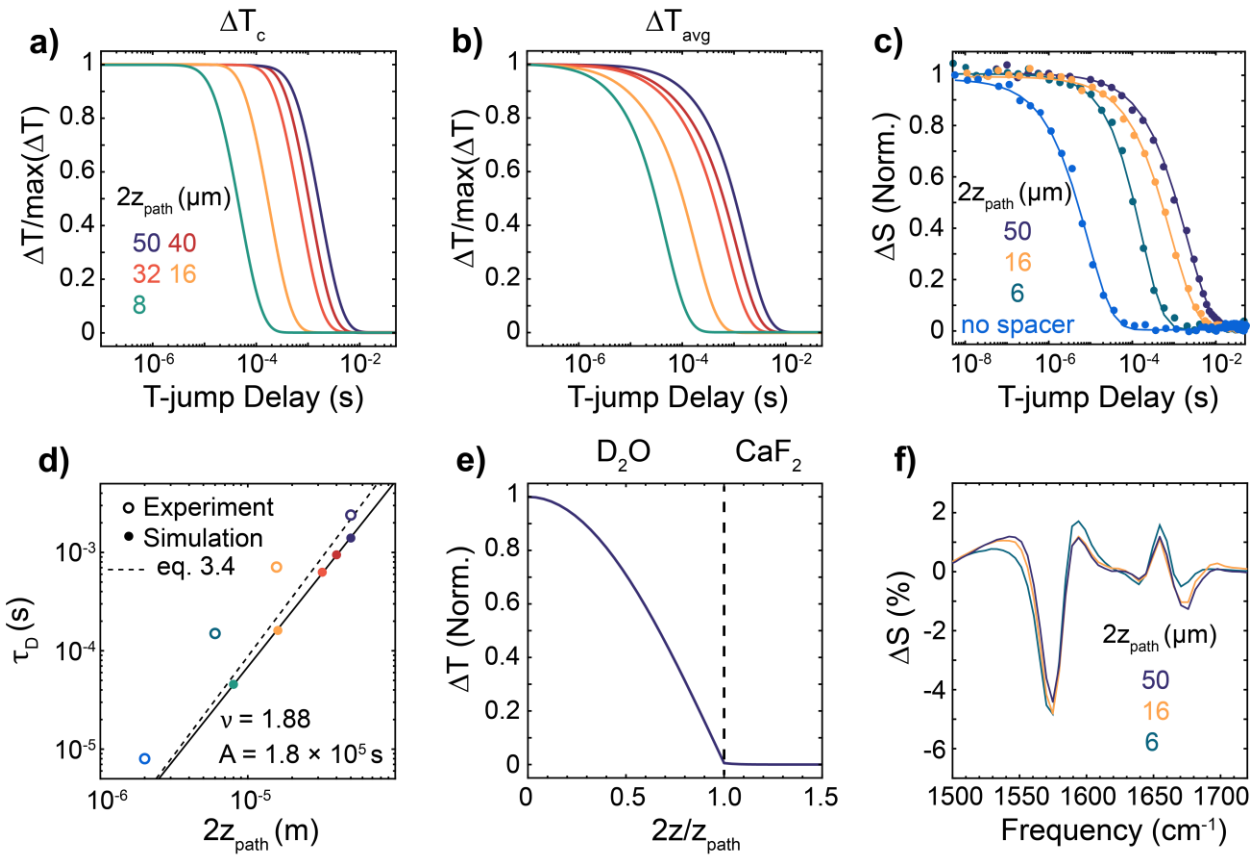

Figure 3.5 Effect of sample pathlength on thermal relaxation. Simulated profiles of **(a)** ΔT_c and **(b)** ΔT_{avg} from sample pathlength ($2z_{path}$) of 50 to 8 μm with CaF_2 windows. **(c)** Experimental cooling profiles measured with pulsed T-jump IR setup for diglycine in D_2O at 50 μm, 16 and 6 μm path lengths and with no spacer (~2 μm) with 1 mm CaF_2 windows. Due to negligible change in mid-IR transmission for small pathlength, cooling profiles are obtained from the t-HDVE signal (Chapter 4.1) of 20 mg/ml diglycine. **(d)** Log-log plot of z_{path} vs. τ_D from simulated ΔT_{avg} (filled circles) and experimental (open circles) results. Simulated results are fit to a power law, $\tau_D = Az_{path}^{\nu}$ (black line). The dashed line corresponds to eq. 3.4. **(e)** Normalized ΔT distribution along z at τ_D. z is normalized by z_{path}. **(f)** t-HDVE signal of diglycine at a 18 ns delay after the T-jump for z_{path} = 25, 8, and 3 μm. Signal is plotted as the percent change relative to the initial temperature spectrum, $\Delta S(t) = (S(t) - S(T_i))/S(T_i)$.

3.3.3 Limitations from cavitation and optical breakdown

It is often desirable to vary the magnitude of a T-jump (ΔT) in order to probe dynamics along the free-energy landscape of a biomolecular transition. However, pulsed T-jumps quickly deposit large amounts of energy into a sample and generate thermal, acoustic, and mechanical effects in the liquid that complicate or inhibit spectroscopy measurements at certain thresholds. At high peak power densities (> 1 GW/cm^2),[56-57] laser excitation will produce a plasma through cascading ionization of water molecules, termed optical breakdown. Our current setup, as well as others,[31, 43, 58] perform T-jumps with ~7 ns, 5-10 mJ pulses that only reach peak power densities of 10-20 MW/cm^2. Although these peak powers are well below the threshold for optical breakdown, other acoustic and mechanical effects may occur and complicate the signal of interest.

T-jumps of ΔT = 5 – 20 °C generate a significant pressure-jump (P-jump) in the heated volume. Full relaxation of the pressure gradient usually occurs more slowly than the T-jump pulse duration and often modulates the spectroscopic observable. The timescale of observed pressure relaxation (τ_p) follows $\tau_p = r/c_W$ where *r* is the radius of the spectroscopic probe beam focus and c_W is the speed of sound in D_2O (~1400 m/s).[59] For our pulsed T-jump setup (Chapter 4.1), *r* ~ 50 μm and τ_p = 36 ns. When the initial P-jump magnitude (ΔP_i) is large enough, pressure relaxation manifests as an apparent increase in mid-IR transmission in our experiments on a 10-100 ns timescale (Fig. 3.5c). ΔP_i is proportional to ΔT and depends on the ratio of the volumetric thermal expansion coefficient (α_T) and isothermal compressibility (β_T) of the liquid, which are both temperature-dependent in H_2O and D_2O.[45, 60]

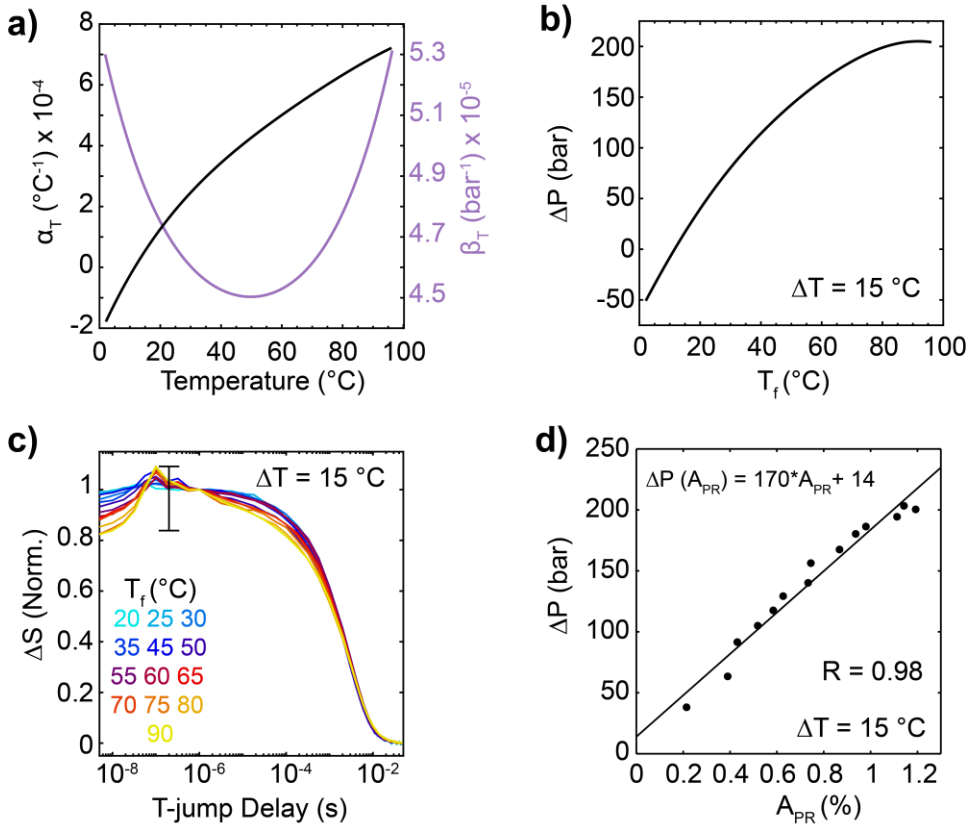

Figure 3.6 Temperature-dependent pressure relaxation in pulsed T-jump experiment. (a) Temperature-dependent volumetric thermal expansion coefficient (α_T, black line) of D_2O determined from empirical density data and isothermal compressibility (β_T, purple line).[45,60] **(b)** Calculated P-jump magnitude immediately after heating (ΔP_i) for T-jumps with ΔT = 15 °C as a function of final temperature (T_f) using eq. 3.4. **(c)** Experimental time-dependent change in mid-IR probe beam transmission following a ΔT ~15 °C T-jump using the setup described in Chapter 4.1. Traces at various T_f are normalized at a delay of 1 μs. A pressure relaxation response is observed from 5 – 100 ns and grows in amplitude (A_{PR}) with increasing T_f **(d)** Scatter plot of A_{PR} vs. predicted ΔP_i shows good correlation (R = 0.98). The black line corresponds to a linear fit.

$$\Delta P(T) = \frac{\alpha_T(T)}{\beta_T(T)} \Delta T \tag{3.5}$$

Figure 3.6 shows the temperature-dependence of α_T, calculated from the reported ρ values, and β_T in D_2O. It is well known that liquid D_2O exhibits a maximum in ρ at ~11.2 °C ($T_{\rho,max}$),

therefore T-jumps with a final temperature (T_f) below $T_{\rho,max}$ induce a small and negative ΔP_i. As T_f increases above $T_{\rho,max}$, ΔP_i becomes increasingly positive before reaching a maximum at ~90 °C. Indeed, the calculated magnitude of ΔP_i is correlated with the experimental increase in mid-IR transmission (Fig. 3.6c-d). The mechanism by which pressure relaxation modulates the mid-IR probe is unclear but likely results from a lensing effect generated by the pressure wave.

Wray et al.[54] performed numerical simulations of pressure relaxation following a T-jump of D_2O in an IR spectroscopy sample cell. For $T_f > T_{\rho,max}$, the P-jump is compressive ($\Delta P_i > 0$) and pressure relaxation leads to a significant tensile pressure change in the center of the heated region by 100 ns, $\Delta P(100\text{ ns}) < 0$. The magnitude of the tensile pressure is proportional to ΔP_i, and if large enough, can initiate bubble formation. This bubble formation process is typically referred to as cavitation. Large T-jumps where $T_f < T_{\rho,max}$ can also initiate cavitation by directly inducing a large tensile pressure upon heating. The threshold for cavitation is highly sensitive to the level of impurities in the solution, the sample window material, and the T-jump beam focus length relative to the solution pathlength. Cavitation typically occurs on a 1 – 10 μs timescale and imparts a much larger distortion to the mid-IR probe than the pressure relaxation, but is only occasionally observed in our group's studies of nucleic acids and proteins in aqueous solution. When observed, replacement of the sample windows and/or filtering of the sample solution is sufficient to remove the cavitation artifact. It is important to note that cavitation and optical breakdown may be limiting factors for future T-jump experiments with larger ΔT.

3.4 CW near-IR T-jump spectroscopy

3.4.1 Predicting the sample temperature profile

The time window of pulsed T-jump spectroscopy is limited by thermal relaxation through the sample windows on a timescale of a few milliseconds. Optical heating with a continuous wave (CW) laser gets around this limitation by providing a constant rate of energy deposition out to arbitrarily long times while being limited to relatively slow heating rise times (τ_R). To date, CW heating has been incorporated to control temperature in single-molecule spectroscopy measurements or in conjunction with fluorescence spectroscopy.[20, 61-62] The mechanism, practical considerations, and potential artifacts from CW optical heating differ substantially from pulsed heating. The CW excitation provides a source of time-dependent energy deposition (Q) dictated by α, l, and the laser power (P_{CW}).

$$\frac{\partial Q}{\partial t} = \alpha(\omega, T) l P_{CW} \tag{3.6}$$

Initially, the rate of temperature change ($d\Delta T/dt$) in the solution depends only on dQ/dt and C_v of the sample. Assuming a cylindrical heating volume defined by Al:

$$\left(\frac{\partial \Delta T}{\partial t}\right)_{t=0} = \frac{\left(\frac{\partial Q}{\partial t}\right)}{C_v(T) A l} = \frac{\alpha(\omega, T) P_{CW}}{C_v(T) A} \tag{3.7}$$

By 0.1 – 1 ms delays after the start of heating, the rate of heating slows due to thermal diffusion through the sample windows and approaches a steady-state temperature (ΔT_0). From Newton's

law of cooling, the sample temperature will approach ΔT_0 with an exponential function described by a time constant τ_R.

$$\frac{\partial \Delta T}{\partial t} = -\frac{\Delta T(t)}{\tau_R} \tag{3.8}$$

$$\Delta T(t) = \Delta T_0 \left(1 - \exp\left(-t / \tau_R\right)\right) \tag{3.9}$$

From eq. 3.9, τ_R may be expressed in terms of the ratio between ΔT_0 and the initial rate of heating.

$$\tau_R = \frac{\Delta T_0}{\left(\frac{\partial \Delta T}{\partial t}\right)_{t=0}} \tag{3.10}$$

Both ΔT_0 and initial rate of heating are proportional to P_{CW}, α, and l, therefore τ_R is essentially independent of P_{CW}.

Numerical simulations of CW optical heating in D_2O are shown in Fig. 3.7. Using eq. 3.6, the CW laser is incorporated into the simulation by adding a constant temperature increase (dT) at each time step (dt).

$$dT\left(r\right) = \frac{P_{CW}\left(r\right)\alpha\left(\omega, T\right)}{C_v(T)A}dt \tag{3.11}$$

P_{CW} has a Gaussian profile along r with a $1/e^2$ diameter of 400 μm and a maximum power of 1.5 W ($P_{CW,max}$). As for the simulations in Section 3.3, we neglect the small z-dependence of dT arising from attenuation of the T-jump through the sample. The simulations show exponential heating with τ_R = 1.7 ms set by thermal diffusion through CaF_2. Once ΔT_0 is reached, the ΔT

distribution along *z* is heterogeneous due to faster thermal diffusion near the windows relative to the center of the sample. This spread in ΔT is equivalent to that at >1 ms delays following pulsed heating (Fig. 3.2) and indicates an important difference between CW and pulsed heating. Following a CW T-jump, the ΔT distribution is heterogeneous but constant in time whereas pulsed T-jumps begin with a uniform ΔT distribution that becomes increasingly heterogeneous on a 50 μs – 1 ms timescale. As a consequence, the T-drop after CW heating follows an exponential rather than the stretched-exponential decay following pulsed heating (Fig. 3.7b).

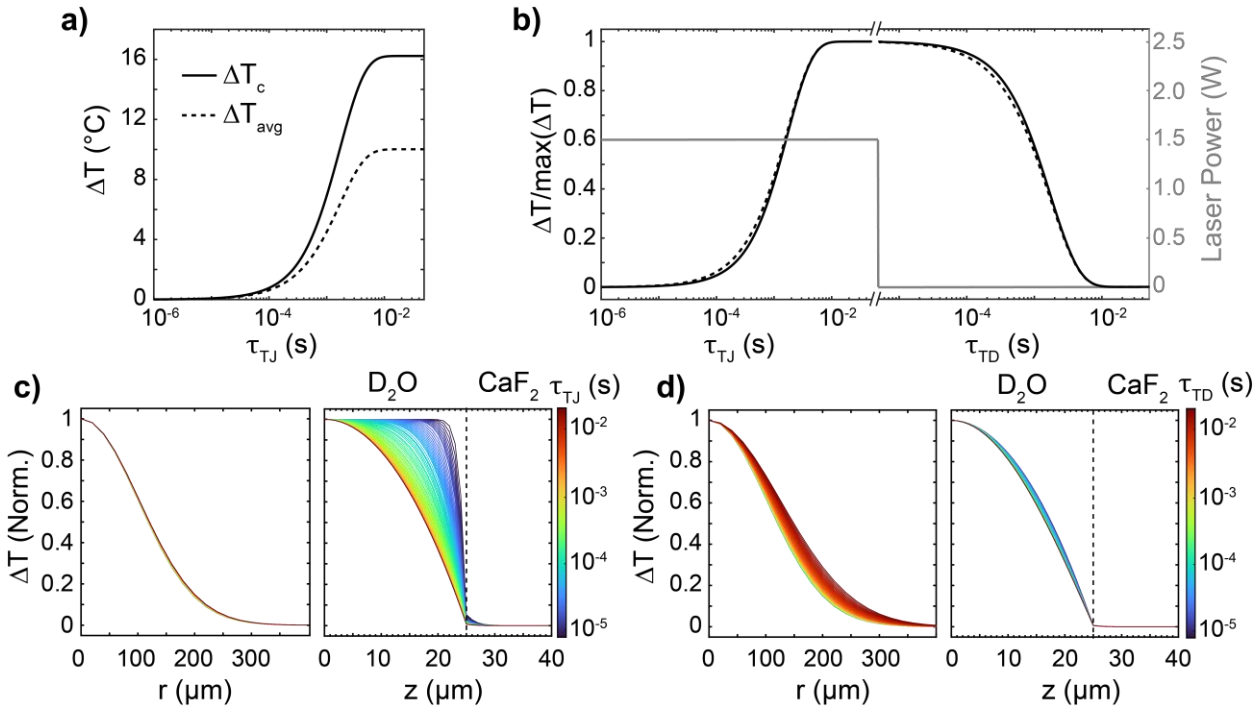

Figure 3.7 Simulations of CW optical heating. (a) Time-dependence of ΔT_c (solid) and ΔT_{avg} (dashed) upon constant heating with a 1.5 W laser input using α at 1.94 μm and 20 °C and with z_{path} = 25 μm. Simulations were propagated in 1 μs time steps, and steps of 20 μm and 1 μm along *r* and *z*, respectively. **(b)** T-jump/T-drop cycle for normalized ΔT_c and ΔT_{avg}. The light gray line indicates the input laser power. **(c)** Normalized ΔT distributions along *r* and *z* during heating and **(d)** during cooling once the input power is turned off.

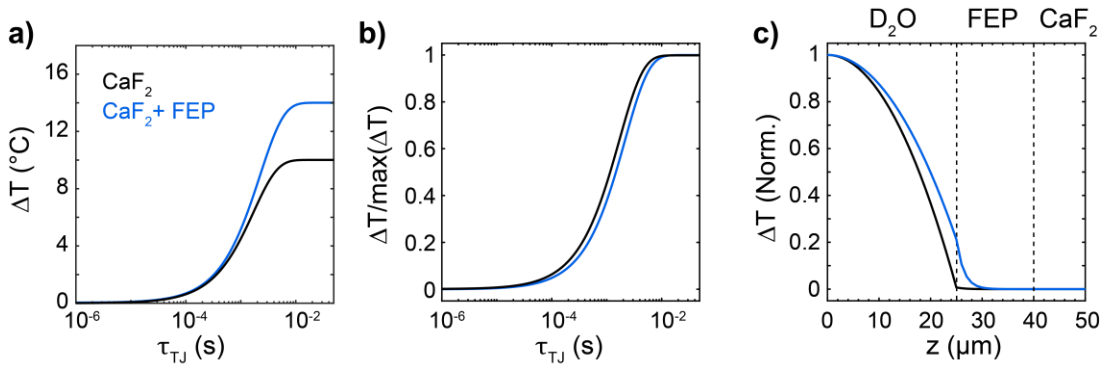

Figure 3.8 Effect of different window material on CW heating. (a) Time-dependence of ΔT_{avg} upon constant heating with a 1.5 W laser for bare CaF_2 and CaF_2 with a 15 μm of FEP. **(b)** Normalized ΔT_{avg} profiles. **(c)** Normalized ΔT distribution along z at τ_{TJ} = 10 ms.

It is evident from eq. 3.7 and the numerical simulations that τ_R, ΔT_0, and heterogeneity in ΔT along z all depend on the sample window properties and solution pathlength as observed for thermal relaxation after a pulsed T-jump. τ_R and τ_D are proportional to each other such that a more insulating window material will extend both timescales. Eq. 3.10 also indicates that changing τ_R and τ_D will proportionally change ΔT_0. This relationship between τ_R and ΔT_0 is exemplified through a 1.4-fold increase in ΔT_0 upon CW heating with bare CaF_2 vs. FEP-coated CaF_2 windows (Fig. 3.8). As for a pulsed T-jump (Fig. 3.3), the numerical simulations underestimate the insulating effect of FEP, and we observe a 2.5-fold increase in ΔT_0 experimentally (Chapter 4.4). ΔT_0 also follows a power-law dependence on z_{path}, $\Delta T_0 = Az_{path}^{\nu}$, with ν = 1.95 under constant P_{CW}, which is nearly identical to the power law relationship between z_{path} and τ_R that arises from thermal diffusion to the windows (Fig. 3.9). While faster heating and cooling at shorter pathlengths is advantageous for studying relaxation kinetics, the nearly quadratic drop off in ΔT_0 is a severe limitation of the CW T-jump approach.

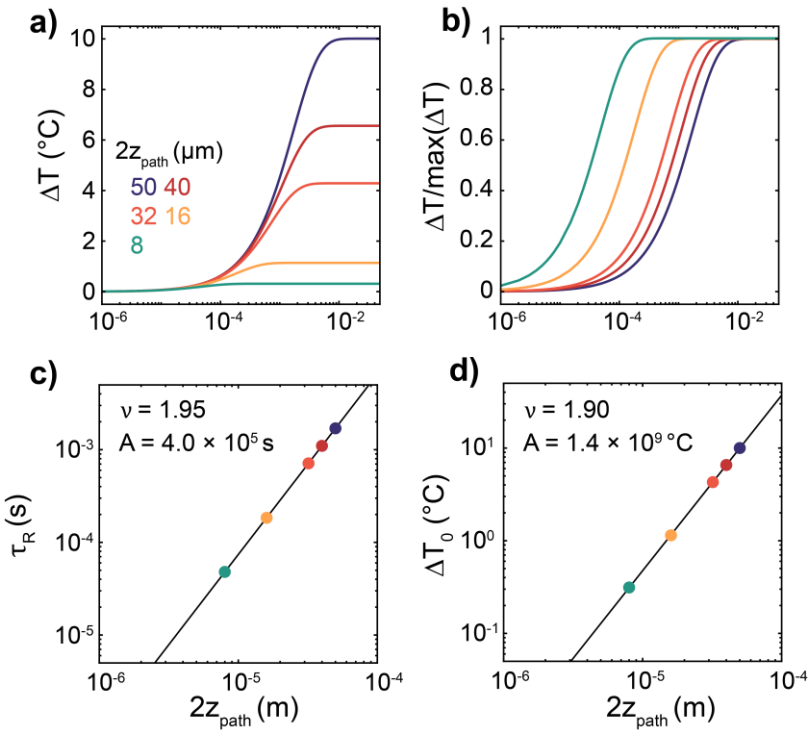

Figure 3.9 Effect of sample pathlength on CW heating. (a) Simulated time-dependence of ΔT_{avg} upon heating for z_{path} = 4 to 25 μm using P_{CW} = 1.5 W. **(b)** Normalized ΔT_{avg} profiles. **(c)** Log-log plot of z_{path} vs. τ_R from ΔT_{avg} profiles and **(d)** z_{path} vs. ΔT_0. Each trend is fit to a power law, $\tau_D = Az_{path}^{\nu}$ and $\Delta T_0 = Az_{path}^{\nu}$, shown as black lines.

One strategy to reduce τ_R without altering τ_D is to modulate P_{CW} in time. P_{CW} can start at a large value, where the instantaneous rate of heating is high, and quickly reduce to a plateau value ($P_{CW,0}$) that sets ΔT_0. Numerical simulations indicate that such a time profile can reduce τ_R by more than 10-fold (Fig. 3.10). The speed up in τ_R will ultimately be limited by how high P_{CW} can be spiked at early time. In practice, P_{CW} is modulated using an acousto-optic modulator, and waveforms more complicated than a step function must be used to minimize time-dependent variation in the final temperature as discussed in Chapter 4.4. Although τ_R can be reduced, the simulations show that the timescale for thermal gradient formation along z is still set by thermal

diffusion to the CaF_2 windows and is similar with or without intensity-modulation. As in pulsed T-jump experiments, the temperature distribution of the sample along *z* changes in time up to ~1 ms even when ΔT_{avg} is essentially constant, and therefore the observed time-dependent thermal profile will likely be sensitive to the shape of the spectroscopic probe volume.

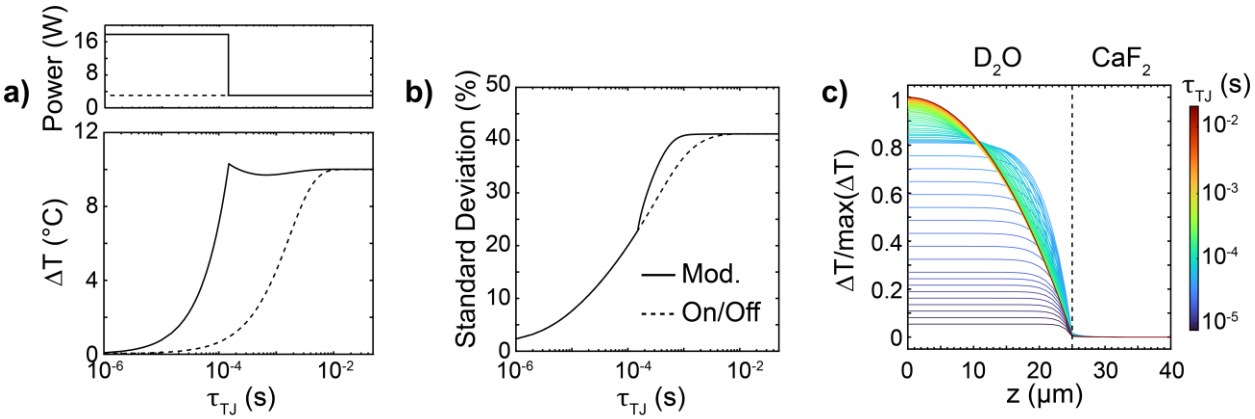

Figure 3.10 Acceleration of CW heating using intensity-modulation. (a) Time-dependence of ΔT_{avg} upon CW heating using a time-dependent (Mod., solid line) and time-independent (On/Off, dashed line) laser power. (Top) The time-dependent power is at 17 W for the first 140 μs and then steps down to a plateau power of 1.5 W used in the time-independent power case. Both simulations were performed with z_{path} = 25 μm. **(b)** Increase in the standard deviation of ΔT standard deviation across z for each type of heating. **(c)** Time-dependent ΔT distribution along *z*.

3.4.2 Thermophoresis

In CW heating, τ_R is long enough that pressure relaxation or cavitation do not contribute to the observed response and the peak laser power densities are well below the threshold for optical breakdown. However, the long-time behavior may still be influenced by thermophoresis, which is the generation of a solute concentration gradient (∇c) proportional to the produced temperature gradient (∇T). Although thermophoresis was discovered more than 150 years ago,[63] its theoretical basis is only partially understood. Under a temperature gradient, the drift velocity (*J*) of a dilute

solute of concentration c depends on both the Brownian diffusion coefficient (D) and thermophoretic mobility (D_T).

$$J = D\nabla c - cD_T\nabla T \tag{3.11}$$

At steady-state (J = 0), ∇c is related to ∇T by the ratio between D_T and D, known as the Soret coefficient (S_T).

$$\nabla c = -cS_T\nabla T \text{ where } S_T = D_T / D \tag{3.12}$$

While D may be accurately predicted from solvent viscosity and solute size and charge, D_T is basically empirical. D_T is often positive, meaning that solute moves from hot to cold regions (thermophobic), but D_T can also be negative (thermophilic). Even for small molecules, D is small enough in water that the timescale for establishing the steady-state ∇c is much slower than τ_R. As a result, thermophoresis typically appears as a time-dependent depletion of signal at long times (>100 ms, Chapter 4.4).

D_T is highly dependent on the solution conditions as well as the type of solute. Since D also depends on many of the same factors, it is most convenient to compare thermophoresis across conditions using S_T. In DNA, S_T is usually positive (S_T ~ 0.01-0.1 °C^{-1} for a 10-mer) but depends on the concentration of counterions, strand length, and temperature.[64] S_T shows a linear increase with temperature, which is thought to arise from hydration effects, and increases proportionally to the square root of strand length ($S_T \propto$ L$^{0.5}$). Also relevant to investigation of nucleic acids is that the concentration of counterions can significantly influence S_T through thermoelectric effects.[65] Under a thermal gradient cations and anions may undergo different thermophoretic migration,

leading to net separation of positive and negatively charged species. In this scenario, a thermoelectric field is generated which can counter or enhance S_T for the molecule of interest. For DNA, S_T has been shown to increase linearly with the Debye length (λ_D). The combination of size, electrostatic, and temperature effects on S_T is commonly described in terms of entropic contributions from ionic shielding and hydration.[64]

$$S_T = \frac{A}{k_B T}\left(-\Delta S_{hyd} + \frac{\beta\sigma_{eff}^2}{4\varepsilon\varepsilon_0}\lambda_D\right) \quad \text{where } \beta = 1 - \frac{T}{\varepsilon}\frac{\partial\varepsilon}{\partial T} \tag{3.13}$$

A is the surface area of the solute, ε is the dielectric constant of water, σ_{eff} is the effective charge of the solute, and ΔS_{hyd} is the entropy of hydration. As shown in Chapter 4.4, the time-dependence of thermophoretic diffusion can contribute signal changes in T-jump spectroscopy that obscure the desired kinetics of the solute. On the other hand, thermophoresis is now established as a highly sensitive probe of binding thermodynamics for proteins and nucleic acids and basis for optical manipulation technology.[66-68]

3.5 Comparison with T-jump experiments that use non-IR probes

We have so far focused on pulsed and CW optical heating for use with mid-IR spectroscopy, but T-jump is highly versatile in the sense that it can be coupled with many different spectroscopic probes. While heating is still typically performed through NIR excitation, the capabilities of T-jump and relevance of certain artifacts will depend on the probe it is implemented with. The two primary factors that differ across T-jump methods are the sample window materials and the size of the heating volumes in the sample. Multiple window materials offer high-quality

optical properties for mid-IR spectroscopy, but most tend to have similar thermal properties. An exception is CVD diamond (Table 3.1), which has the highest thermal conductivity of any material but is also one of the most expensive optical materials. Aside from IR spectroscopy, bulk and single-molecule fluorescence and UV absorption are potentially the most common spectroscopic probes coupled with T-jumps.[16, 43, 69-71] Quartz cuvettes, which have a similar thermal diffusivity to CaF_2 (Table 3.1), are typically used as sample cells for bulk fluorescence and UV absorption. However, these measurements use a sample path length of 100 – 500 μm, which from the scaling found in our numerical simulations would extend τ_D to ~100 ms (Fig. 3.5d), and such timescales are observed experimentally.[16, 69] Rather than directly heat the solvent, T-jumps may be performed through thermal excitation of a metal substrate in contact with the sample.[20, 72] This approach has been utilized in fluorescence microscopy setups to enable τ_R and τ_D values shorter than 10 μs with CW heating. The metal film is only in thermal contact with one half of the sample, therefore thin sample path lengths (<1 μm) are required to minimize ΔT heterogeneity across the probe volume.

When incorporated into a microscope, visible probe lasers can be focused to spot sizes more than 100-fold smaller than the typical 75 – 100 μm $1/e^2$ diameters used in 2D IR spectrometers. The heating beam diameter should be kept much larger than the probe beam diameter to minimize ΔT heterogeneity along *r*. Therefore, a 100-fold reduction in probe beam diameter enables a similar reduction in the T-jump beam diameter to achieve ~50 pL heating volumes[61] whereas those in T-jump 2D IR experiments are ~5 nL (Chapter 4). Small heating volumes are beneficial because they increase heating efficiency and may be combined with a reduction in z_{path} to achieve shorter τ_R (only for CW heating) and τ_D while maintaining large

ΔT.[20, 73] Further, the decrease in heating area along *r* may remove pressure relaxation from the experimental time window for nanosecond pulsed experiments. Optical heating could be incorporated into a 2D IR microscopy setup,[74] enabling a 4-5 fold reduction in the heating volume.

3.6 Modelling of Relaxation Kinetics

3.6.1 Perturbative two-state kinetics

The observed time constants from relaxation experiments contain a combination of microscopic contributions, and a reasonable kinetic model must be chosen to extract meaningful kinetic information from T-jump data. One of the most common assumptions for kinetic modelling of bimolecular reactions is that the induced concentration change is small enough to be "perturbative" and follow a linearized rate equation.[75-76] For example, we can consider a two-state scheme for hybridization between two DNA complementary single-strands (S_1, S_2) to form a duplex (D).

$$S_1 + S_2 \underset{k_d}{\overset{k_h}{\rightleftharpoons}} D \tag{3.14}$$

The rate equations for each species in terms of the hybridization (k_h) and dehybridization (k_d) rate constants

$$\frac{d[S_1]}{dt} = \frac{d[S_2]}{dt} = -k_h[S_1][S_2] + k_d[D] \tag{3.15a}$$

$$\frac{d[D]}{dt} = k_h[S_1][S_2] - k_d[D] \tag{3.15b}$$

During a T-jump experiment, [S_1], [S_2], and [D] are time-dependent and can be expressed in terms of the deviation from the equilibrium value (Δ[X]) at the final temperature ($[X]_{T_f}$).[75]

$$[S_1] = [S_1]_{T_f} + \Delta[S_1] \tag{3.16a}$$

$$[S_2] = [S_2]_{T_f} + \Delta[S_2] \tag{3.16b}$$

$$[D] = [D]_{T_f} + \Delta[D] \tag{3.16c}$$

By mass conservation, we have that the magnitude of concentration change is equivalent for all three species, $\Delta[S_1] = \Delta[S_2] = -\Delta[D] = \Delta c$, and defined by Δc. Then, we can substitute Δc into eqs. 3.15 and 3.16 to re-express the rate equations in terms of concentration change.

$$\frac{d[S_1]}{dt} = \frac{d[S_2]}{dt} = -\frac{d[D]}{dt} = \frac{d\left([S_1]_{T_f} + \Delta c\right)}{dt} \tag{3.17}$$

$[S_1]_{T_f}$ is an equilibrium value such that $\frac{d[S_1]_{T_f}}{dt} = 0$ and the rate equations can be expressed in terms of the rate of concentration change.

$$\frac{d\Delta c}{dt} = -k_h \Delta c^2 - k_h \left([S_1] + [S_2]\right)\Delta c - k_d \Delta c \tag{3.18}$$

Eq. 3.18 is a first-order nonlinear differential equation known as a Ricatti equation, which has been solved previously[75-76] to obtain $\Delta c(t)$.

$$\Delta c(t) = \frac{\Delta c_0}{1 + \frac{k_h \Delta c_0}{\lambda_{obs}}\left[1 - \exp(-\lambda_{obs} t)\right]} \exp(-\lambda_{obs} t) \tag{3.19}$$

$$\text{where, } \lambda_{obs} = k_h\left([S_1] + [S_2]\right) + k_d$$

$$\Delta c_0 = [S_1](t=0) - [S_1]_{T_f} = [S_2](t=0) - [S_2]_{T_f} = [D]_{T_f} - [D](t=0)$$

Δc_0 refers to the max value of Δc that occurs at time zero of the experiment. Removal of the $-k_h \Delta c^2$ term from eq. 3.18 corresponds to linearization or the "perturbative" assumption that is often applied in relaxation spectroscopy.

$$\frac{d\Delta c}{dt} \approx -k_h\left([S_1] + [S_2]\right)\Delta c - k_d \Delta c \tag{3.20}$$

$$\Delta c(t) \approx \Delta c_0 \exp(-\lambda_{obs} t) \tag{3.21}$$

The single-exponential decay of Δc in eq. 3.21 is more simple to apply than eq. 3.19, and the perturbative assumption greatly simplifies derivation of kinetic models beyond two-state scenarios. However, it is important to assess when a T-jump can be considered perturbative.

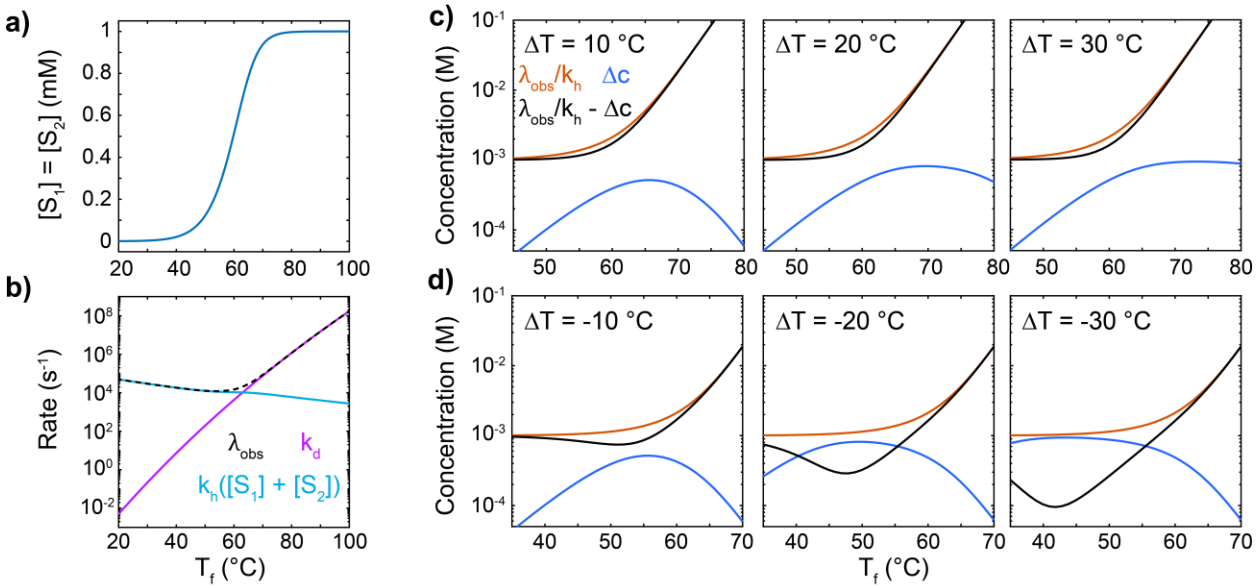

Figure 3.11 Assessment of perturbative kinetics assumption for DNA hybridization. (a) $[S_1]$ and $[S_2]$ as a function of temperature for model data where the dehybridization enthalpy (ΔH_d°) is 310 kJ/mol and entropy (ΔS_d°) is 870 J/molK. Calculations are performed for $[S_1] = [S_2]$ and a total oligonucleotide concentration of 2 mM. **(b)** Model temperature-dependent k_h, k_d, and λ_{obs} from hybridization barriers $\Delta H_h^\ddagger$= -50 kJ/mol and $\Delta S_h^\ddagger$ = -294 J/molK and dehybridization barriers $\Delta H_d^\ddagger$= 260 kJ/mol and $\Delta S_d^\ddagger$ = 576 kJ/mol. **(c)** Calculation of λ_{obs}/k_h (orange), absolute value of Δc (blue), and $\lambda_{obs}/k_h - \Delta c$ (black) as a function of final temperature (T_f) for a T-jump. Each column corresponds to a different ΔT. **(d)** λ_{obs}/k_h, absolute value of Δc, and $\lambda_{obs}/k_h - \Delta c$ for T-drops to a given T_f. For large T-drops, $\lambda_{obs}/k_h \sim \Delta c$ at lower temperatures and violates the perturbative assumption.

The general criteria for a perturbation is $\lambda_{obs}/k_h \gg \Delta c$ such that eq. 3.19 simplifies to eq. 3.21, but when is this true for a real system? Figure 3.11 shows the temperature-dependence of λ_{obs}/k_h and Δc for DNA oligonucleotide hybridization based on data from Chapter 6. For DNA hybridization and protein-protein association, k_d typically follows Arrhenius behavior with a large enthalpic barrier ($\Delta H_d^\ddagger$ > 100 kJ/mol) whereas k_h is much less temperature-sensitive. As a result, λ_{obs}/k_h is nearly constant below the melting transition (T_m) and increases exponentially above it

(Fig. 3.11c-d). The magnitude of Δc depends on ΔT and the width of the melting transition. In our example, T-jumps always fall within the perturbative regime because the Δc is largest where λ_{obs}/k_h is also large. This is further illustrated in Fig. 3.12 where the time-dependence of Δc calculated from linearized (eq. 3.21) and non-linear (eq. 3.19) rate equations are nearly identical under all T-jump conditions. This behavior makes sense because T-jumps with the largest Δc end at final temperatures where k_d dominates and the second-order association process has less impact on the observed kinetics.

It is much more likely for the perturbative assumption to break down in T-drop (ΔT < 0) experiments. Most T-drop experiments are performed with T_f below T_m where association contributes to or dominates the observed kinetics and Δc can approach λ_{obs}/k_h, leading to differences in linearized and non-linearized Δc time profiles even for ΔT = -10 °C (Figs. 3.11d & 3.12b). The non-perturbative Δc decay starts at earlier times than in the perturbative case, but the two overlap at long times when $\lambda_{obs}/k_h \gg \Delta c$. These differences are enhanced as the magnitude of ΔT increases. The relaxation properties of T-drop experiments are similar to those of stopped-flow or microfluidic mixing measurements of association. In those cases, the time-dependence of Δc is often found to deviate from an exponential decay and must be described with second-order kinetics.[77]

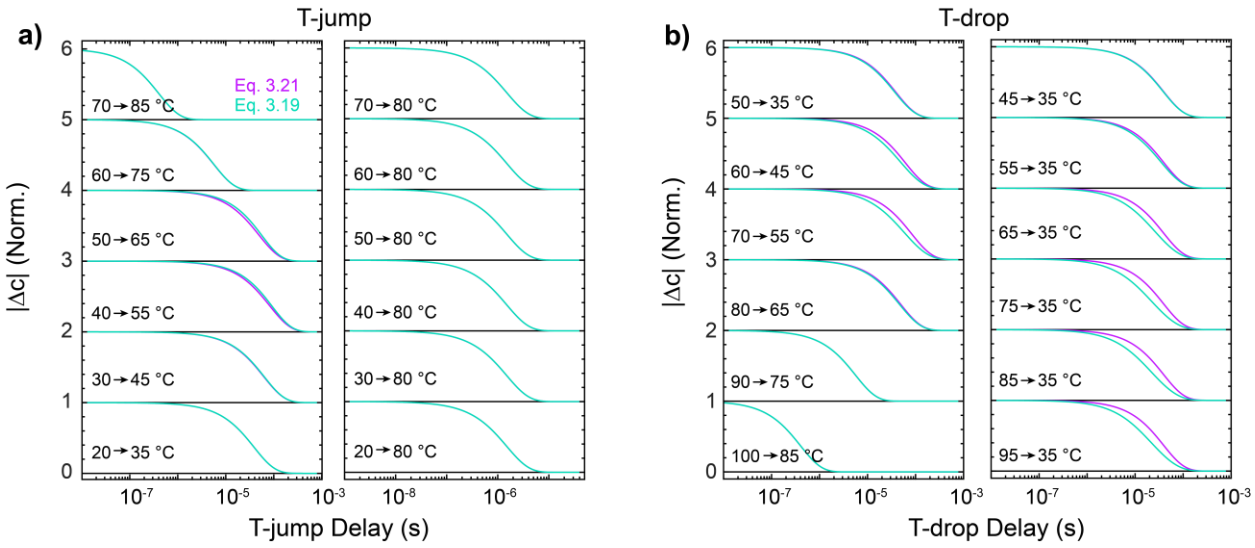

Figure 3.12 Perturbative and non-perturbative T-jump and T-drop profiles. (a) Time-dependent normalized absolute value of Δc following a (left) ΔT = 15 °C T-jump as a function of temperature. Perturbative (eq. 3.21, pink) and non-perturbative (eq. 3.19, cyan) profiles are shown. T-jumps to consistent T_f with variable ΔT are shown on the right. **(b)** (left) Δc during a T-drop with ΔT = -15 °C as a function of temperature. (right) T-drops to a consistent T_f with variable ΔT. Significant deviation between perturbative and non-perturbative time profiles is only observed for T-drops with T_f below T_m.

3.6.2 Extension to three-state kinetics

So far we only considered two-state kinetics for hybridization and dehybridization, but multi-state kinetics may arise from additional stable duplex configurations or competition with other nucleic acid conformations. Solutions for various kinetic schemes are documented,[75-76] and we focus on those most relevant/useful for DNA and protein binding. We start by considering a three-state scheme with an additional duplex state (D_I) that may proceed to the fully intact duplex with forward (k_f) and backward (k_b) rate constants.

$$S_1 + S_2 \underset{k_d}{\overset{k_h}{\rightleftharpoons}} D_I \underset{k_b}{\overset{k_f}{\rightleftharpoons}} D \tag{3.22}$$

As for the two-state case, we first write the rate expression for each species at equilibrium. We assume $[S_1] = [S_2]$ such that the concentration of each single strand is equivalent ($[S_{1/2}]$).

$$\frac{d[D]}{dt} = k_f [D_I] - k_b [D] = 0 \tag{3.23a}$$

$$\frac{d[S_1]}{dt} = \frac{d[S_2]}{dt} = \frac{d[S_{1/2}]}{dt} = k_d [D_I] - k_h [S_{1/2}]^2 = 0 \tag{3.23b}$$

$$\frac{d[D_I]}{dt} = -(k_f + k_d)[D_I] + k_b [D] + k_h [S_{1/2}]^2 = 0 \tag{3.23c}$$

$$c_{tot} = 2[S_{1/2}] + 2[D] + 2[D_I] \tag{3.23d}$$

Upon a change in concentration and assuming a perturbation (neglect $\Delta[S_{1/2}]^2$ terms),

$$\frac{d\Delta[D]}{dt} = k_f \Delta[D_I] - k_b \Delta[D] \tag{3.24a}$$

$$\frac{d\Delta[D_I]}{dt} = -k_f \Delta[D_I] + k_b \Delta[D] - k_d \Delta[D_I] + 2k_h [S_{1/2}]_{T_f} \Delta[S_{1/2}] \tag{3.24b}$$

$$\frac{d\Delta[S_{1/2}]}{dt} = k_d \Delta[D_I] - 2k_h [S_{1/2}]_{T_f} \Delta[S_{1/2}] \tag{3.24c}$$

Using mass conservation, $\Delta[D_I] = -\Delta[D] - \Delta[S_{1/2}]$, the kinetics can be described as a system of two differential equations.

$$\frac{d\Delta[D]}{dt} = -\left(k_f + k_b\right)\Delta[D] - k_f \Delta[S_{1/2}] \tag{3.25a}$$

$$\frac{d\Delta[S_{1/2}]}{dt} = -k_d\Delta[D] - \left(k_d + 2k_h[S_{1/2}]_{T_f}\right)\Delta[S_{1/2}] \tag{3.25b}$$

Systems of differential equations of this form are shown to have solutions that are the sum of two exponential decays.[75-76]

$$\Delta[D] = \Delta[D]_1 \exp\left(-\lambda_{obs,1}t\right) + \Delta[D]_2 \exp\left(-\lambda_{obs,2}t\right) \tag{3.26a}$$

$$\Delta[S_{1/2}] = \Delta[S_{1/2}]_1 \exp\left(-\lambda_{obs,1}t\right) + \Delta[S_{1/2}]_2 \exp\left(-\lambda_{obs,2}t\right) \tag{3.26b}$$

Here concentration changes with subscripts 1 and 2 correspond to changes during the 1st and 2nd exponential components, respectively, and their sum corresponds to the total relaxation ($\Delta[D]_0$, $\Delta[S_{1/2}]_0$). The goal now is to express each observed rate constant ($\lambda_{obs,1}$ & $\lambda_{obs,2}$) for which we assume two particular solutions to the system of differential equations.

$$\Delta[D] = \Delta[D]_0 \exp\left(-\lambda_{obs}t\right) \tag{3.27a}$$

$$\Delta[S_{1/2}] = \Delta[S_{1/2}]_0 \exp\left(-\lambda_{obs}t\right) \tag{3.27b}$$

Combining the derivatives of these solutions, $\frac{d\Delta[D]}{dt} = -\lambda_{obs}\Delta[D]$ and $\frac{d\Delta[S_{1/2}]}{dt} = -\lambda_{obs}\Delta[S_{1/2}]$, into eqs. 3.25 allows us to solve for the solutions of λ_{obs} using the corresponding determinant.

$$\begin{vmatrix} k_f + k_b - \lambda_{obs} & k_f \\ k_d & k_d + 2k_h[S_{1/2}]_{T_f} - \lambda_{obs} \end{vmatrix} = 0 \tag{3.28a}$$

with solutions,

$$\lambda_{obs,1} = \frac{1}{2}\Theta + \left[\left(\frac{1}{2}\Theta\right)^2 - \Phi\right]^{1/2} \tag{3.28b}$$

$$\lambda_{obs,2} = \frac{1}{2}\Theta - \left[\left(\frac{1}{2}\Theta\right)^2 - \Phi\right]^{1/2} \tag{3.28c}$$

where, $\Theta = k_f + k_b + k_d + 2k_h\left[S_{1/2}\right]_{T_f}$

$$\Phi = 2k_h k_f \left[S_{1/2}\right]_{T_f} + k_b\left(k_d + 2k_h\left[S_{1/2}\right]_{T_f}\right)$$

Eqs. 3.28 are valid for perturbative kinetics and any values of $\lambda_{obs,1}$ and $\lambda_{obs,2}$. Often in relaxation measurements, components will be significantly separated in time (such as duplex terminal fraying and full dissociation). Then the fast component will be uncoupled from the slow. For example, if $k_f \,\&\, k_b >> k_d \,\&\, 2k_h\left[S_{1/2}\right]_{T_f}$ then $\left(\frac{1}{2}\Theta\right)^2 >> \Phi$, leading to:

$$\lambda_{obs,1} \approx k_f + k_b \tag{3.29a}$$

$$\lambda_{obs,2} = \frac{\Phi}{\lambda_{obs,1}} = \frac{k_b k_d}{k_f + k_b} + 2k_h\left[S_{1/2}\right]_{T_f} \tag{3.29b}$$

It is common to observe T-jump responses with multiple kinetic components, but the data available is often insufficient to constrain a given multi-state kinetic model. Eq. 3.22 represents a sequential three-state process, but parallel and trap off-pathway kinetics are also possible. Each model requires knowledge of the temperature-dependent equilibrium concentration for the intermediate, which poses a significant practical barrier for this analysis. In practice, the slow observed rate may

be treated with a two-state model or additional assumptions must be made regarding the microscopic rate constants in and out of the intermediate state.

3.7 Acknowledgements

I thank John Hack and Melissa Bodine for helpful feedback on this chapter.

3.8 References

1. Son, I.; Shek, Y. L.; Dubins, D. N.; Chalikian, T. V., Hydration changes accompanying helix-to-coil DNA transitions. *J. Am. Chem. Soc.* **2014,** *136*, 4040-4047.

2. Hickey, D. R.; Turner, D. H., Solvent effects on the stability of A7U7p. *Biochem.* **1985,** *24*, 2086-2094.

3. Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A., Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* **2008,** *47*, 5336-5353.

4. Zimmer, C.; Luck, G.; Venner, H.; Frič, J., Studies on the conformation of protonated DNA. *Biopolymers* **1968,** *6*, 563-574.

5. Lubbe, A. S.; Liu, Q.; Smith, S. J.; De Vries, J. W.; Kistemaker, J. C.; De Vries, A. H.; Faustino, I.; Meng, Z.; Szymanski, W.; Herrmann, A., Photoswitching of DNA hybridization using a molecular motor. *J. Am. Chem. Soc.* **2018,** *140*, 5069-5076.

6. Dumont, C.; Emilsson, T.; Gruebele, M., Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat. Methods* **2009,** *6*, 515-519.

7. Donten, M. L.; Hassan, S.; Popp, A.; Halter, J.; Hauser, K.; Hamm, P., pH-jump induced leucine zipper folding beyond the diffusion limit. *J. Phys. Chem. B* **2015,** *119*, 1425-1432.

8. Nakasone, Y.; Ooi, H.; Kamiya, Y.; Asanuma, H.; Terazima, M., dynamics of inter-DNA chain interaction of photoresponsive DNA. *J. Am. Chem. Soc.* **2016,** *138*, 9001-9004.

9. Narayanan, R.; Zhu, L.; Velmurugu, Y.; Roca, J.; Kuznetsov, S. V.; Prehna, G.; Lapidus, L. J.; Ansari, A., Exploring the energy landscape of nucleic acid hairpins using laser temperature-jump and microfluidic mixing. *J. Am. Chem. Soc.* **2012,** *134*, 18952-18963.

10. Deeg, A. A.; Rampp, M. S.; Popp, A.; Pilles, B. M.; Schrader, T. E.; Moroder, L.; Hauser, K.; Zinth, W., Isomerization-and Temperature-Jump-Induced Dynamics of a Photoswitchable β-Hairpin. *Chem. Eur. J.* **2014,** *20*, 694-703.

11. Jose, D.; Porschke, D., The dynamics of the B− A transition of natural DNA double helices. *J. Am. Chem. Soc.* **2005,** *127*, 16120-16128.

12. Eigen, M., Methods for investigation of ionic reactions in aqueous solutions with half-times as short as 10–9 sec. Application to neutralization and hydrolysis reactions. *Discuss. Faraday Soc.* **1954,** *17*, 194-205.

13. Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A., Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* **2007,** *78*, 063101.

14. Thompson, M. C.; Barad, B. A.; Wolff, A. M.; Sun Cho, H.; Schotte, F.; Schwarz, D. M.; Anfinrud, P.; Fraser, J. S., Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme. *Nat. Chem.* **2019,** *11*, 1058-1066.

15. Wolff, A. M.; Nango, E.; Young, I. D.; Brewster, A. S.; Kubo, M.; Nomura, T.; Sugahara, M.; Owada, S.; Barad, B. A.; Ito, K., Mapping Protein Dynamics at High-Resolution with Temperature-Jump X-ray Crystallography. *bioRxiv* **2022**, 2022.06. 10.495662.

16. Ballew, R.; Sabelko, J.; Reiner, C.; Gruebele, M., A single-sweep, nanosecond time resolution laser temperature-jump apparatus. *Rev. Sci. Instrum.* **1996,** *67*, 3694-3699.

17. Marchand, A.; Czar, M. F.; Eggel, E. N.; Kaeslin, J.; Zenobi, R., Studying biomolecular folding and binding using temperature-jump mass spectrometry. *Nat. Commun.* **2020,** *11*, 566.

18. Khuc, M.-T.; Mendonça, L.; Sharma, S.; Solinas, X.; Volk, M.; Hache, F., Measurement of circular dichroism dynamics in a nanosecond temperature-jump experiment. *Rev. Sci. Instrum.* **2011,** *82*, 054302.

19. Rinnenthal, J.; Wagner, D.; Marquardsen, T.; Krahn, A.; Engelke, F.; Schwalbe, H., A temperature-jump NMR probe setup using rf heating optimized for the analysis of temperature-induced biomacromolecular kinetic processes. *J. Magn. Reson.* **2015,** *251*, 84-93.

20. Polinkovsky, M. E.; Gambin, Y.; Banerjee, P. R.; Erickstad, M. J.; Groisman, A.; Deniz, A. A., Ultrafast cooling reveals microsecond-scale biomolecular dynamics. *Nat. Commun.* **2014,** *5*, 1-7.

21. Lian, T.; Locke, B.; Kholodenko, Y.; Hochstrasser, R. M., Energy flow from solute to solvent probed by femtosecond IR spectroscopy: malachite green and heme protein solutions. *J. Phys. Chem.* **1994,** *98*, 11648-11656.

22. Nölting, B., Distribution of temperature in globular molecules, cells, or droplets in temperature-jump, sound velocity, and pulsed laser experiments. *J. Phys. Chem. B* **1998,** *102*, 7506-7509.

23. Citroni, M.; Fanetti, S.; Falsini, N.; Foggi, P.; Bini, R., Melting dynamics of ice in the mesoscopic regime. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 5935-5940.

24. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

25. Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B., Protein folding and unfolding on a complex energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97*, 2527-2532.

26. Ashwood, B.; Lewis, N. H.; Sanstead, P. J.; Tokmakoff, A., Temperature-Jump 2D IR Spectroscopy with Intensity-Modulated CW Optical Heating. *J. Phys. Chem. B* **2020,** *124*, 8665-8677.

27. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. *Proc. Natl. Acad. Sci. U.S.A.* **2023,** *120*, e2219124120.

28. Sarkar, K.; Meister, K.; Sethi, A.; Gruebele, M., Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe. *Biophys. J.* **2009,** *97*, 1418-1427.

29. Gruebele, M.; Sabelko, J.; Ballew, R.; Ervin, J., Laser temperature jump induced protein refolding. *Acc. Chem. Res.* **1998,** *31*, 699-707.

30. Dyer, R. B.; Maness, S. J.; Franzen, S.; Fesinmeyer, R. M.; Olsen, K. A.; Andersen, N. H., Hairpin folding dynamics: the cold-denatured state is predisposed for rapid refolding. *Biochem.* **2005,** *44*, 10406-10415.

31. Ma, H.; Ervin, J.; Gruebele, M., Single-sweep detection of relaxation kinetics by submicrosecond midinfrared spectroscopy. *Rev. Sci. Instrum.* **2004,** *75*, 486-491.

32. Todisco, M.; Szostak, J. W., Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Res.* **2022,** *50*, 9647-9662.

33. Phan, A. T.; Mergny, J. L., Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. *Nucleic Acids Res.* **2002,** *30*, 4618-4625.

34. Eigen, M.; Kustin, K., The kinetics of halogen hydrolysis. *J. Am. Chem. Soc.* **1962,** *84*, 1355-1361.

35. Pörschke, D.; Eigen, M., Co-operative non-enzymatic base recognition III. Kinetics of the helix—coil transition of the oligoribouridylic· oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **1971,** *62*, 361-381.

36. Pörschke, D.; Uhlenbeck, O. C.; Martin, F. H., Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs. *Biopolymers* **1973,** *12*, 1313-1335.

37. Wetmur, J. G.; Davidson, N., Kinetics of renaturation of DNA. *J. Mol. Biol.* **1968,** *31*, 349-370.

38. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.

39. Beitz, J. V.; Flynn, G. W.; Turner, D. H.; Sutin, N., Stimulated Raman effect. A new source of laser temperature-jump Heating. *J. Am. Chem. Soc.* **1970,** *92*, 4130-4132.

40. Turner, D. H.; Flynn, G. W.; Sutin, N.; Beitz, J. V., Laser Raman temperature-jump study of the kinetics of the triiodide equilibrium. Relaxation times in the 10-8-10-7 second range. *J. Am. Chem. Soc.* **1972,** *94*, 1554-1559.

41. Hoffmann, H.; Yeager, E.; Stuehr, J., Laser temperature-jump apparatus for relaxation studies in electrolytic solutions. *Rev. Sci. Instrum.* **1968,** *39*, 649-653.

42. Chen, S.; Lee, I. S.; Tolbert, W. A.; Wen, X.; Dlott, D. D., Applications of ultrafast temperature jump spectroscopy to condensed phase molecular dynamics. *J. Phys. Chem.* **1992,** *96*, 7178-7186.

43. Kubelka, J., Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem. Photobio. Sci.* **2009,** *8*, 499-512.

44. Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Minnes, L.; Towrie, M.; Parker, A. W.; Hunt, N. T., Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 6188-6199.

45. Vedamuthu, M.; Singh, S.; Robinson, G. W., Accurate mixture-model densities for D2O. *J. Phys. Chem.* **1994,** *98*, 8591-8593.

46. DuPont FEP Film Information Bulletin. DuPont, Ed. DuPont: 2010.

47. Mursalov, B.; Abdulagatov, I.; Dvoryanchikov, V.; Kamalov, A.; Kiselev, S., Isochoric heat capacity of heavy water at subcritical and supercritical conditions. *Int. J. Thermophys.* **1999,** *20*, 1497-1528.

48. Graebner, J., Measurements of specific heat and mass density in CVD diamond. *Diam. Relat. Mater.* **1996,** *5*, 1366-1370.

49. Matsunaga, N.; Nagashima, A., Transport properties of liquid and gaseous D2O over a wide range of temperature and pressure. *J. Phys. Chem. Ref. Data* **1983,** *12*, 933-966.

50. Ballard, S. S.; Combes, L. S.; McCarthy, K. A., A comparison of the physical properties of barium fluoride and calcium fluoride. *J. Opt. Soc. Am.* **1952,** *42*, 684_1-685.

51. Kashnow, R.; McCarthy, K., Phonon scattering in magnesium fluoride. *J. Phys. Chem. Solids* **1969,** *30*, 813-818.

52. Twitchen, D.; Pickles, C.; Coe, S.; Sussmann, R.; Hall, C., Thermal conductivity measurements on CVD diamond. *Diam. Relat. Mater.* **2001,** *10*, 731-735.

53. Hofrichter, J., Laser temperature-jump methods for studying folding dynamics. *Proteins* **2001**, 159-191.

54. Wray, W.; Aida, T.; Dyer, R., Photoacoustic cavitation and heat transfer effects in the laser-induced temperature jump in water. *Appl. Phys. B* **2002,** *74*, 57-66.

55. Crank, J.; Nicolson, P. In *A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type*, Mathematical proceedings of the Cambridge philosophical society, Cambridge University Press: 1947; pp 50-67.

56. Hammer, D. X.; Thomas, R. J.; Noojin, G. D.; Rockwell, B. A.; Kennedy, P. K.; Roach, W. P., Experimental investigation of ultrashort pulse laser-induced breakdown thresholds in aqueous media. *IEEE J. Quantum Elect.* **1996,** *32*, 670-678.

57. Noack, J.; Vogel, A., Laser-induced plasma formation in water at nanosecond to femtosecond time scales: calculation of thresholds, absorption coefficients, and energy density. *IEEE J. Quantum Elect.* **1999,** *35*, 1156-1167.

58. Popp, A.; Scheerer, D.; Heck, B.; Hauser, K., Biomolecular dynamics studied with IR-spectroscopy using quantum cascade lasers combined with nanosecond perturbation techniques. *Spectrochim. Acta A* **2017,** *181*, 192-199.

59. Kiyohara, O.; Halpin, C. J.; Benson, G. C., Ultrasonic velocities for deuterium oxide–water mixtures at 298.15 K. *Can. J. Chem.* **1977,** *55*, 3544-3548.

60. Millero, F. J.; Lepple, F. K., Isothermal compressibility of deuterium oxide at various temperatures. *J. Chem. Phys.* **1971,** *54*, 946-949.

61. Holmstrom, E. D.; Nesbitt, D. J., Real-time infrared overtone laser control of temperature in picoliter H2O samples:"nanobathtubs" for single molecule microscopy. *J. Phys. Chem. Lett.* **2010,** *1*, 2264-2268.

62. De Lorenzo, S.; Ribezzi-Crivellari, M.; Arias-Gonzalez, J. R.; Smith, S. B.; Ritort, F., A temperature-jump optical trap for single-molecule manipulation. *Biophys. J.* **2015,** *108*, 2854-2864.

63. Ludwig, C.; Staatsdruckerei, K. H., *Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösung*. Aus der KK Hof-und Staatsdruckerei, in Commission bei W. Braumüller …: 1856.

64. Duhr, S.; Braun, D., Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 19678-19682.

65. Reichl, M.; Herzog, M.; Götz, A.; Braun, D., Why charged molecules move across a temperature gradient: the role of electric fields. *Phys. Rev. Lett.* **2014,** *112*, 198101.

66. Wienken, C. J.; Baaske, P.; Duhr, S.; Braun, D., Thermophoretic melting curves quantify the conformation and stability of RNA and DNA. *Nucleic Acids Res.* **2011,** *39*, e52-e52.

67. Asmari, M.; Ratih, R.; Alhazmi, H. A.; El Deeb, S., Thermophoresis for characterizing biomolecular interaction. *Methods* **2018,** *146*, 107-119.

68. Chen, Z.; Li, J.; Zheng, Y., Heat-mediated optical manipulation. *Chem. Rev.* **2021,** *122*, 3122-3179.

69. Kuznetsov, S. V.; Sugimura, S.; Vivas, P.; Crothers, D. M.; Ansari, A., Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 18515-18520.

70. Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J., Smaller and faster: The 20-residue Trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.* **2002,** *124*, 12952-12953.

71. Ebbinghaus, S.; Dhar, A.; McDonald, J. D.; Gruebele, M., Protein folding stability and dynamics imaged in a living cell. *Nat. Methods* **2010,** *7*, 319-323.

72. Zondervan, R.; Kulzer, F.; van der Meer, H.; Disselhorst, J. A.; Orrit, M., Laser-driven microsecond temperature cycles analyzed by fluorescence polarization microscopy. *Biophys. J.* **2006,** *90*, 2958-2969.

73. Holmstrom, E. D.; Dupuis, N. F.; Nesbitt, D. J., Pulsed IR heating studies of single-molecule DNA duplex dissociation kinetics and thermodynamics. *Biophys. J.* **2014,** *106*, 220-231.

74. Baiz, C. R.; Schach, D.; Tokmakoff, A., Ultrafast 2D IR microscopy. *Opt. Express* **2014,** *22*, 18724-18735.

75. Bernasconi, C., *Relaxation kinetics*. Academic Press: 1976.

76. Nölting, B., *Protein folding kinetics: biophysical methods*. Springer 2005.

77. Carrillo-Nava, E.; Mejía-Radillo, Y.; Hinz, H.-J. r., Dodecamer DNA duplex formation is characterized by second-order kinetics, positive activation energies, and a dependence on sequence and Mg2+ ion concentration. *Biochem.* **2008,** *47*, 13153-13157.

Chapter 4

Temperature-jump IR spectroscopy in practice

Material for part of this chapter is adapted from:

Ashwood, B.; Lewis, N.H.C.; Sanstead, P.J.; Tokmakoff, A., Temperature-Jump 2D IR Spectroscopy with Intensity-Modulated CW Optical Heating. *J. Phys. Chem. B* **2020** 124, 8665-8677.

4.1 Pulsed T-jump spectrometer

T-jump spectroscopy with an infrared absorption probe (T-jump IR) was first reported in the early 1980s[1] and later adopted by multiple research groups in the 1990s and 2000s for investigation of fast protein folding kinetics.[2-4] Incorporation of T-jump spectroscopy with 2D IR spectroscopy was first demonstrated in the Tokmakoff group[5-6] and has been a workhorse experiment for the study of protein and nucleic acid structural dynamics for nearly twenty years. As we approach the firing of the 2 billionth flashlamp shot, it is worth noting that most aspects of this experiment have remained the same over the years and are well-documented in previous group members' theses.[7-9] With that said, I, like those before me, have spent too much time working with this spectrometer to avoid giving it some discussion.

The pulsed T-jump 2D IR experiment involves synchronizing a conventional ultrafast 2D IR spectrometer with a Q-switched Nd:YAG laser for optical heating. The 2D IR spectrometer can be broken down into three key areas: (1) mid-IR pulse generation, (2) boxcar interferometer to create precisely spatially and temporally separated pulses, and (3) signal generation and detection.

Timing between the T-jump excitation and the mid-IR probe is achieved through electronic synchronization of the Nd:YAG laser and the oscillator that seeds the regenerative amplifier for generating ultrafast pulses.

4.1.1 Mid-IR pulse generation and boxcar interferometer

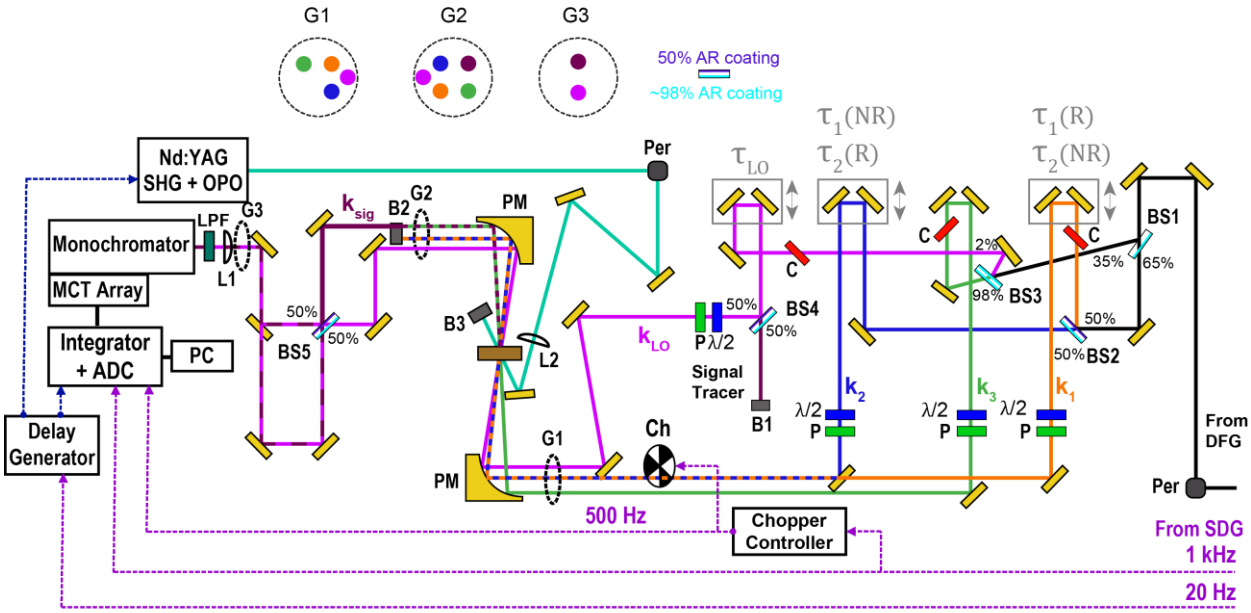

Figure 4.1 Optical layout and electronics for the pulsed T-jump spectrometer. The beam path begins at the output of OPA+DFG (right side, black) and is split into four beams in a boxcar interferometer that are overlapped at the sample: k_1 (orange), k_2 (blue), k_3 (green), k_{LO} (magenta). The signal (k_{sig}, maroon) is heterodyned with a local oscillator (k_{LO}) and detected on a MCT array using a balanced detection scheme. ZnSe beamsplitter (BS) surfaces are color-coded based on their antireflection (AR) coating in the mid-IR: purple – 50%, cyan – 98% transmission. The T-jump laser beam is overlapped with the mid-IR beams at the sample and is electronically synchronized with the mid-IR laser source. Other abbreviations: C – ZnSe compensation, λ/2 – half waveplate, P – polarizer, Per – periscope, L – convex lens, PM – 90° off-axis parabolic mirror, B – beam block, LPF – low-pass filter, Ch – optical chopper, ADC – analog-to-digital converter, PC – local computer.

The details of the spectrometer and ultrafast mid-IR pulse generation have been described previously.[9] In brief, a Ti:Sapphire regenerative amplifier (Libra-HE, Coherent) generates pulses centered at 800 nm with a 90 fs duration, 3-4 mJ pulse energy, and a repetition rate of 1 kHz. Approximately 45% of this energy is used to pump an optical parametric amplifier (OPA, TOPAS-C, Light Conversion) for the pulsed T-jump spectrometer while the remaining portion is used to pump an OPA for the CW T-jump 2D IR spectrometer (Section 4.3). The OPA utilizes collinear Type II phase-matching such that the generated near-IR (NIR) signal and idler are collinear and have orthogonal polarization. The signal and idler are used for difference frequency generation (DFG) to produce mid-IR pulses for 2D IR spectroscopy. Temporal overlap between the signal and idler is controlled using a Michelson interferometer, and they are overlapped in a $AgGaS_2$ crystal to produce mid-IR pulses with 8-12 μJ of pulse energy via DFG. The NIR idler generated during DFG is filtered by a germanium window with an anti-reflection (AR) coating.

Prior to entering the boxcar interferometer, the mid-IR light polarization is rotated from p to s using a 90° periscope. A series of AR-coated ZnSe beamsplitters (BS1-BS4) split the mid-IR input into five separate beams: three for third-order IR signal generation (k_1, k_2, k_3), one for the local oscillator (k_{LO}), and one for alignment (k_4, blocked by B1 during data acquisition). The beamsplitters are arranged and coated such that each of k_1, k_2, and k_3 contain 32-33% of the input pulse energy (not accounting for reflective losses) and just 1-2% is saved for k_{LO}. All beams are aligned into their own retroreflector, which are set to vertically offset k_1 and k_3 above k_2 and k_{LO}. The retroreflectors for k_2, k_1, and k_{LO} are mounted to motorized stages (ANT-90, Aerotech) to adjust the relative timing of each mid-IR pulse at the sample. The polarization of each beam is

independently controlled by a half-waveplate and polarizer. The positions of k_1, k_2, and k_3 are set to three corners of a 1.15 by 1.15 in. square (G1 in Fig. 4.1) prior to being focused to a ~100 μm $1/e^2$ diameter spot at the sample by a 90° off-axis parabolic mirror (10 cm effective focal length). k_{LO} spatially overlaps with the other beams at the sample, but arrives ~100 ps prior to k_1, k_2, and k_3 to avoid interfering in the signal generation process. Passing k_{LO} through the same sample region as the other beams ensures that its spectral and phase characteristics are nearly the same as the signal. Prior to the sample, k_{LO} is set at a height between k_1/k_3 and k_2 and horizontally offset for greater spatial separation with the other beams after the sample such that it can follow a longer path to achieve temporal overlap with the signal in the detection area.

4.1.2 Signal detection

Due to the non-collinear beam geometry of k_1, k_2, and k_3, the third-order signal (k_{sig}) generated from the sample is wavevector-matched in the direction of the fourth corner of the square (G2). After re-collimation by a second off-axis parabolic mirror, the k_1, k_2, and k_3 beams are blocked and k_{LO} and k_{sig} are aligned to spatially and temporally overlap at a beamsplitter (BS5). We employ a balanced-detection scheme where in-phase and π-shifted interference between k_{LO} and k_{sig} are detected simultaneously and subtracted on-the-fly during data acquisition. BS5 is AR-coated so that 50% of k_{sig} is transmitted while 50% reflects off the front-face and accumulates a π-phase shift. This phase-shift occurs because ZnSe has a higher refractive index (n = 2.4) than air (n = 1). In contrast, the reflection of k_{LO} off the back-face of BS5 does not experience a phase shift. The back face is AR-coated for 98% transmission, leading to transmission of k_{LO} through the back-face followed by 50% reflection off the front face. The resulting beams after BS5

(reflected k_{sig} + transmitted k_{LO} & transmitted k_{sig} + reflected k_{LO}) are vertically offset (G3, Fig. 4.1), focused onto and dispersed by a grating monochromator, and detected on separate 64-pixel stripes of HgCdTe (MCT, IR Systems Development). The voltage detected at each pixel is read by a boxcar integrator, and the output signals of each stripe are subtracted for balanced detection on the fly with data acquisition software on a local computer.

Although balanced detection further complicates alignment of the 2D IR spectrometer, it was reported to improve signal-to-noise by 16-fold and enables additional corrections for pulse distortions from linear absorption.[5, 10] The signal-to-noise enhancement largely comes from the ability to perform differential detection with a single laser shot. Each MCT pixel detects the total intensity from k_{sig} and k_{LO} at a given detection frequency (ω_3), which contains homodyne contributions from each field (E_{sig}, E_{LO}) as well as the cross-term that depends on the phase difference between signal and LO ($\Delta\phi$).

$$I_{str1}(\omega_3) = \left| \frac{E_{sig}(\omega_3) + E_{LO}(\omega_3)}{2} \right|^2 = \left[\frac{E_{sig}(\omega_3)}{2} \right]^2 + \left[\frac{E_{LO}(\omega_3)}{2} \right]^2 + E_{sig}(\omega_3) E_{LO}(\omega_3) \cos\left[\Delta\phi(\omega_3) \right] \quad (4.1a)$$

$$I_{str2}(\omega_3) = \left| \frac{E_{sig}(\omega_3) \exp(i\pi) + E_{LO}(\omega_3)}{2} \right|^2 = \left[\frac{E_{sig}(\omega_3)}{2} \right]^2 + \left[\frac{E_{LO}(\omega_3)}{2} \right]^2 - E_{sig}(\omega_3) E_{LO}(\omega_3) \cos\left[\Delta\phi(\omega_3) \right] \quad (4.1b)$$

The π phase difference between the signal on each stripe creates a π-phase shift between the respective cross-terms, giving them opposite signs. Subtraction between the intensity detected on each stripe removes the homodyne terms.

$$I_{str1}(\omega_3) - I_{str2}(\omega_3) = 2E_{sig}(\omega_3) E_{LO}(\omega_3) \cos\left[\Delta\phi(\omega_3) \right] \quad (4.2)$$

Unlike the homodyne terms, the interference signal is linearly dependent on E_{sig} and sample concentration. The time delay between k_{sig} and k_{LO} overlap at BS5 (τ_{LO}) can be adjusted to perform spectral interferometry to extract the amplitude and phase of the signal.[11-12] The resulting complex signal is termed the heterodyned dispersed vibrational echo (HDVE).[11]

$$S_{HDVE}(\omega_3, \tau_{LO}) = 2E_{sig}(\omega_3) E_{LO}(\omega_3) \cos\left[\Delta\phi(\omega_3) + \tau_{LO}\omega_3\right] \tag{4.3}$$

For the work in this thesis, we report the real part of the HDVE signal corrected to τ_{LO} = 0, termed the dispersed pump-probe (DPP).

$$S_{DPP}(\omega_3, \tau_{LO} = 0) = 2\mathbb{R}\left[E_{sig}(\omega_3) E_{LO}(\omega_3) \cos\left[\Delta\phi(\omega_3)\right]\right] \tag{4.4}$$

Although unnecessary in principle, we chop k_2 at 500 Hz (Fig. 4.1) to correct for mismatch in the alignment of each stripe and to remove certain scatter contributions that reach the detector. The actual signal we record on the fly is more complicated than eq. 4.2, and accounts for linear absorption distortions along ω_3 (but not ω_1).[10]

$$S_{obs}(\omega_3) = \frac{I_{str1}^{open} - I_{str2}^{open}}{I_{str1}^{open} + I_{str2}^{open}} - \frac{I_{str1}^{closed} - I_{str2}^{closed}}{I_{str1}^{closed} + I_{str2}^{closed}} \tag{4.5}$$

In eq. 4.5, I_{str1}^{open} and I_{str2}^{open} are the detected intensity on each stripe when k_2 is unchopped, and I_{str1}^{closed} and I_{str2}^{closed} are those with k_2 blocked and, as consequently, no signal is generated.

4.1.3 Incorporation of T-jump

A Q-switched Nd:YAG laser (YG980c, Quantel) is electronically synchronized with the 2D IR spectrometer to perform T-jump spectroscopy. The amplified fundamental (1064 nm, 20

Hz) is frequency-doubled in potassium dihydrogen phosphate (KDP) and pumps a BBO-based optical parametric oscillator (OPO, Opotek) to produce 4 – 10 mJ, ~7 ns pulses of 1.98 μm idler. The signal (727 nm) is not used in the experiment. After the collimation, the beam travels ~2.75 m before being focused and overlapped with the mid-IR probe volume at the sample. The spatial mode of the beam significantly deviates from a Gaussian profile, and a ~7.5 cm focal length is needed to focus to a beam diameter of 400 – 500 μm. Figure 4.2 shows that the spatial mode is elliptical and asymmetric with approximate dimensions of 530 × 380 μm. Measurements of power through different size pinholes estimates a $1/e^2$ diameter of 420 μm, assuming a Gaussian mode.

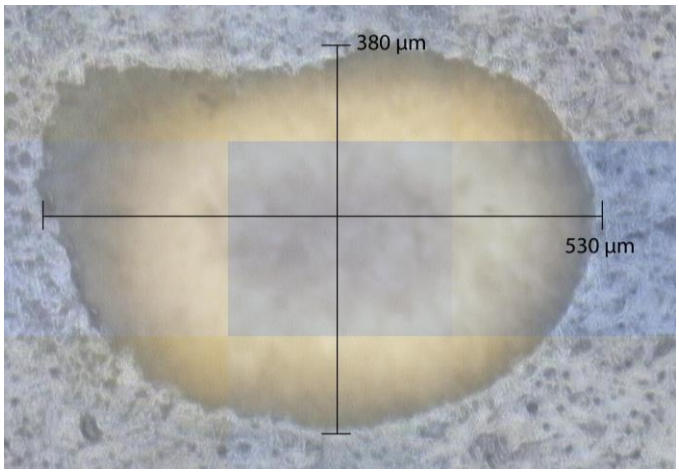

Figure 4.2 Image of T-jump beam focus from burn paper. Microscope image of burn spot from the pulsed T-jump beam at the sample position. Single-shot burns were made as a function of focal position, and the image shows the burn with the smallest area.

The oscillator in the regenerative amplifier serves as the master clock for electronic synchronization between the 2D IR probe and T-jump event. The 80 MHz repetition rate of the oscillator is divided to a 20 Hz signal used to trigger firing of the flashlamps and Q-switching in the T-jump laser.[13] The 20 Hz signal is programmatically delayed from hundreds of picoseconds

to one millisecond using a delay generator (DG535, Stanford Research Systems). Due to the mismatch in repetition rate between the T-jump (20 Hz) and mid-IR probe (1 kHz), 50 mid-IR shots probe the sample in 1 ms intervals between each T-jump. The sub-ms delay from the delay generator sets the time difference between the T-jump and the first mid-IR shot (τ_{TJ}^{1}). In most cases, the 50 ms interval between T-jumps allows for full thermal relaxation of the sample back to its initial condition before the next T-jump pulse arrives.

T-jumps may be used to trigger downhill structural dynamics in proteins and nucleic acids if large and fast enough temperature changes are provided.[14-15] The magnitude of the temperature change (ΔT) is limited by optical breakdown at laser power densities near 1 GW/cm^2 and potentially by cavitation at lower values. Our setup can produce ΔT of 25 °C using a power density of ~20 MW/cm^2, which in theory leaves plenty of room for larger ΔT T-jumps. Our current limitations in ΔT are set by the poor focusing properties of the T-jump beam and the total pulse energy available. One mechanism to achieve larger ΔT is by adding a second pass of the T-jump through the sample. Alternatively, the T-jump source may be upgraded to newly available Q-switched Nd:YAG or Ho:YAG lasers with higher pulse energy and spatial mode quality.[16]

4.2 Pulsed T-jump 2D IR data acquisition

In the current setup, pulsed T-jump experiments may be robustly performed with either HDVE or 2D IR as the spectroscopic probe. Either spectrum may be measured at many T-jump delays by electronically stepping τ_{TJ}^{1}. HDVE acquisition is much faster than 2D IR and therefore more commonly used for finely sampling kinetics.[14, 17-18] As noted in Section 4.1.2, signal acquisition requires differential detection with optical chopping. The chopper subtraction in

steady-state HDVE or 2D IR measurements is performed between adjacent shots with k_2 unchopped and chopped to minimize influence from slow fluctuations in mid-IR pulse energy. However, shot-to-shot subtraction is unsuitable in the T-jump measurement because consecutive shots occur at different points along the temperature and sample response profiles. Instead, data are acquired with two chopper phases spaced by 180°. All odd (τ_{TJ}^1, τ_{TJ}^1+2 ms, …)/even (τ_{TJ}^1+1 ms, τ_{TJ}^1+3 ms, …) shots are chopped/unchopped in one phase and unchopped/chopped in the other. Subtraction is then performed between different chopper phases at each T-jump delay (τ_{TJ}).

$$S_{obs}\left(\tau_{TJ},\omega_3\right)=\left\langle\frac{I_{str1}^{open}-I_{str2}^{open}}{I_{str1}^{open}+I_{str2}^{open}}\right\rangle_{\tau_{TJ},Phase1}-\left\langle\frac{I_{str1}^{closed}-I_{str2}^{closed}}{I_{str1}^{closed}+I_{str2}^{closed}}\right\rangle_{\tau_{TJ},Phase2} \tag{4.6}$$

The brackets in eq. 4.6 denote an average at each chopper phase. For the data in this thesis, 5000 mid-IR shots (or 100 T-jump events) were acquired, sorted by τ_{TJ} and whether k_2 is unchopped or chopped, and then averaged within each chopper phase.

As described previously,[11, 13] data is usually acquired at multiple τ_{LO} values and processed as follows: (1) Fourier transform spectral interferometry (FTSI) to extract real and imaginary components of the HDVE spectrum, (2) phase-adjusted to τ_{LO} = 0, (3) phase-corrected by fitting to a separately collected two-beam pump-probe signal between k_2 and k_{LO}, and (4) converted to DPP or the dispersed vibrational echo (DVE) from the HDVE.

$$S_{DPP}\left(\omega_3,\tau_{LO}=0\right)=\mathbb{R}\left[S_{HDVE}\left(\omega_3,\tau_{LO}=0\right)\right] \tag{4.7a}$$

$$S_{DVE}\left(\omega_3,\tau_{LO}=0\right)=\left|S_{HDVE}\left(\omega_3,\tau_{LO}=0\right)\right|^2 \tag{4.7b}$$

The DVE signal scales quadratically with sample concentration,[11] which makes it harder to interpret and less useful than the DPP signal for studying biomolecular reactions.

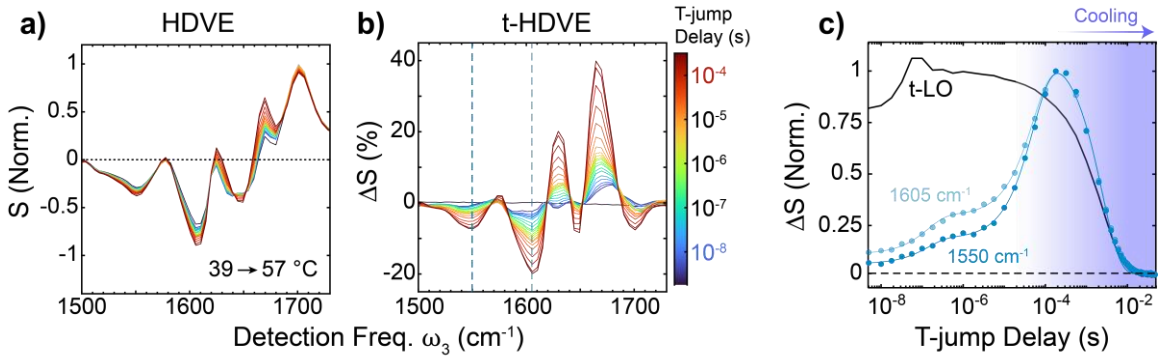

Figure 4.3 Example of pulsed t-HDVE data. (a) HDVE (DPP) spectra (eq. 4.5) of 5′-CGCATATATAT-3′ + complement at various τ_{TJ} delays following a T-jump along the duplex melting transition. Data were acquired with fixed τ_1 = 150 fs and τ_2 = 150 fs. The first spectrum is measured 5 ns prior to the T-jump (τ_{TJ} = -5 ns). **(b)** Difference relative to the initial temperature spectrum, $\Delta S(t) = [S(t) - S(T_i)]/S(T_i)$. **(c)** Normalized t-HDVE time traces at 1550 and 1605 cm^{-1} showing terminal fraying and duplex dissociation components. The solid black line shows the time-dependence of the change in transmission of the LO beam (t-LO).

Figure 4.3 shows an example of T-jump HDVE (t-HDVE) data for DNA duplex dissociation. The DPP signal can be interpreted similarly to the pump-probe spectrum introduced in Chapter 2.2. T-jump data are typically reported as a difference relative to the initial temperature (T_i) to emphasize spectral changes due to the T-jump. There is commonly an offset signal immediately after the T-jump (τ_{TJ} = 5 ns) resulting from sub-ns dynamics, which likely arise from changes in solvation or base-pairing structure. The spectral changes that follow at greater τ_{TJ} then report on terminal fraying (10 – 500 ns) and strand dissociation (10 – 500 μs) before thermal relaxation and re-hybridization occurs. The slowest measurable time constant is ~400 μs for a standard sample with a 50 μm path length and bare CaF_2 windows. Pressure relaxation is observed

from 10 to 100 ns (discussed in Chapter 3.3) and most clearly appears as an increase in transmission of the LO beam (t-LO).

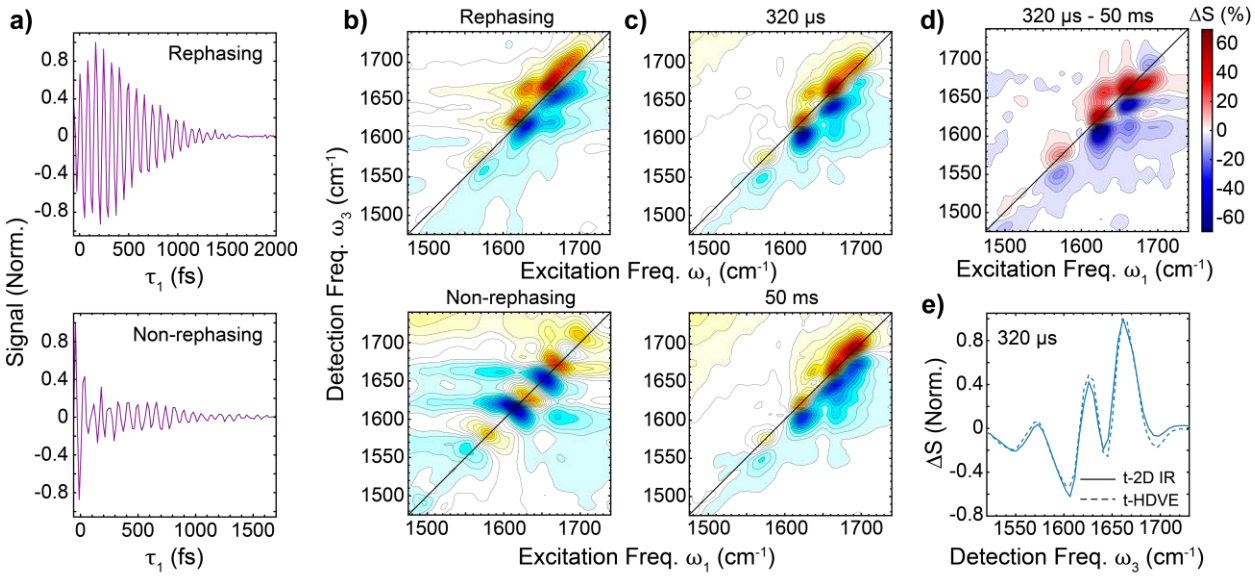

Figure 4.4 Acquisition of pulsed t-2D IR data. (a) Rephasing and non-rephasing FIDs at a τ_{TJ} value of 320 μs and detected at 1670 cm^{-1} for a DNA duplex of 5′-CGCATATATAT-3′ + complement. **(b)** Rephasing and non-rephasing 2D IR spectra. **(c)** Absorptive spectra at 320 μs and 50 ms. **(d)** Difference spectrum between 320 μs and 50 ms highlighting spectral changes from duplex dehybridization. **(e)** Projection of difference spectrum in **(d)** onto the ω_3 axis compared with the t-HDVE spectrum at the same τ_{TJ} value.

T-jump 2D IR (t-2D IR) data are collected by scanning the time delay between k_1 and k_2 (τ_1) and performing a Fourier transform on the resulting free induction decay (FID) to generate the excitation frequency axis (ω_1). In a boxcar beam geometry, rephasing (R) and non-rephasing (NR) signals are wavevector-matched in distinct directions such that only one is detected at a time. However, switching the time ordering of k_1 and k_2 flips the wavevector-matched directions of the R and NR signals. Our detection alignment is set up so that the R signal is measured when k_1 arrives before k_2 and the NR signal is measured when k_2 comes before k_1. In a typical t-2D IR

experiment on nucleic acids, the R FID is first measured by stepping τ_1 from -60 to 2000 fs in 16 fs steps (Fig. 4.4). At each τ_1 step, the signal is acquired identically to t-HDVE with τ_{LO} = 0. The NR FID decays more rapidly and is measured from τ_1 = -60 to 1700 fs. The R and NR FIDs are fully acquired at a single τ_{TJ}^1 before moving to the next. R and NR spectra are produced from Fourier transforming along τ_1 and phase-correcting by fitting the 2D IR projection onto ω_3 to the k_2/k_{LO} pump-probe spectrum. The R and NR spectra are then summed to generate the absorptive t-2D IR spectrum (Fig. 4.4). We report the difference spectrum relative to the maximum of the T_i spectrum as for t-HDVE data. Scanning τ_1 makes acquisition of a t-2D IR spectrum >100-fold slower than t-HDVE, therefore t-2D IR is typically only measured at a few τ_{TJ}^1 values for a given sample.[6, 18-19]

4.3 CW T-jump spectrometer

As shown in Fig. 4.3, the time window for pulsed T-jump IR experiments is limited to a few milliseconds due to temperature relaxation through the sample windows, and many proteins and nucleic acids exhibit large-scale conformational changes on millisecond to second timescales.[20-22] While numerous techniques can probe kinetics occurring on milliseconds or longer timescales, it is ideal for our group to have a single spectroscopic probe that can be used to study processes occurring from nanoseconds to hours. This section contains the description of a T-jump 2D IR spectrometer that utilizes continuous wave (CW) NIR optical heating to heat a sample within hundreds of microseconds and hold the final temperature for arbitrarily long times. Modulation of the CW laser power is synchronized with the 2D IR spectrometer to speed up heating by more than 10-fold and enable robust tunability of the T-jump repetition rate.

4.3.1 2D IR spectrometer

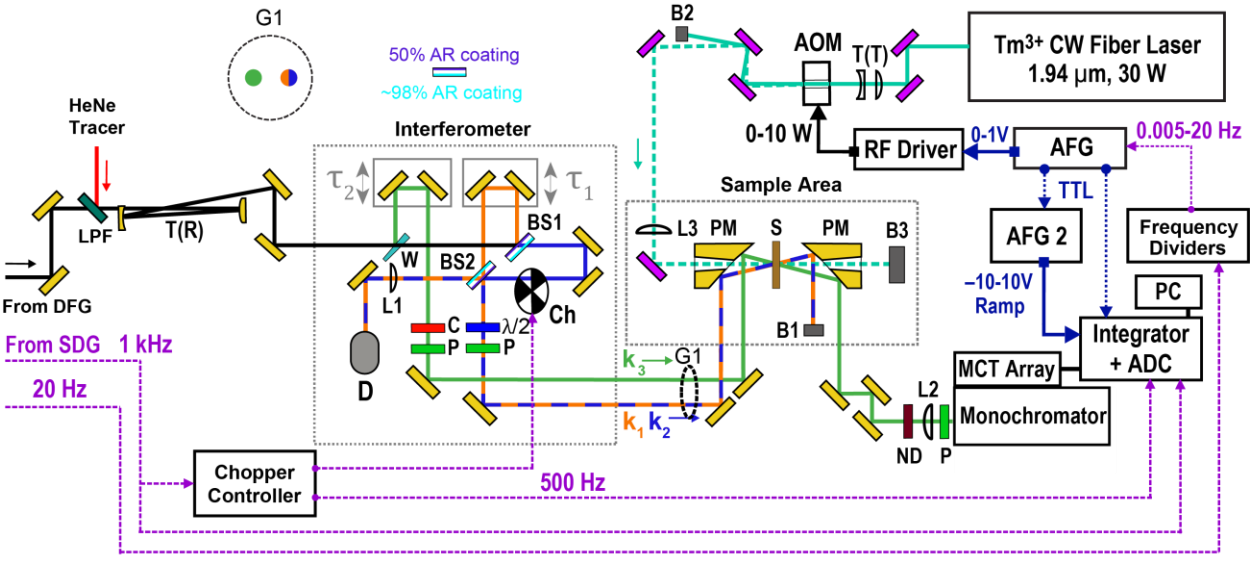

Figure 4.5 Optical layout and electronics for CW T-jump spectrometer. The beam path begins at the output of OPA+DFG (left side, black). The probe beam (k_3, green) is split off by a wedge (W) and the collinear pump pulse pair of k_1 (orange) and k_2 (blue) is generated by a Mach-Zehnder interferometer. The pump and probe beams are overlapped at the sample (S), and the generated signal is detected along with k_3 on an MCT array. Intensity modulation of a CW fiber laser by an acousto-optic modulator (AOM) is electronically synchronized with the mid-IR source and overlapped at the sample for optical heating. Other abbreviations: BS – AR-coated beamsplitter, LP – low-pass filter, T(R) – reflective telescope, T(T) – transmission telescope, L – convex lens, C – CaF_2 compensation material, λ/2 – half waveplate, P – polarizer, D – single-pixel MCT detector, Ch – optical chopper, PM – 90° off-axis parabolic mirror, B – beam block, LPF – low-pass filter, ND – neutral density, AFG – arbitrary waveform generator, ADC – analog-to-digital converter, PC – local computer.

The CW T-jump setup uses a 2D IR spectrometer employing a pump-probe pulse geometry. General aspects of 2D IR in the pump-probe geometry have been discussed previously,[13, 23-24] and this section focuses on elements specific to our setup. The remaining 55% power from the regenerative amplifier not used for the pulsed T-jump spectrometer pumps an OPA+DFG to generate 12-16 µJ of mid-IR pulse energy with s polarization. The NIR idler is

filtered from the mid-IR with AR-coated Ge, which also serves to overlap the mid-IR with a HeNe tracer beam for alignment. The mid-IR beam is collimated with a reflective telescope of spherical Au mirrors. A CaF_2 wedge (0.5°, Altechna) reflects about 5% of the mid-IR to be used as the probe beam (k_3). The transmitted portion enters a Mach-Zehnder interferometer that uses a pair of 50:50 AR-coated CaF_2 beam splitters (Edmund Optics) to split the beam into two arms (k_1, k_2), create a time delay between them (τ_1), and recombine them into a collinear beam. The beam splitters are AR-coated and oriented such that in-phase and π-phase shifted portions are generated. The reflected portion off of BS1 (k_1) experiences a π-phase due to the higher refractive index of CaF_2 (n = 1.4) relative to air (n = 1). The same is true for the reflected portion of k_2 off BS2. In contrast, the reflected portion of k_1 off BS2 does not experience a phase-shift. As a result, beams with k_1 and k_2 in-phase and out-of-phase are generated after BS1. The out-of-phase portion is unsuitable for 2D IR spectroscopy, but is used to measure a pump-pulse interferogram on a single-channel MCT detector as τ_1 is stepped. The pump interferogram is then Fourier transformed and used to correct for timing errors in τ_1 via the Mertz method.[13, 25] The polarization of the in-phase portion of k_1/k_2 is set with a half waveplate and polarizer while that of k_3 is fixed. The pump and probe beams are horizontally offset by 1 in. and focused to a 70-75 μm $1/e^2$ diameter spot (Fig. 4.6c,d) at the sample position using a 90° off-axis parabolic mirror (2 in. effective focal length). The pump and probe pulse duration was determined to be 90-100 fs at the sample by interferometric autocorrelation and intensity cross-correlation measurements, respectively (Fig. 4.6a,b). The pulse duration was minimized by adding additional Ge before the telescope and a 5 mm CaF_2 window into the k_3 path.

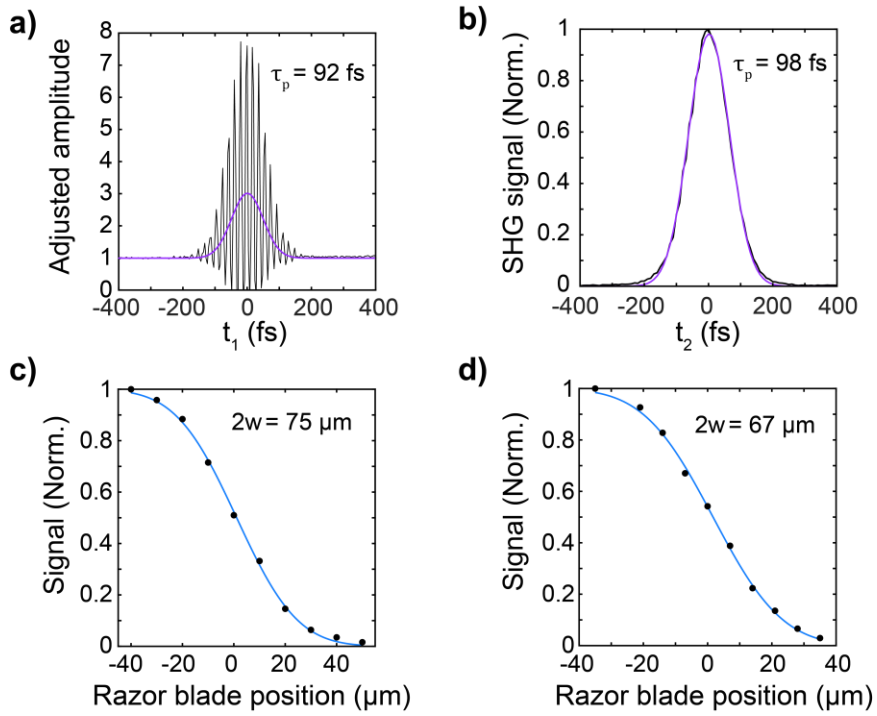

Figure 4.6 Characterization of mid-IR light at the sample. (a) Interferometric autocorrelation (black) in $AgGaS_2$ between k_1 and k_2 at the sample position. The zero-frequency component (intensity autocorrelation) was extracted by Fourier filtering and fitting to a Gaussian function (purple line). The data is shifted and scaled so that the long-time baseline is one and the lowest fringes drop to zero. **(b)** Background-free intensity cross-correlation between k_1/k_2 and k_3 (black) and fit to a Gaussian function (purple). Each measurement was performed with 1 mm CaF_2 added to the optical path to account for transmission into the sample. The full-width at half maximum (FWHM) of the pulse (τ_p) corresponds to the intensity autocorrelation FHWM divided by $\sqrt{2}$. **(c)** Normalized detected signal when scanning a razor blade horizontally across the k_3 at the sample focus. **(d)** Vertical scan across k_3. Scans are fit (blue line) to the integral of Gaussian function $S(x)=\frac{1}{2}\left(1-erf\left[\sqrt{2}\left(x-x_0\right)/w\right]\right)$ to extract the $1/e^2$ diameter (2w).

Unlike for the boxcar geometry described in Section 4.1, the third-order IR signal generated at the sample in a pump-probe geometry is wavevector-matched in the same direction as the probe beam. This leads to self-heterodyning of the signal by the probe and no additional LO is needed. After the sample, the beams are re-collimated with a second off-axis parabolic mirror, the pump is

blocked, and the signal/probe beam is focused onto a grating monochromator and detected on a 64-pixel MCT array. A polarizer set parallel to the polarization of the probe is placed just before the monochromator. The delay between pump pulses (τ_1) and the delay between k_3 and the pump (τ_2) are set by motorized stages (ANT95L, Aerotech) in the optical path of k_1 and k_3, respectively. The sample stage contains motorized actuators that control the sample position in the focus direction (*z*) and perpendicular (*x*).

4.3.2 Pump-probe and 2D IR data acquisition

The desired signal is separated from other contributions to the total detected intensity by differential detection. k_2 is chopped at 500 Hz and subtraction is performed between adjacent mid-IR shots where k_2 is open or blocked. τ_1 is set to 0 fs for pump-probe measurements, and the contributions to the detected intensity contributions are,

$$I_{open}(\tau_2,\omega_3) = \left|E_{k_3}(\omega_3) + E^{(1)}(\omega_3) + E^{(3)}(\tau_2,\omega_3)\right|^2 \tag{4.8a}$$

$$I_{closed}(\tau_2,\omega_3) = \left|E_{k_3}(\omega_3) + E^{(1)}(\omega_3) + \frac{E^{(3)}(\tau_2,\omega_3)}{2}\right|^2 \tag{4.8b}$$

E_{k_3} is the k_3 field, $E^{(1)}$ is the linear absorption signal, and $E^{(3)}$ is the pump-probe signal. We assume that $E_{k_3} >> E^{(3)}, E^{(1)}$, then only the E_{k_3} homodyne (*I*) and signal (*S*, i.e. $E_{k_3}E^{(3)}$ and $E_{k_3}E^{(1)}$) terms are significant in the detected intensity.

$$I_{open}(\tau_2,\omega_3) \approx I_{k_3}(\omega_3) + 2S^{(1)}(\omega_3) + 2S^{(3)}(\tau_2,\omega_3) \tag{4.9a}$$

$$I_{closed}(\tau_2,\omega_3) \approx I_{k_3}(\omega_3) + 2S^{(1)}(\omega_3) + S^{(3)}(\tau_2,\omega_3) \tag{4.9b}$$

Data is acquired on-the-fly as the change in absorption (ΔOD) by subtracting adjacent chopped and unchopped shots.

$$\Delta OD = -\log\left(\frac{I_{open}(\tau_2,\omega_3)}{I_{closed}(\tau_2,\omega_3)}\right) = -\log\left(1+\frac{S^{(3)}(\tau_2,\omega_3)}{I_{k_3}(\omega_3)+2S^{(1)}(\omega_3)+S^{(3)}(\tau_2,\omega_3)}\right) \tag{4.10}$$

It is generally assumed that the intensity of k_3 is much larger than the linear absorption and pump-probe signals, $I_{k_3} >> S^{(3)}, S^{(1)}$, although this may not be true for highly absorbing samples.

$$\Delta OD \approx -\log\left(1+\frac{S^{(3)}(\tau_2,\omega_3)}{I_{k_3}(\omega_3)}\right) \tag{4.11}$$

In a 2D IR experiment, the ΔOD signal is acquired as a function of τ_1, and the resulting FID is Fourier transformed to resolve the excitation frequency axis (ω_1). Pump-probe signals between k_1 and k_3, $E^{(3),PP}(\tau_1+\tau_2,\omega_3)$, and between k_2 and k_3, $E^{(3),PP}(\tau_2,\omega_3)$, are wavevector-matched in the same direction as the signal of interest, $E^{(3),2D}(\tau_1,\tau_2,\omega_3)$.

$$I_{open}(\tau_1,\tau_2,\omega_3) = \left|E_{k_3}(\omega_3)+E^{(1)}(\omega_3)+E^{(3),PP}(\tau_2,\omega_3)+E^{(3),PP}(\tau_1+\tau_2,\omega_3)+E^{(3),2D}(\tau_1,\tau_2,\omega_3)\right|^2 \tag{4.12}$$

When k_2 is chopped, neither $E^{(3),PP}(\tau_2,\omega_3)$ nor $E^{(3),2D}(\tau_1,\tau_2,\omega_3)$ are generated.

$$I_{closed}(\tau_1,\tau_2,\omega_3) = \left|E_{k_3}(\omega_3)+E^{(1)}(\omega_3)+E^{(3),PP}(\tau_1+\tau_2,\omega_3)\right|^2 \tag{4.13}$$

As for the pump-probe measurement, we assume that linear absorption and third-order homodyne terms are weak relative to the cross-terms and I_{k_3}. The resulting signal acquired as ΔOD is,

$$\Delta OD = -\log\left(1+\frac{S^{(3),PP}(\tau_2,\omega_3)+S^{(3),2D}(\tau_1,\tau_2,\omega_3)}{I_{k_3}(\omega_3)+S^{(1)}(\omega_3)+S^{(3)}(\tau_1+\tau_2,\omega_3)}\right) \tag{4.14}$$

Again, assuming that $I_{k_3} >> S^{(3)}, S^{(1)}$ leads to,

$$\Delta OD \approx -\log\left(1+\frac{S^{(3),PP}(\tau_2,\omega_3)+S^{(3),2D}(\tau_1,\tau_2,\omega_3)}{I_{k_3}(\omega_3)}\right) \tag{4.15}$$

Differential detection between shots with k_2 chopped and unchopped enables shot-to-shot removal of the k_1/k_3 pump-probe signal, which is critical due to its dependence on τ_1. In contrast, the k_2/k_3 pump-probe signal is independent of τ_1 and remains as an offset in the FID, which can simply be subtracted in post-processing.

4.3.3 T-jump optical design

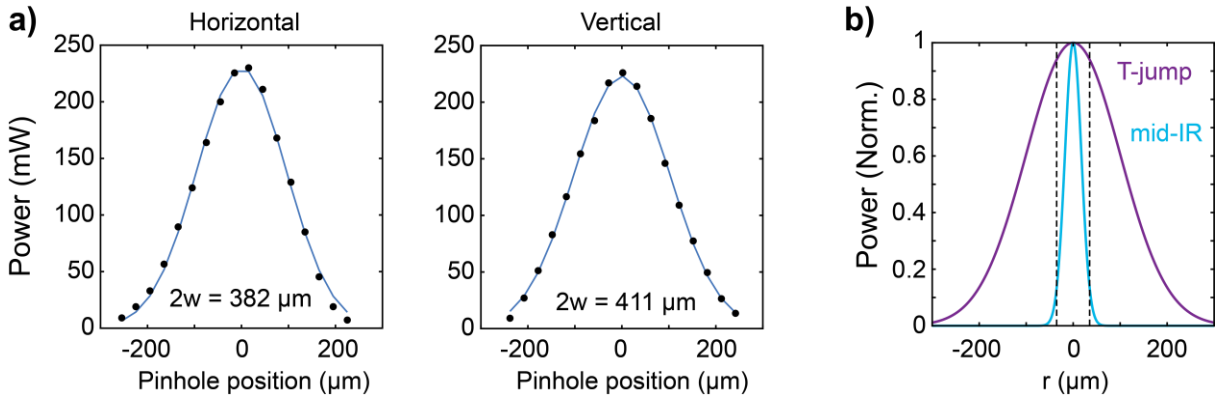

Figure 4.7 Characterization of CW T-jump beam at the sample. (a) Focused CW laser spatial profile at the sample position from measuring the power through a 150 μm wire extruder as it is scanned across the beam in 30 μm steps. Power profiles from scans in the horizontal (left) and vertical (right) directions are well described by a Gaussian function to extract $1/e^2$ beam diameters (2w) of 382 and 411 μm. **(b)** Comparison of T-jump and mid-IR beam sizes at the sample assuming a Gaussian profile and using the average diameter from horizontal and vertical profile measurements.

The T-jump laser is a CW Thulium fiber laser (TLR-30, IPG Photonics) that outputs 30 W of randomly polarized 1.94 μm light. The T-jump laser line uses AR-coated fused silica mirrors with >99% reflectance from 1.9 - 2.2 μm (Edmund Optics). The T-jump beam is first collimated with a Galilean telescope to a ~1.5 mm diameter before passing through a TeO_2 acousto-optic modulator (M1099(M)-T40L-2, Isomet) with a 2 mm active aperture height. An arbitrary waveform generator (AFG, 31101, Tektronix) sends a 40 MHz sine wave function with programmable amplitude (0 – 1 V) to a radiofrequency (RF) driver (531C-7, Isomet) to drive the AOM. The AOM is mounted on a manual rotation stage (RP03, Thorlabs) and rotated to the Bragg angle (9.05 mrad). The rise time (τ_{rise}) of the AOM is ~230 ns assuming a Gaussian spatial mode, $\tau_{rise} = 0.65d/V_a$, where d = 1.5 mm is the beam diameter and V_a = 4.2 × 10^6 mm/s is the acoustic velocity of TeO_2. A first-order diffraction efficiency of 70-75 % is achieved with maximum RF power. The zeroth- and first-order diffracted beams are separated by an angle of 18.10 mrad. After the beams sufficiently separate, the zeroth-order beam is blocked and the first-order beam is routed to the sample for optical heating. The first-order beam is focused by a convex lens (f = 20 cm) and routed through a hole in the first off-axis parabolic mirror to spatially overlap with the mid-IR beams with a $1/e^2$ diameter of ~400 μm (Fig. 4.5). After the sample, the T-jump beam exits through a hole in the second off-axis parabolic mirror and is blocked. The T-jump spot size is >4-fold larger than that of the mid-IR beams such that the spread in ΔT along r across the mid-IR probe is expected to be less than 10%.

4.3.4 T-jump synchronization with mid-IR

In CW T-jump experiments, the CW fiber laser is continuously running at full power, and the properties of the experiment are determined through time-dependent modulation of the power reaching the sample via the AOM. Therefore, the electronics that drive the AOM control experimental properties such as the T-jump repetition rate ($1/\tau_T$), ΔT, the heating (τ_{heat}) and cooling (τ_{cool}) times, and the time spent with the laser on (τ_{on}) and off (τ_{off}) during a cycle (Fig. 4.8). Mid-IR shots arrive at the sample every 1 ms and are sorted based on their time of arrival after the T-jump laser turns on (τ_{TJ}) and off ($\tau_{TD} = \tau_{TJ} - \tau_{off}$). Hundreds to thousands of measurements as a function of τ_{TJ} and τ_{TD} are acquired per T-jump cycle and binned logarithmically. The first mid-IR shot arrives between 0 and 1 ms (τ_{TJ}^{1}) after the T-jump laser turns on and is electronically controlled. The 80 MHz repetition rate from the Ti:Sapphire oscillator seeding the regenerative amplifier serves as the master clock as for the pulsed T-jump spectrometer. The 80 MHz signal is divided to 1 kHz by a delay generator (SDG Elite, Coherent), which is sent to the regenerative amplifier, and the 1 kHz signal is then divided to 20 Hz. This is the same 20 Hz signal used to synchronize firing of the pulsed T-jump laser with the mid-IR (Section 4.1.3). The 20 Hz signal is further divided to a programmable value between 10 and 0.005 Hz ($1/\tau_T$) by a series of two frequency dividers (PRL260NT & PRL220A, Pulse Research Lab) and triggers the AFG to start a waveform of duration τ_T that is converted to an RF signal to drive the AOM.

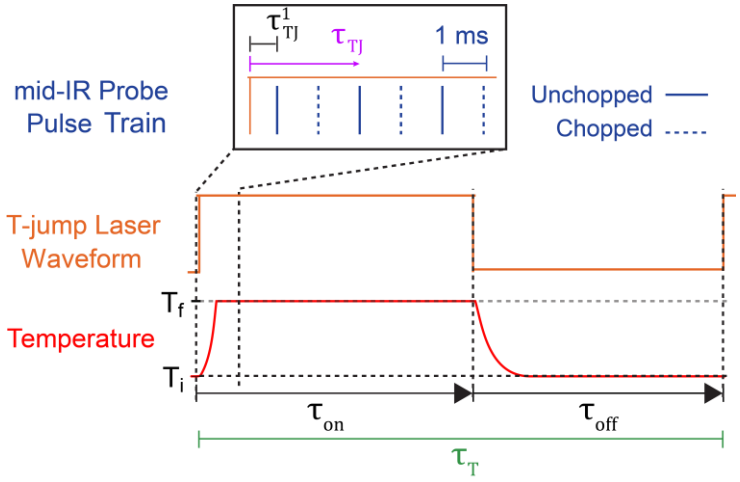

Figure 4.8 Definition of time delays in CW T-jump experiment. Once triggered, the AFG initiates a T-jump/T-drop waveform (orange) of duration τ_T. The waveform consists of electronically controlled time periods where the laser is heating the sample (τ_{on}) or off (τ_{off}). The time delay between the start of heating and the arrival of the first mid-IR shot (τ_{TJ}^1) is electronically controlled and later shots arrive in 1 ms intervals that are expressed in their time delay relative to the start of heating (τ_{TJ}).

As illustrated in Fig. 4.9, we employ two types of data acquisition procedures for $\tau_T \leq 2.5$ s (CW-fast) and $\tau_T > 2.5$ s (CW-slow). For each type of acquisition, the divided signal from the master clock triggers the AFG to initiate heating cycles. At this time, the AFG sends a 1 ms long TTL signal that is sampled by the integrator for the mid-IR detector. The TTL signal is sampled 40 μs after the rising edge of a 1 kHz signal to the integrator (Fig. 4.9b), which triggers integration of the mid-IR pulse detection (542 ns delay, 1.904 μs integration time). Following the 1 ms TTL, a user-defined electronic sub-ms delay controls the start time for sample heating and in turn determines τ_{TJ}^1.

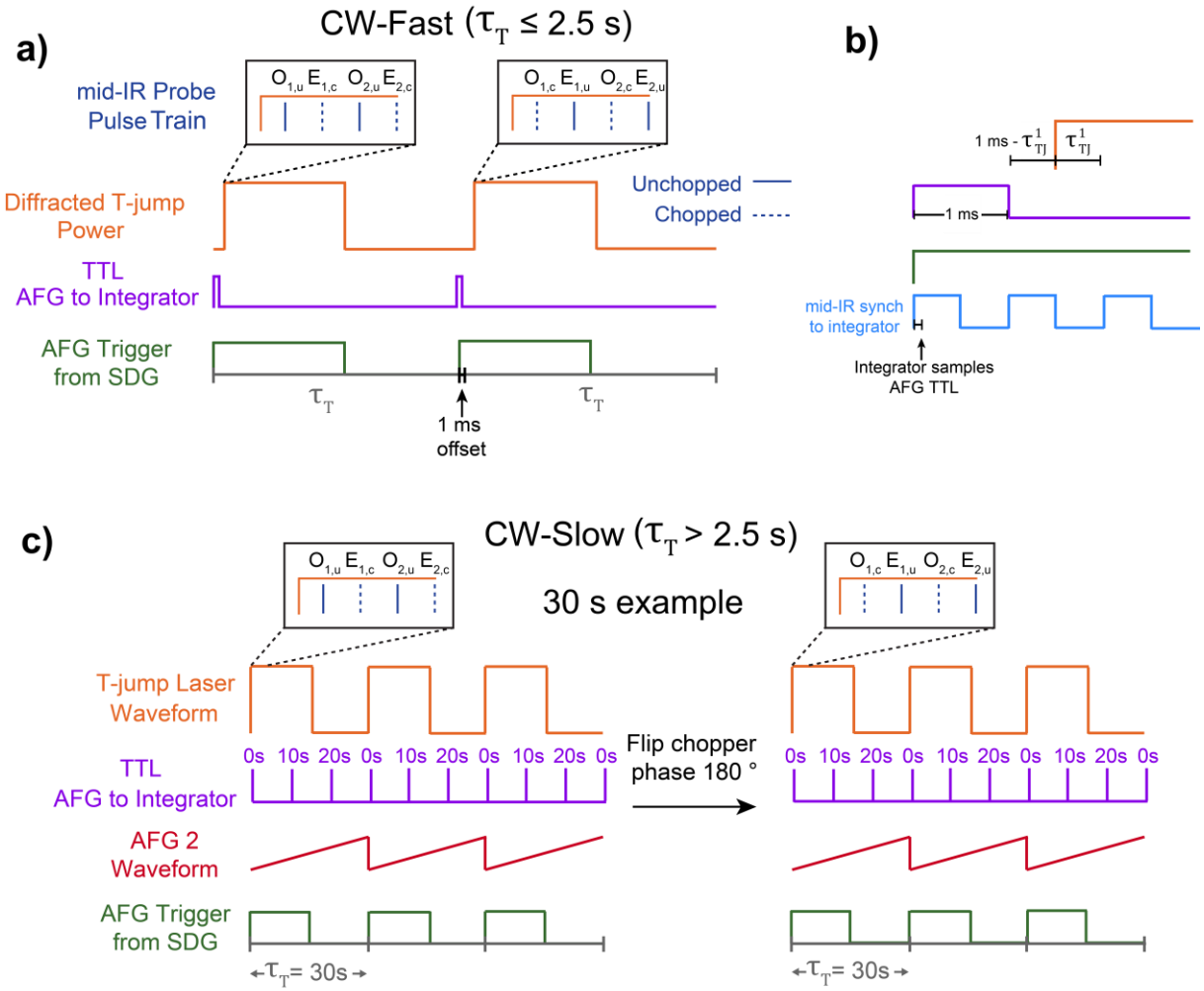

Figure 4.9 Electronic timing control in CW T-jump IR experiments. (a) Timing schematic for T-jump experiments with $\tau_T \leq 2.5$ s (CW-fast). A signal divided from the Ti:Sapphire oscillator master clock triggers the AFG to generate a pair of T-jump/T-drop cycles that are offset 1 ms relative to one another. Due to the 1 ms mismatch, mid-IR shots that have k_2 chopped or unchopped at τ_{TJ} in the first cycle will be flipped to unchopped or chopped in the second, respectively, and can be directly subtracted (eq. 4.16). **(b)** Zoom in on first few ms following AFG triggering. Immediately after being triggered, the AFG sends a 1 ms duration TTL signal that is read by the mid-IR detector integrator. Following the TTL, τ_{TJ}^1 is set by a user-defined delay that adjusts the start of the heating relative to the arrival of the third mid-IR shot. **(c)** Timing schematic for T-jump experiments with $\tau_T > 2.5$ s (CW-slow). Many T-jump/T-drop cycles are acquired at a given chopper phase relative to the AFG trigger such that all odd and even τ_{TJ} are chopped and is obtained by subtraction between data collected with each chopper phase. When $\tau_T > 10$ s, the

Figure 4.9 Electronic timing control in CW T-jump IR experiments, continued

AFG sends a TTL to the integrator every 10 s, and the integrator reads a voltage ramp from AFG 2 to determine which 10 s interval is measured. An example for τ_T = 30 s is shown.

A major consideration for CW-fast and CW-slow acquisition is how to minimize the time between subtraction of chopped and unchopped mid-IR shots, and different methods are used in each case. For T-jump pump-probe (t-PP) and T-jump 2D IR (t-2D IR) measurements, k_2 is chopped at 500 Hz to enable differential detection of the nonlinear signal of interest. To sample τ_{TJ} with resolution finer than 2 ms, we cannot subtract adjacent chopped and unchopped mid-IR shots as in steady-state measurements. Since k_2 is chopped at 500 Hz, half of the detected mid-IR shots will be chopped and half will be unchopped during the T-jump waveform. Figure 4.9a,c shows the scenario where the mid-IR shot at τ_{TJ}^1 is unchopped, and therefore all odd-numbered shots (τ_{TJ}^1, τ_{TJ}^1+2 ms, τ_{TJ}^1+4 ms,…) will be unchopped and all even numbered shots (τ_{TJ}^1+1 ms, τ_{TJ}^1+3 ms, τ_{TJ}^1+5 ms,…) will be chopped. However, the opposite scenario where odd and even numbered shots are chopped and unchopped, respectively, is equally likely to occur. The integrator reads a signal from the chopper to sort whether a given mid-IR shot is chopped or unchopped.

In CW-fast (Fig. 4.9a), a continuous series of mid-IR shots is detected over a pair of T-jump/T-drop cycles. The second cycle is electronically delayed 1 ms relative to the first such that unchopped(chopped) τ_{TJ} delays from the first cycle flip to be chopped(unchopped) in the second cycle, allowing for the ΔOD signal to be determined from direct subtraction between the pair of T-jump/T-drop cycles where *i* and *j* are the cycles with odd and even shots unchopped, respectively.

$$\Delta OD\left(\omega_3,\tau_{TJ}\right)=\begin{cases}-\log\left(\dfrac{I_{u,i}\left(\omega_3,\tau_{TJ}\right)}{I_{c,j}\left(\omega_3,\tau_{TJ}\right)}\right) & \text{if } \tau_{TJ} \text{ odd}\\ -\log\left(\dfrac{I_{u,j}\left(\omega_3,\tau_{TJ}\right)}{I_{c,i}\left(\omega_3,\tau_{TJ}\right)}\right) & \text{if } \tau_{TJ} \text{ even}\end{cases} \tag{4.16}$$

ΔOD data are then collected as a function of τ_{TJ}^{1} to provide fine sampling of sub-ms time delays, but the time resolution for measuring relaxation kinetics is set by the heating time (Section 4.4). The CW-fast approach provides the shortest time interval between subtraction of unchopped from chopped shots while maintaining < 2 ms resolution in τ_{TJ}. However, this approach is limited to $\tau_T \leq 2.5$ s by our current integrator and local PC. Only 10,000 mid-IR shots can be measured sequentially before the data must be moved, which corresponds to a maximum 10 s interval at a 1 kHz mid-IR repetition rate. For CW-fast acquisition, two full T-jump/T-drop cycles must be acquired together within a 10 s interval, which is only guaranteed for $\tau_T \leq 2.5$.

For $\tau_T > 2.5$ s, we employ the CW-slow data acquisition scheme (Fig. 4.9c). Data are acquired for several T-jump/T-drop cycles, n_{cyc}, at a fixed chopper phase relative to the heating such that a given τ_{TJ} is either chopped or unchopped, and then the chopper phase is shifted by 180° so that delays that were previously chopped(unchopped) are then unchopped(chopped). ΔOD signal is obtained from subtraction between data collected with each chopper phase (*i*, *i*+180°).

$$\Delta OD\left(\omega_3,\tau_{TJ}\right)=\begin{cases}-\log\left(\dfrac{\left\langle I_{u,i+180^\circ}\left(\omega_3,\tau_{TJ}\right)\right\rangle_{n_{cyc}}}{\left\langle I_{c,i}\left(\omega_3,\tau_{TJ}\right)\right\rangle_{n_{cyc}}}\right) & \text{if } \tau_{TJ} \text{ odd}\\ -\log\left(\dfrac{\left\langle I_{u,i}\left(\omega_3,\tau_{TJ}\right)\right\rangle_{n_{cyc}}}{\left\langle I_{c,i+180^\circ}\left(\omega_3,\tau_{TJ}\right)\right\rangle_{n_{cyc}}}\right) & \text{if } \tau_{TJ} \text{ even}\end{cases} \tag{4.17}$$

When $\tau_T \leq 10$ s, the entire T-jump/T-drop cycle can be acquired in a continuous series of mid-IR shots, and n_{cyc} is set to 1. If $\tau_T > 10$ s, multiple 10 s series of mid-IR shots ($n_{cyc} > 1$) must be collected to sample all τ_{TJ}. There is a few millisecond delay between the acquisition of each mid-IR shot series that prevents some τ_{TJ} values from being sampled during a given T-jump/T-drop cycle, and n_{cyc} is increased until full sampling of τ_{TJ} is achieved. After acquisition at each chopper phase, chopped and unchopped data at each τ_{TJ} are averaged across n_{cyc} prior to subtraction. A complication of measuring T-jump/T-drop cycles with $\tau_T > 10$ s is that the TTL sent from the AFG at the start of heating may not be sampled during a given 10 s cycle, and therefore mid-IR shots cannot be associated with their τ_T value. To overcome this, the AFG sends a TTL to the integrator at each 10 s interval, and the interval of the TTL is identified by reading a 16-bit voltage ramp (V) of duration τ_T that is triggered at the start of heating (Fig. 4.9c).

$$V\left(\tau_{TJ}\right)=\frac{V\left(\tau_T\right)-V\left(0\right)}{\tau_T}\tau_{TJ} \tag{4.18}$$

Nonlinearities in $V\left(\tau_{TJ}\right)$ and/or inaccurate sampling of the voltage by the integrator external channels makes $V\left(\tau_{TJ}\right)$ unsuitable to associate mid-IR shots with their τ_{TJ} value, yet it is accurate enough to identify which 10 s interval to associate with the TTL signal from the AFG.

4.3.5 Modulation of CW Laser Intensity

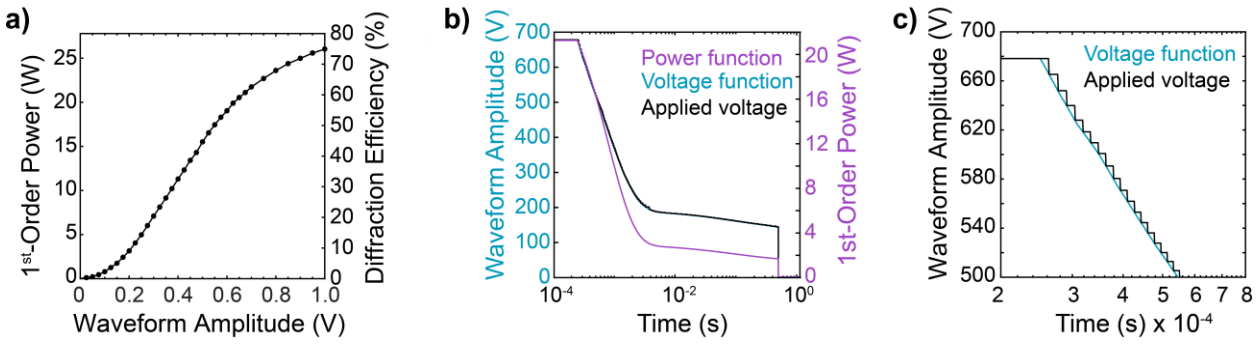

Figure 4.10 Generating seconds-long voltage waveforms for intensity modulation. (a) Measured 1st-order diffracted power as a function of AFG sine-wave voltage amplitude using 34 W of input power. The right axis shows the percentage of input power that is diffracted (diffraction efficiency). **(b)** Comparison of desired laser power profile at the sample (purple), converted voltage profile (blue), and discretized voltage profile built from sequence elements and outputted from the AFG (black). **(c)** Zoom in of **(b)** emphasizing the discretization of the real voltage profile.

The CW T-jump experiment requires precise modulation of the laser intensity at the sample on timescales ranging from microseconds to minutes. To accomplish this control there must be a robust mechanism to apply a user-defined laser power function. The AFG controls the modulation, and therefore the sine-wave amplitude output from the AFG must be calibrated with the actual 1st-order diffracted laser power. Figure 4.10a shows that the 1st-order diffracted power from the AOM follows a nonlinear relationship with sinewave amplitude from the AFG. This relationship is interpolated and saved in the Labview data acquisition program to convert the user-defined power function to AFG voltage (Fig. 4.10b).

Another complication of waveform generation comes from the limited memory available in the AFG. The 40 MHz resonant frequency of TeO_2 requires an AFG sampling rate of 80 MHz, and ≥160 MHz is recommended for best performance (≥2*Nyquist frequency). Our Tektronix

31101 AFG has a total memory of 1.28×10^6 points, which can only produce a 16 ms waveform at a Nyquist sampling rate. Even much more expensive AFG models would only enable a 320 to 640 ms waveform. Modern AFGs have an alternative mode of operation, known as "sequence mode", where short waveform segments are initially saved and then looped without any additional memory cost, and we use this method to assemble waveforms that extend out to minutes. First, a series of 5 μs waveform elements (40 MHz sinewave) with amplitude ranging from 0 to 1 V in 1 mV steps are saved to the AFG. The total memory of these 1001 waveforms with 160 MHz sampling is only $\sim 8 \times 10^5$ points. These elements are then arranged and looped to create a sequence that is sent to the AFG. Our AFG can hold up to 256 elements per sequence, and each element may be looped 1 to 999,999 times (or set to infinity). The small number of elements makes the true voltage waveform from the AFG coarse relative to the input waveform (Fig. 4.10c), but the voltage resolution is fine enough to effectively shape heating and cooling profiles as discussed in Section 4.4.

The AFG sequence not only creates the laser modulation waveform, but also controls the synchronization and timing discussed in Section 4.3.4. The first element of each sequence is a 1 ms (5 μs × 200) 0 V DC signal marked to produce a TTL signal that is sent to the mid-IR detector integrator and a second AFG for >2.5 s timescale experiments (Fig. 4.9). An additional 0 V DC element of duration 1 - τ_{TJ}^1 is added prior to the start of laser heating as a way to set τ_{TJ}^1 (Fig. 4.8), and a typical experiment steps through multiple durations of this element in order to change τ_{TJ}^1. This delay is followed by a series of elements used to generate the heating and cooling waveforms that end with a 0 V DC element set to an infinite loop in the T-drop, and the frequency-divided

signal originating from the mid-IR laser triggers the sequence to restart (Fig. 4.9). There are important differences between the sequence construction for CW-fast and CW-slow operation. In CW-fast, two T-jump cycles made of 126 elements each are combined in a single sequence and offset by a 1 ms element in order to perform a direct chopper subtraction between each T-jump cycle. CW-slow only requires one T-jump cycle but must be modified to generate 1 ms long TTL signals in 10 s intervals. This is accomplished by finding each 10 s interval in the sequence and breaking up the respective element into three elements with an intermediate 1 ms element used for generating the TTL signal.

4.4 Optimization of CW T-jump heating and cooling profile

4.4.1 Speeding up heating and cooling via intensity modulation

CW laser-based T-jump experiments often suffer from slow rise times to a fixed final temperature (T_f) as well as large time-dependent deviations in T_f.[26-27] Typically, optical heating is performed with a square power waveform where the CW laser power switches between on/off position:

$$P(t) = P_0 B\left(t_{on}, t_{off}\right) \tag{4.19}$$

Where the step-on/step-off function is

$$B\left(t_{on}, t_{off}\right) = \begin{cases} 1 & \text{if } t_{on} \le t \le t_{off} \\ 0 & \text{if } t < t_{on} \text{ or } t > t_{off} \end{cases}$$

With a simple on/off power waveform, T_f is achieved once heat deposition from the T-jump laser and thermal diffusion out of the sample reach a steady-state. Our sample is placed between two

CaF_2 windows spaced 50 μm apart. The thermal conductivity of CaF_2 is ~15x greater than that of D_2O, so thermal diffusion out of the sample primarily occurs through the CaF_2 windows. The blue curve in Fig. 4.11b,d shows the ΔT profile, determined from the percent change in k_3 transmission,[28] for a T-jump rise with an on/off waveform and power P_0 = 1.5 W, chosen to give a steady state ΔT = 16 °C (Fig. 4.11a,c). The heating occurs in two phases: a relatively fast phase in which ~75% of the total temperature change occurs within 10 ms, and a slow phase in which heating to a steady-state continues for tens to hundreds of milliseconds. The total rise can be fit to a sum of an exponential and stretched exponential function.

$$T(\tau_{TJ}) = \Delta T\left[1 - A\exp\left(-\left(\tau_{TJ}/\tau_{R1}\right)\right) + BA\exp\left(-\left(\tau_{TD}/\tau_{R2}\right)^{\beta_{R2}}\right)\right] + T_i \qquad (4.20)$$

where τ_{R1} = 6.4 ms, τ_{R2} = 60 ms, and β_R = 0.6 for the blue trace in Fig. 4.11.

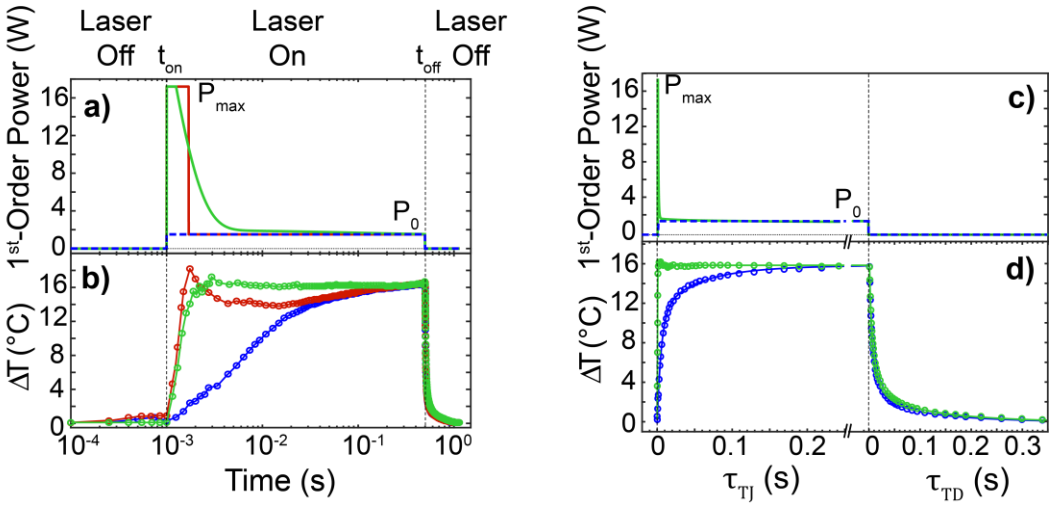

Figure 4.11 Optimization of heating profile with amplitude modulation. **(a)** Semilog plot of T-jump laser power at the sample as a function of time for different amplitude modulation waveforms applied to the AOM. **(b)** Corresponding change in solvent temperature (ΔT) as a function of time for different power waveforms. Measurements were performed with T_i = 20 °C and ΔT was extracted using the change in solvent transmission. **(c,d)** Similar figures to **(a)** and **(b)** on a linear time axis.

When the T-jump laser is turned off, the solvent temperature decays back to T_i, which is set by a re-circulating chiller. Traces of normalized k_3 transmission during the thermal relaxation are shown in Fig. 4.11d. In general, the thermal relaxation is also two phase, with most of the cooling happening quickly within a time scale of $\tau_{D1} \approx 5$ ms and followed by a slower non-exponential phase. The relaxation is well fit by the sum of an exponential and stretched exponential function with minor differences between the datasets with different heating waveforms.

$$T(\tau_{TD}) = \Delta T\left[A\exp\left(-\left(\tau_{TD}/\tau_{D1}\right)\right) + BA\exp\left(-\left(\tau_{TD}/\tau_{D2}\right)^{\beta_{D2}}\right)\right] + T_i \tag{4.21}$$

where $\tau_{D2} \approx 20$ ms and the stretch parameter (β_{D2}) is 0.45.

The thermal profile for on/off CW heating is problematic for T-jump spectroscopy as the solvent temperature rises slowly and varies substantially over most of the measured time window. We optimize the solvent heating profile with time-dependent modulation of the T-jump laser power using an AOM. To increase the rate of heating, the AOM is set to diffract the maximum amount of power (P_{max}) for a short time (<1 ms) and then drop to a plateau power (P_0) identical to the elevated power with the on/off waveform (Fig. 4.11a,c).

$$P(t) = P_{\max}B(0, 0.8\text{ ms}) + P_0B(0.8, 500\text{ ms}) \tag{4.22}$$

The red trace in Fig. 4.11 illustrates a step function between two laser powers with P_{max} = 17 W and P_0 = 1.5 W, which is the simplest waveform for achieving faster heating. While the solvent temperature rises to its max value in < 1 ms, it subsequently drops by nearly 20% following the

step down in laser intensity and then increases back to the max temperature over tens of milliseconds. The temperature drop from 1 to 10 ms is also observed in T-jump simulations[28] and results from an initial uneven balance between thermal diffusion out of the sample window and energy deposition due to the sudden drop in laser power. The subsequent slow rise appears similar to that observed in the on/off heating case.

To flatten the temperature profile at T_f, we employed a more complicated power waveform where the high-to-plateau power transition occurs via a two-component decay (green trace, Fig. 4.11) rather than through a step function. This form varies the input T-jump power to more closely match the thermal diffusion out of the sample region before reaching steady-state heating.

$$P(t) = P_{\max} B\left(0, t_{step}\right) + P_1 \exp\left(-\left(t - t_{step}\right)/t_1\right) + P_2 \exp\exp\left(-\left[\left(t - t_{step}\right)/t_2\right]^{\beta}\right) + P_0 B\left(t_{step}, 500\text{ ms}\right) \quad (4.23)$$

The step-up (t_{step} = 0.25 ms) phase enables rapid heating but is cut short relative to the step function waveform to allow for an exponential phase (P_1 = 15 W, t_1 = 0.8 ms) that contains most of the change in power in the high-to-plateau transition. Shortening t_{step} in exchange for the exponential phase corrects for the 1-10 ms temperature drop observed when the step function waveform is applied but also extends the heating rise time from ~0.8 to ~1 ms. The remainder of the high-to-plateau power transition is a stretched exponential decay that extends for hundreds of milliseconds (P_2 = 0.5 W, t_2 = 100 ms, β = 0.55) and corrects for the slow temperature increase observed when the step function and on/off waveforms are used. Once heated within ~1 ms, the sample maintains T_f to ±3% (°C) over the entire T-jump time window, minimizing potential temperature-dependent artifacts in T-jump spectroscopy. Due to the time-dependent CW laser

power, the solvent heating rise is not a simple exponential. The temperature profile can be fit to a compressed exponential function.

$$T\left(\tau_{TJ}\right) = T_f - \Delta T \exp\left(-\left(\tau_{TJ} / \tau_R\right)^{\beta_R}\right) \tag{4.24}$$

The data in Fig. 4.10b,d was fit to a time constant (τ_R) of 0.5 ms and compression parameter (β_R) of 1.7.

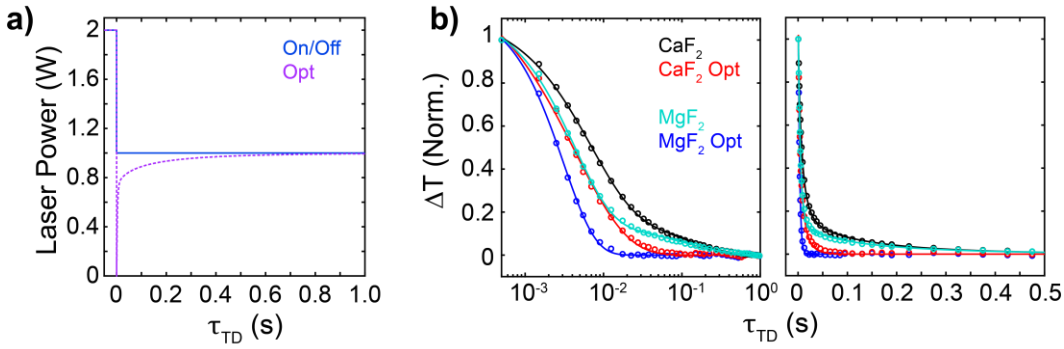

Figure 4.12 Optimization of cooling profile with amplitude modulation. **(a)** Plot of T-jump laser power at the sample for a step T-drop (On/Off) and optimized for fastest cooling (Opt) using MgF_2 and CaF_2 windows. **(b)** Normalized ΔT as a function of time with and without optimization using MgF_2 and CaF_2 windows. Data is presented on a (left) log and (right) linear time axis.

Modulation of T-jump laser power may also be used to speed up the cooling of the solvent temperature back to T_i (Fig. 4.12). In this scenario, a non-zero laser power (P_{0D}) fixes T_i above the set point of the recirculating sample chiller. Rather than applying a step for the T-drop, the laser power drops to 0 and recovers to P_{0D} with a sum of exponential and stretched exponential decays.

$$P(t) = P_{0D} - \left[P_{1D} \exp\left(-t / t_{D1}\right) + P_{2D} \exp\left(-\left(t / t_{D2}\right)^{\beta}\right)\right] \tag{4.25}$$

The role of each exponential is to account for both components in the on/off waveform as done for heating (eq. 4.23). In Fig. 4.12a, P_{0D} = 1 W and the first exponential (P_{1D} = 0.7 W, t_{D1} = 3 ms) carries most of the amplitude and rapidly decays while the second component (P_{2D} = 0.3 W, t_{D2} = 100 ms, β = 0.6) is lower amplitude and long-lived. Application of this waveform nearly removes the long-lived component and the cooling profile can be fit to a single compressed exponential function.

$$T\left(\tau_{TD}\right) = T_i + \Delta T \exp\left(-\left(\tau_{TD} / \tau_D\right)^{\beta_D}\right) \tag{4.26}$$

where the profile taken with CaF_2 windows in Fig. 4.12b was fit to a time constant (τ_D) of 10 ms and compression parameter (β_D) of 1.1.

4.4.2 Heating and cooling with different window materials

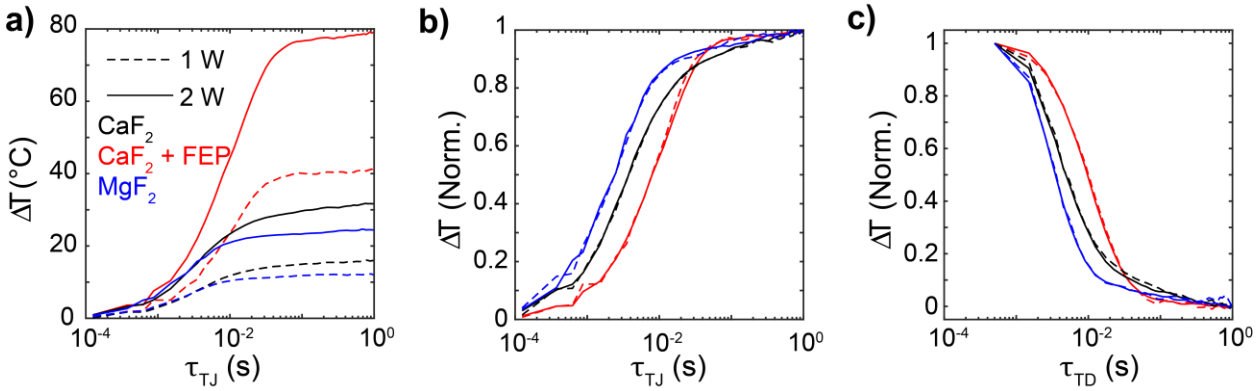

Figure 4.13 Effect of window material on CW heating and cooling. **(a)** ΔT for step heating (no optimization) of D_2O using CaF_2, CaF_2+FEP, and MgF_2 windows and 1 W (dashed lines) or 2 W (solid lines) laser power. A 50 μm sample path length is used for all measurements **(b)** Normalized heating profiles. **(c)** Normalized cooling profiles after switching off laser power at the sample.

The simulations in Chapter 3 demonstrated that the heating and cooling properties of the sample are sensitive to different window materials, and here I discuss how these different materials may be used to improve CW T-jump and T-drop measurements. In addition to bare CaF_2, we currently use bare MgF_2 windows and CaF_2 windows coated with a melted 12.7 μm layer of fluorinated ethylene propylene (FEP, DuPont). Figure 4.13 illustrates the differences in D_2O heating and cooling profiles with each window material using an on/off laser power waveform, and Tables 4.1 and 4.2 summarize the profile properties. The overall heating and cooling times are shortest for the material with the highest thermal diffusion coefficient (MgF_2) and longest for the most insulating windows (CaF_2+FEP). Slower thermal diffusion out of the sample also leads to a larger steady-state temperature change (ΔT_0) for a given laser power (P_0). A laser power of 1 W produces ΔT_0 ~ 40 °C with CaF_2+FEP but only ΔT_0 ~ 15 °C with bare CaF_2, which is a much more dramatic for effect than predicted from the heat diffusion simulations in Chapter 3. MgF_2 and CaF_2+FEP are also shown to reduce the amplitude of the slow stretched exponential heating and cooling component relative to bare CaF_2. The origin of this response is already unclear, and this trend suggests that the slow component amplitude is not directly related to the rate of thermal diffusion through the window. Instead, the slow component may arise from the window-D_2O interface that is altered across the materials.

Table 4.1 Fit parameters for CW heating (on/off) of D_2O with different window materials.

Window	ΔT_0 (°C)	A (%)	τ_{R1} (ms)	B (%)	τ_{R2} (ms)	β_R
MgF_2	12	84	2.8	16	86	0.39
CaF_2	16	67	4.2	33	32	0.42
CaF_2 + FEP	41	92	11.4	8	100	0.44

Table 4.2 Fit parameters for CW cooling (on/off) of D_2O with different window materials.

Window	A (%)	τ_{D1} (ms)	B (%)	τ_{D2} (ms)	β_D
MgF_2	92	3.7	8	110	0.76
CaF_2	80	4.9	20	56	0.60
CaF_2 + FEP	94	12.1	6	110	0.91

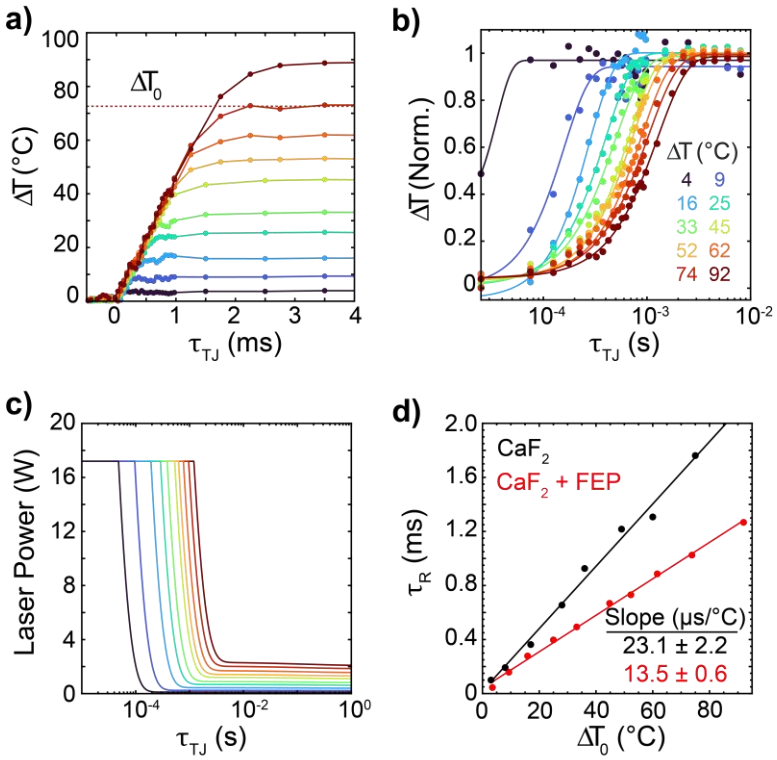

Figure 4.14 Relationship between ΔT and τ_R. **(a)** Optimized heating profiles of D_2O using variable P_0 and CaF_2+FEP windows. **(b)** Normalized heating profiles fit to a compressed exponential rise: $\Delta T(t) = 1 - \exp\left[-(t/\tau_R)^{\beta}\right]$ where $\beta \geq 1$. **(c)** Laser power waveforms used to generate heating profiles in **(a,b).** **(d)** τ_R as a function of steady-state temperature change (ΔT_0) using bare CaF_2 and CaF_2+FEP. Data follow a linear relationship and the slope for each window material is reported.

The different window materials provide distinct advantages for T-jump and T-drop spectroscopy. Our T-drop measurements are limited to probing kinetics slower than the multiple

millisecond cooling time. Switching to MgF_2 windows with laser intensity modulation shortens the cooling profile to a compressed exponential (eq. 4.26) with τ_D = 4 ms and β_D = 1.3. Faster cooling can be achieved with a shorter sample path length (Fig. 3.9) or window materials with greater thermal diffusivity, but each of these effects will proportionally reduce ΔT_0 and increase heterogeneity in ΔT across the probe volume.

The CaF_2+FEP window presents numerous advantages over bare CaF_2 for T-jump measurements. The thermal insulation from FEP reduces the steady-state thermal gradient along the focal direction of the sample, which reduces heterogeneity in ΔT. Further, the insulating effect enables faster heating than bare CaF_2 when laser intensity modulation is used (Fig. 4.14). The optimized heating time constant scales linearly with ΔT_0, and the slope for bare CaF2 and CaF_2+FEP are 23.1 and 13.5 μs/°C, respectively, such that the heating timescales are similar with each window material for small T-jumps ($\Delta T_0 \leq 10$ °C) but become further separated as ΔT_0 increases. CaF_2+FEP windows also make it much easier to create a flat temperature profile from 10 ms to 1 s. Correcting for the slow heating component requires fine tuning of the laser power waveform, and the optimal waveform changes from sample to sample. CaF_2+FEP windows reduce the amplitude of this component 4-fold, and therefore much less waveform optimization is necessary to achieve a time-independent ΔT profile beyond 1 ms.

4.4.3 Long-time heating and cooling artifacts

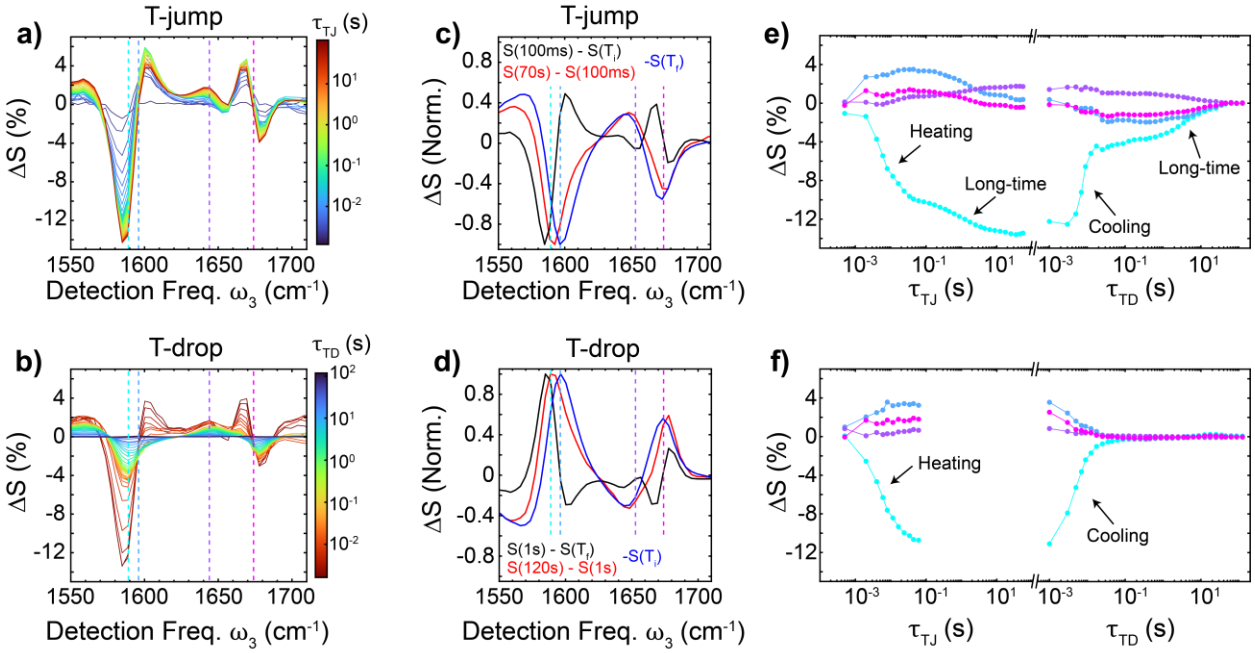

Figure 4.15 Observation of ms-to-s signal change during heating and cooling. (a) Percent change in pump-probe signal of 10 mg/ml diglycine in D_2O from τ_{TJ} = 0.5 ms to 75 s following a step T-jump from 20 to 45 °C. Measurements are performed with bare CaF_2 windows and a 50 μm path length. **(b)** Same signal during the subsequent T-drop from τ_{TD} = 1 ms to 125 s. **(c)** Comparison of the T-jump difference spectrum between τ_{TJ} = 100 ms and T_i (black), difference spectrum between τ_{TJ} = 70 s and 100 ms (red), and the negative of the pump-probe signal at T_i (blue). **(d)** Comparison of the T-drop difference spectrum between τ_{TD} = 1 s and T_f (black), difference spectrum between τ_{TD} = 120 s and 1 s (red), and the negative of the pump-probe signal at T_f (blue). **(e)** Pump-probe signal change at select frequencies indicated in **(a-d)** for a T-jump cycle with τ_{on} = 75 s and τ_{off} = 125 s. **(f)** Similar plot for a T-jump cycle with τ_{on} = 100 ms and τ_{off} = 125 s.

The previous sections focused on the characterization and optimization of the heating and cooling profiles up to delays of 1 s. Before moving on to discuss experiments performed over longer timescales, it is important to note that T-jump and T-drop measurements exhibit additional components spanning from 100 ms to 100 s considered as artifacts for our purposes. Figure 4.15 shows an example of diglycine T-jump and T-drop PP data collected out to τ_{TJ} = 75 s and

τ_{TD} = 125 s, which does not exhibit any large-scale structural changes or assembly. After heating or cooling on a tens of ms timescale, a second component is observed from ~100 ms to ~50 s that may correspond to the fully resolved τ_{R2} and τ_{D2} components. The signal change associated with the component is distinct from heating/cooling and resembles a loss of PP signal in the T-jump and gain of signal in the T-drop (Fig. 4.15c,d). Stopping heating prior to τ_{TJ} = 100 ms completely removes the long-time component from the T-drop, confirming that the origin of the long-time component is the same for T-jump and T-drop. A nearly identical response is observed in oligonucleotides and proteins (not shown), and these samples show an additional component on timescales longer than 10 s that may correspond to thermophoresis (Chapter 3.4). However, further characterization is required to determine the origin of these long-time components, and caution should be taken to distinguish these components from kinetics of interest.

4.5 Acquisition and processing of CW T-jump data

4.5.1 CW T-jump linear IR absorption

The simplest experiment to carry out on the CW T-jump spectrometer monitors the change in transmission of the k_3 beam following the T-jump in order to extract the change in sample linear absorption (t-A). This is the most common type of T-jump IR experiment.[3, 29-31] Our setup is not optimized for t-A spectroscopy, but it may be collected for free in t-PP or t-2D IR measurements. We do not use a separate reference detector, so the t-A signal is acquired as the difference between spectral intensity at a given τ_{TJ}, $I(\omega, \tau_{TJ})$, and that at the initial temperature, $I(\omega, T_i)$.

$$\Delta OD\left(\omega,\tau_{TJ}\right)=-\log\left(\frac{I\left(\omega,\tau_{TJ}\right)}{I\left(\omega,T_i\right)}\right) \tag{4.27}$$

In practice, $I(\omega,T_i)$ is taken as the average spectral intensity at long τ_{TD} delays following re-equilibration of the of the sample at T_i. The measured difference spectrum then contains the time-dependent change in the solvent and sample absorption. When measured on its own, the t-A signal can be acquired without optical chopping. However, when acquired simultaneously with the t-PP or t-2D IR signal, the t-A signal is extracted only from mid-IR shots where k_2 is chopped to avoid distortions from the k_2/k_3 PP and 2D IR signals.

In the mid-IR, t-A spectra contain a broad background due to the temperature-dependent transmission of the bend-libration combination absorption band of D_2O. When the solvent difference signal is comparable to or much larger than the t-A signal of interest, the t-A spectra will be distorted by time-dependent changes in temperature. The D_2O background must be removed to extract the t-A spectrum of interest. The most common method to remove the transient D_2O background in T-jump IR spectroscopy is to subtract subsequent datasets acquired on the sample and the solvent, which doubles the data acquisition time.[2] Another approach is to apply a discrete wavelet transform (DWT) to the raw t-A data to separate the broad D_2O background from the narrow spectral changes arising from the sample of interest. The DWT method is advantageous because it does not require an additional D_2O dataset and has previously been used to separate broad backgrounds from signals of interest in IR and Raman spectroscopy.[32-34]

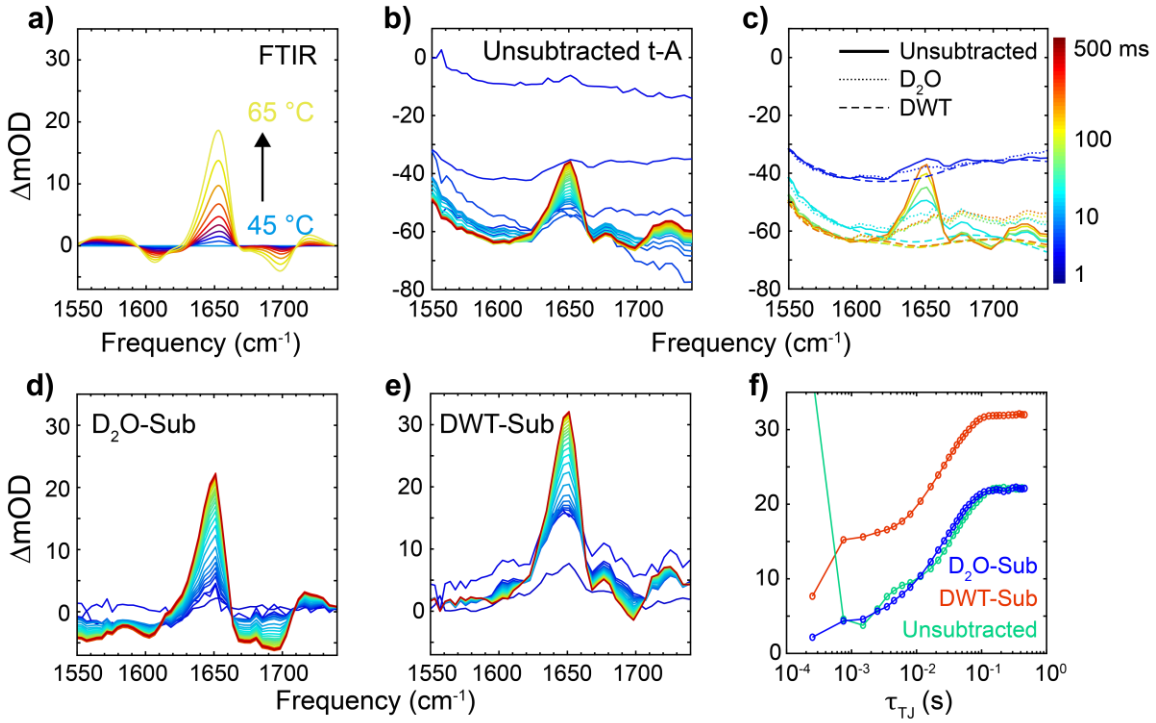

Figure 4.16 Extraction of T-jump linear absorption signal from change in mid-IR transmission. **(a)** FTIR difference spectra of C3T3 at pH* 4.5 from 45 to 65 °C. **(b)** Change in k_3 absorption as a function of τ_{TJ} for C3T3 following a T-jump from 45 to 65 °C. **(c)** t-A spectra of D_2O (dotted lines) and DWT reconstructed background spectra (dashed lines) plotted with the t-A spectra of C3T3 at select τ_{TJ} values. **(d)** C3T3 t-A spectra after subtraction of D_2O t-A spectra. **(e)** C3T3 t-A spectra after subtraction of DWT reconstructed background spectra. **(f)** Time traces of unsubtracted, D_2O-subtracted, and DWT-subtracted t-A data at a probe frequency of 1656 cm^{-1}. The unsubtracted time trace is vertically shifted to the D_2O-subtracted amplitude.

We compare each method of D_2O background removal to extract the t-A signal for unfolding of a model DNA i-motif, 5′-CCCTTTCCCTTTCCCTTTCCC-3′ (C3T3). The Mallat algorithm[35] is used for performing the DWT on the spectra, and the Daubechies wavelet family[36] was found to be most effective for separating the D_2O background. The unsubtracted t-A data for C3T3 following a T-jump from 45 to 65 °C is shown in Fig. 4.16b. A ~60 mOD loss of D_2O absorption occurs within ~1 ms, and the changes due to C3T3 unfolding are observed from 1 to

500 ms. t-A spectra of D_2O from a separate measurement (dotted lines) as well as reconstructed backgrounds (dashed lines) from the DWT method are shown with the C3T3 t-A data (solid lines) in Fig. 4.16c. The background spectra differ primarily between 1620 and 1740 cm^{-1}. The DWT reconstructed background is almost independent of frequency in this range while the D_2O spectra increase from 1620 to 1720 cm^{-1}. The t-A data corrected with the D_2O dataset resembles the FTIR difference spectra of C3T3 between 45 and 65 °C (Fig. 4.16a). In contrast, the main feature at 1650 cm^{-1} of the DWT processed t-A data is offset by ~10 mOD, indicating poor background subtraction in that frequency region.

Proper D_2O background subtraction is particularly important for intensity-modulated CW T-jump t-A data because the temperature profile near T_f may exhibit time-dependent structure that will introduce kinetic artifacts into the raw data. In Fig. 4.16f, a time trace from the unsubtracted t-A data of C3T3 probed at 1656 cm^{-1} exhibits two apparent components occurring from 1 - 10 ms and 10 – 100 ms. However, the first component is removed in both the D_2O- and DWT-subtracted data, giving a single-component rise from 1 to 100 ms that is consistent with the t-PP data in Fig. 4.17a. The kinetic response from the D_2O- and DWT-subtracted data are essentially equivalent, indicating that the DWT method can properly account for time-dependent changes in the t-A background. Therefore, the DWT method may still be useful for extracting t-A kinetic data in the absence of an additional background measurement of D_2O.

4.5.2 CW T-jump pump-probe

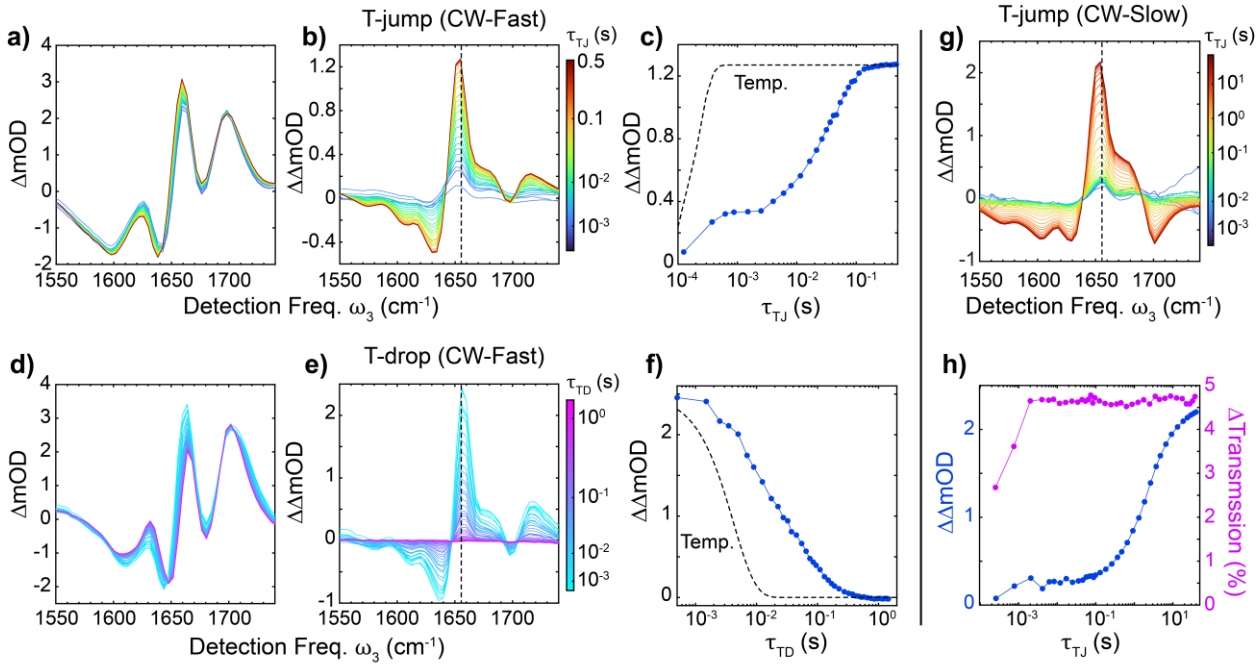

Figure 4.17 T-jump pump-probe with CW-fast and CW-slow acquisition. **(a)** PP spectra of C3T3 at pH* 4.5 from τ_{TJ} = 0.1 to 500 ms for a T-jump from 55 to 65 °C using CaF_2+FEP windows. Data were acquired with the CW-fast method, and τ_2 is fixed at 150 fs. **(b)** Difference spectra relative to the spectrum at T_i. **(c)** PP signal change at 1656 cm^{-1} and solvent temperature profile (dashed line). **(d-f)** Similar plots for a T-drop from 76 to 61 °C using MgF_2 windows. **(g)** PP difference spectra of C3T3 at pH* 6.3 from τ_{TJ} = 0.5 ms to 50 s following a T-jump from 28 to 40 °C. Data were acquired with the CW-slow method. **(h)** PP signal change at 1656 cm^{-1} and change in solvent transmission profile (pink, 1550 cm^{-1}) used to determine ΔT.

Pump-probe spectroscopy is the most commonly employed probe with our CW T-jump spectrometer. The PP spectrum does not require solvent background subtraction and is more sensitive to changes in nucleic acid secondary structure than the linear absorption spectrum (Chapter 2). Acquisition of t-PP data is also more efficient than t-2D IR. Relative to the t-HDVE measurement used for pulsed T-jump (Section 4.1), the t-PP data requires minimal processing. Importantly, there is no phasing step for PP spectra. Many more time points are collected in CW

T-jump experiments than for pulsed T-jump since the kinetics of interest are typically slower than the 1 kHz repetition rate of the mid-IR probe. Data are finely binned along τ_{TJ} and τ_{TD} on-the-fly to minimize data file sizes and then more coarsely binned in post-processing. Data at each τ_{TJ}^{1} delay are acquired in separate iterations and averaged in post-processing.

T-jumps of C3T3 at different pH conditions are shown to illustrate acquisition of t-PP data (Fig. 4.17). Under acidic conditions, unfolding occurs on 1 – 100 ms timescales and the CW-fast method (Fig. 4.9) is used to acquire PP spectra from τ_{TJ} = 0.1 to 500 ms. The PP spectrum of C3T3 contains three GSB features at 1632, 1665, and 1700 cm^{-1} that correspond to cytosine and thymine carbonyl and ring stretch vibrations (Chapter 2), and the 1632 and 1665 cm^{-1} features interfere with ESA bands. As for pulsed T-jump data, we typically report the t-PP data as a difference relative to the T_i spectrum to emphasize changes after the T-jump.

$$\Delta\Delta\mathrm{OD}(\omega_3,\tau_{TJ}) = -\log\left(\frac{I_{open}(\omega_3,\tau_{TJ})}{I_{closed}(\omega_3,\tau_{TJ})}\right)+\log\left(\frac{I_{open}(\omega_3,T_i)}{I_{closed}(\omega_3,T_i)}\right) \tag{4.28}$$

The largest amplitude difference feature is the gain at 1656 cm^{-1}, which shows kinetic components from τ_{TJ} = 0.1 – 1 ms and 3 – 100 ms that correspond to solvent heating and i-motif unfolding, respectively. The solvent heating response contains spectral change from all structural changes faster than the heating and therefore is highly dependent on the sample. t-PP T-drop measurements are shown under a similar condition (Fig. 4.17d-f) using MgF_2 windows. The two kinetic components are partially overlapped in time due to the slower time resolution relative to heating, and as a result T-drop kinetic data are typically less reliable. Increasing the solution pH* to 6.3 slows i-motif unfolding and folding by over an order of magnitude, requiring a switch to

acquisition with the CW-slow method (Fig. 4.17g,h). The change in solvent transmission at 1550 cm^{-1} indicates that a constant T_f is maintained for the full 50 s delay range. The data quality at τ_{TJ} ≤ 10 ms is significantly lower with CW-slow method due to the lower repetition rate of the T-jump and larger time interval between shots in the chopper subtraction. An alternative approach (not shown) for acquiring slow T-jump kinetics is to simply use a shot-to-shot chopper subtraction as for steady-state PP and 2D IR (Section 4.3.2) that will limit the time resolution along τ_{TJ} and τ_{TD} to 2 ms.

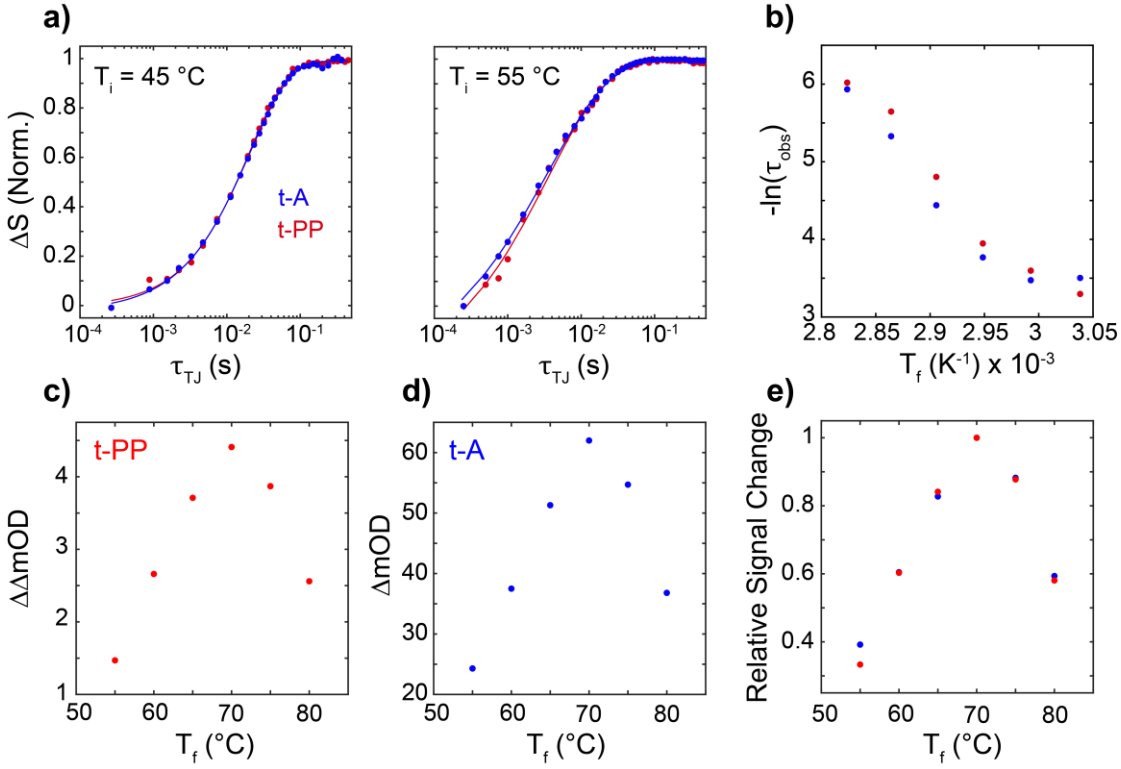

Figure 4.18 Comparison of t-PP and t-A kinetic data. **(a)** Normalized time traces at 1656 cm^{-1} for (blue) t-A spectra following D_2O background subtraction and (red) t-PP spectra of C3T3 for T-jumps with T_i 45 and 55 °C. Data (circles) are fit (solid lines) to a sum of one or two exponentials with a constant offset. **(b)** Observed rates extracted from fits plotted as a function of T_f. **(c)** Maximum signal change in t-PP spectra as a function of T_f. **(d)** Same for t-A. **(e)** Normalized trends in maximum signal change for t-PP and t-A.

The detected t-PP spectrum scales linearly with the amplitude of the signal electric field, therefore the t-PP spectrum should linearly map to sample population as shown previously for HDVE spectroscopy.[37] To confirm that the t-PP signal scales linearly with sample population, we compare the t-PP and t-A response for C3T3. Following background subtraction, the t-A data was compared with the t-PP response of C3T3. We compare the observed rate and amplitude of the t-A and t-PP response across the C3T3 melting transition because each observable depends on the change in population between unfolded and folded species. Figure 4.18 shows t-A and t-PP time traces at 1656 cm^{-1} normalized to the same scale. Each signal response nearly overlaps for both the 45 to 65 and 55 to 75 °C T-jumps. The observed relaxation rate ($1/\tau_{obs}$) reporting on unfolding/folding kinetics is obtained by fitting the time traces to a sum of one or two exponentials. $1/\tau_{obs}$ from t-A and t-PP exhibit a similar increase with temperature, consistent with an activated process like unfolding. The fit value of $1/\tau_{obs}$ from t-A is slightly lower (10-20 s^{-1}) than that from t-PP data.

For a series of constant ΔT T-jumps with various T_i across the unfolding/folding transition, the amplitude of the observed relaxation response is expected to follow the first derivative of the thermal melting curve: increase until the population change is greatest and then decrease as T_i moves further along the unfolding transition. As shown in Fig. 4.18c-e, the maximum of the t-PP and t-A difference spectra rises until T_f = 70 °C and decays at higher temperatures. When each temperature-trend is normalized, the relative signal change in the t-PP response tracks that of the t-A response, suggesting that each are similarly related to changes in population of the sample. Overall, both the similarities in relaxation kinetics and signal magnitude between the t-PP and t-A

probes demonstrates that the t-PP spectrum is linearly related to unfolded and folded populations and that t-PP and t-A measurements should yield the same kinetic information.

4.5.3 CW T-jump 2D IR

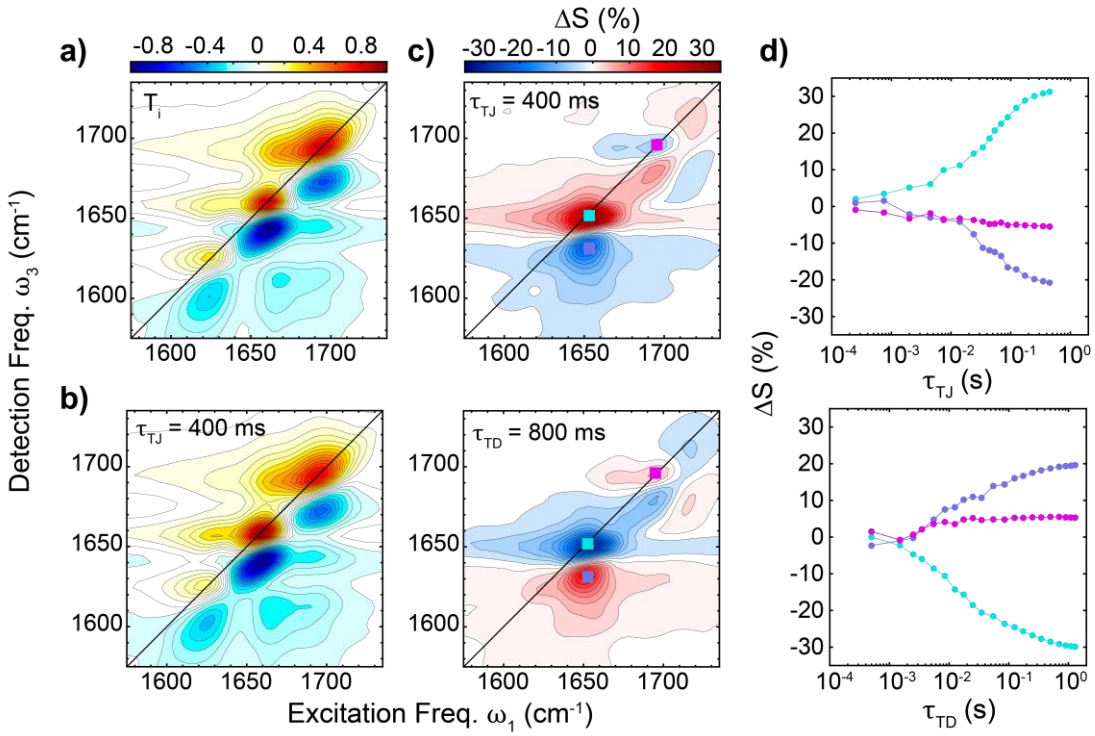

Figure 4.19 CW T-jump 2D IR of i-motif unfolding and folding. (a) 2D IR spectrum of C3T3 in pH* 4.5 solution at T_i = 44 °C using parallel pulse polarization (ZZZZ) and τ_2 = 150 fs. **(b)** t-2D IR spectrum at τ_{TJ} = 400 ms following a T-jump from 44 to 64 °C. **(c)** Difference spectra at (top) τ_{TJ} = 400 ms relative to the 44 °C spectrum and (bottom) for the T-drop at τ_{TD} = 800 ms relative to the 64 °C spectrum. Data are plotted in percent change relative to the initial spectrum, $\Delta S(\tau) = [S(\tau) - S(T_i)]/\max(S(T_i))$. **(d)** Time traces for (top) T-jump and (bottom) T-drop at select frequencies that are color-coded and marked in **(c)**.

CW t-2D IR data were acquired by stepping τ_1 and averaging the signal over multiple T-jump cycles before moving the next delay. As for pulsed t-2D IR, we typically undersample along τ_1 to reduce data acquisition time, and the data shown in Fig. 4.19 was collected with 24 fs steps from -160 to 2000 fs. The mid-IR k_1/k_2 interferogram is simultaneously acquired from

detection of the π-shifted output of the Mach-Zehnder interferogram and used to correct for phase errors in 2D IR surfaces at each τ_{TJ} and τ_{TD} via the Mertz method.[13, 25]

The 2D IR spectra of C3T3 resolves the PP content along ω_1, revealing the cross-peaks and lineshape information (Fig. 4.19). Following the T-jump, unfolding most noticeably leads to an increase in signal of the doublet at 1656 cm^{-1} as well as a loss and gain at 1700 and 1715 cm^{-1}, respectively. The reverse changes our observed following the T-drop. In contrast to our pulsed T-jump setup, the processes probed with CW T-jump are slow enough that 2D IR spectra are acquired at numerous τ_{TJ} and τ_{TD} points of interest with our 1 kHz probe source (Fig. 4.19d). Therefore, millisecond-to-minute kinetics may be finely sampled with t-2D IR in a relatively time efficient manner. These different roles of t-2D IR in our pulsed and CW T-jumps setups suggest that switching to higher repetition rate mid-IR sources will improve the data quality of T-jump PP and 2D IR and expand the range of processes that may be studied. Indeed, recent improvements in ytterbium (Yb)-based laser technology and mid-IR detection have enabled a few groups to perform transient 2D IR experiments with 100 kHz repetition rates.[38-39]

4.6 Accessible time windows for kinetics from pulsed and CW T-jump setups

4.6.1 Convolution of sample response with temperature profile

The relaxation response of interest in T-jump and T-drop spectroscopy is typically assumed to occur at a fixed final temperature where heating or cooling is instantaneous.[40] This is often a reasonable assumption, but it becomes problematic when the time scale of the temperature rise or drop overlaps with the kinetics of interest. In our pulsed T-jump experiment, processes on

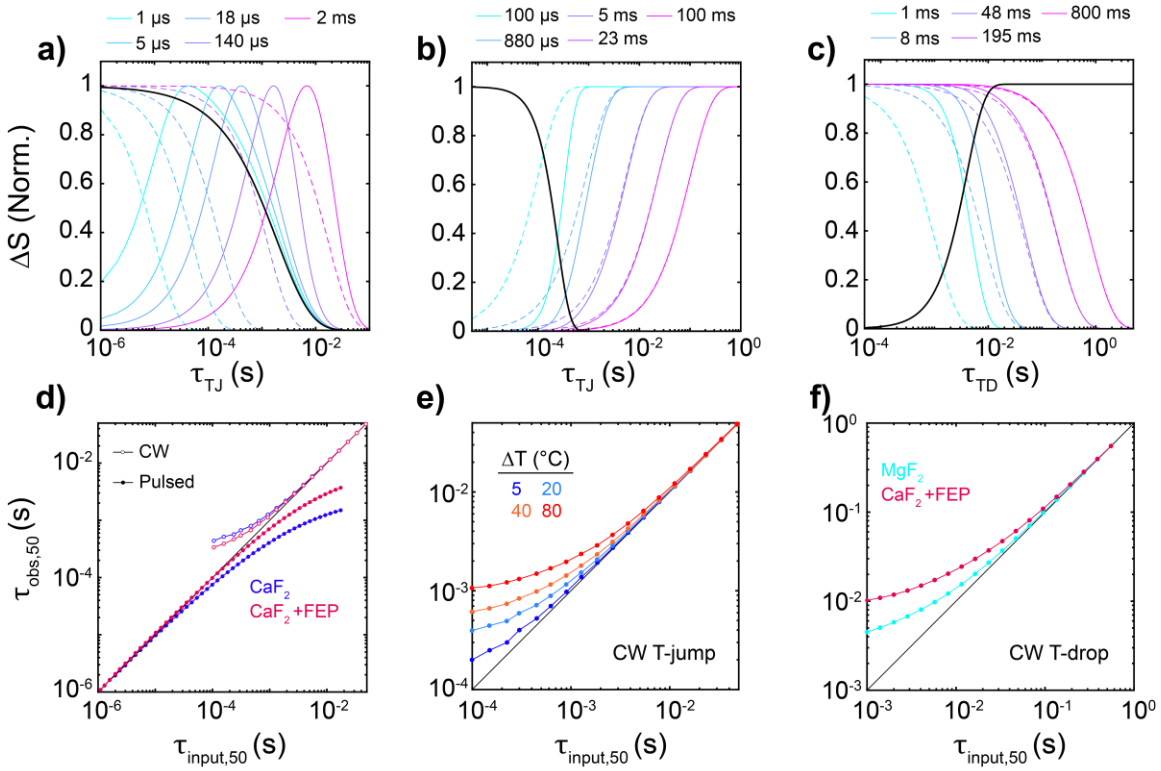

Figure 4.20 Convolution of sample cooling and heating profiles with model kinetic responses. **(a)** Model sample response (R, dashed lines), temperature profile (black dashed line), and observed signal from convolution (colored solid lines) for pulsed T-jump with CaF_2 windows and 50 µm path length. Each color is for a different model sample time constant. Similar plots are shown for **(b)** CW T-jump with $\Delta T = 15$ °C and CaF_2+FEP windows and **(c)** CW T-drop with MgF_2 windows. **(d)** Observed time constant ($\tau_{obs,50}$) from convolution as a function of input model response time constant ($\tau_{input,50}$) for pulsed T-jump (solid circles) and CW T-jump (open circles) with CaF_2 and CaF_2+FEP windows. Data for CW T-jump is at $\Delta T = 15$ °C. Time constants are defined as the time where signal has reached 50% of the rise value. Similar plots are shown for **(e)** CW T-jump with various ΔT using CaF_2+FEP windows and **(f)** CW T-drop with MgF_2 windows and amplitude modulation and CaF_2+FEP windows without amplitude modulation.

picosecond timescales are completely unresolved and those on 5 – 20 ns timescales are likely distorted. At the other end of the temperature profile, cooling back to T_i occurs as nucleic acids dissociate or unfold on 0.1 – 10 ms timescales. The slow heating rise time of CW T-jump causes full or partial temporal overlap between the heating and sample kinetics, and this is even more of

a concern for extracting kinetics from the T-drop response. If the sample kinetics fully occur within the heating or cooling timescale, then the kinetics or dynamics of interest cannot be extracted. However, minor distortions to the relaxation response arising from partial overlap may be corrected.

The timescales and dynamics of nucleic acid reactions are typically temperature dependent, yet we will start by treating the system as linear and time-invariant (LTI), which assumes its kinetics are independent of temperature. While an LTI system is an oversimplification for a real nucleic acid, it is a useful starting point because its observed time-dependent behavior (O) can be simply described as the convolution between the true system response (S) and the instrument response function (IRF).

$$O(\tau) = S(\tau) * IRF(\tau) = \int_0^\infty IRF(t)\, S(\tau - t)\, dt \tag{4.29}$$

The IRF is the solvent temperature profile for T-jump and T-drop spectroscopy. Figure 4.20a shows the convolution (colored solid lines) between the pulsed T-jump solvent cooling profile (solid black line) with an input single exponential response from the system (colored dashed lines). The observed signal change rises with the reaction timescales and drops back to zero due to cooling. When the reaction time constant is much shorter than the cooling time, the observed reaction kinetics are undistorted and follow the cooling profile at late times. As the reaction time constant overlaps with cooling, the observed reaction time appears faster than its true value. The relationship between true ($\tau_{input,50}$) and observed ($\tau_{obs,50}$) time constants are shown in Fig. 4.20d using temperature profiles measured for D_2O with a 50 μm path length between bare CaF_2 or CaF_2-

FEP windows (Fig. 3.4). With bare CaF_2, $\tau_{obs,50}$ significantly deviates from $\tau_{input,50}$ for $\tau_{input,50}$ > 100 μs as observed previously.[8] The 4-fold slowdown of cooling with CaF_2+FEP windows extends the agreement between $\tau_{obs,50}$ and $\tau_{input,50}$ by a similar magnitude.

We strive to achieve significant overlap between the relaxation timescales accessible in pulsed T-jump and CW T-jump such that there is no gap in the millisecond regime. Figure 4.20d compares the deviation in $\tau_{obs,50}$ from $\tau_{input,50}$ with that from CW T-jump with optimized heating times. The temperature profile for CW T-jump and the relationship between $\tau_{obs,50}$ from $\tau_{input,50}$ depends on ΔT (Figs. 4.14 and 4.20e), and we compare with ΔT = 15 °C since that is a common T-jump size in our pulsed experiments. The solvent heating profile is flipped to a decay function prior to being convoluted with the sample response to obtain a single component rise. When $\tau_{input,50} \leq 1$ ms, the observed response is delayed relative to real reaction time and the two progressively overlap as $\tau_{input,50}$ increases. For a given input time constant, the observed response is primarily distorted at the beginning of its rise and achieves better overlap with the input profile at longer τ_{TJ}. CW and pulsed T-jump trends with CaF_2+FEP windows do not meet on the diagonal but instead show identical deviation (on a log scale) in $\tau_{input,50}$ from $\tau_{obs,50}$ at $\tau_{input,50}$ = 0.62 ms, suggesting that there may be a gap from $\tau \sim 0.5 - 1$ ms that cannot be measured accurately with either setup.

Cooling is more than an order of magnitude slower than CW heating, leading to great risk for distortions in T-drop kinetics. Figure 4.20f shows convolution calculations with the slowest (CaF_2+FEP without optimization) and fastest (MgF_2 with optimization) cooling for our T-drop

experiments. Even with the fastest cooling, significant deviation in T-drop kinetics is shown for $\tau_{input,50} \leq 20$ ms, limiting the applicability for T-drop measurements.

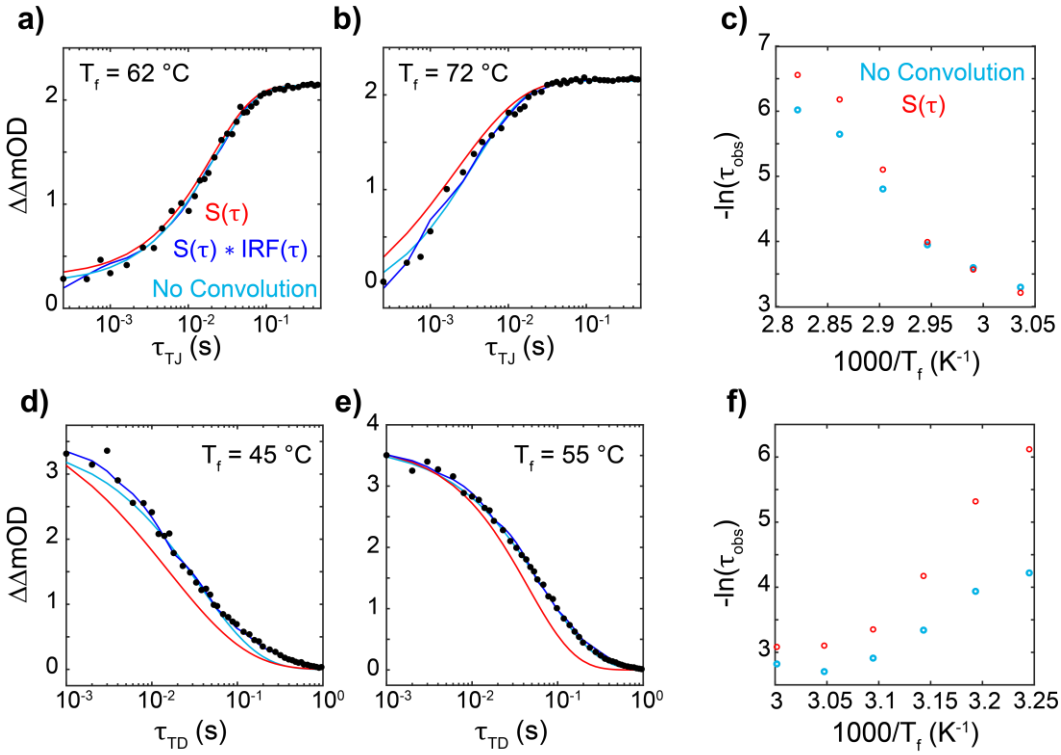

Figure 4.21 Application of convolution with temperature profile to CW T-jump and T-drop data. t-PP T-jump time traces of C3T3 at 1656 cm^{-1} for T-jumps from **(a)** 45 to 65 °C and **(b)** 55 to 75 °C acquired using bare CaF_2 windows. The t-PP data (black circles) are fit to a sum of exponentials (light blue line) and then convoluted with the solvent heating profile (IRF, dark blue line). The sample response (S, red line) extracted from the convolution is also shown. **(c)** Observed T-jump relaxation rate ($1/\tau_{obs}$) as a function of T_f extracted from fitting t-PP traces with (red) and without (light blue) convolution with the solvent heating profile. **(d-f)** Similar plots for T-drop measurements.

Convolution with an IRF may be incorporated into fitting of the T-jump and T-drop data to correct for minor distortions, and Fig. 4.21 shows an example application toward C3T3 at pH* 4.5. T-jump relaxation timescales at low T_i are essentially unaffected by the heating profile but data at higher temperatures show possible distortion. Assuming an LTI system, the C3T3 T-jump

data was fit to an exponential rise convoluted with the solvent heating profile (performed numerically in MATLAB) to account for the distortion. Observed time constants with and without the convolution are shown as a function of temperature. At the three lowest temperatures, each method gives the same time constant, suggesting the heating profile does not artificially delay the observed kinetics. At higher temperatures, the convolution gives a faster time constant due to greater overlap between the sample response and heating profile. The same method may be applied to extract T-drop kinetics, and Fig. 4.21d-f shows an example for C3T3 using bare CaF_2 windows without laser power modulation. This is potentially the most distortive scenario because the cooling profile is multi-exponential with a large amplitude slow component. At low temperatures when the folding is fastest, the observed T-drop response cannot be fit to a single-exponential decay because of this slow cooling component. Convoluting with the full cooling profile leads to a high quality fit over all times and extracts a significantly faster time constant for folding than without the convolution. The slow cooling component leads to less distortion at higher temperatures with slower folding, but fitting with a convolution still extracts a faster folding decay.

4.6.2 Influence of heating and cooling profiles on two-state kinetics

Treating the observed T-jump and T-drop signals as a convolution between the solvent profile and true sample response partially accounts for distortions to the signal. In reality, the system response strongly depends on temperature and the LTI assumption must be relaxed to accurately describe distortions to the observed signal. When the LTI assumption is relaxed, the folding (k_f), unfolding (k_u), and observed rate constants ($1/\tau_{obs}$) are time-dependent along the T-jump and T-drop solvent profiles.

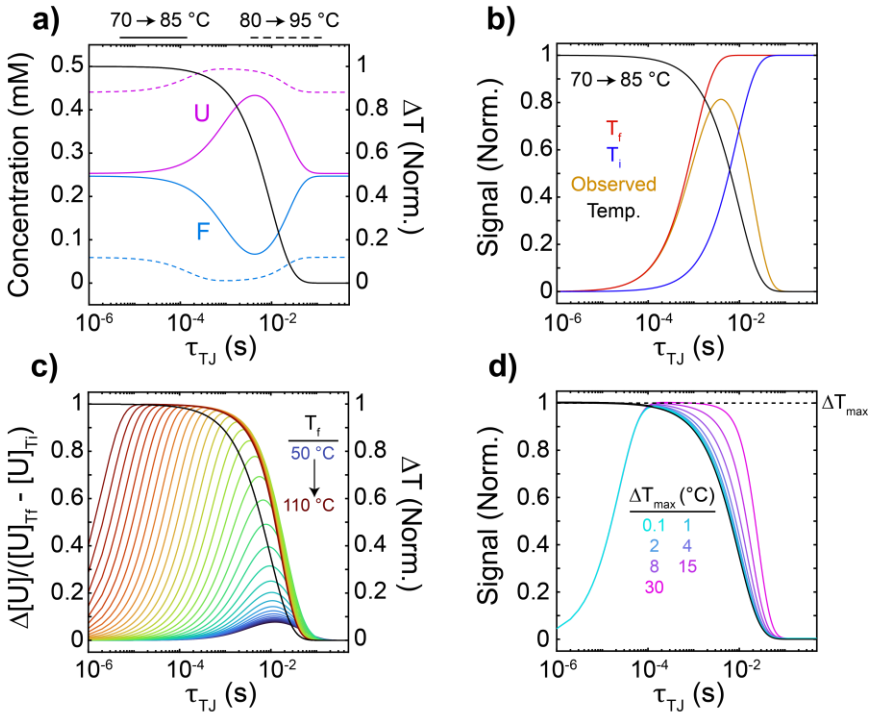

Figure 4.22 Simulation of pulsed T-jump signals for two-state kinetics. (a) Concentration profiles of unfolded (U, magenta) and folded (F, blue) species as a function of τ_{TJ} for T-jumps from 70 to 85 °C (solid line) and 80 to 95 °C (dashed line) using the temperature profile measured with CaF_2+FEP windows (solid black line). **(b)** Observed concentration profile of U (orange) normalized by the equilibrium concentration change between T_i and T_f, $\Delta[U]/([U]_{Tf} - [U]_{Ti})$. Expected sample responses at a constant temperature of T_i (blue) and T_f (red) are shown as well. **(c)** Observed concentration profiles as a function of T_f using the kinetic parameters described in the text. **(d)** Observed signal for T_f = 100 °C with variable ΔT.

T-jump and T-drop responses are simulated for a model two-state unfolded (U)/folded (F) system using the experimental heating and cooling temperature profiles.

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} F \tag{4.30}$$

The observed T-jump and T-drop rate can be related to k_f and k_u.[40]

$$1/\tau_{obs} = k_f + k_u \tag{4.31}$$

Both k_f and k_u rate constants are described by a Kramers-like equation in the high friction limit.[41]

$$k_{u/f}(T) = \frac{k_B T}{h}\frac{\eta(T_{ref})}{\eta(T)}\exp\left(\frac{\Delta S^{\ddagger}_{u/f}}{R}\right)\exp\left(-\frac{\Delta H^{\ddagger}_{u/f}}{RT}\right) \tag{4.32}$$

In eq. 4.32, k_B is the Boltzmann constant, h is the Planck constant, R is the ideal gas constant, $\eta(T)$ is the temperature-dependent viscosity of D_2O and $\eta(T_{ref})$ is the value at a reference temperature of 37 °C,[42] $\Delta S^{\ddagger}_{u/f}$ is the activation entropy for unfolding or folding, and $\Delta H^{\ddagger}_{u/f}$ is the activation enthalpy for unfolding or folding. Temperature-dependence in $\Delta S^{\ddagger}_{u/f}$ and $\Delta H^{\ddagger}_{u/f}$ is neglected. Values of $\Delta H^{\ddagger}_{u}$ = 183 kJ/mol, $\Delta H^{\ddagger}_{f}$ = -16 kJ/mol, $\Delta S^{\ddagger}_{u}$ = 254 J/molK, and $\Delta S^{\ddagger}_{f}$ = -327 J/molK were determined from temperature-dependent T-jump data of C3T3 at pH* 4.5 and are used for the calculations. Numerous biomolecular binding and folding systems exhibit similar barrier heights and temperature-dependence, so these results are generally applicable to many systems studied in our group.[14, 43-44] The concentrations of U and F then propagate in time according to eq. 4.33.

$$\left[U(\tau_{i+1})\right] = dt\left(k_u(T)\left[F(\tau_i)\right] - k_f(T)\left[U(\tau_i)\right]\right) + \left[U(\tau_i)\right] \tag{4.33}$$

$$\text{where } \left[F(\tau_i)\right] = c_{tot} - \left[U(\tau_i)\right]$$

A time step (dt) of 1 μs was used for all simulations. The total sample concentration (c_{tot}) was set to 0.5 mM. At each time point, k_u and k_f are re-evaluated based on the solvent temperature. Starting concentrations of U and F are determined from the ratio of k_u to k_f at T_i.

Simulations of pulsed T-jump responses are shown in Fig. 4.21 using the measured temperature profile with a sample path length of 50 μm and CaF_2+FEP windows and the kinetic parameters listed above. Similar calculations were performed previously with bare CaF_2 windows.[8] After an instantaneous switch from T_i to T_f, an increase and decrease of [U] and [F], respectively, are observed before re-folding due to cooling back to T_i. Cooling of the solution on a 1 – 10 ms timescale causes the sample response to be cutoff when τ_{obs} is in a similar time window, causing the amplitude of the observed signal to diminish as it mixes with the temperature decay. Figure 4.23a compares $\tau_{obs,50}$ with the predicted time constant at T_f ($\tau_{Tf,50}$), and $\tau_{obs,50}$ artificially shortens when mixed with the temperature decay but to a lesser degree than in $\tau_{obs,50}$ vs. $\tau_{input,50}$ from the convolution. It is important to note that the time-dependent populations of U and F do not decay with the ΔT profile but instead lag behind. This arises from a nonlinear relationship between the effective time-dependent folding constant ($K_f(\tau_{TJ}) = [F(\tau_{TJ})]/[U(\tau_{TJ})]$) and $\Delta T(\tau_{TJ})$ that results from the large enthalpy difference between the folded and unfolded state of the model system ($\Delta H_f^\circ = 199$ kJ/mol) and the significant ΔT of 15 °C (See Chapter 6 of Ref 8). Reducing ΔT leads to better matching between [U] and ΔT (Fig. 4.21d) and the same trend is observed when reducing ΔH_f° (not shown).

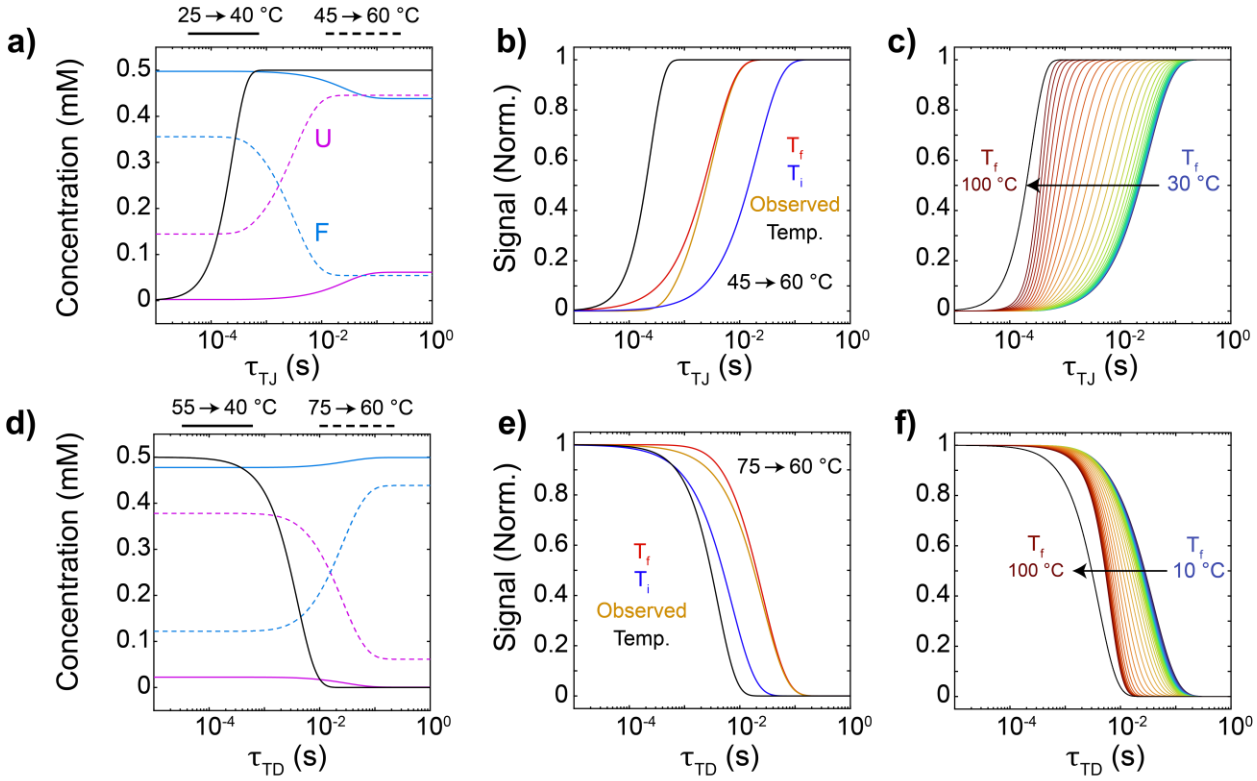

Figure 4.23 Simulation of CW T-jump and T-drop signals for two-state kinetics. (a) Concentration profiles of species U (magenta) and F (blue) as a function of τ_{TJ} for T-jumps from 25 to 40 °C (solid line) and 45 to 60 °C (dashed line) using the CW heating profile measured with CaF_2+FEP windows (solid black line). **(b)** Normalized observed concentration profile U (orange). Expected sample responses at a constant temperature of T_i (blue) and T_f (red) are shown as well. **(c)** Normalized concentration profile of U for T-jumps of ΔT = 15 °C ranging from T_f = 30 to 100 °C. **(d-f)** Similar plots of T-drops using the CW cooling profile measured with MgF_2 and optimization.

CW T-jump simulations using CaF_2+FEP windows, ΔT = 15 °C, and the same kinetic parameters as for pulsed T-jump are outlined in Fig. 4.23. Concentration profiles for U and F are shown for T-jumps from 25 to 40 °C (solid lines) and 45 to 60 °C (dashed lines). As for the convolution results, the mixing of the sample response and heating profile leads to a delay in the observed kinetics and a sharpening of the concentration profile from an exponential to compressed-

exponential rise. The deviation of $\tau_{obs,50}$ from $\tau_{Tf,50}$ is greater than from $\tau_{input,50}$ (convolution) by ~$10^{0.08}$ across the 0.1 – 1 ms window. The observed kinetics from T-drop simulations are similarly delayed when mixed with the cooling of the sample. However, the deviation of $\tau_{obs,50}$ from $\tau_{Tf,50}$ is less than from $\tau_{input,50}$ under the specified conditions. The trends for both T-jump and T-drop show weak sensitivity to minor changes in the temperature-dependence of folding and unfolding barriers (Fig. 4.24c,d).

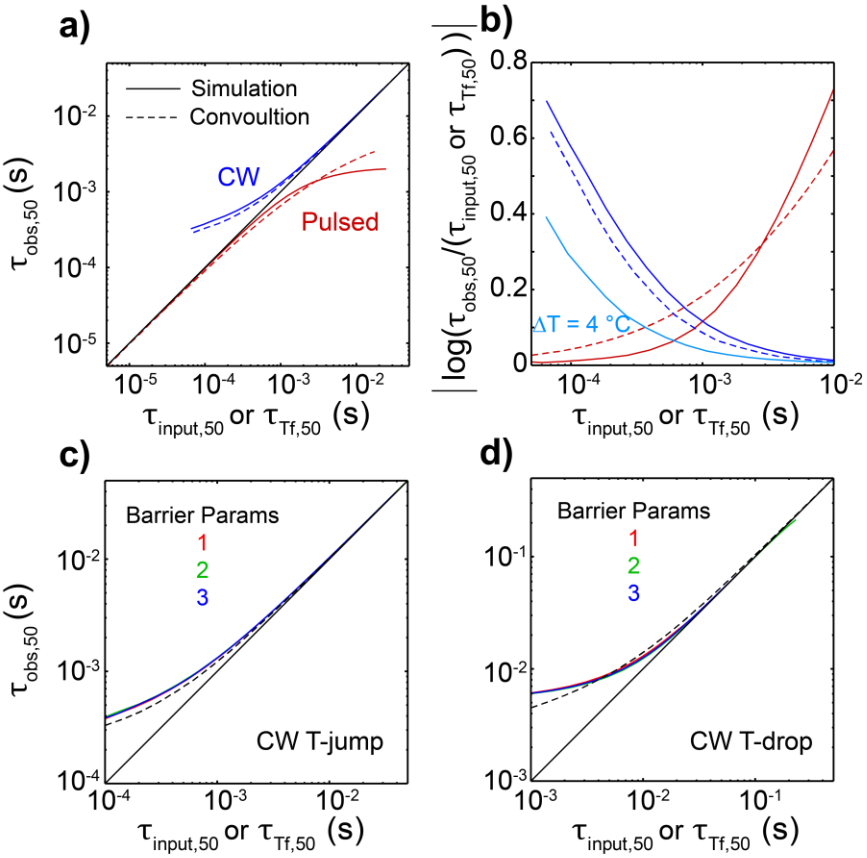

Figure 4.24 Observed vs. input rates from simulation of two-state kinetics. (a) Comparison of $\tau_{obs,50}$ vs. $\tau_{input,50}$ from convolution (dashed lines) and $\tau_{obs,50}$ vs. $\tau_{Tf,50}$ from temperature-dependent two-state kinetic simulations (solid lines) for pulsed and CW T-jump (ΔT = 15 °C) using temperature profiles measured with CaF_2+FEP windows and a 50 μm sample path length. **(b)** Difference between observed and input rate for convolution, $\left|log\left(\tau_{obs,50}/\tau_{input,50}\right)\right|$, and

Figure 4.24 Observed vs. input rates from simulation of two-state kinetics, continued
two-state kinetic simulations, $\left|log\left(\tau_{obs,50}/\tau_{Tf,50}\right)\right|$. **(c)** $\tau_{obs,50}$ vs. $\tau_{Tf,50}$ from CW T-jump simulations using different input barriers: (1, red) $\Delta H_u^{\ddagger}$ = 183 kJ/mol, $\Delta H_f^{\ddagger}$ = -16 kJ/mol, $\Delta S_u^{\ddagger}$ = 254 J/molK, and $\Delta S_f^{\ddagger}$ = -327 J/molK, (2, green) $\Delta H_u^{\ddagger}$= 213 kJ/mol, $\Delta H_f^{\ddagger}$= 14 kJ/mol, $\Delta S_u^{\ddagger}$= 342 J/molK, and $\Delta S_f^{\ddagger}$= -239 J/molK, (3, blue) $\Delta H_u^{\ddagger}$= 198 kJ/mol, $\Delta H_f^{\ddagger}$= -1.3 kJ/mol, $\Delta S_u^{\ddagger}$= 298 J/molK, and $\Delta S_f^{\ddagger}$= -283 J/mol. The convolution result is shown as a black dashed.

Comparison of the $\tau_{obs,50}$ vs. $\tau_{Tf,50}$ deviation for CW and pulsed T-jumps predicts that our experimental methods are least accurate (with ΔT = 15 °C) for measuring processes with τ_{50} ~ 1 ms with a deviation of $10^{0.1}$ from the true value, and a similar degree of error is predicted over the ~0.5 – 5 ms range (Fig. 4.24b). If the kinetics of interest are insensitive to ΔT_0, which is typical for high barrier processes, then CW T-jump may be performed with identical T_f and smaller ΔT to obtain more accurate time constants in this time window. The observed response may also deviate from single-exponential kinetics in this time window, particularly with the CW T-jump approach, but this is not directly reported through the time constants alone.

4.7 Combining kinetics from T-jump and T-drop spectroscopy

Kinetic studies of nucleic acids or proteins are often limited to one experiment that is primarily sensitive to either folding/association or unfolding/dissociation, but there is much benefit from combining data with sensitivity to each direction. T-drop or mixing experiments do not offer as fast of time resolution as pulsed T-jump, but they can be combined with CW T-jump experiments to obtain direct sensitivity to folding/association and unfolding/dissociation kinetics. Such approaches may be used to obtain more accurate kinetic parameters or to identify deviations from standard two-state kinetics.[43, 45]

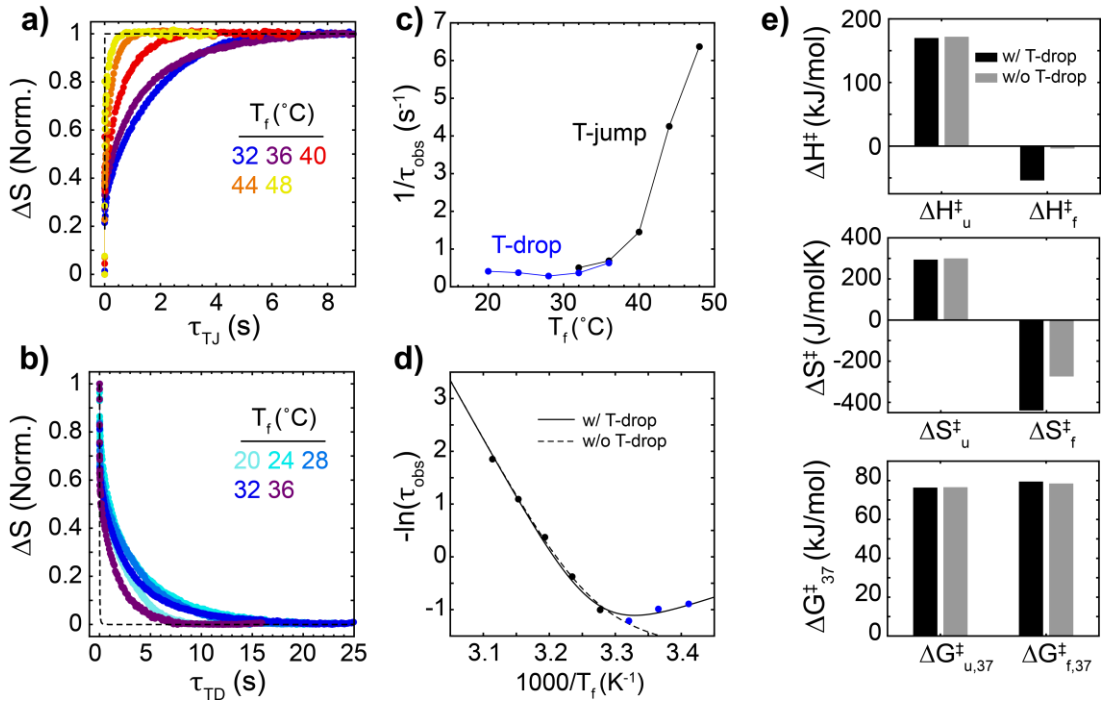

Figure 4.25 Improved characterization of two-state kinetics by combining T-jump and T-drop spectroscopy. (a) CW T-jump signal change of C3T3 at 1656 cm^{-1} for various T_f with ΔT = 12 °C. Oligonucleotide was prepared at a concentration of 0.4 mM in pH* 6.3 80 mM sodium phosphate buffer. Dashed line corresponds to the solvent temperature profile. **(b)** T-drop profiles acquired in the same measurement with T-jump data. **(c)** Observed rates from T-jump (black) and T-drop (blue). **(d)** Global fitting of T-jump and T-drop rates with the PP thermal melting curve (not shown) to a two-state model (eqs. 4.30-4.32 and 4.34). Dashed line corresponds to a fit omitting the T-drop rates **(e)** Comparison of barriers $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, and $\Delta G^{\ddagger}$ for folding and unfolding extracted with and without the T-drop observed rates.

Figure 4.25 illustrates how combining T-jump and T-drop results enables the extraction of more accurate two-state kinetic parameters for i-motif folding and unfolding. For systems with relatively sharp thermal melting transitions, T-jump measurements occur at T_f values where $1/\tau_{obs}$ is dominated by k_u. The forward and backward enthalpic barriers are usually very different for folding and binding transitions, leading to stark differences in the temperature-dependence of k_u and k_f. In T-jump data, this usually appears as a slight deviation from exponential behavior in

$1/\tau_{obs}$ at the lowest T_f values due to similar weighting of k_u and k_f, which is shown for C3T3 at pH* 6.3 in Fig. 4.25d. If applying a two-state model, $1/\tau_{obs}$ must be broken up into k_u and k_f, where the temperature-dependent unfolding equilibrium constant K_u is provided from the thermal melting curve or calorimetry.

$$k_f = \tau_{obs}^{-1}\left(1+K_u\right)^{-1} \text{, where} \quad K_u = \frac{k_u}{k_f} \tag{4.34}$$

Fitting the C3T3 T-jump data alone leads to a large enthalpic barrier ($\Delta H_u^{\ddagger}$ ~ 170 kJ/mol) and favorable entropy ($\Delta S_u^{\ddagger}$ ~ 300 J/molK) for unfolding and a negligible enthalpy ($\Delta H_f^{\ddagger}$ ~ 0) and large entropic barrier ($\Delta S_f^{\ddagger}$ ~ -280 J/molK) for folding. However, expanding the temperature range of $1/\tau_{obs}$ using T-drops reveals a much larger shift in slope at T_f< 30 °C. Fitting the T-jump and T-drop data together gives identical values of $\Delta H_u^{\ddagger}$ and $\Delta S_u^{\ddagger}$ to using T-jump only, but $\Delta H_f^{\ddagger}$ and $\Delta S_f^{\ddagger}$ are decreased by ~50 kJ/mol and ~140 J/molK, respectively. It is also important to note that $1/\tau_{obs}$ is essentially identical for T-jump and T-drop experiments performed at the same T_f, which is a requirement of two-state kinetics.

In addition to refining two-state kinetic analysis, T-jump and T-drop data may be combined to reveal deviations from two-state kinetics. For example, differences in $1/\tau_{obs}$ obtained from T-jump and T-drop measurements at a consistent T_f indicate a deviation from standard two-state kinetics. Additionally, different relaxation methods may be more sensitive to certain states along the folding or binding free-energy landscape.[43, 46]

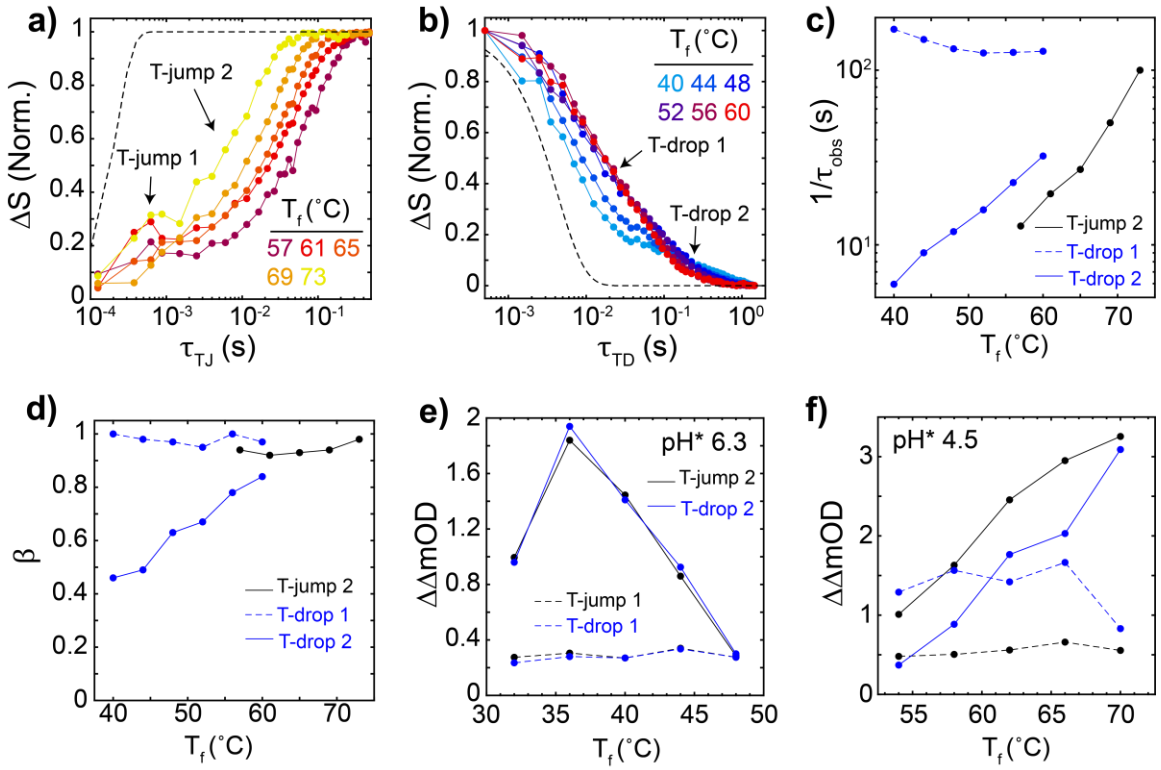

Figure 4.26 Discrepancy between T-jump and T-drop indicates multi-state kinetics. (a) CW T-jump signal change of C3T3 at 1656 cm^{-1} for various T_f with ΔT = 12 °C. Oligonucleotide was prepared at a concentration of 0.4 mM in pH* 4.5 80 mM sodium phosphate buffer. Dashed line corresponds to the solvent temperature profile. **(b)** T-drop profiles acquired in separate measurements with MgF_2 windows. **(c)** Observed rates and **(d)** stretch parameters (β) from fits of T-jump (black) and T-drop (blue) data to a sum of stretched exponential decays. **(f)** Signal change associated with 1st (dashed line) and 2nd (solid line) kinetic component from T-jump (black) and T-drop (blue) data of C3T3 at pH* 6.3. **(e)** Same comparison for data at pH* 4.5 where T-jump and T-drop was measured together using MgF_2 windows.

Figure 4.26 shows that C3T3 exhibits distinct T-jump and T-drop kinetics at pH* 4.5. T-jump traces contain a small signal change within the heating time followed by a single exponential unfolding component on 1 – 100 ms timescales that speeds up exponentially with temperature. The T-jump data alone can be well described by two-state kinetics. In contrast, T-drop measurements reveals clear multi-component folding at low T_f with time windows of 1 – 50 ms

and 0.1 to >1 s. The first component is highly mixed with the sample cooling, and therefore its time constant is likely overestimated. The first component slows and the second component speeds up as T_f increases such that they overlap at higher T_f. Additionally, the signal change associated with the first T-drop component is much greater than in the heating component from T-jump (Fig. 4.26e), further indicating that some degree of folding occurs within the first T-drop component. For the two-state kinetics at pH* 6.3, the heating component from T-jump and cooling component from T-drop have identical amplitudes. The T-drop data at pH* 4.5 indicates at least a 3-state folding landscape that would have otherwise been treated as two-state from T-jump data alone.

4.8 Acknowledgements

I thank Nicholas Lewis and Paul Sanstead for their assistance and feedback in designing and building the initial version of the CW T-jump spectrometer, and I thank Yumin Lee for helping to implement improvements over the past few years. I also thank John Hack and Melissa Bodine for feedback on this chapter.

4.9 References

1. Möller, K. O.; De Maeyer, L. C., Temperature-jump apparatus with infrared detection for relaxation studies in thin samples. *Rev. Sci. Instrum.* **1982,** *53,* 1596-1601.

2. Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B., Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **1996,** *35,* 691-697.

3. Ma, H.; Ervin, J.; Gruebele, M., Single-sweep detection of relaxation kinetics by submicrosecond midinfrared spectroscopy. *Rev. Sci. Instrum.* **2004,** *75,* 486-491.

4. Hauser, K.; Krejtschi, C.; Huang, R.; Wu, L.; Keiderling, T. A., Site-specific relaxation kinetics of a tryptophan zipper hairpin peptide using temperature-jump IR spectroscopy and isotopic labeling. *J. Am. Chem. Soc.* **2008,** *130,* 2984-2992.

5. Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A., Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* **2007,** *78*, 063101.

6. Chung, H. S.; Ganim, Z.; Jones, K. C.; Tokmakoff, A., Transient 2D IR spectroscopy of ubiquitin unfolding dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104*, 14237-14242.

7. Chung, H. S. Thermal unfolding dynamics of proteins probed by nonlinear infrared spectroscopy. Massachusetts Institute of Technology, 2007.

8. Jones, K. C. Temperature-jump 2D IR spectroscopy to study protein conformational dynamics. Massachusetts Institute of Technology, 2012.

9. Stevenson, P. Membrane and Membrane Protein Dynamics Studied with Time-Resolved Infrared Spectroscopy. Massachusetts Institute of Technology, Department of Chemistry, 2017.

10. Jones, K. C.; Ganim, Z.; Peng, C. S.; Tokmakoff, A., Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* **2012,** *29*, 118-129.

11. Jones, K. C.; Ganim, Z.; Tokmakoff, A., Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 14060-14066.

12. Lepetit, L.; Chériaux, G.; Joffre, M., Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **1995,** *12*, 2467-2474.

13. Sanstead, P. J. C., Investigation of DNA Dehybridization through Steady-State and Transient Temperature-Jump Nonlinear Infrared Spectroscopy. The University of Chicago: 2018.

14. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

15. Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B., Protein folding and unfolding on a complex energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97*, 2527-2532.

16. Popp, A.; Scheerer, D.; Heck, B.; Hauser, K., Biomolecular dynamics studied with IR-spectroscopy using quantum cascade lasers combined with nanosecond perturbation techniques. *Spectrochim. Acta A* **2017,** *181*, 192-199.

17. Stevenson, P.; Tokmakoff, A., Time-resolved measurements of an ion channel conformational change driven by a membrane phase transition. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 10840-10845.

18. Peng, C. S.; Baiz, C. R.; Tokmakoff, A., Direct observation of ground-state lactam–lactim tautomerization using temperature-jump transient 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 9243-9248.

19. Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A., Oxidized derivatives of 5-methylcytosine alter the stability and dehybridization dynamics of duplex DNA. *J. Phys. Chem. B* **2020,** *124*, 1160-1174.

20. Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J., i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018,** *46*, 8038-8056.

21. Gray, R. D.; Trent, J. O.; Chaires, J. B., Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.* **2014,** *426*, 1629-1650.

22. Jeon, J.; Wilson, C. B.; Yau, W.-M.; Thurber, K. R.; Tycko, R., Time-resolved solid state NMR of biomolecular processes with millisecond time resolution. *J. Magn. Reson.* **2022,** *342*, 107285.

23. DeFlores, L. P.; Nicodemus, R. A.; Tokmakoff, A., Two-dimensional Fourier transform spectroscopy in the pump-probe geometry. *Opt. Lett.* **2007,** *32*, 2966-2968.

24. Helbing, J.; Hamm, P., Compact implementation of Fourier transform two-dimensional IR spectroscopy without phase ambiguity. *J. Opt. Soc. Am. B* **2011,** *28*, 171-178.

25. Mertz, L., Auxiliary computation for Fourier spectrometry. *Infrared Phys.* **1967,** *7*, 17-23.

26. Laouer, K.; Schmid, M.; Wien, F.; Changenet, P.; Hache, F., Folding dynamics of DNA G-quadruplexes probed by millisecond temperature jump circular dichroism. *J. Phys. Chem. B* **2021,** *125*, 8088-8098.

27. Polinkovsky, M. E.; Gambin, Y.; Banerjee, P. R.; Erickstad, M. J.; Groisman, A.; Deniz, A. A., Ultrafast cooling reveals microsecond-scale biomolecular dynamics. *Nat. Commun.* **2014,** *5*, 1-7.

28. Ashwood, B.; Lewis, N. H.; Sanstead, P. J.; Tokmakoff, A., Temperature-Jump 2D IR Spectroscopy with Intensity-Modulated CW Optical Heating. *J. Phys. Chem. B* **2020,** *124*, 8665-8677.

29. Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Minnes, L.; Towrie, M.; Parker, A. W.; Hunt, N. T., Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 6188-6199.

30. Scheerer, D.; Chi, H.; McElheny, D.; Keiderling, T. A.; Hauser, K., Isotopically site-selected dynamics of a three-stranded β-sheet peptide detected with temperature-jump infrared-spectroscopy. *J. Phys. Chem. B* **2018,** *122*, 10445-10454.

31. Dyer, R. B.; Gai, F.; Woodruff, W. H.; Gilmanshin, R.; Callender, R. H., Infrared studies of fast events in protein folding. *Acc. Chem. Res.* **1998,** *31*, 709-716.

32. Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A., Oxidized Derivatives of 5-Methylcytosine Alter the Stability and Dehybridization Dynamics of Duplex DNA. *J. Phys. Chem. B* **2020,** *124*, 1160-1174.

33. Ramos, P. M.; Ruisánchez, I., Noise and background removal in Raman spectra of ancient pigments using wavelet transform. *J. Raman. Spectrosc.* **2005,** *36*, 848-856.

34. Galloway, C.; Le Ru, E.; Etchegoin, P., An iterative algorithm for background removal in spectroscopy by wavelet transforms. *Appl. Spectrosc.* **2009,** *63*, 1370-1376.

35. Mallat, S. G., A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989,** *11*, 674-693.

36. Daubechies, I., *Ten lectures on wavelets*. Siam: 1992; Vol. 61.

37. Jones, K. C.; Ganim, Z.; Tokmakoff, A., Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 14060-14066.

38. Donaldson, P. M.; Greetham, G. M.; Middleton, C. T.; Luther, B. M.; Zanni, M. T.; Hamm, P.; Krummel, A. T., Breaking Barriers in Ultrafast Spectroscopy and Imaging Using 100 kHz Amplified Yb-Laser Systems. *Acc. Chem. Res.* **2023,** *56*, 2062-2071.

39. Hamm, P., Transient 2D IR spectroscopy from micro-to milliseconds. *J. Chem. Phys.* **2021,** *154*, 104201.

40. Bernasconi, C., *Relaxation kinetics*. Academic Press: 1976.

41. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7*, 284-304.

42. Cho, C.; Urquidi, J.; Singh, S.; Robinson, G. W., Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **1999,** *103*, 1991-1994.

43. Narayanan, R.; Zhu, L.; Velmurugu, Y.; Roca, J.; Kuznetsov, S. V.; Prehna, G.; Lapidus, L. J.; Ansari, A., Exploring the energy landscape of nucleic acid hairpins using laser temperature-jump and microfluidic mixing. *J. Am. Chem. Soc.* **2012,** *134*, 18952-18963.

44. Goldberg, J. M.; Baldwin, R. L., Kinetic mechanism of a partial folding reaction. 2. Nature of the transition state. *Biochem.* **1998,** *37*, 2556-2563.

45. Chen, C.; Wang, W.; Wang, Z.; Wei, F.; Zhao, X. S., Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* **2007,** *35*, 2875-2884.

46. Wirth, A. J.; Liu, Y.; Prigozhin, M. B.; Schulten, K.; Gruebele, M., Comparing fast pressure jump and temperature jump protein folding experiments and simulations. *J. Am. Chem. Soc.* **2015,** *137*, 7152-7159.

Chapter 5

DNA hybridization from coarse-grained simulations

5.1 Introduction

5.1.1 Molecular dynamics for investigating nucleic acid structural transitions

A primary goal of the work in this thesis is to expand our insight into nucleic acid base-pairing and conformational changes through interrogation of the underlying dynamics. Although our T-jump experiments provide a means to measure kinetics over a wide range of timescales with the structural sensitivity of IR spectroscopy (described in Chapter 4), they cannot directly probe the dynamics of transitions between single-strand and duplex ensembles (i.e. transition paths). As briefly discussed in Chapters 1 and 2, only single-molecule experiments can in principle measure transition paths. However, even when transition paths are measured for biomolecular transitions, the information is limited to an average transition-path duration, $\langle \tau_{TP} \rangle$, and/or must be interpreted in terms of coarse experimental observables such as FRET efficiency or molecular extension.[1-3] Although transition paths of DNA hairpin folding and base-pair zipping have been reported,[4-8] transition paths for hybridization between separate oligonucleotides have only been measured with slowed motion near a graphene surface yet not in free solution or cellular environments.[9]

Given the current experimental bottlenecks for measuring nucleic acid structural dynamics, we utilize molecular dynamics (MD) simulations throughout this thesis in tandem with our experiments. As introduced in Chapter 1, there are also difficulties with accessing accurate

hybridization dynamics from MD simulations. In addition to the complications of sufficiently sampling rare transitions, the inaccuracies of all-atom force fields for nucleic acids are potentially the most limiting factor for simulating hybridization.[10] Recent advances have produced force fields that accurately model canonical DNA duplex structure over many microsecond simulations.[11-13] Since these force fields are parameterized for B-DNA, they still struggle to model most non-duplex structures accurately.[14-17] As a result, usage of all-atom MD simulations for studying hybridization or other nucleic acid conformational transitions has been limited.[18-20]

While all-atom force fields continue to progress, coarse-grained models of DNA have been developed and applied to study hybridization dynamics.[21-27] These force fields typically condense nucleotides to just a few sites and lack explicit solvent molecules, enabling enables more efficient sampling of slow structural transitions like folding and assembly. Since their development, numerous types of structural transitions have been investigated with DNA coarse-grained models such as hybridization,[21, 24, 27-28] competition between hybridization and hairpin folding,[29] toehold strand displacement,[30] and even assembly of DNA origami structures.[31]

Multiple investigations of duplex hybridization with coarse-grained MD simulations have revealed diverse encounter and nucleation behavior across oligonucleotides of varying sequence,[21, 24, 27] but there is still a poor understanding of many dynamic aspects of hybridization. Potentially most unclear are the structural and mechanistic differences between successful and unsuccessful hybridization events. Do failed events tend to start with different base pairs or strand orientations than successful events? Previous studies routinely identify out-of-register base pairing during hybridization,[21, 24, 27-28] and such configurations may either speed up hybridization by keeping

oligonucleotides in close proximity or act as a kinetic trap that leads to failure. A predictive understanding of how to tune hybridization success will enhance oligonucleotide sequence design in nanotechnology and biotechnology and calls for an investigation into the dynamics of successful and unsuccessful hybridization.

5.1.2 Summary of 3SPN.2 coarse-grained model

In this thesis, we use the 3-site-per-nucleotide (3SPN.2) coarse-grained model of DNA,[23] which has previously been employed to investigate mechanisms of oligonucleotide hybridization.[24, 27, 32-33] Each nucleotide is replaced with a set of three spherical interaction sites placed at the center-of-mass of the phosphate, deoxyribose, and nucleobase. The bonded potential includes contributions from chemical bonds, bond angles, and dihedrals (only between deoxyribose and phosphate sites). The non-bonded potential includes excluded volume, base-stacking, cross-stacking, base-pairing, and electrostatic interactions. Base-pairing, base-stacking, and cross-stacking interactions are anisotropic as they are controlled by the distance and angles between sites. Mismatch base-pairing interactions are not included in the model, potentially affecting the stability of out-of-register configurations. Counterion and solvation effects are treated implicitly. Phosphate sites are given an effective charge of -0.6 in accordance with Manning counterion condensation theory,[34] and electrostatic repulsion between phosphate sites is modeled via Debye-Hückel theory with a Debye screening length set by the simulation temperature and ionic strength. Solvent effects are incorporated using Langevin dynamics with an experimentally motivated per-site friction coefficient of 9.94×10^{-11} m^2/s.[35] Simulations are carried out in the NVT ensemble with a Langevin thermostat.

3SPN.2 MD simulations are utilized to map the free-energy landscape of small oligonucleotides and probe the transition paths between duplex and single-strand ensembles. In this chapter, we particularly focus on how nucleobase sequence alters the free-energy landscape and dynamics of hybridization. We and others have previously studied sequence-dependent hybridization mechanisms using 3SPN.2 MD simulations.[24, 27] The content in this chapter is complementary to these previous studies and carries a distinct focus on trying to identify key nucleobase contacts in the transition-state ensemble. We begin by assessing the properties of simple base-pairing collective variables and move to describe the sequence of base-pairing events during hybridization through analysis of transition paths.

5.2 Hybridization along base-pairing collective variables

Characterization of hybridization dynamics from MD simulations requires the application of appropriate collective variables (CVs) that describe a collection of select atoms, angles, or distances used to distinguish duplexes from single strands and the configurations in between. These CVs may be used to enhance the sampling of rare transitions and/or in the analysis of unbiased MD simulations. The CV must separate transitions between duplex and single-strand ensembles from structural motion within each basin, an essential property for distinguishing successful hybridization events from non-reactive trajectories and for identifying the ensemble of transition-state configurations during hybridization. Common CVs for bimolecular association include the center-of-mass (COM) separation between monomers, root-mean-square-deviation (RMSD) from the native structure, and the number of intermolecular contacts below a given distance threshold,[24, 36] and a large body of research is dedicated to methods that identify a small

number of CVs optimized for describing the slowest structural motions of a system.[37-40] These methods have been employed to identify metastable states and their rates of interconversion in short oligonucleotides,[27] but the dynamics of interconversion between states remained elusive. To examine the dynamics of hybridization, we need to identify CVs for direct analysis of oligonucleotide trajectories.

CVs that report on base-pair contacts are commonly applied to examine the free-energy landscape or mechanism for hybridization, but it is important to assess how well these CVs capture the relevant degrees of freedom to hybridization. Numerous CVs can distinguish duplex and single-strand states and show a single free-energy barrier in between, yet that observation alone does not mean that the dynamics of hybridization are accurately described along the coordinate. For example, some processes are not well-described by a single CV, and projection of the multi-dimensional dynamics onto a 1D reaction coordinate can introduce free-energy barriers where non-existent or suggest a low-barrier path where it's impossible to traverse. Such a projection is also likely to introduce a position-dependent diffusion coefficient along the 1D free-energy landscape as off-coordinate motions are compressed into small segments of the CV.[5, 41-42] However, describing reaction dynamics along a single CV provides the best connection to experiments because most single-molecule measurements are along one or two coordinates, and fast ensemble relaxation methods contain a small number of observables. Further, analysis of barrier crossing trajectories in multi-dimensional CV space requires thorough sampling that is potentially prohibitive even with coarse-grained models and adds extra complication to interpretation. There are many examples in which biomolecular folding dynamics appear to be

well-described with a single simple CV,[43-45] and we take a similar approach here for DNA hybridization.

5.2.1 Free-energy profile for hybridization

We particularly focus on CVs that report on in-register base pairing between strands. r_{bp} is the average base-pair separation across all in-register sites *i*, where *N* is the total number of possible base pairs (10 or 11 in this chapter). 3SPN.2 represents each nucleobase as a single spherical site centered at the nucleobase center-of-mass (COM), therefore the base-pair separations correspond to distances between the COM of each nucleobase.

$$r_{bp} = \frac{1}{N}\sum_{i=1}^{N} r_{bp,i} \tag{5.1a}$$

$1/r_{bp}$ is the average inverse base-pair separation across all in-register sites *i*, providing higher resolution to small changes in base-pair separation than r_{bp}.

$$1/r_{bp} = \frac{1}{N}\sum_{i=1}^{N} 1/r_{bp,i} \tag{5.1b}$$

n_{bp} is the number of intact in-register base pairs determined using a separation threshold of r_{cutoff} = 0.7 nm for each index.

$$n_{bp} = \frac{1}{N}\sum_{i=1}^{N} \begin{cases} 1 & r_{bp,i} \le r_{cutoff} \\ 0 & r_{bp,i} > r_{cutoff} \end{cases} \tag{5.1c}$$

Various system properties show minor sensitivity to the choice of r_{cutoff} (Appendix 6.E).

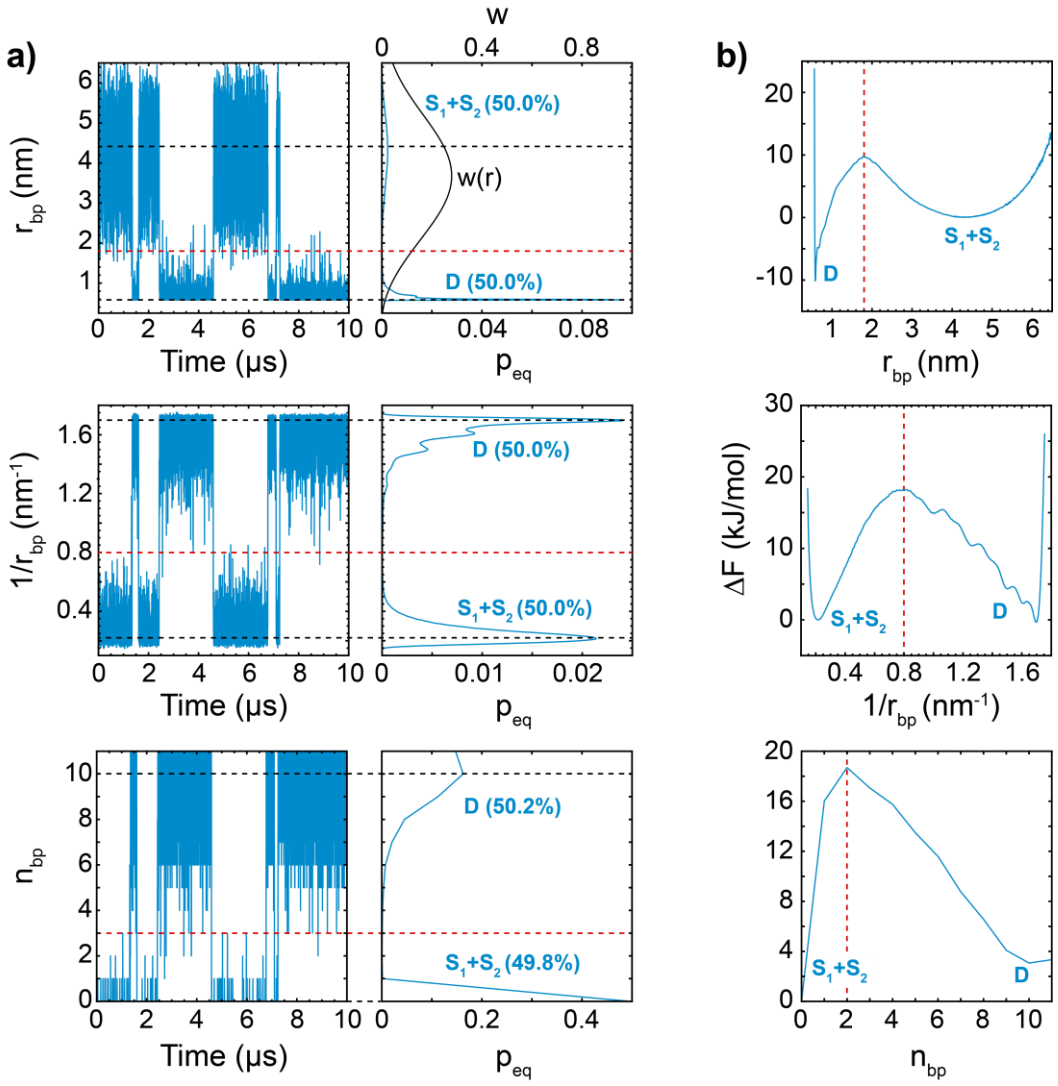

Figure 5.1 3SPN.2 MD trajectories projected onto base-pairing collective variables. (a) Example 3SPN.2 MD trajectories of CGCcap at 330 K plotted along (top) average in-register base-pair separation (r_{bp}), (middle) average inverse in-register base-pair separation ($1/r_{bp}$), and (bottom) number of intact in-register base pairs (n_{bp}). Column to the right shows the probability distribution (p_{eq}) along each collective variable from 25 × 10 μs simulations where duplex (D) and single-strand (S_1+S_2) basins are labeled. The probability that the nearest neighbor molecule is a distance r away, $w(r)$, is shown as the black line on the r_{bp} plot.[48] The integrated population over D and S_1+S_2 basins is indicated for each CV. **(b)** Free-energy profiles (FEPs) calculated from probability distributions as $\Delta F = -RT\ln(p_{eq})$ relative to the single-strand free-energy minimum. The peak of the hybridization free-energy barrier is marked with a red-dashed line.

Figure 5.1 illustrates how each base-pairing CV offers distinct sensitivity to the process of hybridization and structural changes within the duplex basin. An example 10 μs 3SPN.2 MD trajectory for 5′-CGCATATATAT-3′ + complement (CGCcap) is projected onto each CV, demonstrating that the system spends most of its time in the single-strand or duplex states with rapid transition paths between states. Regardless of the CV, free-energy profiles (FEPs, Fig. 5.1b) calculated from 25×10 μs of trajectory data and plotted relative to the single-strand basin minimum (ΔF) indicate that the single-strand and duplex states are separated by a single free-energy barrier, consistent with nucleation-zipper models of hybridization.[46-47] Along r_{bp}, the position of the single-strand free-energy minimum is determined by the size of the simulation box (Table 5.1). The position of the minimum is shifted to a higher r_{bp} value relative to the maximum of the probability distribution that a molecule's nearest neighbor is located a distance r away, $w(r)$.[48] The single-strand basin along r_{bp} is much broader (2.5 – 6 nm) than the transition region (1 – 2.5 nm) or the duplex basin (0.57 – 1 nm). The duplex free-energy minimum, corresponding to the fully intact duplex, is located at 0.58 nm as the equilibrium separation for A:T and G:C base pairs in 3SPN.2 are 0.55 and 0.60 nm, respectively.[23] However, the duplex state is much broader than this minimum because the termini can fray without disrupting the rest of the duplex. The $1/r_{bp}$ CV provides higher resolution of the landscape topology at short base-pair separations, leading to similar widths for the transition region (0.5 – 1.2 nm^{-1}), single-strand state (0.15 – 0.5 nm^{-1}), and duplex state (1.2 – 1.7 nm^{-1}). Integration over the duplex and single-strand basins, performed from either side of the free-energy barrier, indicates that r_{bp} and $1/r_{bp}$ generate identical two-state populations. n_{bp} is a discrete CV that places the entire single-strand ensemble

into the n_{bp} = 0 state, putting all emphasis on the transition region and duplex state. Due to its usage in statistical models of hybridization,[47, 49-50] n_{bp} is one of the most common CVs for describing hybridization and serves as an important CV for comparing our results with previous studies.

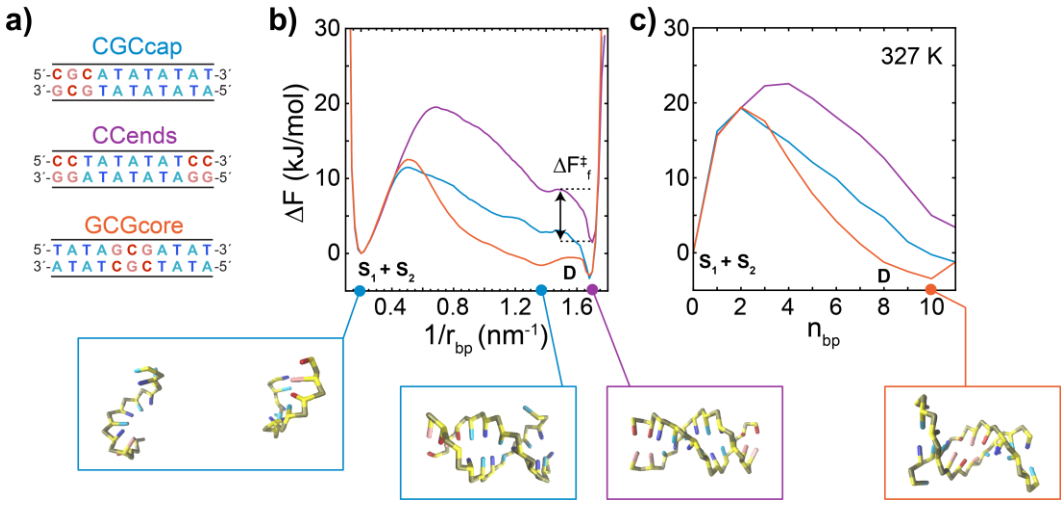

Figure 5.2 Sequence-dependent free-energy profiles for hybridization. (a) Model DNA sequences CGCcap, CCends, and GCGcore used throughout this chapter. FEPs as a function of **(b)** $1/r_{bp}$ and **(c)** n_{bp} at 327 K for each sequence determined from 3SPN.2 MD simulations with well-tempered metadynamics (WTMetaD).[9] Different arrangements of G:C and A:T base pairs alter the structure of the barrier region and duplex basin.

We use three model sequences to assess sequence-dependent effects on the dynamics and FEPs of hybridization. Our sequences vary the relative arrangement of G:C and A:T base pairs, which can significantly alter the duplex ensemble and pathways for hybridization.[27, 49, 51-52] CGCcap contains three G:C base pairs on one end to create asymmetry in base pairing stability across the duplex, CCends contains G:C termini which minimize fraying, and GCGcore contains G:C base pairs at the duplex center. In this chapter we neglect discussion of fully repetitive

sequences that have been reported to bind out-of-register and hybridize through non-canonical mechanisms.[19, 21, 24, 27] Figure 5.2 shows FEPs for each sequence along $1/r_{bp}$ and n_{bp} determined from 3SPN.2 MD simulations employing well-tempered metadynamics (WTmetaD),[53] and details of the sampling procedure were previously described.[32] Each sequence contains a narrow duplex minimum centered at 1.7 nm^{-1} along $1/r_{bp}$ that corresponds to the fully base-paired duplex (n_{bp} = 11). A free-energy barrier ($\Delta F_f^{\ddagger}$) separates this minimum from another at ~1.4 nm^{-1} that contains configurations with one frayed base pair, similar to the n_{bp} = 10 state. The magnitude of $\Delta F_f^{\ddagger}$ is progressively smaller from CCends to CGCcap to GCGcore due to weaker base pairing of A:T termini relative to G:C termini. The duplex state of GCGcore is also broader than the other sequences (~1.0 – 1.7 nm^{-1}) due to the propensity for fraying of multiple base pairs. These sequence-dependent trends in the width of the duplex free-energy basin are consistent with previous lattice model calculations and temperature-jump experiments.[27, 49, 52, 54] The FEPs also reveal variation in the hybridization free-energy barrier across the three sequences that is qualitatively similar along $1/r_{bp}$ and n_{bp} CVs. The sequence-dependence of the free-energy barrier is a primary focus of this chapter and will be discussed throughout. The barrier maximum for all sequences involves the formation of at least two base pairs. GCGcore and CGCcap each exhibit a barrier maximum at $1/r_{bp}$ = 0.52 nm^{-1} and n_{bp} = 3 that primarily involves the formation of the G:C base pairs. The barrier of CGCcap is broader than in GCGcore, presumably due to the weaker G:C base pairing at the terminus and the high configurational freedom of the unbound A:T region relative to those in GCGcore where base pairs are formed in the center of the duplex. The

barrier maximum of CCends is shifted to $1/r_{bp}$ = 0.68 nm^{-1} and n_{bp} = 4 – 5 due to a need to form more, yet less stable, A:T base pairs to successfully hybridize.

Table 5.1 Summary of unbiased 3SPN.2 MD simulations used in this thesis. Additional simulation details have been described previously.[27, 32]

Sequence	T (K)	Time (μs)	Strand conc. (mM)	Ionic strength (mM)[b]	K_d[d] (mM)	# Transition paths
CGCcap 5′-CGCATATATAT-3′ + comp.	330	25 × 10	5.3 (614 nm^3)[a]	600	1.33	170
CCends 5′-CCTATATATCC-3′ + comp.	327	25 × 10	5.3 (614 nm^3)	600	0.83	55
GCGcore 5′-TATAGCGATAT-3′ + comp.	334	25 × 10	5.3 (614 nm^3)	600	0.98	75
GCmix 5′-ATGATATCAT-3′	324	40 × 25	7 (475 nm^3)	240	2.27	194
GCends 5′-GATATATATC-3′	317	40 × 25	7 (475 nm^3)	240	1.18	140
GCcore 5′-ATATGCATAT-3′	324	40 × 25	7 (475 nm^3)	240	2.07	290

[a]Volume of cubic periodic box. [b]Counterions are treated implicitly. [c]Includes hybridization and dehybridizaiton transition paths. [d]Determined using integrated populations over single-strand and duplex basins along $1/r_{bp}$.

5.2.2 Quality of collective variables for probing hybridization dynamics

Here we assess the quality of our base-pairing CVs for describing hybridization dynamics with the use of transition-path statistics through methods originally developed by Hummer and Best with application to protein folding.[55-57] In this framework, a high-quality CV separates the equilibrium dynamics within duplex and single-strand basins from the transitions between basins. A transition path is defined as the segment of a trajectory that exits from a point *A* and crosses another point *B* before going back across *A*. For duplex hybridization, we place these points on the

edges of the duplex and single-strand basins adjacent to the transition region. Once all transition paths are extracted from the trajectory data, the conditional probability for being on a transition path when at a position x along the CV, $p(TP|x)$, is calculated using a Bayesian relation of conditional probabilities.[55]

$$p(TP \mid x) = \frac{p(x \mid TP)\, p(TP)}{p_{eq}(x)} \tag{5.2}$$

Here, p_{eq} is the equilibrium probability distribution from the whole trajectory introduced in Fig. 5.1, $p(x|TP)$ is the conditional probability of a trajectory being at point x given that it is during a transition path, and $p(TP)$ is the fraction of time spent in a transition path during the trajectory. $p(TP|x)$ is related to committor values (ϕ) that are commonly used to identify transition states. For diffusive Langevin dynamics,[55]

$$p(TP \mid x) = 2\phi_A(x)\phi_B(x) = 2\phi_B(x)\left[1-\phi_B(x)\right] \tag{5.3}$$

$$\text{with } \phi_A(x) = 1-\phi_B(x)$$

where $\phi_A(x)$ is the probability that a trajectory starting from *r* will cross point A before point B and $\phi_B(x)$ corresponds to the opposite. For a two-state system, $p(TP|x)$ should have a Gaussian shape peaked at the free-energy barrier maximum $p(TP|x) = \phi_A(x) = \phi_B(x) = 0.5$, meaning that the system has an equal probability of going to state A or B at the barrier maximum. However, eq. 5.3 assumes that the diffusion coefficient is independent of x, which is generally inaccurate for 1D projections of biomolecular folding and binding dynamics and indicates a typical limitation of committor analysis to identify high-quality 1D CVs.[5, 41-42] Therefore, the quality of a CV is

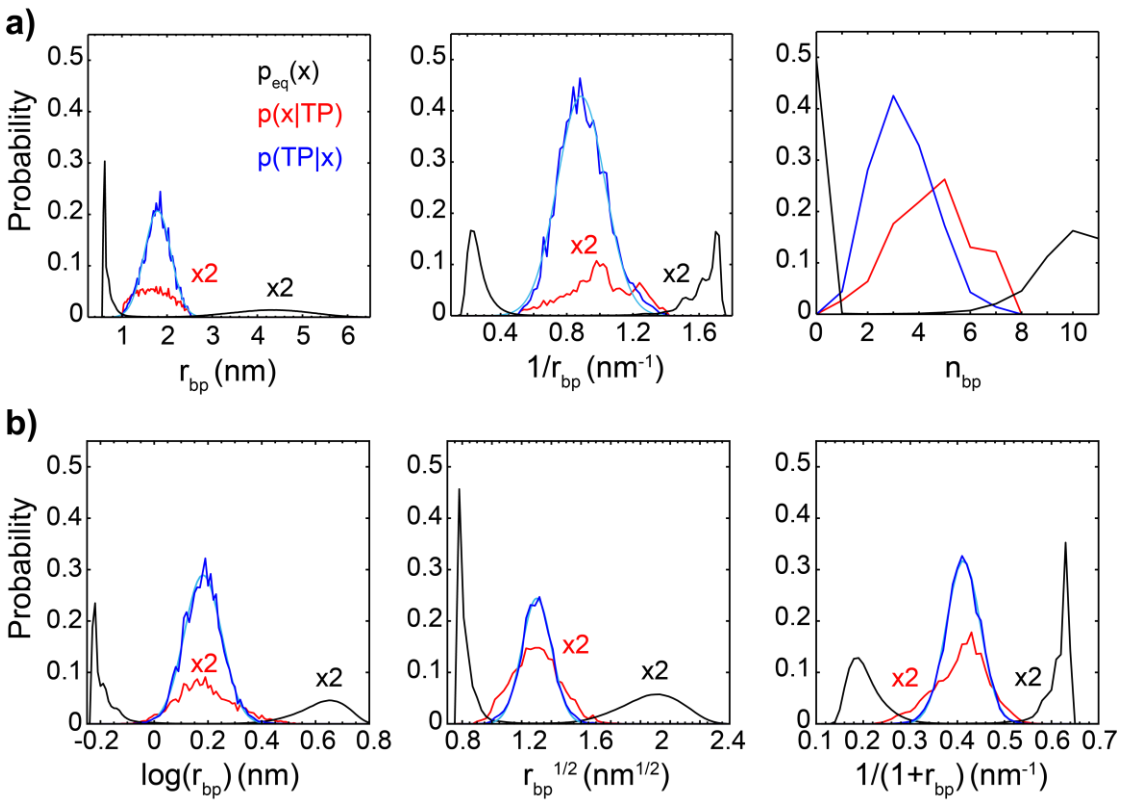

Figure 5.3 Assessing the quality of collective variables from transition paths. (a) Probability distributions (p_{eq}) of CGCcap along r_{bp}, $1/r_{bp}$, and n_{bp} CVs from 25 × 10 μs of unbiased 3SPN.2 MD simulations. $p(x|TP)$ corresponds to the probability distribution during transition paths between duplex and single-strand states, and $p(TP|x)$ is the probability of being on a transition path at point x along the CV calculated using eq. 5.2. p_{eq} and $p(x|TP)$ distributions are normalized but $p(TP|x)$ is not normalized. Fits of $p(TP|x)$ to a Gaussian function are shown for r_{bp} and $1/r_{bp}$ (light blue line). p_{eq} and $p(x|TP)$ are vertically scaled by a factor of 2 in all but n_{bp} plots for clarity. The probability of being on a transition path, $p(TP)$, is approximately 5×10^{-4} for each CV. Higher peak values of $p(TP|x)$ in $1/r_{bp}$ and n_{bp} indicate that these are higher quality CVs than r_{bp} for monitoring hybridization dynamics. **(b)** Same plots for additional CVs: average log of base-pair separation $\log(r_{bp})$, average square root of base-pair separation $r_{bp}^{1/2}$, and $1/(1 + r_{bp})$. p_{eq} and $p(x|TP)$ are vertically scaled by a factor of 2 for each CV.

determined based on how well $p(TP|x)$ conforms to a single sharp peak with a maximum near 0.5, where the transition-state ensemble is collapsed to a single point along x. A poor CV will

exhibit a very broad maximum or a peak well below 0.5, signifying that excursions within single-strand and duplex basins overlap with transition paths and that the transition-state ensemble is not distinctly identified along the CV.

Figure 5.3a shows $p(TP|x)$ calculated along r_{bp}, $1/r_{bp}$, and n_{bp} CVs for CGCcap. $p(TP|x)$ exhibits a Gaussian shape along each of the CVs, yet a notably smaller maximum value for r_{bp} (0.2) than $1/r_{bp}$ (0.43) and n_{bp} (0.43). The sharp features in $p(TP|x)$ and $p(x|TP)$ primarily come from poor sampling of the transition-path region (50 – 200 transition paths per sequence) as transition paths only make up a ~5×10^{-4} portion of the total simulation.

The transition region between duplex and single-strand states is characterized by small changes in base-pair separation that lead to base-pair formation and breaking, and these are better separated from single-strand heterogeneity along $1/r_{bp}$ than with r_{bp}. For example, consider a scenario where each base pair is separated by 3 nm (r_{bp} = 3 nm) and then three base pairs form at one end to nucleate the duplex. Assuming the other base pairs maintain an average separation of 3 nm, this nucleated configuration has r_{bp} = 2.4 nm, a value that may also be achieved by fully dissociated single-strands (for example if all base pairs are separated by 2.4 nm). The same nucleation leads to a change from 0.33 to 0.63 nm^{-1} along $1/r_{bp}$ whereas configurations with all base pairs separated by 2.4 nm are well isolated from the transition region at $1/r_{bp}$ = 0.42 nm^{-1}. From the same reasoning, CVs such as the average log-space base-pair separation, $\log(r_{bp})$, exhibit a maximum $p(TP|x)$ value between that of r_{bp} and $1/r_{bp}$ (Fig. 5.3b). The n_{bp} CV similarly isolates small changes in base-pair separation among the dissociated single strands from those that lead to base-pair formation and breaking via the cutoff distance of 0.7 nm. Given that we are

interested in exploring the dynamics of hybridization, we will analyze trajectories using $1/r_{bp}$ throughout this chapter because it provides the best ability to describe barrier crossing.

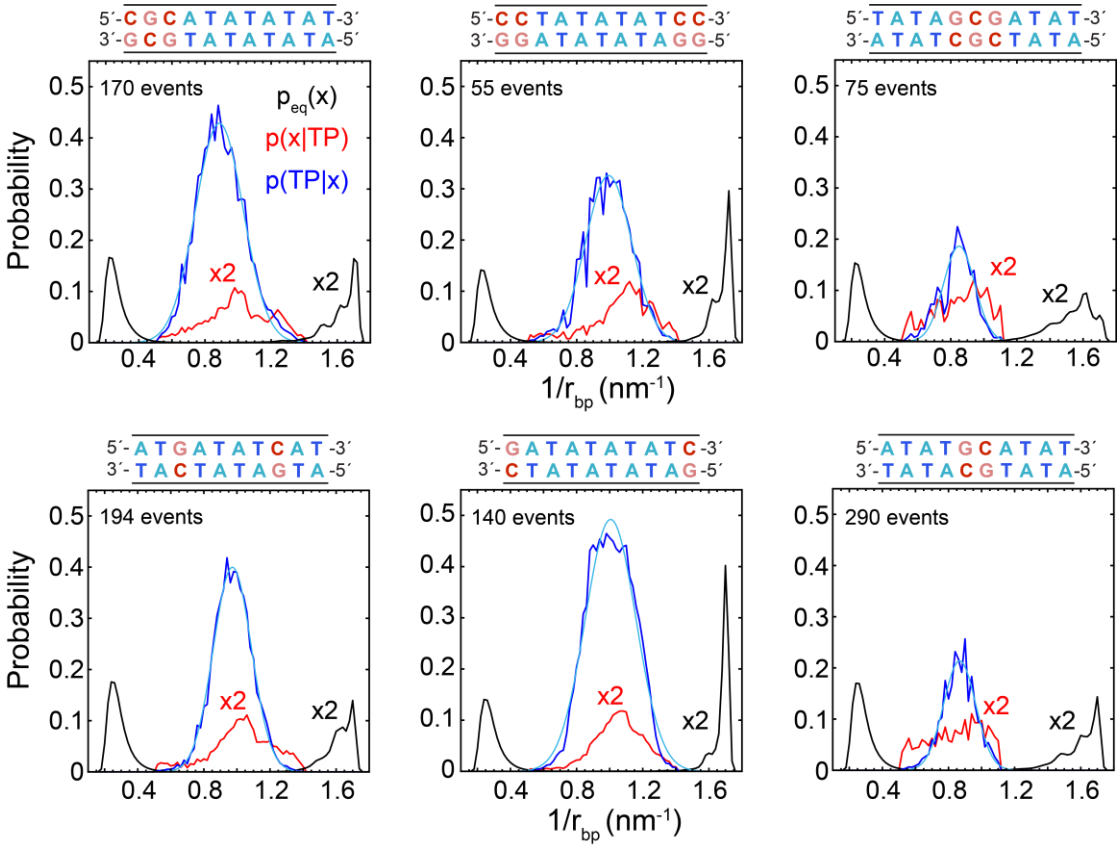

Figure 5.4 Sequence-dependent quality of $1/r_{bp}$ collective variable. $p(x|TP)$, $p(TP|x)$, and p_{eq} for (top) CGCcap, CCends, and GCGcore along $1/r_{bp}$ calculated from 25 × 10 μs of simulation data (conditions listed in Table 5.1). The same plots are shown in the bottom row for three additional sequences from 40 × 25 μs of simulation data: (left) GCmix, (center) GCends, and (right) GCcore. p_{eq} and $p(x|TP)$ are vertically scaled by a factor of 2 in all plots. The total number of transition paths sampled for each sequence is listed.

Calculation of $p(TP|x)$ for other sequences reveals that the quality of $1/r_{bp}$ and n_{bp} CVs is sequence-dependent. The $p(TP|x)$ maximum along $1/r_{bp}$ decreases from CGCcap (0.43) to CCends (0.33) to GCGcore (0.20). We also analyzed self-complementary sequences 5′-ATGATATCAT-3′ (GCmix), 5′-GATATATATC-3′ (GCends), and 5′-ATATGCATAT-3′

(GCcore). GCmix and GCends each have high $p(TP|x)$ maxima, but GCcore shows a maximum at ~0.2. Although we do not study enough sequences to obtain general sequence trends, the low $p(TP|x)$ values for GCcore and GCGcore suggest that large amounts of terminal fraying reduce the quality of $1/r_{bp}$ and n_{bp}. Fraying extends the duplex state to lower values of $1/r_{bp}$ and n_{bp} such that some fraying events may overlap with transition paths along these CVs. Simply removing the terminal base pairs from $1/r_{bp}$ partially accounts for fraying effects (Fig. 5.5), suggesting that $1/r_{bp}$ should be altered for specific sequence types.

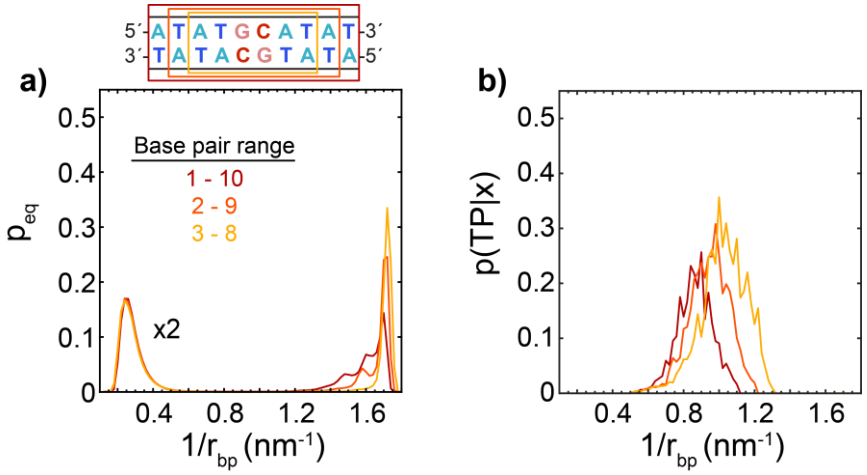

Figure 5.5 Influence of the range of included base pairs on $1/r_{bp}$ quality. (a) p_{eq} for GCcore along $1/r_{bp}$ modified as an average over all base pairs (red), base pairs 2 – 9 (orange), and base pairs 3 – 8 (yellow). Data is vertically scaled by a factor of 2. Removing terminal base pairs from the CV narrows the observed duplex basin. **(b)** Corresponding $p(TP|x)$ distributions calculated along each CV. The $p(TP|x)$ maximum value increases as terminal base pairs are eliminated from $1/r_{bp}$ due to the minimization of fraying effects on the CV.

5.2.3 Optimization of collective variables

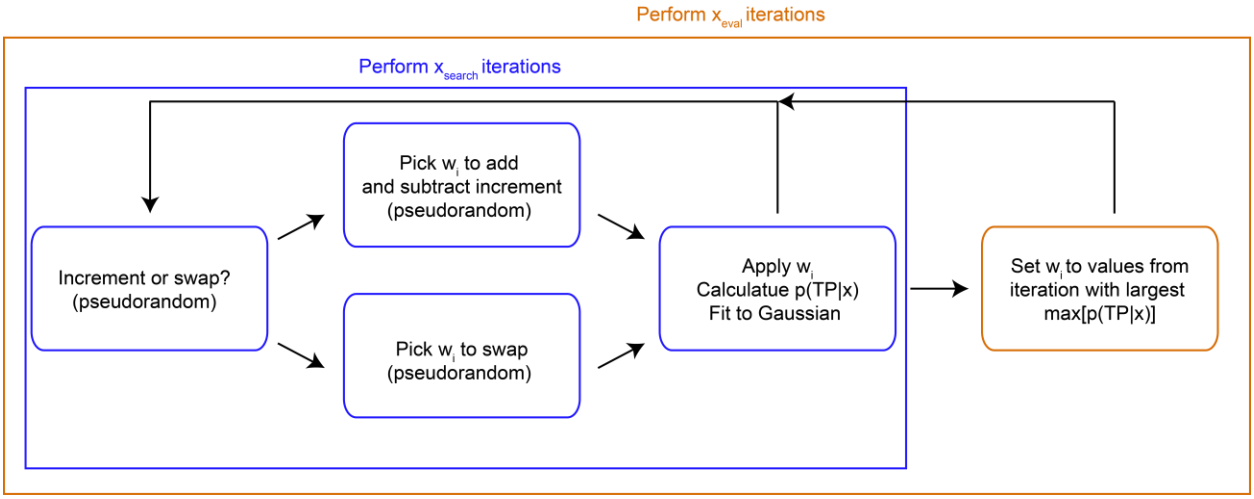

Figure 5.6 Flowchart of Monte Carlo optimization of collective variable based on $p(TP|x)$. During each iteration, weights (w) for two base pair indices i are incremented or swapped in pseudorandom manner and $p(TP|x)$ is calculated along the modified CV. The process is repeated for x_{search} iterations without any penalties and then the weights are set to values that achieve the largest maximum value of $p(TP|x)$ during the search. The optimization is then repeated for x_{eval} iterations of the search process.

The high sensitivity of $p(TP|x)$ to oligonucleotide sequence and choice of base pairs suggests that the $1/r_{bp}$ and n_{bp} CVs for a given sequence may be optimized based on $p(TP|x)$ by weighting contributions from each base-pair index. As a continuous CV, $1/r_{bp}$ can be more robustly optimized than n_{bp} and will be the focus of this section. In principle, the full matrix of inter-nucleobase separations could be included in the CV to optimize for certain strand orientation effects or out-of-register base pairing, but we limit our analysis to in-register base-pair separations to minimize the complexity of the optimized CV. We adopt a Monte Carlo optimization procedure (Fig. 5.6) introduced by Best & Hummer for protein folding and later applied to protein-protein

association.[43, 56, 58] The $1/r_{bp}$ CV is now a weighted average over individual inverse base-pair separations ($1/r_{bp,opt}$) where w_i is the weight applied to each index.

$$1/r_{bp,opt} = \frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{r_{bp,i}} \tag{5.4}$$

$$\text{where } \sum_{i=1}^{N} w_i = N$$

The sum of weights is conserved to ensure that the range of $1/r_{bp,opt}$ does not significantly change during the optimization process. During each iteration of the optimization, w_i for two indices are either swapped or incremented. The operation and indices are picked pseudorandomly using the MATLAB randi function. If incremented, one index is increased and the other is decreased by a constant value. An increment of 0.1 was used for all optimizations. $p(TP|x)$ is calculated along $1/r_{bp,opt}$ during each iteration and fit to a Gaussian to obtain the maximum value. This process repeats for x_{search} unconstrained iterations of increments or swaps to find a combination of w_i that increases the maximum value of $p(TP|x)$. After x_{search} iterations, w_i resets to the values that gave the largest maximum value of $p(TP|x)$ with all $w_i > 0$. This process is then repeated x_{eval} times. x_{search} needs to be large enough to get out of local maxima yet small enough to maintain computational efficiency. Optimizations presented in this chapter were performed with x_{search} = 100 and x_{eval} = 1000.

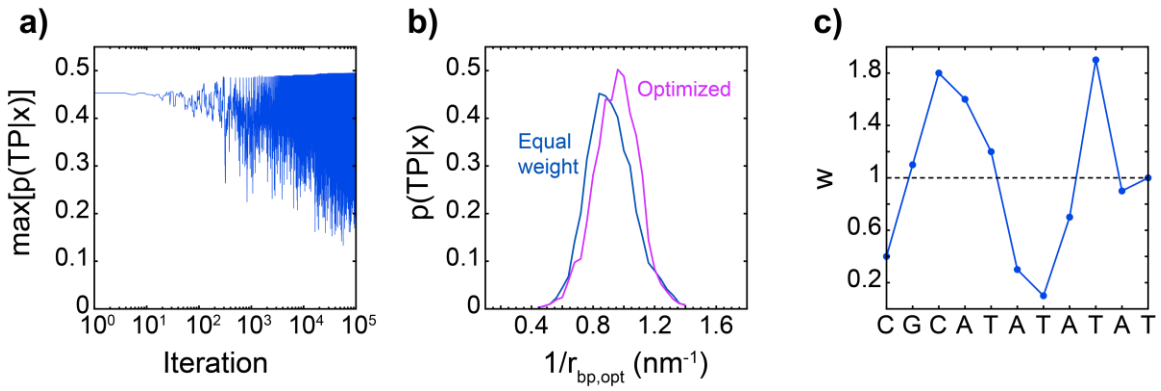

Figure 5.7 Example of Monte Carlo optimization of $1/r_{bp}$ based on $p(TP|x)$. (a) Maximum value of $p(TP|x)$ across $x_{search} \times x_{eval}$ iterations for CGCcap. **(b)** Comparison of $p(TP|x)$ prior to optimization and afterward. **(c)** Final base pair weights (w). Based on eq. 5.4, indices with $w > 1$ contribute more to $1/r_{bp,opt}$ than those with $w < 1$.

An example optimization of CGCcap is illustrated in Fig. 5.7. The maximum $p(TP|x)$ value is shown to fluctuate throughout $x_{search} \times x_{eval}$ iterations. Reductions in the $p(TP|x)$ maximum occur during the unconstrained search phases, but the overall $p(TP|x)$ maximum is shown to progressively increase and level off during the optimization. Although CGCcap already starts from a maximum value of 0.44 with uniform weighting, optimization leads to a maximum value of 0.49. The optimized w values tune $1/r_{bp,opt}$ to increase weight from base pair indices 3 and 4 near the G:C-rich end and de-weight indices 6 – 8 in the A:T region. The optimized CGCcap $p(TP|x)$ distribution and w values are similar for trials initialized from different w values (Fig. 5.8). The optimized w values are most variable for base pairs 9 – 11, suggesting that $p(TP|x)$ is less sensitive to these weights.

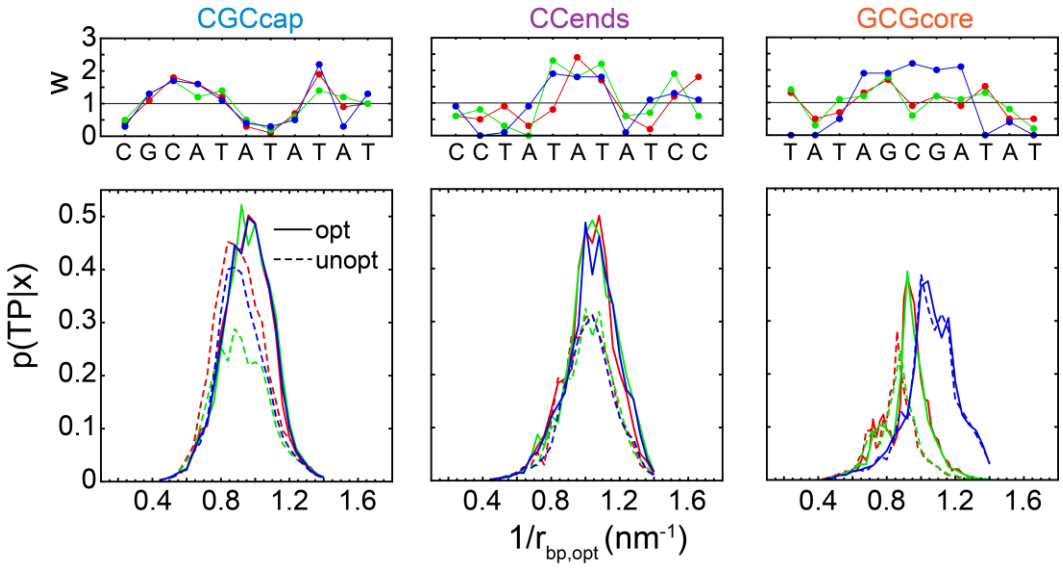

Figure 5.8 Reproducibility of $1/r_{bp}$ optimization. $p(TP|x)$ prior to optimization (unopt) and after optimization for different initial w values (denoted by different colors). Optimized w values are shown on the top. Similar optimized $p(TP|x)$ distributions and w values are obtained for unoptimized $p(TP|x)$ distributions with similar maximum positions along $1/r_{bp,opt}$, but the optimization may get stuck in different local maxima if initialized from a significantly shifted $p(TP|x)$ distribution (GCGcore, blue trace).

CCends and GCGcore show greater sensitivity to the initial w values, demonstrating limitations and inaccuracies of the optimization process (Fig. 5.8). Optimization leads to a significant increase in the $p(TP|x)$ maximum for each sequence, but the final $p(TP|x)$ distribution and w values are sensitive to the initial conditions. Starting the optimization from a highly shifted $p(TP|x)$ distribution for GCGcore leads to an optimized $p(TP|x)$ distribution shifted relative to when starting from the unweighted $1/r_{bp}$ (red curves in Fig. 5.9), yet each contains an identical maximum $p(TP|x)$ value. These two optimization results correspond to distinct local maxima with different w values. Additionally, separate optimization trials for CCends indicate that large spreads of w values can give rise to equivalent $p(TP|x)$ distributions. It is possible that $1/r_{bp}$ is simply a

poor CV to describe GCGcore hybridization and no re-weighting can produce a $p(TP|x)$ maximum near 0.5 even if additional optimization time is provided. In contrast, each trial for CCends gets close to the desired $p(TP|x)$ maximum of 0.5, suggesting that there is no unique solution of w values for this sequence. Therefore, further improvement of the optimization process may be achieved by optimizing on additional properties of the $p(TP|x)$ distribution beyond its maximum value.

Keeping the limitations of the optimization process in mind, $1/r_{bp,opt}$ was generated for all six model sequences (Fig. 5.9). The $p(TP|x)$ maximum reaches 0.4 – 0.5 for CGCcap, CCends, and GCends, suggesting $1/r_{bp,opt}$ fully separates transition paths from equilibrium excursions within single-strand or duplex states for these sequences. The w patterns are qualitatively similar for CCends and GCends where weight is added at the central base pairs and reduced at A:T contacts near the G:C termini. In contrast, sequences that undergo significant terminal fraying exhibit lower quality $1/r_{bp,opt}$ CVs. The $p(TP|x)$ maximum for GCmix remains at ~0.4 and only improves to ~0.3 for GCGcore and GCcore. We also compare $1/r_{bp,opt}$ with the leading CVs obtained from time-lagged independent component analysis (TICA) and state-reversible VAMPnets (SRV),[37, 40] which are optimized for the slowest kinetics in the system (Appendix 5.A).

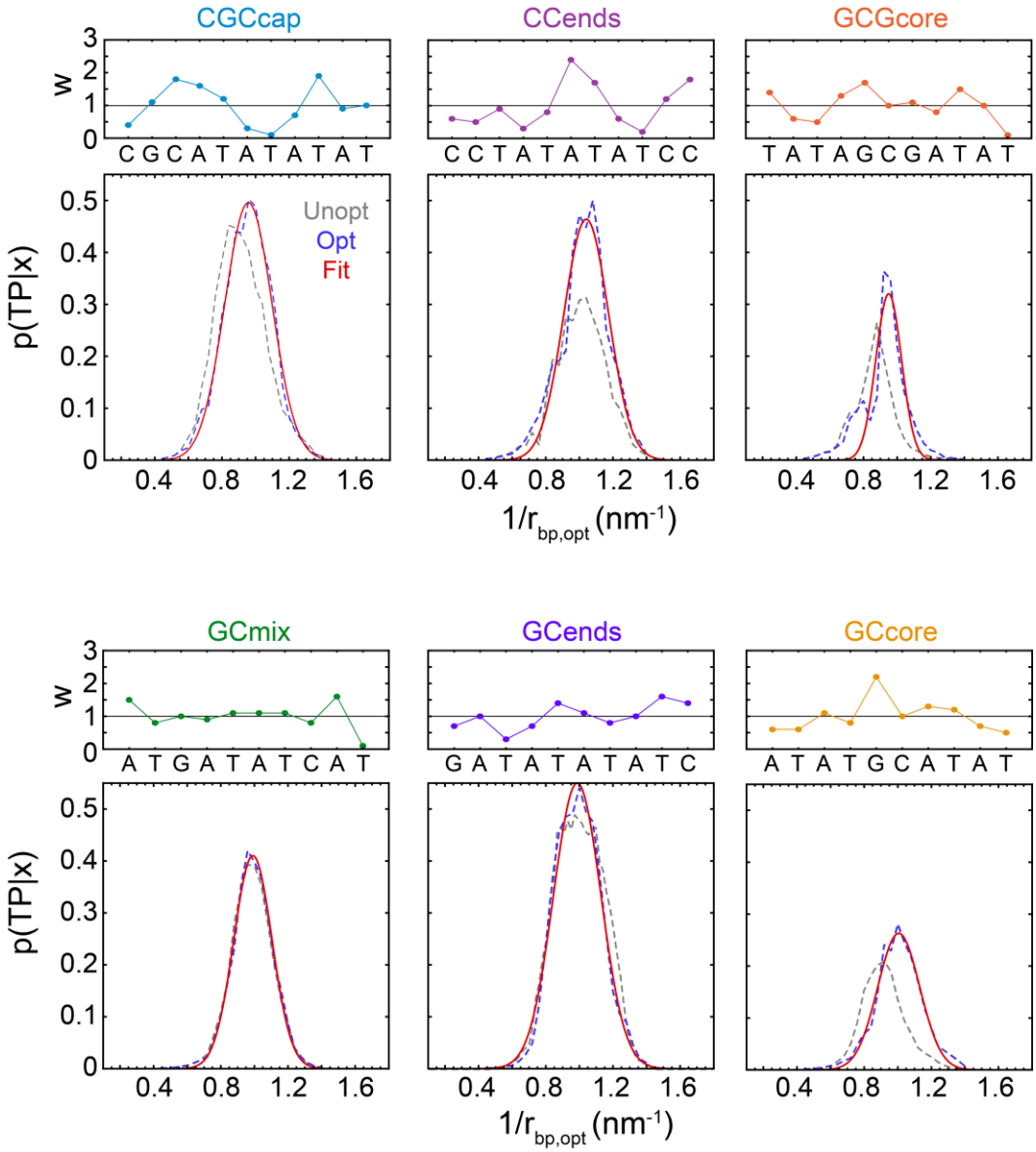

Figure 5.9 Optimized $1/r_{bp}$ CVs for all sequences. $p(TP|x)$ distribution along uniformly-weighted $1/r_{bp}$ (gray dashed line) and $1/r_{bp,opt}$ (blue dashed line) for each sequence. w values are shown above each distribution. Red solid lines correspond to Gaussian fits.

5.3 Structural properties during hybridization and dehybridization

5.3.1 Hybridization mechanism along $1/r_{bp.opt}$

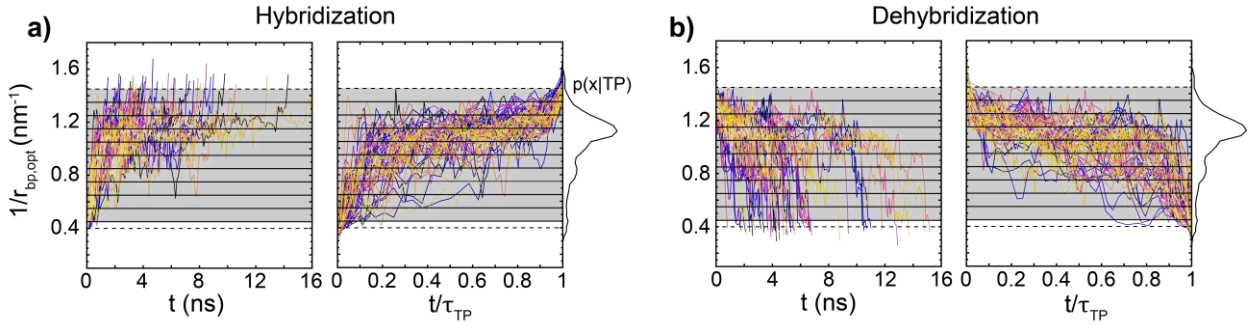

Figure 5.10 Binning transition paths along $1/r_{bp,opt}$. Transition paths for **(a)** hybridization and **(b)** dehybridization observed from CGCcap. Each color corresponds to a distinct transition path starting from 0.4 nm^{-1} and ending at 1.45 nm^{-1} (dashed lines) for hybridization or vice versa for dehybridization. Gray horizontal bars indicate 0.1 nm^{-1} bins along $1/r_{bp,opt}$. Plots on the right show the same transition paths where the time axis is normalized by the duration of each transition path (τ_{TP}). The conditional probability $p(x|TP)$, is shown on the right of each plot.

In this section, we analyze trajectories along $1/r_{bp,opt}$ to study the series of base-pairing events and strand configurations during hybridization. We start by analyzing the properties of hybridization transition paths. Our goal is to describe the configurational and base-pairing properties of DNA as a function of transition-path time. It would be most simple to bin $1/r_{bp,opt}$ across all transition paths at each time point; however, the duration of each transition path, τ_{TP}, is different and follows a distribution that depends on the free-energy landscape and CV.[59] We instead bin across all time points in all transition paths that share a given value of $1/r_{bp,opt}$ (Fig. 5.10). Since we only sample 50 – 300 transition paths (including hybridization and dehybridization) in the unbiased simulations for each sequence, we use a bin size of 0.1 nm^{-1} along $1/r_{bp,opt}$ that encompasses 100 – 5000 simulation frames per bin. Transition paths are defined

using boundaries of $1/r_{bp,opt}$ = 0.4 nm^{-1} and 1.45 nm^{-1}. As illustrated in Fig. 5.10 for CGCcap, the progress in time and progress along $1/r_{bp,opt}$ during a transition path share a non-linear relationship. There is roughly a linear relationship within the 0.4 – 0.8 nm^{-1} and 1.1 – 1.45 nm^{-1} regions that each encompass ~20% of the transition path duration. There is a large spread of time values from 0.8 – 1.1 nm^{-1} due to numerous recrossings near the hybridization free-energy barrier, leading to a maximum in the conditional probability for being at position x given the trajectory is on a transition path, $p(x|TP)$. Similar relationships between transition-path progress in time and along a defined CV have been reported from single-molecule extension trajectories of DNA hairpin folding as well as from simulations of diffusive motion along a 1D double-well potential.[6, 60-61] Additional insight into the underlying dynamics may be available from the shape of the transition path,[6, 60] but here we show it to illustrate that each bin along $1/r_{bp,opt}$ contains a different spread of transition-path time points.

We use contact-separation maps to show the average distances between all nucleobases within each $1/r_{bp,opt}$ bin (Fig. 5.11). Distances along the diagonal correspond to in-register contacts while those off the diagonal correspond to out-of-register contacts. Examples of CGCcap and CCends from 0.45 to 1.45 nm^{-1} demonstrate a clear progression of reduction in base-pair separation, particularly along the diagonal, as $1/r_{bp,opt}$ increases. The behavior of CGCcap is consistent with the nucleation analysis in Appendix 6.E where separation first reduces near the G:C-rich end and the A:T end remains highly separated. In contrast, CCends follows a greater diversity of pathways as separations drop almost uniformly across the central contacts at $1/r_{bp,opt}$ = 0.85 to 1.15 nm^{-1}. The diversity of base-pair contacts is more clearly visualized with contact

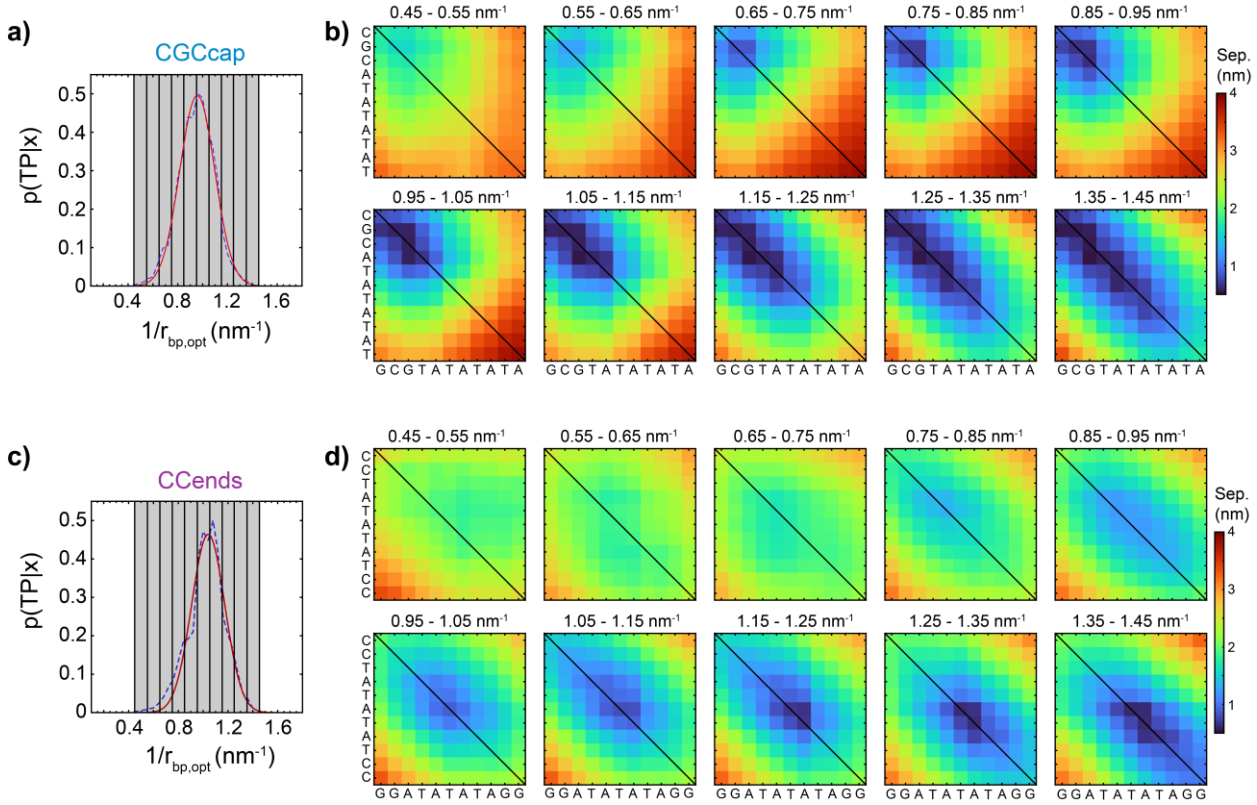

Figure 5.11 Progress of hybridization transition paths visualized with contact-separation maps. (a) $p(TP|x)$ along $1/r_{bp,opt}$ for CGCcap at 330 K where vertical gray regions correspond to slices used for averaging transition paths. **(b)** Separation maps plotting the distance between all nucleobases at different points throughout (de)hybridization transition paths. Each plot corresponds to an average over all transition-path simulation frames within the denoted $1/r_{bp,opt}$ window. Diagonal lines correspond to in-register base pairs **(c-d)** Similar plots for CCends at 327 K.

maps that show the probability of each distance being below 0.7 nm (Fig. 5.12). The contact maps reveal the presence of out-of-register contacts from 0.45 to 0.95 nm^{-1} for CGCcap and 0.45 to 1.25 nm^{-1} for CCends. In CGCcap, out-of-register contacts are primarily observed between the A:T end of the CGC strand and the central A:T bases of the GCG strand. In CCends, out-of-register contacts are scattered from 0.45 to 0.85 nm^{-1}, but appear localized to a two-base-pair shifted configuration

from 0.85 to 1.25 nm^{-1}. The out-of-register contacts lead to non-linear relationship between $1/r_{bp,opt}$ and the number of intact in-register base pairs (n_{bp}).

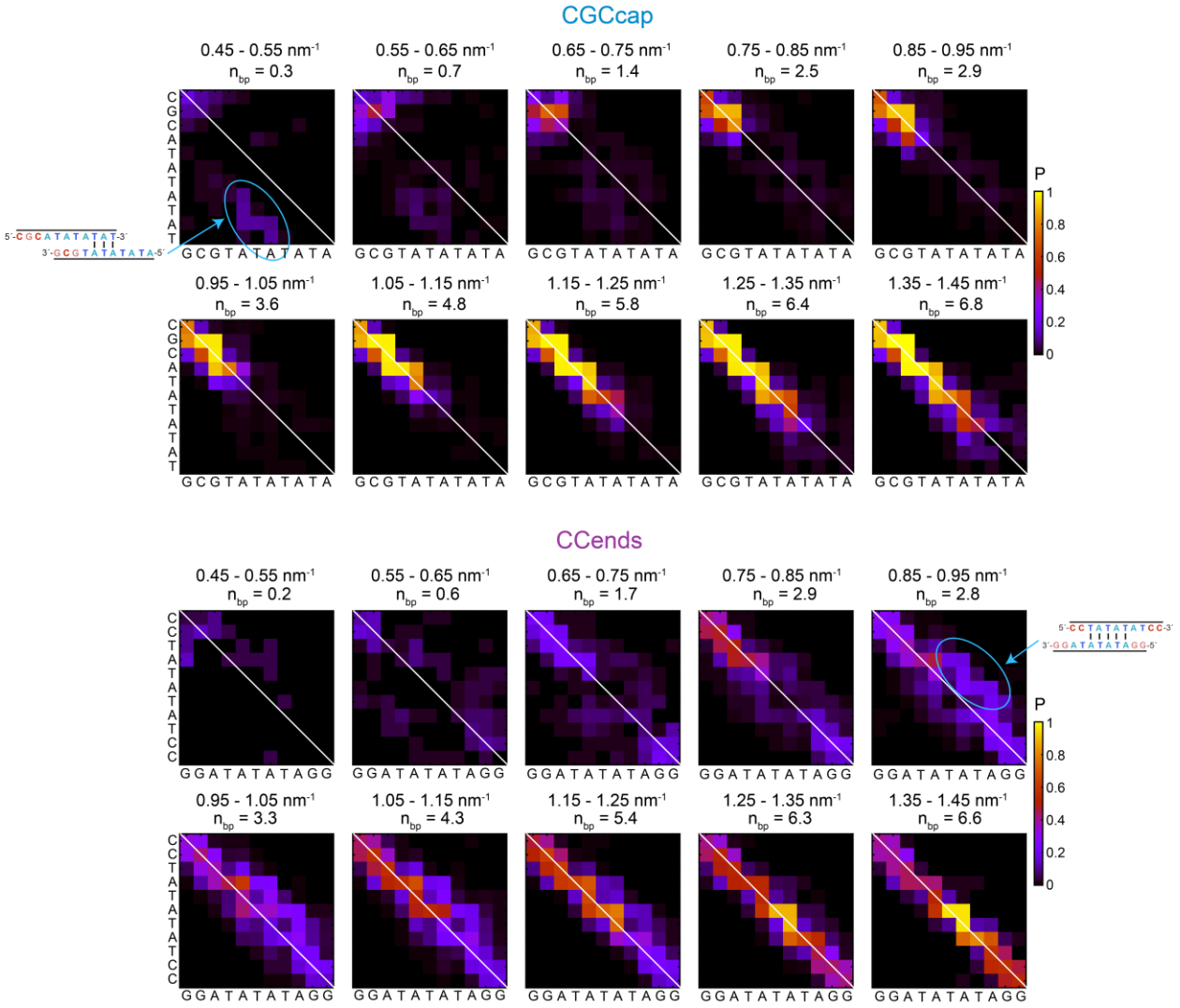

Figure 5.12 Progress of hybridization transition paths visualized with contact-probability maps. Base-pair contact maps for (top) CGCcap and (bottom) CCends indicating the probability that given nucleobase separation is below 0.7 nm at different points along (de)hybridization transition paths. Separation maps plotting the distance between all nucleobases at different points throughout transition paths. Each plot corresponds to an average over all transition path simulation frames within the denoted $1/r_{bp,opt}$ window. The average number of intact in-register base pairs (n_{bp}) is also listed for each window.

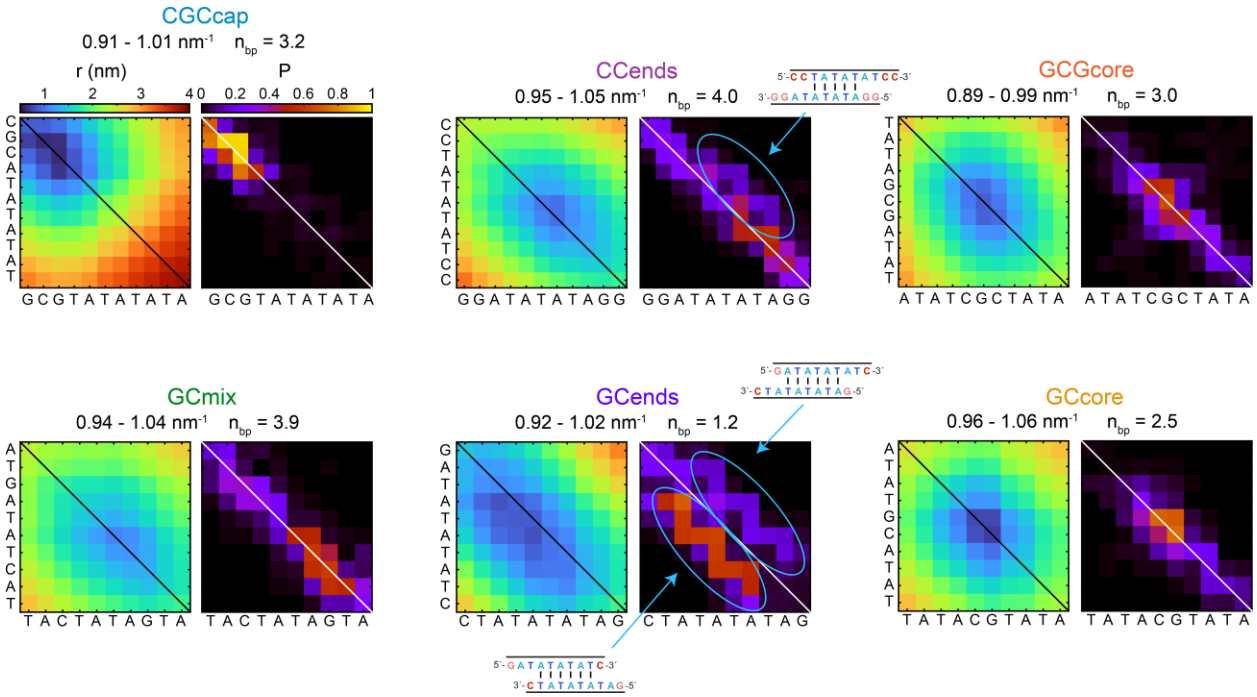

Figure 5.13 Base-pairing properties at the $p(TP|x)$ maximum. Separation-contact and contact-probability maps averaged across a 0.1 nm^{-1} window centered at the maximum of $p(TP|x)$ along $1/r_{bp,opt}$ (Fig. 5.9) for each sequence. Simulations temperatures are listed in Table 5.1.

To investigate which base-pair contacts are most common in the hybridization transition-state ensemble, we plot the contact-separation and contact-probability maps for a 0.1 nm^{-1} bin centered at the maximum of $p(TP|x)$ along $1/r_{bp,opt}$ (Fig. 5.13). The $p(TP|x)$ maximum best reports of the transition-state ensemble when having a value close to 0.5 (CGCcap, CCends, GCmix, and GCends) and less so for sequences with a lower value maximum (GCGcore and GCcore). Since we are only binning over transition paths, the maps only contain configurations that successfully cross the hybridization transition state. In general, the base-pairing patterns indicate a high prevalence of G:C base pairing in the transition-state ensemble, as exhibited by CGCcap, GCmix, GCGcore, and GCcore. CCends and especially GCends exhibit significant

probability for pairing out-of-register in configurations shifted by two base pairs. Interestingly shifted configurations of CCends only occur in the direction with 3′ overhangs whereas GCends exhibits particular preference for shifting with 5′ overhangs (~50%) vs. 3′ overhangs (20%). These shifted configurations of GCends were also found as on-pathway macrostates between the duplex and single-strand states in a Markov state model constructed from the same 3SPN.2 MD simulation data.[27] The two-base-pair shifted configurations of GCends are stabilized relative to CCends largely because they can form an extra A:T base pair. However, it is unclear why CCends only binds with a 3′ overhang orientation while GCends prefers the 5′ overhang configuration. Empirical trends from thermodynamic studies of oligonucleotides indicate that 5′ overhangs are generally more stabilizing to the duplex than 3′ overhangs,[62-63] but the physical origin for these trends remains unclear and may not be incorporated into 3SPN.2. Additionally, explicit mismatch base-pairing interactions would likely influence the relative stability shifted configurations,[64] but 3SPN.2 only includes excluded volume interactions for mismatches that may produce inaccurate stability of non-canonical structures.[23, 27]

5.3.2 Reactive and non-reactive hybridization events

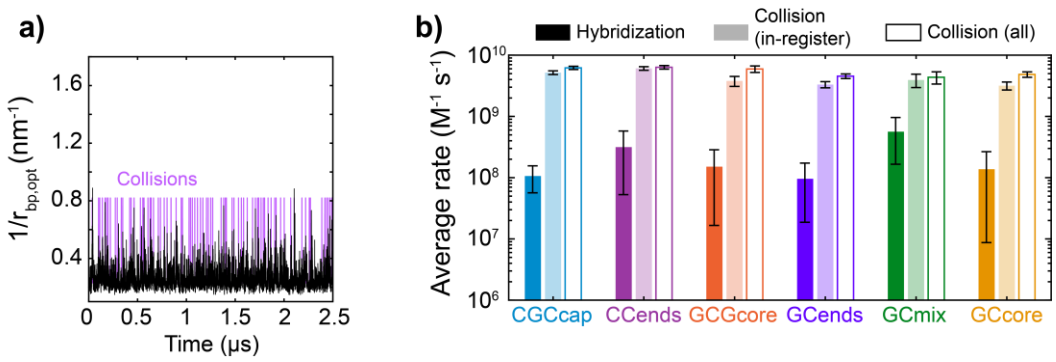

Figure 5.14 Rates of oligonucleotide collisions and hybridization. (a) Example trajectory of CGCcap along $1/r_{bp,opt}$ (black) in the single-strand state. Purple lines indicate time points where single strands collide. **(b)** Average hybridization rate (dark-filled bars), collision rate based on in-register contact (light-filled bars), and collision rate using any contact for each sequence (open bars). Rates are divided by the effective oligonucleotide concentration (Table 5.1) to give the respective bimolecular rate constant. Error bars indicate standard deviation across distributions of hybridization rates.

So far we have only considered transition paths, or the trajectories where hybridization and dehybridization are successful. However, hybridization is a rare event where many failed attempts occur on average before successful duplex formation. A simple question that arises is whether there are structural differences between successful and failed hybridization events, and this is a general question of interest across various bimolecular reactions. Such dynamics are rarely investigated in biomolecular systems due to the difficulties of sufficiently sampling successful and failed events,[58, 65-66] and we are not aware of any reports for nucleic acid systems.

We first compare the kinetics of collisions and successful hybridization and observe behavior similar to experiment. The diffusion-limited hybridization rate constant for an 8-mer oligonucleotide is estimated to be 2×10^9 M^{-1} s^{-1} in aqueous solution based on an experimental

diffusion coefficient of 1×10^6 cm^2/s.[67] This rate constant is 100 – 1000-fold larger than hybridization rate constants measured with both single-molecule and ensemble spectroscopy,[67-68] suggesting that numerous diffusive encounters between strands fail prior to hybridization. Figure 5.14 shows the rate of strand collisions and hybridization determined from 3SPN.2 MD simulations. Unbiased trajectories are divided into single-strand and duplex dwell periods, and collisions are identified within the single-strand dwell periods. Collisions are defined as going from a completely dissociated configuration to forming at least one inter-nucleobase contact. To minimize apparent collisions arising from rapid strand reorientation, we only consider collisions that occur at least 10 ns after the previous collision. The collision rate constant is then calculated as the total oligonucleotide concentration divided by the mean time interval between collisions, $k_{col} = c_{tot}/\langle\tau_{col}\rangle$. The value of k_{col} ranges from $3 – 6 \times 10^9$ M^{-1} s^{-1} for each sequence, comparable to the experimental diffusion-limited rate constant. k_{col} is similar whether all contacts or only in-register contacts are considered a collision. The hybridization rate constant is calculated as the total oligonucleotide concentration divided by the mean single-strand dwell time, $k_h = c_{tot}/\langle\tau_{ss}\rangle$, and ranges from $1 – 4 \times 10^8$ M^{-1} s^{-1}. k_h is determined only from 20 – 40 single-strand dwell times per sequence, so the sequence-dependent trends may not be reliable. We observe a general ~20-fold reduction in k_h relative to k_{col}. While this difference is smaller than from experiment, it was previously shown that (de)hybridization timescales from 3SPN.2 MD simulations are 10-fold faster than in experiment.[27] Given this acceleration of hybridization in 3SPN.2, our comparison of collision and hybridization timescales are consistent with a picture where numerous binding attempts fail prior to successful hybridization.

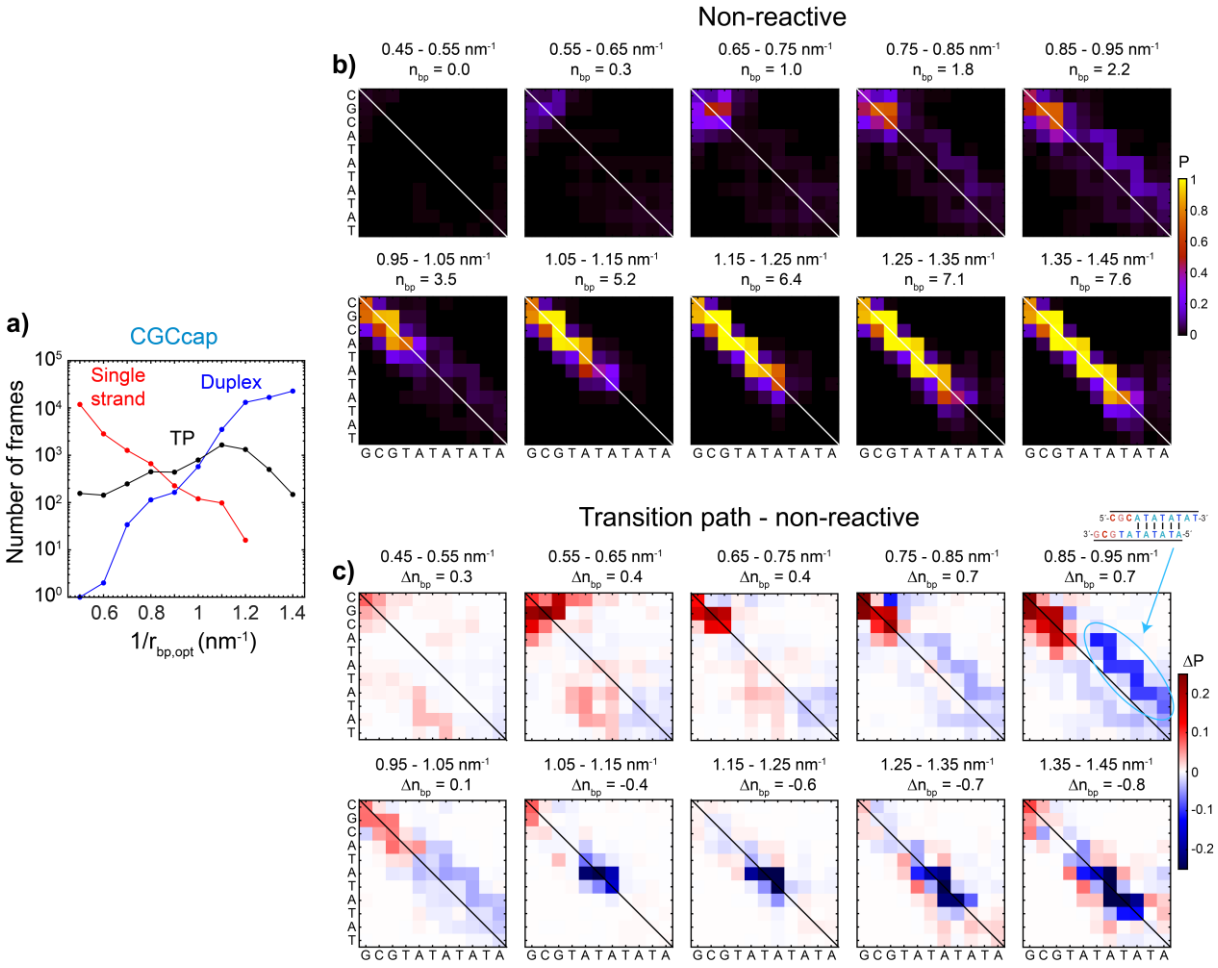

Figure 5.15 Comparing successful and unsuccessful (de)hybridization events. (a) Number of frames contributing to each bin along $1/r_{bp,opt}$ for CGCcap. The black markers map out $p(x|TP)$. Red corresponds to failed hybridization events originating from the single-strand state and blue corresponds to failed dehybridization events originating from the duplex state. **(b)** Contact-probability maps for failed (de)hybridization trajectories from 0.45 to 1.25 nm^{-1}. **(c)** Difference between transition-path (Fig. 5.13) and failed-trajectory contact-probability maps. The difference in the number of intact in-register base pairs is also listed (Δn_{bp}). Red indicates greater base-pairing probability in the transition-path ensemble and vice versa for blue.

We next aim to characterize the spatial and base-pairing properties that differentiate a failed vs. successful hybridization or dehybridization event. We generate contact-separation and contact-probability maps by binning along $1/r_{bp,opt}$ using all frames that are not part of transition paths.

As shown for CGCcap in Fig. 5.14a, the bins from 0.45 – 0.85 nm^{-1} are dominated by dynamics within the single-strand state while bins from 0.95 – 1.45 nm^{-1} are dominated by the duplex state, and the corresponding contact-probability maps are shown in Fig. 5.15b. To isolate differences with transition paths, we take the difference between transition-path and non-reactive contact-probability maps at each bin. From 0.45 – 0.85 nm^{-1}, transition paths show a clear preference for base pairing at the G:C end with an increase in base-pairing probability of $\Delta P = 0.1 – 0.2$. Also, the configurations with out-of-register A:T base pairing are only found in transition paths (Fig. 5.11), suggesting it may serve to keep strands in proximity for eventual hybridization. A recent kinetic modelling study of hybridization also suggested that out-of-register contacts facilitate duplex formation by extending the lifetime of the encounter complex between strands.[69] Transition paths still show a preference for G:C base pairing in the barrier region from 0.75 – 1.05 nm^{-1} and contain an average of 0.7 more in-register base pairs than non-reactive trajectories (Δn_{bp}). Additionally, non-reactive paths have a 20% probability of binding in a two-base-pair shifted configuration (Fig. 5.14b) that acts as an off-pathway kinetic trap to hybridization. On the duplex side of the barrier (1.05 – 1.45 nm^{-1}), transition paths instead contain less base pairs on average relative to non-reactive paths ($\Delta n_{bp} < 0$), particularly in the central A:T region. Since dehybridization typically starts with unzipping of the A:T end, trajectories that break the central A:T contacts are more likely to proceed to full dissociation of the duplex.

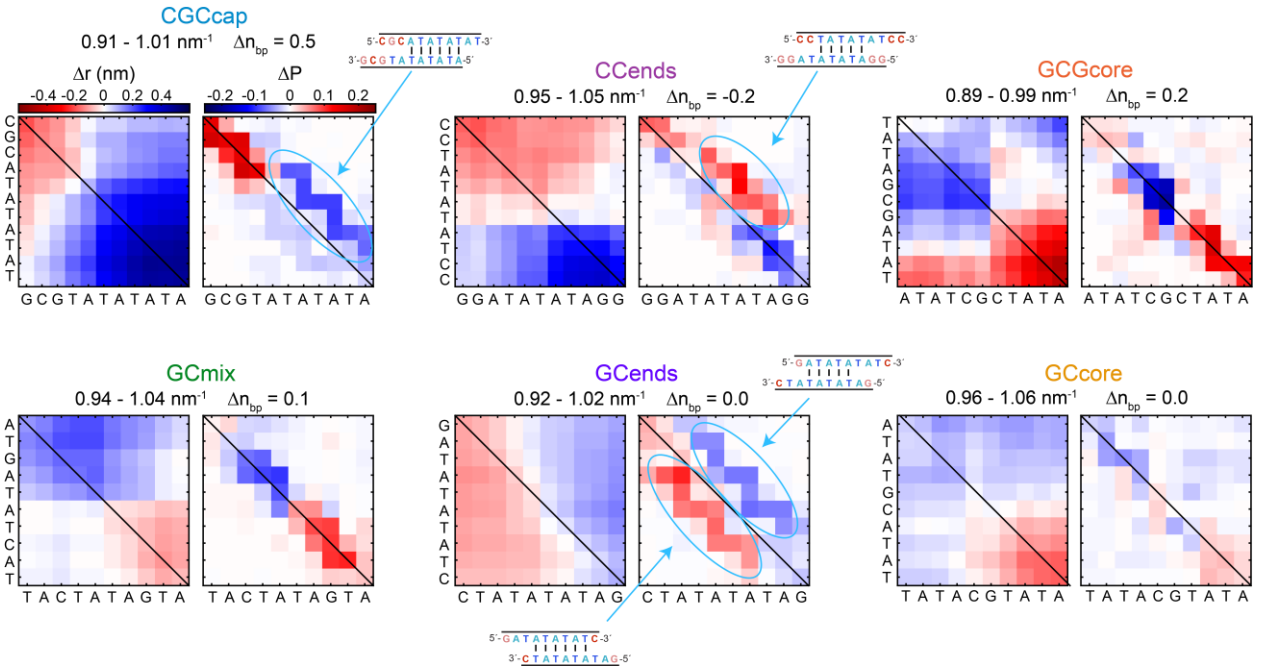

Figure 5.16 Difference between successful and failed events at the maximum of $p(TP|x)$. Difference between transition-path (Fig. 5.14) and non-reactive (left) contact-separation and (right) contact-probability maps across a 0.1 nm^{-1} window centered at the maximum of $p(TP|x)$ along $1/r_{bp,opt}$ for each sequence. Red indicates greater base-pairing probability and closer distance in the transition-path ensemble and vice versa for blue.

Figure 5.16 highlights the differences between transition paths and non-reactive paths at the peak of $p(TP|x)$ along $1/r_{bp,opt}$ for each sequence. For sequences where the $p(TP|x)$ maximum is close to 0.5, the barrier maximum contains nearly an equal number of simulation frames from transition paths and non-reactive trajectories, meaning there is an equal chance for (de)hybridization to fail or succeed from this point. The number of in-register base pairs is nearly identical between transition paths and non-reactive paths ($\Delta n_{bp} \sim 0$), but the arrangement of base pairs can differ significantly. For instance, the two-base-pair shifted configuration of CCends is more likely to occur in transition paths than non-reactive paths. In GCends, the shifted configuration with 5′ overhangs is more likely in transition paths while the flipped configuration

with 3′ overhangs is more significant in non-reactive paths. There can also be different preference for in-register contacts as observed in CGCcap, GCGcore, and GCmix. Overall, the difference between successful and unsuccessful barrier crossing is highly sequence-dependent.

Lastly, we compare the base-pairing properties of transition paths and non-reactive trajectories on the single-strand and duplex sides of the barrier. These difference maps give insight into the early events that may enable a (de)hybridization event to be successful. For hybridization, we examine a bin centered from 0.55 – 0.7 nm^{-1} depending on the sequence (Fig. 5.17a). Transition paths generally contain more base pairing than non-reactive paths at this point, with a particular preference for G:C base pairs in CGCcap, CCends, GCGcore, and GCmix. In CGCcap and CCends, the increased probability of binding at the 5′-CGC or 5′-CC ends coincides with greater separation between nucleobase at the other end of the oligonucleotides. Out-of-register contacts also promote successful hybridization in CGCcap (as noted in Fig. 5.15) and GCends, potentially because they keep strands in close proximity.

While successful hybridization is promoted by early formation of key base-pair contacts, dehybridization is most successful after early disruption of in-register base pairs. Figure 5.17b compares contact-separation and contact-probability maps at a bin centered from 1.15 to 1.20 nm^{-1}. Out-of-register CCends and GCends (5′ or 3′) duplex configurations are more probable to successfully dehybridize, likely due to their lower binding stability relative to an in-register duplex. In GCGcore and GCmix, transition paths exhibit a reduction of base pairing in the duplex center and G:C sites, yet an increase at the A:T termini relative to non-reactive paths. This comparison highlights that non-reactive paths in this region along $1/r_{bp,opt}$ contain significant terminal fraying

while transition paths involve unzipping of the G:C and central base pairs from one side or the other. In contrast, transition and non-reactive paths in CGCcap involve a similar amount of terminal A:T fraying, but transition paths proceed to disrupt central A:T base pairs.

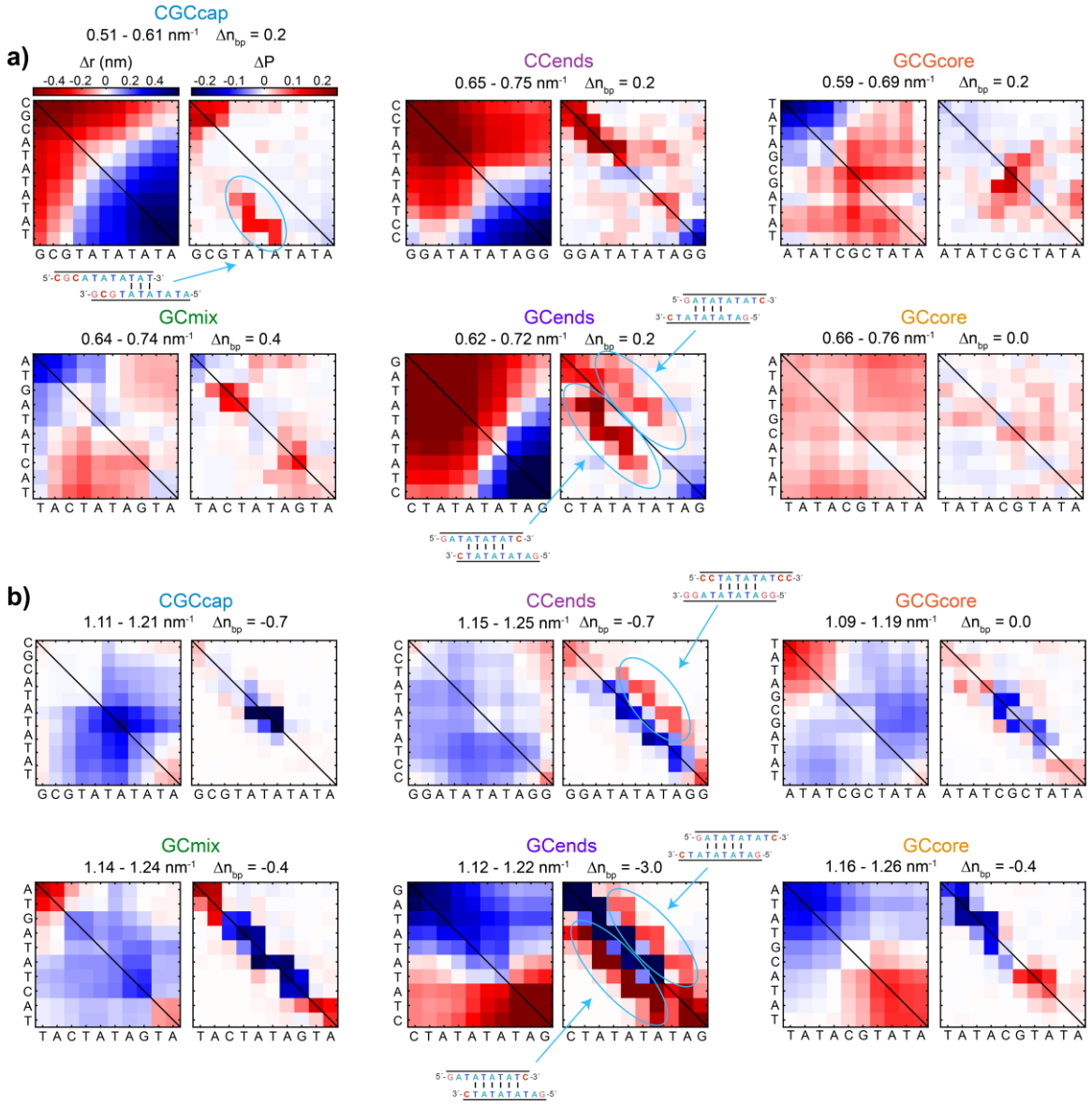

Figure 5.17 Difference between successful and failed events at their early stages. (a) Difference between transition-path and failed-trajectory (left) contact-separation and (right) contact-probability maps at a 0.1 nm^{-1} bin centered on the single-strand side (0.55 – 0.7 nm^{-1}) of the $p(TP|x)$ maximum. **(b)** Same plots for a 0.1 nm^{-1} bin centered on the duplex side (1.1 – 1.25 nm^{-1}) of the $p(TP|x)$ maximum.

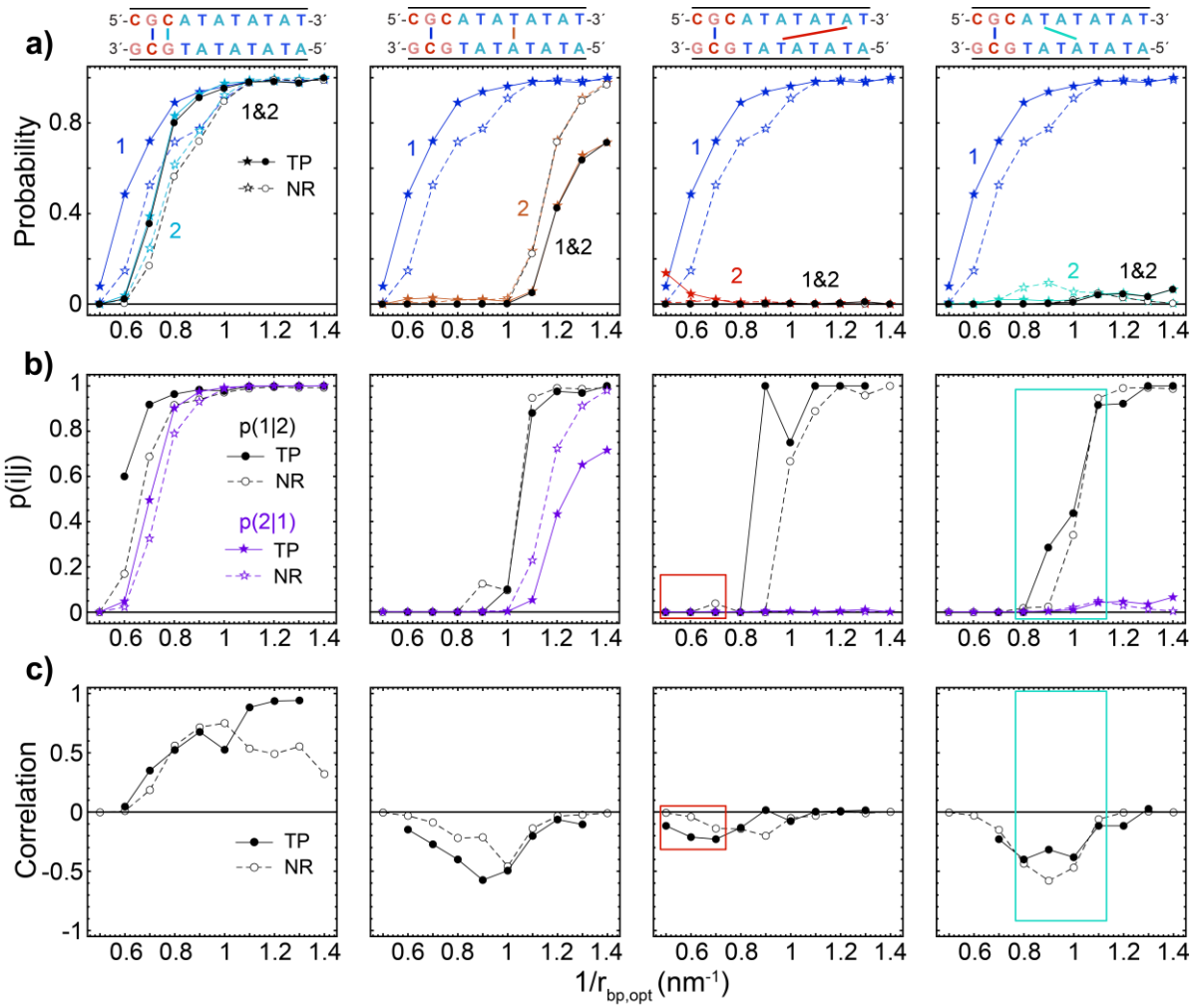

Figure 5.18 Relationship between select base-pair contacts during hybridization of CGCcap. **(a)** Probability of base-pair formation for contacts marked in the sequence schematic as a function of 0.1 nm^{-1} bins along $1/r_{bp,opt}$ for CGCcap. Black data points correspond to the joint probability of each base pair being intact. Data are shown for transition paths (TP, filled marker, solid line) and non-reactive trajectories (NR, open marker, dashed line). **(b)** Conditional probabilities that base pair i is intact given that base pair j is intact, $p(i|j)$. Conditional probabilities are shown between contacts marked 1 and 2 in panel **(a)**. **(c)** Correlation between the base-pair formation in marked bases at each bin along $1/r_{bp,opt}$ (eq. 5.5). Adjacent in-register base pairs show positive correlation at all stages of hybridization while those many sites apart have negative correlation in the early phases of hybridization. Popular out-of-register contacts noted in Figs. 5.16 and 5.17 are negatively correlated with formation of in-register contacts at the G:C-rich end.

The contact maps in Figs. 5.15 – 5.17 suggest both beneficial and detrimental roles of out-of-register base pairing during hybridization, but the maps alone do not describe the full base-pairing structure of these configurations. For example, transition paths of CGCcap show a preference relative to non-reactive events for out-of-register A:T base pairing and in-register G:C base pairing at $1/r_{bp,opt} < 0.75$ nm^{-1}, but it is unclear whether these correspond to a single binding geometry or two distinct configurations. To gain insight into the relationship between base pairing regions, we examine the conditional probability that a base pair i is intact given that another base pair j is intact, $p(i|j)$, as well as the correlation (C) between contacts over each interval of $1/r_{bp,opt}$ (Δx).

$$C\left(i, j, \Delta x\right) = \frac{\left\langle \Delta q_i \cdot \Delta q_j \right\rangle_{\Delta x}}{\sqrt{\left\langle \Delta q_i \right\rangle_{\Delta x}^2} \cdot \sqrt{\left\langle \Delta q_j \right\rangle_{\Delta x}^2}} \tag{5.5}$$

$$\text{where } \Delta q = q - \left\langle q \right\rangle_{\Delta x}$$

The variable q corresponds to a value of 1 or 0 depending on whether the base pair is intact or broken, respectively. Figure 5.18 shows conditional probabilities and correlations between four different contact pairs for CGCcap. The contact closer to the 5′ terminus of the top strand is denoted contact 1 and the other contact 2. Adjacent in-register G:C contacts show a similar increase in each conditional probability, the probability that base-pair 1 is intact given that base-pair 2 is intact, $p(1|2)$, and vice versa, $p(2|1)$. Both $p(1|2)$ and $p(2|1)$ approach unity at $1/r_{bp,opt}$ = 1.0 nm^{-1} and a similar increase is observed in the correlation and joint probability of base-pairing, and such

characteristics indicate that these adjacent contacts tend to form and break together. The second G:C base pair and an A:T base pair on the other duplex half show a negative correlation at low-$1/r_{bp,opt}$ bins and mismatch between $p(1|2)$ and $p(2|1)$. $p(1|2)$ grows at lower $1/r_{bp,opt}$ values than $p(2|1)$ because the A:T base-pair formation occurs when the G:C base pair is already intact, but the opposite is not true until $1/r_{bp,opt}$ > 1.1 nm^{-1}. Negative correlation at low-$1/r_{bp,opt}$ bins arises from a favorable entropy for opening of the A:T side upon nucleation of the G:C-rich region.

Conditional probabilities and correlation between the most abundant pairs of in-register and out-of-register base pairs help identify whether they exist together or separately during hybridization. CGCcap exhibits four-base-pair shifted A:T base pairing during transition paths from $1/r_{bp,opt}$ = 0.5 – 0.7 nm^{-1} (Fig. 5.12). These contacts show negative correlation and negligible conditional and joint probabilities with the second in-register G:C base pair (red box, Fig. 5.18), suggesting that the base pairs are formed in separate configurations. Non-reactive trajectories of CGCcap show a different two-base-pair shifted configuration from $1/r_{bp,opt}$ = 0.8 – 1.1 nm^{-1} in addition to in-register base pairing at the G:C-rich end. These contacts occur in separate configurations from 0.8 – 0.9 nm^{-1} as they show negative correlation and negligible conditional and joint probabilities. However, an increase in conditional and joint probabilities and a reduction of negative correlation from 1.0 – 1.1 nm^{-1} suggests the presence of configurations with each set of contacts. We also examine the relationship between in-register and the two-base-pair shifted configuration configurations in CCends and GCends (Fig. 5.19). As introduced in Chapter 1.3, CCends may transition from out-of-register to fully in-register configurations through structures that contain both types of base pairs, which is supported by the 0.2 – 0.5 conditional probability

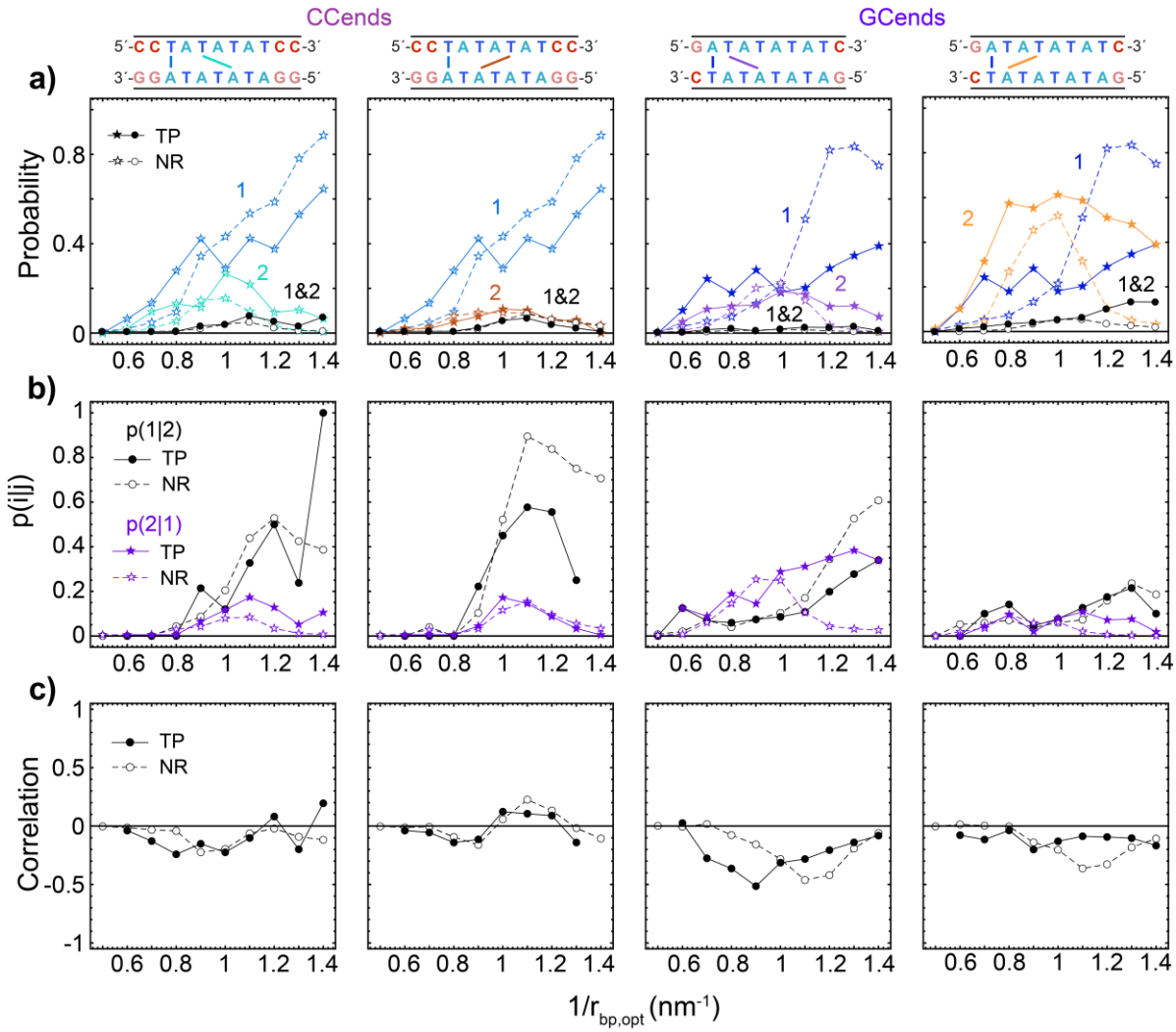

Figure 5.19 Relationship between in-register and out-of-register base-pair contacts during hybridization of CCends and GCends. (a) Probability of base-pair formation for contacts marked in the sequence schematic as a function of 0.1 nm^{-1} bins along $1/r_{bp,opt}$ for CCends and GCends. Black data points correspond to the joint probability of each base pair being intact. Data are showed for transition paths (TP, filled marker, solid line) and non-reactive trajectories (NR, open marker, dashed line). **(b)** Conditional probabilities that base pair i is intact given that base pair j is intact, $p(i|j)$. Conditional probabilities are shown between contacts marked 1 and 2 in panel **(a)**. **(c)** Correlation between the base-pair formation in marked bases at each bin along $1/r_{bp,opt}$ (eq. 5.5).

and peak in joint probability between contacts from 0.9 – 1.2 nm^{-1}. Although out-of-register configurations are much more stable in GCends relative to CCends, there is lower probability to

form mixed in-register/out-of-register configurations. Overall, our analysis indicates that out-of-register and mixed in-register/out-of-register configurations can play beneficial and detrimental roles in the dynamics of hybridization.

5.4 Acknowledgements

I would like to thank Michael Jones for carrying out the 3SPN.2 and all-atom MD simulations used in this thesis. I am also grateful to Michael Jones and Prof. Andrew Ferguson for helpful discussions regarding many of the analysis techniques employed in this chapter. I thank Seung Yeon Lee, Nicholas Lewis, and Abhirup Guha for providing feedback on the chapter.

5.5 References

1. Kim, J.-Y.; Chung, H. S., Disordered proteins follow diverse transition paths as they fold and bind to a partner. *Science* **2020,** *368*, 1253-1257.

2. Hoffer, N. Q.; Woodside, M. T., Probing microscopic conformational dynamics in folding reactions by measuring transition paths. *Curr. Opin. Chem. Biol.* **2019,** *53*, 68-74.

3. Chung, H. S., Transition path times measured by single-molecule spectroscopy. *J. Mol. Biol.* **2018,** *430*, 409-423.

4. Neupane, K.; Wang, F.; Woodside, M., Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 1329-1334.

5. Hoffer, N. Q.; Neupane, K.; Woodside, M. T., Observing the base-by-base search for native structure along transition paths during the folding of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U.S.A.* **2021,** *118*, e2101006118.

6. Hoffer, N. Q.; Neupane, K.; Woodside, M. T., Measuring the average shape of transition paths during the folding of a single biological molecule. *Proc. Natl. Acad. Sci. U.S.A.* **2019,** *116*, 8125-8130.

7. Truex, K.; Chung, H. S.; Louis, J. M.; Eaton, W. A., Testing landscape theory for biomolecular processes with single molecule fluorescence spectroscopy. *Phys. Rev. Lett.* **2015,** *115*, 018101.

8. Neupane, K.; Hoffer, N. Q.; Woodside, M., Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* **2018,** *121*, 018102.

9. Wang, H.; Li, B.; Kim, Y.-J.; Kwon, O.-H.; Granick, S., Intermediate states of molecular self-assembly from liquid-cell electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 1283-1292.

10. Sponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P., RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **2018,** *118*, 4177-4338.

11. Tucker, M. R.; Piana, S.; Tan, D.; LeVine, M. V.; Shaw, D. E., Development of force field parameters for the simulation of single-and double-stranded DNA molecules and DNA–protein complexes. *J. Phys. Chem. B* **2022,** *126*, 4442-4457.

12. Love, O.; Galindo-Murillo, R.; Zgarbová, M.; Šponer, J. i.; Jurečka, P.; Cheatham III, T. E., Assessing the Current State of Amber Force Field Modifications for DNA─ 2023 Edition. *J. Chem. Theory Comput.* **2023,** *19*, 4299-4307.

13. Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M., How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* **2017,** *45*, 4217-4230.

14. Oweida, T. J.; Kim, H. S.; Donald, J. M.; Singh, A.; Yingling, Y. G., Assessment of AMBER force fields for simulations of ssDNA. *J. Chem. Theory Comput.* **2021,** *17*, 1208-1217.

15. Kim, H.; Pak, Y., Improving all-atom force field to accurately describe DNA G-quadruplex loops. *J. Phys. Chem. B* **2022,** *126*, 6199-6209.

16. Zgarbová, M.; Jurečka, P.; Sponer, J.; Otyepka, M., A-to B-DNA transition in AMBER force fields and its coupling to sugar pucker. *J. Chem. Theory Comput.* **2018,** *14*, 319-328.

17. Zgarbová, M.; Sponer, J.; Jurecka, P., Z-DNA as a touchstone for additive empirical force fields and a refinement of the alpha/gamma DNA torsions for AMBER. *J. Chem. Theory Comput.* **2021,** *17*, 6292-6301.

18. Piomponi, V.; Fröhlking, T.; Bernetti, M.; Bussi, G., Molecular simulations matching denaturation experiments for N6-Methyladenosine. *ACS Cent. Sci.* **2022,** *8*, 1218-1228.

19. Xiao, S.; Sharpe, D. J.; Chakraborty, D.; Wales, D. J., Energy landscapes and hybridization pathways for DNA hexamer duplexes. *J. Phys. Chem. Lett.* **2019,** *10*, 6771-6779.

20. Yang, C.; Kulkarni, M.; Lim, M.; Pak, Y., In silico direct folding of thrombin-binding aptamer G-quadruplex at all-atom level. *Nucleic Acids Res.* **2017,** *45*, 12648-12656.

21. Ouldridge, T. E.; Šulc, P.; Romano, F.; Doye, J. P.; Louis, A. A., DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **2013,** *41*, 8886-8895.

22. Ouldridge, T. E.; Louis, A. A.; Doye, J. P., Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011,** *134*, 02B627.

23. Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J., An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013,** *139*, 10B604_1.

24. Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J., Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* **2014,** *141*, 07B613_1.

25. Chakraborty, D.; Hori, N.; Thirumalai, D., Sequence-dependent three interaction site model for single-and double-stranded DNA. *J. Chem. Theory Comput.* **2018,** *14*, 3763-3779.

26. Markegard, C. B.; Fu, I. W.; Reddy, K. A.; Nguyen, H. D., Coarse-grained simulation study of sequence effects on DNA hybridization in a concentrated environment. *J. Phys. Chem. B* **2015,** *119*, 1823-1834.

27. Jones, M. S.; Ashwood, B.; Tokmakoff, A.; Ferguson, A. L., Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, markov state models, and infrared spectroscopy. *J. Am. Chem. Soc.* **2021,** *143*, 17395-17411.

28. Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A., DNA duplex formation with a coarse-grained model. *J. Chem. Theory Comput.* **2014,** *10*, 5020-5035.

29. Schreck, J. S.; Ouldridge, T. E.; Romano, F.; Šulc, P.; Shaw, L. P.; Louis, A. A.; Doye, J. P., DNA hairpins destabilize duplexes primarily by promoting melting rather than by inhibiting hybridization. *Nucleic Acids Res.* **2015,** *43*, 6181-6190.

30. Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **2013,** *41*, 10641-10658.

31. Snodin, B. E.; Romano, F.; Rovigatti, L.; Ouldridge, T. E.; Louis, A. A.; Doye, J. P., Direct simulation of the self-assembly of a small DNA origami. *ACS Nano* **2016,** *10*, 1724-1737.

32. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. *Proc. Natl. Acad. Sci. U.S.A.* **2023,** *120*, e2219124120.

33. Jones, M. S.; McDargh, Z. A.; Wiewiora, R. P.; Izaguirre, J. A.; Xu, H.; Ferguson, A. L., Molecular Latent Space Simulators for Distributed and Multimolecular Trajectories. *J. Phys. Chem. A* **2023,** *127*, 5470-5490.

34. Manning, G. S., Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. *J. Chem. Phys.* **1969,** *51*, 924-933.

35. Nkodo, A. E.; Garnier, J. M.; Tinland, B.; Ren, H.; Desruisseaux, C.; McCormick, L. C.; Drouin, G.; Slater, G. W., Diffusion coefficient of DNA molecules during free solution electrophoresis. *Electrophoresis* **2001,** *22*, 2424-2432.

36. Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F., Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **2017,** *9*, 1005-1011.

37. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013,** *139*, 015102.

38. Pinamonti, G.; Paul, F.; Noé, F.; Rodriguez, A.; Bussi, G., The mechanism of RNA base fraying: Molecular dynamics simulations analyzed with core-set Markov state models. *J. Chem. Phys.* **2019,** *150*, 154123.

39. Mardt, A.; Pasquali, L.; Wu, H.; Noé, F., VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018,** *9*, 5.

40. Chen, W.; Sidky, H.; Ferguson, A. L., Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019,** *150*, 214114.

41. Chodera, J. D.; Pande, V. S., Splitting probabilities as a test of reaction coordinate choice in single-molecule experiments. *Phys. Rev. Lett.* **2011,** *107*, 098102.

42. Best, R. B.; Hummer, G., Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 1088-1093.

43. Zheng, W.; Best, R. B., Reduction of all-atom protein folding dynamics to one-dimensional diffusion. *J. Phys. Chem. B* **2015,** *119*, 15247-15255.

44. Neupane, K.; Manuel, A. P.; Woodside, M. T., Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat. Phys.* **2016,** *12*, 700-703.

45. Neupane, K.; Manuel, A. P.; Lambert, J.; Woodside, M. T., Transition-path probability as a test of reaction-coordinate quality reveals DNA hairpin folding is a one-dimensional diffusive process. *J. Phys. Chem. Lett.* **2015,** *6*, 1005-1010.

46. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.

47. Poland, D.; Scheraga, H. A., *Theory of helix-coil transitions in biopolymers*. Academic Press.: 1970.

48. Chandrasekhar, S., Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **1943,** *15*, 1.

49. Araque, J.; Robert, M., Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* **2016,** *144*, 125101.

50. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.

51. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

52. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

53. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008,** *100*, 020603.

54. Sanstead, P. J.; Tokmakoff, A., A lattice model for the interpretation of oligonucleotide hybridization experiments. *J. Chem. Phys.* **2019,** *150*, 185104.

55. Hummer, G., From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **2004,** *120*, 516-523.

56. Best, R. B.; Hummer, G., Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 6732-6737.

57. Best, R. B.; Hummer, G.; Eaton, W. A., Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 17874-17879.

58. Robustelli, P.; Piana, S.; Shaw, D. E., Mechanism of coupled folding-upon-binding of an intrinsically disordered protein. *J. Am. Chem. Soc.* **2020,** *142*, 11092-11101.

59. Berezhkovskii, A. M.; Makarov, D. E., On distributions of barrier crossing times as observed in single-molecule studies of biomolecules. *Biophys. Rep.* **2021,** *1*, 100029.

60. Makarov, D. E., Barrier crossing dynamics from single-molecule measurements. *J. Phys. Chem. B* **2021,** *125*, 2467-2476.

61. Berezhkovskii, A. M.; Makarov, D. E., Communication: Transition-path velocity as an experimental measure of barrier crossing dynamics. *J. Chem. Phys.* **2018,** *148*, 201102.

62. Dickman, R.; Manyanga, F.; Brewood, G. P.; Fish, D. J.; Fish, C. A.; Summers, C.; Horne, M. T.; Benight, A. S., Thermodynamic contributions of 5'-and 3'-single strand dangling-ends to the stability of short duplex DNAs. *J. Biophys. Chem.* **2012,** *3*, 1-15.

63. Bommarito, S.; Peyret, N.; SantaLucia Jr, J., Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **2000,** *28*, 1929-1934.

64. SantaLucia Jr, J.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-440.

65. Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E., Atomic-level characterization of protein–protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2019,** *116*, 4244-4249.

66. Blöchliger, N.; Xu, M.; Caflisch, A., Peptide binding to a PDZ domain by electrostatic steering via nonnative salt bridges. *Biophys. J.* **2015,** *108*, 2362-2370.

67. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.

68. Todisco, M.; Szostak, J. W., Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Res.* **2022,** *50*, 9647-9662.

69. Phan, T. T.; Phan, T. M.; Schmit, J. D., Beneficial and detrimental effects of non-specific binding during DNA hybridization. *Biophys. J.* **2023,** *122*, 835-848.

Appendix 5.A Comparison of $1/r_{bp,opt}$ to TICA and SRV CVs

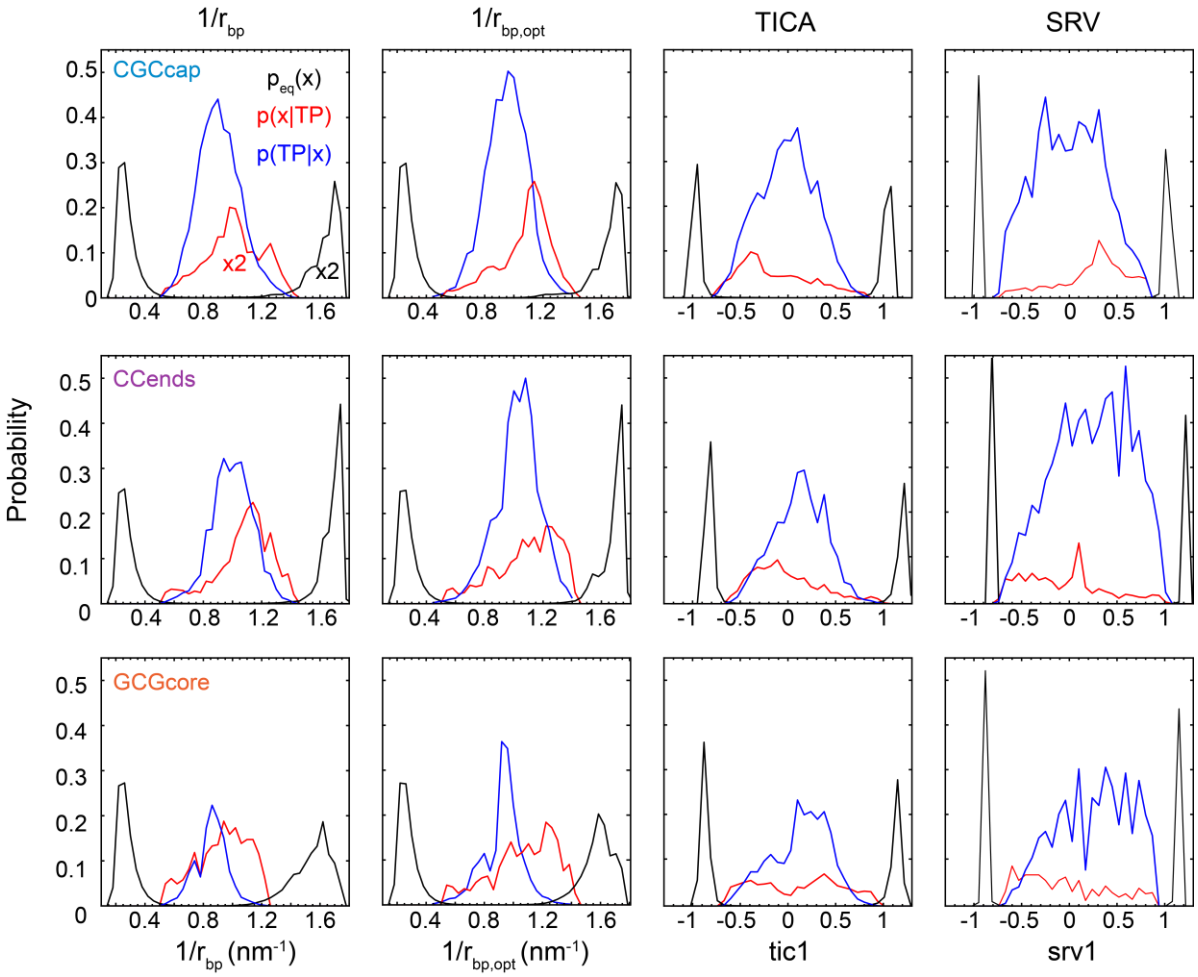

Figure 5.A1 Transition-path test of TICA and SRV CVs. p_{eq}, $p(x|TP)$, and $p(TP|x)$ calculated along $1/r_{bp}$, $1/r_{bp,opt}$, the leading TICA CV (tic1), and the leading SRV CV (srv1). p_{eq} and $p(x|TP)$ are vertically scaled by a factor of 2 for clarity.

The $1/r_{bp,opt}$ CV is obtained by optimization of $p(TP|x)$ maximum as described in Section 5.2.3, and here we compare with other CVs that are optimized for describing the kinetics of hybridization. In one approach, time-lagged independent component analysis (TICA)[37] is used to project trajectories of all inverse interbase distances between strands onto a low-dimensional

space where each CV is a linear combination of input distances. In another method, state-reversible VAMPnets (SRV),[40] the leading CV is constructed from a non-linear combination of inverse distances. In each method, the low-dimensional space is optimized to approximate the eigenvectors associated with the slowest eigenvalues of the transfer operator for propagating molecular dynamics. The leading CV (tic1 or srv1) is associated with the slowest timescale dynamics in the system, which is hybridization for our DNA oligonucleotides. Previous studies have found that the leading SRV CV describes transitions between the single-strand and duplex states while additional CVs capture out-of-register shifting or large-scale fraying.[27, 32]

The probability distributions along tic1 and srv1 appear quite different from $1/r_{bp,opt}$. The single-strand and duplex basins p_{eq} are each narrow relative to $1/r_{bp,opt}$ because fraying and out-of-register shifting are associated with higher-order TICA and SRV CVs (tic2, tic3, etc.). As a result, most of the space along tic1 and srv1 corresponds to the transition region between duplex and single-strand states, and the $p(TP|x)$ distributions are broad. Along tic1, the $p(TP|x)$ maximum is similar to those along the uniformly-weighted $1/r_{bp}$ CV. The $p(TP|x)$ maximum along srv1 reaches values comparable to $1/r_{bp,opt}$, but the maxima are broad and noisy due to limited sampling. Since tic1 and srv1 are optimized with distinct criteria from $1/r_{bp,opt}$, it is not surprising that their apparent quality based on $p(TP|x)$ is different.

Chapter 6

Disruption of base-pairing cooperativity in DNA duplexes by an abasic site

The material for this chapter is adapted from:

Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. *Proc. Natl. Acad. Sci. U.S.A.* **2023,** *120,* e2219124120.

6.1 Introduction

Numerous modifications to canonical base pairs, including epigenetic chemical modifications, mismatches, and lesions, alter local base pairing interactions in double-stranded DNA which may in turn shift global duplex thermodynamic stability by interrupting cooperative interactions. Cooperativity refers to the non-additive energetic advantages of forming base pairs in a stretch of multiple adjacent nucleotides concertedly rather than individually.[1] This phenomenon underlies the quantitatively accurate nearest neighbor (NN) models of DNA thermodynamic stability.[2-3] While cooperativity in base pairing is understood to be rooted in the stacking interactions between nucleobases, the molecular factors bridging local perturbations and global destabilization are not fully understood. Identifying these factors require site-specific characterization of the hybridization free-energy landscape and duplex structural dynamics over wide time windows. This motivates the combination of temperature-jump infrared (T-jump IR) spectroscopy and coarse-grained molecular dynamics (MD) simulations that are employed in this

chapter to access thermodynamic, kinetic, and dynamic insight into how an abasic site disrupts base pairing and stacking in short oligonucleotides.

The loss of a nucleobase to form an abasic site (apurinic/apyrimidinic or AP site) is among the most naturally abundant DNA modifications due to spontaneous or enzymatic cleavage of the glycosidic bond.[4-5] Previous work showed that an AP site destabilized the DNA duplex in a sequence-dependent fashion[6-11] that was correlated with the identity of base pairs adjacent to the modified site; however, an AP site may also disrupt base pairing well beyond nearest neighbors.[12] Since duplexes containing an AP site retain a B-form conformation on average,[13-15] an AP site influences duplex stability and dynamics through disruption of cooperative interactions rather than a large-scale structural change.

We characterize the disruption of cooperativity in various sequence contexts by investigating the impact of central AP sites in three 11-mer template oligonucleotides that are expected to exhibit wide variation in dynamic behavior due to different relative placement of G:C and A:T base pairs.[16-22] T-jump IR spectroscopy separately monitors changes in A:T and G:C base pairing from nanoseconds to milliseconds following a laser T-jump that induces duplex melting. MD simulations employing the 3-site-per-nucleotide (3SPN.2) coarse-grained model[16, 23] are used to parameterize a kinetic Markov state model (MSM)[24-25] that allows us to interpret the experiments in terms of the temperature-induced reshaping of the duplex free-energy landscape. Our results indicate that duplex destabilization primarily stems from the loss of base pairing and stacking interactions around the AP site as well as disruption of base-pairing cooperativity within the duplex. The AP site splits the base pairing dynamics of the duplex into two segments that must

each overcome a significant free-energy barrier to hybridize. Although our study is limited to oligonucleotides, length-dependent trends in duplex destabilization behavior enable us to address the impact of an AP site in biological DNA.

6.2 Destabilization of the DNA duplex from an abasic site

6.2.1 Experimental characterization of duplex thermodynamic stability

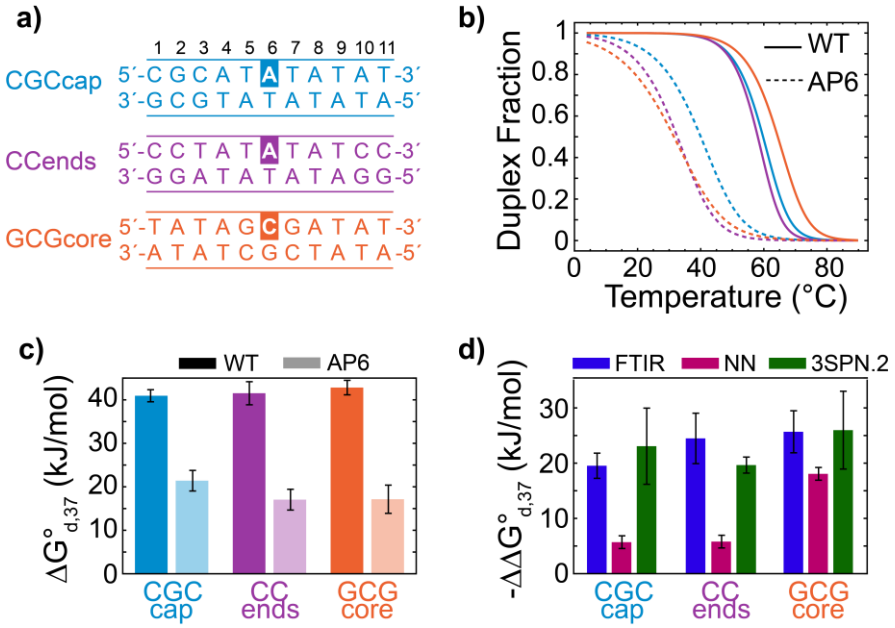

Figure 6.1 DNA duplex destabilization induced by an AP site. (a) CGCcap, CCends, and GCGcore sequences studied in this work, identifying the 6-position of the AP site along the top strand (shaded). **(b)** Duplex melting curves determined from two-state analysis of an FTIR temperature series for WT (solid lines) and AP6 (dashed lines) sequences (Appendix 6.A). **(c)** Duplex dehybridization free energy at 37°C ($\Delta G^{\circ}_{d,37}$) determined from FTIR data for WT (dark) and AP6 (light) sequences. **(d)** Drop in $\Delta G^{\circ}_{d,37}$ for each AP6 sequence with respect to WT sequence, $\Delta\Delta G^{\circ}_{d,37} = \Delta G^{\circ}_{d,37}(AP) - \Delta G^{\circ}_{d,37}(WT)$, compared to those determined from the NN model. Also shown are the change in dehybridization Helmholtz free energy ($\Delta\Delta F^{\circ}_{d,37}$) obtained from melting curves generated from 3SPN.2 MD simulations. FTIR and 3SPN.2 error bars correspond to 95% confidence intervals from fits to a two-state model (Fig. 6.2) while NN error bars correspond to the reported error in the model.[2]

The DNA sequences shown in Fig. 6.1a were chosen to vary the position of three or four stabilizing G:C base pairs within an otherwise A:T sequence. Sequence "CGCcap" places them on one end to create asymmetric base-pairing stability, "CCends" places them on both ends to minimize terminal base pair fraying,[19] and "GCGcore" centers them to promote frayed A:T terminal base pairs.[19-20, 26] For each oligonucleotide, we compare the hybridization of the wild-type (WT) sequence to its complement with one containing a central AP site at the 6th base position (AP6). DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) at desalt-grade purity. AP sites were incorporated as a tetrahydrofuran group (dSpacer). Samples were purified with 3 kDa centrifugal filters (Amicon). All oligonucleotides were prepared in pH* 6.8 400 mM sodium phosphate buffer (SPB, [Na^+] = 600 mM).

Experiments showing how a central AP site destabilizes the duplex are presented in Figs. 6.1 and 6.2. Duplex melting curves obtained by Fourier-transform infrared spectroscopy (FTIR) show a 20–30 °C reduction in melting temperature (T_m) that equates to a drop of 20-30 kJ/mol in the dehybridization free energy between AP6 and WT sequences ($\Delta\Delta G^\circ_{d,37} = \Delta G^\circ_{d,37}(AP) - \Delta G^\circ_{d,37}(WT)$), consistent with previous studies.[6, 8, 11] Calculations of $\Delta\Delta G^\circ_{d,37}$ using Santa Lucia's NN model[2, 27] merely removing the two central NN interaction parameters involved with the AP site only accounts for a portion of the experimental value, suggesting that duplex destabilization is not simply additive in the NN energies (Fig. 6.1c).[28] An additional free-energy penalty affecting both the dehybridization enthalpy (ΔH°_d) and entropy (ΔS°_d) must arise from the AP site (Fig. 6.2). We also find that $\Delta\Delta G^\circ_{d,37}$ is ~5 kJ/mol larger for CCends-AP6 than CGCcap-AP6 even though the

modified and adjacent base pairs to the AP site are identical in each sequence, suggesting that the oligonucleotide sequence further from the AP site contributes to $\Delta\Delta G^{\circ}_{d,37}$.

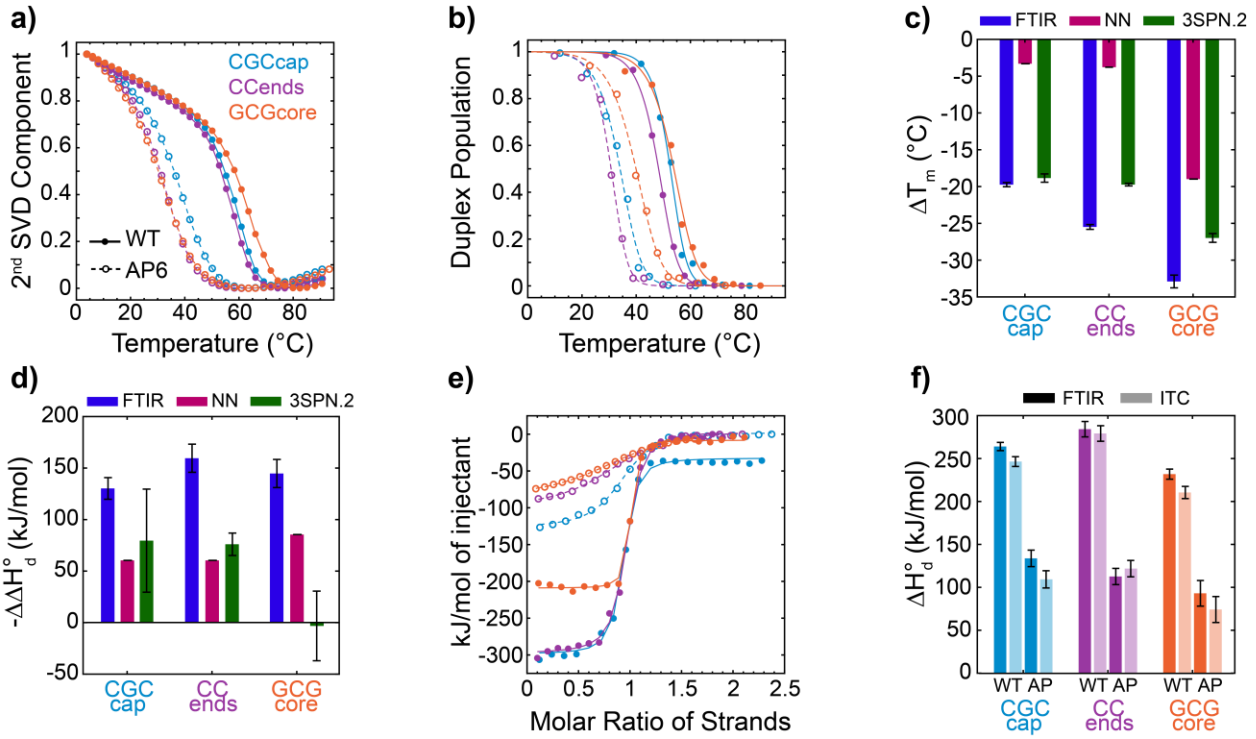

Figure 6.2 DNA duplex stability measured with FTIR temperature series and ITC. (a) Normalized 2nd SVD components from FTIR temperature series for each WT and AP6 sequence. **(b)** Temperature-dependent fraction of intact duplexes from 3SPN.2 MD simulations employing WTmetaD. Data are fit to two-state thermodynamic model (solid and dashed lines, Appendix 6.A). **(c)** Difference in T_m between AP6 and WT sequences (ΔT_m) from FTIR melting curves, NN model, and 3SPN.2 melting curves. **(d)** Difference in dehybridization enthalpy (ΔH°_d) between AP6 and WT sequences ($\Delta\Delta H^{\circ}_d$) from FTIR melting curves and NN model. Values from 3SPN.2 melting curves correspond to the change in dehybridization internal energy $\Delta\Delta U^{\circ}_d$. **(e)** isothermal titration calorimetry (ITC) thermograms plotted as molar ratio between complement DNA strands. Data are fit to a two-state model (solid and dashed lines, Appendix 6.B). ITC sample conditions are listed in Table 6.B1. **(f)** Comparison of ΔH°_d determined from FTIR melting curves (dark) and ITC (light). The measured change in heat capacity between dissociated and hybridized states (ΔC_p) of WT sequences is applied to the FTIR-determined ΔH°_d values to account for differences between T_m and the ITC temperature. Error bars for all FTIR, ITC, and 3SPN.2 parameters indicate 95% confidence intervals from two-state model fits. Error bars for NN values correspond to the previously reported error in the NN model.

6.2.2 Base-pairing free-energy landscape

Molecular dynamics simulations with the 3SPN.2 model and well-tempered metadynamics (WTMetaD) were used to gain insight into the molecular factors governing the thermodynamic stability of various oligonucleotide structures.[28] Structurally 3SPN.2 predicts that duplexes with the AP site are virtually unchanged from the B-form conformation of the WT duplex and the nucleobase opposite the AP site is predominantly intrahelical, consistent with previous experimental studies and all-atom MD simulations.[13-14, 28] The extrahelical configuration of GCGcore-AP6 is negligible, yet there is minor population for CCends-AP6 (0.3%) and CGCcap-AP6 (0.8%). This sequence-dependence presumably arises from the 2-fold larger cross-stacking energy between guanine nucleobases relative to thymines in 3SPN.2.[23] The stability of each duplex was assessed using simulations as a function of temperature to compute duplex-fraction melting curves and the Helmholtz free energy for dehybridization $\Delta\Delta F^{\circ}_{d,37}$ (Figs. 6.1d, 6.3). The simulations show a similar destabilization by the AP site as in experiment, suggesting that the coarse-grained model is properly accounting for the duplex destabilization by the AP site. We note that the simulated sequence-dependence of $\Delta\Delta F^{\circ}_{d,37}$ differs from the experimental $\Delta\Delta G^{\circ}_{d,37}$, but any differences lie within the uncertainty of the data. Previous studies have verified the accuracy of 3SPN.2 for predicting T_m in canonical duplexes with near-physiological salt concentrations ([Na^+] = 120 mM),[23] but we find that they are consistently 5-10 °C lower than experimental values, which is likely due to the high salt concentrations ([Na^+] = 600 mM) used in this work.

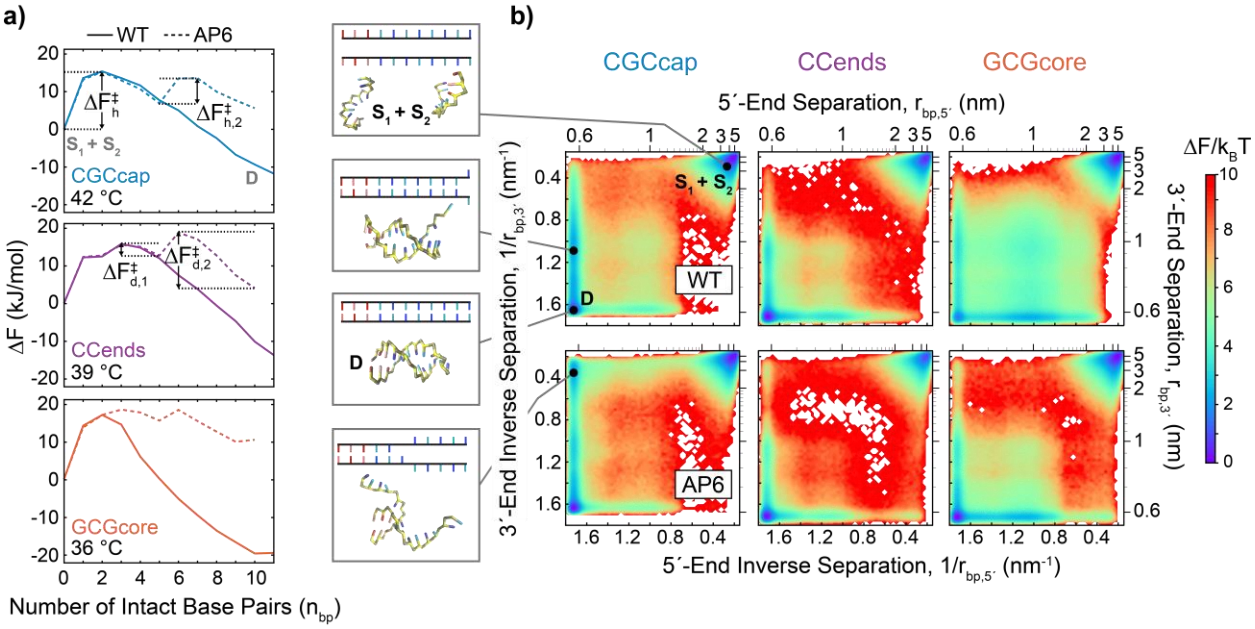

Figure 6.3 Influence of AP site on base-pairing free-energy landscape. (a) Free-energy profiles as a function of the number of intact base pairs, $\Delta F(n_{bp}) = -RT \ln P(n_{bp})/P(0)$, for WT (solid lines) and AP6 (dashed lines) sequences extracted from 3SPN.2 MD simulations with WTMetaD at temperatures 10-20°C below $T_{m,MD}$ of each WT sequence. Base pairs are determined using a 0.7 nm radial cutoff between the center-of-mass of a nucleobase site and that of its complement. Hybridization barriers in n_{bp} for forming the first ($\Delta F_h^{\ddagger}$) and second ($\Delta F_{h,2}^{\ddagger}$) base-pair segments of the AP6 sequence are indicated for CGCcap. Barriers along the reverse dehybridization direction for dissociating the second ($\Delta F_{d,2}^{\ddagger}$) and first ($\Delta F_{d,1}^{\ddagger}$) base pair segments are shown for CCends-AP6. **(b)** Free-energy surfaces for all sequences plotted as a function of the average separation between base pairs 1, 3, and 5 on the 5′-end ($r_{bp,5'}$) and 7, 9, and 11 on the 3′-end ($r_{bp,3'}$). 5′- and 3′-ends refer to the top strand in Fig. 6.1a. Free energy (ΔF) is plotted relative to the minimum of the single-strand well (S_1 + S_2). Surfaces were generated using 2.5 million frames from 25 unbiased 10 μs simulation trajectories near the 3SPN.2-determined melting temperature of each sequence and are binned and plotted on a r_{bp}^{-1} scale to better resolve different base-pairing configurations. Addition of an AP6 site creates local minima (top left and bottom corners) in the free-energy landscape corresponding to configurations with only the 5′- or 3′-end hybridized that make up the $n_{bp} = 5$ minimum in panel **(b)**.

The computed hybridization free-energy profile (FEP) as a function of the number of intact base pairs (n_{bp}) is compared for WT and AP6 sequences in Fig. 6.3a. These profiles overlap for

the early steps in hybridization (n_{bp} = 0-5 for CGCcap and CCends and n_{bp} = 0-2 for GCGcore) and span the barrier to nucleating the duplex ($\Delta F_h^{\ddagger}$) at n_{bp} = 2-3. The free energy of duplex formation continues downhill with increasing n_{bp} in the WT sequences, as expected for a two-state nucleation and zipping process,[29-31] but the AP6 sequences display a minimum at n_{bp} = 5 and must pass through an additional free-energy barrier to fully hybridize ($\Delta F_{h,2}^{\ddagger}$). The dominant base-pairing configurations of AP6 sequences at n_{bp} = 5 are those with one side of the AP site fully intact and the other side completely dissociated, indicating a stepwise process of hybridizing one segment and then the other (Fig. 6.4). The nucleobase opposite the AP site is predominantly intrahelical at n_{bp} = 5. Therefore, an addition free-energy penalty for base-pair nucleation on the second side of the AP site contributes to $\Delta\Delta G_d^o$, which is absent in the NN model.

We also observe that $\Delta F_{h,2}^{\ddagger} < \Delta F_h^{\ddagger}$, reflecting the difference in hybridization mechanism for the first and second segments. The temperature-dependence of $\Delta F_h^{\ddagger}$ and $\Delta F_{h,2}^{\ddagger}$ indicates each arises from an entropic penalty that is partially balanced by a favorable internal-energy change for base pairing (Fig. 6.5). The reductions in translational and orientational entropy for bimolecular association contribute to $\Delta F_h^{\ddagger}$ but not $\Delta F_{h,2}^{\ddagger}$, where conformational entropy of the partially bound chains dominates. For a $n_{bp} = 5$ configuration with one fully hybridized segment, the separation of the unbound segment base pairs adjacent to the AP site is more constrained than the terminal base pairs. As a result, nucleation of the second segment usually occurs by forming a base pair adjacent to the AP site, as shown below.

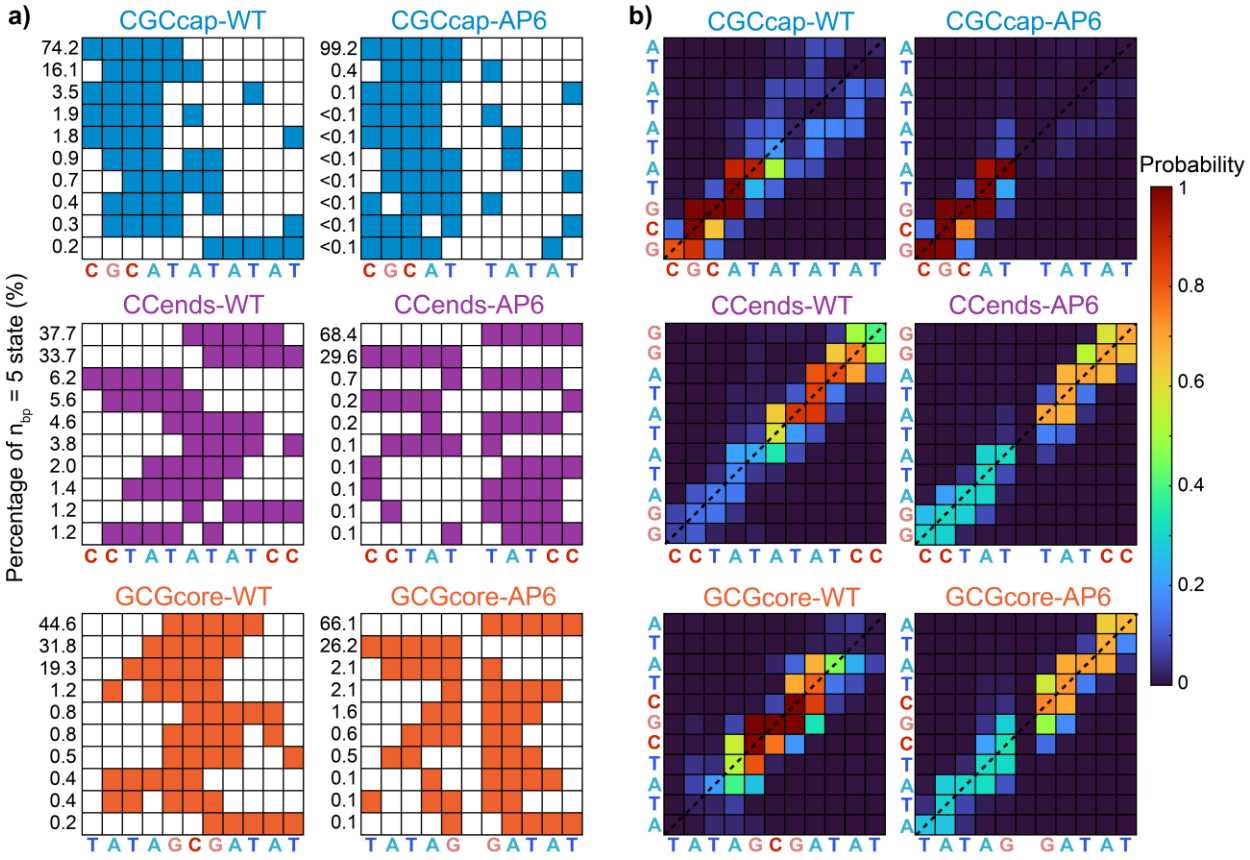

Figure 6.4 Distribution of duplex configurations containing five intact base pairs from 3SPN.2 MD simulations. (a) Ten most probable base-pairing configurations contributing to the n_{bp} = 5 state from the FEPs in Fig. 6.3a for each WT and AP6 sequence. The probability of a given configuration (%) is listed on the left. Filled squares correspond to an intact base pair and white squares indicate a broken base pair. A base-pair separation cutoff of 0.7 nm was used. **(b)** Contact-probability maps indicating the probability for each nucleobase intermolecular distance to be below 0.7 nm in the n_{bp} = 5 state. Squares along the diagonal (dashed line) correspond to probabilities for in-register distances. Contact-probability maps indicate that the n_{bp} = 5 state is dominated by configurations with in-register base-pairs.

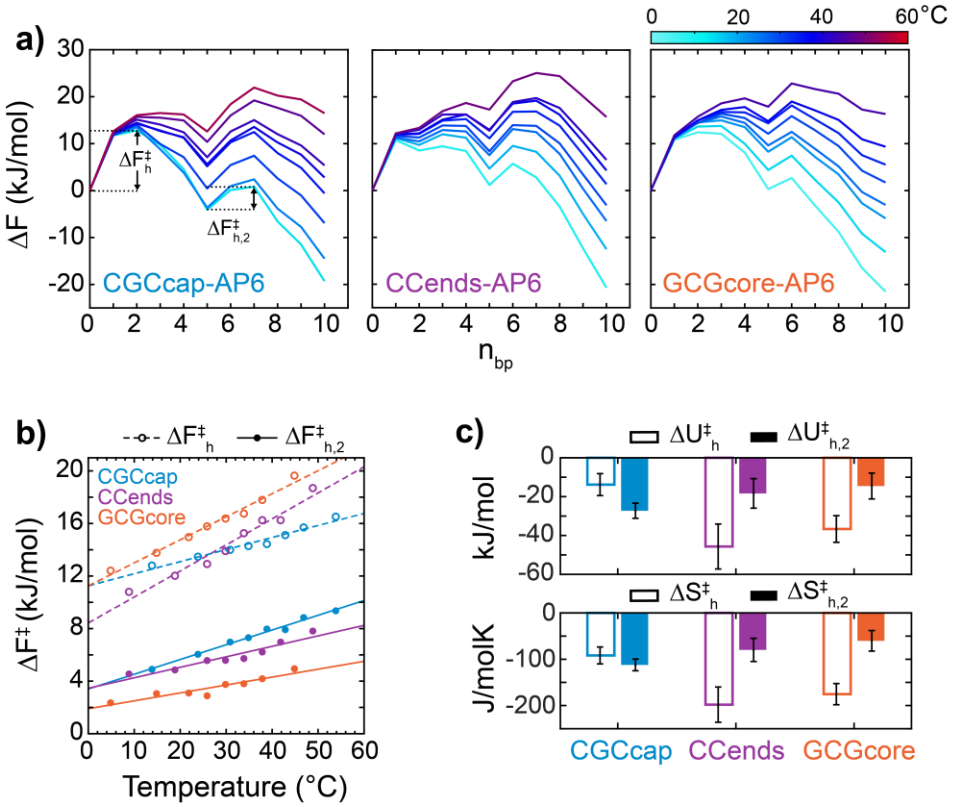

Figure 6.5 Temperature-dependence of nucleation barriers determined from 3SPN.2. (a) FEPs of AP6 sequences determined from 3SPN.2 MD simulations with WTMetaD at temperatures from $T_{m.MD}$-20°C to $T_{m.MD}$+20°C. **(b)** Temperature-dependent values of $\Delta F_h^{\ddagger}$ (open circles) and $\Delta F_{h,2}^{\ddagger}$ (filled circles) determined from **(a)**. Dashed and solid lines indicate linear fits to $\Delta F^{\ddagger} = \Delta U^{\ddagger} - T\Delta S^{\ddagger}$. **(c)** Internal energy ($\Delta U_h^{\ddagger}$ and $\Delta U_{h,2}^{\ddagger}$) and entropic ($\Delta S_h^{\ddagger}$ and $\Delta\Delta S_{h,2}^{\ddagger}$) contributions to $\Delta F_h^{\ddagger}$ and $\Delta F_{h,2}^{\ddagger}$ from fits in in **(b)**. Error bars correspond to 95% confidence intervals from fits.

Additional sequence dependent details can be observed in the free-energy surfaces (FESs) in Fig. 6.3b, which are computed at the 3SPN.2-determined melting temperature ($T_{m,MD}$) and projected into the average base pair separation (r_{bp}) on either end of the duplex. FESs are binned and plotted on a r_{bp}^{-1} scale to better resolve different basins within the duplex state. Here 5′- and 3′-sides of the duplex refer to the top strand in Fig. 6.1a. Fully intact sides have $r_{bp} < 0.7$ nm whereas

well-separated strand segments have r_{bp} = 3-5 nm. Addition of an AP6 site creates metastable configurations with only the 5′- or 3′-segment hybridized. This provides a structural rationalization for the emergence of the $\Delta F_{h,2}^{\ddagger}$ barriers observed in Fig. 6.3a. The relative stability of 5′-hybridized-to-3′-hybridized configurations qualitatively follows the expected stability of those segments predicted by the NN model.[28]

6.3 Direct observation of internal nucleation barrier

6.3.1 Segment-dehybridization kinetics from T-jump IR

We experimentally test for the presence of $\Delta F_{h,2}^{\ddagger}$ and a half-hybridized metastable state by characterizing duplex dehybridization with T-jump IR spectroscopy (described in Chapter 4).[19, 32-33] We equilibrate the sample at a temperature 15°C below T_m, where most of the DNA is duplexed and use a 15°C optical T-jump to heat the solution within 7 ns and promote duplex dehybridization. Changes in base-pairing interactions are probed at varying time delays after the T-jump using heterodyned dispersed vibrational echo spectroscopy (HDVE),[32] which provides a transient IR spectrum reporting on changes in nucleobase ring and carbonyl vibrational bands (Fig. 6.6a). The transient spectrum that we report is the change of the spectrum at a given T-jump delay relative to the initial temperature, T_i. Each nucleobase contains a distinct IR fingerprint of ring and carbonyl vibrations[34-35] that are reshaped upon base pair formation,[20, 36] providing separate probes for G:C and A:T base pairing using the guanine band at 1550 cm^{-1} and adenine band at 1605 cm^{-1}.

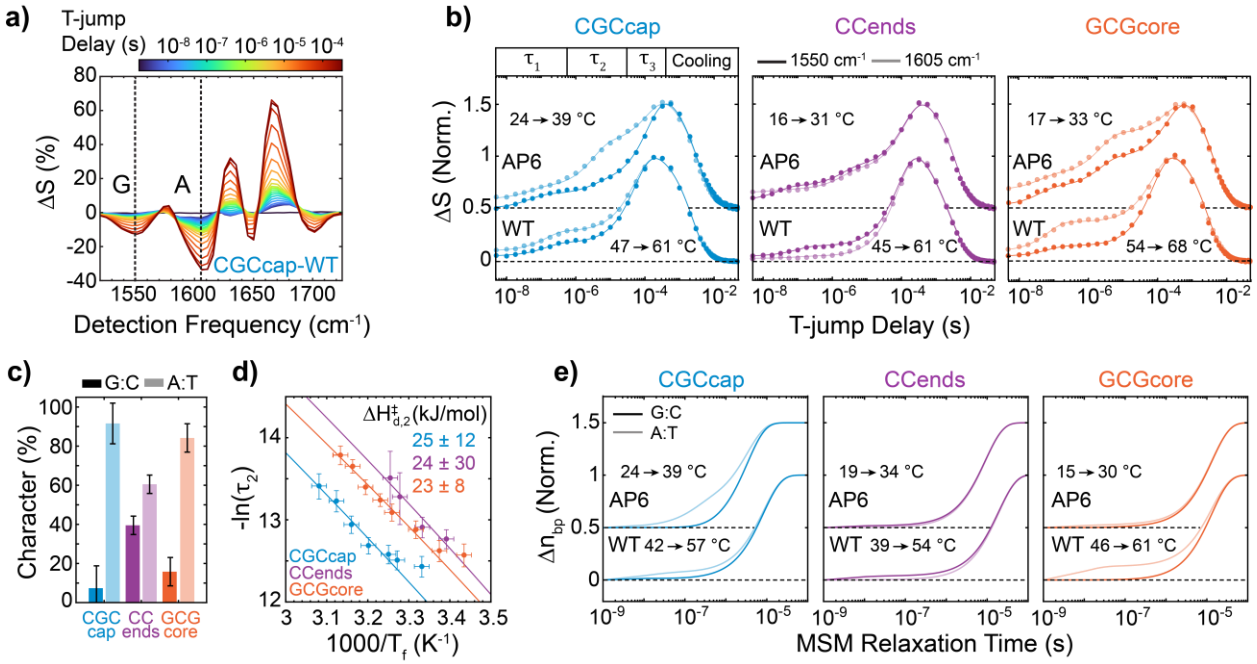

Figure 6.6 Direct observation of segment-nucleation barrier from T-jump IR spectroscopy. **(a)** T-jump IR difference (t-HDVE) spectra of CGCcap-WT from 5 ns to 1 ms delays after the T-jump, plotted as the change relative to the maximum of the spectrum before the T-jump: $\Delta S(t) = [S(t) - S(T_i)]/\max[S(T_i)]$. **(b)** Normalized $\Delta S(t)$ time traces at 1550 cm^{-1} (dark, G:C) and 1605 cm^{-1} (light, A:T) cm^{-1} plotted for the CGCcap, CCends, GCGcore sequences, showing the initial and final temperatures for each sequence ($T_i \rightarrow T_f$). T-jumps are performed from approximately T_m–15°C to T_m for each sequence. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Solid lines correspond to three- or four-component fits from global lifetime analysis. **(c)** Percentage of G:C (dark) and A:T (light) base pair loss character observed in half-dehybridization t-HDVE responses. Percentages are derived from t-HDVE amplitudes of the 1550 and 1605 cm^{-1} bands, respectively, determined from global lifetime fitting. Error bars indicate 95% confidence intervals propagated from global fits. **(d)** Temperature-dependent observed rates for half-dehybridization responses (τ_2^{-1}) with fit to a Kramers-like equation in the high-friction limit (solid lines, eq. 6.1). Vertical error bars correspond to 95% confidence intervals from global lifetime fitting and horizontal error bars are the standard deviation in T-jump magnitude. The enthalpic barrier for half-dehybridization ($\Delta H_{d,2}^{\ddagger}$) is reported. **(e)** Markov state model (MSM) T-jump simulations of the normalized change in intact G:C (dark) and A:T (light) base pairs, $\Delta n_{bp}(t) = [n_{bp}(t) - n_{bp}(T_i)]/[n_{bp}(T_f) - n_{bp}(T_i)]$. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. T-jumps were initialized from a population distribution determined at $T_i = T_{m,MD}$–15°C, projected into the

Figure 6.6 Direct observation of segment-nucleation barrier from T-jump IR spectroscopy, continued

microstates of an MSM constructed at $T_f \approx T_{m,MD}$, and allowed to relax to the ensemble determined at T_f. The presence of a base pair is based on a $r_{bp,eq} + 0.3$ nm radial cutoff between a nucleobase and its complement, where $r_{bp,eq}$ is the equilibrium base-pair separation for G:C (0.55 nm) and A:T (0.6 nm) base pairs. The relaxation kinetics are insensitive to small changes in the radial cutoff value. The MSM relaxation profiles contain two to three kinetic components spanning from 1 ns to 100 μs, and CGCcap-AP6 is the only sequence that exhibits a significant amplitude of the half-dehybridization component that appears from 10 ns to 1 μs.

T-jump IR measurements of WT sequences in Fig. 6.6b show two distinct kinetic components on ~100 ns (τ_1) and ~100 μs (τ_3) timescales, followed by thermal relaxation of the sample back to T_i at ~1 ms. The transient IR spectra are well described by global lifetime fitting with exponentially damped spectral components, and similar results are obtained from inverse-Laplace-transform rate distribution spectra.[28] These processes have previously been characterized in similar sequences and they are illustrated in Fig. 6.8 through changes to the hybridization free-energy profile induced by the T-jump.[18-19, 37-38] The T-jump destabilizes the duplex state, resulting in barrierless terminal base pair fraying (τ_1). The signal change observed during τ_1 increases in order from CCends-WT to CGCcap-WT to GCGcore-WT due to an increased propensity for fraying in A:T base pairs relative to G:C.[28] This fast step is followed by complete dissociation into single strands by an activated barrier crossing (τ_3). The temperature-dependence of dehybridization rates obtained from τ_3^{-1} follows Arrhenius behavior with a large enthalpic barrier ($\Delta H_d^{\ddagger}$ > 100 kJ/mol)[28] that is consistent with full strand dissociation.[31, 39]

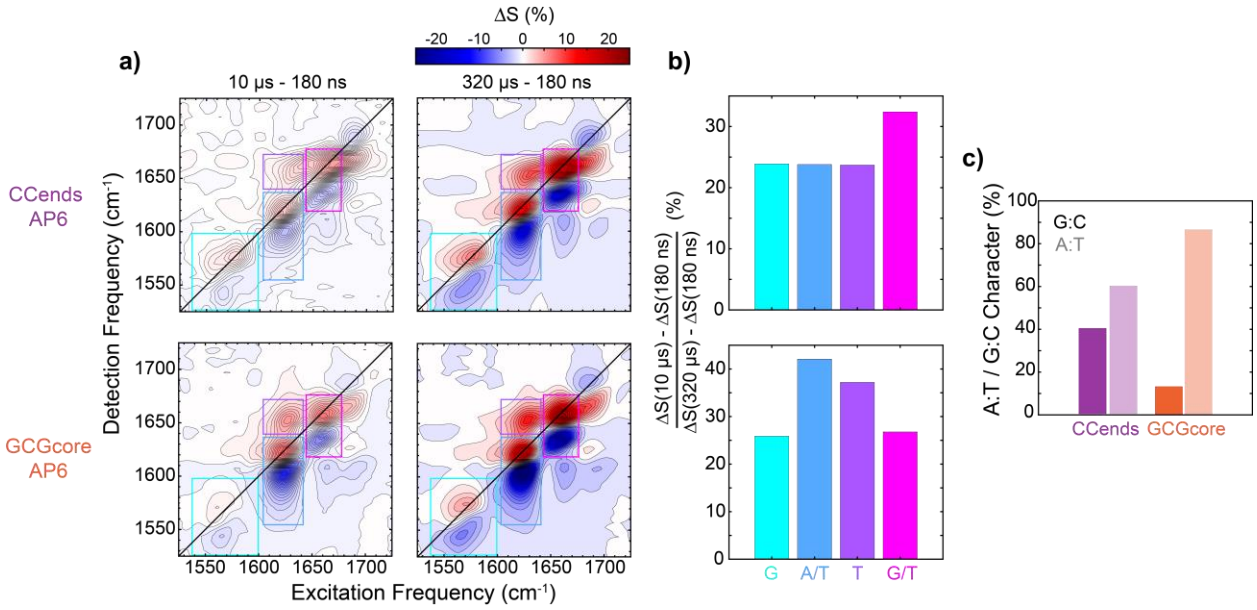

Figure 6.7 Determination of relative A:T and G:C spectral content in τ_2 response from t-2D IR. **(a)** t-2D IR spectra taken as the difference between (left) 10 μs and 180 ns and (right) 320 μs and 180 ns for CCends-AP6 and GCGcore-AP6 sequences. Spectral change amplitude is plotted as the percent change relative to the T_i spectrum. 320 μs – 180 ns spectra are plotted with 14 contours using uniform ΔS = 1.4 % spacing for CCends-AP6 and GCGcore-AP6. 10 μs – 180 ns spectra are plotted with 14 contours using uniform 0.4 and 0.7 % spacing for CCends-AP6 and GCGcore-AP6, respectively. **(b)** Ratio of integrated spectral change between 10 μs – 180 ns and 320 μs – 180 ns t-2D IR spectra in the spectral regions denoted by boxes in **(a)**. Integrations were performed over the marked regions on the absolute value t-2D IR spectra to avoid cancellation between positive and negative amplitudes. **(c)** Percentage of G:C (dark) and A:T (light) base pair loss character observed in τ_2 T-jump responses for CCends-AP6 and GCGcore-AP6 determined from t-2D IR measurements.

The T-jump IR responses of CGCcap-AP6, CCends-AP6, and GCGcore-AP6 reveal an additional kinetic component that occurs on a ~1 μs timescale (τ_2) and corresponds to dehybridization of the full stretch of base pairs on one side of the AP site (half-dehybridization). Using the contribution of G (1550 cm^{-1}) and A (1605 cm^{-1}) to the spectral response during τ_2 relative to the full time window, we determine the fraction of A:T and G:C base-pair loss character

during τ_2 (Fig. 6.6c, 6.7 and Appendix 6.D). The observed A:T character of 60% for CCends-AP6 and 85% for GCGcore-AP6 match the expectation for dissociating either the 5′- or 3′-half of the duplex based on the fractional A:T content of these palindromic sequences. On the other hand, the 91% A:T character for CGCcap-AP6 indicates a clear preference for dissociating the pure A:T segment. We find that τ_2^{-1} increases exponentially with temperature (Fig. 6.6d) and can also be described by a Kramers-like equation (eq. 6.1) as expected for an activated process.

$$1/\tau = \frac{k_b T}{h}\frac{\eta_{37}}{\eta(T)}\exp\left(-\frac{\Delta H^{\ddagger} - T\Delta S^{\ddagger}}{RT}\right) \tag{6.1}$$

Here $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are the enthalpic and entropic barriers, respectively, η is the temperature-dependent viscosity of D2O.[40] Together, the relative A:T/G:C character and activated nature of the τ_2 response directly reveal the presence of a free-energy barrier in AP6 sequences for half-dehybridization ($\Delta G_{d,2}^{\ddagger}$), which in the reverse direction corresponds to the barrier for nucleating the remaining base pairs across the AP site ($\Delta G_{h,2}^{\ddagger}$). Figure 6.8b illustrates how the AP site introduces the metastable half-hybridized state (n_{bp} = 5) into the free-energy profile, and the resulting kinetics following the T-jump.

6.3.2 Simulation of T-jump relaxation with Markov state models

To provide a molecular-level interpretation of the T-jump IR responses, we constructed Markov state models (MSMs) from 250 μs of unbiased 3SPN.2 simulations near $T_{m,MD}$ to furnish kinetic models of the T-jump relaxation behavior.[28] Using base-pairing assignments for the 200 microstates, we computed the fractional change in the number of intact A:T and G:C base pairs

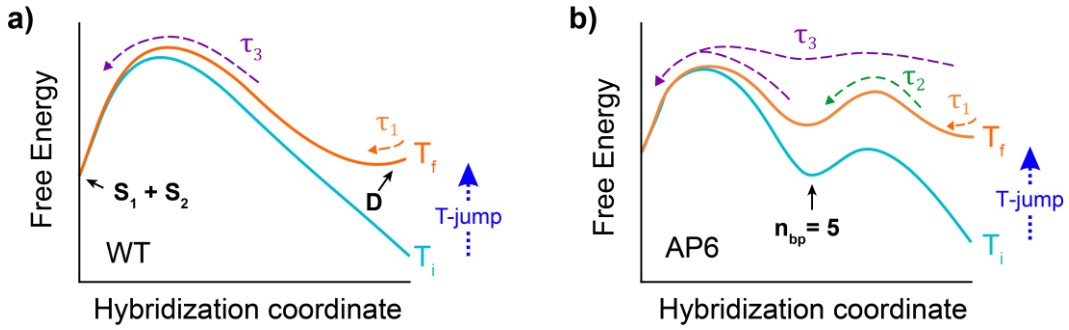

Figure 6.8 Free-energy profile reshaping by a T-jump for WT and AP6 sequences. (a) An increase in temperature promotes broadening of the duplex free-energy minimum (D) by terminal base pair fraying that occurs on a ~100 ns timescale (τ_1, orange arrows). This is followed by cooperative dissociation of the duplex to single-strands ($S_1 + S_2$) on a ~100 µs timescale (τ_3, purple arrows). **(b)** In AP6 sequences, cooperative dissociation is split into multiple steps. Following minor terminal fraying (τ_1), base pairs on one side of the AP site dehybridize on a ~1 µs timescale (τ_2, green arrows) prior to full dissociation of the duplex (τ_3).

(Δn_{bp}) as a function of relaxation time at $T_{m,MD}$ after initiating the MSM kinetics with a state occupancy obtained at $T_{m,MD}$ – 15°C (Fig. 6.6e). The 3SPN.2 coarse-grained model is known to accelerate computational timescales by ~1 order of magnitude relative to experiment,[18] but maintains a clear temporal separation between terminal fraying (~20 ns), half-dehybridization (~500 ns), and complete dehybridization (~10 µs). Consistent with the FEPs in Fig. 6.3a, we observe significant amplitude of τ_2 response for CGCcap-AP6, the only sequence for which $\Delta F^{\ddagger}_{d,2} < \Delta F^{\ddagger}_{d,1}$ at the final temperature used for the simulation. In contrast, $\Delta F^{\ddagger}_{d,2} > \Delta F^{\ddagger}_{d,1}$ for GCGcore-AP6 and CCends-AP6 sequences, leading to overlap of half-dehybridization and complete dissociation timescales and negligible τ_2 response amplitude. Experimentally, half-dehybridization responses are observed for all three AP6 sequences, so although our computational model predicts a free-energy minimum at n_{bp} = 5 and accurately captures $\Delta\Delta F^{o}_{d,37}$, we surmise that small errors in

the ratio of $\Delta F_{d,1}^{\ddagger}$ to $\Delta F_{d,2}^{\ddagger}$ – potentially due to inaccurate modeling of the local interactions around the AP site for which the 3SPN.2 model was not directly parameterized – may lead to substantial changes in the kinetic predictions of the computational model. Nevertheless, the internal consistency of the model in resolving significant τ_2 response as a result of $\Delta F_{d,2}^{\ddagger} < \Delta F_{d,1}^{\ddagger}$ provides a mechanistic understanding of the origin of this observed kinetic behavior.

6.4 Impact of AP site in longer oligonucleotides

The global impact of the AP site across these 11-mer oligonucleotides raises the question of how an AP site may disrupt cooperativity in longer oligonucleotides. We compared duplexes of A:T-rich sequences similar to GCGcore containing 11, 15, 19, and 27 possible base pairs (*N*) with and without a central AP site using the same destabilization metrics as above (Figs. 6.9 and 6.10). The observed duplex destabilization, measured with the difference in T_m between AP and WT sequences (ΔT_m), is greatest for the shortest sequences and drops with increasing *N*, but still remains substantial for *N* = 27. In comparison, the NN model yields values of ΔT_m that are roughly half, but qualitatively follows the same trend with *N*. The NN model applies only a constant free-energy penalty independent of *N* to account for the AP site. We find that this quantity can be corrected to nearly reproduce the experimental data by fitting the NN ΔT_m trend to that from FTIR using an entropic factor of 34.5 J/molK. This correction corresponds to a free-energy penalty of -10.7 kJ/mol at 37 °C that approximately matches the difference between NN and experimental $\Delta\Delta G_{d,37}^{\circ}$values for sequences with the GCG motif. This observation suggests that the AP-site defect introduces a single thermodynamic penalty that is "diluted out" with increasing *N*. In contrast to the trends in ΔT_m, the experimental $\Delta\Delta G_{d,37}^{\circ}$ decreases with *N* (Fig. 6.9d) by an amount

dependent on temperature (Fig. 6.10d). This could be a non-additive contribution to the AP site destabilization, which may reflect that the fraction of $\Delta G^{\circ}_{d,37}$ lost by incorporating an AP site ($\Delta\Delta G^{\circ}_{d,37}/\Delta G^{\circ}_{d,37}(WT)$) is much higher for short sequences than long ones.

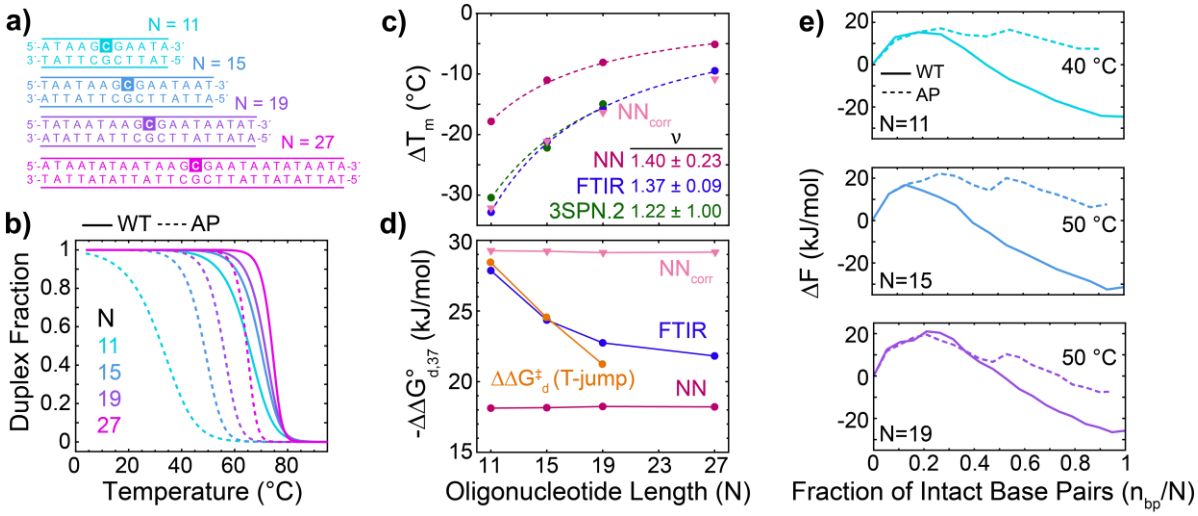

Figure 6.9 Influence of oligonucleotide length on duplex destabilization induced by an AP site. (a) GCGcore-like sequences of variable length in terms of number of possible base pairs (*N*), identifying the central AP site in the top strand (shaded). Sequences contain a GCG motif with a variable number of A:T base pairs on each side leading to *N* = 11, 15, 19, and 27. **(b)** FTIR melting curves for GCGcore-like sequences where solid and dashes lines correspond to WT and AP sequences, respectively. All sequences were measured with a 1:1 strand ratio, total strand concentration of 2 mM, and in pH 7 400 mM SPB. **(c)** $\Delta T_m = T_m(AP) - T_m(WT)$ and **(d)** $-\Delta\Delta G^{\circ}_{d,37}$ between AP and WT sequences. Values from FTIR melting curves are compared with those from NN calculations. Dashed lines indicate fits to a power law $\Delta T_m(N) = AN^{-\nu}$, and ν is listed where error corresponds to 95% confidence interval from fit. The NN ΔT_m trend was corrected to the experimental data (NN_{corr}) by fitting with an entropic factor. **(e)** FEPs for WT (solid lines) and AP (dashed lines) N = 11, 15, and 19 sequences extracted from 3SPN.2 MD WTMetaD simulations at the indicated temperatures that are approximately 20 °C below $T_{m,MD}$ of each WT sequence. FEPs are plotted as a function of the fraction of intact base pairs (n_{bp}/N) within a base pair separation cutoff of 0.7 nm.

The sizeable influence of a single lesion on the stability of a 27-mer duplex led us to investigate the predictions of the NN model for much larger oligonucleotides (Fig. 6.10e-f). These calculations predict that ΔT_m does not drop to an offset, but appears to follow power law behavior, $\Delta T_m \sim N^{-\nu}$, with ν ~ 1.1 in A:T-rich sequences. The same power law behavior with ν ~ 1.4 fits the experimental ΔT_m trend (Fig. 6.9c), and similar behavior is calculated for $\Delta\Delta G^{\circ}_{d,37}/\Delta G^{\circ}_{d,37}(WT)$. Of course, this NN calculation does not account for the fact that long stretches of A:T base pairs (>20 bp) may form bubbles,[41] or that an AP site could significantly impact bubble behavior as observed for base-pair mismatches.[42] The ΔT_m trend is expected to be sensitive to sequence and A:T/G:C content. For example, NN calculations predict ΔT_m to approach zero more sharply with an exponent of ν ~ 3.5 when extending a CCends template with GC steps due to the greater $\Delta G^{\circ}_{d,37}$ associated with GC steps.

For comparison, 3SPN.2 simulations employing WTMetaD were performed for the GCGcore-like sequences of length N = 11, 15, and 19. We find that one central AP site has a negligible effect on the end-to-end distance of duplex DNA, regardless of N, but does have a small lengthening effect on single strands that becomes negligible as N increases.[28] The computed FEPs as a function of the fraction of intact base pairs have a similar free-energy barrier $\Delta F^{\ddagger}_{h,2}$ at $n_{bp}/N = 0.5$ with a barrier height of ~4 kJ/mol for all lengths (Fig. 6.9e), indicating that the disruption of hybridization cooperativity and second nucleation penalty arising from $\Delta F^{\ddagger}_{h,2}$ persists even in the long duplex limit.

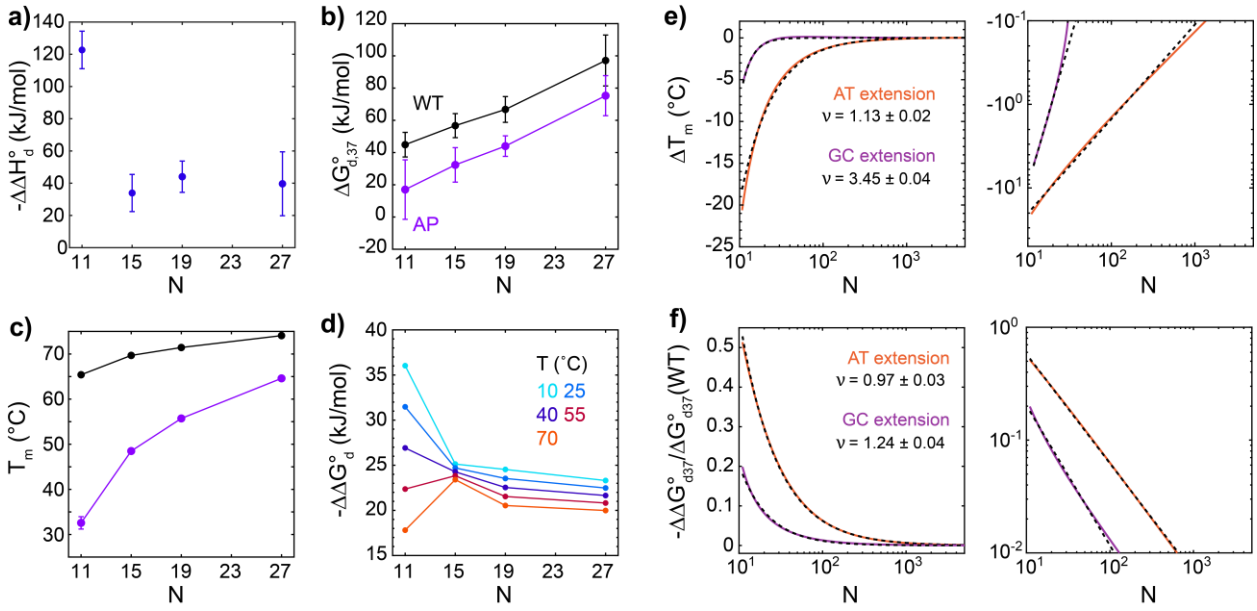

Figure 6.10 Oligonucleotide length effects on hybridization thermodynamics. (a) $\Delta\Delta H^\circ_d$ **(b)** $\Delta G^\circ_{d,37}$ and **(c)** T_m from FTIR melting curves and NN model predictions for length-dependent GCGcore-like oligonucleotides shown in Fig. 6.9a. FTIR error bars correspond to 95% confidence intervals from two-state fits. **(d)** $\Delta\Delta G^\circ_{d,37}$ from FTIR melting curves at various temperatures. **(e)** NN model predictions of ΔT_m for GCGcore and CCends template sequences extended from N = 11 to 5,000 shown on (left) log-linear and (right) log-log scales. The GCGcore and CCends sequences were symmetrically extended with AT and GC dinucleotide repeats, respectively. The NN predictions approximately follow power-law behavior as illustrated by fits to $\Delta T_m(N) = AN^{-\nu}$ (dashed black lines), and ν is listed for each sequence with an error that corresponds to the 95% confidence intervals of the fit. The log-log plots indicate minor deviations from power-law behavior for ΔT_m>0.1˚C. **(f)** NN model predictions of $\Delta\Delta G^\circ_{d,37}/\Delta G^\circ_{d,37}(WT)$ on (left) log-linear and (right) log-log scales.

While measurable in the N = 11 AP sequence, the half-dehybridization T-jump response decreases in magnitude in the N = 15 AP sequence and is undetectable for N = 19,[28] indicating that $\Delta G^\ddagger_{d,2}$ increases and ultimately exceeds $\Delta G^\ddagger_{d,1}$ for large N. Similar to $\Delta\Delta G^\circ_{d,37}$, a decrease with N is observed in $\Delta\Delta H^\circ_d$ (Fig. 6.10a) as well as $\Delta\Delta G^\ddagger_{d,37}$ from T-jump measurements (Fig. 6.9d). As indicated by the FEPs in Fig. 6.9e, these observations indicate that the role of half-hybridized states

in duplex stability decreases as *N* increases, yet remains important even for oligonucleotides far longer than those investigated here.

6.5 Disruption of cooperativity in hybridization dynamics

Both T-jump measurements and computed FEPs reveal the presence of free-energy barriers for hybridizing each half of AP6 sequences, but they do not directly describe if or how an AP site affects the dynamics of recognition and binding during hybridization. Even among studies of unmodified DNA many of the early events during hybridization are poorly understood, and it is clear that oligonucleotides may hybridize via multiple mechanisms.[16-18, 43] We examined the dynamics of base pairing from thousands of hybridization events generated by 3SPN.2 MD simulations at 20 °C for each 11-mer sequence.

To efficiently sample hybridization events to characterize nucleation behavior, 1000-2500 simulations of each sequence were initiated from a dissociated configuration, generated by shifting equilibrium duplexes by 2 nm in the x, y, and z directions, and were declared complete when reaching a defined duplex state cutoff of r_{bp} = 0.6 nm, which corresponds to a fully hybridized duplex. An umbrella potential with a force constant k=100 kJ/nm^{2}mol was applied for r_{bp} > 3 nm prevent diffusive separation of the oligonucleotide strands. We verified that this cutoff was sufficiently large that it did not influence the binding and association events by confirming negligible differences in the hybridization mechanism when comparing trajectories harvested under restraining potentials with (r_{bp} > 3 nm, k = 100 kJ/nm^{2}mol) and (r_{bp} > 4 nm, k = 10 kJ/ nm^{2}mol). Trajectories were trimmed to analyze the time window between when at least one in-

register base pair separation distance falls below 2.5 nm to when all base pair separation distances fall below 0.7 nm.

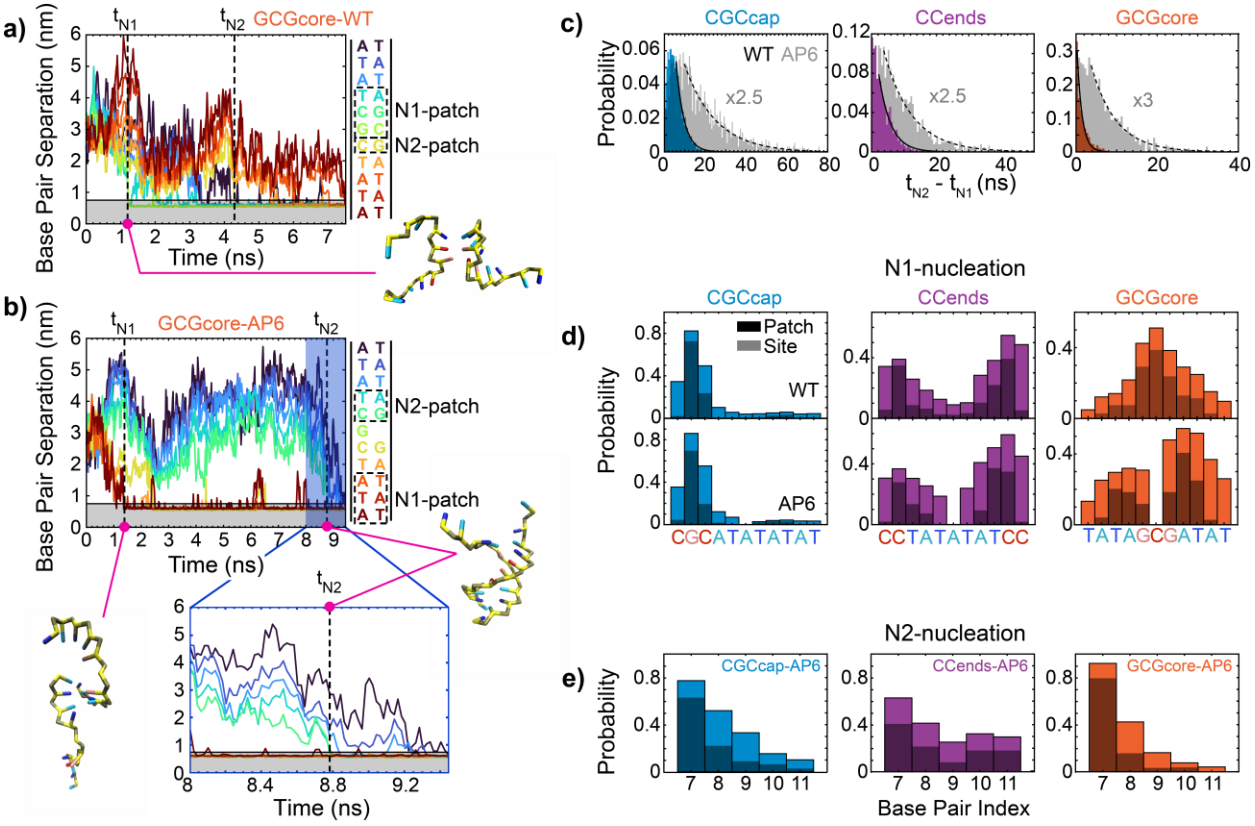

Figure 6.11 Splitting of hybridization dynamics into two nucleation events by an AP site. (a) Example hybridization trajectory at 20 °C of GCGcore-WT plotting the separation of each in-register base-pair contact (color coded). The horizontal solid line indicates the base-pair separation cutoff of 0.7 nm, and vertical dashed lines indicate the time points of nucleation (t_{N1} and t_{N2}) in the example trajectories. N1-nucleation is defined as the first permanent formation of any base pair during the hybridization event and N2-nucleation is the first permanent formation of a base pair on the other half of the duplex. N1- and N2-patches correspond to all base pairs with a separation below 0.7 nm at t_{N1} and t_{N2}, respectively. Dashed boxes denote the base pairs within the N1- and N2-patches for the example trajectories. **(b)** Example hybridization trajectory of GCGcore-AP6 at 20 °C. The bottom panel magnifies the portion of the trajectory near t_{N2} for GCGcore-AP6. The hybridization event is declared complete when the average base pair separation reaches 0.6 nm. The GCGcore-WT trajectory exhibits a single nucleation event where t_{N2} is defined just for comparison with GCGcore-AP6, which undergoes a distinct nucleation event for each half of the duplex. **(c)** Probability distribution of time difference between N2- and N1-nucleation ($t_{N2} - t_{N1}$).

Figure 6.11 Splitting of hybridization dynamics into two nucleation events by an AP site, continued

Grey distributions correspond to AP6 sequences scaled up by a factor of 2.5-3 for clarity. Distribution tails are fit to an exponential decay for WT (solid) and AP6 (dashed) sequences. AP6 sequences exhibit wider $t_{N2} - t_{N1}$ distributions than WT sequences due to the added free-energy barrier for N2-nucleation. **(d)** Probability that a given base pair site is part of the N1-patch (colored bars) or the N1-nucleation site (gray bars) determined from >1000 hybridization trajectories for each sequence. The AP site reshapes the N1-patch in a sequence-dependent fashion. **(e)** Probability that a base-pair site is part of the N2-patch (colored bars) or the N2-nucleation site (gray bars). Base-pair index refers to the distance in sequence from the terminal base pair on the duplex half where nucleation occurs (Fig. 6.12b). Regardless of sequence, the most probable site for N2-nucleation is the base pair adjacent to the AP site.

Trajectories of example hybridization events for GCGcore-WT and -AP6 sequences are shown in Fig. 6.11a,b, plotting the separation between all in-register base pairs as a function of time. Trajectories are trimmed to the time interval between the first base-pair separation distance falling below 2.5 nm and when all distances fall below 0.7 nm. From each trajectory, we isolate the time point for first permanent base pair formation, the N1-nucleation point t_{N1}, and define a nucleation patch (N1-patch) as any intact base pairs present at that point, whether they are transient or permanent.

As illustrated in Fig. 6.11a,b, WT sequences exhibit a single nucleation event—here through a N1-patch of three adjacent bases—followed by concerted formation of the remaining base pairs on either side of the patch within a few nanoseconds in a classic nucleation and zipping process.[31, 39] In contrast, AP6 sequences tend to hybridize one half of the duplex and remain in the half-hybridized state with frayed termini for many nanoseconds before nucleating the second half (N2-nucleation) at t_{N2}. This is illustrated through the changes in the distribution of $t_{N1} - t_{N2}$ time-intervals between WT and AP6 sequences in Fig. 6.11c where t_{N2} is defined for WT sequences as

the time point for nucleating a base pair on the half opposite from where t_{N1} occurs. These dynamics are consistent with the presence of two barriers to hybridization, $\Delta F_h^{\ddagger}$ and $\Delta F_{h,2}^{\ddagger}$, and suggest that n_{bp} is a good coordinate to describe hybridization. Further details of nucleation for unmodified sequences are explored in Appendix 6.E.

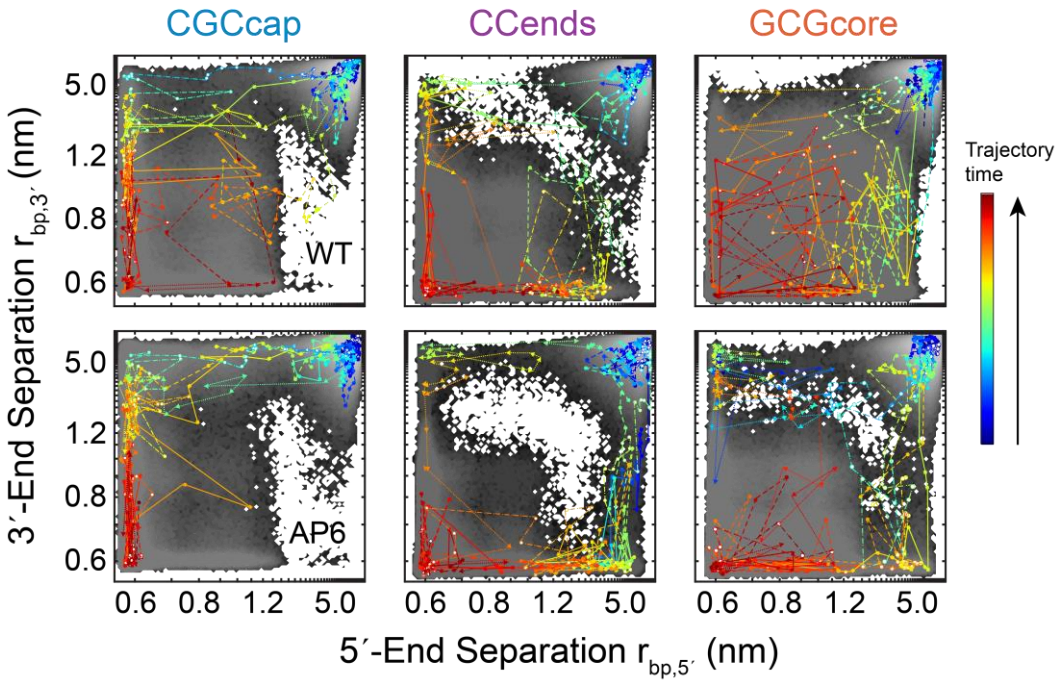

Figure 6.12 Hybridization trajectories along 2D FES. Examples hybridization events from unbiased 3SPN.2 MD simulations projected onto $r_{bp,3'}$ vs. $r_{bp,5'}$ coordinates (defined in Fig. 6.3b). FESs from Fig. 6.3b are shown as the gray scale background. Four hybridization events (different symbols for each) are shown for each sequence progressing from blue to red. Color is normalized to the length of the trajectory. Inclusion of the AP site biases hybridization dynamics through half-hybridized configurations in the top-left and bottom-right corners of the FESs.

We can relate the hybridization dynamics back to the 2D FES in Fig. 6.3b by projecting hybridization trajectories onto $r_{bp,3'}$ vs. $r_{bp,5'}$ coordinates as shown in Fig. 6.12. Trajectories start from the dissociated state in the top-right corner and are fully hybridized when reaching the bottom-left corner. Trajectories for WT sequences cross through both the edges and center of the

configurational space. In contrast, the AP site biases trajectories to proceed along the edges of the space through half-hybridized configurations, consistent with the nucleation dynamics in Fig. 6.11. The continuous motion along the $r_{bp,3'}$ vs. $r_{bp,5'}$ space suggest that these coordinates describe the dynamics, as well as thermodynamic, properties of hybridization for these sequences.

N1- and N2-nucleation events have qualitatively different properties that strongly depend on nucleobase sequence. Quantifying the N1-nucleation position and N1-patch size shows that initial nucleation most likely occurs near G:C base pairs with a N1-patch spread over 1-5 contacts (Fig. 6.11d). A central AP site localizes the N1-patch into one half of the duplex (Fig. 6.13), which is most disruptive to GCGcore and reshapes the nucleation free-energy barrier shown in Fig. 6.3a. The N1-patch of CCends and CGCcap are less perturbed by a central AP site, and their corresponding WT and AP6 FEPs have nearly identical profiles from n_{bp} = 0-5.

N2-nucleation typically occurs with all base pairs in the first half intact, which constrains the dissociated segment structure relative to single-stranded DNA[44-45] and reduces the barrier to N2-nucleation. As a result, N2-nucleation is most probable adjacent to the AP site but with characteristics that depend on the identity of the base pair adjacent to the AP site as well as the sequence of the entire segment (Fig. 6.11e). GCGcore-AP6 demonstrates that an adjacent G:C base pair promotes N2-nucleation next to the AP site relative to sequences with an adjacent A:T base pair and does not require as many base pairs in the N2-patch (Fig. 6.13c). Although each have an adjacent A:T base pair, the terminal G:C base pairs of CCends-AP6 promote a higher probability of N2-nucleation away from the AP site than in CGCcap-AP6 and exemplify that N2-nucleation

dynamics depend on the sequence of the entire segment, not just the nucleobases adjacent to the AP site.

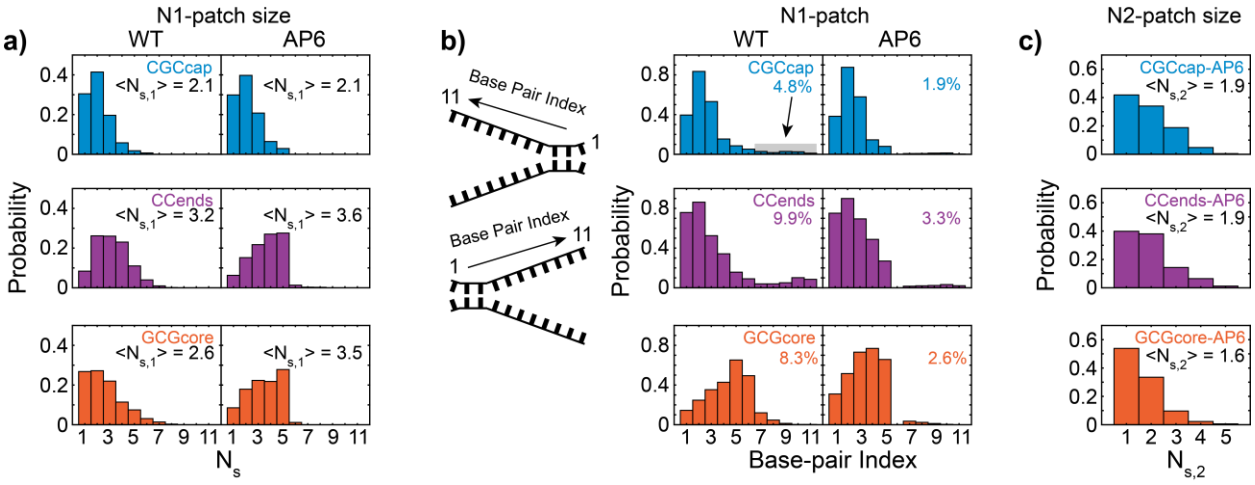

Figure 6.13 Nucleation-site and -patch properties. (a) Probability distribution for N1-patch size ($N_{s,1}$), which corresponds to the number of intact base-pairs at t_{N1}, determined with a base-pair separation cutoff of 0.7 nm. The average value of $N_{s,1}$, $\langle N_{s,1} \rangle$, is listed for each sequence and tends to increase in the presence of the AP site depending on the sequence. **(b)** Probability that a base pair is involved in the N1-patch determined from 1000-25000 hybridization trajectories for each sequence. Base-pair index refers to the distance (in number of base-pairs) from the terminal base-pair on the duplex half where N1-nucleation occurs (illustrated on the left). The sum over base-pair index 7-11 is shown for each sequence. **(c)** Probability distribution for N2-patch size ($N_{s,2}$) in AP6 sequences, which corresponds to the number of intact base-pairs at t_{N2} on the duplex half unbound at t_{N1} (base-pair index 7-11) The average value of $N_{s,2}$, $\langle N_{s,2} \rangle$, is listed for each sequence and is smaller for GCGcore-AP6 than CGCcap-AP6 and CCends-AP6.

6.6 Conclusions

Our study combines T-jump IR spectroscopy and coarse-grained MD simulations to reveal a hierarchy of consequences arising from a single AP site in duplex DNA, spanning thermodynamic destabilization, hybridization pathways, and melting kinetics. An AP site reduces the global cooperativity by effectively splitting the base-pairing energetics and dynamics of the

duplex into two segments. The level of impact is sequence-dependent where both the loss of base-pairing interactions and height of the barrier for hybridizing across the AP site depend on the identity of surrounding base pairs. This disruption of base pairing energetics and dynamics may be important for the recognition of AP sites by AP endonuclease[5, 46-47] and subsequent repair that plays a critical role in maintaining genome integrity.[48] Additionally, damage-induced AP sites may form in clusters[49-50] that will likely amplify the magnitude of the disruption observed from a single AP site in this work.

The hallmark of DNA duplex stability is the nonlinearity of cooperative interactions that depend on the concerted formation of multiple continuous contacts, but we see that even one perturbation in this cooperativity can unravel duplex stability and base-pairing dynamics. We have focused on an AP site, but there are non-canonical base pairs and epigenetic and damage-induced nucleobase modifications that impose an energetic penalty similar to that of an AP site ($-\Delta\Delta G^{\circ}_{d,37}$ = 5-20 kJ/mol) and likely disrupt base-pairing cooperativity within the duplex.[3, 51-55] Base-pairing stability and dynamics themselves tune the interactions between DNA and other molecules yet also give rise to mechanical dynamics and distortions that facilitate protein recognition.[56-57]

The contribution of cooperativity to stability is recognized not only among nucleic acids but for any binding partners that rely on a continuous stretch of interactions between subunits such as for protein-protein association and self-assembly of material and biological subunits. Even though the nature of these interactions are completely different, a disruption through mutation or defect may reshape stability and dynamics, for better or for worse.

6.7 Acknowledgements

I thank Michael Jones for carrying out the MD simulations and building the MSMs used in this chapter as well as for many helpful discussions. I also thank Yumin Lee and Lukas Whaley-Mayda for helpful discussions and feedback on the manuscript that covers the material in this chapter.

6.8 References

1. Dill, K.; Bromberg, S., *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*. Garland Science: 2010.

2. SantaLucia, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci. U.S.A.* **1998,** *95*, 1460-1465.

3. SantaLucia Jr, J.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-440.

4. Lindahl, T., Instability and decay of the primary structure of DNA. *Nature* **1993,** *362*, 709-715.

5. Thompson, P. S.; Cortez, D., New insights into abasic site repair and tolerance. *DNA Repair* **2020,** *90*, 102866.

6. Gelfand, C. A.; Plum, G. E.; Grollman, A. P.; Johnson, F.; Breslauer, K. J., Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence. *Biochem.* **1998,** *37*, 7321-7327.

7. Minetti, C. A.; Sun, J. Y.; Jacobs, D. P.; Kang, I.; Remeta, D. P.; Breslauer, K. J., Impact of bistrand abasic sites and proximate orientation on DNA global structure and duplex energetics. *Biopolymers* **2018,** *109*, e23098.

8. Sági, J.; Guliaev, A. B.; Singer, B., 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines. *Biochem.* **2001,** *40*, 3859-3868.

9. Ide, H.; Shimizu, H.; Kimura, Y.; Sakamoto, S.; Makino, K.; Glackin, M.; Wallace, S. S.; Nakamuta, H.; Sasaki, M.; Sugimoto, N., Influence of. alpha.-Deoxyadenosine on the Stability and

Structure of DNA. Thermodynamic and Molecular Mechanics Studies. *Biochem.* **1995,** *34*, 6947-6955.

10. Pyshnyi, D. V.; Lomzov, A. A.; Pyshnaya, I. A.; Ivanova, E. M., Hybridization of the bridged oligonucleotides with DNA: Thermodynamic and kinetic studies. *J. Biomol. Struct. Dyn.* **2006,** *23*, 567-579.

11. Vesnaver, G.; Chang, C.-N.; Eisenberg, M.; Grollman, A. P.; Breslauer, K. J., Influence of abasic and anucleosidic sites on the stability, conformation, and melting behavior of a DNA duplex: correlations of thermodynamic and structural data. *Proc. Nat. Acad. Sci. U.S.A.* **1989,** *86*, 3614-3618.

12. Morten, M. J.; Lopez, S. G.; Steinmark, I. E.; Rafferty, A.; Magennis, S. W., Stacking-induced fluorescence increase reveals allosteric interactions through DNA. *Nucleic Acids Res.* **2018,** *46*, 11618-11626.

13. Chen, J.; Dupradeau, F.-Y.; Case, D. A.; Turner, C. J.; Stubbe, J., DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* **2008,** *36*, 253-262.

14. Zálešák, J.; Lourdin, M.; Krejčí, L.; Constant, J.-F.; Jourdan, M., Structure and dynamics of DNA duplexes containing a cluster of mutagenic 8-oxoguanine and abasic site lesions. *J. Mol. Biol.* **2014,** *426*, 1524-1538.

15. Lukin, M.; de Los Santos, C., NMR structures of damaged DNA. *Chem. Rev.* **2006,** *106*, 607-686.

16. Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J., Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* **2014,** *141*, 07B613_1.

17. Ouldridge, T. E.; Šulc, P.; Romano, F.; Doye, J. P.; Louis, A. A., DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **2013,** *41*, 8886-8895.

18. Jones, M. S.; Ashwood, B.; Tokmakoff, A.; Ferguson, A. L., Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, Markov state models, and infrared spectroscopy. *J. Am. Chem. Soc.* **2021,** *143*, 17395-17411.

19. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

20. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

21. Schickinger, M.; Zacharias, M.; Dietz, H., Tethered multifluorophore motion reveals equilibrium transition kinetics of single DNA double helices. *Proc. Nat. Acad. Sci. U.S.A.* **2018,** *115*, E7512-E7521.

22. Hertel, S.; Spinney, R.; Xu, S.; Ouldridge, T. E.; Morris, R.; Lee, L., The stability and number of nucleating interactions determine DNA hybridization rates in the absence of secondary structure. *Nucleic Acids Res.* **2021,** *50*, 7829-7841.

23. Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; De Pablo, J. J., An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013,** *139*, 10B604_1.

24. Pande, V. S.; Beauchamp, K.; Bowman, G. R., Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010,** *52*, 99-105.

25. Wehmeyer, C.; Scherer, M. K.; Hempel, T.; Husic, B. E.; Olsson, S.; Noé, F., Introduction to Markov state modeling with the PyEMMA software. *Living J. Comput. Mol. Sci.* **2019,** *1*, 5965.

26. Araque, J.; Robert, M., Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* **2016,** *144*, 125101.

27. Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A., Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* **2008,** *47*, 5336-5353.

28. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. *Proc. Natl. Acad. Sci. U.S.A.* **2023,** *120*, e2219124120.

29. Poland, D.; Scheraga, H. A., *Theory of helix-coil transitions in biopolymers*. Academic Press.: 1970.

30. Thompson, P. A.; Eaton, W. A.; Hofrichter, J., Laser Temperature Jump Study of the Helix⇌ Coil Kinetics of an Alanine Peptide Interpreted with a 'Kinetic Zipper'Model. *Biochem.* **1997,** *36*, 9200-9210.

31. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.

32. Jones, K. C.; Ganim, Z.; Tokmakoff, A., Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 14060-14066.

33. Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A., Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* **2007,** *78*, 063101.

34. Peng, C. S.; Jones, K. C.; Tokmakoff, A., Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 15650-15660.

35. Banyay, M.; Sarkar, M.; Gräslund, A., A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **2003,** *104*, 477-488.

36. Lee, C.; Cho, M., Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **2006,** *125*, 114509.

37. Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Minnes, L.; Towrie, M.; Parker, A. W.; Hunt, N. T., Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 6188-6199.

38. Dale, J.; Howe, C. P.; Toncrova, H.; Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Towrie, M.; Parker, A. W.; McLeish, T. C.; Hunt, N. T., Combining steady state and temperature jump IR spectroscopy to investigate the allosteric effects of ligand binding to dsDNA. *Phys. Chem. Chem. Phys.* **2021,** *23*, 15352-15363.

39. Pörschke, D.; Eigen, M., Co-operative non-enzymatic base recognition III. Kinetics of the helix—coil transition of the oligoribouridylic· oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **1971,** *62*, 361-381.

40. Cho, C.; Urquidi, J.; Singh, S.; Robinson, G. W., Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **1999,** *103*, 1991-1994.

41. Zeng, Y.; Montrichok, A.; Zocchi, G., Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.* **2004,** *339*, 67-75.

42. Zeng, Y.; Zocchi, G., Mismatches and bubbles in DNA. *Biophys. J.* **2006,** *90*, 4522-4529.

43. Yin, Y.; Zhao, X. S., Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* **2011,** *44*, 1172-1181.

44. Israels, B.; Albrecht, C. S.; Dang, A.; Barney, M.; von Hippel, P. H.; Marcus, A. H., Submillisecond Conformational Transitions of Short Single-Stranded DNA Lattices by Photon Correlation Single-Molecule Förster Resonance Energy Transfer. *J. Phys. Chem. B* **2021,** *125*, 9426-9440.

45. Qu, P.; Yang, X.; Li, X.; Zhou, X.; Zhao, X. S., Direct measurement of the rates and barriers on forward and reverse diffusions of intramolecular collision in overhang oligonucleotides. *J. Phys. Chem. B* **2010,** *114*, 8235-8243.

46. Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A., DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature* **2000,** *403*, 451-456.

47. Kuznetsova, A. A.; Matveeva, A. G.; Milov, A. D.; Vorobjev, Y. N.; Dzuba, S. A.; Fedorova, O. S.; Kuznetsov, N. A., Substrate specificity of human apurinic/apyrimidinic endonuclease APE1 in the nucleotide incision repair pathway. *Nucleic Acids Res.* **2018,** *46*, 11454-11465.

48. Huskova, A.; Landova, B.; Boura, E.; Silhan, J., The rate of formation and stability of abasic site interstrand crosslinks in the DNA duplex. *DNA Repair* **2022,** *113*, 103300.

49. Bignon, E.; Gattuso, H.; Morell, C.; Dehez, F.; Georgakilas, A. G.; Monari, A.; Dumont, E., Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion. *Nucleic Acids Res.* **2016,** *44*, 8588-8599.

50. Georgakilas, A. G.; O'Neill, P.; Stewart, R. D., Induction and repair of clustered DNA lesions: what do we know so far? *Radiat. Res.* **2013,** *180*, 100-109.

51. Singh, S. K.; Szulik, M. W.; Ganguly, M.; Khutsishvili, I.; Stone, M. P.; Marky, L. A.; Gold, B., Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Res.* **2011,** *39*, 6789-6801.

52. Dubini, R. C. A.; Korytiaková, E.; Schinkel, T.; Heinrichs, P.; Carell, T.; Rovó, P., 1H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives. *ACS Phys. Chem. Au* **2022,** *2*, 237-246.

53. Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M., Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **2006,** *314*, 1001-1004.

54. Cisse, I. I.; Kim, H.; Ha, T., A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* **2012,** *19*, 623-627.

55. Rangadurai, A.; Shi, H.; Xu, Y.; Liu, B.; Abou Assi, H.; Boom, J. D.; Zhou, H.; Kimsey, I. J.; Al-Hashimi, H. M., Measuring thermodynamic preferences to form non-native conformations in nucleic acids using ultraviolet melting. *Proc. Nat. Acad. Sci. U.S.A.* **2022,** *119*, e2112496119.

56. Afek, A.; Shi, H.; Rangadurai, A.; Sahay, H.; Senitzki, A.; Xhani, S.; Fang, M.; Salinas, R.; Mielko, Z.; Pufall, M. A.; Poon, G. M. K.; Haran, T. E.; Schumacher, M. A.; Al-Hashimi, H. M.; Gordan, R., DNA mismatches reveal conformational penalties in protein–DNA recognition. *Nature* **2020,** *587*, 291-296.

57. Ngo, T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016,** *7*, 1-9.

58. Mergny, J.-L.; Lacroix, L., Analysis of thermal melting curves. *Oligonucleotides* **2003,** *13*, 515-537.

59. Marky, L. A.; Breslauer, K. J., Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **1987,** *26*, 1601-1620.

60. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with DNA duplex formation: salt-and sequence-dependent effects. *Biochem.* **2006,** *45*, 604-616.

61. Lang, B. E.; Schwarz, F. P., Thermodynamic dependence of DNA/DNA and DNA/RNA hybridization reactions on temperature and ionic strength. *Biophys. Chem.* **2007,** *131*, 96-104.

62. Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; Record, M. T., Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochem.* **1999,** *38*, 8409-8422.

63. Hadži, S.; Lah, J., Origin of heat capacity increment in DNA folding: The hydration effect. *Biochim. Biophys. Acta* **2021,** *1865*, 129774.

64. Hughesman, C. B.; Turner, R. F.; Haynes, C. A., Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. *Biochem.* **2011,** *50*, 5354-5368.

65. Dragan, A.; Privalov, P.; Crane-Robinson, C., Thermodynamics of DNA: heat capacity changes on duplex unfolding. *Eur. Biophys. J.* **2019,** *48*, 773-779.

66. Freire, E.; Schön, A.; Velazquez-Campoy, A., Isothermal titration calorimetry: general formalism using binding polynomials. *Meth. Enzymol.* **2009,** *455*, 127-155.

67. van Stokkum, I. H.; Larsen, D. S.; Van Grondelle, R., Global and target analysis of time-resolved spectra. *Biochim. Biophys. Acta* **2004,** *1657*, 82-104.

68. Mullen, K. M.; Vengris, M.; van Stokkum, I. H., Algorithms for separable nonlinear least squares with application to modelling time-resolved spectra. *J. Glob. Optim.* **2007,** *38*, 201-213.

69. Noid, W., Perspective: Advances, challenges, and insight for predictive coarse-grained models. *J. Phys. Chem. B* **2023,** *127*, 4174-4207.

70. Papavasileiou, K. D.; Peristeras, L. D.; Bick, A.; Economou, I. G., Molecular dynamics simulation of pure n-alkanes and their mixtures at elevated temperatures using atomistic and coarse-grained force fields. *J. Phys. Chem. B* **2019,** *123*, 6229-6243.

71. Fritz, D.; Koschke, K.; Harmandaris, V. A.; van der Vegt, N. F.; Kremer, K., Multiscale modeling of soft matter: scaling of dynamics. *Phys. Chem. Chem. Phys.* **2011,** *13*, 10412-10420.

72. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.

73. Hammond, G. S., A correlation of reaction rates. *J. Am. Chem. Soc.* **1955,** *77*, 334-338.

74. Bechtold, B. *Violin Plots for MATLAB, Github Project*, 2016.

Appendix 6.A Two-state thermodynamic model of hybridization

Most of the oligonucleotide systems studied in this thesis exhibit a single sigmoidal melting transition. These measurements can be most simply described using a two-state thermodynamic model for duplex (*D*) dissociation to single strands (*S*) for non-self-complementary or self-complementary oligonucleotides. This is a standard treatment of optical melting curves,[58-59] but we provide details here to clarify the analysis used throughout this thesis.

$$D \rightleftharpoons S_1 + S_2 \quad \text{(Non-self-complementary)} \qquad \text{(6.A1a)}$$

$$D \rightleftharpoons 2S \quad \text{(Self-complementary)} \qquad \text{(6.A1b)}$$

In the two-state description, the D state represents an average over the ensemble of all base-pairing configurations (frayed, shifted, etc.). The equilibrium constant for dissociation (K_d) is related to the fraction of intact duplexes (θ_D).

$$K_d = \frac{[S_1][S_2]}{[D]} = \frac{(1-\theta_D)^2 c_{tot}}{2\theta_D} \quad \text{(Non-self-complementary)} \quad \text{(6.A2a)}$$

$$K_d = \frac{[S]^2}{[D]} = \frac{2(1-\theta_D)^2 c_{tot}}{\theta_D} \quad \text{(Self-complementary)} \qquad \text{(6.A2b)}$$

$$\theta_D = \frac{2[D]}{c_{tot}} \tag{6.A3}$$

c_{tot} is the total oligonucleotide concentration.

$$c_{tot} = 2[D] + [S_1] + [S_2] \quad \text{(Non-self-complementary)} \tag{6.A4a}$$

$$c_{tot} = 2[D] + [S] \quad \text{(Self-complementary)} \tag{6.A4b}$$

Eq. 6.A2a assumes a 1:1 molar ratio between complementary single strands ($[S_1] = [S_2]$), and all FTIR temperature series in this thesis were performed under this condition. Eqs. 6.A2 can be arranged to express θ_D in terms of K_d.

$$\theta_D = 1 + \frac{K_d - \sqrt{K_d^2 + 2c_{tot}K_d}}{c_{tot}} \quad \text{(Non-self-complementary)} \tag{6.A5a}$$

$$\theta_D = 1 + \frac{K_d - \sqrt{K_d^2 + 8c_{tot}K_d}}{4c_{tot}} \quad \text{(Self-complementary)} \tag{6.A5b}$$

The temperature dependence of θ_D and K_d are determined by the enthalpy (ΔH_d°) and entropy (ΔS_d°) difference between the D and S states.

$$K_d(T) = \exp\left[-\frac{\Delta H_d^\circ(T)}{RT} + \frac{\Delta S_d^\circ(T)}{R} \right] \tag{6.A6}$$

In reality, both ΔH_d° and ΔS_d° are temperature-dependent due to a change in heat capacity (ΔC_p) between the D and S states, which is thought to arise from temperature-dependent stacking interactions in each of the single-strands and a change in solvent accessible surface area upon

dehybridization.[60-63] For dissociation of short oligonucleotides, ΔC_p follows a length-scaling of 150-250 J/molK per base-pair[64-65] and depends on the sequence and counterion concentration and can also be temperature-dependent.[60-61] Assuming ΔC_p is constant over the temperature range of interest, we can express ΔH_d° and ΔS_d° as a function of temperature.

$$\Delta H_d^\circ(T) = \Delta H_d^\circ(T_m) + \Delta C_p(T - T_m) \tag{6.A7}$$

$$\Delta S_d^\circ(T) = \Delta S_d^\circ(T_m) + \Delta C_p \ln\left(\frac{T}{T_m}\right) \tag{6.A8}$$

ΔH_d° and ΔS_d° are expressed relative to their values at the melting temperature (T_m), which are the van't Hoff enthalpy and entropy. T_m, which here is defined as the temperature where θ_D = 0.5, can be re-cast in terms of ΔH_d° and ΔS_d°.

$$T_m = \frac{\Delta H_d^\circ(T_m)}{\Delta S_d^\circ(T_m) - R\ln(c_{tot}/4)} \qquad \text{(Non-self-complementary)} \tag{6.A9a}$$

$$T_m = \frac{\Delta H_d^\circ(T_m)}{\Delta S_d^\circ(T_m) - R\ln(c_{tot})} \qquad \text{(Self-complementary)} \tag{6.A9b}$$

ΔH_d° and T_m also determine the slope of the melting transition at T_m:

$$\left(\frac{\partial \theta_D}{\partial T}\right)_{T_m} = -\frac{\Delta H_d^\circ(T_m)}{6RT_m^2} \tag{6.A10}$$

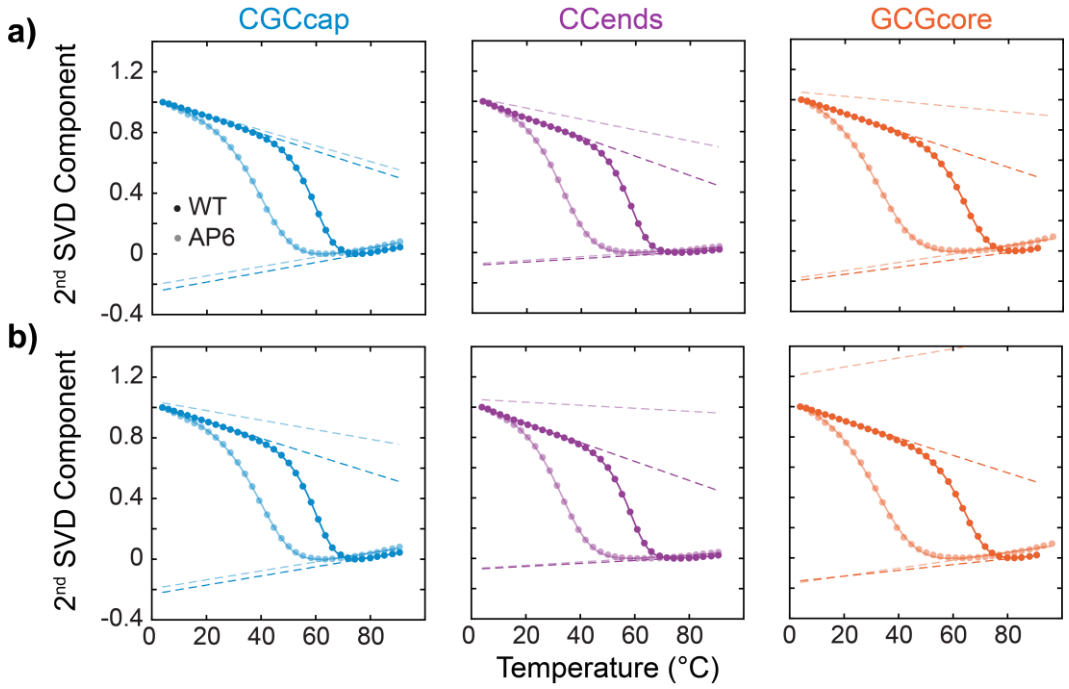

Figure 6.A1 Two-state fitting to determine duplex melting curves. 2nd SVD components from FTIR temperature series fit (solid lines) to 2-state thermodynamic model (eq. 6.A11) **(a)** with ΔC_p = 0 and **(b)** with ΔC_p fixed at the value determined from ITC. Dashed lines indicate high (L_D) and low (L_S) baselines from the fits.

To fit the FTIR melting data, ΔH_d° and T_m were used as free parameters and ΔC_p was set to 0 initially. The 2nd SVD components from FTIR data were then fit to θ_D with baselines described by additional slope (m_D, m_s) and intercept (b_D, b_s) terms.

$$V_{Fit}^{(2)} = \theta_D \left(L_D - L_S \right) + L_S \tag{6.A11}$$

$$L_D = m_D T + b_D \tag{6.A12a}$$

$$L_S = m_S T + b_S \tag{6.A12b}$$

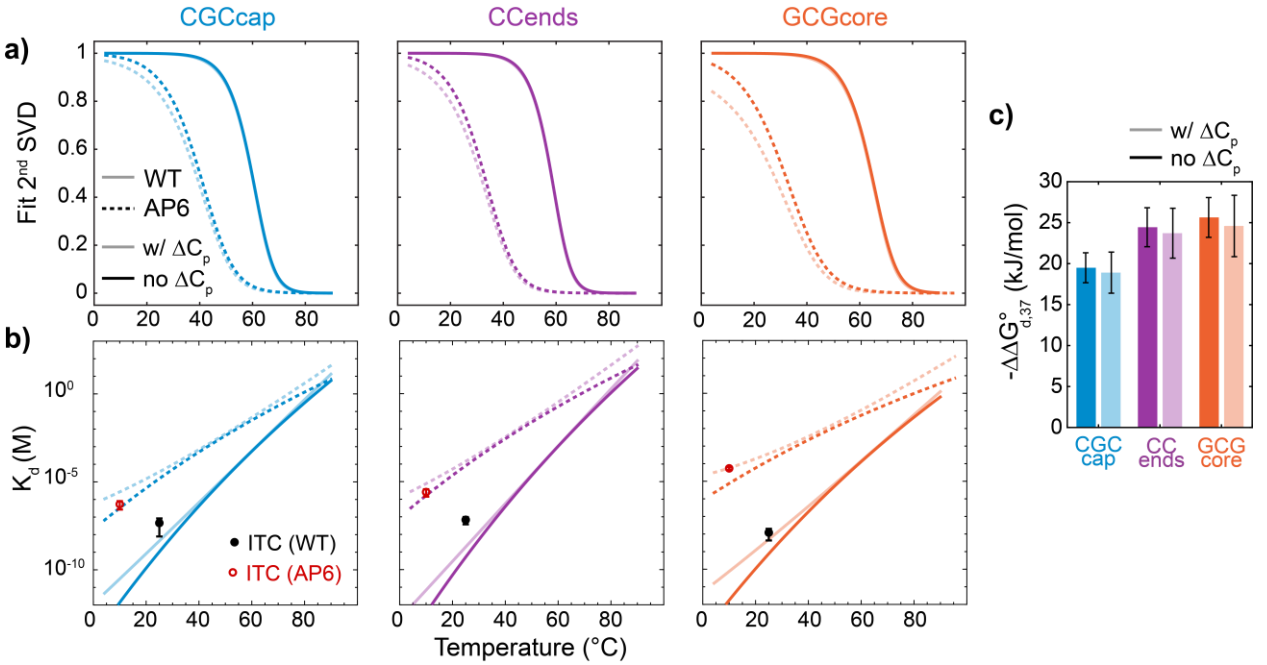

Figure 6.A2 Impact of temperature-dependent ΔH°_{d} and ΔS°_{d} on melting curves. θ_D determined from fitting FTIR 2nd SVD components to a two-state model (eq. 6.A11) **(a)** with ΔC_p = 0 and **(b)** ΔC_p fixed at the value determined from ITC. **(c)** Temperature-dependent equilibrium constant for duplex dissociation (K_d) determined from two-state fits to FTIR 2nd SVD components with ΔC_p = 0 (solid lines) and with ΔC_p set to value from ITC (transparent lines). K_d values determined at select temperatures with ITC are shown as circles. Error bars indicate 95% confidence interval from fitting ITC thermograms to a single-site binding model.

Fig. 6.A1a shows fits of the FTIR 2nd SVD components to eq. 6.A10 with ΔC_p = 0. We can also fit the melting data using ΔC_p determined from ITC (Fig. 6.B1), which is shown in Fig. 6.A1b. Compared to profiles of θ_D determined from fits with ΔC_p = 0 (Fig. 6.A1a), those including ΔC_p exhibit reduced hybridization at lower temperature due to higher K_d values. Overall, K_d values determined from ITC agree better with the melting curve fits that include ΔC_p, which is expected since the ΔC_p values were determined ITC. The largest discrepencies are observed for the lowest K_d measurements (ITC K_d < 10^{-7}), but these ITC determined values are expected to have the

highest error due to having very sharp thermogram transitions. $\Delta\Delta G^{\circ}_{d,37}$ values are nearly the same with or without inclusion of ΔC_p in the thermodynamic fitting (Fig. 6.A2) but this comparison does not account for minor differences in ΔC_p between WT and AP6 sequences.

Appendix 6.B Isothermal titration calorimetry for duplex hybridization

Isothermal titration calorimetry (ITC) experiments are used throughout this thesis as a complementary measurement of duplex hybridization thermodynamics. Unlike optical melting curves, ITC (as well as differential scanning calorimetry) directly measures ΔH°_{d}. Therefore, ITC may be used to verify interpretation of FTIR and other spectroscopic melting curves.

All measurements in this thesis were performed using a MicroCal iTC200 (Malvern Panalytical) housed in the BioPhysics core facility at the University of Chicago. Following each injection, the power required to re-equilibrate the sample and reference cells was monitored as a function of time. Each measurement started with an initial 0.4 μL injection followed by 19 2 μL aliquot injections of the titrant oligonucleotide solution into the cell. The duration of the 2 μL injections were varied from 4 – 210 seconds depending on the sample in order to avoid signal saturation. The power profile of each injection was integrated over time to obtain the change in heat from the injection (ΔQ). The heat of dilution was measured for the titrant at each experimental temperature and subtracted from ΔQ after integration. Curves of ΔQ vs molar ratio of titrant and cell oligonucleotide were fit to a single site binding model[66] with free parameters for binding enthalpy (ΔH°_{d}), binding constant (K_a), and reaction stoichiometry (n).

$$\Delta Q_i = Q_i - Q_{i-1} + \frac{V_{inj}}{V_o}\left[\frac{Q_i + Q_{i-1}}{2}\right] \tag{6.B1}$$

$$Q_i = \frac{nM_{c,i}V_o\Delta H^\circ}{2}\left[1 + \frac{T_{c,i}}{nM_{c,i}} - \frac{1}{nK_aM_{c,i}} - \sqrt{\left(1 + \frac{T_{c,i}}{nM_{c,i}} + \frac{1}{nK_aM_{c,i}}\right)^2 - \frac{4T_{c,i}}{nM_{c,i}}}\right] \tag{6.B2}$$

Q_i is the total heat exchanged after injection i of titrant. $M_{c,i}$ and $T_{c,i}$ are the concentration of cell oligonucleotide and titrant oligonucleotide, respectively, in the cell after injection i. V_o is the initial volume of sample in the cell, which was 200 µL for all measurements, and V_{inj} is the volume of titrant used in each injection (2 µL).

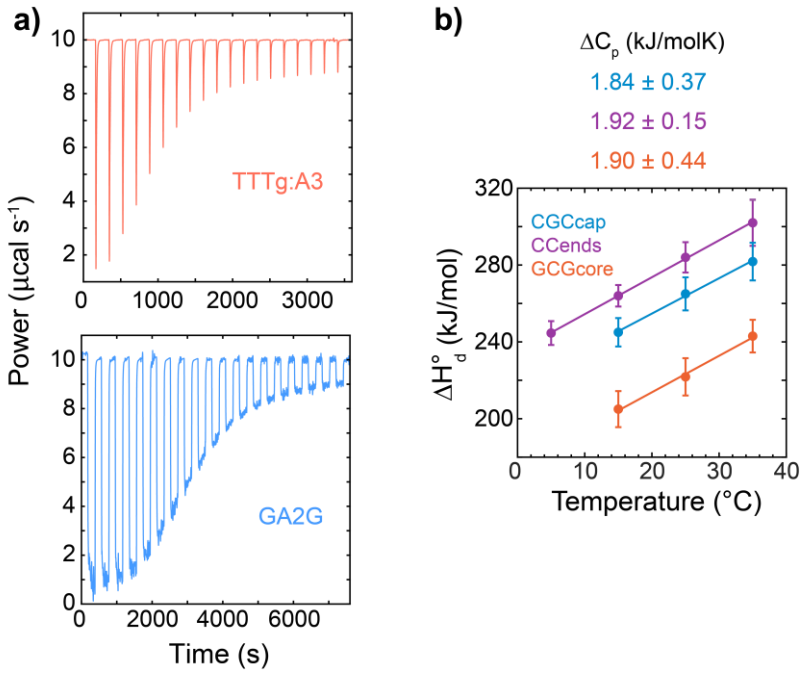

Figure 6.B1 Isothermal titration calorimetry of DNA hybridization. (a) Raw ITC thermograms for TTTg:A3 and GA2G sequences (Chapter 9) showing the measured power vs. time. A slower injection rate was used for GA2G to avoid signal saturation. **(b)** Enthalpy of dissociation (ΔH_d°) determined with ITC as a function of temperature for CGCcap, CCends, and GCGcore WT sequences. Vertical error bars indicate 95% confidence intervals from nonlinear least squares fitting of ITC thermograms. Temperature-dependent ΔH_d° data are fit to linear functions (solid lines) and the slope (ΔC_p) is listed for each sequence.

Appendix 6.C Global lifetime analysis of t-HDVE data

t-HDVE is analyzed using global lifetime analysis, which globally fits the time-domain traces at all probe wavelengths to a user-defined number of kinetic components.[67] The number of components are decided upon inspection of the time-domain data as well as the number of apparent components in rate spectra.[28] We applied global lifetime analysis to the t-HDVE data ($\psi_{t\text{-}HDVE}$) using three or four exponential components (N_{comp}) based on an examination of the rate- and time-domain data.

$$\psi_{t-HDVE}\left(\omega,\tau_{TJ}\right)=\sum_{n=1}^{N_{comp}}A_n\left(\omega\right)\exp\left(-\frac{\tau_{TJ}}{\tau_n}\right) \tag{6.C1}$$

The first two or three components account for the T-jump response while the final component accounts for thermal relaxation and re-hybridization of the sample. Global fitting was performed by minimizing the objective function (R) in eq. 6.C2 using a nonlinear least-squares solver where I is an identity matrix, C is the matrix of time-dependent concentration profiles, and C^+ denotes the pseudoinverse of C.[68]

$$R=\left\|\left(I-CC^{+}\right)\psi_{t-HDVE}\right\|^2 \tag{6.C2}$$

Figure 6.C1 shows an example of the global lifetime fitting applied to CGCcap-AP6. CCends-AP4, CCends-AP6, and GCGcore-AP6 where each dataset is fit to a sum of four exponential decays. The frequency-dependent amplitudes associated with each exponential component make up the decay-associated spectra (DAS) shown in Fig. 6.C1c. The final DAS corresponds to the

thermal relaxation response and has the opposite amplitude sign relative to the T-jump components.

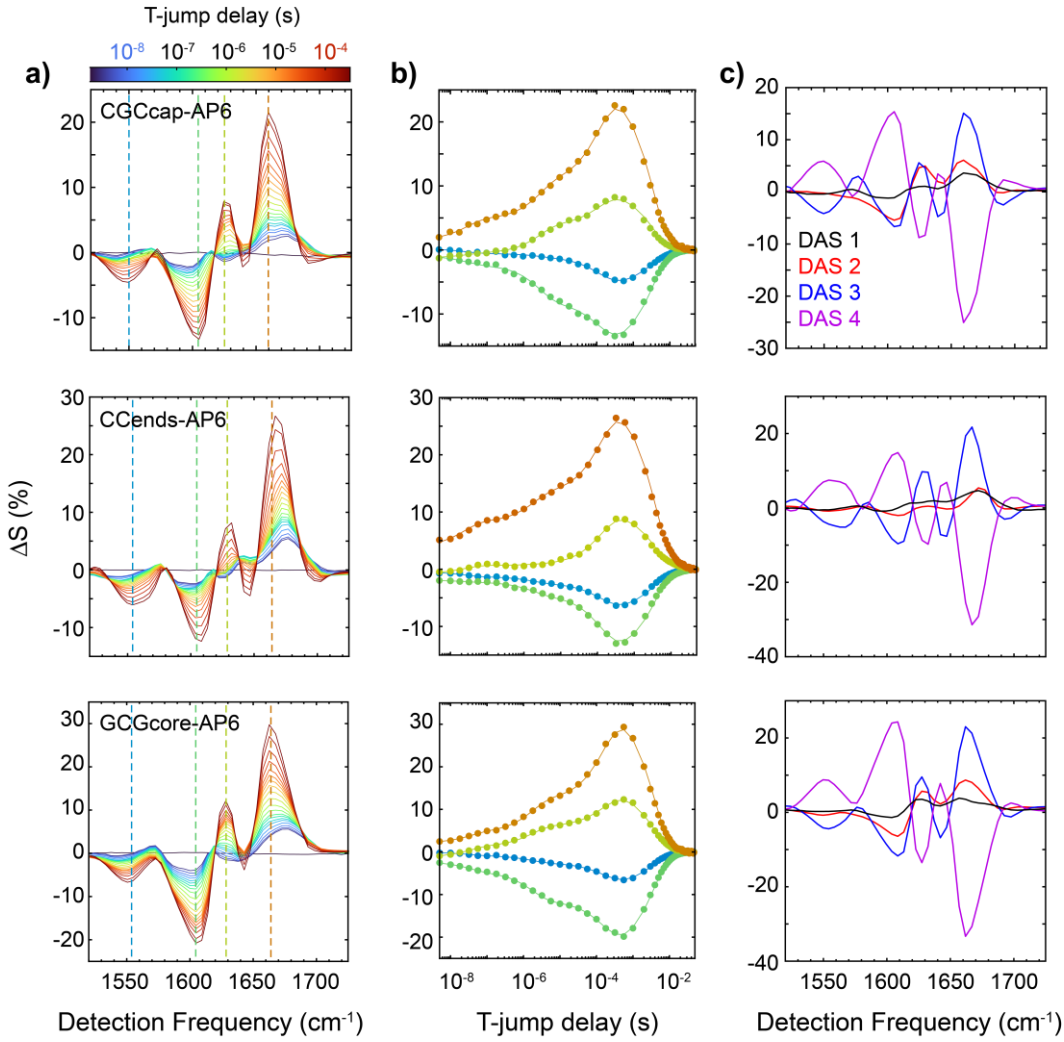

Figure 6.C1 Example of global lifetime fitting applied to t-HDVE data. **(a)** t-HDVE spectra from 5 ns to 560 μs. **(b)** Representative time traces probed at 1554, 1605, 1625, and 1665 cm^{-1} with fits (solid lines) from global fitting to a series of four exponentials. **(c)** Decay-associated spectra (DAS) from four-component global fitting. The final component of opposite sign amplitude corresponds to thermal relaxation and rehybridization of the duplex.

Fig. 6.C2 shows the DAS for each sequence determined from three- or four-component global lifetime fitting. All sequences exhibit a first component with time constant from tens to

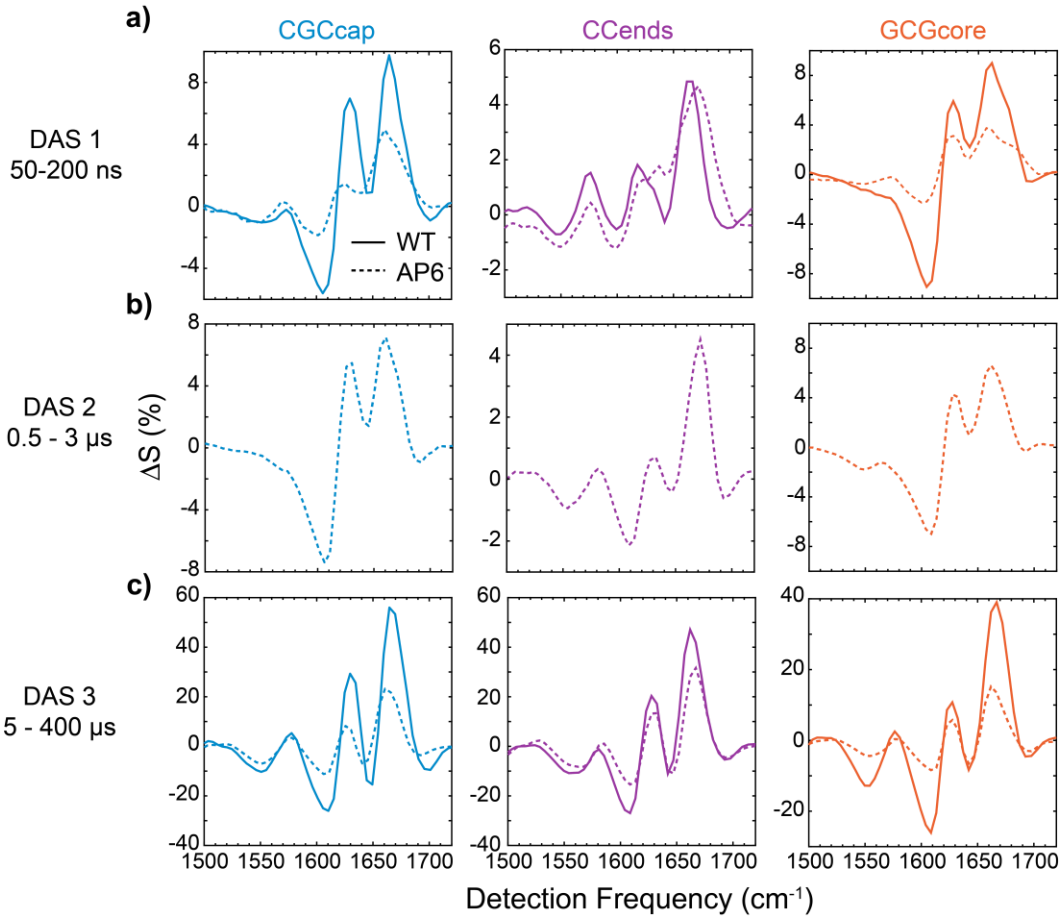

Figure 6.C2 Decay-associated spectra determined from global lifetime analysis of t-HDVE data. DAS corresponding to thermal relaxation are not shown. DAS are shown for a T-jump with T_f near the respective T_m of each sequence. **(a)** The first spectral component (DAS 1) has a time constant ranging from tens to hundreds of ns and is observed for all sequences. **(b)** An intermediate component (DAS 2) is only observed for CGCcap-AP6, CCends-AP6, and GCGcore-AP6 with a

hundreds of nanoseconds. Previous studies of short oligonucleotide dissociation with T-jump IR demonstrated that the structural dynamics on this timescale can primarily be assigned to terminal base pair fraying.[18-20, 37-38] The first component spectra in Fig. 6.C2 (DAS 1) are consistent with this assignment. A second component (DAS 2) with time constant from hundreds of ns to many µs is only needed for CGCcap-AP6, CCends-AP6, and GCGcore-AP6. The final T-jump

component (DAS 3) occurs from a few to hundreds of μs depending on the temperature and sequence and corresponds to full strand dissociation and association.

Appendix 6.D Determination of A:T and G:C character in τ_2 process

Figure 6.6c shows the fraction of A:T and G:C character in the τ_2 response observed for CGCcap-AP6, CCends-AP6, and GCGcore-AP6. For these calculations, we assume that the t-HDVE signals at 1550 and 1605 cm^{-1} report exclusively on changes of the guanine ring and A/T ring ESA bands, respectively. In reality, there are also small overlapping contributions from changes in the ESA band of the guanine carbonyl mode to the A/T region and the A/T ESA band to the G region. The τ_2 T-jump amplitudes are shown in Fig. 6.D1 for values of T_f where a response was measurable. The percentage (P) of G:C and A:T dehybridization during the intermediate response is determined from the ratio of total dissociation amplitude ($A_2 + A_3$) and τ_2 response amplitude (A_2) at 1550 and 1605 cm^{-1}.

$$P_{G/A} = \frac{A_{2,G/A}}{A_{2,G/A} + A_{3,G/A}} \tag{6.D1}$$

The values of P at 1550 and 1605 cm^{-1} are shown in Fig. 6.D1a. P generally decreases in going to higher temperatures due to a greater loss of duplex population (i.e. larger A_3). Then, the A:T and G:C character (C) are determined by the ratio of $P_{A/G}$ to the sum of P_A and P_G weighted by the number of A:T (N_A) and G:C (N-N_A) base pairs in the sequence. N is the total number of starting base pairs in the sequence, which is assumed to be 10 for sequences containing an AP site.

$$C_{A:T} = \frac{N_A P_A}{\left(N - N_A\right) P_G + N_A P_A} \tag{6.D2a}$$

$$C_{G:C} = \frac{\left(N - N_A\right) P_G}{\left(N - N_A\right) P_G + N_A P_A} = 1 - C_{A:T} \tag{6.D2b}$$

The values of $C_{A:T}$ are shown in Fig. 6.D1c and are nearly the same, within experimental error, across the temperature ranges measured. The mean value of $C_{A:T}$ and $C_{G:C}$ across measured temperature points are plotted in Fig. 6.6c.

We also determined $C_{A:T}$ and $C_{G:C}$ for CCends-AP6 and GCGcore-AP6 using t-2D IR spectroscopy in order to test the validity of the G:C and A:T spectral markers in the t-HDVE data. Spectra were acquired at T-jump delays of 180 ns, 10 μs, and 320 μs to reflect the fast solvation and terminal fraying dynamics, half-dehybridization response, and duplex-to-single-strand transition, respectively (Fig. 6.7). We used the difference between 10 μs and 180 ns spectra as the τ_2 response and the difference between 320 μs and 180 ns surfaces for the total dissociation. The integrated signal change over certain regions of the absolute value difference spectra was used to determine G:C and A:T response. We used the G ring mode region (cyan) for the isolated G:C response and the A/T ring mode region (blue) for A:T. We also compare with the signal change of the T intrabase cross-peak (purple) between the carbonyl/ring vibration near 1660 cm^{-1} and ring vibration at 1630 cm^{-1} as well as the overlapping signal changes from G carbonyl and T carbonyl/ring vibrations (pink). The ratio between G:C (cyan) and A:T (blue) amplitudes determined on the 10 μs – 180 ns surface and 320 μs – 180 ns surface should be nearly equivalent to the $P_{G/A}$ values determined from t-HDVE data. Then, $C_{A:T}$ and $C_{G:C}$ can be determined from the

t-2D IR data using eqs. 6.D2a,b. The *C* values determined with t-2D IR and t-HDVE data are nearly identical (Figs. 6.C1c and 6.6), which indicates that the minor interference between A/T and G/C cross-peak and ESA amplitude changes is negligible at 1550 and 1605 cm^{-1} in the t-HDVE spectra.

In addition to calculating the relative A:T and G:C loss contributions to the intermediate T-jump response, the fraction of total base pair dissociation during the intermediate response (D_2) can be estimated. We simply calculate D_2 as the weighted mean of G:C and A:T intermediate dissociation percentages.

$$D_2 = \frac{\left(N - N_A\right) P_{G:C} + N_A P_{A:T}}{N} \tag{6.D3}$$

In contrast to $C_{G:C}$ and $C_{A:T}$, D_2 strongly depends on temperature and typically decreases sharply towards higher temperatures. The temperature trends of D_2 are shaped by the temperature-dependence of the duplex-to-single-strand response (A_3), which is determined by the change in duplex and single-strand concentration during the T-jump.

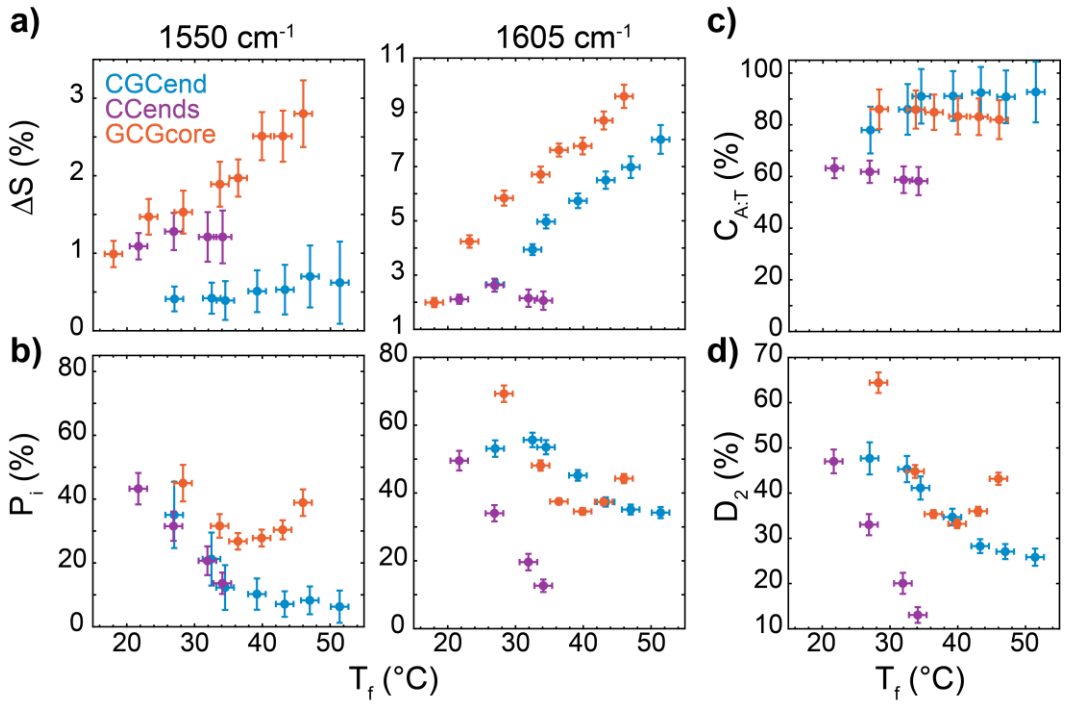

Figure 6.D1 Calculation of A:T/G:C character in τ_2 response. (a) Amplitudes of second component (DAS 2) probed at 1550 and 1605 cm^{-1} from global lifetime fitting of t-HDVE data for CGCcap-AP6, CCends-AP6, and GCGcore-AP6 as a function of T_f. **(b)** Amplitudes in **(a)** divided by the sum of DAS 2 and DAS 3 amplitudes (eq. 6.D1). **(c)** A:T character of τ_2 response at each temperature from eq. 6.D2. **(d)** Fraction of total dehybridization response occurring during τ_2 determined using eq. 6.D3.

Appendix 6.E Characterization of hybridization trajectories

6.E.1 Nucleation properties from hybridization trajectories

We analyze transition paths for hybridization by monitoring the trajectories for each individual base-pair separation. Trajectories are cut to the time interval between the first base-pair separation distance falling below 2.5 nm and when all distances fall below 0.7 nm. A few trajectories are shown for CGCcap in Fig. 6.E1, exemplifying the diversity of possible hybridization pathways. In one example, base pairs first form (defined by crossing below a 0.7 nm

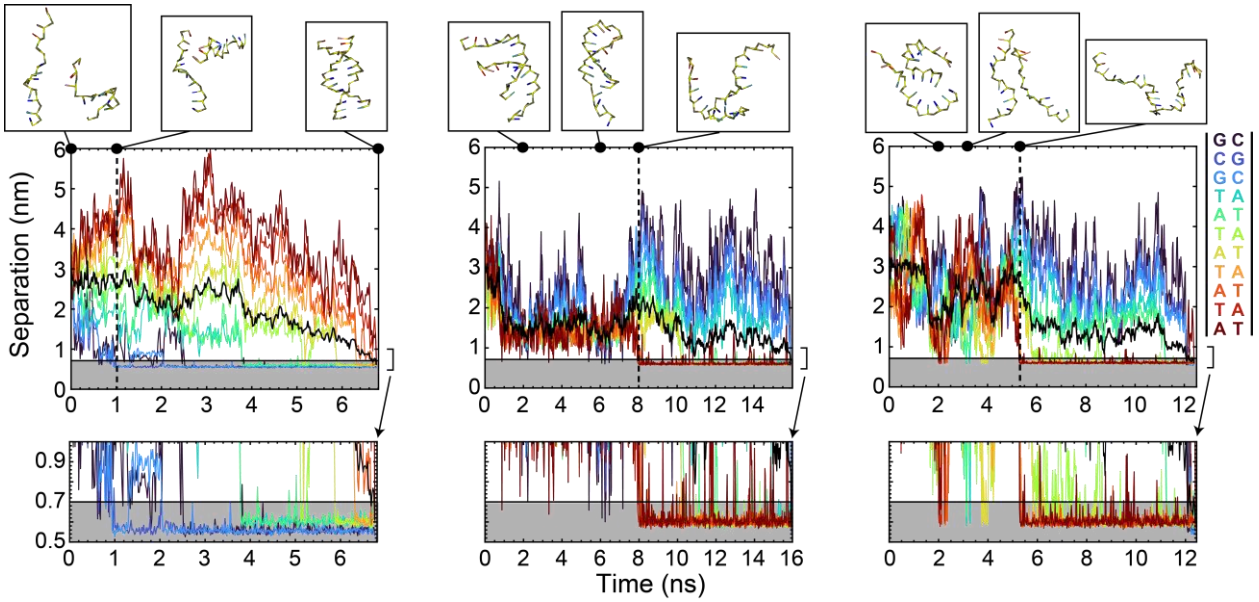

Figure 6.E1 Diversity of hybridization pathways from 3SPN.2 MD simulations. Different types of hybridization pathways using CGCcap as an example. Each color trajectory corresponds to a different base-pair separation distance (color code on the right) and the black line corresponds to the average over all base-pair separations (r_{bp}, eq. 5.1a). Trajectories are cut to the time interval between the first base pair separation falling below 2.5 nm and when all distances fall below 0.7 nm. Trajectories are plotted with a 20 ps time step. The solid horizontal line and shaded area indicate the 0.7 nm threshold for base pairing. Snapshots at indicated time points are shown above each trajectory. The bottom row shows an enlarged view of separations 0.5 – 1.0 nm. Vertical dashed lines indicate the time point of the first permanent base-pair formation (nucleation). (Left) Nucleation of G:C-rich region followed by zippering of the remaining base pairs. (Center) Initial formation of non-native base pairs brings both strands into close contact followed by nucleation on the A:T end followed by zippering of the remaining base pairs. (Right) Multiple failed encounters prior to nucleation on the A:T end.

threshold) in the G:C-rich cap while the remaining parts of the duplex are disordered. One of the G:C base pairs remains intact for the rest of the trajectory, and we refer to this permanent base-pair formation as nucleation. The remaining base pairs form in order as expected for a nucleation-zipper process.[31] Other trajectories show a more complex encounter where base pairs initially form out-of-register or strands collide in configurations that lack in-register base pairs. These

configurations bring contacts into close proximity such that nucleation can occur on the weaker A:T end of the duplex.

The study of duplex hybridization is largely rooted in the concepts of the nucleation-zipper model (introduced in Chapter 1) where formation of a small number of base pairs is the bottleneck for hybridization. The characteristics of nucleation certainly depend on the oligonucleotide sequence as evidenced by the variable free-energy barrier positions observed in Figs. 5.2 and 6.3 and experimental reports of sequence-dependent association rates.[21,22] However, properties of nucleation are rarely examined directly from hybridization trajectories. As shown in Fig. 6.11, we identify the nucleation position in each trajectory as the first permanent base pair to form (N-site) and define a nucleation patch (N-patch) as all base pairs that are intact at the time of nucleation (t_N). Histograms over all trajectories reveal that each sequence contains a different distribution of N-patch sizes (N_s, number of base pairs in the N-patch) and N-patch positions. The average N-patch size, $\langle N_s \rangle$, increases from CGCcap (2.10) to GCGcore (2.64) to CCends (3.20) as shown in Fig. 6.13. This trend roughly matches the peak position of the free-energy barrier along n_{bp} (Fig. 5.2), although we do not have FEPs at 295 K for direct comparison. It is important to note that $\langle N_s \rangle$ is sensitive to the choice of base-pair separation cutoff, and we choose 0.7 nm because it is the shortest separation in a relatively flat region of sensitivity (Fig. 6.E2). The N-patch position distributions show that nucleation is most likely to occur near G:C base pairs regardless of their location. Terminal G:C base pairs often participate in the N-patch but are rarely the N-site due to their tendency to fray during the zippering process (see magnified trajectories in Fig. 6.E1). The greater free-energy benefit from forming G:C base pairs leads to smaller N-patches when

nucleating in G:C-rich regions. Nucleation is increasingly localized onto G:C-regions in order from CCends to GCGcore to CGCcap, contributing to a reduction in $\langle N_s \rangle$ across these sequences.

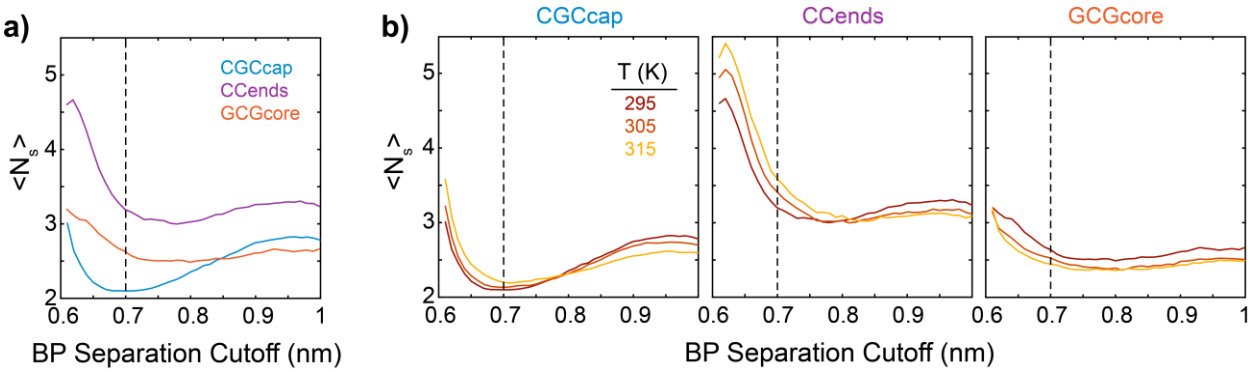

Figure 6.E2 Dependence of N-patch size on base-pair separation cutoff. (a) $\langle N_s \rangle$ determined using base pair separation cutoffs ranging from 0.61 to 1 nm for each sequence at 295 K. A vertical dashed line at 0.7 nm indicates the cutoff used for analysis in this chapter. **(b)** $\langle N_s \rangle$ at 295, 305, and 315 K over the same range of base pair separation cutoffs for each sequence.

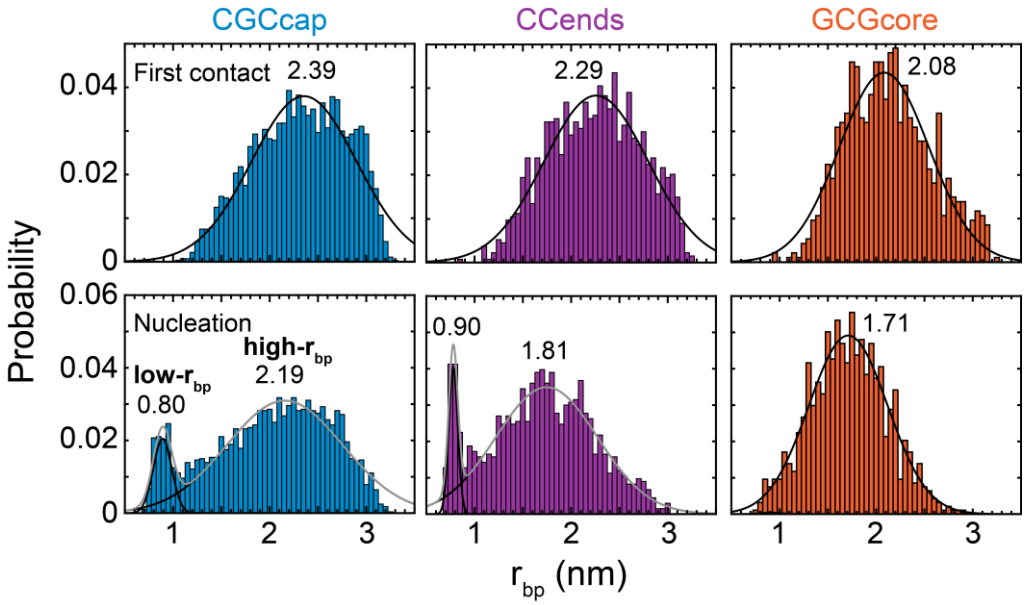

Figure 6.E3 Average base-pair separation during first contact and nucleation. (top) Probability distributions of r_{bp} at the time point of first base-pair formation (not permanent) from simulations at 295 K. Black solid line is a single Gaussian fit. (bottom) Probability distributions of r_{bp} at t_N. Distributions of CGCcap and CCends are fit to a sum of two Gaussians (gray solid line, low-r_{bp} and high-r_{bp}) and that of GCGcore is fit to a single Gaussian. Peak center positions are listed for each fit.

In addition to a distribution of possible base pairs, the N-patch may contain a spread of strand configurations. Figure 6.E3 shows the distribution of r_{bp} at the time of first base-pair formation (not necessarily permanent) and nucleation. The first-contact distributions have similar widths (FWHM ~1.5 nm) for each sequence and are shifted 0.2 – 0.4 nm higher in r_{bp} than the distributions at nucleation. In contrast, the nucleation distributions show great variability across sequences. Each sequence shows a broad Gaussian-like region centered at r_{bp} = 1.7 – 2.2 nm, but CGCcap and CCends contain an extra peak at r_{bp} < 1 nm. The center of the broad peak shifts lower from CGCcap to CCends to GCGcore due to different positions of nucleation. Nucleation at the G:C-rich end in CGCcap allows highly separated A:T ends that lead to large r_{bp}. In contrast, CCends and particularly GCGcore tend to nucleate near the center of the oligonucleotides, producing lower r_{bp} values because each dissociated segment is short relative to the A:T segment in CGCcap. The low-r_{bp} peak corresponds to the formation of highly aligned duplex configurations before nucleation, and the peak persists for distributions at the time points of second and third permanent base pair formation as well as when using different base-pair separation cutoffs (Figs. 6.E4 and 6.E5). These low-r_{bp} configurations likely arise when strands form out-of-register contacts (Fig. 6.E1 and Chapter 5.3) or undergo numerous failed base-pairing events before nucleation. We define the number of pre-nucleation encounters (n_{pre}) as the number of transitions between strands containing no contacts (all individual separations above 0.7 nm) and those with at least one contact prior to t_N (Fig. 6.E6a). While the high-r_{bp} peak is dominated by events with n_{pre} = 0 – 10, the low-r_{bp} peak exhibits a broad distribution of n_{pre} spanning from 0 to >70. Visualizations of trajectories with large n_{pre} often indicate out-of-register contacts or

highly aligned strand configurations as exemplified in Fig. 6.E1 for CGCcap. Simple DNA nearest-neighbor calculations (ignoring contributions from dangling ends and overhangs) indicate that the possible out-of-register configurations from CGCcap and CCends are 12 – 15 kJ/mol more stable than with GCGcore, potentially explaining the lack of low-r_{bp} peak in GCGcore.

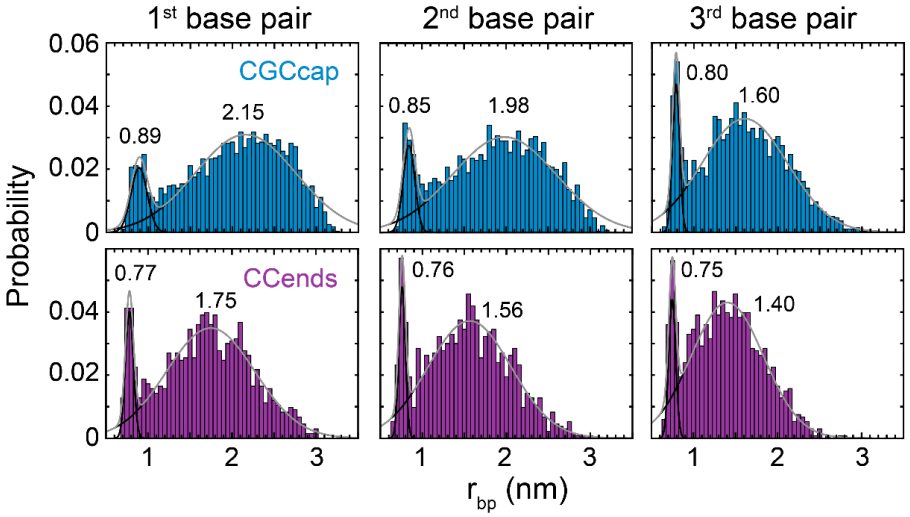

Figure 6.E4 Average base-pair separation across permanent formation of multiple base pairs. Probability distributions of r_{bp} for CGCcap and CCends at the time point of (left) nucleation and permanent formation of (center) the second base pair and (right) the third base pair from simulations at 295 K. Distributions are fit to a sum of two Gaussians (gray solid line) and peak

The data in Fig. 6.E6b indicates that the position of first contact and nucleation are linked to the r_{bp} nucleation distribution. For example, nucleation in the A:T region of CGCcap typically occurs at lower r_{bp} relative to nucleation in the G:C region (Figs. 6.E1 and 6.E6b). First contact at A:T base pairs 6 – 10 also tends toward nucleation with r_{bp} < 1.5 nm even if nucleation occurs at the G:C side, which may result from a tendency for initial A:T contacts to briefly form out-of-register. Such position-dependent trends are less pronounced for CCends. First contact and nucleation near the center leads to slightly lower r_{bp} than at the ends, but nucleation with r_{bp} < 1.0

nm appears to dominantly come from first contact and nucleation at the G:C ends. Lastly, there is essentially no dependence of nucleation r_{bp} on first contact or nucleation position for GCGcore within the error of our data.

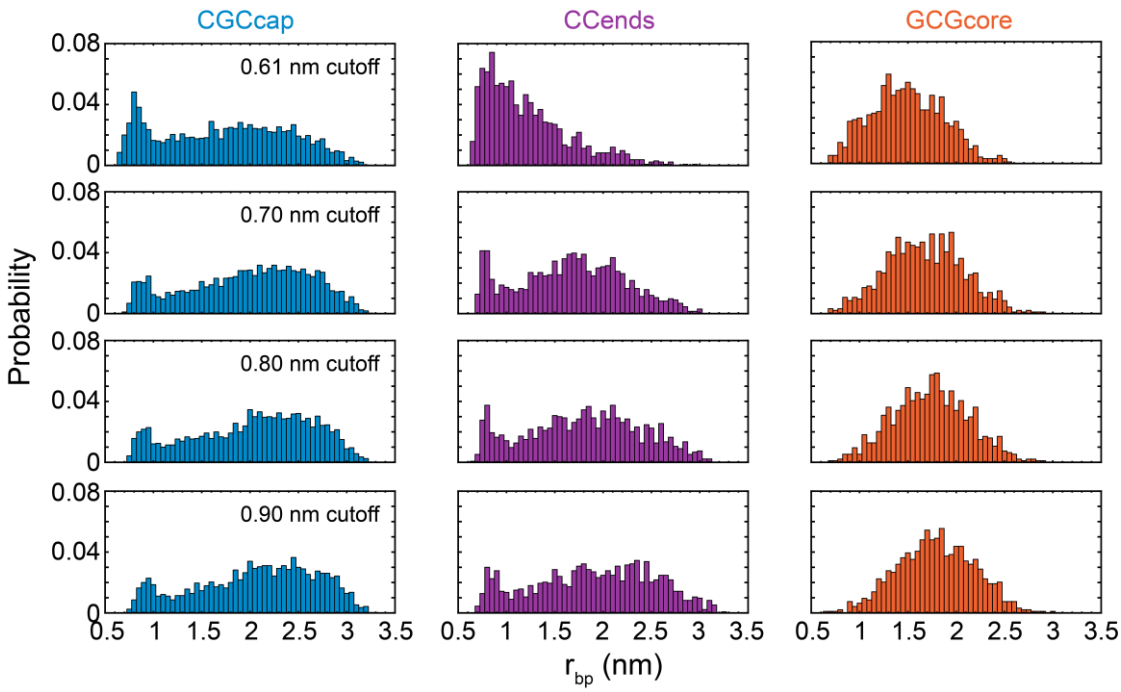

Figure 6.E5 Dependence of r_{bp} distributions at nucleation on separation cutoff for base pairing. r_{bp} distributions at t_N calculated using base-pair separation cutoffs of 0.61, 0.70, 0.80, and 0.90 nm. The greatest change is observed between 0.61 and 0.70 nm.

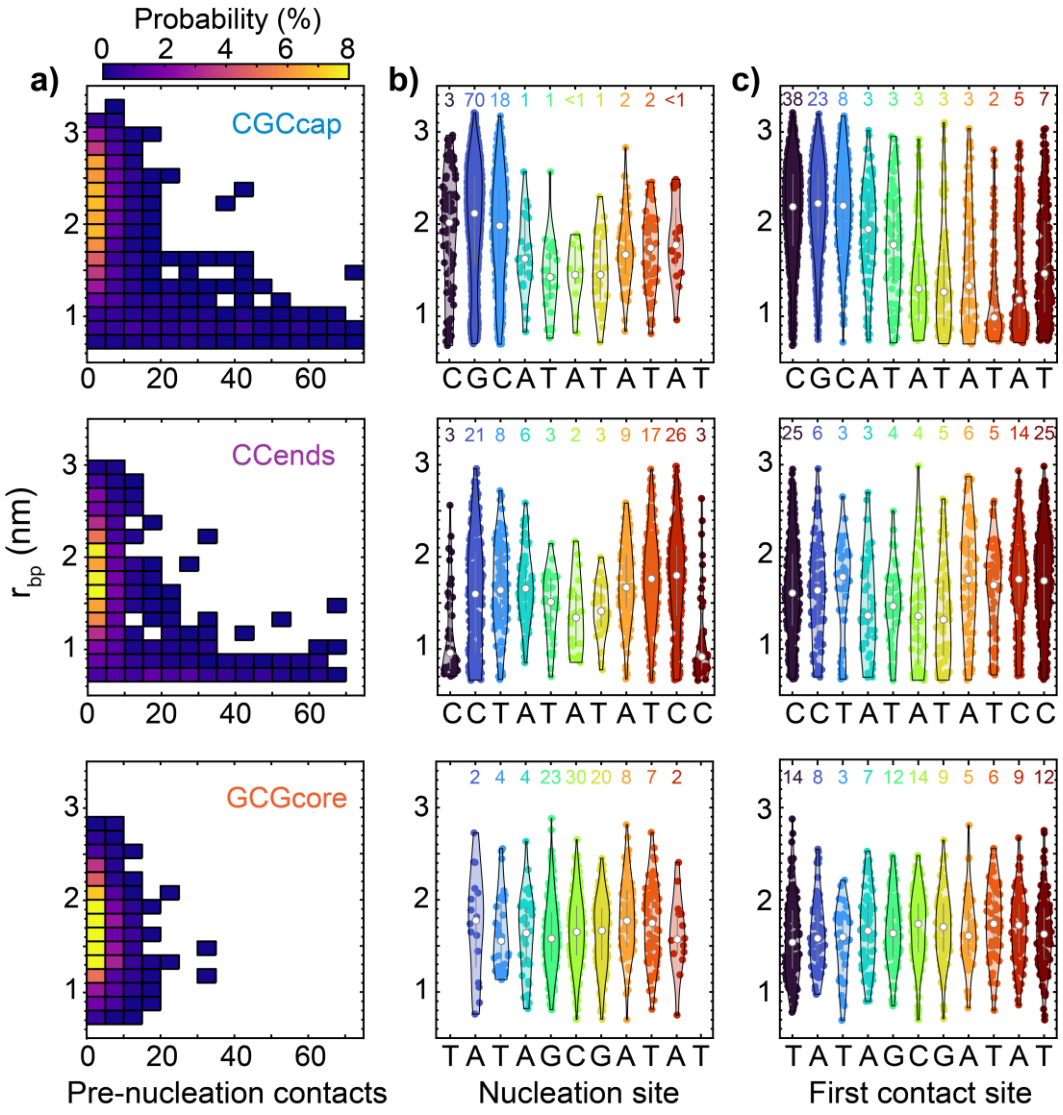

Figure 6.E6 Connection between nucleation position and average base-pair separation. (a) 2D probability distributions as a function of the number of pre-nucleation contacts (n_{pre}) and r_{bp}. n_{pre} corresponds to the number of times a duplex transitions from having all base pairs broken to at least one base pair intact prior to t_N. **(b)** Violin plots showing the distribution of r_{bp} at t_N upon nucleation at each base-pair site. White circles correspond to median values and gray lines correspond to interquartile ranges. Only data points within the violin area are plotted. Violin plots were generated using the MATLAB function from https://github.com/bastibe/Violinplot-Matlab.[74] **(c)** Similar violin plots for distribution of r_{bp} at time of first base-pair formation. Numbers above each distribution indicate the percentage of events at that position.

6.E.2 Temperature-dependent nucleation properties

Thermodynamics and kinetics from 3SPN.2 simulations

The balance of enthalpic and entropic effects in DNA hybridization makes its thermodynamics and kinetics highly sensitive to temperature. Temperature may be tuned experimentally to infer aspects of the free-energy landscape as illustrated throughout this thesis. Here we examine the effect of temperature on mechanistic aspects of hybridization through 3SPN.2 MD simulations. It is important to note that MD force fields, particularly coarse-grained force fields, are often poorly parameterized to describe temperature-dependent characteristics of the desired free-energy landscape.[69-70] Therefore, it is first important to evaluate how accurately 3SPN.2 predicts temperature-dependent thermodynamic and kinetic observables for hybridization.

As illustrated in Fig 6.E7, a two-state analysis of FEPs generated from 3SPN.2 simulations at multiple temperatures produces duplex melting curves.[18] The melting temperature of each sequence is 5 – 10 °C lower than from experimental melting curves. 3SPN.2 was previously shown to most accurately predict melting temperatures under medium ionic strength conditions (I = 60 – 100 mM),[16] but our simulations and experiments are performed at I = 600 mM. Although the melting temperatures are shifted, the melting curve shapes are qualitatively similar to the experimental data. Fitting the total probability density of the duplex state at each temperature to a two-state hybridization model allows us to extract the internal-energy change (ΔU_h°) and entropy change (ΔS_h°) for hybridization (Appendix 6.A). These parameters are used to describe the temperature-dependence of the Helmholtz free energy for hybridization, $\Delta F_h^\circ = \Delta U_h^\circ - T\Delta S_h^\circ$. The ΔU_h° and ΔS_h° values are of similar magnitude to the hybridization enthalpy (ΔH_h°) and ΔS_h°

extracted from experimental melting curves (Table 6.E1), but sequence-dependent trends are only partially captured. GCGcore is accurately found to have the lowest ΔU_h° and ΔS_h° values among the three sequences. Values of CGCcap are greatly overestimated, leading to a sharper melting curve than in experiment.

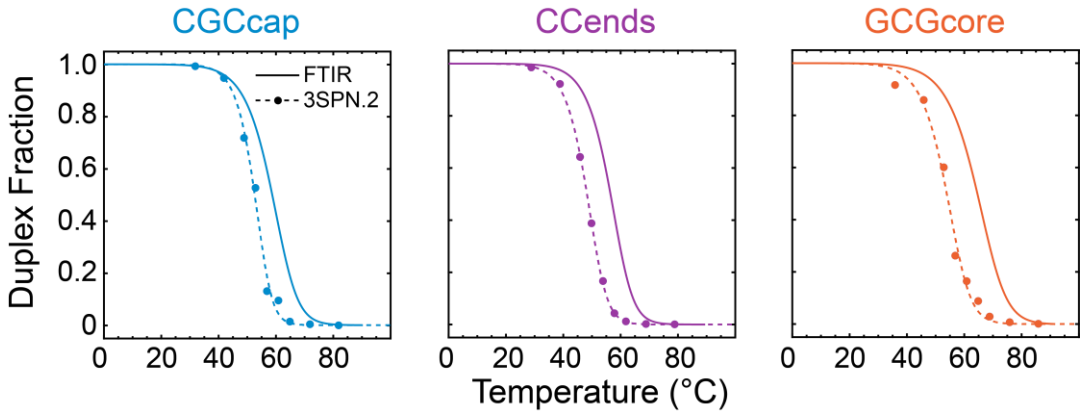

Figure 6.E7 Comparison of experimental and 3SPN.2 duplex melting curves. Integrated probability density of the duplex state ($1/r_{bp} > 1/r_{bp,max}$) along $1/r_{bp}$ at each temperature (circles) and fits to a two-state model for hybridization of non-self-complementary oligonucleotides (dashed lines, Appendix 6.A). Solid lines correspond to melting curves extracted from two-state analysis of FTIR temperature series (Appendix 6.A). Experiments were performed with an effective oligonucleotide concentration of 2 mM and [Na^+] = 600 mM. Simulated FEPs were analytically extended to reduce the oligonucleotide concentration from 5.3 to 2 mM.[18]

Table 6.E1 Comparison of hybridization thermodynamics and kinetics from experiment and 3SPN.2 MD simulations. ΔH, ΔU, ΔF, and ΔG are in units of kJ/mol and ΔS is units of J/molK. ΔU and ΔF are reported for 3SPN.2 rather than ΔH and ΔG, respectively, because simulations were carried out in the NVT ensemble.

Sequence	ΔH°_h	ΔU°_h	ΔS°_h	ΔS°_h	$\Delta G^\circ_{h,37}$	$\Delta F^\circ_{h,37}$	$\Delta H^\ddagger_h$	$\Delta U^\ddagger_h$	$\Delta H^\ddagger_d$	$\Delta U^\ddagger_d$
	Exp.[a]	3SPN.2	Exp.[a]	3SPN.2	Exp.[a]	3SPN.2	Exp.[b]	3SPN.2	Exp.[b]	3SPN.2
CGCcap	-310	-409	-868	-1192	-40.9	-39.2	-55	-71	255	319
CCends	-342	-345	-970	-1010	-41.5	-31.7	-59	-101	283	251
GCGcore	-284	-303	-779	-863	-42.8	-35.3	-44	-36	240	215

[a]Determined from two-state analysis of FTIR temperature series. [b]Determined from two-state kinetic analysis of T-jump data.

Temperature-dependent hybridization kinetics from 3SPN.2 are more difficult to properly evaluate relative to melting thermodynamics. Recent work demonstrated that the 3SPN2-determined rate constant for oligonucleotide dehybridization (k_d) follows an exponential temperature-dependence with an internal-energy barrier ($\Delta U^\ddagger_d$) comparable to the experimental enthalpic barrier ($\Delta H^\ddagger_d$), yet the dependence of $\Delta U^\ddagger_d$ on sequence poorly matched experiment.[18] Further, the rate constant for terminal A:T fraying (k_{fray}) from 3SPN.2 follows a linear temperature-dependence as observed in T-jump IR experiments.[18] The study also found that k_d is accelerated 10-fold and terminal fraying 120-fold relative to experiment, which is a common characteristic of coarse-grained force fields due to smoothing of the underlying free-energy landscape.[71]

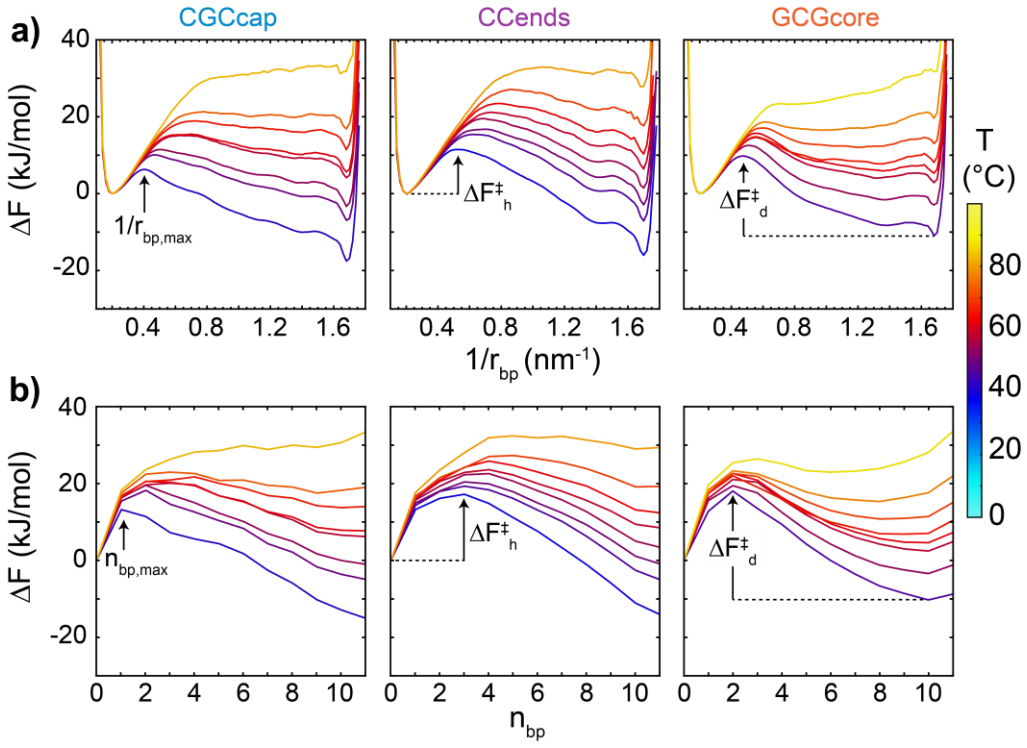

Figure 6.E8 Temperature-dependent free-energy profiles for hybridization. FEPs along **(a)** $1/r_{bp}$ and **(b)** n_{bp} at 8 – 9 temperatures for each sequence and calculated from 3SPN.2 MD simulations with WTMetaD. The hybridization barrier maximum ($1/r_{bp,max}$ and $n_{bp,max}$) is indicated for CGCcap, the hybridization free-energy barrier height ($\Delta F_h^{\ddagger}$) is indicated for CCends, and the dehybridization free-energy barrier height ($\Delta F_d^{\ddagger}$) is indicated for GCGcore.

Figure 6.E8 shows FEPs along $1/r_{bp}$ and n_{bp} determined from 3SPN.2 simulations carried out with WTMetaD at 8 – 9 temperature points ranging from 20 °C below to 20 °C above the respective melting temperature of each sequence. FEPs are shifted such that the single-strand free-energy minimum corresponds to ΔF = 0. For each CV, the hybridization ($\Delta F_h^{\ddagger}$) and dehybridzation ($\Delta F_d^{\ddagger}$) free-energy barriers are defined as the difference between the single-strand and duplex minimum free energy, respectively, and the maximum free-energy point between single-strand duplex states. These quantities are purely thermodynamic and do not necessarily have

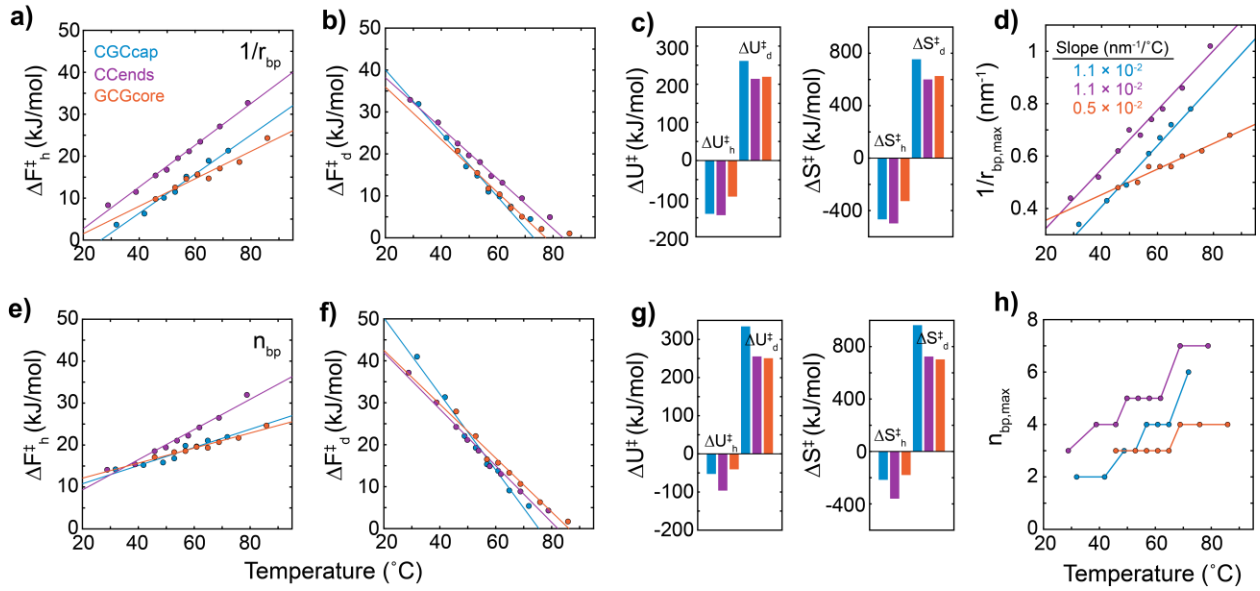

Figure 6.E9 Temperature-dependent trends in hybridization free-energy barrier. (a) Hybridization ($\Delta F_h^{\ddagger}$) and **(b)** dehybridization free-energy barrier height ($\Delta F_d^{\ddagger}$) extracted from FEPs along $1/r_{bp}$ as a function of temperature. **(c)** Internal-energy barriers ($\Delta U^{\ddagger}$) and entropic barriers ($\Delta S^{\ddagger}$) extracted from linear fits in **(a)** and **(b)** to $\Delta F^{\ddagger} = \Delta U^{\ddagger} - T\Delta S^{\ddagger}$. **(d)** Position of free-energy barrier maximum along $1/r_{bp}$ ($1/r_{bp,max}$) as a function of temperature. Solid lines correspond to linear fits and the slope from each fit is listed. **(e – h)** Similar plots for FEPs along n_{bp}.

any correlation to the actual timescales of hybridization or dehybridization. However, the dynamics of hybridization are well described along $1/r_{bp}$ and n_{bp} (at least for CGCcap and CCends), therefore we expect $\Delta F_h^{\ddagger}$ and $\Delta F_d^{\ddagger}$ to be reasonably correlated with timescales for hybridization and dehybridization. $\Delta F_h^{\ddagger}$ increases linearly with temperature while $\Delta F_d^{\ddagger}$ decreases linearly with temperature. The negative slope of $\Delta F_h^{\ddagger}$ reflects negative internal-energy ($\Delta U_h^{\ddagger}$) and entropic ($\Delta S_h^{\ddagger}$) contributions to the hybridization barrier that are commonly observed experimentally and similar in magnitude to values for our model sequences (Table 6.E1).[19,72] In contrast, the internal-energy ($\Delta U_d^{\ddagger}$) and entropic ($\Delta S_d^{\ddagger}$) contributions to $\Delta F_d^{\ddagger}$ are large and positive.

Although the sequence-dependence of each barrier contribution is qualitatively similar along $1/r_{bp}$ and n_{bp}, internal-energy and entropic values from n_{bp} FEPs are ~50 kJ/mol and ~100 J/molK more positive than from $1/r_{bp}$ FEPs. The $\Delta U_h^{\ddagger}$ and $\Delta U_d^{\ddagger}$ values from n_{bp} FEPs better agree with experimental values but do not accurately capture trends across sequences.

As $\Delta F_h^{\ddagger}$ increases and $\Delta F_d^{\ddagger}$ decreases with increasing temperature, the position of the barrier maximum ($1/r_{bp,max}$ or $n_{bp,max}$) shifts toward the duplex state in accordance with the Hammond postulate.[73] This shift indicates that the transition state for hybridization contains more base pairing as the temperature increases. The shift in $1/r_{bp,max}$ is roughly linear as a function of temperature with a slope of 0.5×10^{-2} nm^{-1} K^{-1} for GCGcore and 1.1×10^{-2} nm^{-1} K^{-1} for CGCcap and CCends. A similar sequence-dependence is found for $n_{bp,max}$, but the n_{bp} CV is too coarse to extract a meaningful slope. The sequence-dependent shifts in barrier position may reflect real differences in the transition-state ensemble as temperature changes, or it may be an artifact arising from the poor quality of $1/r_{bp}$ and n_{bp} CVs for describing GCGcore hybridization.

Temperature-dependent nucleation dynamics

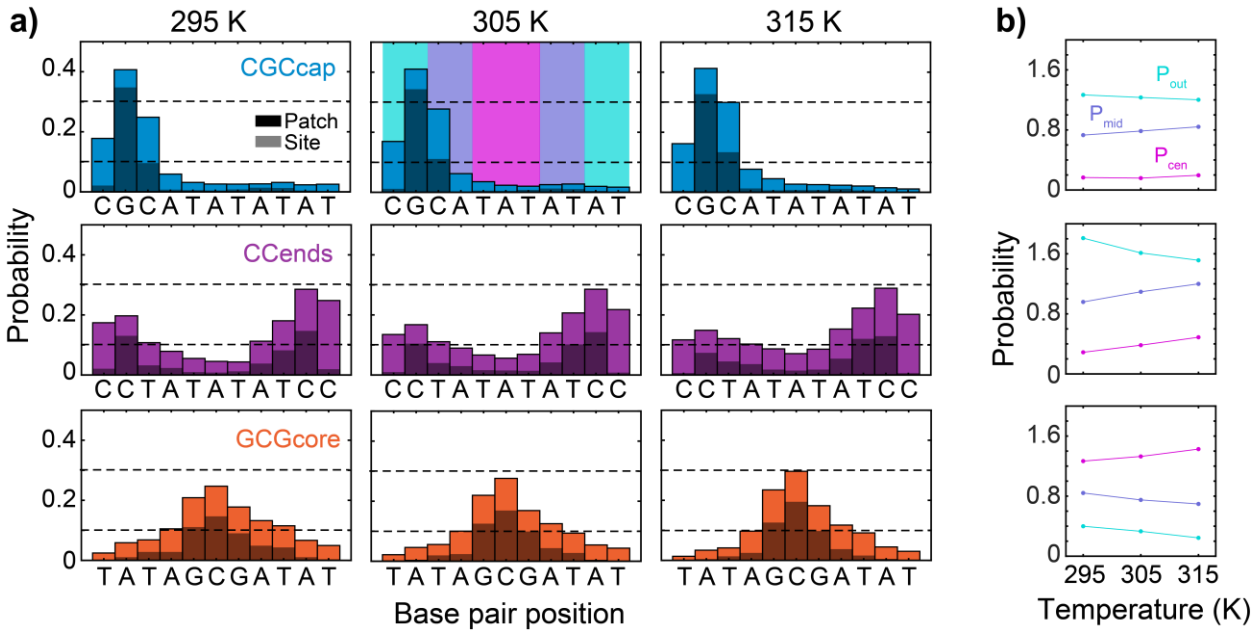

Figure 6.E10 Temperature-dependent nucleation patch. (a) Probability that a given base-pair site is part of the N-patch (colored bars) or the nucleation site (gray bars) for simulations performed at (left) 295 K, (center) 305 K, and (right) 315 K. Horizontal dashed lines at 0.1 and 0.3 are drawn as reference points to compare across temperatures. **(b)** Probability that outer (P_{out}, cyan, base pairs 1-2 and 10-11), intermediate (P_{mid}, purple, base pairs 3-4 and 8-9), and central (P_{cen}, magenta, base pairs 5-7) base pairs are part of the N-patch as a function of temperature. The different regions are marked in the top-center panel of **(a)**. Probability of being in the N-patch shifts from outer to central base pairs as temperature increases.

Thousands of hybridization events were sampled for each sequence at 305 and 315 K using the same approach introduced for simulations at 295 K, and N-patch distributions at each temperature are shown in Fig. 6.E10. The N-patch distributions remain qualitatively similar from 295 to 315 K for all sequences, but there is a clear shift in probability density in CCends and GCGcore from base pairs near the termini to intermediate and central contacts (Fig. 6.E10b). This shift likely results from an entropic preference to nucleate in the center of the duplex due to the possibility of more zippering pathways, and the entropic preference receives greater weight as the

temperature increases. The favorable internal energy for base pairing at the G:C end in CGCcap outweighs the benefit of nucleating in the duplex center over this temperature range such that only a minor shift of probability density from outer to intermediate contacts is observed.

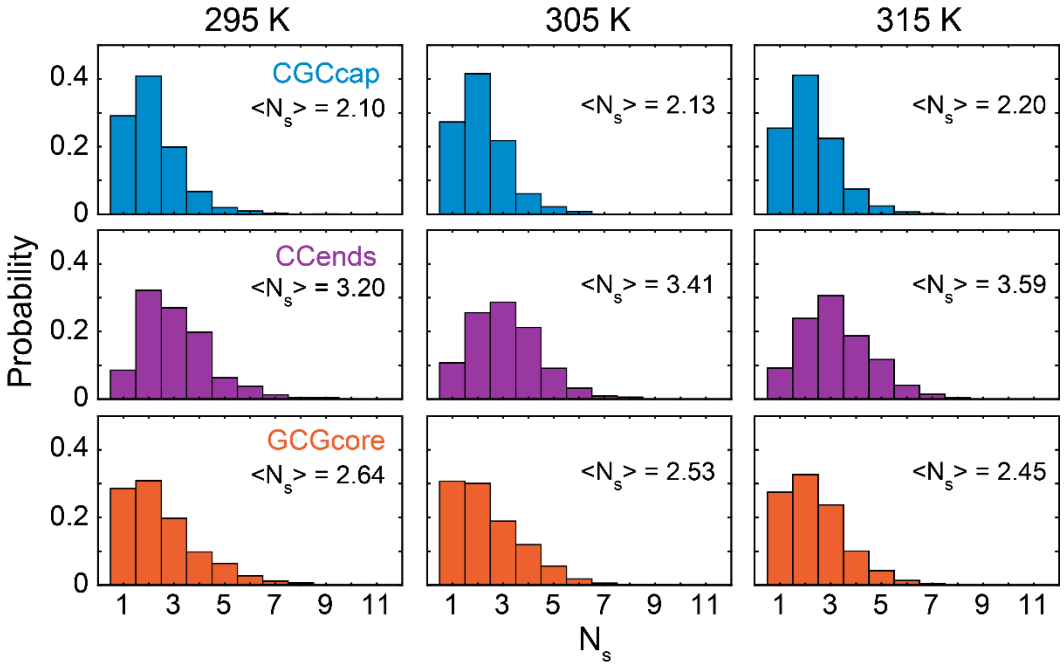

Figure 6.E11 Nucleation patch size from 295 to 315 K. Probability distribution of N-patch size (N_s) for each sequence at 295, 305, and 315 K. $\langle N_s \rangle$ is listed for each condition.

We also find that the average size of the N-patch shifts with temperature (Fig. 6.E11). CCends exhibits a 0.2 increase in $\langle N_s \rangle$ over each 10 K interval and reaches $\langle N_s \rangle$ = 3.6 at 315 K whereas CGCcap only shifts 0.1 in $\langle N_s \rangle$ from 295 to 315 K and GCGcore decreases by 0.2. The same qualitative trends are observed for any base-pair separation cutoff from 0.6 – 0.8 nm (Fig. 6.E2). These trends are consistent with shifts in the A:T/G:C content of the N-patch as more A:T contacts are needed to provide the same N-patch stability as a lesser number of G:C contacts. For example, the total percent contribution of G:C base pairs in the N-patch of CCends declines from

60% at 295 K to 49% at 315 K while GCGcore contains an increase from 43% to 53%. Assuming that the linear trend in barrier position vs. temperature from FEPs holds down to 295 K, we would expect $\langle N_s \rangle$ to increase linearly as temperature increases if it is correlated with $n_{bp,max}$ and/or $1/r_{bp,max}$. The N-patch trends of CGCcap and GCGcore are inconsistent with this expectation. Additionally, there is essentially no change in the width or center position of the high-r_{bp} distribution at t_N as a function of temperature for any sequence (Fig. 6.E12). A minor temperature-dependence in the average value of r_{bp} at nucleation occurs regardless of the base-pair separation cutoff (Fig. 6.E13). These observations suggest that our nucleation metric is poorly correlated with the ensemble at $1/r_{bp,max}$ or $n_{bp,max}$, particularly for CGCcap and GCGcore.

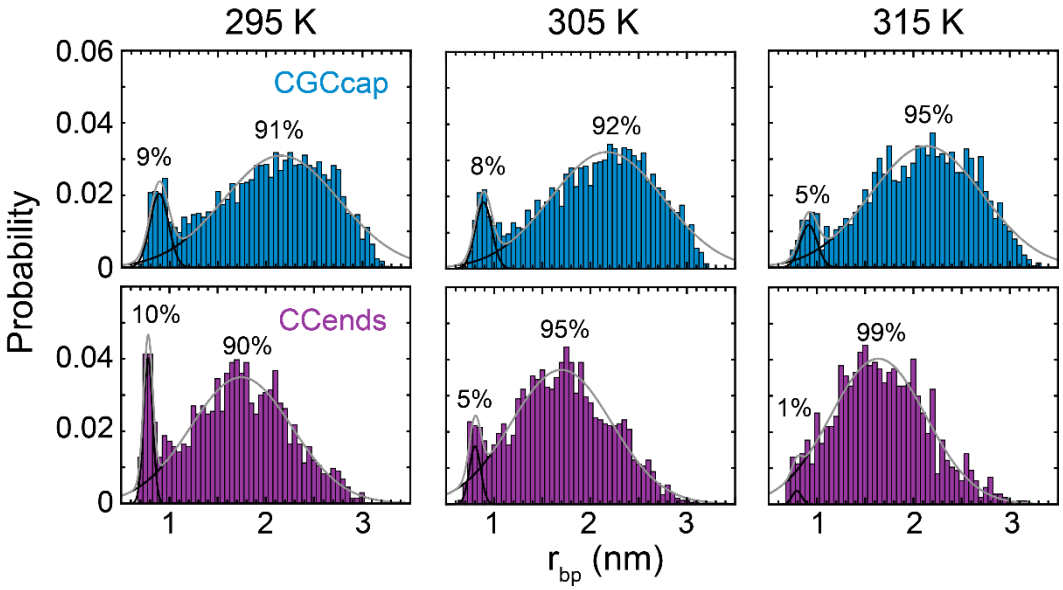

Figure 6.E12 Temperature-dependent r_{bp} distributions during nucleation. (top) Probability distributions of r_{bp} at t_N for CGCcap and CCends from simulations at 295, 305, and 315 K. Distributions are fit to a sum of two Gaussians (gray solid line). The percentage that each Gaussian contributes to the total integrated amplitude of the probability distribution is listed at each temperature. The population of the low-r_{bp} peak decreases as temperature increases and the position of the high-r_{bp} peak remains unchanged.

Although the high-r_{bp} peak is weakly sensitive to temperature, there is a clear reduction in the amplitude of the low-r_{bp} peak as temperature increases (Fig. 6.E12). In CCends, this peak is almost fully eliminated by 315 K whereas the peak remains with 5% population for CGCcap. Based on the discussion in Chapter 5.3 and Appendix 6.E.1, the loss of the low-r_{bp} peak at higher temperature suggests a reduction in out-of-register base-pairing configurations prior to nucleation, consistent with an entropic penalty for duplex formation. CGCcap contains more stable potential out-of-register configurations than CCends, which is consistent with the persistence of the low-r_{bp} peak at higher temperatures.

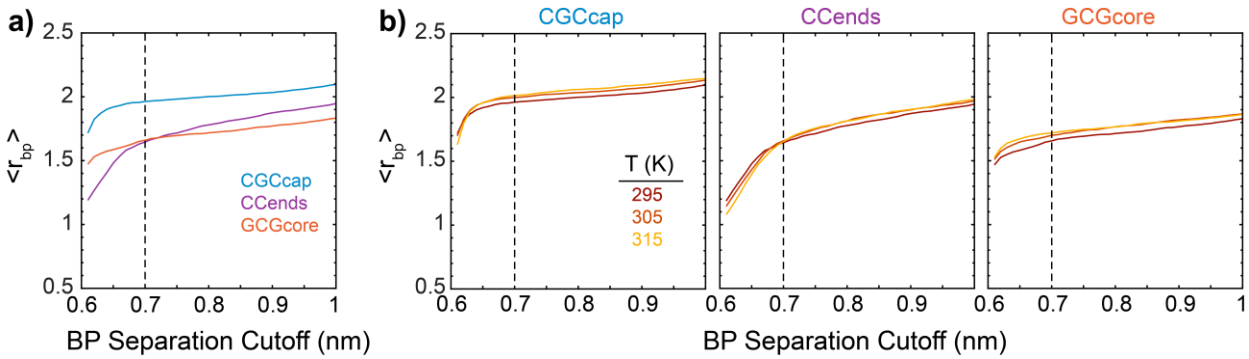

Figure 6.E13 Dependence of average r_{bp} at nucleation on separation cutoff for base pairing and temperature. (a) Average r_{bp} at t_N, $\langle r_{bp} \rangle$, as a function of base pair separation cutoff from 0.61 to 1 nm. A vertical dashed line at 0.7 nm indicates the cutoff used for analysis presented in Section 5.3. **(b)** $\langle r_{bp} \rangle$ at 295, 305, and 315 K over the same range of base pair separation cutoffs for each sequence.

Chapter 7

Abasic-site position and sequence effects on DNA duplex stability and dynamics

The material in this chapter is adapted from a manuscript that is under review:

Ashwood, B.; Jones, M. S.; Lee, Y.; Sachleben, J. R.; Ferguson, A. L.; Tokmakoff, A., Molecular insight into how the position of an abasic site and its sequence environment influence DNA duplex stability and dynamics. *BioRxiv* **2023**, https://doi.org/10.1101/2023.07.22.550182.

7.1 Introduction

Design of nucleic acid technologies such as microarrays for sequencing and gene expression profiling, DNA-PAINT for imaging, dynamic DNA devices, and CRISPR-Cas systems for gene editing rely on a predictive understanding of duplex hybridization and base pairing in oligonucleotides.[1-4] Most important for these areas is a fundamental understanding of hybridization thermodynamics, kinetics, and dynamics as a function of nucleobase sequence and for non-canonical base-pairing interactions. Although molecular interpretations for sequence-dependent duplex stability are still developing,[5-6] nearest-neighbor (NN) models provide an empirical yet quantitative prediction of duplex binding free energy,[7-8] and efforts have been undertaken to predict hybridization kinetics.[9-12] Numerous types of nucleobase chemical modifications and lesions are also known to influence the stability of base pairing,[13-15] and these effects are often quantified as a function of local nucleobase sequence around the modified site in NN models.[16-18] Additionally, end effects are likely to cause the position of a modification within a short oligonucleotide to influence the binding stability. Base-pair mismatches have been reported to be

more destabilizing when located progressively inward from the duplex termini,[19-22] and this behavior may be general for any destabilizing modification. Such position dependence likely results from an interplay of position-dependent entropic penalties and disruption of local base-pairing dynamics, but details of the underlying molecular mechanisms at play are not yet resolved.

Formation of an abasic site (apurinic/apyrimidinic or AP site) results from the loss of a nucleobase through spontaneous or enzymatic cleavage of the glycosidic bond and is one of the most naturally abundant DNA lesions.[23-24] Relative to mismatches, which may exhibit an array of non-canonical base-pairing configurations and structural deformations depending on their sequence context,[25-27] AP sites exhibit minimal perturbation to DNA duplex structure.[28-29] Therefore, AP sites can be used to probe positional effects on duplex stability and base-pairing dynamics with minimal complexity added from non-canonical base pairing interactions. Previous work showed that duplex destabilization arising from an AP site at the center of an oligonucleotide strongly depends on the identity of the adjacent base pairs.[30-32] We recently demonstrated that the degree of destabilization from a central AP site depends on the full oligonucleotide sequence in addition to the adjacent base pairs, a consequence arising from a sequence-dependent free-energy penalty for nucleating base pairs on each side of the AP site.[33] However, it remains unclear how the position of the AP site will tune this penalty, the overall duplex destabilization, and the dynamics of base pairing in the oligonucleotide. These position-dependent dynamics have implications for the repair of AP sites and their mutagenic properties, such as in the stalling of transcription.[34]

We investigate the impact of shifting the AP-site position on the disruption of base-pairing stability and dynamics in multiple 11-mer template oligonucleotides with variable arrangement of guanine:cytosine (G:C) and adenine:thymine (A:T) base pairs. Temperature-dependent infrared (IR) and ^{1}H NMR spectroscopy demonstrate that AP sites increasingly disfavor duplex hybridization as they move inward from the terminus. A cooperative helix-coil model and molecular dynamics (MD) simulations employing the 3-site-per-nucleotide (3SPN.2) coarse-grained model[35-36] reveal that the position-dependent destabilization stems from a nucleation penalty for base pairing on each side of the AP site, which leads to a transition from multi-base-pair fraying when an AP site is near the termini to metastable half-dehybridization with central AP sites. The presence of these metastable partially-dehybridized duplex segments is directly resolved through temperature-jump IR spectroscopy (T-jump IR). Sequence-specific effects complicate the generality of these observations. For instance, ^{1}H NMR spectroscopy and all-atom MD simulations indicate that certain segment sequences may circumvent this nucleation penalty by forming out-of-register base pairs. Experimental results can consistently be interpreted with 3SPN.2 MD simulations to provide detailed insight into the mechanism by which sequence and AP-site position alter the free-energy landscape for duplex hybridization.

7.2 Position-dependent destabilization of the duplex by an AP site

7.2.1 Duplex destabilization depends on the position of an AP site

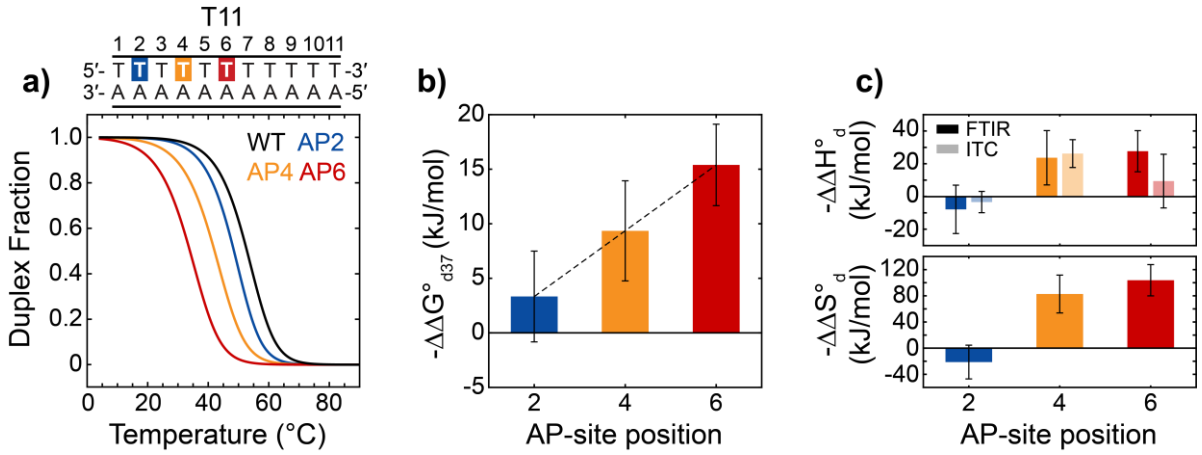

Figure 7.1 DNA duplex destabilization depends on AP-site position. (a) Duplex melting curves for a homogeneous A:T sequence that is unmodified (WT, black) or contains an AP site at the 2 (AP2), 4 (AP4), or 6 (AP6) position. Melting curves were extracted from a two-state analysis of FTIR temperature series (Appendix 6.A). **(b)** Change in hybridization free energy at 37 °C for each AP sequence with respect to the WT sequence, $\Delta\Delta G^{\circ}_{d37} = \Delta G^{\circ}_{d37} - \Delta G^{\circ}_{d37}(WT)$. The dashed line corresponds to a linear fit with a slope of 3 kJ/mol per base-pair index. **(c)** Change in dehybridization enthalpy ($\Delta\Delta H^{\circ}_{d}$, top) and dehybridization entropy ($\Delta\Delta S^{\circ}_{h}$, top) with respect to the WT sequence. Light $\Delta\Delta H^{\circ}_{d}$ bars correspond to values measured with ITC in non-deuterated solution.[42] FTIR and ITC $\Delta\Delta H^{\circ}_{d}$ values were shifted to 25 °C using the change in heat capacity (ΔC_p) of 1.05 J/molK measured for T11-WT with ITC. FTIR and ITC error bars correspond to 95% confidence intervals from two-state fits. An AP site is least destabilizing near the duplex termini and becomes increasingly destabilizing when moved inward.

An AP site destabilizes the duplex, and the magnitude of destabilization depends on the position of the AP site. We compare the thermodynamic impact of incorporating an AP site at the second (AP2), fourth (AP4), and central sixth (AP6) position relative to the unmodified form (WT) of a homogenous sequence 5′-TTTTTTTTTTT-3′ + complement (T11). Duplex melting curves extracted from a two-state thermodynamic analysis of FTIR temperature series show that the

melting transition midpoint (T_m) decreases by 5 to 20 °C as the AP site is shifted from the second to sixth position (Fig 7.1a). The change in the hybridization free energy between WT and AP sequences ($\Delta\Delta G^\circ_{d37} = \Delta G^\circ_{d37}(AP) - \Delta G^\circ_{d37}(WT)$) increases linearly in 6 kJ/mol increments from AP2 to AP4 to AP6 (Fig. 7.1b). The change in dehybridization van't Hoff enthalpy ($\Delta\Delta H^\circ_d$) and entropy ($\Delta\Delta S^\circ_d$) also increase by ~45 kJ/mol and ~120 J/molK over this range, respectively. After correcting to an equivalent temperature of 25 °C, isothermal titration calorimetry (ITC, Appendix 6.B) gives similar $\Delta\Delta H^\circ_d$ values to FTIR for T11-AP2 and T11-AP4 but a smaller value for T11-AP6 (Fig. 7.1c).

7.2.2 Position-dependent destabilization from segment nucleation barrier

The position-dependent destabilization from an AP site may be understood through a statistical treatment of duplex hybridization. Helix-coil (HC) models provide one of the simplest statistical descriptions of cooperative DNA duplex melting thermodynamics.[37-38] Base-pairing thermodynamics are described by two equilibrium constants: *s*, which is the Boltzmann factor of the free energy for adding an individual base pair next to an intact base pair (ΔG°_{int}) and σ, which is the Boltzmann factor of the free energy for nucleating a stretch of base pairs (ΔG°_{nuc}).

$$s = e^{-\Delta G^\circ_{\text{int}}/RT} \tag{7.1}$$

$$\sigma = e^{-\Delta G^\circ_{nuc}/RT} \tag{7.2}$$

Since this is a mean-field model, we apply the average value of *s* across all possible contacts within the duplex, $\langle s \rangle$, as determined from nearest-neighbor (NN) enthalpy (ΔH°_h) and entropy (ΔS°_h)

parameters.[7] The large penalty to nucleation makes σ << 1, which is the source of cooperativity in the model. We assume ΔG°_{nuc} is purely entropic such that σ is independent of temperature.

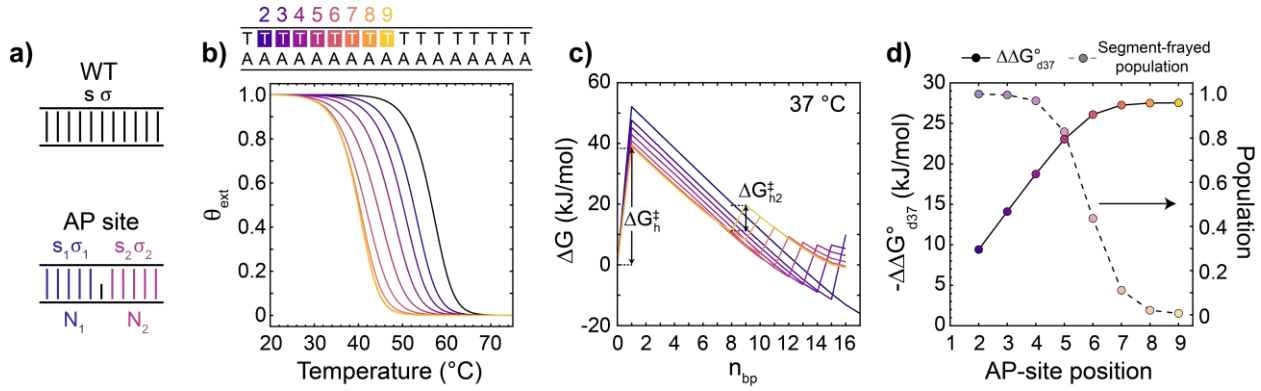

Figure 7.2 Helix-coil modeling of position-dependent duplex destabilization from an AP site. **(a)** Schematic of two-stretch helix-coil (HC) model used to model the impact of AP-site position on duplex stability. The WT duplex contains a single nucleation penalty (σ) and base-pair stability constant (s) applied uniformly to each base-pair site. An AP site splits the duplex into two base-pair stretches with their own s and σ values. **(b)** Fraction of duplexes containing one or more base pairs (θ_{ext}) for an unmodified T17:A17 sequence (black) and those containing an AP site at the 2 – 9 positions (purple to yellow). $\sigma_1 = \sigma_2 = 10^{-4}$ for all sequences. **(c)** Free-energy profiles at 37 °C for each sequence, illustrating barriers for strand association ($\Delta G^{\ddagger}_{h}$) and nucleation of the second-segment ($\Delta G^{\ddagger}_{h2}$). **(d)** $\Delta\Delta G^{\circ}_{d37}$ (dark circles and solid line) and the corresponding fraction of duplexes with the short base-pair segment completely frayed (light circles and dashed line). Shifting the AP site inward from the terminus leads to an increase in the number of frayed base pairs and greater destabilization of the duplex. The degree of destabilization levels off once the binding stability of the short segment is large enough to overcome the nucleation penalty.

It is generally unfavorable to have multiple disconnected stretches of base pairs in short duplexes, so we describe the WT system with an internal partition function ($Z_{int,D}$) limited to a single stretch of base pairs.[37-39]

$$Z_{\mathrm{int},D} = 1 + \sigma \sum_{n_{bp}=1}^{N} (N - n_{bp} + 1) \langle s \rangle^{n_{bp}} \tag{7.3}$$

N is the maximum number of possible base pairs, and n_{bp} is the number of intact base pairs for a given microstate. As detailed in Appendix 7.A, we can extend the HC model to treat an AP site as a defect that splits the duplex into two stretches of N_1 and N_2 base pairs (Fig. 7.2a). The internal partition function for the duplex can then be written as a product of two partition functions similar to eq. 7.3 for the two stretches, $Z_{int,D} = Z_{int,D1}Z_{int,D2}$. Each stretch has its own nucleation penalty (σ_1, σ_2) and $\langle s \rangle$ value, which corresponds to the average base-pair equilibrium constant across the respective stretch. Our usage of a second nucleation penalty is similar to incorporating a 'defect' free-energy penalty as in previous statistical models of duplexes containing a mismatch.[19]

To calculate melting curves from the HC model, we determine the fraction of oligonucleotide strands that contain one or more base pairs (θ_{ext}) as a function of temperature. θ_{ext} depends on both external and internal degrees of freedom of the system. We evaluate the external partition functions of the single-strand and duplex states as the number of possible ways of arranging single-strand and duplex molecules as a self-avoiding walk on a 3D cubic lattice,[40] allowing us to compute θ_{ext} in terms of $Z_{int,D}$.

$$\theta_{ext} = 1 + \frac{1 - \sqrt{1 + 2c_{tot}\gamma Z_{\mathrm{int},D}}}{c_{tot}\gamma Z_{\mathrm{int},D}} \tag{7.4}$$

$$\text{where } \gamma = 6c^{\circ} N_A V_{ss}$$

c° is the standard-state concentration of 1 M, N_A is Avogadro's constant, V_{ss} is the volume of a single-strand molecule, and c_{tot} is the total concentration of oligonucleotides. We neglect the

temperature-dependence of V_{ss} such that the temperature-dependence of θ_{ext} comes only from $Z_{int,D}$.

Figure 7.2b shows calculated θ_{ext} melting curves for an extended homogeneous sequence, T17:A17. Moving an AP site from the second to sixth base pair site leads to a gradual reduction in the duplex melting temperature, consistent with the experimental observation for T11. The shift in melting temperature becomes minor and eventually levels off as the AP site moves closer to the duplex center. This trend in the θ_{ext} curves may be understood from examining the free-energy profile (FEP) as a function on the number of intact base pairs (n_{bp}) calculated from the HC model (Fig. 7.2c). The WT sequence exhibits a single free-energy barrier to hybridization ($\Delta G_h^{\ddagger}$) between the single-strand (n_{bp} = 0) and duplex states that arises from the reduction in translation and configurational (set by σ) entropy upon binding of single strands, and formation of the remaining base pairs is cooperative and downhill in energy. In the HC model, the AP site disrupts this cooperativity and introduces an additional entropic barrier for forming base pairs on each side of the AP site ($\Delta G_{h2}^{\ddagger}$), and we recently experimentally and computationally verified the presence of $\Delta G_{h2}^{\ddagger}$ in duplexes with a central AP site.[33] The AP-site position controls the position of $\Delta G_{h2}^{\ddagger}$ along the hybridization FEP. When an AP site is at the second base-pair position, the adjacent terminal base pair does not have enough binding stability to overcome $\Delta G_{h2}^{\ddagger}$ and remains highly frayed at 37 °C. As the AP site shifts from the second to sixth base-pair position, the shorter duplex segment increases in length and base-pairing stability but remains highly frayed (Fig. 7.2c). The number of frayed base pairs increases with segment length, leading to the position-dependent decrease in duplex stability observed from the θ_{ext} curves and $\Delta\Delta G_{d37}^{\circ}$. Once the short segment reaches a

length of seven base pairs, the binding stability overcomes $\Delta G^{\ddagger}_{h2}$ and duplex destabilization levels off. $\Delta G^{\ddagger}_{h2}$ is fully encompassed into the duplex state at these AP-site positions, creating a local free-energy minimum along n_{bp} that corresponds to configurations with one segment dehybridized. The leveling off is also facilitated by a reduction in the magnitude of $\Delta G^{\ddagger}_{h2}$ from 14 to 9 kJ/mol as the AP site moves inward from the termini.

The HC calculations suggest that the position-dependent destabilization from an AP site depends only on the magnitude of the nucleation barrier and the binding stability of the short segment. An additional minor effect is that the number of base-pair arrangements possible over the whole duplex decreases as the AP site moves closer to the duplex, resulting in a reduction of base-pairing combinational entropy and a free-energy penalty of 1 – 2 kJ/mol (Fig. 7.A.1). Given the small magnitude of this effect, duplex destabilization is essentially only position-dependent when an AP site is close to the termini. This range is roughly ~1 – 9 base pairs from our model calculations but depends on the molecular properties of $\Delta G^{\ddagger}_{h2}$ and the short duplex segment.

7.2.3 Interplay of nucleobase sequence and AP-site position

Beyond the factors that influence a homogeneous duplex, destabilization by an AP site will depend on multiple aspects of the nucleobase sequence. Previous reports showed that $\Delta\Delta G^{o}_{h37}$ is highly sensitive to the identity of the nucleobase being removed as well as the bases adjacent to the AP site.[31-32] Further, nucleobase sequence will influence the binding stability of the short duplex segment as well as the magnitude of $\Delta G^{\ddagger}_{h,2}$. We use three additional 11-mers with distinct base-pairing properties to investigate sequence-dependent destabilization from an AP site (Fig.

7.3a). "CGCcap" contains three G:C base pairs on one end to create asymmetry in the duplex base-pairing stability, "CCends" places a pair of G:C base pairs at the termini to minimize base-pair fraying, and "GCGcore" has central G:C base pairs to promote fraying of the A:T termini.[41-42]

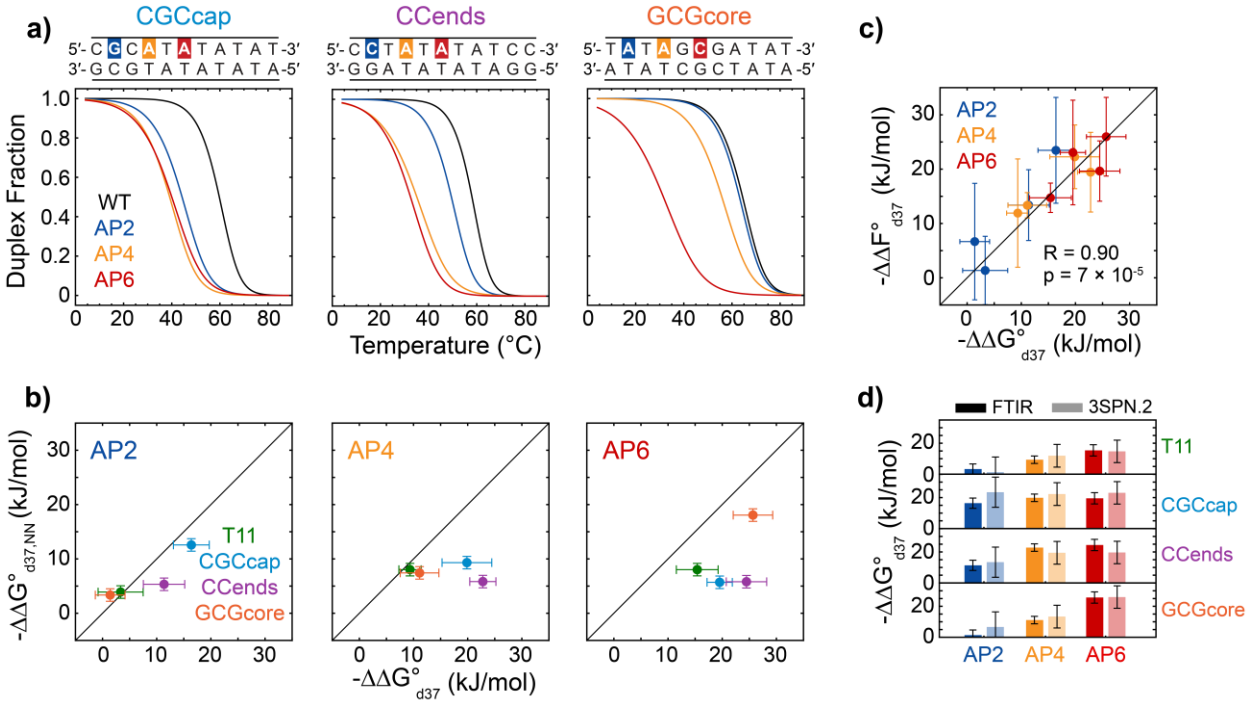

Figure 7.3 Sequence-dependent duplex destabilization from an AP site. (a) Duplex melting curves for CGCcap, CCends, and GCGcore sequences determined from a two-state analysis of FTIR temperature series. **(b)** Scatter plots of $\Delta\Delta G^{\circ}_{d37}$ from FTIR vs. nearest-neighbor (NN) model calculations of dehybridization free-energy change from an AP site ($\Delta\Delta G^{\circ}_{d37,NN}$) for all sequences. Plots are separated by AP-site position. The comparison indicates that the NN model systematically underestimates $\Delta\Delta G^{\circ}_{d37}$ and poorly predicts sequence-dependent destabilization from an AP site. **(c)** Scatter plot of $\Delta\Delta G^{\circ}_{d37}$ from FTIR vs. $\Delta\Delta F^{\circ}_{d37}$ from 3SPN.2-determined melting curves for all sequences with an AP site. The Pearson correlation coefficient (R) and p value is listed. **(d)** Comparison of $\Delta\Delta G^{\circ}_{d37}$ (dark bars) and $\Delta\Delta F^{\circ}_{d37}$ (light bars) for each sequence. Error bars for $\Delta\Delta G^{\circ}_{d37}$ and $\Delta\Delta F^{\circ}_{d37}$ are propagated from two-state fits, and NN error bars correspond to the reported standard deviation from comparison with experimental data.[7]

An AP site is generally more destabilizing when shifted inward from the termini, but Fig. 7.3a illustrates how the relative placement of G:C and A:T base pairs significantly tunes this position-dependent trend. At identical position, incorporating an AP site at a G:C base pair is more destabilizing than at an A:T base pair, which is most notable for AP2 sites as observed in the comparison of $\Delta\Delta G^{\circ}_{h37}$ values extracted for CGCcap-AP2/CCends-AP2 and T11-AP2/GCGcore-AP2 (Fig. 7.3b). However, the removal of an interior A:T base pair can be equivalent or even more destabilizing than removing a near-terminal G:C base pair as found in comparing CCends-AP2 and CCends-AP6. For CGCcap, the strength of nearest-neighbor base-pairing and stacking interactions around the AP site decreases from AP2 to AP4 to AP6 modifications. The destabilization from disrupting the G:C-rich region in CGCcap-AP2 almost matches the combination of destabilization from A:T disruption and incorporation of a nucleation penalty in CGCcap-AP4 and CGCcap-AP6, nearly cancelling of the position-dependence in $\Delta\Delta G^{\circ}_{h37}$.

The AP-site position and its local sequence context alone do not fully capture the sequence-dependent trends in $\Delta\Delta G^{\circ}_{d37}$. This point is illustrated in Fig. 7.3b by comparing $\Delta\Delta G^{\circ}_{d37}$ values from FTIR and those calculated with Santa Lucia's NN model ($\Delta\Delta G^{\circ}_{d37,NN}$).[7] $\Delta\Delta G^{\circ}_{d37,NN}$ values for AP4 and AP6 sequences are calculated by removing the two NN parameters associated with the AP site. For AP2 sequences, the 3′ NN parameter is removed and the 5′ parameter is halved to account for having just a single base-pair and stacking interaction at the terminal base. The NN model systematically underestimates the magnitude of $\Delta\Delta G^{\circ}_{d37}$ due to the lack of a free-energy penalty for base pairing on each side of the AP site. Further, NN effects do not predict the sequence-dependence of $\Delta\Delta G^{\circ}_{d37}$ at fixed AP-site position. For example, $\Delta\Delta G^{\circ}_{d37}$ spans 14 kJ/mol

for AP4 sequences yet $\Delta\Delta G^{\circ}_{d37,NN}$ values for these sequences fall within a ~3 kJ/mol window. In another example, CCends-AP6 and GCGcore-AP6 exhibit $\Delta\Delta G^{\circ}_{d37}$ values within 1 kJ/mol and their $\Delta\Delta G^{\circ}_{d37,NN}$ predictions are separated by 12 kJ/mol. These observations suggest the presence of additional sequence-dependent contributions to the free-energy penalty from an AP site.

7.3 Sequence-dependent nucleation barrier and stability of the weak segment

We next aim to evaluate how the properties of the segment-nucleation barrier and binding stability of the short segment depend on nucleobase sequence. The HC model qualitatively captures the effect of sequence on duplex destabilization (Appendix 7.A), but it neglects numerous molecular factors that are necessary for an accurate description of the underlying base-pairing free-energy landscape. Therefore, we next characterized the underlying hybridization free-energy landscape as a function of AP-site position using coarse-grained MD simulations. In particular, the threshold segment length for stable hybridization, its sequence-dependence, and the molecular behavior of few-base-pair segments were examined.

7.3.1 Nucleation barrier and stability of the weak base-pair segment

We conducted MD simulations with the 3SPN.2 coarse-grained model employing well-tempered metadynamics (WTMetaD) to sample the hybridization free-energy landscape at 7 – 8 temperatures from $T_{m,MD} - 20$ °C to $T_{m,MD} + 20$ °C, where $T_{m,MD}$ is the melting temperature determined with 3SPN.2. Incorporation of an AP site in 3SPN.2 has negligible impact on the B-DNA duplex structure, and the free nucleobase remains predominantly intrahelical,[33, 43] consistent with previous structural characterization of duplexes containing AP sites.[28] Two-state analysis of

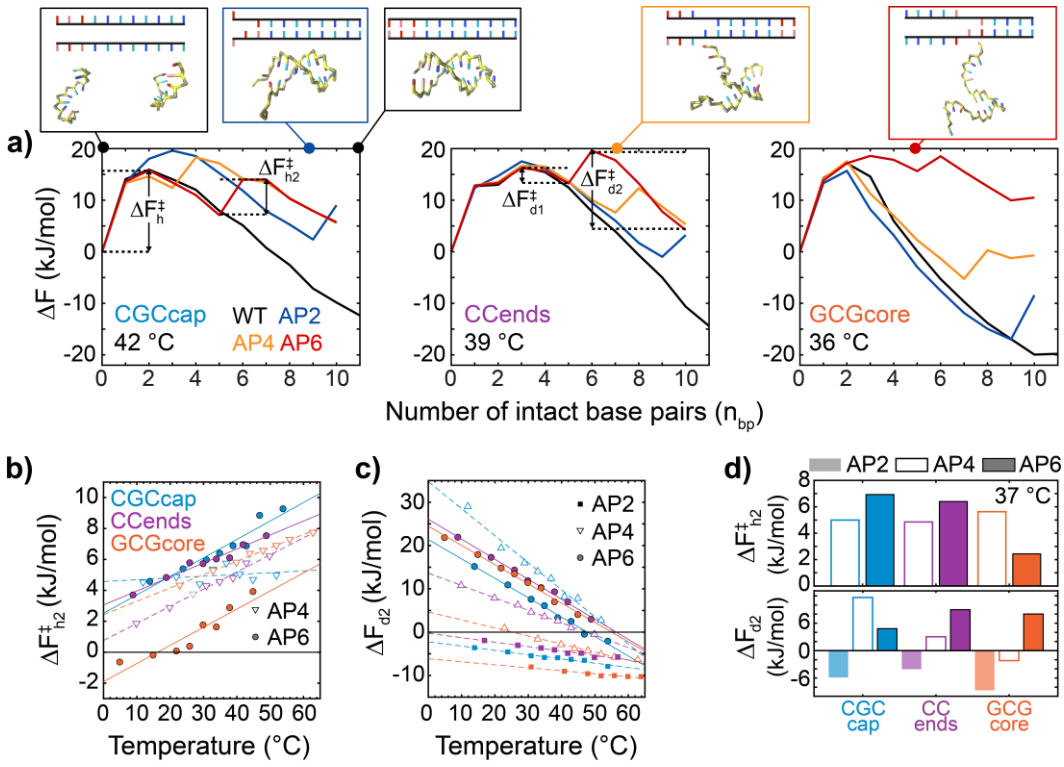

Figure 7.4 Sequence-dependent free-energy profiles for duplex hybridization. (a) Free-energy profiles as a function of the number of intact base pairs, $\Delta F(n_{bp}) = -RT \ln P(n_{bp})/P(0)$, for each sequence from 3SPN.2 MD simulations employing WTMetaD. Base pairs were assigned using a 0.7 nm radial separation cutoff. Simulations were carried out at temperatures 10-20 °C below the 3SPN.2-determind T_m ($T_{m,MD}$). AP2, AP4, and AP6 sequences show an additional hybridization barrier for $n_{bp} > 2$ for base pairing on both sides of the AP site ($\Delta F^{\ddagger}_{h2}$). **(b)** $\Delta F^{\ddagger}_{h2}$ for AP4 (triangles) and AP6 (circles) sequences as a function of temperature. **(c)** Free-energy difference between states with weak-segment frayed and intact, $\Delta F_{h2} = -RT \ln P(n > n_{sd})/P(n_{sd})$, for all sequences where n_{sd} is the number of base-pairs in the segment-dehybridized state. Dark-solid, dashed, and light-solid lines correspond to linear fits for AP6, AP4, and AP2 sequences, respectively. Extracted internal-energy and entropy changes are provided in Fig. 7.5. **(d)** (top) $\Delta F^{\ddagger}_{h2}$ and (bottom) ΔF_{h2} at 37 °C for each sequence. 3SPN.2 MD simulations suggest that two to five base pairs, depending on the G:C content, provide enough stability for a segment to hybridize.

3SPN.2-determined melting curves produces hybridization Helmholtz free energies (ΔF°_{h37} and $\Delta\Delta F^{\circ}_{h37}$) that are well-correlated (R = 0.90) with experimental values (Figs. 7.3cd), indicating that

3SPN.2 reasonably predicts the sequence-dependent free-energy penalty from an AP site. In contrast, ΔH_h° and ΔS_h° are poorly captured by 3SPN.2, particularly for AP sequences.[43]

Figure 7.4 shows FEPs computed from the probability distribution along n_{bp} for CGCcap, CCends, and GCGcore sequences that qualitatively agree with those from the HC model (Figs. 7.2c & 7.A.1). WT sequences show a single free-energy barrier to hybridization ($\Delta F_h^\ddagger$) peaked at n_{bp} = 2 or 3 and hybridization is energetically downhill for n_{bp} > 3. AP sequences must overcome a second barrier ($\Delta F_{h2}^\ddagger$) to form a fully intact duplex. Just as for $\Delta G_{h2}^\ddagger$ from the HC model, $\Delta F_{h2}^\ddagger$ arises from the energy penalty for nucleating the second base-pair segment and creates a local free-energy minimum that corresponds to duplex configurations containing intact base pairs on only one side of the AP site. $\Delta F_{h2}^\ddagger$ lies between n_{bp} = 9 and n_{bp} = 10 in AP2 sequences and is not observable along the n_{bp} coordinate, yet a barrier for forming the terminal base pair is observed as a function of average base-pair separation (r_{bp}).[43] Both $\Delta F_h^\ddagger$ and $\Delta F_{h2}^\ddagger$ arise from entropic penalties for nucleation and are partially balanced by favorable enthalpy changes for base pairing (Figs. 7.4b and 7.5a). $\Delta F_h^\ddagger$ is predominantly a reflection of the reduction in translational and orientational entropy upon bimolecular association, which are absent from $\Delta F_{h2}^\ddagger$. Instead, $\Delta F_{h2}^\ddagger$ is dominated by a reduction in conformational entropy of the unhybridized segment and is therefore 2-to-5-fold smaller than $\Delta F_h^\ddagger$ (Fig. 7.5d). $\Delta F_{h2}^\ddagger$ is also sensitive to the sequence of nucleating base pairs adjacent to the AP site (Fig. 7.4d). $\Delta F_{h2}^\ddagger$ is smallest for GCGcore-AP6, particularly at low temperatures, and involves formation of a G:C base pair adjacent to the AP site whereas $\Delta F_{h2}^\ddagger$

requires formation of one or two A:T base pairs on the second segment in the other sequences, leading to a larger barrier height.

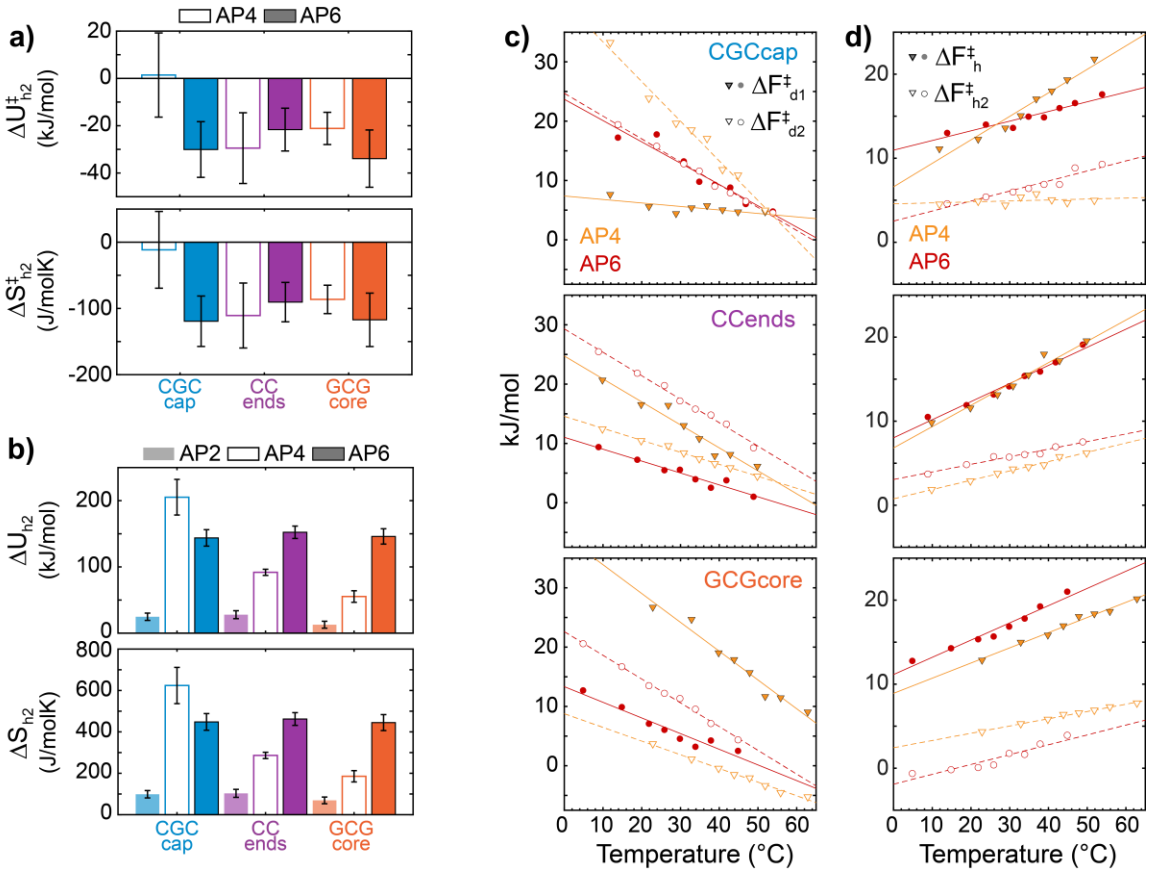

Figure 7.5 Temperature dependence of ΔF_{d2}, $\Delta F^{\ddagger}_{h2}$, $\Delta F^{\ddagger}_{d1}$, and $\Delta F^{\ddagger}_{d2}$ from FEPs. (a) Internal-energy ($\Delta U^{\ddagger}_{h2}$) and entropic ($\Delta S^{\ddagger}_{h2}$) barriers for nucleation of the second base-pair segment determined from linear fits of the temperature-dependence of $\Delta F^{\ddagger}_{h2}$ (Fig. 7.4b), $\Delta F^{\ddagger}_{h2} = \Delta U^{\ddagger}_{h2} - T\Delta S^{\ddagger}_{h2}$. **(b)** Internal-energy (ΔU_{d2}) and entropy (ΔS_{d2}) change for dehybridization of the second base-pair segment determined from linear fits of the temperature dependence of ΔF_{d2} (Fig. 7.4c). Error bars correspond to 95% confidence intervals from fits. **(c)** Temperature-dependence of $\Delta F^{\ddagger}_{d1}$ (closed symbols and solid lines) and $\Delta F^{\ddagger}_{d2}$ (open symbols and dashes lines) for AP4 and AP6 sequences**. (d)** Temperature-dependence of $\Delta F^{\ddagger}_{h}$ (closed symbols and solid lines) and $\Delta F^{\ddagger}_{h2}$ (open symbols and dashed lines) for AP4 and AP6 sequences**.**

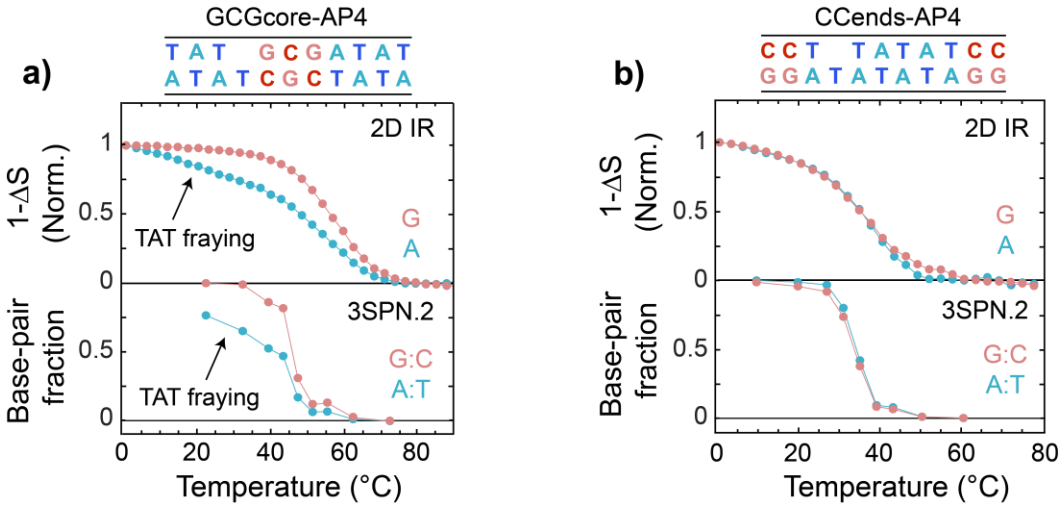

Figure 7.6 Sequence-dependent pre-melting of three-base-pair segments. (a) (top) Normalized temperature-dependent change in two-dimensional infrared (2D IR) signals relative to the initial temperature, $1 - \Delta S(T) = [S(T) - S(1℃)]/S(1℃)$, at guanine (G) and adenine (A) ring vibrational bands for GCGcore-AP4. (bottom) Fraction of intact G:C and A:T base pairs from 3SPN.2 MD simulations carried out at 8 – 9 temperatures across the duplex melting transition. Base pairs were assigned using a radial cutoff of 0.7 nm. **(b)** Same plot for CCends-AP4. Fraying of the TAT segment is experimentally observed in GCGcore-AP4 but not for the CCT segment in CCends-AP4, consistent with the sequence dependence of weak-segment stability predicted by the simulations.

The AP-site position and nucleobase sequence determine the position and magnitude of the $\Delta F_{h2}^{\ddagger}$ barrier along n_{bp} and the base-pairing stability of the second segment. The combination of these factors determine the free energy for dehybridization of the weakly bound segment (ΔF_{d2}). The terminal base pair adjacent to the AP-site in AP2 sequences has an insufficiently stabilizing potential energy for base pairing to overcome the unfavorable entropy, leading to a positive free-energy change for forming the last base pair ($\Delta F_{d2} < 0$, Fig. 7.4c). The threshold length of the weak segment to overcome the nucleation penalty and reach $\Delta F_{h2} < 0$ depends on the sequence of the segment. If the weak segment contains multiple G:C base pairs like in CCends-AP4, then three base pairs are sufficient to form a stable segment. In contrast, the pure A:T segment of

GCGcore-AP4 is largely frayed at 37 °C and four to six base pairs are needed to form a stable segment as evidenced by CGCcap-AP4 and CGCcap-AP6. These sequence-dependent length thresholds for forming a stably-bound segment are also observed experimentally from FTIR and 2D IR temperature series that separately track changes in adenine and guanine ring modes that report on changes in A:T and G:C base pairing, respectively (Fig. 7.6). GCGcore-AP4 exhibits fraying of the short segment at temperatures below full dissociation of the duplex whereas such fraying is negligible for CCends-AP4. The resulting experimental melting profiles qualitatively agree with the temperature-dependent fraction of intact G:C and A:T base pairs from simulation. Further analysis of the total spectral change from FTIR and 2D IR melting experiments suggests that the short segments of GCGcore-AP2 and GCGcore-AP4 have a significant frayed population even below 4 °C.[43] This sequence-dependent length scale for forming a stably bound segment contributes to the mismatch between $\Delta\Delta G^{\circ}_{d37}$ and $\Delta\Delta G^{\circ}_{d37,NN}$ observed in Fig. 7.3b. For instance, the spread in $\Delta\Delta G^{\circ}_{d37}$ for AP4 sequences may primarily occur because the A:T-rich short segments of T11-AP4 and GCGcore-AP4 are more weakly bound than the short segments of CGCcap-AP4 and CCends-AP4.

7.3.2 Kinetics of segment-dehybridization

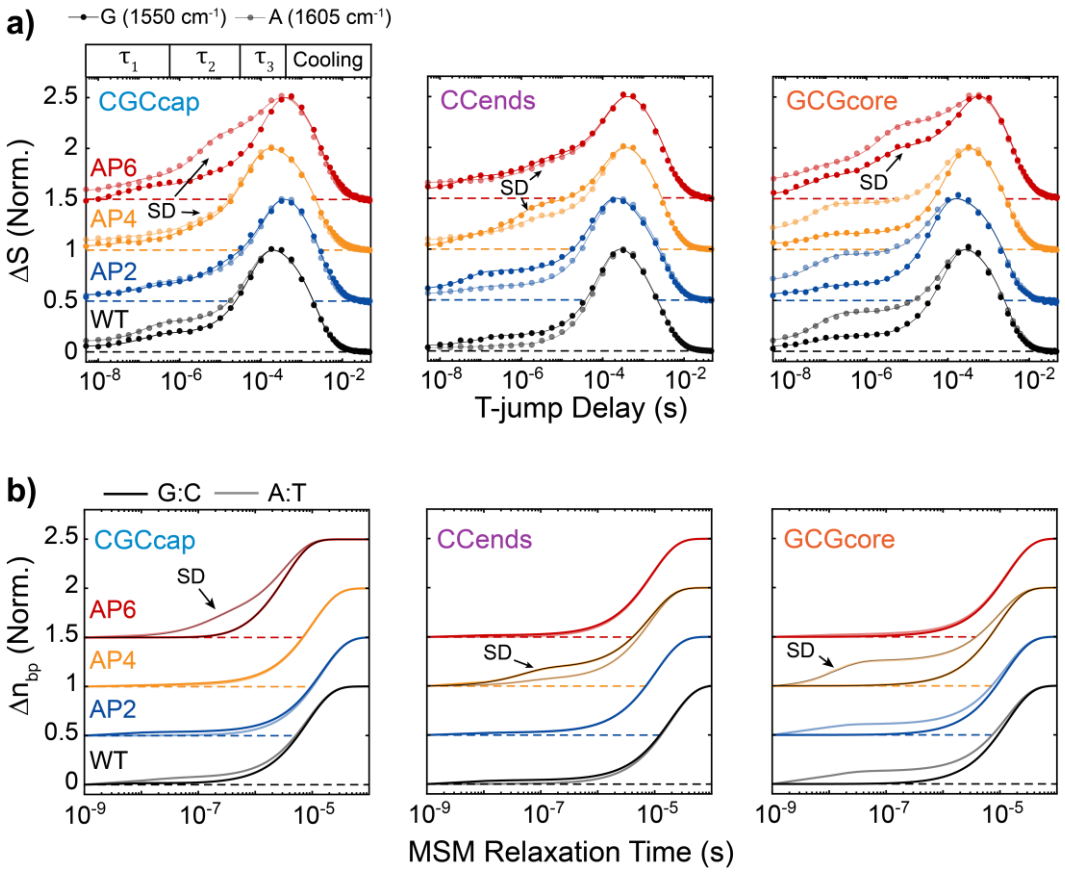

Figure 7.7 T-jump relaxation from experiment and simulation. (a) T-jump IR (t-HDVE) time traces plotted for CGCcap, CCends, and GCGcore sequences as the normalized signal change relative to the maximum of the initial-temperature spectrum, $\Delta S(t) = [S(t) - S(T_i)]/\max[S(T_i)]$. ΔS time traces probed at 1550 cm^{-1} (dark) report on changes in G:C base pairing and those at 1605 cm^{-1} (light) report on changes in A:T base pairing. T-jumps are performed from approximately T_m–15°C to T_m for each sequence. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Solid lines correspond to three- or four-component fits from global lifetime analysis (Appendix 6.C). A microsecond timescale segment-dehybridization response (SD) is observed for CCends-AP4 and CCends-AP6. **(b)** Markov state model (MSM) T-jump simulations of the normalized change in intact G:C (dark) and A:T (light) base pairs, $\Delta n_{bp}(t) = [n_{bp}(t) - n_{bp}(T_i)]/[n_{bp}(T_f) - n_{bp}(T_i)]$, from $T_{m,MD}$–15°C to $T_{m,MD}$. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Black/gray solid lines for CGCcap-AP6, CCends-AP4, and GCGcore-AP4 correspond to fits to a sum of two exponential components.

Only some sequences exhibit a large enough population of configurations with fraying of the weak segment to be observed in IR melting experiments, and the frayed configurations in other sequences are expected to be metastable and located between free-energy barriers for hybridization on each side of the AP site. To test for the presence of metastable partially-hybridized configurations and a second barrier to hybridization, we turned to T-jump IR spectroscopy, which can be used to directly measure the kinetics of short-segment dehybridization.[33] In these experiments, the sample is equilibrated at a temperature below T_m and optically heated by ~15 °C ($\Delta T = T_f - T_i$) within 7 ns. The ensuing changes in base pairing of the duplex are monitored with heterodyned dispersed vibrational echo spectroscopy (HDVE)[44] from nanosecond-to-millisecond time delays. HDVE spectroscopy reports on changes in ring and carbonyl vibrational bands similarly to an IR pump-probe spectrum, and we report the spectra as the difference in signal at a given time delay after the T-jump relative to the maximum of the initial-temperature spectrum, $\Delta S(t) = [S(t) - S(T_i)]/\max[S(T_i)]$. We use the change in excited-state absorptions of the guanine ring mode at 1550 cm^{-1} and of the adenine ring mode at 1605 cm^{-1} as reporters of G:C and A:T base pairing, respectively, as in previous studies.[33, 41-42]

T-jump measurements reveal distinct kinetic behavior between WT and AP sequences (Fig. 7.7a). WT sequences show two dehybridization kinetic components on ~100 ns (τ_1) and ~100 μs (τ_3) timescales followed by thermal relaxation back to T_i at ~1 ms. The T-jump IR data are well described by global lifetime fitting with exponentially damped spectral components (Appendix 6.C), and the same number of components and similar kinetics are obtained from inverse-Laplace-transform rate distribution spectra.[43] The τ_1 and τ_3 processes were previously assigned to T-jump

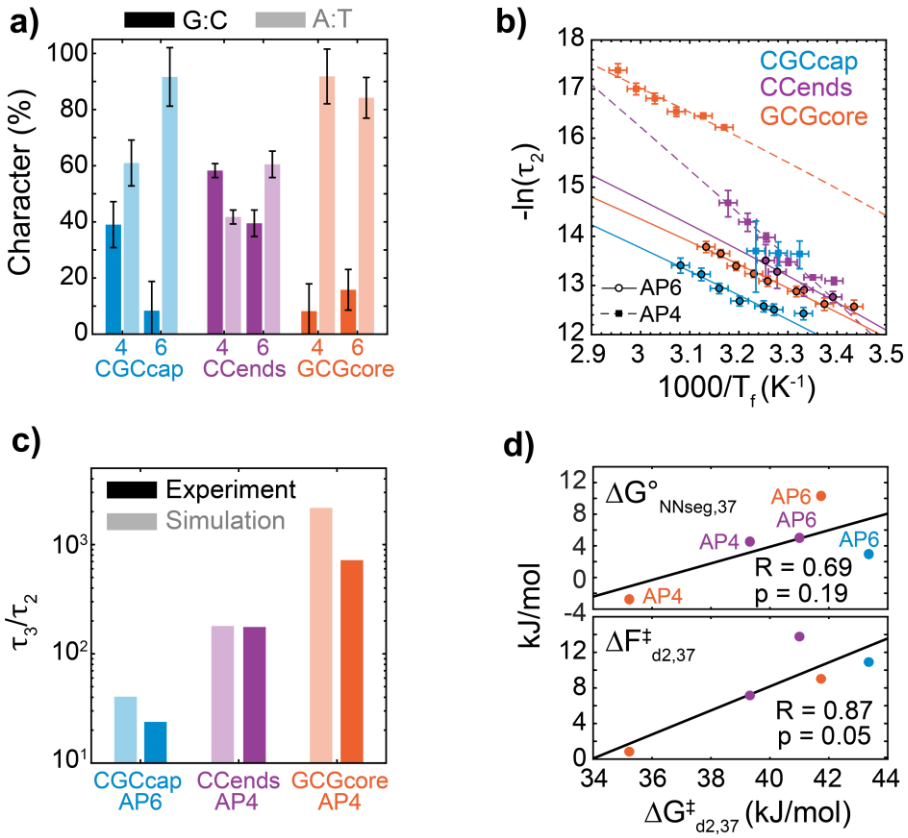

Figure 7.8 Comparison of segment-dehybridization kinetics from experiment and simulation. (a) Percentage of G:C (dark, 1550 cm^{-1}) and A:T (light, 1605 cm^{-1}) base-pair loss character observed in segment-dehybridization t-HDVE responses. Percentages are derived from amplitudes determined from global lifetime fitting (Appendix 6.D). Error bars indicate 95% confidence intervals propagated from global fits. **(b)** Temperature-dependent observed rate constants for segment-dehybridization ($1/\tau_2$) determined from T-jump IR. Data are fit to a Kramers-like equation in the high-friction limit (eq. 7.B3). The rate trend for CGCcap-AP4 is not fit due to insufficient data. Vertical error bars correspond to 95% confidence intervals from global lifetime fitting and horizontal error bars are the standard deviation in T-jump magnitude. **(c)** Ratio of full-strand dissociation time constant (τ_3) to τ_2 from T-jump IR (dark) and MSM T-jump simulations (light) indicating qualitative agreement between simulation and experiment. **(d)** Scatter plots of the segment-dehybridization free-energy barrier determined from T-jump IR ($\Delta G^{\ddagger}_{d2}$) vs. (top) the dehybridization free energy for the respective segment calculated from the NN model (ΔG°_{NNseg})[7] and (bottom) $\Delta F^{\ddagger}_{d2}$ determined from 3SPN.2 MD simulations at 37 °C. Linear fits (solid black line) are shown for each plot with the respective Pearson correlation coefficient (R) and p value. $\Delta G^{\ddagger}_{d2}$ shows better correlation with $\Delta F^{\ddagger}_{d2}$ than ΔG°_{NNseg}.

induced terminal fraying of the duplex and full-strand dehybridization, respectively,[33] consistent with reports for similar sequences.[41] The temperature-dependent trends in timescale and amplitude of τ_3 support its assignment to complete strand dissociation (Appendix 7.B). Many of the AP4 and AP6 sequences show an additional kinetic component labeled τ_2 between 500 ns to 10 μs with sequence-dependent amplitude and timescale. The τ_2 response corresponds to dehybridization of a complete stretch of base pairs on one side of the AP site (segment dehybridization) and reveals the presence of a free-energy barrier from the fully intact duplex to segment-dehybridized configurations ($\Delta G_{d2}^{\ddagger}$). The observation of $\Delta G_{d2}^{\ddagger}$ verifies the presence of $\Delta G_{h2}^{\ddagger}$ in the reverse direction, as illustrated from the FEPs in Fig. 7.4a and previously reported for the AP6 sequences.[33]

The spectral properties and kinetics of the τ_2 process support its assignment to segment-dehybridization and provide insight into the properties of $\Delta G_{h2}^{\ddagger}$. The fraction of A:T and G:C base-pair loss character during τ_2 determined from global lifetime fitting (Fig. 7.8a) match the expected signal change for dehybridizing either half of CCends-AP6 and GCGcore-AP6, the 5′-TATAT-3′ segment in CGCcap-AP6, and the 5′-CCT-3′ segment in CCends-AP4. Additionally, the segment dehybridization rate τ_2^{-1} increases exponentially with temperature as expected for an active process, illustrated by the Arrhenius plot in Fig. 7.7b, which we analyze using a Kramers-like equation (eq. 7.B3). The τ_2 response of CGCcap-AP4 is more difficult to interpret. It has low amplitude and mixed G:C and A:T character even though each segment is purely composed of G:C or A:T base pairs. Due to the similar binding stability of 5′-CGC-3′ and 5′-TATATAT-3′ segments, it is possible that τ_2 contains similar amplitude from both segment-hybridization pathways. Lastly, we observe that GCGcore-AP4 does not have a τ_2 response; however, unlike all

other sequences, τ_1^{-1} increases exponentially with temperature in an active manner. We therefore assign τ_1 in GCGcore-AP4 to dehybridization of the 5′-TAT-3′ segment and note that it may overlap in time with terminal fraying dynamics from the other end of the duplex.

We find large variation in segment-dehybridization kinetics across sequence and segment length from T-jump IR measurements. To more directly connect these measurements to simulation, we built kinetic models of T-jump relaxation using Markov state models (MSMs) constructed from 250 μs of unbiased 3SPN.2 simulations near $T_{m,MD}$ for each sequence as previously reported for WT and AP6 sequences in Chapter 6.[33, 43] The change in the fraction of intact A:T and G:C base pairs (Δn_{bp}) is computed as a function of relaxation time at $T_{m,MD}$ after initiating the MSM kinetics with a population distribution obtained at $T_{m,MD}$–15°C to replicate the initial experimental T-jump ensemble. Relaxation traces in Figs. 7.7b show two to three kinetic components that correspond to terminal fraying (1 – 30 ns, τ_1), segment-dehybridization (50 – 500 ns, τ_2), and complete strand dissociation (1 – 50 μs, τ_3). The MSM relaxation timescales are accelerated by ~1 order of magnitude relative to T-jump IR measurements, as previously reported for the 3SPN.2 model,[45] and therefore we use the ratio between τ_3 and τ_2 (or τ_1 for GCGcore-AP4) to compare simulation and experimental timescales. Experimental and simulated τ_3/τ_2 values show qualitative agreement and an increase from CGCcap-AP6 to CCends-AP4 to GCGcore-AP4, suggesting that 3SPN.2 accurately captures sequence-dependent variation in $\Delta G_{d2}^{\ddagger}$ (Fig. 7.8c). We also find correlation (R = 0.87) between the $\Delta G_{d2}^{\ddagger}$ and $\Delta F_{d2}^{\ddagger}$ from the FEPs in Fig. 7.4a at 37 °C (Fig. 7.8d). The correlation between $\Delta G_{d2}^{\ddagger}$ and the calculated NN dehybridization free energy of the respective unbinding segment (ΔG_{NNseg}°) is substantially poorer (R = 0.69),

suggesting that both the dehybridization energy of the segment (ΔG_{d2}) and its nucleation barrier ($\Delta G_{h2}^{\ddagger}$) significantly contribute to the sequence-dependence of $\Delta G_{d2}^{\ddagger}$.

MSM relaxation profiles generally exhibit a lower segment-dehybridization response amplitude than in experiment, and negligible response is observed for CGCcap-AP4, CCends-AP6, and GCGcore-AP6 even though each sequence possesses a metastable segment-dehybridized state (Figs. 7.4a and 7.7a). For each of these three sequences, we observe $\Delta F_{d2}^{\ddagger}$ to be larger than the free-energy barrier for dehybridization of the final base-pair segment ($\Delta F_{d1}^{\ddagger}$) at $T_{m,MD}$ (Fig. 7.5). When $\Delta F_{d2}^{\ddagger} > \Delta F_{d1}^{\ddagger}$, the waiting time to surpass $\Delta F_{d2}^{\ddagger}$ is the rate-limiting step for dehybridization, and we do not expect to observe a segment-dehybridization response in the T-jump kinetics. In contrast, we observe $\Delta F_{d2}^{\ddagger} \leq \Delta F_{d1}^{\ddagger}$ for sequences where there is a large amplitude segment-dehybridizaiton response (CCends-AP4, CGCcap-AP6, GCGcore-AP4). Segment-dehybridization is only observed experimentally for CGCcap-AP4 and CCends-AP6 at the few lowest measured temperatures, which may be explained by $\Delta F_{d2}^{\ddagger}$ increasing relative to $\Delta F_{d1}^{\ddagger}$ and surpassing it at higher temperatures (Fig. 7.5). The overall lower amplitude of segment-dehybridization response in simulations therefore suggests that the ratio of $\Delta F_{d2}^{\ddagger}$ to $\Delta F_{d1}^{\ddagger}$ is generally overestimated, and as $\Delta F_{d2}^{\ddagger}$ is fairly correlated with $\Delta G_{d2}^{\ddagger}$, we suggest that sequence-dependent error in this ratio primarily arises from $\Delta F_{d1}^{\ddagger}$. Error in $\Delta F_{d1}^{\ddagger}$ may stem from inaccurate modeling of base-pair and stacking interactions around the AP site for which 3SPN.2 is not parametrized. Further, 3SPN.2 simulations tend to poorly predict sequence-dependent variation in the dehybridization barrier among canonical oligonucleotides.[45]

7.4 Contributions from non-canonical base pairing

7.4.1 Out-of-register base pairing at near-terminal AP sites

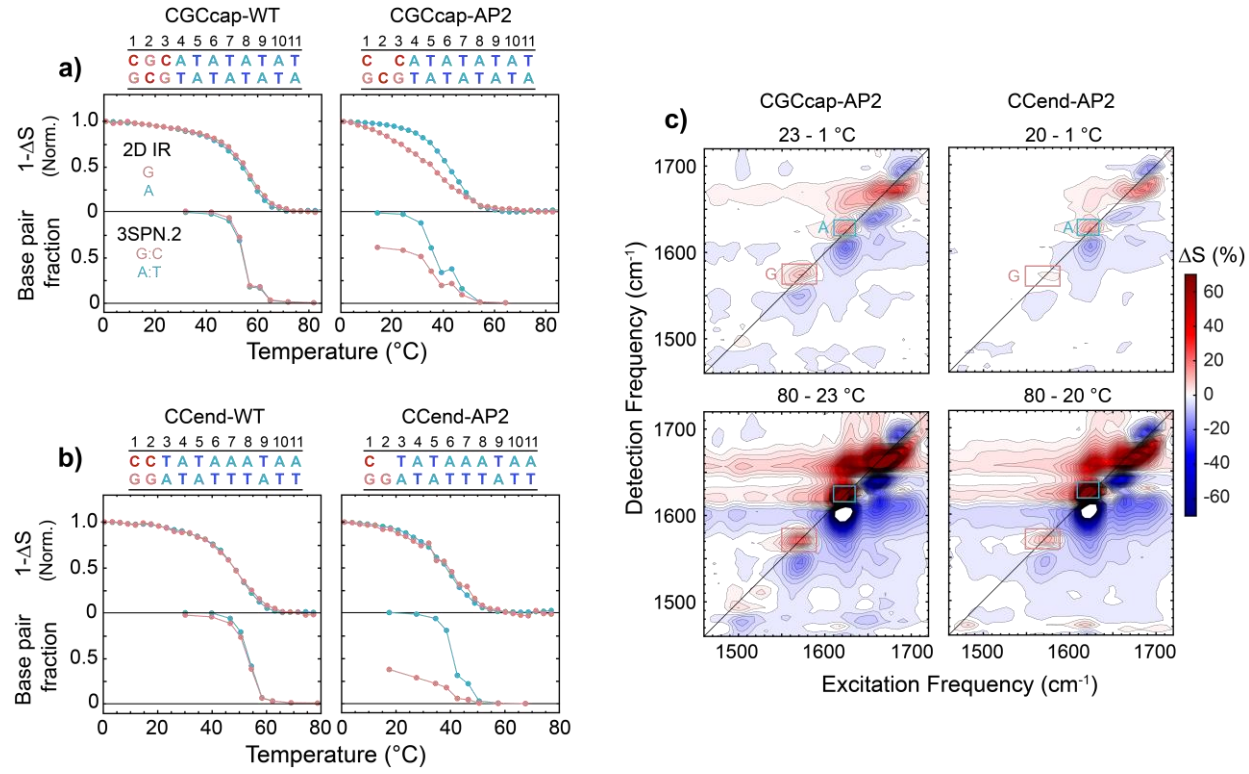

Figure 7.9 Partial dehybridization of CGCcap-AP2 observed with 2D IR. Normalized temperature-dependent change in two-dimensional infrared (2D IR) signals relative to the initial temperature, $1 - \Delta S(T) = [S(T) - S(1℃)]/S(1℃)$, at guanine (G) and adenine (A) ring vibrational bands for **(a)** CGCcap-AP2 and **(b)** CCend-AP2. Bottom sections of each panel show the fraction of intact G:C and A:T base pairs from 3SPN.2 MD simulations carried out at 8 – 9 temperatures across the duplex melting transition. Base pairs were assigned using a radial cutoff of 0.7 nm. **(c)** 2D IR difference spectra of (top) low-temperature (1 to ~23-30 °C) and (bottom) high-temperature (~23-30 to 80 °C) changes. Spectra are plotted as the difference relative the maximum of the 1 °C, $\Delta S(T) = [S(T_2) - S(T_1)]/S(1℃)$. Contours are plotted with uniform 2.1% spacing. A and G ring mode regions are indicated for each sequence. Low-temperature fraying of the short segment is observed in experiment and simulation for CGCcap-AP2, yet not observed experimentally for CCend-AP2.

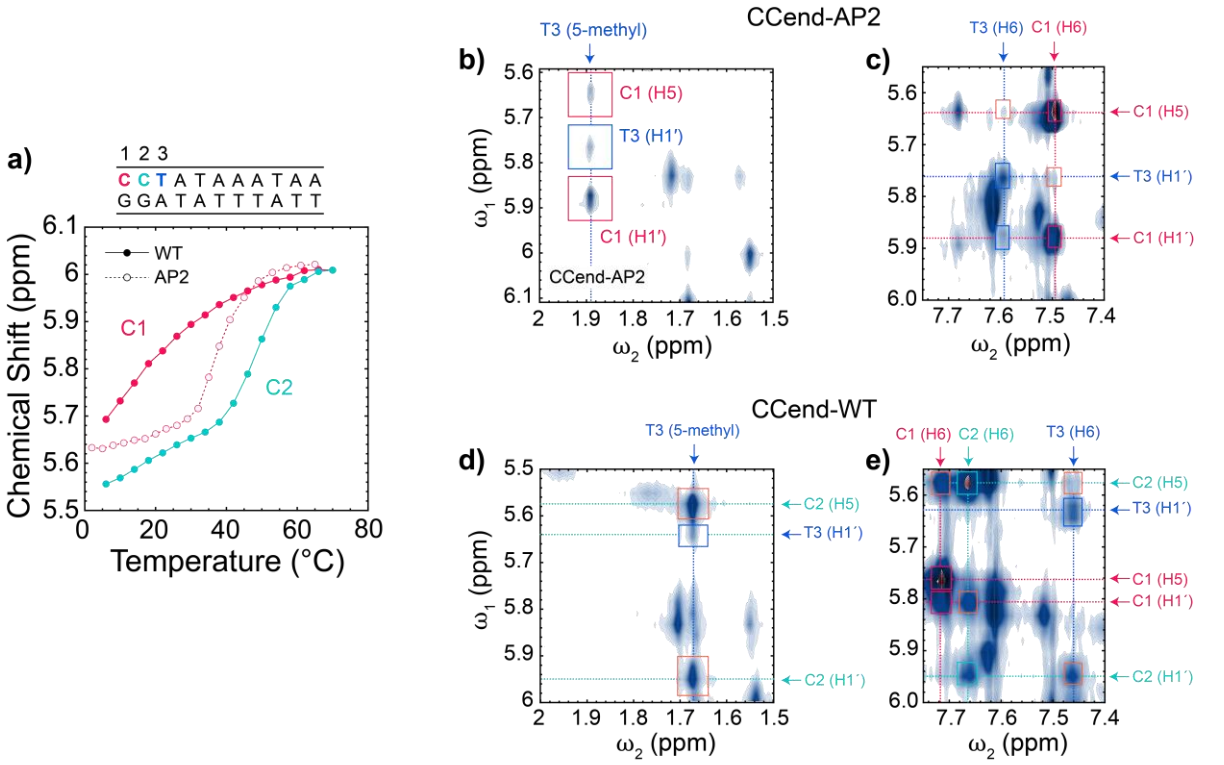

Figure 7.10 Out-of-register base pairing in CCend-AP2. (a) NMR chemical shifts of cytosine H5 nuclei for CCend-WT (filled circles, solid lines) and CCend-AP2 (open circles, dashed lines) determined as a function of temperature from TOCSY measurements. The terminal cytosine (C1) shows low-temperature fraying in CCend-WT whereas the second cytosine (C2) signal primarily changes during duplex dissociation. C1 does not exhibit fraying CCend-AP2 and instead tracks duplex dissociation. **(b)** NOESY cross peaks between H5/H1′ nuclei and 5-methyl protons of thymine for CCend-AP2. Cross-peaks are observed between C1 and T3 (red boxes). **(c)** NOESY and TOCSY cross peaks between H6 and H5/H1′ (5.55 – 6.0 ppm). Cross-peaks **(d)** between H5/H1′ nuclei and 5-methyl protons and **(e)** between H6 and H5/H1′ are shown for CCend-WT. Orange boxes denote interbase cross-peaks while blue, red, cyan boxes correspond to intrabase thymine, C1, and C2 cross-peaks, respectively. Numerous cross-peaks between C1 and T3 in CCend-AP2 indicate out-of-register base pairing of C1 with G2. In CCend-WT, cross-peaks are instead observed between C2 and T3.

The simulations presented in Fig. 4 suggest that the terminal base pair next to an AP2 site is highly frayed at physiological temperatures regardless of sequence, but our experiments indicate that the terminal base pair can stably bind in certain sequence contexts. As introduced in Fig. 7.6,

we monitor the temperature-dependence of G:C and A:T features in the IR spectrum to check for the presence of segment-dehybridization at temperatures well below T_m in CGCcap-AP2 and CCend-AP2, a modified form of CCends with only a single CC terminus. Relative to FTIR, 2D IR temperature series are found to provide greater contrast between spectral changes arising from base-pair disruption and those giving rise to melting curve baselines.[43] The A:T and G:C temperature profiles of WT sequences are essentially identical and suggest that duplex melting is nearly all-or-nothing (Fig. 7.9ab). CGCcap-AP2 shows a clear deviation from all-or-nothing behavior where G:C base-pair character is lost at temperatures 20-30 °C below A:T melting, likely corresponding to disruption of the terminal G:C base pair as predicted from 3SPN.2 MD simulations (Fig. 7.9a). CCend-AP2, a modified version of CCends with only a single CC terminus, exhibits concerted changes of G:C and A:T signatures and no indication of melting below 1 °C,[43] suggesting that the terminal G:C base pair is stabilized relative to CGCcap-AP2.

Comparison of partial-dehybridization signatures across our sequences corroborates thresholds for stable segment hybridization from 3SPN.2 simulations. The identical G:C and A:T melting behavior of CCends-AP4 indicates that the three base-pair segment, 5´-CCT-3´, remains hybridized with the rest of the duplex.[43] In contrast, the 5´-TAT-3´ segment of GCGcore-AP4 unbinds at temperatures well below the rest of the duplex. These observations support the sequence-dependent trends in ΔF_{h2} from 3SPN.2 MD simulations and confirm that the length threshold for stable binding ($\Delta G_{h2} < 0$) of pure A:T segments is in the range of 4 – 7 base pairs at physiological temperatures while that for G:C segments is as low as 2 – 3 base pairs. The nucleation free-energy penalty for the second duplex segment is significantly lower than for the

initial duplex nucleation, therefore it is important to note that these values underestimate thresholds for stable hybridization of free duplexes and are most applicable for forming segments that are covalently linked to another bound segment.

The lack of partial dehybridization in CCend-AP2 suggests its terminal base pair is stabilized relative to the other sequences, and we used ^{1}H NMR spectroscopy to further examine the dominant duplex configurations of CCend and CGCcap sequences. Melting profiles for individual G:C base pairs were extracted from the temperature-dependent chemical shift of cytosine H5 nuclei isolated using ^{1}H-^{1}H TOCSY temperature series (Fig. 7.10a).[43] Interior cytosines (C2, C3) of CCend-WT and CGCcap-WT show sigmoidal transitions over the temperature range of the duplex melting transition measured with IR spectroscopy, but most of the frequency change in the terminal C1 H5 nuclei occurs at temperatures well below the melting transition. We assign the low-temperature behavior at C1 to fraying of the terminal base pair, which is further supported by the lack of such changes in GCGcore-WT where all cytosines are at the center of the duplex.[43] The temperature-dependence of the C1 H5 frequency in CCend-AP2 follows the duplex melting transition, suggesting that the terminal G:C base pair is intact and dissociates together with the rest of the duplex as indicated by 2D IR temperature series. The ^{1}H-^{1}H NOESY spectrum reveals intense cross peaks between C1 and T3 nuclei, indicating that C1 is stacked with T3 in the duplex state and presumably base paired with G2 rather than G1 (Figs. 7.10bc). This out-of-register base pairing creates a duplex with one stretch of base pairs, circumventing the nucleation penalty for forming a new segment. Similar out-of-register configurations have been proposed previously around AP sites.[28, 46] In contrast, the temperature-dependent cytosine H5 frequencies in CGCcap-AP2 indicate partial dehybridization of the G:C

cap more than 10 °C below the duplex melting transition. The TOCSY spectra of CGCcap-AP2 also show additional cytosine peaks from 2 to 26 °C that suggest multiple slowly interconverting duplex configurations may be present at low temperature,[43] yet the assignment of these structures are still unclear.

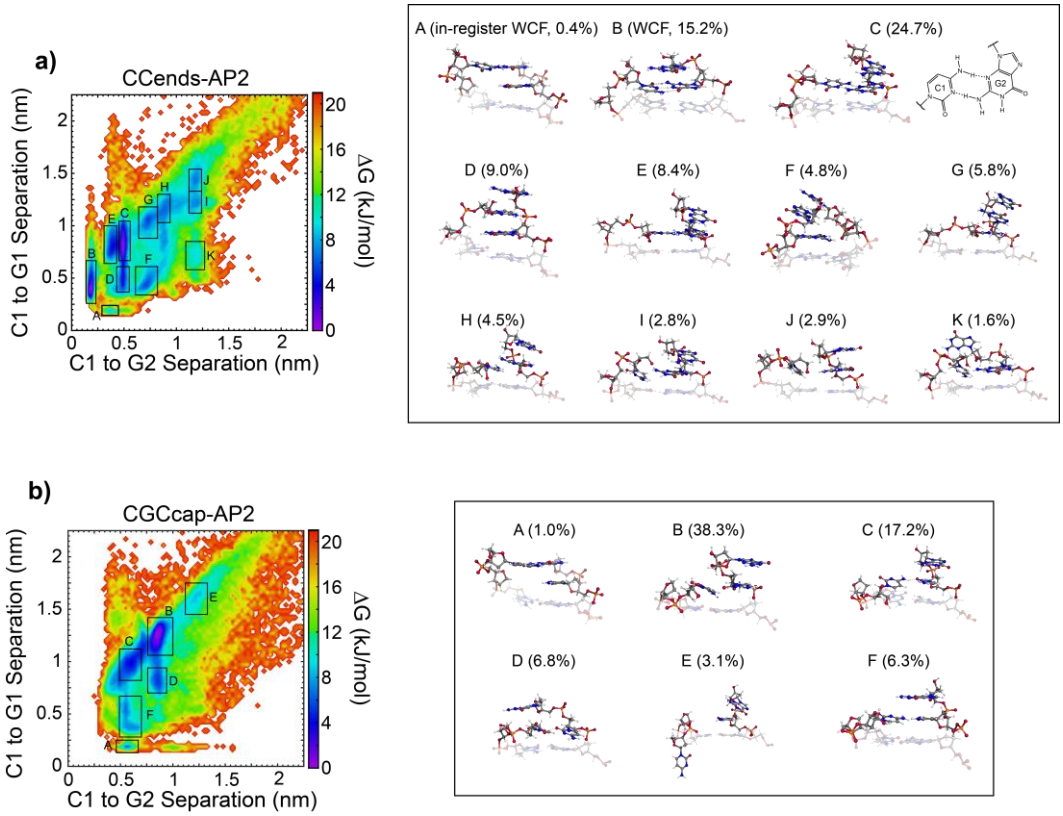

Figure 7.11 Terminal base-pair configurations from all-atom MD simulations. (a) Free-energy surface from 20 × 1 μs all-atom MD simulations of CCend-AP2 using the AMBER-bsc1 force field at 30 ˚C. The x-axis is the average separation between hydrogen-bonding atoms between C1 and G2 (out-of-register), and the y-axis is the average separation between those atoms in C1 and G1 (in-register). Snapshots of base pairs 1-3 from each free-energy minima are shown, and the population from a given minimum is determined from integrating over the marked rectangles. In agreement with experiment, C1 primarily base pairs out-of-register yet contains multiple free-energy minima with distinct base-pairing and stacking configurations. **(b)** Similar plot for CGCcap-AP2 indicating a rugged energy landscape with numerous non-canonical configurations.

All-atom MD simulations provide insight into the possible base-pairing and stacking configurations adopted by CCend-AP2 and CGCcap-AP2. Free-energy surfaces (FESs) constructed from 20 × 1 μs simulations indicate that each sequence adopts a variety of non-canonical configurations that interconvert on sub-microsecond timescales (Fig. 7.11). Interpretation of the simulations results must be done with care since the AMBER-bsc1 force is not parameterized for AP sites. While we cannot anticipate accurate predictions of the relative stabilities of the various states, the predictions are nonetheless useful for identifying putative configurational states and their relative propensities to help interpret experimental measurements. Only 1% of CGCcap-AP2 configurations exhibit in-register base pairing between C1 and G1, and C1 instead prefers to stack with C3 and interact with C2. Similar for CCend-AP2, configurations with in-register base pairing between C1 and G1 only make up 0.4% of the population (Fig. 7.11a). The most populated states are those with C1 paired to G2 and the free deoxyribose group fully extrahelical (states B and C). State B (15.2%) is characterized by a Watson-Crick-Franklin (WCF) base pair between C1 and G2 with stacking of G1 over the base pair as often observed for duplexes with dangling ends.[47] The C1 nucleotide rotates approximately 180° in state C (24.7%) and forms a base pair composed of two amino-imino hydrogen bonds, N3(C) to N2H(G) and N4H(C) to N3(G), and G1 stacks only with G2. This type of base pair has previously been observed for mononucleotides binding to RNA templates and in other folded RNA structures.[48-49] Another common configuration involves C1 stacked between bases on the other strand (state D). State B is most consistent with the NOESY spectrum for CCend-AP2 because it is the only configuration where the H5 nuclei of C1 is close enough to the T3 methyl group (2.7 Å vs. 8.8 Å in state C and 8-10 Å in state D) to enable significant dipolar coupling to produce an intense cross-peak. While

the NOESY data suggests a configuration like state B has the largest population experimentally, we cannot rule out the presence of states C and D as they interconvert with state B too quickly to be resolved as separate sets of NMR peaks.

7.4.2 Out-of-register duplex configurations in repetitive sequences

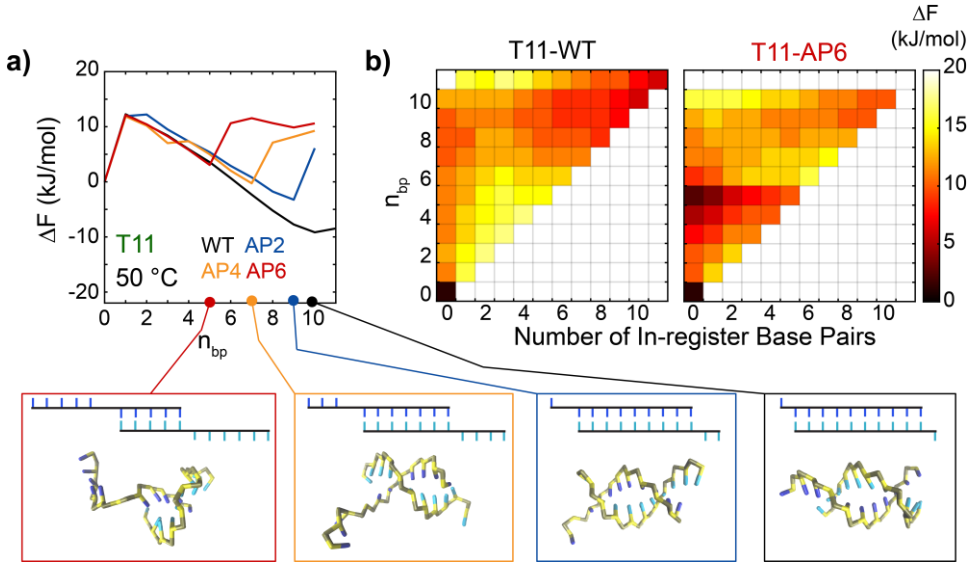

Figure 7.12 AP-site induced out-of-register base pairing in T11. (a) FEPs as a function of n_{bp} for T11 sequences at 50 °C from 3SPN.2 MD simulations with WTMetaD. Representative structures show out-of-register shifted configurations at duplex free-energy minima. **(b)** 2D free-energy surfaces for (left) T11-WT and (right) T11-AP6 of the number of in-register base pairs vs. n_{bp}. The free-energy minimum at n_{bp} = 5 for T11-AP6 corresponds to a completely out-of-register configuration.

The stability of the duplex for the repetitive T11 sequence can be affected by out-of-register base pairing, and 3SPN.2 MD simulations confirm the role of these configurations. Figure 7 shows duplex free-energy minima at n_{bp} = 10 for T11-WT, n_{bp} = 9 for T11-AP2, n_{bp} = 7 for T11-AP4, and n_{bp} = 5 for T11-AP6 corresponding to duplex configurations shifted by one, two, four, and six base-pair indices, respectively. The shifted configurations for T11-AP2 and T11-AP4 are the

dominant duplex species even at $T_{m,MD} - 20$ °C, but the population of those for T11-WT and T11-AP6 grow with increasing temperature until the duplex fully dissociates (Figs. 7.12a). Relative to unbound segments of in-register segment-dehybridized configurations found in CGCcap, CCends, and GCGcore sequences, the single-strand overhangs of out-of-register configurations have greater conformational entropy and a lower electrostatic penalty that lead to enhanced stability of the duplex region.[50]

Out-of-register base pairing is commonly observed in coarse-grained MD simulations and statistical modeling of repetitive DNA oligonucleotides,[36, 45, 51-53] yet it is challenging to verify the population of such configurations experimentally. The IR spectral features of in- and out-of-register A:T base pairing are expected to be essentially identical, therefore our experimental data does not directly report on the presence of out-of-register base pairing. Additionally, the experimental temperature-dependent melting behavior of T11 sequences does not indicate a significant population of out-of-register configurations and suggest that 3SPN.2 simulations overestimate the thermodynamic stability of out-of-register configurations for T11 AP sequences. The probability of out-of-register hybridization pathways may be overestimated due to smoothening of the free-energy landscape generally induced by coarse-graining.[54] The total change in adenine FTIR and 2D IR signal upon thermal dissociation is nearly constant in T11-WT and each AP sequence, suggesting each duplex contains a similar degree of A:T base pairing at lowest measured temperature in the duplex state.[43] Further, 3SPN.2 predicts that partial dehybridization of T11-AP4 and T11-AP6 occurs at temperatures below $T_{m,MD}$, which broadens

the overall melting curve (reduces slope at $T_{m,MD}$ by 20% relative to T11-WT). However, only a 6% reduction in the melting curve slope is observed experimentally.

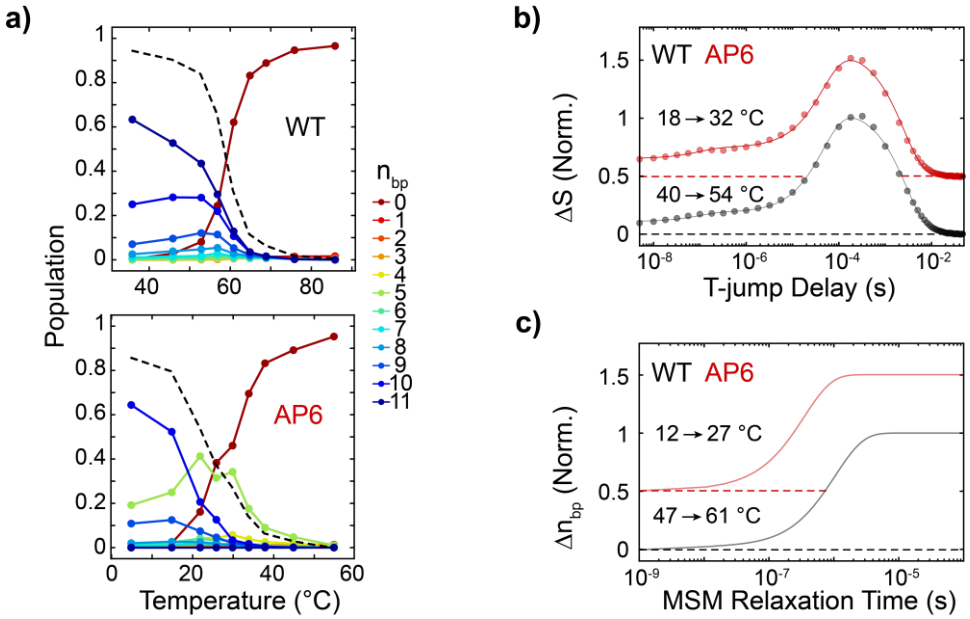

Figure 7.13 Testing for temperature-induced out-of-register shifting. (a) Temperature-dependent population profiles of duplexes with n_{bp} ranging from 0 to 11 for (top) T11-WT and (bottom) T11-AP6. The black dashed line corresponds to the total fraction of intact base pairs. **(b)** Normalized $\Delta S(t)$ time traces at 1605 cm^{-1} (A:T) plotted for the T11-WT and T11-AP6. T-jumps are performed from approximately T_m–15°C to T_m for each sequence. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Solid lines correspond to three-component fits from global lifetime analysis. **(c)** MSM T-jump simulations of the normalized change in A:T base pairs, $\Delta n_{bp}(t)$, using the same method as in Fig. 7.6b. Due to overlapping timescales for partial dehybridization to the out-of-register n_{bp} = 5 state and full-strand dissociation, only a single dehybridization kinetic component is observed in the relaxation profiles.

We also find that MSM and T-jump IR relaxation are insensitive to the kinetics of out-of-register shifting (Figs. 7.13bc). Although a thermally-induced increase of the n_{bp} = 5 state is predicted for T11-AP6 and $\Delta F_{d2}^{\ddagger} < \Delta F_{d1}^{\ddagger}$, the MSM relaxation timescale for shifting overlaps with full-strand dissociation (Fig. 7.13c), which was observed previously in similar canonical

oligonucleotides.[45] Shifting from the n_{bp} = 10 to n_{bp} = 5 state first entails breaking of all in-register contacts followed by the formation of five out-of-register base pairs. The barrier height for shifting is likely much larger than indicated by $\Delta F_{d2}^{\ddagger}$ in Fig. 7.12a, and n_{bp} is a poor coordinate to describe the shifting transition. As a result, no partial-dehybridization response is observed in MSM relaxation profiles. Similarly, T-jump IR measurements only show responses for terminal fraying and full-strand dissociation and cannot indicate nor rule out a role of out-of-register base pairing in these sequences (Fig. 7.12b).

7.5 Conclusions

Our study reveals molecular details underpinning the position-dependent duplex destabilization from an AP site within DNA oligonucleotides by combining temperature-dependent IR and NMR spectroscopy, T-jump IR kinetics, and coarse-grained MD simulations. An AP site destabilizes the duplex through a loss of stacking and base-pairing interactions that disrupt cooperative base pairing and introduce a free-energy barrier for nucleating base-pair segments on each side of the AP site. This nucleation barrier leads to a position-dependent duplex destabilization from the AP site. Experiments and simulations show that the nucleation barrier promotes fraying of the weak segment when the AP site is near the termini. As the AP site moves inward by three to six base-pair sites, the short segment has enough stability to bind, and the frayed configuration becomes metastable as observed in T-jump measurements. At this position, the nucleation barrier is fully encompassed into the duplex and the AP site is maximally destabilizing. The positions over which this transition occurs depends on the nucleobase sequence in the short segment and adjacent to the AP site, which leads to poor prediction of duplex destabilization by

the NN model even at a fixed AP-site position. While the transition occurs at short-segment lengths of two to five base pairs depending on the G:C content of the segment, certain short-segment sequences may be able to circumvent the nucleation barrier by forming base pairs out-of-register as found experimentally for CCends-AP2.

This work has focused on AP sites, yet there are numerous non-canonical base pairs, nucleobase modifications, and damaged-induced lesions that impart a comparable energetic penalty to duplex formation ($\Delta\Delta G^{\circ}_{h37}$ = 5-20 kJ/mol).[8, 13, 55-56] Although they alter local base-pairing and stacking interactions in different ways, such modifications and non-canonical base pairs may disrupt base-pairing cooperativity and follow a position-dependence similar to an AP site. For example, base-pair mismatches impose a sequence-dependent free-energy penalty to duplex formation that follows a similar position-dependence as for AP sites.[8, 19-20] Greater reduction in duplex stability occurs as the mismatch moves inward 3 – 4 base pairs from the termini. While the base-pairing geometry of mismatches is highly sequence-dependent, the 'defect penalty' used to model their behavior is essentially the same as the nucleation penalty described in our work and supports that the position-dependent thermodynamic and dynamic consequences from destabilizing modifications may be similar to what we find for AP sites.

7.6 Acknowledgements

I thank Michael Jones for carrying out the MD simulations and building the MSMs used in this chapter as well as for many helpful discussions. I also thank Joseph Sachleben for assistance with NMR measurements. I thank Yumin Lee for helpful comments on the manuscript that covers the material in this chapter.

7.7 References

1. Bumgarner, R., Overview of DNA microarrays: types, applications, and their future. *Curr. Protoc. Mol. Biol.* **2013,** *101*, 22.1. 1-22.1. 11.

2. Strauss, S.; Jungmann, R., Up to 100-fold speed-up and multiplexing in optimized DNA-PAINT. *Nat. Methods* **2020,** *17*, 789-791.

3. Song, J.; Li, Z.; Wang, P.; Meyer, T.; Mao, C.; Ke, Y., Reconfiguration of DNA molecular arrays driven by information relay. *Science* **2017,** *357*, eaan3377.

4. Klein, M.; Eslami-Mossallam, B.; Arroyo, D. G.; Depken, M., Hybridization kinetics explains CRISPR-Cas off-targeting rules. *Cell Rep.* **2018,** *22*, 1413-1423.

5. Altun, A.; Garcia-Ratés, M.; Neese, F.; Bistoni, G., Unveiling the complex pattern of intermolecular interactions responsible for the stability of the DNA duplex. *Chem. Sci.* **2021,** *12*, 12785-12793.

6. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D., Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006,** *34*, 564-574.

7. SantaLucia, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci.* **1998,** *95*, 1460-1465.

8. SantaLucia Jr, J.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-440.

9. Hertel, S.; Spinney, R.; Xu, S.; Ouldridge, T. E.; Morris, R.; Lee, L., The stability and number of nucleating interactions determine DNA hybridization rates in the absence of secondary structure. *Nucleic Acids Res.* **2021,** *50*, 7829-7841.

10. Zhang, J. X.; Fang, J. Z.; Duan, W.; Wu, L. R.; Zhang, A. W.; Dalchau, N.; Yordanov, B.; Petersen, R.; Phillips, A.; Zhang, D. Y., Predicting DNA hybridization kinetics from sequence. *Nat. Chem.* **2018,** *10*, 91-98.

11. Todisco, M.; Szostak, J. W., Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Res.* **2022,** *50*, 9647-9662.

12. Phan, T. T.; Phan, T. M.; Schmit, J. D., Beneficial and detrimental effects of non-specific binding during DNA hybridization. *Biophys. J.* **2023,** *122*, 835-848.

13. Dubini, R. C. A.; Korytiaková, E.; Schinkel, T.; Heinrichs, P.; Carell, T.; Rovó, P., 1H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives. *ACS Phys. Chem. Au* **2022,** *2*, 237-246.

14. Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A., Oxidized derivatives of 5-methylcytosine alter the stability and dehybridization dynamics of duplex DNA. *J. Phys. Chem. B* **2020,** *124*, 1160-1174.

15. Liu, B.; Shi, H.; Rangadurai, A.; Nussbaumer, F.; Chu, C.-C.; Erharter, K. A.; Case, D. A.; Kreutz, C.; Al-Hashimi, H. M., A quantitative model predicts how m6A reshapes the kinetic landscape of nucleic acid hybridization and conformational transitions. *Nat. Commun.* **2021,** *12*, 1-17.

16. Watkins Jr, N. E.; SantaLucia Jr, J., Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.* **2005,** *33*, 6258-6267.

17. Szabat, M.; Prochota, M.; Kierzek, R.; Kierzek, E.; Mathews, D. H., A Test and Refinement of Folding Free Energy Nearest Neighbor Parameters for RNA Including N6-Methyladenosine. *J. Mol. Biol.* **2022**, 167632.

18. Hopfinger, M. C.; Kirkpatrick, C. C.; Znosko, B. M., Predictions and analyses of RNA nearest neighbor parameters for modified nucleotides. *Nucleic Acids Res.* **2020,** *48*, 8901-8913.

19. Naiser, T.; Kayser, J.; Mai, T.; Michel, W.; Ott, A., Position dependent mismatch discrimination on DNA microarrays–experiments and model. *BMC Bioinform.* **2008,** *9*, 1-12.

20. Naiser, T.; Ehler, O.; Kayser, J.; Mai, T.; Michel, W.; Ott, A., Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges. *BMC Biotechnol.* **2008,** *8*, 1-23.

21. Pozhitkov, A.; Noble, P. A.; Domazet-Lošo, T.; Nolte, A. W.; Sonnenberg, R.; Staehler, P.; Beier, M.; Tautz, D., Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.* **2006,** *34*, e66-e66.

22. Ding, F.; Cocco, S.; Raj, S.; Manosas, M.; Nguyen, T. T. T.; Spiering, M. M.; Bensimon, D.; Allemand, J.-F.; Croquette, V., Displacement and dissociation of oligonucleotides during DNA hairpin closure under strain. *Nucleic Acids Res.* **2022,** *50*, 12082-12093.

23. Lindahl, T., Instability and decay of the primary structure of DNA. *Nature* **1993,** *362*, 709-715.

24. Thompson, P. S.; Cortez, D., New insights into abasic site repair and tolerance. *DNA Repair* **2020,** *90*, 102866.

25. Kimsey, I. J.; Szymanski, E. S.; Zahurancik, W. J.; Shakya, A.; Xue, Y.; Chu, C.-C.; Sathyamoorthy, B.; Suo, Z.; Al-Hashimi, H. M., Dynamic basis for dG• dT misincorporation via tautomerization and ionization. *Nature* **2018,** *554*, 195-201.

26. Afek, A.; Shi, H.; Rangadurai, A.; Sahay, H.; Senitzki, A.; Xhani, S.; Fang, M.; Salinas, R.; Mielko, Z.; Pufall, M. A.; Poon, G. M. K.; Haran, T. E.; Schumacher, M. A.; Al-Hashimi, H. M.; Gordan, R., DNA mismatches reveal conformational penalties in protein–DNA recognition. *Nature* **2020,** *587*, 291-296.

27. Rossetti, G.; Dans, P. D.; Gomez-Pinto, I.; Ivani, I.; Gonzalez, C.; Orozco, M., The structural impact of DNA mismatches. *Nucleic Acids Res.* **2015,** *43*, 4309-4321.

28. Chen, J.; Dupradeau, F.-Y.; Case, D. A.; Turner, C. J.; Stubbe, J., DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* **2008,** *36*, 253-262.

29. Lukin, M.; de Los Santos, C., NMR structures of damaged DNA. *Chem. Rev.* **2006,** *106*, 607-686.

30. Vesnaver, G.; Chang, C.-N.; Eisenberg, M.; Grollman, A. P.; Breslauer, K. J., Influence of abasic and anucleosidic sites on the stability, conformation, and melting behavior of a DNA duplex: correlations of thermodynamic and structural data. *Proc. Nat. Acad. Sci.* **1989,** *86*, 3614-3618.

31. Gelfand, C. A.; Plum, G. E.; Grollman, A. P.; Johnson, F.; Breslauer, K. J., Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence. *Biochem.* **1998,** *37*, 7321-7327.

32. Sági, J.; Guliaev, A. B.; Singer, B., 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines. *Biochem.* **2001,** *40*, 3859-3868.

33. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of Energetic and Base Pairing Cooperativity in DNA Duplexes by an Abasic Site. *Proc. Nat. Acad. Sci.* **2023,** *120*, e2219124120.

34. Wang, W.; Walmacq, C.; Chong, J.; Kashlev, M.; Wang, D., Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proceedings of the National Academy of Sciences* **2018,** *115*, E2538-E2545.

35. Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; De Pablo, J. J., An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013,** *139*, 10B604_1.

36. Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J., Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* **2014,** *141*, 07B613_1.

37. Poland, D.; Scheraga, H. A., *Theory of helix-coil transitions in biopolymers*. Academic Press.: 1970.

38. Applequist, J.; Damle, V., Thermodynamics of the helix-coil equilibrium in oligoadenylic acid from hypochromicity studies. *J. Am. Chem. Soc.* **1965,** *87*, 1450-1458.

39. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.

40. Sanstead, P. J.; Tokmakoff, A., A lattice model for the interpretation of oligonucleotide hybridization experiments. *J. Chem. Phys.* **2019,** *150*, 185104.

41. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

42. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

43. Ashwood, B.; Jones, M. S.; Lee, Y.; Sachleben, J. R.; Ferguson, A. L.; Tokmakoff, A., Molecular insight into how the position of an abasic site and its sequence environment influence DNA duplex stability and dynamics. *BioRxiv* **2023**, https://doi.org/10.1101/2023.07.22.550182.

44. Jones, K. C.; Ganim, Z.; Tokmakoff, A., Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 14060-14066.

45. Jones, M. S.; Ashwood, B.; Tokmakoff, A.; Ferguson, A. L., Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, Markov state models, and infrared spectroscopy. *J. Am. Chem. Soc.* **2021,** *143*, 17395-17411.

46. Au, R. Y.; Ng, K. S.; Chi, L. M.; Lam, S. L., Effect of an abasic site on strand slippage in DNA primer-templates. *J. Phys. Chem. B* **2012,** *116*, 14781-14787.

47. Isaksson, J.; Chattopadhyaya, J., A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochem.* **2005,** *44*, 5390-5401.

48. Zhang, W.; Tam, C. P.; Wang, J.; Szostak, J. W., Unusual base-pairing interactions in monomer–template complexes. *ACS Cent. Sci.* **2016,** *2*, 916-926.

49. Lemieux, S.; Major, F., RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.* **2002,** *30*, 4250-4263.

50. Di Michele, L.; Mognetti, B. M.; Yanagishima, T.; Varilly, P.; Ruff, Z.; Frenkel, D.; Eiser, E., Effect of inert tails on the thermodynamics of DNA hybridization. *J. Am. Chem. Soc.* **2014,** *136*, 6538-6541.

51. Ouldridge, T. E.; Šulc, P.; Romano, F.; Doye, J. P.; Louis, A. A., DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **2013,** *41*, 8886-8895.

52. Araque, J.; Robert, M., Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* **2016,** *144*, 125101.

53. Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A., DNA duplex formation with a coarse-grained model. *J. Chem. Theory Comput.* **2014,** *10*, 5020-5035.

54. Xiao, S.; Sharpe, D. J.; Chakraborty, D.; Wales, D. J., Energy landscapes and hybridization pathways for DNA hexamer duplexes. *J. Phys. Chem. Lett* **2019,** *10*, 6771-6779.

55. Roost, C.; Lynch, S. R.; Batista, P. J.; Qu, K.; Chang, H. Y.; Kool, E. T., Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* **2015,** *137*, 2107-2115.

56. Singh, S. K.; Szulik, M. W.; Ganguly, M.; Khutsishvili, I.; Stone, M. P.; Marky, L. A.; Gold, B., Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Res.* **2011,** *39*, 6789-6801.

57. Jost, D.; Everaers, R., A unified Poland-Scheraga model of oligo-and polynucleotide DNA melting: salt effects and predictive power. *Biophys. J.* **2009,** *96*, 1056-1067.

58. Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A., Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* **2008,** *47*, 5336-5353.

59. Zeng, Y.; Montrichok, A.; Zocchi, G., Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.* **2004,** *339*, 67-75.

60. Applequist, J.; Damle, V., Theory of the Effects of Concentration and Chain Length on Helix—Coil Equilibria in Two-Stranded Nucleic Acids. *J. Chem. Phys.* **1963,** *39*, 2719-2721.

61. Sim, A. Y.; Lipfert, J.; Herschlag, D.; Doniach, S., Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E* **2012,** *86*, 021901.

62. Bernasconi, C., *Relaxation kinetics*. Academic Press: 1976.

63. Cho, C.; Urquidi, J.; Singh, S.; Robinson, G. W., Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **1999,** *103*, 1991-1994.

64. Neupane, K.; Hoffer, N. Q.; Woodside, M., Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* **2018,** *121*, 018102.

65. Rauzan, B.; McMichael, E.; Cave, R.; Sevcik, L. R.; Ostrosky, K.; Whitman, E.; Stegemann, R.; Sinclair, A. L.; Serra, M. J.; Deckert, A. A., Kinetics and thermodynamics of DNA, RNA, and hybrid duplex formation. *Biochem.* **2013,** *52*, 765-772.

66. Ashwood, B.; Sanstead, P. J.; Dai, Q.; He, C.; Tokmakoff, A., 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex. *J. Phys. Chem. B* **2019,** *124*, 627-640.

67. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.

68. Pyshnyi, D. V.; Lomzov, A. A.; Pyshnaya, I. A.; Ivanova, E. M., Hybridization of the bridged oligonucleotides with DNA: Thermodynamic and kinetic studies. *J. Biomol. Struct. Dyn.* **2006,** *23*, 567-579.

Appendix 7.A Two-stretch helix-coil model

To gain insight into the sequence-dependent melting behavior of DNA duplexes containing an AP site, we need to go beyond to the two-state thermodynamic description of DNA hybridization and assess the characteristics of the duplex state. Along this line, we employ a cooperative helix-coil transition model based on the perfect-matching helix-coil models of Poland and Scheraga and others.[37-38] Helix-coil models describe the statistics of base pairing in the DNA duplex and have been extended for improved prediction of DNA melting behavior.[57] Base-pairing thermodynamics are primarily described by a nucleation parameter (σ) that contains the statistical weight for forming the first base pair in an unbound region as well as an equilibrium constant (s) for each base pairing formation step adjacent to an already formed base pair.

$$s_i = e^{-\Delta G^{\circ}_{\mathrm{int},i}/RT} \tag{7.A1}$$

$$\sigma_i = e^{-\Delta G^{\circ}_{\mathrm{nuc},i}/RT} \tag{7.A2}$$

ΔG°_{int} and ΔG°_{nuc} are the free energies changes for base-pair formation and nucleating a stretch of base pairs. To fully account for DNA sequence effects, s and σ will be different for each base-pair site i within the duplex.

7.A1 Duplex melting without an AP site

The most general form of the base-pairing partition function, $Z_{int,D}$, describes the statistics of base pairing configurations for a duplex with *N* possible base pairs and accounts for all possible dissociation bubble configurations:

$$Z_{\mathrm{int},D}(N) = \sum_{n_{bp}=0}^{N} \sum_{\nu=0}^{\nu_{\max}} \left[\prod_{j=1}^{n_{bp}} s_{i,j} \prod_{\ell=1}^{\nu} \sigma_{i,\ell} \right] \tag{7.A3}$$

For a given number of intact base pairs, n_{bp}, there are base-pairing configurations with variable number of continuous base-pairing stretches, ν, that depend on n_{bp} and *N* and the maximum number of possible base-pair stretches (ν_{max}). The energy of a given microstate is determined by products of *s* across each intact base pair site, *j*, and σ across each stretch nucleation site, ℓ. While eq. 7.A3 accounts for all possible base-pair configurations, it is too general to be of practical use and can be significantly simplified by making a couple of key assumptions. First, we assume that the thermodynamics of the duplex-to-single-strand transition can largely be captured by a single *s* parameter, $\langle s \rangle$, that is an average over all in-register base-pairing interactions, and that σ is independent of base-pairing position. In practice, $\langle s \rangle$ is computed from the average nearest-neighbor (NN) free-energy parameter across the sequence.

$$\langle s \rangle = e^{-\Delta G^{\circ}_{\mathrm{int}}/RT} = \frac{\sum_{i}^{N-1} \Delta H^{\circ}_{NN,i} - T \sum_{i}^{N-1} \Delta S^{\circ}_{NN,i}}{N-1} \tag{7.A4}$$

In eq. 7.A4, ΔH°_{NN} and ΔS°_{NN} come from Santa Lucia's NN parameters for duplex DNA with a correction for counterion concentration.[7, 58] The use of a single average *s* will partially account for sequence-dependent behavior of full dissociation, but it cannot properly capture pre-melting effects such as terminal fraying. The simplified form of $Z_{int,D}$ with $\langle s \rangle$ and site-independent σ is:

$$Z_{\mathrm{int},D}(N) = \sum_{n_{bp}=0}^{N} \sum_{\nu=0}^{\nu_{\max}} \left[W\left(N, n_{bp}, \nu\right) \left\langle s \right\rangle^{n_{bp}} \sigma^{\nu} \right] \tag{7.A5}$$

In eq. 7.A5, a degeneracy term, *W*, corresponds to the number of microstates with given values of *N*, n_{bp}, and *v*. Eq. 7.A5 can be further simplified by neglecting microstates with multiple stretches of base pairs. The entropic penalty of forming short loops of dissociated base pairs makes configurations with multiple base pairing stretches highly unfavorable in short canonical sequences like those studied in this work (*N* = 11). Previous statistical models and simulations have observed negligible population of configurations with loops in short canonical sequences,[40, 45, 52] and experiments have indicated that a minimum of ~20 A:T base pairs flanked by G:C regions are needed to generate non-negligible bubble populations.[59] Therefore, we limit $Z_{int,D}$ to only consider single stretch configurations (ν = 1) with in-register base pairing.[37-38, 60] For a single base-pairing stretch, the degeneracy factor, *W*, is simplified to $N - n_{bp}$+1.

$$Z_{\mathrm{int},D}(N) = 1 + \sigma \sum_{n_{bp}=1}^{N} \kappa^{n_{bp}} \left(N - n_{bp} + 1\right) \left\langle s \right\rangle^{n} \tag{7.A6}$$

κ is an entropic correction factor that is used to shift the calculated T_m for WT sequences to match experimental values. The choice of κ will depend on σ. From $Z_{int,D}$, the average fraction of intact base pairs per duplex (θ_{int}) can be determined.

$$\theta_{\text{int}} = \frac{\langle s \rangle}{N} \frac{\partial \ln Z_{\text{int},D}}{\partial \langle s \rangle} \tag{7.A7}$$

Short duplexes typically undergo full-strand dissociation at temperatures below the point at which $\theta_{int} \sim 0$ due to non-negligible contributions from the external partition functions ($Z_{ext,i}$) of the single-strand and duplex species. The dissociation equilibrium constant (K_d) defined in Appendix 6.A can be re-cast in terms $Z_{ext,i}$ and $Z_{int,i}$ for the single-strands and duplex.

$$K_d = \frac{[S_1][S_2]}{[D]} = \frac{Z_{\text{int},S1}^{N_{S1}} Z_{ext,S1} Z_{\text{int},S2}^{N_{S2}} Z_{ext,S2}}{Z_{\text{int},D}^{N_D} Z_{ext,D}} \tag{7.A8}$$

N_{S1} and N_{S2} are the number of each single-strand molecules in solution. The external partition functions account for the greater translational entropy in the single-strand state compared to the duplex as well as the increased stability of the duplex as c_{tot} increases. Similar to a lattice gas model,[37] we calculate the external partition functions using the number of possible ways to arrange single-strand and duplex DNA molecules on a 3D lattice. The volume of a single lattice site (V_{ss}) is set to that of a single-strand molecule where $V_{ss} = 4\pi R_G^3/3$. We use R_G = 1.47 nm for N = 11 and [Na^+] = 600 mM based on the experimentally determined length and ionic strength scaling relationships for the radius of gyration (R_G) of single-strand DNA.[61] We neglect any minor sequence-dependence to R_G and use the same value of V_{ss} for all oligonucleotides in this work.

$Z_{ext,S1}$ and $Z_{ext,S2}$ are determined from the number of ways of placing $N_{S1/S2}$ single-strand molecules on a 3D lattice with M sites, which in the dilute limit ($N_{S1/S2} << M$) is:

$$Z_{ext,S1/S2} \approx \frac{M^{N_{S1/S2}}}{N_{S1/S2}!} \text{ when } N_{S1/S2} << M \tag{7.A9}$$

M is defined by $M = V/V_{ss}$ where V is the total system volume. Duplexes are defined when a single-strand occupies any of the 6 lattice sites adjacent to another single strand. $Z_{ext,D}$ then corresponds to the number of ways to arrange the number of duplexed oligonucleotides, N_D, on M lattice sites where N_D are restricted to the 6 adjacent sites around an already occupied site.

$$Z_{ext,D} \approx \frac{M^{N_D} 6^{N_D}}{N_D!} \text{ when } N_D << M \tag{7.A10}$$

We can put the external partition functions in terms of known parameters by first determining the chemical potential (μ) of each species i.

$$\mu_i = -k_B T \left(\partial \ln Z_{tot,i} / \partial N_i \right) \text{ where } Z_{tot,i} = Z_{ext,i} Z_{\text{int},i}^{N_i} \tag{7.A11}$$

Combining eq. 7.A11 with the equilibrium definition $\mu_{S1} + \mu_{S2} = \mu_D$, the following relation can be derived.

$$\frac{N_{S1} N_{S2}}{N_D} = \frac{M Z_{\text{int},S1} Z_{\text{int},S2}}{6 Z_{\text{int},D}} \tag{7.A12}$$

Using eq. 7.A12 and relating the number of molecules in the system to concentration, $N_i = \mathrm{N}_A [i] V$, we can re-express K_d (eq. 7.A8).

$$K_d = \frac{1}{\gamma}\frac{Z_{\text{int},S1}Z_{\text{int},S2}}{Z_{\text{int},D}} \quad \text{where} \quad \gamma = 6c^{\circ}\mathrm{N}_A V_{ss} \tag{7.A13}$$

$\mathrm{N_A}$ is Avogadro's constant and c° is the standard state concentration of 1 M. Our treatment of the internal partition functions exclusively consider base pairing configurations, therefore $Z_{int,S1}$ and $Z_{int,S2}$ each have a statistical weight of 1 relative to $Z_{int,D}$ and eq. 7.A13 can be further simplified.

$$K_d = \frac{1}{\gamma Z_{\text{int},D}} \tag{7.A14}$$

Just as for the two-state thermodynamic model described in Appendix 6.A, the fraction of oligonucleotide strands in the duplex state (θ_{ext}) is determined by K_d and c_{tot}.

$$\theta_{ext} = 1 + \frac{K_d - \sqrt{K_d^2 + 2c_{tot}K_d}}{c_{tot}} \tag{7.A15}$$

Eqs. 7.A14 and 7.A15 can be combined to give an expression for θ_{ext} in terms of $Z_{int,D}$.

$$\theta_{ext} = 1 + \frac{1 - \sqrt{1 + 2c_{tot}\gamma Z_{\text{int},D}}}{c_{tot}\gamma Z_{\text{int},D}} \tag{7.A16}$$

7.A.2 Duplex melting with an AP site

Our helix-coil model of DNA duplexes containing an AP site is illustrated in Fig. 7.2. We model the AP site as a defect in the duplex that splits base pairing into two possible stretches of length N_1 and N_2, where $N_1 + N_2 = N - 1$. For simplicity, each stretch is considered independent and contains its own $\langle s \rangle$ and σ parameters. The $\langle s \rangle$ parameter for each stretch is determined from average NN parameters across each stretch.[7]

$$\left\langle s_j \right\rangle = e^{-\Delta G^{\circ}_{\mathrm{int},j}/RT} = \frac{\sum_{i}^{N_j-1} \Delta H^{\circ}_{NN,i} - T\sum_{i}^{N_j-1} \Delta S^{\circ}_{NN,i}}{N_j - 1} \quad j = 1,2 \text{ when } N_j > 1 \tag{7.A17a}$$

$$\left\langle s_j \right\rangle = e^{-\Delta G^{\circ}_{\mathrm{int},j}/RT} = \Delta H^{\circ}_{NN} - T\Delta S^{\circ}_{NN} \quad j = 1,2 \text{ when } N_j = 1 \tag{7.A17b}$$

When a stretch has only a single base pair (N_1 or $N_2 = 1$), $\langle s \rangle$ is approximated by the NN parameter for a single dinucleotide step. By the design of $\langle s \rangle$, then, $Z_{int,D}$ can be described as a product of internal partition functions for stretches 1 and 2.

$$Z_{\mathrm{int},D1} = 1 + \sigma_1 \sum_{n_{bp,1}=1}^{N_1} \kappa^{n_{bp,1}} (N_1 - n_{bp,1} + 1) \left\langle s_1 \right\rangle^{n_{bp,1}} \tag{7.A18a}$$

$$Z_{\mathrm{int},D2} = 1 + \sigma_2 \sum_{n_{bp,2}=1}^{N_2} \kappa^{n_{bp,2}} (N_2 - n_{bp,2} + 1) \left\langle s_2 \right\rangle^{n_{bp,2}} \tag{7.A18b}$$

$$Z_{\mathrm{int},D} = Z_{\mathrm{int},D1} Z_{\mathrm{int},D2} \tag{7.A18c}$$

The form of θ_{int} is split into a term for each stretch while θ_{ext} still follows eq. 7.A15.

$$\theta_{\mathrm{int}} = \frac{\left\langle s_1 \right\rangle}{N_1 + N_2} \frac{\partial \ln Z_{\mathrm{int},D1}}{\partial \left\langle s_1 \right\rangle} + \frac{\left\langle s_2 \right\rangle}{N_1 + N_2} \frac{\partial \ln Z_{\mathrm{int},D2}}{\partial \left\langle s_2 \right\rangle} \tag{7.A19}$$

A key effect of splitting the duplex into two stretches is that the total number of possible base-pairing arrangements (W_{tot}) is reduced. For a single base-pairing stretch, W_{tot} is determined from the sum over W in eq. 7.A6.

$$W_{tot}(N) = \frac{N(N+1)}{2} \tag{7.A20}$$

For sequences containing an AP site, W_{tot} is computed from the number of possible base-pairing configurations in each stretch.

$$W_{tot}(N_1, N_2) = \frac{N_1(N_1+1)}{2} + \frac{N_2(N_2+1)}{2} \tag{7.A21}$$

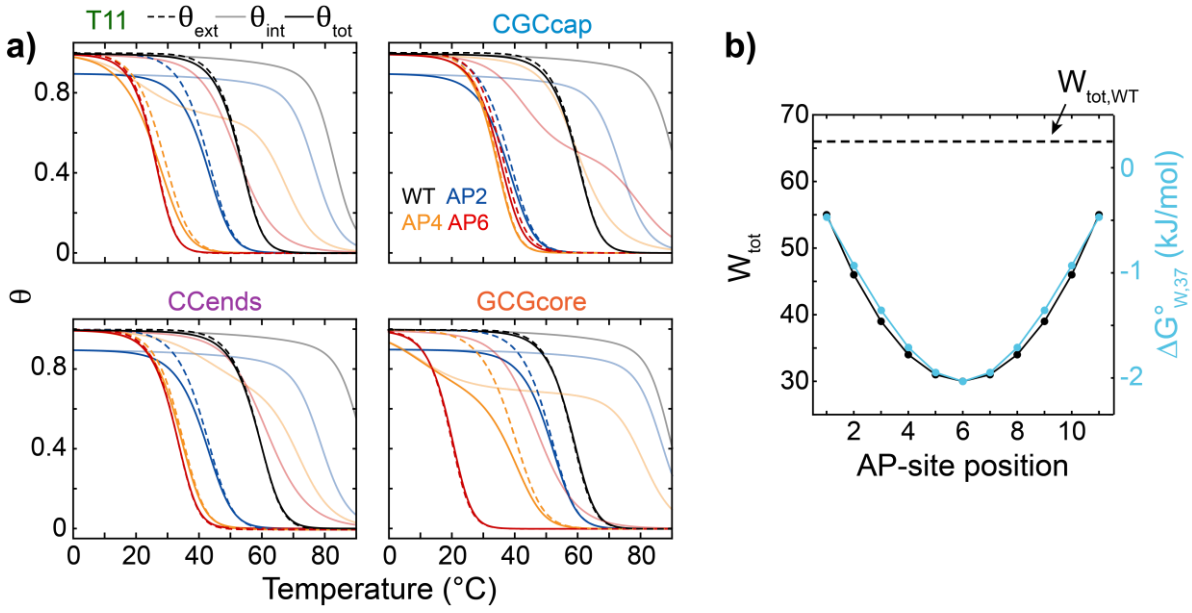

Figure 7.A1 Melting profiles from two-stretch helix-coil model. (a) External (θ_{ext}, dashed line), internal (θ_{int}, light solid line) and total base-pair ($\theta_{tot} = \theta_{int}\theta_{ext}$, dark solid line) melting curves from eqs. 7.A7, 7.A16, 7.A19. A single nucleation parameter ($\sigma = \sigma_1 = \sigma_2$) of 10^{-4} was used for each sequence. An entropic scaling factor (κ) of 2.1, 2.3, 2.3, 2.6 was applied to T11, CGCcap, CCends, and GCGcore sequences, respectively, to match the WT sequence T_m to the experimental value. **(b)** Number of possible in-register base-pairing configurations (W_{tot}) as a function of AP-site position from eqs. 7.A20 and 7.A21. The dashed line indicates W_{tot} for the WT duplex. W_{tot} decreases as an AP site is shifted further from the duplex termini. The free-energy penalty for the reduction in W_{tot} at 37 °C, $\Delta G^{\circ}_{W,37} = -RTln(W_{tot,AP}/W_{tot,WT})$, is shown in blue.

7.A.3 Base-pairing probability distributions and free-energy profiles

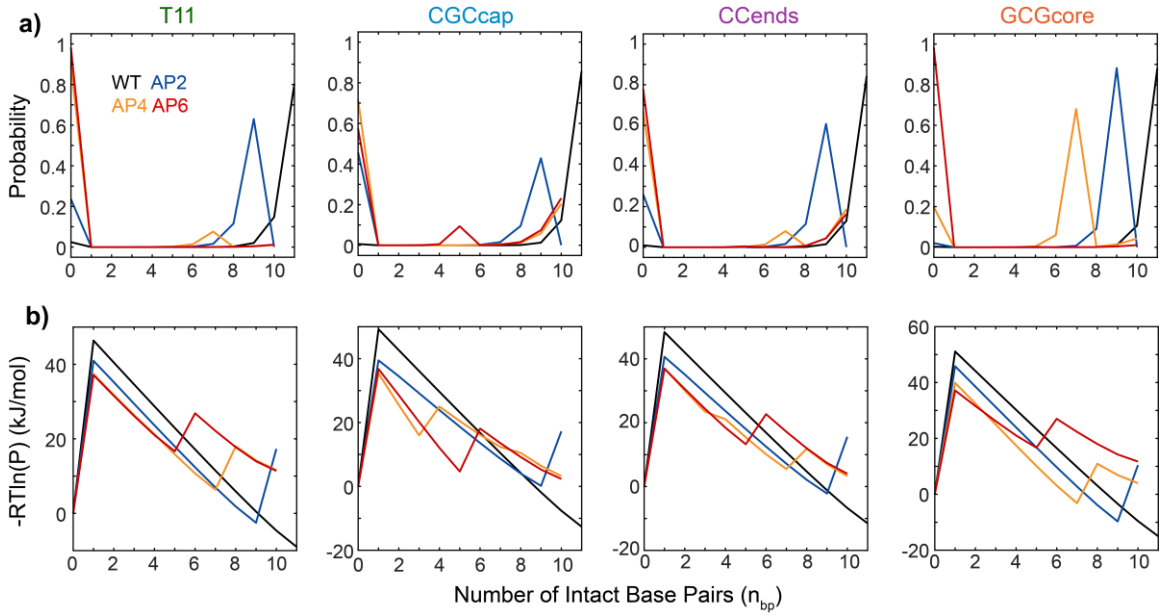

Figure 7.A2 Probability distributions and free-energy profiles from two-stretch helix-coil model. (a) Probability distributions as a function of the number of intact base pairs (n_{bp}) at 37 °C determined from the helix-coil model (eqs. 7.A.22 and 7.A.23). σ, σ_1, and σ_2 were set to 10^{-4} for all sequences and κ was set to 2.1, 2.3, 2.3, and 2.6 for T11, CGCcap, CCends, and GCGcore sequences, respectively. **(b)** Free-energy profiles determined from the probability distributions (P) in **(a)** by $-RTln(P)$, where R is the ideal gas constant.

The probability distribution across n_{bp} can be determined for WT sequences using the statistical weight for a given n_{bp} divided by the internal partition function. The probability of having n_{bp} base pairs is scaled by θ_{ext} to account for full-strand dissociation.

$$P\left(N,n_{np}\right)=\frac{\sigma\left(N-n_{bp}+1\right)\left\langle s\right\rangle^{n_{bp}}}{Z_{\mathrm{int},D}\left(N\right)}\theta_{ext}\ \text{ for } n_{bp}\geq 1 \tag{7.A22a}$$

$$P\left(N,0\right)=1-\sum_{n_{bp}=1}^{N}P\left(N,n_{bp}\right) \tag{7.A22b}$$

In section 7.A2, sequences containing an AP site were described in terms of independent stretches with $n_{bp.1}$ and $n_{bp,2}$ base pairs. To determine the probability distribution across $n_{bp} = n_{bp,1} + n_{bp,2}$ (Fig. 7.A2), we take into account all possible combinations of $n_{bp,1}$ and $n_{bp,2}$ that sum to n_{bp}.

$$P\left(N, n_{bp}\right) = \theta_{ext} \sum_{n_{bp,1}=x}^{y} P_1\left(N_1, n_{bp,1}\right) P_2\left(N_2, n_{bp} - n_{bp,1}\right) \text{ for } n_{bp} \geq 1 \tag{7.A23}$$

$$\text{where, } P_m\left(N_m, n_{bp,m}\right) = \frac{\sigma_m \left(N_m - n_{bp,m} + 1\right) \left\langle s_i \right\rangle^{n_{bp,i}}}{Z_{\text{int},Dm}\left(N_m\right)} \text{ for } m = 1 \text{ or } 2$$

$$\begin{array}{cccc} & & x & y \\ & n_{bp} \leq N_1 \,\&\, n_{bp} \leq N_2 & 0 & n_{bp} \\ \text{and} & n_{bp} \leq N_1 \,\&\, n_{bp} > N_2 & n_{bp} - N_2 & n_{bp} \\ & n_{bp} > N_1 \,\&\, n_{bp} \leq N_2 & 0 & N_1 \\ & n_{bp} > N_1 \,\&\, n_{bp} > N_2 & n_{bp} - N_2 & N_1 \end{array}$$

Appendix 7.B Two-state kinetics of hybridization

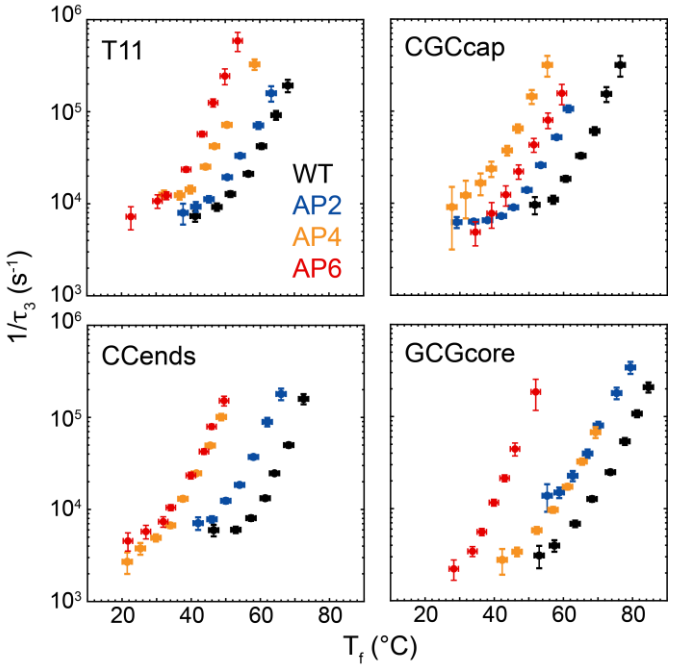

Figure 7.B1 Rate of τ_3 response from T-jump IR data. Observed rates for full-duplex dissociation ($1/\boldsymbol{\tau_3}$) determined from global lifetime fitting of the t-HDVE data as a function of T_f. Vertical error bars indicate 95% confidence intervals from global fits and horizontal error bars correspond to the measured standard deviation in T-jump magnitude.

The temperature-dependence of the T-jump response during τ_3 (Fig. 7.B1) provides insight into how an AP-site alters the kinetics of duplex dissociation and association. Dissociation (k_d) and hybridization (k_h) rate constants were derived by applying a two-state relaxation kinetics model to the observed rates ($1/\tau_3$) (eq. 7.B1, Appendix 3.A). FTIR melting curves (Figs. 7.1 and 7.3) were used to determine single-strand concentrations ([S1] and [S2]) and the dissociation constant (K_d) at each T_f.[62]

$$1/\tau_3 = k_d + k_h\left(\left[S_1\right]_{T_f} + \left[S_2\right]_{T_f}\right) \tag{7.B1}$$

$$1/\tau_3 = \frac{k_d K_{D,T_f}}{K_{d,T_f} + [S_1]_{T_f} + [S_2]_{T_f}}, \text{ where } K_{d,T_f} = \frac{k_{d,T_f}}{k_{h,T_f}} \tag{7.B2}$$

Although many of the AP4 and AP6 sequences exhibit three-state relaxation kinetics overall, we can still apply a two-state treatment to $1/\tau_3$. In this case, reports on the average dehybridization kinetics between the duplex ensemble after terminal fraying and segment-dehybridization during τ_1 and τ_2 and the single-strand ensemble. The temperature-dependence of k_d and k_h (Fig. 7.B2) can be described by a Kramers-like expression in the high-friction limit to extract enthalpic ($\Delta H^{\ddagger}$) and entropic ($\Delta S^{\ddagger}$) barriers where the temperature-dependence of the D_2O solvent viscosity (η) is taken into account.[63]

$$k_{d/h}(T) = \frac{k_b T}{\hbar} \frac{\eta_{37}}{\eta(T)} \exp\left(-\frac{\Delta H^{\dagger}_{d/h} - T\Delta S^{\dagger}_{d/h}}{RT} \right) \tag{7.B3}$$

In eq. 7.B3, we assume a transmission coefficient of 1, yet the true value is likely much smaller.[64] Overestimating the transmission coefficient leads to a reduction in $\Delta S^{\ddagger}$ and therefore an increase in free-energy barrier ($\Delta G^{\ddagger}$) for both dehybridization and hybridization. However, assuming the transmission coefficient or speed limit of dehybridization and hybridization are independent of sequence, the relative trends of $\Delta S^{\ddagger}$ and $\Delta G^{\ddagger}$ across sequence and AP-site position are still informative.

Consistent with previous studies, we find a large enthalpic dehybridization barrier ($\Delta H_d^{\ddagger}$ >150 kJ/mol, Fig. 7.B2) for all sequences. Incorporation of an AP site has a negligible effect or reduces $\Delta H_d^{\ddagger}$ by as much as 120 kJ/mol ($\Delta\Delta H_d^{\ddagger}$). $\Delta G_{d37}^{\ddagger}$ and ΔG_{d37}° are well correlated with a

Pearson correlation coefficient (R) of 0.98 and a slope of 0.90 ± 0.08 (Fig. 7.B2e). Slopes of $\Delta G^{\ddagger}_{d37}$ vs. ΔG°_{d37} near 1 have been observed across canonical oligonucleotides of various sequence or environmental conditions,[65-66] which suggests that the change in $\Delta G^{\ddagger}_{d37}$ is primarily due to destabilization of the duplex rather than a change in transition state energy. In contrast, the hybridization free-energy barrier ($\Delta G^{\ddagger}_{h37}$) shows a poor correlation with ΔG°_{d37} (R = 0.35) and a slope of -0.10 ± 0.08. $\Delta G^{\ddagger}_{h37}$ only varies by ~6 kJ/mol across all sequences, such that all values are within the error of each other. $\Delta G^{\ddagger}_{h37}$ is generally more weakly dependent on sequence or modification than $\Delta G^{\ddagger}_{d37}$,[11, 67] but has been shown to increase slightly from an AP site.[68] Most of our T-jump measurements are performed at temperatures where k_d dominates $1/\tau_3$ such that our determination of k_h is rather indirect and model-dependent. Therefore, a more direct measurement of k_h is required to accurately assess small changes from an AP site.

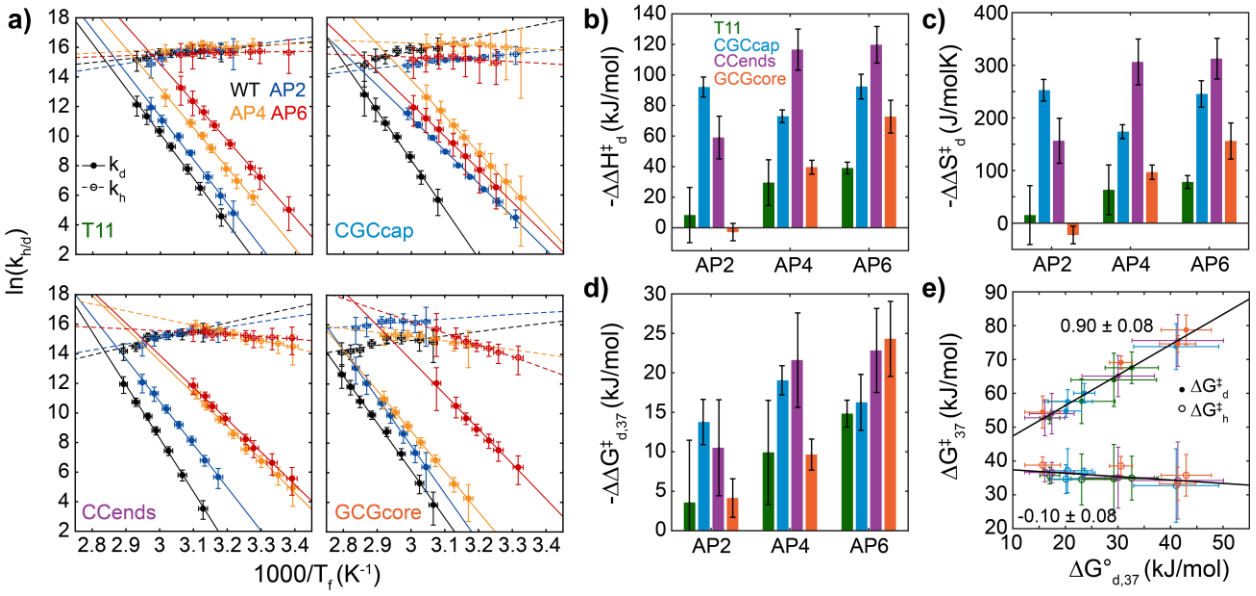

Figure 7.B2 Kinetics of hybridization and dehybridization. (a) Temperature-dependence of dehybridization (k_d, filled circles) and hybridization (k_h, open circles) rate constants determined from two-state analysis of ($1/\tau_3$, Fig. 7.B1). Temperature-trends are fit to eq. 7.B3 (solid and dashed lines). Vertical error bars indicate 95% confidence intervals propagated from global fits and FTIR melting curves and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. Change relative to WT sequence in **(b)** enthalpic ($\Delta\Delta H_d^{\ddagger}$), (c) entropic ($\Delta\Delta S_d^{\ddagger}$), and **(d)** dehybridization free-energy barrier at 37 °C ($\Delta\Delta G_{d,37}^{\ddagger}$) extracted from fits to eq. 7.B3 in **(a)**. Error bars correspond to 95% confidence intervals from fit. **(e)** Scatter plots for $\Delta\Delta G_{d,37}^{\circ}$ vs. $\Delta G_{d,37}^{\ddagger}$ (closed circles) and $\Delta G_{d,37}^{\circ}$ vs. the free-energy hybridization barrier ($\Delta G_{h,37}^{\ddagger}$, open circles) for all sequences with linear fits (black solid lines).

Chapter 8

Disruption of base-pairing cooperativity from cytosine modifications and protonation

The material in this chapter is adapted from:

Ashwood, B.; Sanstead, P. J.; Dai. Q.; He. C.; Tokmakoff, A. 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex. *J. Phys. Chem. B* **2020**, 124, 627-640.

8.1 Introduction

As introduced in Chapter 1, the physical properties of nucleic acids are highly sensitive to environmental factors such as temperature,[1] counterions,[2-3] and pH.[4-5] In particular, pH effects can greatly impact nucleic acids through direct protonation or deprotonation of specific nucleobase sites, leading to changes in secondary structure[5-7] and duplex dissociation.[4, 8] Even before the determination of DNA's double helical structure,[9-10] it had been shown that polymeric DNA dissociates under alkaline and acidic conditions.[11] Since then, many details related to the identities of protonated species, conformational changes, and their thermodynamic impact on the DNA duplex have been investigated.[4-5, 12-13] It has been demonstrated that the thermal melting temperature of DNA is greatly reduced under acidic conditions and largely stems from perturbations to G:C base pairing.[5, 14-15] As a result, the degree of protonation tolerated in dsDNA and its thermodynamic impact is highly-dependent on the level of G:C content.

In general, the pH-dependent properties of nucleic acids observed in biology are driven by protonation at the N3 position (Fig. 8.1) of cytosine (C) due to its relatively high and tunable pK_a

compared to other protonatable bases.[6] In addition to initiating mismatches and G:C Hoogsteen base pairs,[16] N3 protonation drives the formation of C quadruplex structures known as intercalated motifs (i-motifs).[6] Recently, i-motif structures have been observed *in vivo*, particularly in human promoter regions, suggesting that they may play a significant role in gene regulation.[17]

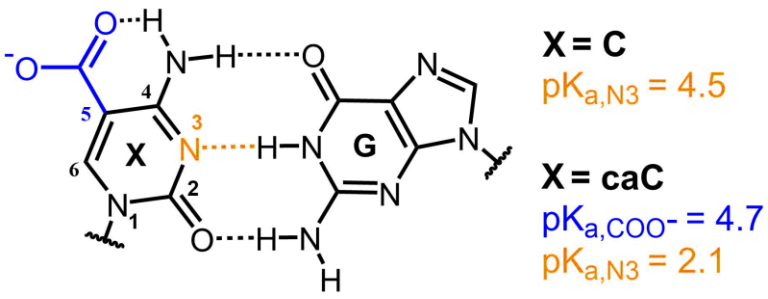

Figure 8.1 Chemical structure of X:G Watson-Crick-Franklin base pair. X corresponds to cytosine (C) or 5-carboxylcytosine (caC). Titratable groups for C and caC are shown with their respective pK_a values for the isolated nucleoside in solution.[38]

While N3 protonation of C is correlated to most pH-induced function of canonical nucleic acids, numerous modified nucleobases exist *in vivo* that have unique pH-dependent properties. For example, 5-bromination lowers the pK_a of uracil from ~10 to ~8,[18] allowing for significant N3 deprotonation under physiological conditions that may facilitate damage to nucleic acids.[19] More recently, cytosine derivatives involved in the active DNA demethylation pathway, where 5-methylcytosine (mC) is sequentially oxidized to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC),[20-21] have been discovered in mammals. In this context fC and caC are selectively excised by thymine DNA glycosylase (TDG) to give an abasic site which is repaired with C by base excision repair to complete the demethylation cycle. Neither mC nor hmC are associated with significant shifts in N3 pK_a (p$K_{a,N3}$). In contrast, fC exhibits a reduced pK$_{a,N3}$ of 2.4 but shows no pH-dependence in its excision activity, while caC

has two titratable groups with pK_a values of 2.1-2.4 (p$K_{a,N3}$) and 4.2-4.8 (C5-COO$^-$, pK_{a,COO^-}),[22-23] and shows accelerated excision under mildly acidic conditions.[22, 24] The pH-dependent behavior of the caC nucleobase stands out among the epigenetic cytosine derivatives since it has a pK_a in a similar range to C, but the titratable group associated with this protonation is the exocyclic carboxyl group rather than N3. While addition of a proton at N3 occupies a hydrogen bond acceptor site that blocks the formation of a Watson-Crick-Franklin (WCF) base pair (Fig. 8.1), protonation of the exocyclic carboxyl group situated in the major groove would appear less perturbative, but could nevertheless lead to unique pH-dependent thermodynamic or kinetic effects in modified dsDNA.

In nucleic acid nanotechnology, the sensitivity of DNA to C protonation has motivated the design of numerous pH-driven nanodevices.[25-26] In particular, the formation of i-motif and DNA triple helices in C- and CG-rich sequences, respectively, require C protonation and offer tunability of nucleic acid secondary structure.[6, 25] These secondary structures are highly sensitive to sequence composition and environmental factors, promoting shifts p$K_{a,N3}$ of by >1 units.[6] Such shifts allow for the construction of uniquely pH-sensitive DNA tools for a variety of applications. For example, pH-driven conformational switches have been used to enhance the location specificity of drug-delivery agents,[27] increase control over DNA strand displacement equilibria,[28] and develop sensors to map the pH of biological environments.[29] While C N3 protonation is the primary mechanism behind these pH-devices, more recent work has demonstrated enhanced tunability through the incorporation of modified nucleobases.[30-32] In particular, substitution of C to caC has been shown to destabilize i-motif formation in many sequences with high sensitivity to the position of modification as well as destabilize DNA triplex formation under physiological conditions.[32-33]

Destabilization of each type of structure is most likely due to the reduction in p$K_{a,N3}$ upon 5-carboxylation. However, the dynamic properties, such as folding rate and mechanism, of pH-driven devices are also essential to their function. As has been shown for 5-bromocytidine,[30] caC may be able to tune specific aspects of i-motif folding or DNA triplex association dynamics important for nanotechnology.

The pH-dependent properties of canonical and modified DNA utilized in biology and nanotechnology are determined by the impact of nucleobase protonation on local and global base-pairing properties. Therefore, a detailed understanding of the pH-dependence of hybridization thermodynamics, kinetics, and structural dynamics is warranted. The incorporation of caC may lead to unique pH-dependences in these properties compared to canonical DNA that are nevertheless tunable over the same pH range due to the fact that pK_{a,COO^-} is similar to p$K_{a,N3}$ in C, the 5-position is oriented in the major groove, and protonation of the exocyclic carboxyl group may not directly disturb caC:G base pairing. So far, the biophysical impact of caC has primarily been studied at neutral pH, showing only minor perturbations to double-stranded DNA.[33-36] However, the protonated carboxyl group is much more electron-withdrawing,[37] leading to a reduction in p$K_{a,N3}$ of caC proposed weakening of G:caC base pairing.[38] Additionally, simulations have suggested that 5-carboxyl protonation may tune local solvation and increase local base-pair fluctuations.[36]

Despite interest in the pH-dependent properties of C and caC, a detailed understanding of nucleobase protonation on duplex dissociation, particularly the kinetics and base-pairing dynamics involved in the process, remains incomplete. Here, we characterize the role of 5-carboxyl and N3

protonation on DNA duplex thermodynamics and dehybridization dynamics of a ten-nucleotide sequence using FTIR and temperature-jump infrared (IR) spectroscopy. The mid-IR spectrum of nucleic acids is sensitive to nucleobase-specific changes in base pairing and protonation,[38-41] making IR spectroscopy particularly useful for the study of pH-dependent hybridization of DNA. We demonstrate that N3 protonation of C drastically shifts the duplex/single-strand equilibrium and reduces the barrier to dissociation without significantly impacting the hybridization transition state energy. In contrast, caC-modified dsDNA can handle essentially complete protonation of the 5-carboxyl sites without significantly shifting the duplex-to-single-strand transition. Instead, 5-carboxyl protonation leads to highly sloped melting curve baselines that are interpreted as a reduction in the internal base pairing within duplexes. Protonation of caC still leads to a large reduction in the dissociation barrier, but, in contrast to N3 protonation of C, does so through lowering the transition state free energy rather than destabilizing the duplex state. Our results demonstrate that caC and C impart distinct pH-dependent properties onto dsDNA, providing insight into the biological function of these protonation events as well as their potential utility in nucleic acid nanotechnology.

8.2 pH-dependent duplex melting of oligonucleotides with C or caC

Figure 8.2 shows FTIR temperature series of the sequence 5′-TAXGXGXGTA-3′, with X = caC or X = C, prepared at pH* 6.8 and 3.7 and ranging from 3 to 97 °C in ~4.5 °C steps. The solution pD is 0.44 greater than the measured pH*,[42] so the deuteron concentration at pH* 6.8 corresponds to standard physiological values. The in-plane ring vibrations, carbonyl stretches,

and -ND_2 bends of the nucleobases absorb in the 1500-1750 cm^{-1} frequency range and each base contributes a unique vibrational fingerprint to the oligonucleotide spectrum.[39, 43-44]

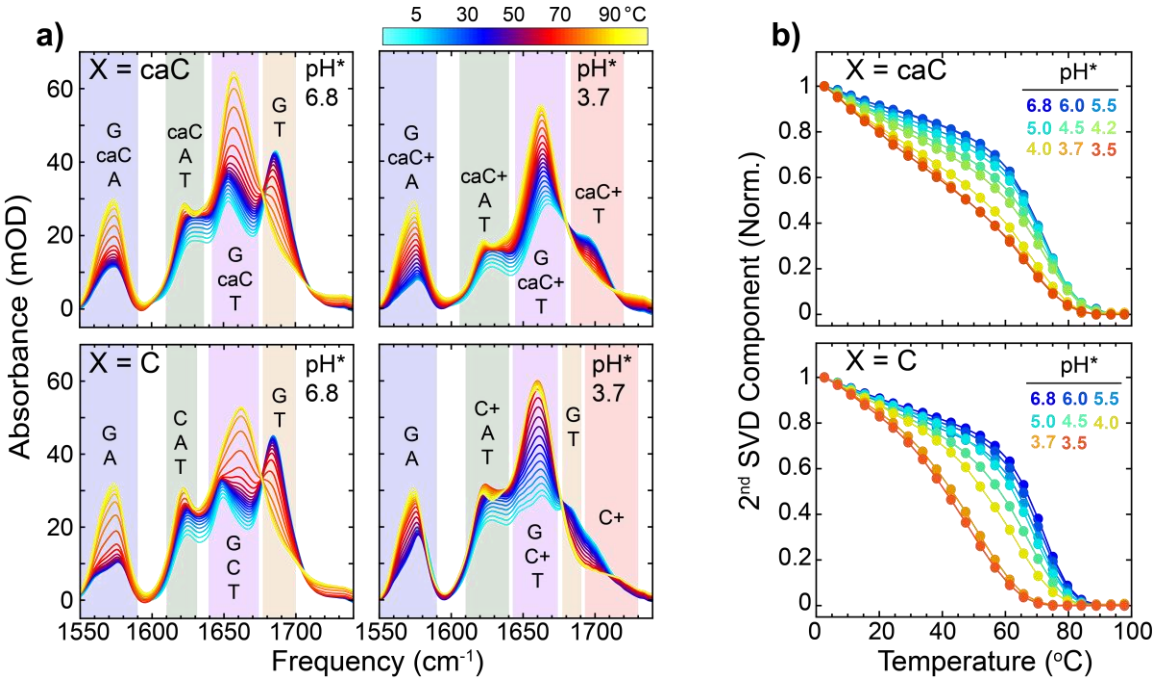

Figure 8.2 Influence of C and caC protonation on duplex melting. (a) FTIR temperature series of 5′-TAXGXGXGTA-3′ for (top) X = caC and (bottom) X = C prepared at pH* 6.8 and pH* 3.7 from 3 to 97 °C. Samples are prepared at an oligonucleotide concentration of 1 mM in 20 mM SPB with 16 mM NaCl and adjusted to the reported pH* with the addition of DCl. Colored windows indicate in-plane nucleobase vibration contributions to each frequency region with the degree of contribution decreasing from top to bottom within a region. **(b)** Normalized 2nd components from singular value decomposition (SVD) across FTIR temperature series from pH* 6.8 to 3.5.

At pH* 6.8 and 3 °C, both sequences exist overwhelmingly as duplexed DNA. The corresponding spectra of these XG-rich duplexes (Fig. 8.2) appear similar to one another, although the carboxyl group of caC mixes strongly with the in-plane base vibrations and produces an additional absorption near 1570 cm^{-1}.[38] As the temperature increases, the spectrum in this frequency range is reshaped significantly and several prominent changes are indicative of

dehybridization. Below 1600 cm^{-1}, the spectrum is dominated by G ring mode absorptions whose intensity increases as X:G base pairing is disrupted. A similar increase corresponding to A ring mode absorption is observed near 1625 cm^{-1}. At higher frequencies, the spectrum is congested with multiple overlapping peaks, with the growth in intensity near 1665 cm^{-1} corresponding to T, X, and G carbonyl absorptions. The increase due to G at this frequency is accompanied by a drop in absorption near 1685 cm^{-1} since this G carbonyl mode is shifted 20 cm^{-1} when engaged in a WCF base pair. The highest frequency T carbonyl at 1690 cm^{-1} can also contribute to this intensity reduction upon the loss of T:A base pairing.

Reducing the pH* of the oligonucleotide solution significantly alters the FTIR spectrum, particularly at low temperature, as evident through comparison of the pH* 6.8 and pH* 3.7 spectra in Fig. 8.2. FTIR spectra measured at intermediate pH* points at both low and high temperature are shown in Fig. 8.3. When X = caC, a loss in absorbance is observed near 1575 and 1625 cm^{-1} as well as a shift of the 1650 cm^{-1} feature to higher frequency and gain at 1700 cm^{-1}. These spectral changes are consistent with the pH*-dependent FTIR spectra of 2′-deoxy-5′-carboxylcytidine (dcaC) (Fig. 8.3). Therefore, we can assign the pH*-dependent changes of X = caC to protonation of the caC nucleobase. Adenosine (N1 pK_a: 3.9-4.0) and guanosine (N7 pK_a: 3.0-3.3) can also be protonated under acidic conditions,[45-46] but comparison of the spectra in Fig. 8.2 against the pH-dependent FTIR spectra of these nucleoside monophosphates suggests there is little of these protonation events over the studied pH* range (Fig. 8.4). The FTIR spectra of X = caC additionally reveal a reduction in amplitude and shift of the 1685 cm^{-1} guanine mode to lower frequency as caC is protonated. The loss of this feature at neutral pH reports on DNA duplex dissociation, and is always accompanied by an increase in absorbance of the G peaks centered at 1575 and 1665 cm^{-1},

as discussed above. For X = caC, the absorption at 1575 cm^{-1} instead decreases in amplitude due to shifts in the caC ring vibration upon protonation, suggesting that the loss of the 1685 cm^{-1} feature at low temperature does not result from duplex dissociation. The frequency and amplitude of the 1685 cm^{-1} peak is known to be sensitive to DNA conformation as well as the local base-pairing and stacking environment,[39, 44, 47-48], therefore numerous factors could be responsible for its modulation as discussed previously.[49]

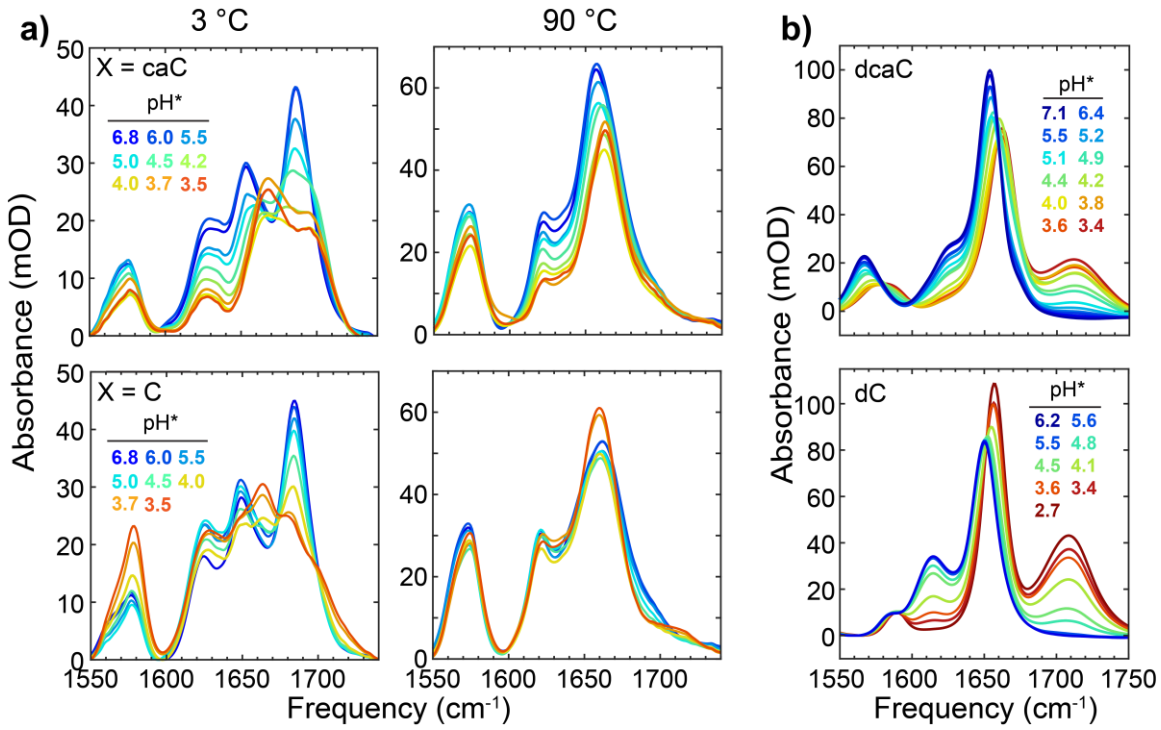

Figure 8.3 pH-dependent FTIR spectra of duplexes and nucleosides. (a) FTIR spectra of (top) X = caC and (bottom) X = C at 3 °C and 90 °C from pH* 6.8 to 3.5. **(b)** FTIR spectra of 2´-deoxy-5-carboxylcytidine (dcaC, top) from pH* 7.1 to 3.4 and 2´-deoxycytidine (dC, bottom) from pH* 6.2 to 2.7 at room temperature in D_2O.

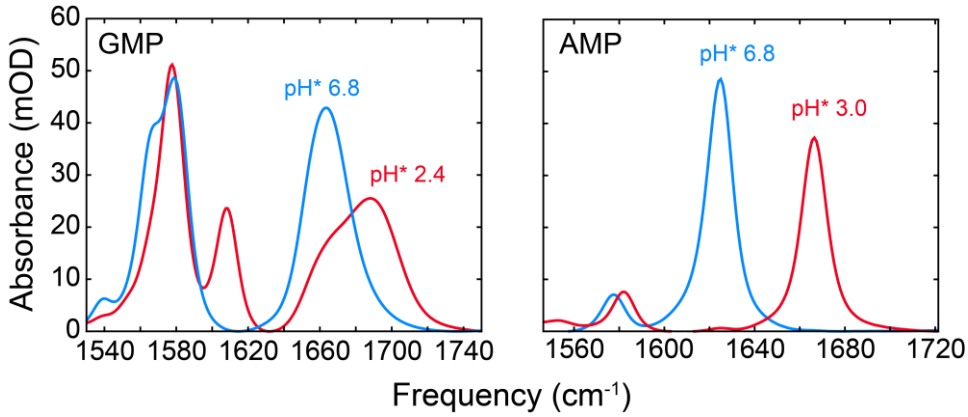

Figure 8.4 FTIR spectra of protonated AMP and GMP. FTIR spectra of 2′-deoxyguanosine 5′-monophosphate (GMP, left) at pH* 6.8 and 2.4 and 2′-deoxyadenosine 5′-monophosphate (AMP, right) at pH* 6.8 and 3.0. Samples were prepared at 3 mg/ml in deuterated 20 mM SPB with 16 mM NaCl.

From pH* 6.8 to 3.5, X = C exhibits a shift in the 1650 cm^{-1} peak to higher frequency and growth of a band near 1700 cm^{-1} at low temperature (Fig. 8.2). Both of these spectral changes are consistent with pH*-dependent FTIR spectra of 2′-deoxycytidine (dC), suggesting N3 of C is protonated when X = C over this pH* range, in agreement with the reported p$K_{a,N3}$ of 4.2-4.5.[13, 23, 38] The significant loss and gain at 1685 and 1660 cm^{-1}, respectively, as well as the intensity gain at 1575 cm^{-1} also suggest changes to guanine either through protonation or a reduction in base pairing. The pK_a of guanine's N7 position (3 – 3.3) is much lower than p$K_{a,N3}$ of C,[45] and signatures of guanine protonation (Fig. 8.4) are not present in either the low or high-temperature oligomer spectra. Instead, the intensity changes in the guanine features must be attributed to the loss of G:C base pairing due to N3 protonation of C.

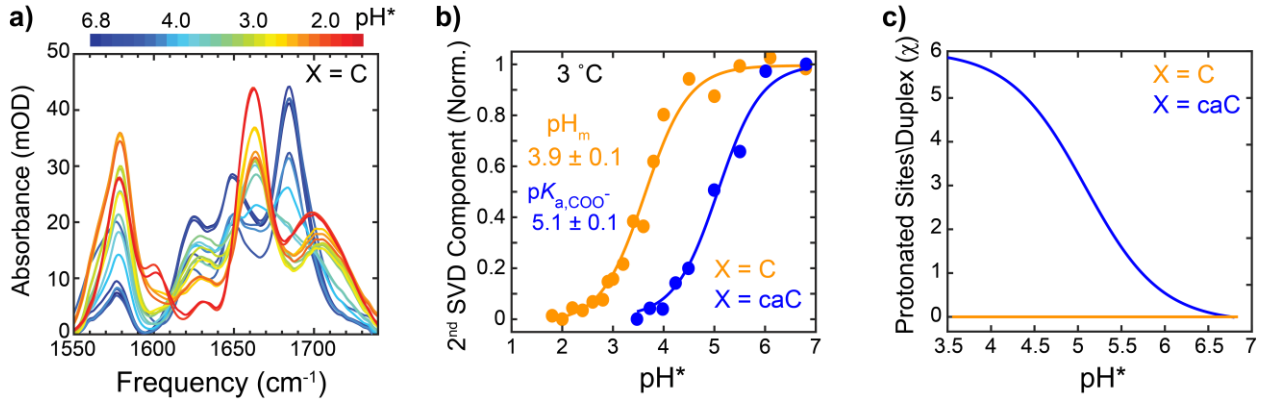

Figure 8.5 pH titration of X = C monitored with FTIR. (a) FTIR titration of X = C from pH* 6.8 to 1.8 performed at 3 °C. Measurements were carried out using a flow cell constructed by Ram Itani. **(b)** 2nd SVD component of FTIR titration on X = C (orange) and X = caC (blue). The 2nd SVD component of X = C and X = caC are fit to eq. 8.A10 and eq. 8.A6, respectively, to extract the inflection point and degree of protonation in the duplex and single strand (Appendix 8.A). The error in pH_m and pK_{a,COO^-} represent the 95% confidence interval of the fit. **(c)** Protonated sites per duplex (χ) calculated for X = caC (blue) and X = C (orange) as a function of pH*.

To further characterize the pH*-dependent spectral changes from N3 protonation, we performed an FTIR titration of the X = C sequence from pH* 6.8 to 1.8 at 3 °C (Fig. 8.5a). Over this pH* range, the DNA duplex dissociates to single strands due to protonation of C, A, G, and the phosphate backbone. The spectral changes as the pH* of the solution decreases in Fig. 8.5a are dominated by features that indicate the loss of G:C base pairing coincident with N3 protonation of C, as assigned above. Evidence of N1 and N7 protonation of A and G, respectively, is only observed at the lowest pH* points sampled (Fig. 8.6). Therefore, the 2nd SVD component of the FTIR spectra across this pH* range should report on the N3 C protonation-driven duplex-to-single-strand transition and can be used to estimate the fraction of intact duplexes at 3 °C as a function of pH*. When fit to a Henderson-Hasselbalch equation, the 2nd SVD component duplex-to-single-strand and N3 protonation trend show an inflection point of 3.9 ± 0.1 that corresponds to the acid

denaturation melting point (pH_m) of the duplex/single-strand equilibrium at 3 °C. In analogy to the melting temperature (T_m) in thermal denaturation experiments, pH_m corresponds to the pH where half of the possible duplex species are intact. The value of pH_m for X = C is consistent with those previously measured for polymeric DNA.[4]

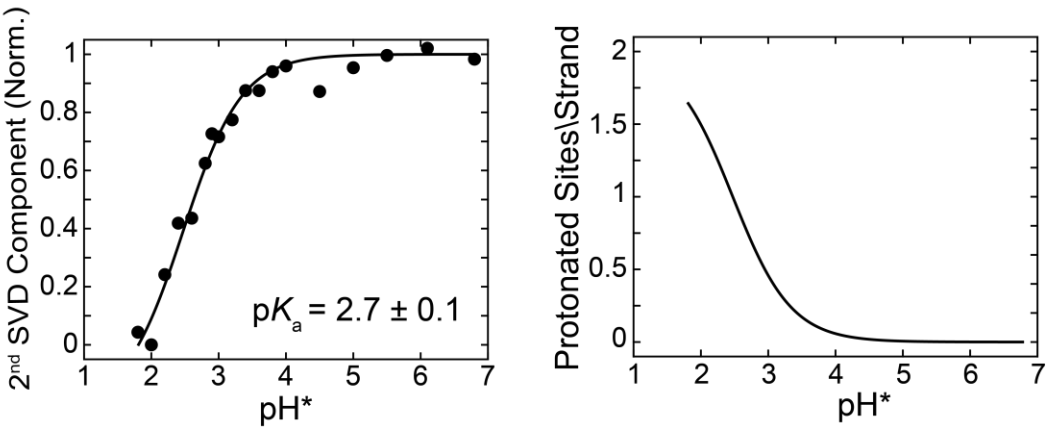

Figure 8.6 Protonation of adenine in X = C. (Left) 2nd SVD component for FTIR titration of X = C over the frequency window 1600-1640 cm^{-1}, which is dominated by changes in the adenine and cytosine ring modes. Data are fit to the Henderson-Hasselbalch equation (solid line). (Right) Number of protonated adenine sites per X = C strand as a function of pH* determined from the fit pK_a. The apparent pK_a for N1 adenine protonation is shifted more than 1 pH unit lower relative to AMP, suggesting that most protonation from pH* 3 – 4 occurs at cytosine N3 sites.

Our pH*-dependent FTIR measurements of X = C and X = caC may be used to estimate the number of protonated C and caC sites, respectively, per duplex (χ) as a function of pH*. For X = caC, an apparent pK_{a,COO^-} of 5.1 is determined from SVD across pH* at 3 °C (Fig. 8.5). The apparent pK_{a,COO^-} shows no measurable change from 3 to 97 °C,[49] suggesting that pK_{a,COO^-} of the X = caC duplex and single-strand are similar. Therefore, we assume that the 5-carboxyl protonation equilibrium for X = caC is identical in the duplex and single-strand and does not strongly depend on temperature. The small change in pK_{a,COO^-} with temperature is consistent with

previous temperature-dependent measurements of nucleic acid protonation.[50] Additionally, the X = caC sequence is assumed to be highly duplexed at low temperature across the pH* range studied here based on the spectroscopic signatures consistent with extensive caC:G base pairing observed at pH* 3.7 at 3 °C discussed above. The resulting degree of X = caC duplex protonation as a function of pH* is shown in Fig. 8.5c. The X = caC duplex is almost completely deprotonated at pH* 6.8 whereas an average of 5+ sites per duplex are protonated below pH* 4.4. Further details regarding our estimation of the degree of protonation are presented in Appendix 8.A.

As discussed above, the pH*-induced duplex-to-single-strand transition in the X = C sequence appears to be driven by N3 protonation of C. The convolution of these two processes indicates that N3 protonation of C occurs over the same pH*-range as duplex denaturation, but it is not clear how protonation is distributed among duplex and single-strand species. Assuming the effective p$K_{a,N3}$ and pH_m at 3 °C are equivalent for X = C, the fraction of intact duplex species (θ_{ext}) and degree of N3 protonation in the duplex and single-strand can be related as a function of pH* (Appendix 8.A). As shown in Fig. 8.5c, X = C is essentially only protonated when dissociated while the duplex cannot tolerate N3 protonation at even one site. This result may also be interpreted as a significant reduction of p$K_{a,N3}$ in the duplex relative to the single-strand,[51] leading to a sharp decline in duplex fraction at pH* values where single-strand protonation becomes favorable.

Having identified the spectroscopic signatures of dehybridization and nucleobase protonation, we can assess the pH* dependent melting behavior of the two model sequences. To take into account the global changes to the spectrum upon melting, SVD is performed on the FTIR temperature series at each pH* to determine a melting curve. The melting curves measured from

pH* 6.8 to 3.5 are shown in Fig. 8.2 for the X = caC and C sequences. While the two sequences have nearly identical melting profiles at pH* 6.8, they exhibit starkly different trends with decreasing pH*. The X = caC melting profile shows only minor shifts in the inflection point of the sigmoidal melting transition with descending pH*, but the low-temperature baseline slope steepens significantly under acidic conditions. In contrast, the melting transition of X = C shifts drastically to lower temperature and broadens as pH* is reduced. Baseline slopes are often observed in thermal melting curves and are attributed to factors such as evaporation, temperature-dependent changes in sample path length, or DNA base solvation.[52-53] However, a comparison of spectral changes along the low-temperature baseline when X = caC (Fig. 8.7) reveals gains in intensity at 1575 and 1660 cm^{-1}, indicating the loss of G:C base pairing, at pH* 3.7 that are not observed under neutral conditions. This observation as well as the abnormally large changes in baseline slope with decreasing pH* when X = caC suggest that the low-temperature baseline is related to changes in base pairing and base stacking within the duplex state. Based on the pH* and temperature dependence of the mid-IR spectra, the pH* dependence of the melting curves, and the assignment of the extent of protonation in the duplex state, we conclude that there is distinct pH-dependent melting behavior due to divergent protonation events in C- and caC-containing duplexes. Whereas the X = C sequence protonates at N3 C over this pH* range and appears unable to tolerate protonation in the duplex state, the X = caC sequence protonates at the exocyclic 5-carboxyl group, resulting in a duplex much more robust to reductions in pH.

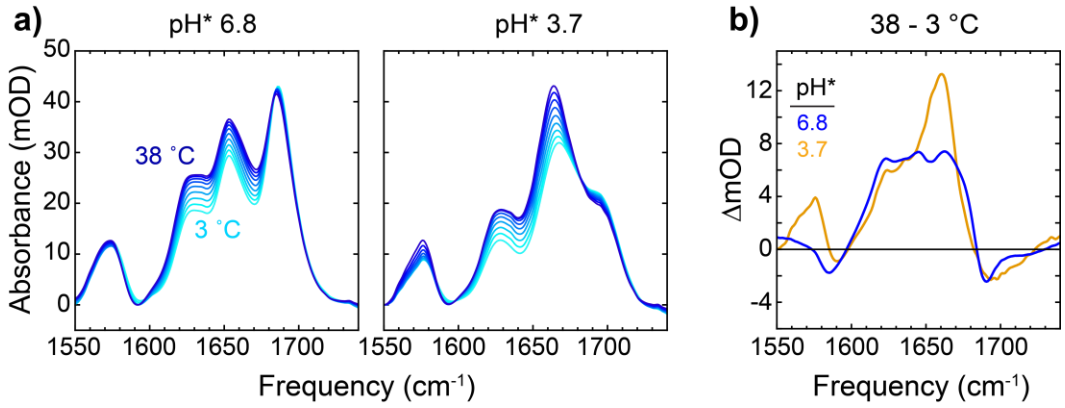

Figure 8.7 FTIR spectral changes along the low-temperature baseline of X = caC. (a) FTIR temperature series of X = caC from 3 to 38 °C at pH* 6.8 and pH* 3.7. **(b)** Difference spectra between 38 and 3 °C.

8.3 Influence of protonation on T-jump IR relaxation kinetics

In addition to equilibrium melting experiments, we employed transient temperature jump (T-jump) measurements to assess the kinetic and dynamic impact of nucleobase protonation in the X = C and caC oligonucleotides. The difference heterodyned dispersed vibrational echo (t-HDVE) spectrum,[54] which can be interpreted similarly to a pump-probe difference spectrum, is used to track changes to the DNA ensemble at delays following the T-jump. Illustrative time traces tracked at the most intense difference feature of 1670 cm^{-1}, which contains contributions from G:C and A:T base pairing, are shown for the X = C and caC sequences at pH* 6.8 and 3.7 in Fig. 8.8. For both sequences and at each pH*, three distinct timescales are observed: (1) a small amplitude rise within ~200 ns (λ_{ns}), (2) a larger rise near ~100 μs ($\lambda_{\mu s}$), and (3) decay of the difference signal due to thermal relaxation and re-hybridization in ~ 2 ms. λ_{ns} is primarily assigned to fraying of the A:T termini while $\lambda_{\mu s}$ corresponds to the duplex-to-single-strand transition, which is supported

by T-jump measurements of short DNA oligonucleotides in Chapter 6 as well as in previous reports.[47, 55-57]

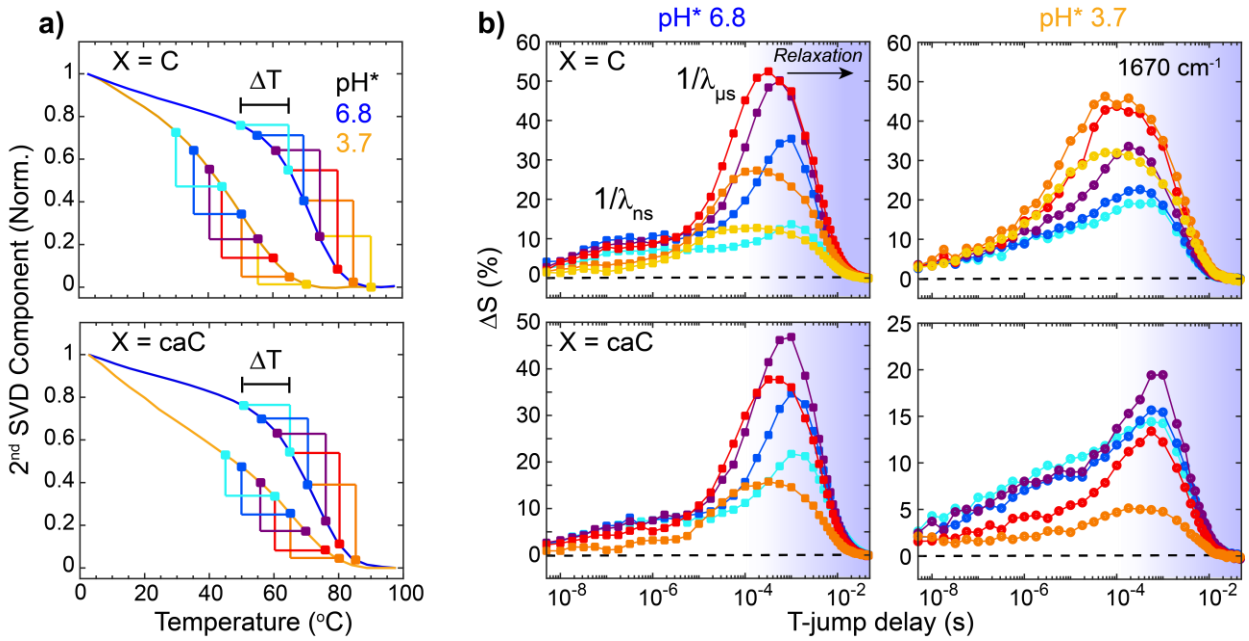

Figure 8.8 T-jump IR spectroscopy of duplex dehybridization for X = C and X = caC. (a) Temperature ranges for T-jump experiments of (top) X = caC and (bottom) X = C along the 2nd SVD components from FTIR temperature series at pH* 6.8 and 3.7. **(b)** t-HDVE kinetic traces probed at 1670 cm^{-1} for each sequence and pH* 6.8 and 3.7 for the T-jump ranges marked in **(a)**. Traces are presented as the percent change in signal relative to the spectrum maximum at T_i, $\Delta S = [S(t) - S(T_i)]/\max[S(T_i)]$. Data at intermediate pH* conditions were previously reported.[49]

For each sequence and pH* condition, T-jump experiments were performed with a fixed jump magnitude (ΔT ≈15 °C) and varying initial temperature (T_i) along the duplex-to-single-strand transition region (Fig. 8.8). Under neutral conditions, X = C and X = caC show similar temperature-dependent relaxation kinetics. In each case, $\lambda_{\mu s}$ increases exponentially with temperature, as observed in previous T-jump experiments of DNA dissociation.[3, 55, 58-59] Additionally, the magnitude of the $\lambda_{\mu s}$ response varies with the expected change in equilibrium

base pairing between T_i and T_f. In contrast to $\lambda_{\mu s}$, the temperature-dependence of λ_{ns} is negligible within our experimental resolution. The magnitude of the λ_{ns} response remains unchanged across the low T_i sampled, but sharply reduces at high temperature. Since λ_{ns} primarily corresponds to fraying of the A:T termini as well as other changes in base pairing, the amplitude of its response is expected to decrease as T_i approaches a condition where the remaining duplexes in solution are largely frayed at equilibrium.

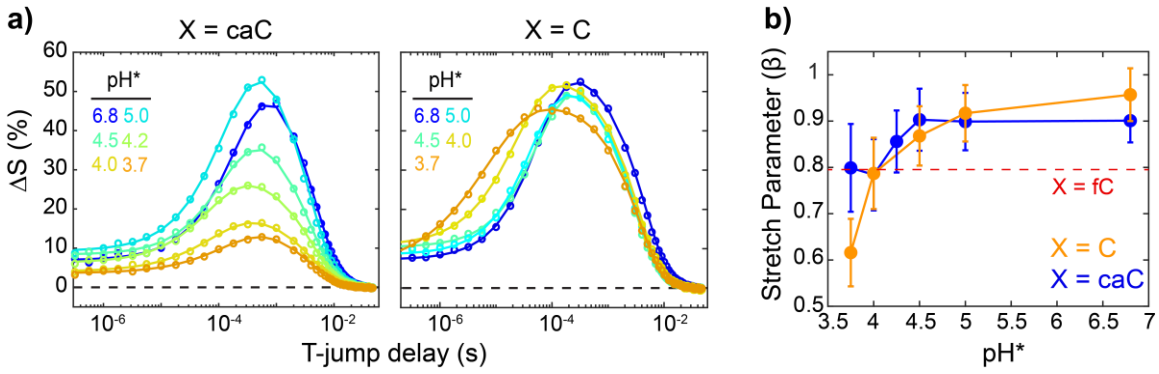

Figure 8.9 Stretched exponential relaxation at acidic pH. (a) T-jump time traces of X = caC and X = C probed at 1670 cm^{-1} from pH* 6.8 to 3.7. Traces are shown for T-jumps across the middle of the duplex-to-single-strand transition. Traces were truncated at 500 ns, and solid lines correspond to fits to a sum of stretched-exponential and exponential decays, $\Delta S(t) = Aexp\left[-(t\lambda_{\mu s})^{\beta}\right] + Bexp[-(t\lambda_{cool})]$. **(b)** Stretch parameter (β) as a function of pH*. β values were averaged across experiments for the middle three T_i conditions. Error bars indicate the 95% confidence interval from the fits averaged over the three temperatures. The red dashed line corresponds to the value from a sequence where cytosine is replaced by 5-formylcytosine (X = fC) at pH* 6.8.[57]

As the solution pH* is reduced, both X = C and X = caC exhibit drastically different relaxation kinetics compared to those at neutral conditions. The $\lambda_{\mu s}$ response becomes increasingly non-exponential (Fig. 8.9) and faster. Additionally, the variation of $\lambda_{\mu s}$ and its associated signal

amplitude with temperature are greatly reduced, signifying a reduction in the barrier to dissociation and cooperativity of the duplex-to-single-strand transition, respectively. For X = caC, the signal change of the duplex/single-strand response decreases at low pH*, consistent with the apparent change in the melting curve shown in Fig. 8.8a and adds further support that the highly sloped low-temperature baseline observed upon 5-carbxoyl protonation arises from changes in internal duplex base pairing. Overall, our results demonstrate N3 and 5-carboxyl protonation have distinct and significant impacts on DNA duplex stability and dehybridization kinetics.

8.4 Self-consistently modeling hybridization thermodynamics and kinetics

To quantify the impact of protonation on the hybridization of our model sequences and to provide a consistent framework with which to discuss all of the experimental results, we propose a model that self-consistently describes the thermodynamics and kinetics of duplex formation. The melting of short oligonucleotides is typically assumed to occur in an all-or-nothing fashion, where DNA strands are fully base-paired or separated (Chapter 6.A).[53, 60] However, the apparent loss of A:T and G:C base pairing at low temperature and asymmetric melting curves exhibited by X = caC suggests a degree of pH-dependent pre-melting within the DNA duplex. To account for both the sigmoidal duplex-to-single-strand transition and the loss of base pairing at lower temperatures that gives rise to a sloping baseline, we interpret the melting curve as a total base-pairing fraction, θ_{tot} that can be separated into external (θ_{ext}) and internal (θ_{int}) contributions.[57, 61]

$$\theta_{tot}(T) = \theta_{\text{int}}(T)\theta_{ext}(T) \tag{8.1}$$

Here, θ_{ext} refers to the fraction of intact duplexes containing at least one base pair and is related to the duplex/single-strand ($D \rightleftharpoons 2S$) equilibrium constant, K_d, that can be described by an external enthalpy (ΔH°_{ext}) and entropy (ΔS°_{ext}). The average fraction of intact base pairs within the duplex is given by θ_{int}. For polymeric DNA, reductions in θ_{int} are the dominant factor in duplex dissociation and typically exhibit a more gradual and less cooperative dependence on temperature relative to θ_{int}.[61]w The low temperature changes in base pairing that manifest as asymmetry in the melting curve also appear to accumulate non-cooperatively with temperature, consistent with the expected profile of θ_{int}. For the purpose of modelling θ_{int}, we derive an expression for the average fraction of intact base pairs with respect to the equilibrium constant for forming or breaking a single base pair that is described by an internal enthalpy (ΔH°_{int}) and entropy (ΔS°_{int}).[57] However, it is an oversimplification to assign low-temperature loss of base pairing to the discrete loss of base-pair contacts for such a short duplex. Instead, it is more likely that the steepening duplex baseline of X = caC upon 5-carboxyl protonation corresponds to the loosening of caC:G base pairing along continuous structural coordinates or an increase in base-pair structural fluctuations as suggested previously.[36]

The model can be extended to incorporate the T-jump results by assuming a two-state equilibrium of self-complementary oligomers, where the observed relaxation rate $\lambda_{\mu s}$ is related to the association (k_a) and dissociation (k_d) rates (Chapter 3.6):[62]

$$\lambda_{\mu s} = k_d + 4\left[S\right]_{T_f} k_a \tag{8.2}$$

The single-strand concentration, [*S*], at the final temperature (T_f) is obtained from the melting curve. We use a maximum entropy inverse-Laplace transform approach (MEM-iLT)[63] to extract λ_{ns} and $\lambda_{\mu s}$, the observed rates of each process. To determine both the activation enthalpy ($\Delta H^{\ddagger}$) and entropy ($\Delta S^{\ddagger}$) of hybridization, we fit k_a and k_d to a Kramers-like model in the high friction limit.[64]

$$k_d = \frac{C_d^o \lambda_{ns}}{\eta(T)} \exp\left(\frac{\Delta S_d^{\ddagger}}{R}\right) \exp\left(\frac{-\Delta H_d^{\ddagger}}{RT}\right) \tag{8.3a}$$

$$k_a = \frac{C_a^o \lambda_{ns}}{\eta(T)} \exp\left(\frac{\Delta S_a^{\ddagger}}{R}\right) \exp\left(\frac{-\Delta H_a^{\ddagger}}{RT}\right) \tag{8.3b}$$

where λ_{ns} is the observed rate of the 10-100 ns response measured in the t-HDVE kinetic traces, and $\eta(T)$ is the temperature-dependent viscosity of D_2O.[65] Here, λ_{ns} is taken as an estimate for the rate of diffusive hybridization in analogy to similar treatments in protein folding studies.[66-68] This stems from our assignment of λ_{ns} as largely corresponding to terminal A:T fraying dynamics, which have been shown to be diffusive in nature.[55] The temperature-dependence of λ_{ns} is negligible within the error of our measurement over the surveyed temperature range, therefore the mean over the lowest three T_i is used in the Kramers analysis. An additional unit parameter is included in the dissociation (C_d°) and association (C_a°) pre-exponential factor fixed at 1 Pa·s and 1 Pa·s·M^{-1}, respectively, which amounts to assuming that the pre-exponential factor is equivalent for association and dissociation, effectively placing any additional contributions into $\Delta S^{\ddagger}$.

For a two-state process on a 1D free-energy profile, the standard free-energy difference between the duplex and single-strand states (ΔG°_{ext}) is equal to the difference between the dissociation ($\Delta G^{\ddagger}_{d}$) and association ($\Delta G^{\ddagger}_{a}$) activation free energies. Therefore, the duplex/single-strand equilibrium constant can be expressed in terms of enthalpic and entropic barriers:

$$K_d(T) = \exp\left(\frac{\Delta S^{\ddagger}_{d} - \Delta S^{\ddagger}_{a}}{R}\right)\exp\left(\frac{\Delta H^{\ddagger}_{a} - \Delta H^{\ddagger}_{d}}{RT}\right) \tag{8.4}$$

where the numerators in the arguments of the exponentials are equal to ΔS°_{ext} and ΔH°_{ext}, respectively. Using eqs. 8.1 – 8.4, the thermodynamic and kinetic data can be globally fit to self-consistently describe the pH*-dependent hybridization thermodynamics and kinetics of X = C and X = caC. In total, the thermodynamic and kinetic data is described by six parameters: $\Delta H^{\ddagger}_{d}$, $\Delta H^{\ddagger}_{a}$, $\Delta S^{\ddagger}_{d}$, $\Delta S^{\ddagger}_{a}$, ΔH°_{int}, ΔS°_{int}. An additional parameter (*A*) that scales the melting 2nd SVD components by a value between 0 and 1 at the lowest temperature measured is also needed.[49] In an all-or-nothing melting model, thermal melting curves are normalized to 1 along the duplex baseline, reflecting that all duplexes are intact at the lowest temperature (Appendix 8.A). However, sequences with substantial changes in internal base pairing prior to duplex dissociation may not be fully intact at the lowest temperature measured, and the normalization amplitude (*A*) used in

fitting the thermal melting curves is proposed to report on the degree of internal base pairing at the lowest temperature sampled (3 °C).

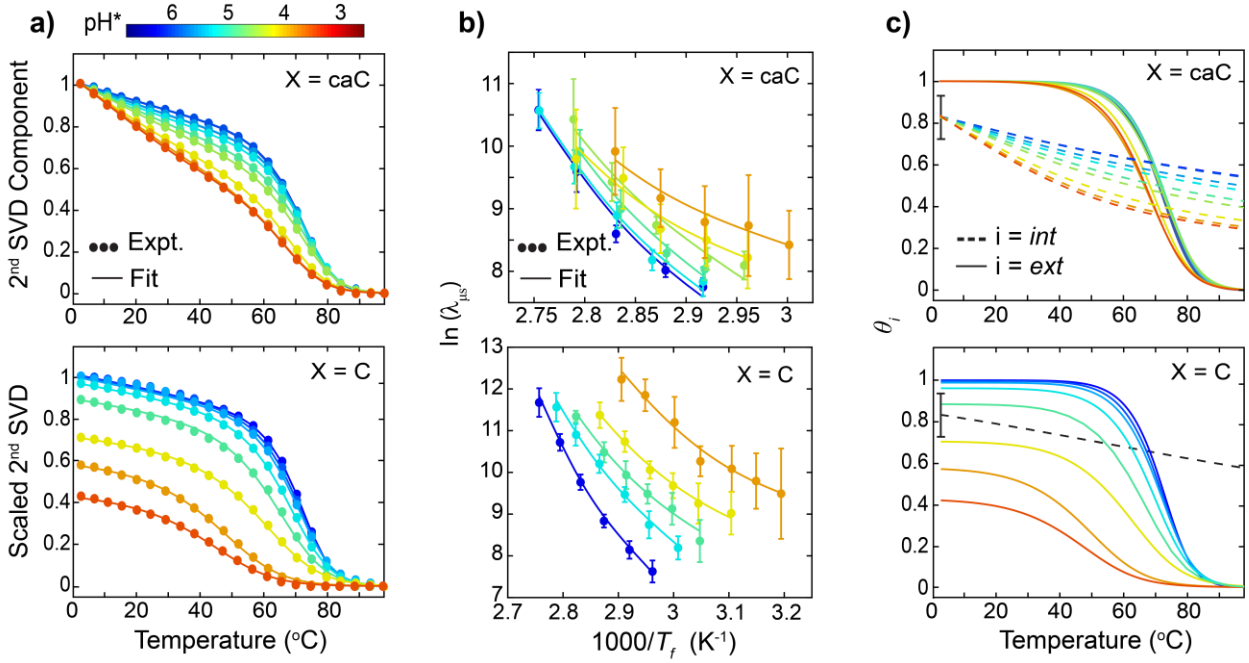

Figure 8.10 Global fitting of thermodynamics and kinetics for duplex melting. Fits to the **(a)** 2nd SVD components from FTIR temperature series and **(b)** observed relaxation rates ($\lambda_{\mu s}$) from T-jump measurements of (top) X = caC and (bottom) X = C using eqs. 8.1 – 8.4. 2nd SVD components at pH* 6.0, 5.5, and 3.5 were not globally fit with $\lambda_{\mu s}$ because T-jump measurements were not performed at those pH* conditions. Error bars in **(b)** represent the amplitude-weighted standard deviation in $\lambda_{\mu s}$ across probe frequency. **(c)** Extracted profiles of θ_{ext} (solid line) and θ_{int} (dashed line) from pH* 6.8 to 3.5. θ_{int} is normalized to 0.825, which is the average value of 1/*A* across pH* for X = caC.[49] The error bar in the value of θ_{int} at 3 °C is the mean error in 1/*A* across pH*.

Figure 8.10 shows fits to the 2nd SVD component melting curves and $\lambda_{\mu s}$ across temperature for X = caC and X = C using eqs. 8.1 – 8.4. T-jump experiments were performed at select pH* points among the measured equilibrium melting experiments. Therefore, melting 2nd SVD components acquired at pH* 6.0, 5.5, and 3.5 were fit without $\lambda_{\mu s}$ to eq. 8.1 using ΔH°_{ext},

ΔS°_{ext}, ΔH°_{int}, and ΔS°_{int}. As shown previously,[49] consistent thermodynamic values are determined when fitting the 2nd SVD components with or without kinetic data. The profiles of θ_{int} and θ_{ext} determined from the fits for X = caC are shown in Fig. 8.10c. Upon a reduction in pH*, the profile of θ_{ext} exhibits a small shift toward lower temperature and broadens slightly while θ_{int} becomes more sharply decreasing along temperature. The value of *A* for X = caC is assumed to account for deviations in θ_{int} from 1 at 3 °C and was found to be insensitive to pH* within the accuracy of our measurement and model.[49] Therefore, the average value of 1/*A* across pH* (0.825) was applied to each 2nd SVD component and the average fit error across pH* is used to estimate the error in *A*.

A comparison of the FTIR temperature series from pH* 6.8 to 3.5 for X = C (Fig. 8.2) shows substantial broadening and shifting of the melting transition as pH* is reduced, indicating a loss of melting cooperativity and overall destabilization of the duplex, respectively. This provides further evidence that the low pH* samples do not begin from a duplex fraction of unity at 3 °C. The thermal 2nd SVD components measured for the X = C sequence can be modeled using eq. 8.1, where both θ_{int} and θ_{ext} contribute to the melting profile. However, our FTIR titration analysis suggests that the X = C sequence cannot tolerate N3 protonation in the duplex state, and it is a reasonable approximation that θ_{int} is independent of pH*. Therefore, we fit the pH* 6.8 2nd SVD component to eq. 8.1, determining θ_{int}, θ_{ext}, and a 3 °C normalization offset of 0.825. Under neutral conditions, the slope of θ_{int} is expected to arise from fraying of the A:T termini. The normalized 2nd SVD components at all pH* points were scaled at 3 °C (Fig. 8.10a) by the value of θ_{ext} determined from the FTIR titration of X = C (Fig. 8.5) and the form of θ_{int} at pH* 6.8 (dashed

line in Fig. 8.10c) was assumed constant across all values of pH*. As shown in Fig. 8.10, both the thermal 2nd SVD components and temperature trend in $\lambda_{\mu s}$ are well fit within this treatment.

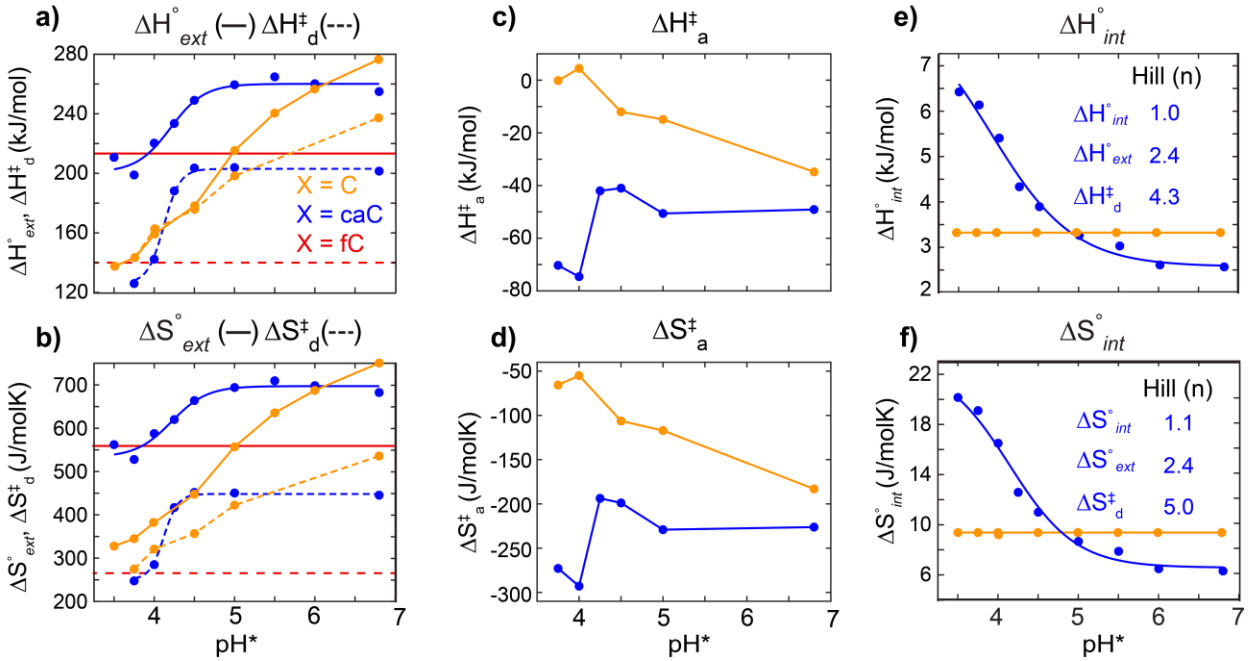

Figure 8.11 Parameters from global fits to melting curves and T-jump relaxation rates. (a) External enthalpy (ΔH°_{ext}, solid lines) and enthalpic dissociation barrier ($\Delta H^{\ddagger}_{d}$, dashed lines) for X = C (orange) and X = caC (blue). Red lines correspond to the value for X = fC at pH* 6.8.[57] **(b)** External entropy (ΔS°_{ext}, solid lines) and entropic dissociation barrier ($\Delta S^{\ddagger}_{d}$, dashed lines). **(c)** Enthalpic association barrier ($\Delta H^{\ddagger}_{a}$). **(d)** Entropic association barrier ($\Delta S^{\ddagger}_{a}$). **(e)** Internal enthalpy (ΔH°_{int}). **(f)** Internal entropy (ΔS°_{int}). Internal and external thermodynamic parameters as well as dissociation barriers for X = caC are fit to a Hill equation (solid lines). Hill coefficients (n) from fits are listed in **(e)** and **(f)**.

Figure 8.11 presents the trends in thermodynamic and kinetic parameters determined from the fits to thermal 2nd SVD components and $\lambda_{\mu s}$ across pH* for X = caC and X = C. Over the studied pH* range, X = caC exhibits reductions in ΔH°_{ext} and ΔS°_{ext} of ~40 kJ/mol and 150 J/molK, respectively. X = C shows substantially greater reductions in ΔH°_{ext} and ΔS°_{ext} of ~140 kJ/mol and 400 kJ/mol, further demonstrating the large disruption of duplex formation upon N3 protonation

shown here and by others.[4, 8] 5-carboxyl protonation also leads to a reduction in θ_{int} as well as increase in ΔH°_{int} and ΔS°_{int}. The dissociation barriers $\Delta H^{\ddagger}_{d}$ and $\Delta S^{\ddagger}_{d}$ both decrease upon 5-carboxyl and N3 protonation while opposite trends are observed in $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ between the two types of protonation. At pH* 6.8, each sequence exhibits negative association barriers as observed for DNA hybridization in many previous works.[3, 55, 58-59, 69] However, $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ each become more negative upon 5-carboxyl protonation and less negative upon N3 protonation. Overall, the activation parameters resemble those in activation energy and pre-exponential factor determined from an Arrhenius analysis of the kinetic data.[49] Further discussion of the hybridization thermodynamic and kinetic parameter trends with pH* follows.

Examining the trends in thermodynamic and kinetic parameters in Fig. 8.11 across pH* reveals distinct behavior due to 5-carboxyl and N3 protonation. For X = caC, each parameter remains almost unchanged between pH* 6.8 and 5.0 and then changes sharply across lower pH* values with titration-like behavior. In contrast, X = C exhibits large, but more gradual changes across the pH* window studied. The trends for X = caC can each be well fit by the Hill equation[70] with consistent apparent pK_a values of ~4.5 and variable Hill coefficients (n). For fitting of binding curves, n > 1 indicates positive cooperativity between binding sites that leads to sharpening of the binding transition. The trends in external thermodynamics as well as enthalpic and entropic barriers reveal sharp transitions that fit to values of n > 1.

8.5 Discussion of results

8.5.1 Impact of protonation on the thermodynamic stability of the duplex

Similar to previous studies,[4-5, 8] the X = C duplex is greatly destabilized as a result of N3 protonation, and our analysis suggests that this specific sequence essentially cannot tolerate any protonation in the duplex state. However, the degree of duplex destabilization from N3 protonation and maximum number of protonated C bases is expected to vary with sequence composition. Previous work with polymeric DNA and short RNA sequences has shown that duplex stability is dependent on the GC content in the sequence.[4, 8, 14] UV CD measurements suggested that up to 50% of C bases in polymeric DNA could be protonated prior to signs of complete duplex dissociation at low temperature.[5] Instead, C protonation leads to disruption of many GC base pairs before full dissociation of the duplex.[14] It is clear that N3 protonation generally destabilizes the duplex state, but the degree and nature of the destabilization will depend on the context of the GC base pairs. Therefore, while X = C completely dissociates upon a N3 protonation event, other sequences may be able to tolerate some degree of protonation depending on the number and location of GC base pairs present. Regardless, it is well-established that the thermodynamics of nucleic acid hybridization in both polymers and oligonucleotides is additive and can be decomposed into the nearest-neighbor contributions of discrete dinucleotide steps.[71-72] It is thus reasonable to expect the present results to apply generally for these base pairs and their local sequence context whether in short duplexes or longer polymeric DNA.

In contrast to the X = C sequence, protonation of the exocyclic carboxyl group of caC is much less perturbative to the duplex/single-strand equilibrium, shifting the melting inflection point

down by ~6 °C in going from pH* 6.8 to 3.5. As a result, the X = caC duplex can tolerate near complete protonation of caC sites (Fig. 8.5), and this leads to measurable changes in internal base pairing in the duplex. As caC is protonated, clear signatures of the loss of G:C base pairing appear along the low-temperature baseline of X = caC (Fig. 8.7). For short oligonucleotides like those studied here, it is unlikely that the low-temperature G:C changes correspond to complete loss of discrete base-pair contacts. Instead, these signatures suggest loosening of hydrogen bonding or base stacking along continuous structural coordinates. Recent MD simulations of caC-containing duplexes demonstrated that 5-carboxyl protonation increases the degree of structural fluctuations within the modified X:G base pair.[36] Motions along these structural coordinates alter the intramolecular base-pair distance and orientation, which are expected to alter guanine and C vibrational modes. Therefore such structural fluctuations may account for the observed melting curve baselines.[44] The physical origin for this base pair loosening may stem from weakened caC:G base pairing due to the electron-withdrawing nature of the protonated carboxylic acid group, consistent with similar base pair loosening observed for DNA containing 5-formylcytosine.[57]

The assignment of the two pK_a values observed for caC (2.1-2.4 and 4.2-4.8) has been debated in recent reports.[38, 73-74] Using FTIR-monitored titrations and simulated IR spectra of 2´-deoxy-5-carboxylcytosine, we previously concluded that the pK_a values at 4.2-4.8 and 2.1-2.4 correspond to protonation of the 5-carboxyl and N3 positions, respectively.[38] The pH-dependent behavior of X = C and X = caC observed in the current study further support these assignments. Overall, protonation of X = C is shown to disrupt the DNA duplex to a much greater degree than protonation of X = caC from pH* 6.8 – 3.5. If X = caC was protonated at the N3 position in the studied pH* range, similar pH-dependent behavior to X = C would be expected. Therefore, we

confidently assign the pK_a value at 4.2-4.8 in the caC nucleobase and within duplex DNA to protonation of the 5-carboxyl position.

8.5.2 Influence of N3 and 5-carboxyl protonation on hybridization kinetics

To date, the impact of nucleobase protonation on dissociation and hybridization has remained elusive from a dynamical perspective, despite the importance of these processes in biology and nanotechnology applications. Not only do our results add insight to how N3 protonation impacts the stability of short oligonucleotides as well as the tunability of base-pairing stabilities through caC protonation, but we also characterized the impact on the kinetics of hybridization.

Just as N3 protonation greatly destabilizes the DNA duplex, it also increases the rate of duplex dissociation. For example, k_d increases by nearly a factor of 500 at T_i = 55 °C between pH* 6.8 and 3.7. 5-carboxyl protonation leads to a far more modest increase in the dissociation rate at T_i = 55 °C by a factor of 7-10 between pH* 6.8 and 3.7. Global fitting of the kinetics and thermodynamics indicate that both N3 and 5-carboxyl protonated oligonucleotides are still well described by a two-state duplex/single-strand model. Over the pH* range studied, each type of protonation results in a similar reduction of 80-100 kJ/mol in $\Delta H_d^{\ddagger}$ and 200-250 J/molK in $\Delta S_d^{\ddagger}$. The trend upon N3 protonation suggests that $\Delta H_d^{\ddagger}$ and $\Delta S_d^{\ddagger}$ would continue to decrease at pH* <3.5. However, due to significant protonation of adenine and eventually guanine at pH* <3.5, we have restricted our study to the pH* window between 6.8 and 3.7. In contrast to the dissociation barriers, both $\Delta H_a^{\ddagger}$ and $\Delta S_a^{\ddagger}$ become less negative for X = C and more negative for X = caC as pH* decreases.

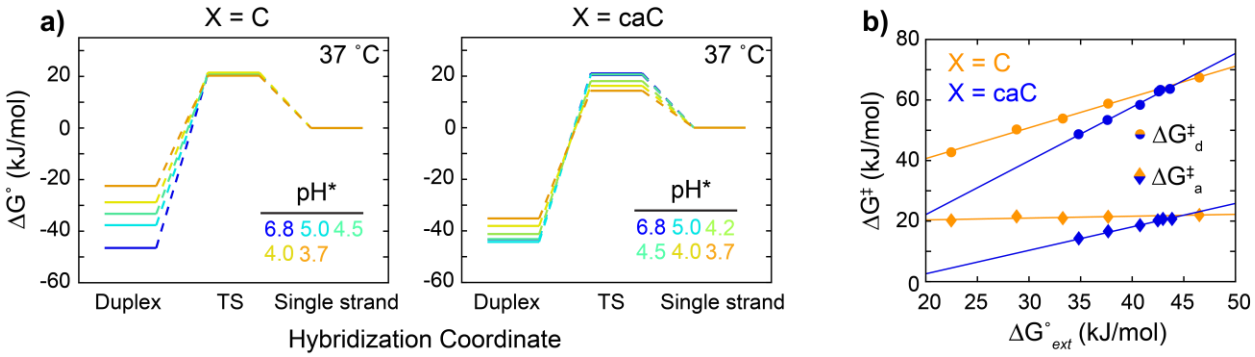

Figure 8.12 pH-dependent hybridization free-energy plots. Representative 1D free-energy plots for X = C and X = caC at 37 °C constructed using the association ($\Delta G_a^{\ddagger}$) and dissociation ($\Delta G_d^{\ddagger}$) free energy barriers determined from global fitting between T-jump and FTIR temperature series results. **(b)** Relationship between $\Delta G_d^{\ddagger}$ (circles) and $\Delta G_a^{\ddagger}$ (diamonds) and ΔG_{ext}° at 37 °C for X = C (orange) and X = caC (blue). Solid lines indicate linear fits.

For both dissociation and association, the enthalpy and entropy are highly pH*-dependent, and therefore simplistic free energy diagrams may be informative for understanding the impact of each type of protonation. Figure 8.12 presents free energy plots along a 1D hybridization coordinate for X = C and X = caC at 37 °C. These three-level profiles are composed of the duplex/single-strand free energy change (left), hybridization transition state (TS, center), and single-strand (right). The single-strand free energy is assumed to be independent of pH* and is set as the reference state. At 37 °C, X = C and X = caC exhibit reductions in $\Delta G_d^{\ddagger}$ of 25 and 20 kJ/mol, respectively, in going from pH* 6.8 to 3.7. $\Delta G_a^{\ddagger}$ is essentially independent of N3 protonation at this temperature. This trend for X = caC is quantitatively consistent with the reduction of ~7 kJ/mol in $\Delta G_d^{\ddagger}$ observed upon protonation of a 5′-GCGATXGATCGC-3′, which contains a third of the modified sites in our model sequence.[75] Across the studied pH* range, linear relationships are observed between $\Delta G_d^{\ddagger}/\Delta G_a^{\ddagger}$ and ΔG_{ext}° for each sequence (Fig. 8.12c). For X = C, a slope of 1.04 is observed when plotting $\Delta G_d^{\ddagger}$ against ΔG_{ext}° and a slope of 0.06 is observed when plotting $\Delta G_a^{\ddagger}$ against

ΔG°_{ext}, indicating that reduction in $\Delta G^{\ddagger}_{a}$ at low pH* is due to destabilization of the duplex relative to the single strand rather than a change in the transition-state free energy. In contrast, X = caC exhibits a slope of 1.77 for $\Delta G^{\ddagger}_{d}$ vs. ΔG°_{ext} and 0.77 for $\Delta G^{\ddagger}_{a}$ vs. ΔG°_{ext}, indicating that both destabilization of the duplex relative to the single strand and a reduction in transition-state free energy occur upon 5-carboxyl protonation.

8.5.3 Proposed mechanistic impact of protonation on DNA hybridization

One can gain further insight into the mechanistic implications of the thermodynamic and kinetic results if we assume that the hybridization transition state involves the formation of some small subset of native contacts followed by rapid downhill formation of the remaining base pairs, in analogy to the classic nucleation-zipper mechanism of helix-to-coil transitions in biomolecules.[58-59] Both sequences exhibit negative values of $\Delta H^{\ddagger}_{a}$ at neutral pH* conditions (Fig. 8.11), which is routinely observed for oligonucleotide hybridization and proposed to arise from the formation of base-pairing and/or stacking interactions in the transition-state ensemble.[55, 58-59, 69] A large entropic penalty due to loss of translational and configuration freedom accompanies this initial base-pair formation, causing the rate of association to decrease as temperature increases.

Interestingly, $\Delta H^{\ddagger}_{a}$ approaches 0 kJ/mol and $\Delta S^{\ddagger}_{a}$ becomes less negative with increasing N3 protonation of the X = C oligonucleotide. Multiple possibilities could account for the reduction in magnitude of $\Delta H^{\ddagger}_{a}$. The X = C single-strand is expected to become highly protonated as pH* decreases, as evidenced by the pH-driven denaturation observed in Fig. 8.5. For GC-rich sequences such as the one studied here, hybridization is assumed to initiate at C:G base pairs, but protonation at N3 interferes with the ability to form C:G contacts. Therefore, the number of contacts in the

transition state may remain unchanged but offer less enthalpic stabilization due to the competition between base-pair hydrogen bonding and protonation at N3. As more N3 sites are protonated, the negative charge density of the single-strand is reduced, and the number of bound sodium cations and excluded chloride and phosphate anions decrease. Analysis of the FTIR titration on X = C suggests that N3 protonation can only occur in the single strand. As a result, the charge density and counterion environment around the protonated single-strand and deprotonated duplex are quite different, and hybridization of the former must involve the loss of N3 protons, gain of associated sodium cations, and exclusion of additional anions. These changes likely contribute to the thermodynamic and kinetic effects observed for N3 protonation in the X = C sequence.

The decrease in the magnitude of $\Delta H_a^{\ddagger}$ and $\Delta S_a^{\ddagger}$ at low pH* could also reflect a reduction in the number of base pairs formed in the hybridization transition-state ensemble. $\Delta H_a^{\ddagger}$ ranges from -35 to 0 kJ/mol over the pH* range studied here. The average enthalpic benefit associated with forming a base pair in the CG region of X = C is estimated to be -21 kJ/mol based on the unified nearest-neighbor parameters for DNA hybridization.[71] Therefore, the change in $\Delta H_a^{\ddagger}$ suggests that the number of base-pair contacts in the transition-state ensemble shifts from ~2 to 0 upon protonation. For the X = C sequence, the reduction of $\Delta H_a^{\ddagger}$ to 0 kJ/mol would suggest that no base pairs are formed in the transition-state ensemble at low pH*. Even in this case, k_a is 100 times slower (~10^7 $M^{-1}s^{-1}$) than the estimated diffusion-limited association rate constant determined from fluorescence correlation spectroscopy (FCS) measurements of short oligonucleotides,[3] indicating that an entropic barrier to duplex formation dictates DNA hybridization kinetics even without the formation of base-pair contacts in the transition state.

While protonation of the X = caC sequence also increases the rate of duplex association, $\Delta H_a^{\ddagger}$ and $\Delta S_a^{\ddagger}$ are both observed to increase in magnitude as pH* is lowered. Again, this could be interpreted as changes to the energetics of initial contact formation or reflect a change in the structure of the hybridization transition state. 5-carboxyl protonation has the potential to impact intra- and inter-strand electrostatics, caC:G base pairing, base stacking, and major groove solvation. Our pH*-dependent FTIR results indicate that caC sites within X = caC are essentially completely deprotonated at pH* 6.8, adding substantial negative charge to the duplex and single-strand states. Protonation of caC should reduce electrostatic repulsion between the exocyclic carboxyl group and the phosphate backbone. However, the absence of negative charge may alter the degree of hydration in the major groove. Our model for internal base pairing suggests that 5-carboxyl protonation increases the enthalpic stabilization of base-pair formation (Fig. 8.11) even though contacts become loosened overall. Such an increase in ΔH_{int}° may explain the increase in magnitude of $\Delta H_a^{\ddagger}$ as pH* decreases, but not the increase in entropic penalty. The change in $\Delta S_a^{\ddagger}$ is more complex as 5-carboxyl protonation can alter factors such as base hydration and single-strand flexibility.[76] Both the increase in magnitude of $\Delta H_a^{\ddagger}$ and $\Delta S_a^{\ddagger}$ and decrease in $\Delta H_a^{\ddagger}$ and $\Delta S_a^{\ddagger}$ could also be explained by an increase in the number of base-pair contacts associated with the hybridization transition state, which could in turn result from the increased structural fluctuation of caC:G base pairs upon 5-carboxyl protonation.[36]

Our results demonstrate that N3 and 5-carboxyl protonation significantly perturb DNA hybridization and dissociation kinetics, but in a nucleobase-specific manner that can be rationalized in terms of duplex destabilization and changes to the transition state for a two-state

process. However, it is important to note that while the nature of the transition state may change, so too may the distribution of transition states. As shown in Fig. 8.9, both sequences exhibit non-exponential relaxation kinetics at low pH* that may stem from inhomogeneity among the duplex ensemble or the transition state. This spread of rates may be modulated by base-pairing fluctuations, solvent and ion interactions, as well as distributions of protonated species.

8.5.4 Apparent cooperative impact of protonated caC sites

5-carboxyl protonation reduces the overall magnitude of the thermodynamic parameters for duplex dissociation in an apparently cooperative manner. As shown in Fig. 8.11, trends in ΔH°_{ext}, ΔS°_{ext}, and ΔG°_{ext} for X = caC follow a Hill profile with n >1, signifying positive cooperativity. Typically, Hill equations are applied to identify cooperativity in binding problems,[77] but here the protonation of X = caC is noncooperative (n = 1, Fig. 8.5). Instead, 5-carboxyl protonation seems to cooperatively shift the thermodynamics of the duplex-to-single-strand transition. Our analysis indicates that nearly all caC sites in the duplex state become protonated from pH* 6.8 to 3.5, and therefore it is possible that multiple protonated caC:G base pairs interact. It seems most likely that such cooperativity would arise from adjacent protonated base pairs, where increased structural fluctuations of each may amplify the weakening of hydrogen bonding and base stacking interactions. Similar effects were predicted for C methylation in which simulations of adjacent methylated CpG dinucleotides influenced base-pair fluctuations in a more than additive manner.[78]

Similar to the shifts in duplex/single-strand equilibrium, 5-carboxyl protonation appears to alter DNA hybridization kinetics in a cooperative manner as both dissociation and association

barriers show sharp transitions as a function of pH* and exhibit a Hill coefficient greater than 1 (Fig. 8.11). Additionally, the pH* trends in the dissociation barrier show a sharper transition (larger Hill coefficient) than observed for the external thermodynamic terms. As shown in the 1D free-energy diagrams (Fig. 8.12), the pH*-dependence of $\Delta G_d^{\ddagger}$ at 37 °C has contributions from duplex destabilization and a reduction in free energy of the transition state as pH* is lowered. The former corresponds to ΔG_{ext}° while the latter is only captured by $\Delta G_d^{\ddagger}$ and $\Delta G_a^{\ddagger}$. Therefore, the more cooperative manner in which $\Delta G_d^{\ddagger}$ varies as a function of pH* relative to ΔG_{ext}° suggests that protonation of multiple 5-carboxyl sites leads to a cooperative change in the free energy of the hybridization transition state in addition to the duplex state. A comparison of $\Delta H_d^{\ddagger}$ with ΔH_{ext}° and $\Delta S_d^{\ddagger}$ with ΔS_{ext}° reveal the same trends in apparent cooperativity with 5-carboxyl protonation.

N3 protonation leads to a larger reduction in DNA duplex stability and dissociation barrier than 5-carboxyl protonation, but it does so almost linearly as a function of pH* (Fig. 8.11) over the range studied. This more gradual reduction reflects a non- or anti-cooperative impact of N3 protonation on the duplex-to-single-strand transition and is consistent with our FTIR titration results that estimate the X = C duplex cannot tolerate N3 protonation, making cooperative interactions between multiple protonated sites in the duplex very unlikely.

8.6 Conclusions

We have investigated the impact of cytosine and 5-carboxylcytosine protonation on the thermodynamics and kinetics of DNA dissociation and hybridization using FTIR and optically-induced temperature-jump IR spectroscopy. Our results demonstrate that between pH* 6.8 and 3.5, DNA containing C and caC are predominantly protonated at the N3 and 5-carboxyl positions,

respectively. Each protonation is shown to have a distinct impact on the thermodynamics and kinetics of dissociation and association between model DNA oligonucleotides. Protonation at N3 completely disrupts the ability of the DNA to duplex, shifts the melting inflection point by >20 °C, and reduces the cooperativity of the duplex-to-single-strand transition. These thermodynamic properties are accompanied by a speedup in the rate of duplex dissociation and increasingly stretched exponential kinetics at low pH*, with the reduction in dissociation barrier primarily achieved through destabilization of the duplex state. 5-carboxyl protonation leads to highly sloped melting baselines that reflect an accumulation of disrupted base-pairing contacts in the duplex state, but perturbations to the duplex/single-strand equilibrium are comparatively minor. Regardless of the identity of X, the dissociation free-energy barrier is reduced to a similar degree over the pH* range studied. However, N3 protonation does reduce the barrier through destabilization of the DNA duplex while 5-carboxyl protonation reduces the dehybridization transition state energy. Unlike with N3 protonation, X = caC duplexes can be highly protonated, and protonation of multiple sites is observed to alter duplex/single-strand thermodynamics and kinetics in a cooperative manner.

Protonation of N3 in cytosine plays critical roles in DNA damage, triplex association, and i-motif formation and is utilized to create pH-driven nanodevices. Additionally, the potential importance of the protonation equilibrium of 5-carboxylcytosine in cytosine demethylation has recently come to light. Each of these functions and applications rely on protonation-driven perturbations to double helical DNA, and this work demonstrates the unique ability of N3 and 5-carboxyl protonation to alter the stability, hybridization kinetics, and base-pairing dynamics of nucleic acids. The loosening of caC:G base pairing and reduction in dissociation barrier initiated

by caC protonation may assist in its selective recognition within the cytosine demethylation pathway.[21-22, 24, 79] Such pH-dependent base-pair loosening as well as the overall reduction of p$K_{a,N3}$ upon substitution of caC for C may also prove useful in tuning the physical properties of DNA triplex and i-motif formation. Recent work has shown that caC can effectively fine tune the thermodynamic stability of both structures,[32-33] and the results and analysis in the present work provide physical insight into the observed thermodynamic response of the DNA triplex and i-motif to the incorporation of caC as well as the prediction of pH-dependent dynamical behavior upon protonation of the 5-carboxyl group.

8.7 Acknowledgements

I thank Paul Sanstead and Sam Peng for initiating our group's work on modified cytosine nucleobases. I am grateful to Paul Sanstead for his mentorship at the time of this project and for numerous helpful discussions regarding the design of the study and interpretation of the data. I thank Qing Dai for helpful discussions and for synthesizing the X = caC sample used in this chapter.

8.8 References

1. Yin, Y.; Zhao, X. S., Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* **2011,** *44*, 1172-1181.

2. Sim, A. Y.; Lipfert, J.; Herschlag, D.; Doniach, S., Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E* **2012,** *86*, 021901.

3. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.

4. Zimmer, C.; Venner, H., Protonation of cytosine in DNA. *Biopolymers* **1966,** *4*, 1073-1079.

5. Zimmer, C.; Luck, G.; Venner, H.; Frič, J., Studies on the conformation of protonated DNA. *Biopolymers* **1968,** *6*, 563-574.

6. Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J., i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018,** *46*, 8038-8056.

7. Mergny, J.-L.; Lacroix, L.; Han, X.; Leroy, J.-L.; Helene, C., Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *J. Am. Chem. Soc.* **1995,** *117*, 8887-8898.

8. Mariani, A.; Bonfio, C.; Johnson, C. M.; Sutherland, J. D., pH-Driven RNA strand separation under prebiotically plausible conditions. *Biochem.* **2018,** *57*, 6382-6386.

9. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **1953,** *171*, 737-738.

10. Franklin, R. E.; Gosling, R. G., Molecular configuration in sodium thymonucleate. *Nature* **1953,** *171*, 740-741.

11. Gulland, J. M.; Jordan, D.; Taylor, H., 213. Deoxypentose nucleic acids. Part II. Electrometric titration of the acidic and the basic groups of the deoxypentose nucleic acid of calf thymus. *J. Chem. Soc.* **1947**, 1131-1141.

12. Courtois, Y.; Fromageot, P.; Guschlbauer, W., Protonated polynucleotide structures: 3. An optical rotatory dispersion study of the protonation of DNA. *Eur. J. Biochem.* **1968,** *6*, 493-501.

13. González‐Olvera, J. C.; Durec, M.; Marek, R.; Fiala, R.; Morales‐García, M. d. R. J.; González‐Jasso, E.; Pless, R. C., Protonation of Nucleobases in Single‐and Double‐Stranded DNA. *ChemBioChem* **2018,** *19*, 2088-2098.

14. Smol'janinova, T.; Zhidkov, V.; Sokolov, G., Analysis of difference spectra of protonated DNA: determination of degree of protonation of nitrogen bases and the fractions of disordered nucleotide pairs. *Nucleic Acids Res.* **1982,** *10*, 2121-2134.

15. Tajmir‐Riahi, H.; Ahmad, R.; Naoui, M.; Diamantoglou, S., The effect of HCl on the solution structure of calf thymus DNA: a comparative study of DNA denaturation by proton and metal cations using Fourier transform IR difference spectroscopy. *Biopolymers* **1995,** *35*, 493-501.

16. Nikolova, E. N.; Goh, G. B.; Brooks III, C. L.; Al-Hashimi, H. M., Characterizing the protonation state of cytosine in transient G· C Hoogsteen base pairs in duplex DNA. *J. Am. Chem. Soc.* **2013,** *135*, 6766-6769.

17. Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D., I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018,** *10*, 631-637.

18. Privat, E. J.; Sowers, L. C., A proposed mechanism for the mutagenicity of 5-formyluracil. *Mutat. Res. Fundam. Mol.* **1996,** *354*, 151-156.

19. Driggers, P.; Beattie, K., Effect of pH on the base-mispairing properties of 5-bromouracil during DNA synthesis. *Biochem.* **1988,** *27*, 1729-1735.

20. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011,** *333*, 1300-1303.

21. Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.-L.; Luo, C.; Jiang, H.; He, C., Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **2012,** *8*, 328-330.

22. Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C., Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **2013,** *135*, 15813-15822.

23. La Francois, C. J.; Jang, Y. H.; Cagin, T.; Goddard, W. A.; Sowers, L. C., Conformation and proton configuration of pyrimidine deoxynucleoside oxidation damage products in water. *Chem. Res. Toxicol.* **2000,** *13*, 462-470.

24. Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X., Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012,** *40*, 10203-10214.

25. Idili, A.; Vallée-Bélisle, A.; Ricci, F., Programmable pH-triggered DNA nanoswitches. *J. Am. Chem. Soc.* **2014,** *136*, 5836-5839.

26. Dong, Y.; Yang, Z.; Liu, D., DNA nanotechnology based on i-motif structures. *Acc. Chem. Res.* **2014,** *47*, 1853-1860.

27. Porchetta, A.; Idili, A.; Vallée-Bélisle, A.; Ricci, F., General strategy to introduce pH-induced allostery in DNA-based receptors to achieve controlled release of ligands. *Nano Lett.* **2015,** *15*, 4467-4471.

28. Amodio, A.; Zhao, B.; Porchetta, A.; Idili, A.; Castronovo, M.; Fan, C.; Ricci, F., Rational design of pH-controlled DNA strand displacement. *J. Am. Chem. Soc.* **2014,** *136*, 16469-16472.

29. Modi, S.; Nizak, C.; Surana, S.; Halder, S.; Krishnan, Y., Two DNA nanomachines map pH changes along intersecting endocytic pathways inside the same cell. *Nat. Nanotechnol.* **2013,** *8*, 459-467.

30. Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H., Tuning the pH Response of i-Motif DNA Oligonucleotides. *ChemBioChem* **2015,** *16*, 1647-1656.

31. Zamiri, B.; Mirceta, M.; Bomsztyk, K.; Macgregor Jr, R. B.; Pearson, C. E., Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC) 8•(GGCCCC) 8 repeat: effect of CpG methylation. *Nucleic Acids Res.* **2015,** *43*, 10055-10064.

32. Wright, E. P.; Abdelhamid, M. A.; Ehiabor, M. O.; Grigg, M. C.; Irving, K.; Smith, N. M.; Waller, Z. A. E., Epigenetic modification of cytosines fine tunes the stability of i-motif DNA. *Nucleic Acids Res.* **2020,** *48*, 55-62.

33. Sumino, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M., Synthesis and properties of oligodeoxynucleotides containing 5-carboxy-2′-deoxycytidines. *Bioorganic & medicinal chemistry letters* **2008,** *18*, 274-277.

34. Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T., Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chem. Eur. J.* **2011,** *17*, 13782-13788.

35. Szulik, M. W.; Pallan, P. S.; Nocek, B.; Voehler, M.; Banerjee, S.; Brooks, S.; Joachimiak, A.; Egli, M.; Eichman, B. F.; Stone, M. P., Differential stabilities and sequence-dependent base pair opening dynamics of Watson–Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochem.* **2015,** *54*, 1294-1305.

36. Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016,** *7*, 10813.

37. Hansch, C.; Leo, A.; Taft, R., A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991,** *91*, 165-195.

38. Dai, Q.; Sanstead, P. J.; Peng, C. S.; Han, D.; He, C.; Tokmakoff, A., Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chem. Biol.* **2016,** *11*, 470-477.

39. Banyay, M.; Sarkar, M.; Gräslund, A., A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **2003,** *104*, 477-488.

40. Peng, C. S. Two-dimensional infrared spectroscopy of nucleic acids: application to tautomerism and DNA aptamer unfolding dynamics. Massachusetts Institute of Technology, 2014.

41. Stelling, A. L.; Xu, Y.; Zhou, H.; Choi, S. H.; Clay, M. C.; Merriman, D. K.; Al‐Hashimi, H. M., Robust IR‐based detection of stable and fractionally populated G‐C+ and A‐T Hoogsteen base pairs in duplex DNA. *FEBS Lett.* **2017,** *591*, 1770-1784.

42. Krężel, A.; Bal, W., A formula for correlating pKa values determined in D2O and H2O. *J. Inorg. Biochem.* **2004,** *98*, 161-166.

43. Peng, C. S.; Jones, K. C.; Tokmakoff, A., Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 15650-15660.

44. Lee, C.; Cho, M., Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **2006,** *125*, 114509.

45. Benoit, R. L.; Fréchette, M., Protonation of hypoxanthine, guanine, xanthine, and caffeine. *Can. J. Chem.* **1985,** *63*, 3053-3056.

46. Major, D. T.; Laxer, A.; Fischer, B., Protonation studies of modified adenine and adenine nucleotides by theoretical calculations and 15N NMR. *J. Org. Chem.* **2002,** *67*, 790-802.

47. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

48. Krummel, A. T.; Zanni, M. T., DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: Insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* **2006,** *110*, 13991-14000.

49. Ashwood, B.; Sanstead, P. J.; Dai, Q.; He, C.; Tokmakoff, A., 5-Carboxylcytosine and cytosine protonation distinctly alter the stability and dehybridization dynamics of the DNA duplex. *J. Phys. Chem. B* **2019,** *124*, 627-640.

50. Siegfried, N. A.; O'Hare, B.; Bevilacqua, P. C., Driving forces for nucleic acid p K a shifting in an A+· C wobble: Effects of helix position, temperature, and ionic strength. *Biochem.* **2010,** *49*, 3225-3236.

51. Crothers, D. M., Statistical thermodynamics of nucleic acid melting transitions with coupled binding equilibria. *Biopolymers* **1971,** *10*, 2147-2160.

52. Amunson, K. E.; Anderson, B. A.; Kubelka, J., Temperature effects on the optical path length of infrared liquid transmission cells. *Appl. Spectrosc.* **2011,** *65*, 1307-1313.

53. Mergny, J.-L.; Lacroix, L., Analysis of thermal melting curves. *Oligonucleotides* **2003,** *13*, 515-537.

54. Jones, K. C.; Ganim, Z.; Tokmakoff, A., Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 14060-14066.

55. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

56. Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Minnes, L.; Towrie, M.; Parker, A. W.; Hunt, N. T., Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 6188-6199.

57. Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A., Oxidized derivatives of 5-methylcytosine alter the stability and dehybridization dynamics of duplex DNA. *J. Phys. Chem. B* **2020,** *124*, 1160-1174.

58. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.

59. Pörschke, D.; Eigen, M., Co-operative non-enzymatic base recognition III. Kinetics of the helix—coil transition of the oligoribouridylic· oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **1971,** *62*, 361-381.

60. Owczarzy, R., Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys. Chem.* **2005,** *117*, 207-215.

61. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.

62. Bernasconi, C., *Relaxation kinetics*. Academic Press: 1976.

63. Kumar, A. T.; Zhu, L.; Christian, J.; Demidov, A. A.; Champion, P. M., On the rate distribution analysis of kinetic data using the maximum entropy method: Applications to myoglobin relaxation on the nanosecond and femtosecond timescales. *J. Phys. Chem. B* **2001,** *105*, 7847-7856.

64. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7*, 284-304.

65. Cho, C.; Urquidi, J.; Singh, S.; Robinson, G. W., Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **1999,** *103*, 1991-1994.

66. Chung, H. S.; Tokmakoff, A., Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy. *Proteins* **2008,** *72*, 474-487.

67. Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M., The folding mechanism of a β-sheet: The WW domain. *J. Mol. Biol.* **2001,** *311*, 373-393.

68. Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M., Mapping the transition state of the WW domain β-sheet. *J. Mol. Biol.* **2000,** *298*, 283-292.

69. Chen, C.; Wang, W.; Wang, Z.; Wei, F.; Zhao, X. S., Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* **2007,** *35*, 2875-2884.

70. Hill, A. V., The combinations of haemoglobin with oxygen and with carbon monoxide. I. *Biohcem. J.* **1913,** *7*, 471.

71. SantaLucia Jr, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 1460-1465.

72. Xia, T.; SantaLucia Jr, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H., Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson− Crick base pairs. *Biochem.* **1998,** *37*, 14719-14735.

73. Hardwick, J. S.; Lane, A. N.; Brown, T., Epigenetic modifications of cytosine: biophysical properties, regulation, and function in mammalian DNA. *BioEssays* **2018,** *40*, 1700199.

74. Drohat, A. C.; Coey, C. T., Role of base excision "repair" enzymes in erasing epigenetic marks from DNA. *Chem. Rev.* **2016,** *116*, 12711-12729.

75. Dubini, R. C.; Korytiaková, E.; Schinkel, T.; Heinrichs, P.; Carell, T.; Rovó, P., 1H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives. *ACS Phys. Chem. Au* **2022,** *2*, 237-246.

76. Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L., Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 799-804.

77. Weiss, J. N., The Hill equation revisited: uses and misuses. *FASEB J.* **1997,** *11*, 835-841.

78. Teng, X.; Hwang, W., Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.* **2018,** *114*, 1791-1803.

79. Pidugu, L. S.; Dai, Q.; Malik, S. S.; Pozharski, E.; Drohat, A. C., Excision of 5-carboxylcytosine by thymine DNA glycosylase. *J. Am. Chem. Soc.* **2019,** *141*, 18851-18861.

Appendix 8.A Degree of protonation in X = C and X = caC duplexes

The degree of 5caC protonation in X = caC can be directly determined from the 2nd SVD component across pH* shown in Fig. 8.5. The 2nd SVD component is well fit to the Henderson-Hasselbach equation. At 3 °C, the duplex fraction (θ_{ext}) is assumed to be essentially independent of pH*, and therefore the fit 2nd SVD component describes the degree of protonation in the duplex state. The number of protonated 5caC sites per duplex (χ) is given by eq. 8.A1:

$$\chi = \frac{[HA]\theta_{ext}}{[D]} \tag{8.A1}$$

where [HA] and [D] are the concentration of protonated 5caC sites and duplex, respectively. There are three 5caC bases per DNA strand, so the total concentration of sites is three times greater than the oligonucleotide concentration (c_{tot}):

$$[HA]+[A^-] = 3c_{tot} \tag{8.A2}$$

Rearranging the Henderson-Hasselbach equation gives an expression for [HA]:

$$[HA] = \frac{[A^-]}{10^{pH-pK_a}} \tag{8.A3}$$

which can be expressed in terms of c_{tot} using 8.A3:

$$[HA] = \frac{3c_{tot}}{1+10^{pH-pK_a}} \tag{8.A4}$$

The DNA duplex concentration ([D]) can be cast in terms of c_{tot} and θ_{ext}:

$$[D] = \frac{\theta_{ext} c_{tot}}{2} \tag{8.A5}$$

Then, eqs. 8.A2, 8.A4, and 8.A5 can be used to determine χ for X = caC:

$$\chi = \frac{6}{1+10^{pH-pK_a}} \tag{8.A6}$$

The plot of χ for X = caC is shown in Fig. 8.5. Over the pH* range studied here, X = caC varies from fully deprotonated at pH* 6.8 to near complete protonation of all six sites at pH* 3.5.

Unlike X = caC, θ_{ext} of X = C is highly sensitive to pH* and requires a modified treatment to determine the degree of duplex protonation. An FTIR titration of X = C (Fig. 8.5) was performed, where the 2nd SVD component is assumed to primarily report on the pH*-induced duplex-to-single-strand transition. The duplex-to-single-strand transition and cytosine N3 protonation appear to occur over the same pH* range, indicating that pH_m, the pH at which θ_{ext} = 0.5 at 3 °C, is equivalent to the apparent pK_a of N3 protonation. Therefore, the duplex fraction can be explicitly related to degree of N3 protonation in the single-strand and duplex at a specific pH*. The percentage of protonated N3 sites is given by eq. 8.A7:

$$\theta_p = \frac{[HA]}{[HA]+[A^-]} = \frac{1}{1+10^{pH-pK_a}} \tag{8.A7}$$

Here we define [HA] in terms of single-strand ([S]) and duplex ([D]) concentrations:

$$[HA] = c_S[S] + c_D[D] \tag{8.A8}$$

where c_S and c_D represent the average number of protonated N3 sites in the single-strand and duplex states, respectively. By defining [S] and [D] in terms of c_{tot} and $\theta_{ext,}$ and using eqs. 8.A2 and 8.A7.

$$\theta_p = \frac{1}{3}\left[c_S(1-\theta_{ext}) + \frac{c_D\theta_{ext}}{2} \right] \tag{8.A9}$$

By re-arranging this expression for θ_p and substituting in eq. 8.A7, an expression relating θ_{ext} to single-strand and duplex protonation is given:

$$\theta_{ext} = \frac{\frac{3}{10^{pH-pK_a}+1} - c_S}{0.5c_D - c_S} \tag{8.A10}$$

Assuming that the pH*-induced duplex-to-single-strand transition follows the N3 titration curve, eq. 8.A10 describes the relation between θ_{ext} and N3 protonation at low temperature. Equation 8.A10 can be used to fit the 2nd SVD component from the FTIR titration of X = C, where pK_a, $c_{S,}$ and c_D are used as fit parameters. The 2nd SVD component is well fit (Fig. 8.5) only when c_S = 2.92 and c_D = 0.02, indicating that the X = C duplex cannot tolerate any degree of N3 protonation. This suggests that the thermodynamic and kinetic consequences of N3 protonation purely arise from shifting in the duplex-to-single-strand transition equilibrium at low pH*.

Thermodynamics and kinetics of dinucleotide hybridization to gaps and overhangs

The material in this chapter is adapted from:

Ashwood, B.; Jones, M. S.; Radakovic, A.; Khanna, S.; Lee, Y.; Sachleben, J. R.; Szostak, J. W.; Ferguson, A. L.; Tokmakoff, A., Thermodynamics and kinetics of DNA and RNA dinucleotide hybridization to gaps and overhangs. *Biophys. J.* **2023,** *122*, 3323-3339.

9.1 Introduction

DNA and RNA duplex hybridization has been investigated for more than sixty years with the aim of developing predictive models for how thermodynamic and kinetic properties will vary with molecular factors such as strand length, sequence, and chemical modifications.[1-3] The development of numerous models, such as the quantitatively accurate nearest-neighbor (NN) models[8-9] for thermodynamics or the kinetic-zipper model,[11-12] shape the modern understanding of nucleic acid properties as well as the development of nucleic acid technology.[13-14] However, most of these models were developed for oligonucleotides, and the hybridization properties of nucleic acid segments shorter than six nucleotides in length have largely been neglected. Hybridization of such short segments is often perceived as irrelevant due to their poor binding stability in aqueous solution, yet there are biological processes where segments as short as mononucleotides must hybridize. Many of these processes, such as the binding of 2′-deoxynucleotide triphosphates to single-strand DNA (ssDNA) during strand replication or binding between three nucleotide anticodons of tRNA and their complement mRNA, utilize additional protein-nucleic acid interactions to promote binding.[15-16] However, processes such as non-enzymatic replication of

nucleic acids and toehold-mediated strand displacement require hybridization of 1-4 base-pair patches without the help of proteins.[13, 17-19]

A conserved feature among the examples of short oligonucleotide hybridization is that a short segment binds to a single-strand overhang adjacent to a duplex region or in a gap between two duplex segments. The adjacent duplexes offer coaxial stacking interactions that stabilize the hybridization of the short oligonucleotide. The thermodynamic benefit of coaxial stacking has been measured for DNA and RNA oligonucleotides and is often found to be similar to stacking in B-DNA or A-RNA, respectively.[20-25] The structural constraints of the sugar-phosphate backbone at a nick site are relaxed relative to a covalently-linked base-pair step and may enable more stabilizing stacking configurations.[26-27] Additionally, the single-strand segments in overhang and gap templates may have different physical properties from free single-strands such as reduced configurational flexibility, increased stacking, and different hydration that influence the stability and dynamics of binding with short oligonucleotides. A few studies have shown evidence for increased rigidity of gap and overhang single-stranded regions relative to free single-strands as well as bending of duplexes containing 1 or 2 nucleotide gaps,[28-34] and these properties are likely highly sensitive to nucleobase sequence and cation concentrations.[35-37]

The binding stability of mononucleotides with overhang and gap templates has recently been studied. NMR-monitored titrations of mononucleotides with overhangs revealed dissociation constants (K_d) ranging from ~10 mM for G:C base pairing to ~200 mM for A:T base pairing.[38-39] Isothermal titration calorimetry (ITC) and FRET-monitored titrations have shown that the K_d for binding onto a gap decreases nearly 3 orders of magnitude to values of 0.2-0.6 mM,[40-41] which

corresponds to a ~12 kJ mol^{-1} increase in the dissociation free energy (ΔG_d°) and is consistent with forming an additional nearest-neighbor step.[8, 42] These studies give a sense for the stability of short oligonucleotide hybridization to overhangs and gaps and indicate that it depends on sequence and whether DNA or RNA is used, yet, with the exception of ITC, the thermodynamic information is limited to dissociation constants (K_d or ΔG_d°) at the measurement temperature. A detailed understanding of hybridization onto overhangs and gaps requires knowledge of enthalpic and entropic contributions to stability and the role of relevant environmental variables such as temperature and counterion concentration.[43-45] More importantly, no studies have directly monitored the time-dependence of hybridization and dehybridization for short oligonucleotides onto templates, and these timescales determine whether reactions such as non-enzymatic extension or toehold-mediated strand displacement are possible.

Here we report the temperature-dependent thermodynamics and kinetics of short oligonucleotide dehybridization from overhangs and gaps using steady-state and temperature-jump infrared (T-jump IR) spectroscopy. IR spectroscopy is sensitive to base pairing and stacking interactions and distinctly resolves changes in A:T and G:C base pairing without the need for synthetic labels.[46-47] Characterization of temperature-dependent binding stability enables accurate determination of enthalpic and entropic contributions and is particularly relevant in non-enzymatic replication where temperature variations may play a key role.[43-44] We focus on the binding of DNA and RNA dinucleotides and demonstrate our approach with a model system of an adenine-adenine (AA) dinucleotide binding next to pure G:C duplex regions. This model system was chosen to maximize spectral contrast between binding of the dinucleotide and the structural changes and

dynamics within the template itself, but our approach is applicable for any nucleobase sequence. We find quantitative agreement between thermodynamic results obtained from IR spectroscopy and those from complementary temperature-dependent ^{1}H NMR and ITC measurements of binding. From T-jump IR measurements, we extract a time constant ranging from 200 ns to 40 μs for dissociation of AA from the template that depends on the temperature and template, while the time constant for association is a few microseconds and shows a minor dependence on temperature. We also study the binding of 3-mer (AAA) and 4-mer (AAAA) adenine DNA oligonucleotides to assess the length-dependence of association in this short oligonucleotide regime. Our experimental results as well as all-atom molecular dynamics (MD) simulations suggest that differences in the structure and dynamics of the bound complex and template compared to free oligonucleotides lead to significantly different binding thermodynamics and kinetics for overhangs and gaps.

9.2 Measurements of dinucleotide dehybridization

9.2.1 Low-temperature hybridization of dinucleotides to overhangs and gaps

Figure 9.1a illustrates the model sequences that we designed to study the hybridization of a 2′-deoxyadenosine dinucleotide onto G:C rich DNA templates incorporating a single-stranded thymine overhang or thymine gap. Overhang templates (TTo) are composed of a 6-mer primer strand bound to the 3′ end of a 14-mer template strand. Gap templates also have a 6-mer helper strand bound to the 5′-end of the template. Additionally, we studied the binding of adenosine dinucleotide onto a single-stranded uracil RNA gap template (UUg). The abbreviation AA will be used to refer to both the 2′-deoxyadenosine dinucleotide for DNA and adenosine dinucleotide for RNA samples throughout. The primer and helper segments consist of six guanine:cytosine (G:C)

base pairs in order to maximize their binding stability while also keeping their size relatively small to minimize IR absorption from the template itself. The hybridized product between the template and AA is termed the AA-overhang (TTo:AA) or AA-gap complex (TTg:AA, UUg:AA). Although the formation of a fully formed gap or overhang complex can potentially involve numerous binding equilibria, the choice of a weakly bound AA to a far more stable GC-rich template means that the dissociation of the complex is well described through two equilibrium constants: K_d for the dissociation of AA from the template and $K_{d,Temp}$ for dissociation of the template.

We first test for evidence that AA binds to gaps and overhangs at low temperature (1 °C) using FTIR spectroscopy. Figure 9.1 compares the FTIR spectra in the 1540 – 1720 cm^{-1} region of isolated TTg, TTo, and UUg templates with samples containing an equimolar mixture of template and AA. This spectral region contains in-plane carbonyl and ring stretching vibrations of the nucleobases that are sensitive to base-pairing and stacking interactions and have previously been studied in detail.[46-49] The templates are fully intact at 1 °C such that spectral changes from adding AA report on the binding of AA to the template. Each mixed sample exhibits an increase in absorbance near 1575 and 1625 cm^{-1} resulting from the A ring vibrations of AA. The 1650 – 1700 cm^{-1} region primarily contains carbonyl vibrations. The shifting of intensity from 1665 to 1685 cm^{-1} indicates increased stacking and base pairing of guanines adjacent to the gap, and the overlapping gain in intensity of the thymine carbonyl band near 1695 cm^{-1} indicates the formation of A:T base pairing.[46] The intensity of the 1625 cm^{-1} band directly reports on the stacking and base-pairing interactions experienced by AA, and its suppression in TTg:AA and UUg:AA relative to TTo:AA indicates greater binding of AA in the gap at 1°C. Further, FTIR-monitored titrations

of AA with each template indicate an increase in fraction of TTg bound to AA at equimolar conditions ($\theta_{1:1}$) progressing from TTo:AA to UUg:AA to TTg:AA (Figs. 9.1e & 9.A1). K_d decreases by over an order of magnitude in switching the template from an overhang to a gap as previously observed for binding of guanosine monophosphate (GMP) to RNA.[40]

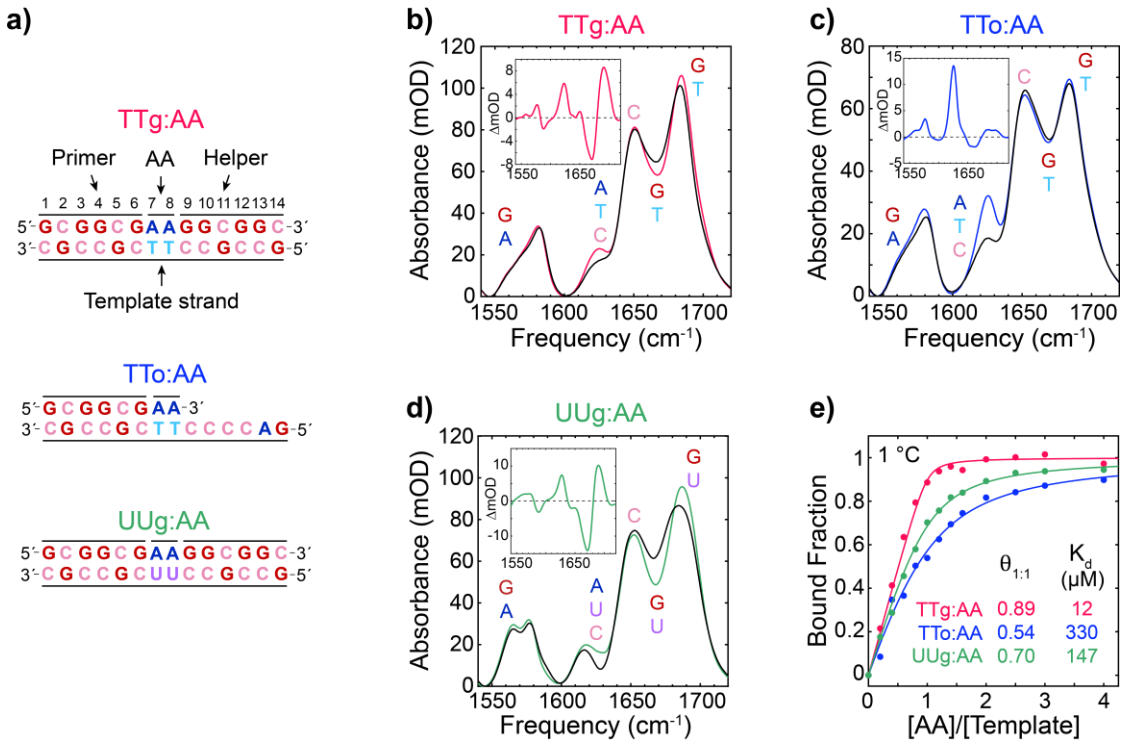

Figure 9.1 Low-temperature hybridization of AA dinucleotide to gap and overhang. (a) Model sequences studied in this work. Gap templates are comprised of a 14-mer template strand bound to 6-mer primer and 6-mer helper strands whereas the overhang templates are only bound to the primer strand. **(b, c, d)** FTIR spectra at 1 °C of 1 mM gap and overhang templates in pH* 6.8 400 mM sodium phosphate buffer (SPB) without AA (black lines) and after adding 1 mM AA (color lines). Difference spectra upon addition of AA are shown as insets. Labels indicate the nucleobases contributing to each frequency region of the FTIR spectra. **(e)** Titrations of AA against 1 mM TTg, TTo, and UUg templates at 1 °C monitored with FTIR. The fraction of template bound to AA is determined from fitting the second singular value decomposition (SVD) component of FTIR data between 1650 – 1720 cm^{-1} to a two-state binding model (solid lines, eq. 9.A2). The fraction of AA bound to template at a 1:1 molar AA:template ratio ($\theta_{1:1}$) and the dissociation constant (K_d) are listed for each complex. Additional spectra and fitting details are provided in Appendix 9.A.

9.2.2 AA binding equilibrium monitored with temperature-dependent IR and NMR spectroscopy

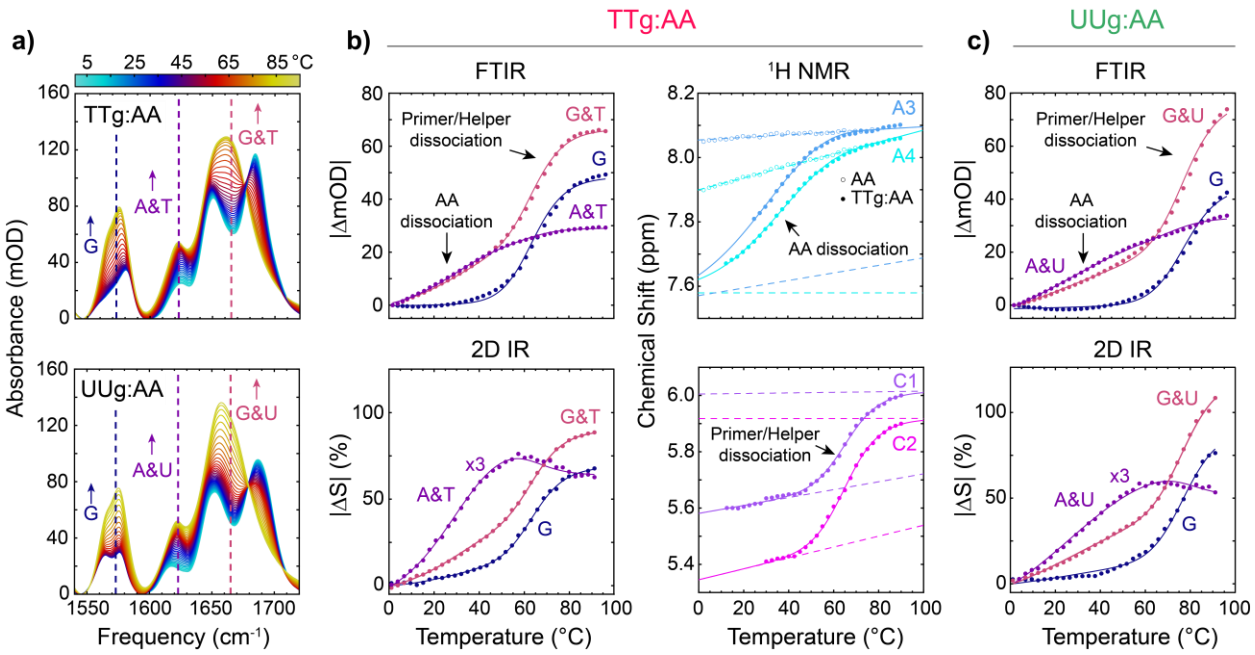

Figure 9.2 Temperature-induced dissociation of AA. (a) FTIR spectra of TTg:AA and UUg:AA from 1 to 96 °C in ~2.6 °C steps. Solutions contain 1 mM each of AA, primer, helper, and template strands. **(b)** Thermal AA and template dissociation of the TTg:AA complex monitored with FTIR, 2D IR, and ^{1}H NMR. FTIR traces are plotted as the absolute value change in absorption relative to 1 °C for select frequencies marked in **(a)** that primarily report on intensity changes in G (1573 cm^{-1}, blue), A & T (1623 cm^{-1}, purple), and G & T (1665 cm^{-1}, pink). 2D IR traces at select frequencies similar to the FTIR data are plotted as the absolute value percent change in signal relative to 1 °C (|ΔS|). From ^{1}H NMR, the AA dissociation transition is monitored using the chemical shift of two aromatic protons from adenine (A3 and A4), and H5 aromatic protons of cytosine (C1 and C2). The temperature-dependent chemical shifts of A3 and A4 in free AA are shown as open circles. The corresponding ^{1}H NMR spectra are shown in Fig. 9.5. **(c)** Thermal AA and template dissociation transitions of the UUg:AA complex monitored with (top) FTIR and (bottom) 2D IR. Solid lines for each dataset correspond to fits to a three-state sequential model (Appendix 9.B). FTIR and 2D IR data are globally fit together. The four ^{1}H NMR peaks are globally fit to the same three-state model and include upper and lower baselines (dashed lines).

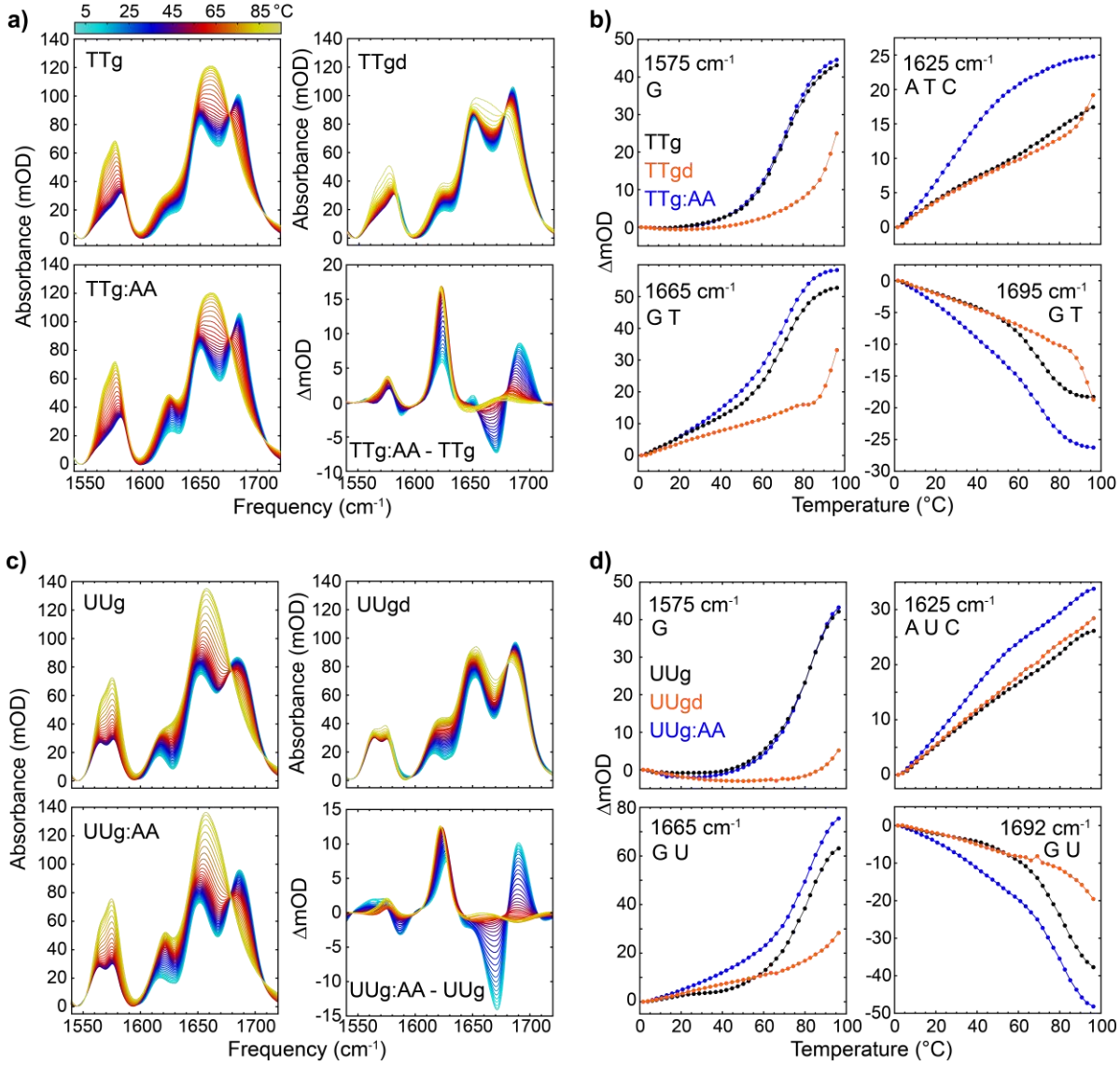

Figure 9.3 Temperature-dependent FTIR spectra of DNA and RNA gap templates. (a) FTIR temperature series for TTg, TTg:AA, and a fully complementary duplex 5′-GCGGCGAAGGCGGC-3′/5′-GCCGCCTTCGCCGC-3′ (TTgd) from 1 to 96 °C. The bottom-right panel shows the differences spectra between TTg:AA and TTg at each temperature. **(b)** Change in absorbance as a function of temperature at select FTIR frequencies for TTg (black), TTg:AA (blue), and TTgd (orange). **(c-d)** Similar plots for RNA gap sequences.

To accurately measure temperature-dependent AA binding stability, IR and NMR spectroscopic probes are employed to monitor changes in base pairing as a function of temperature that are then modeled to extract melting curves for AA dissociation. FTIR temperature series were performed to track AA binding stability from 1 to 96 °C (Fig. 9.2a,b). Both AA-gap complexes show two melting transitions that correspond to dissociation of AA from the template followed by dissociation of the primer and helper from the template strand. The AA dissociation transition in TTg:AA and UUg:AA is observed between 1 and 60 °C as an increase in intensity of the adenine ring mode at 1625 cm^{-1} and the change of carbonyl bands at 1685-1695 and 1665 cm^{-1}. The sharper dissociation transition of the primer and helper – characterized exclusively by changes in the guanine ring and carbonyl bands – occurs at higher temperature as a single step melting transition since the binding stability of the primer and helper to the template strand are nearly identical. The midpoint for primer and helper dissociation is shifted 13 °C higher in UUg:AA due to the more stabilizing nature of GC base-pair steps in A-RNA relative to B-DNA.[8-9] AA unbinding in TTo:AA is already halfway complete at 1 °C as indicated in Fig. 9.1e and is completed by ~40 °C.[5]

The sigmoidal character of the AA thermal dissociation transition is superimposed on other slowly varying temperature-dependent changes, primarily a result of solvation. We account for additional independent spectral changes by measuring FTIR temperature series of TTg, TTo, and UUg and fully complementary duplexes (TTgd, UUgd) for reference, as shown in Fig. 9.3 and discussed previously.[5] It is clear that linear background changes contribute heavily to AA dissociation curve shape at 1625 nm^{-1}, particularly for UUgd. In contrast, we find that these linear background changes are only minor contributions in temperature series monitored by two-dimensional IR (2D IR) spectroscopy (Fig. 9.4). The 2D IR temperature series show profiles for

AA dissociation and primer and helper dissociation transitions that correlate closely the transitions observed in FTIR data but with reduced linear baseline contribution to the sigmoidal transitions. The relative reduction of the linear baseline primarily arises from the fourth-order proportionality of the 2D IR signal to the vibrational transition dipole moment (μ^4) as introduced in Chapter 2.

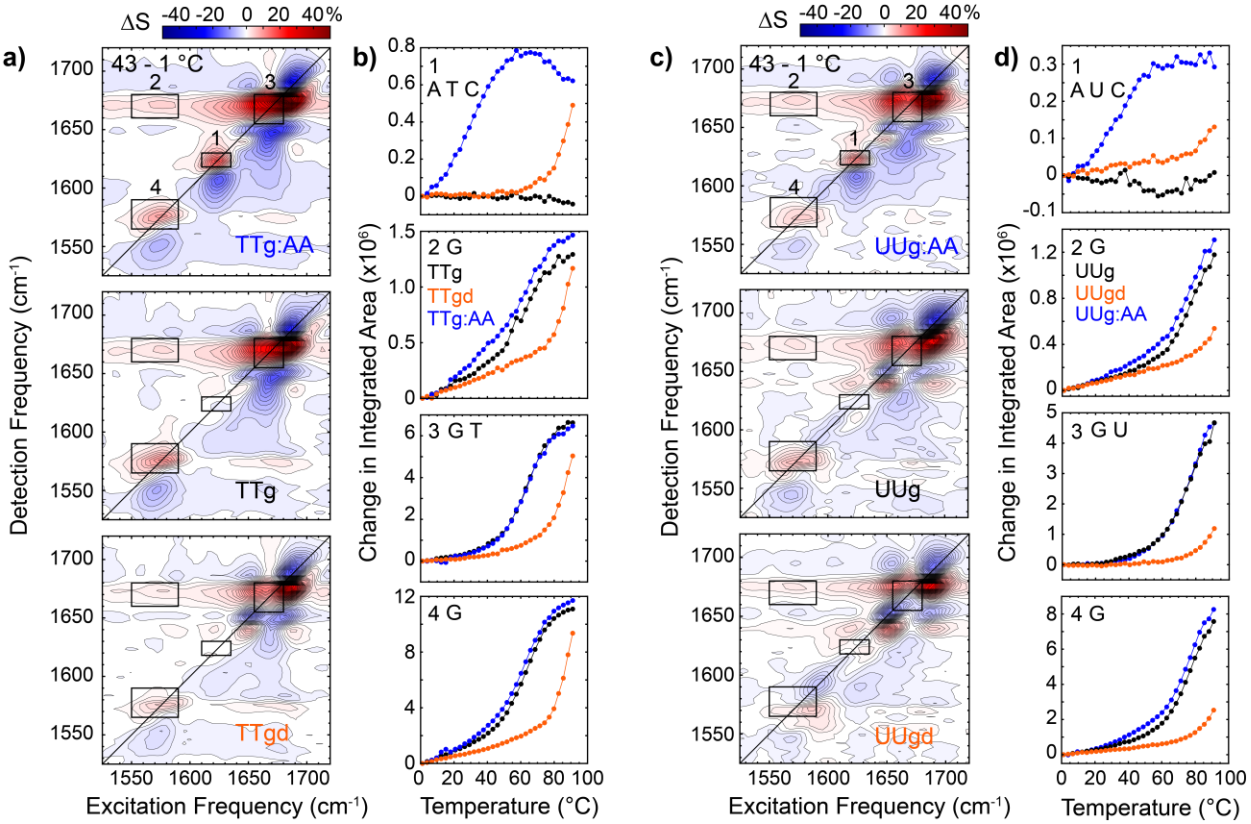

Figure 9.4 Temperature-dependent 2D IR spectral changes of DNA and RNA gap templates. **(a)** 2D IR difference spectra between 43 and 1 °C for TTg:AA, TTg, and TTgd. **(b)** Temperature-dependent integrated amplitudes over the regions marked in **(a)** plotted relative to the spectrum at 1 °C. **(c-d)** Corresponding temperature-dependent 2D IR data for UUg:AA, UUg, and UUgd.

We also obtain similar AA dissociation and primer and helper dissociation transitions from ^{1}H NMR of TTg:AA (Figs. 9.2b and 9.5), lending confidence to the profiles of the IR-monitored dissociation transitions. ^{1}H NMR temperature series have previously been used to extract nucleic

acid binding thermodynamics and contain base-specific information in short oligonucleotides.[50] We performed 1H NMR temperature series using a Bruker AVANCE III 500 MHz spectrometer. Temperature series were performed in 2.5 °C steps and the sample was equilibrated for 5 min and auto gradient and lock shimmed with TopShim at each temperature. prior to acquiring spectra. Numerous resonances are sensitive to dissociation of AA and/or primer and helper dissociation, and a few signals may be used to selectively track each process. The H2 and H8 protons of adenine specifically probe the AA dissociation equilibrium while H5 and H6 protons of cytosine will primarily be sensitive to primer and helper dissociation. Many of the peaks are broad and overlap at low temperature, likely due to slow tumbling of the AA-gap complex, and become more narrow and separated as the temperature increases. The adenine protons of AA exhibit a ~0.4 ppm (200 Hz) frequency difference between the bound and dissociated configurations. As discussed in Section 9.3.2, the exchange rate for AA binding and unbinding is ~50,000 Hz at 15 °C, therefore the four aromatic adenine protons from 7.6 to 8.4 ppm show a single set of peaks whose chemical shift report on the relative population of bound and unbound states (Fig. 9.5). From 10 to 50 °C, identification of the aromatic adenine peaks is hindered by overlap with resonances from aromatic protons on guanine, thymine, and cytosine, but the chemical shift of two adenine peaks (labeled A3 and A4) can be determined over most of the temperature range through comparison with temperature-dependent spectra of free AA and TTg. The temperature-dependence of A3 and A4 chemical shift follows a single sigmoidal transition that reports on the AA binding equilibrium and exhibits a midpoint and width similar to the IR-monitored transition (Fig. 9.2). Due to the congestion of aromatic peaks from 7.6 to 8.4 ppm, we use the H5 cytosine protons located from 5.2 to 6.2 ppm to monitor the primer and helper binding equilibrium. Two H5 cytosine signals (C1

and C2), that appear insensitive to AA dissociation, are distinguished from H1′ peaks of the deoxyribose groups by TOCSY cross-peaks to H6 cytosine signals[5] and exhibit sigmoidal temperature-dependent trends in chemical shift that report on primer and helper dissociation.

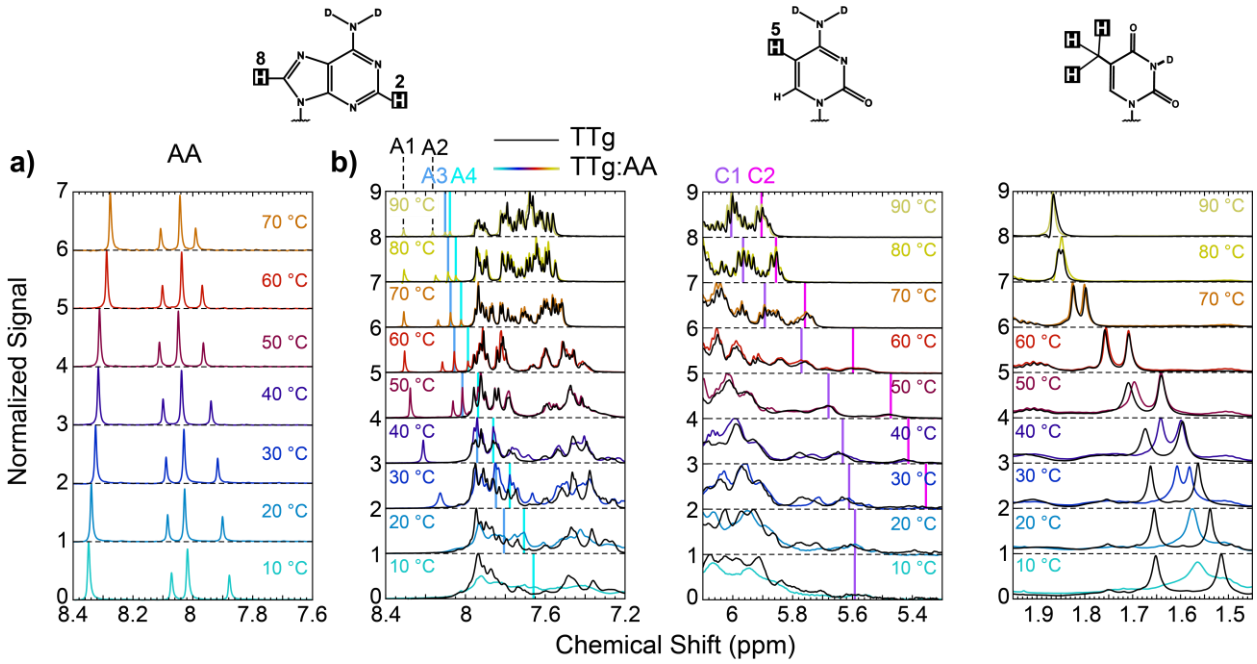

Figure 9.5 Temperature-dependent ^{1}H NMR spectra of TTg:AA. (a) Normalized aromatic proton spectra of AA. **(b)** Normalized ^{1}H NMR spectra of TTg:AA from 10 °C to 90 °C. (Left) Frequency window from 8.4 to 7.6 ppm that contains H8/H2 signals of adenine, H8 signals of guanine, and H6 signals of thymine and cytosine. Each adenine proton signal is labeled A1-A4 in the 90 °C spectrum, and the A3 and A4 chemical shift values are marked at each temperature with dark and light blue vertical lines, respectively. The spectra of TTg are shown in black and those of TTg:AA are colored. A3 and A4 chemical shifts are indicated with vertical lines. (Center) Frequency window from 6.2 to 5.2 ppm that contains H1′ signals from the 2′-deoxyribose moiety of each nucleotide and H5 signals from cytosine. Two cytosine H5 signals are denoted C1 and C2 and marked with purple and magenta vertical lines, respectively, at each temperature. (Right) Frequency window from 1.45 to 1.95 ppm window showing the methyl protons of thymine. All oligonucleotides are present at 1 mM concentration in pH* 6.8 400 mM SPB in D_2O.

9.2.3 Temperature-jump IR spectroscopy of AA dissociation kinetics

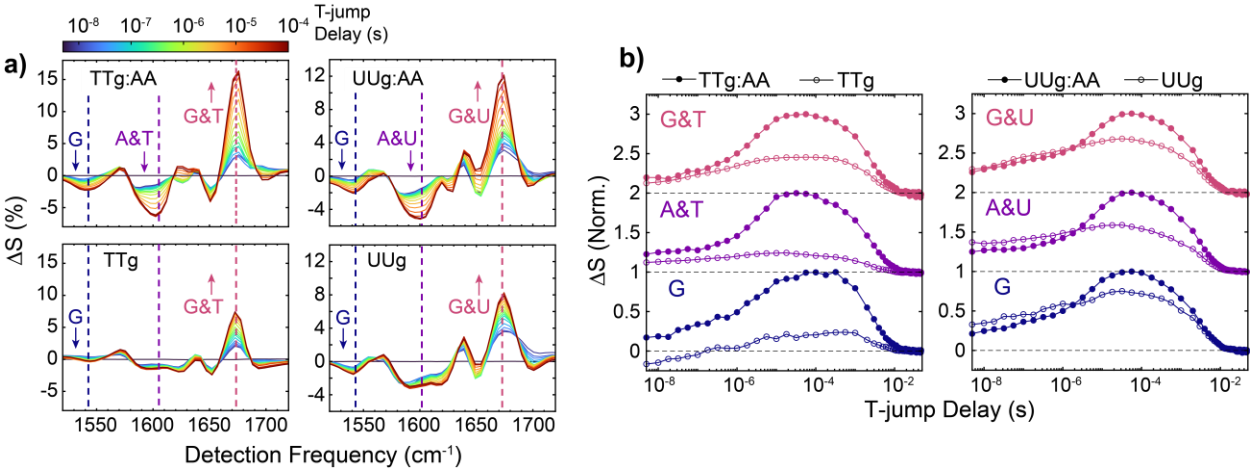

Figure 9.6 Time-dependence of AA dissociation from gaps observed with T-jump IR spectroscopy. (a) Heterodyned dispersed vibrational echo difference spectra (t-HDVE) at T-jump delays ranging from 5 ns to 180 μs for TTg:AA, TTg, UUg:AA, and UUg following a T-jump from 19 to 33 °C. Spectra are plotted as the change relative to the maximum signal value of the initial temperature spectrum ($\Delta S(t) = S(t)/\max(S(T_i))$). **(b)** Normalized $\Delta S(t)$ time traces at 1545 cm^{-1} (blue, guanine ring mode), 1605 cm^{-1} (purple, adenine ring mode), and 1672 cm^{-1} (red, guanine and thymine carbonyl modes) for TTg:AA and UUg:AA (filled circles) as well as TTg and UUg (open circles). The probe frequencies are marked as vertical dashed lines in the t-HDVE spectra. Traces at different frequencies are shifted vertically with respect to one another and dashed lines indicate respective baselines. AA dissociation is observed as a single kinetic component from 1 to 10 μs, and base-pairing dynamics from the template itself occur on an overlapping timescale.

To directly monitor the time-dependence of AA dehybridization from gaps and overhangs, we employed T-jump IR spectroscopy[51-52] using T-jumps of ΔT ≈13 °C across the center of the AA melting transition (Fig. 9.6). The absorption change at 1605 cm^{-1} by the adenine ring mode and at 1545 cm^{-1} by the guanine ring mode provide selective reporters for changes in A:T and G:C base pairing, whereas the 1672 cm^{-1} band tracks both A:T and G:C base-pair disruption through the thymine and guanine carbonyl vibrations.[7, 53] For all complexes, time traces at these three frequencies exhibit what appears to be a single kinetic component from 500 ns to 10 μs prior to

decay of the signal due to thermal re-equilibration of the sample from 1 to 10 ms (Fig. 9.6b). The large amplitude of the A:T response at 1605 cm^{-1} and its microsecond timescale support assignment of the process to AA unbinding.

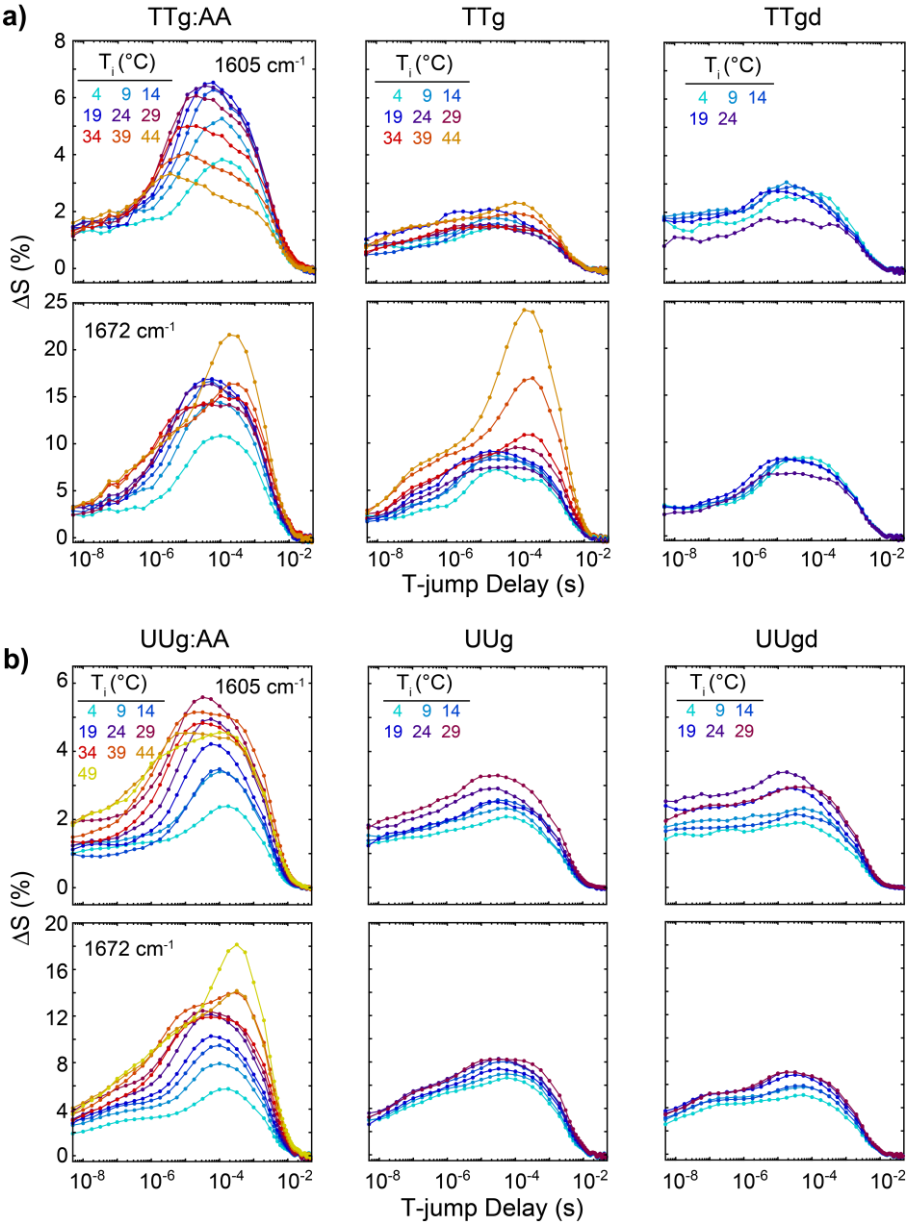

Figure 9.7 Temperature-dependent T-jump measurements of gap sequences. (a) t-HDVE time-domain responses of TTg:AA, TTg, and TTgd probed at 1605 (top) and 1672 cm^{-1} (bottom) for multiple temperatures indicated on the plot. $\Delta T \sim 13$ °C for all measurements. **(b)** Similar plots for UUg:AA, UUg, and UUgd.

Without AA, the gap and overhang templates also exhibit a response over a similar time window but with lower amplitude, particularly at the adenine ring mode, relative to the AA-template complexes. The dominant G:C character of the response suggests that it may arise from base-pair disruption adjacent to the gap or at the primer and helper termini; however, the same response is also observed in the fully formed duplexes (Fig. 9.7), which indicates it arises from fraying of the primer and helper G:C termini. This observation is consistent with previous NMR studies that found G:C base pairs adjacent to a gap to be more stable than those at the duplex termini,[33] and T-jump experiments indicating G:C fraying in DNA oligonucleotides on microsecond timescales.[54] G:C fraying occurs faster than AA dissociation yet with enough temporal overlap that only a single kinetic component is resolved at each frequency. AA dissociation dominates the observed kinetic component at most frequencies, but the signal change at 1672 cm^{-1} contains enough amplitude from G:C fraying that the extracted observed rate is faster than at 1605 cm^{-1} for TTg:AA at low temperature.[5]

9.3 Thermodynamic and kinetic modeling

9.3.1 Thermodynamic analysis of AA dissociation

The results demonstrate steady-state and time-resolved thermal dissociation of AA, which we self-consistently model to extract quantitative thermodynamic and kinetic information. In general, complete dissociation of the AA-template complex may proceed through multiple pathways differentiated by the order in which AA, primer, or helper segments dissociate. Exact treatment of these coupled equilibria leads to a fourth or higher-order dependence of each binding fraction on the concentration of a given species, which we find does not have a stable solution over

the full temperature range of each transition. However, the much greater binding stability of the primer and helper allows us to neglect species with AA bound to the template strand without the primer and helper. We also model primer and helper dissociation from the template strand as a unimolecular equilibrium between the template and "dissociated" template that is equivalent to treating binding of the primer and helper as folding of hairpins.[21] This results in a three-state sequential model described by equilibrium constants for AA dissociation (K_d) and dissociation of the primer and helper ($K_{d,Temp}$). Using TTg:AA as an example:

$$\text{TTg:AA} \overset{K_\text{d}}{\rightleftharpoons} \text{TTg} + \text{AA} \tag{9.1a}$$

$$\text{TTg} \overset{K_\text{d,Temp}}{\rightleftharpoons} \text{dTTg} \tag{9.1b}$$

where dTTg is the "dissociated" template. The assumption in eq. 9.1b sharpens the primer and helper melting profile relative to its true bimolecular,[5] but this is not a serious concern because this only affects our analysis in the temperature range where the transitions overlap.

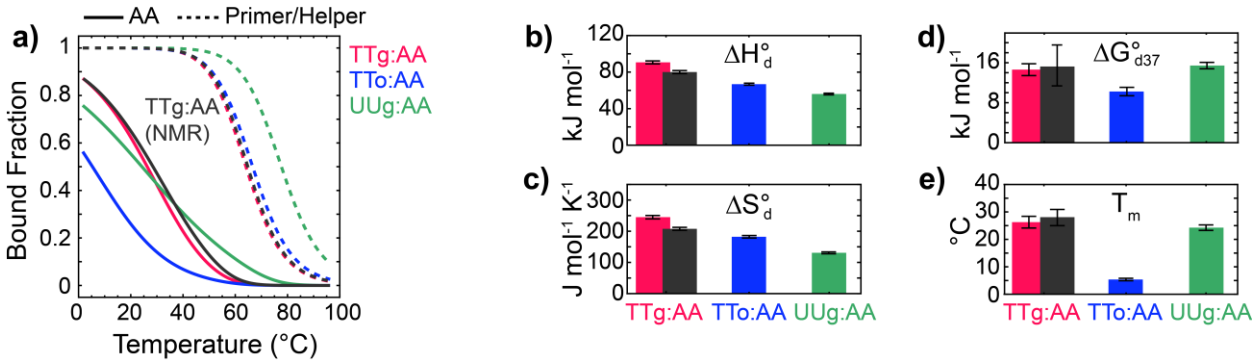

Figure 9.8 Thermodynamics of AA dissociation from gap and overhang templates extracted with three-state global fitting of FTIR and 2D IR temperature series. (a) AA (solid) and primer and helper (dashed) dissociation curves for TTg:AA, TTo:AA, and UUg:AA extracted from global fitting of FTIR and 2D IR temperature series to a three-state sequential model (eqs. 9.1a-b, Appendix 9.B). Dark gray lines correspond to dissociation components of TTg:AA obtained from ^{1}H NMR temperature series (Fig. 9.5). AA dissociation **(b)** enthalpy (ΔH°_d), **(c)** entropy (ΔS°_d), **(d)** free energy at 37 °C (ΔG°_{d37}), and **(e)** melting temperature (T_m) determined for each sequence. Dark gray bars indicate values from ^{1}H NMR. Error bars are derived from 95% confidence intervals in fit parameters ΔH°_d and ΔS°_d.

To extract melting curves for AA dissociation, FTIR and 2D IR temperature series for a given sequence are globally fit to eqs. 9.1 using the spectral decomposition described in Appendix 9.C and constraints from the titration at 1°C. The resulting melting curves are shown in Fig. 9.8 and thermodynamic parameters are shown in Fig. 9.8 and Table 9.1, and the corresponding spectral components are shown in Figs. 9.C2. Overall, the global fit describes both data sets well (Fig. 9.2). The first spectral component corresponds to AA dissociation and contains signatures of A:T base-pair breaking whereas the second component contains only G:C base-pair breaking features. The AA dissociation curves resemble the raw temperature-dependent change in absorbance and start from a bound fraction less than 1 at 1 °C.

We also find quantitative agreement between the TTg:AA dissociation curves determined from FTIR and 2D IR and from global fitting of temperature-dependent ^{1}H NMR chemical shifts (Fig. 9.5), lending support to the accuracy of the IR dissociation curves. The temperature-dependence trends in chemical shift of A3, A4, C1, and C2 resonances (Fig. 9.2) were globally fit to the sequential three-state model (eq. 9.B14). A3 and A4 peak chemical shifts were used for the first component (θ_{A_n}) and those of C1 and C2 for the second component (θ_{Temp}). Linear temperature-dependent changes in chemical shift also contribute to each peak and are corrected for using baseline fits. The high-temperature baselines of A3 and A4 are constrained to those measured for free AA (Fig. 9.5a). The extracted melting curves and thermodynamic parameters for the AA-gap complex and template dissociation quantitatively agree with those determined from global fitting of FTIR and 2D IR temperature series (Fig. 9.8) as well as ITC measurements.[5]

Table 9.1 Thermodynamic parameters for dissociation of AA from overhang and gap templates. Values determined from global fitting of spectroscopic temperature series to three-state sequential model (eqs. 9.1a-b). Error bars are derived from 95% confidence intervals in fit parameters ΔH_d° and ΔS_d°. ΔH_d° and the dissociation free energy at 8 °C (ΔG_{d8}°) from ITC measurements are also listed where errors correspond to 95% confidence intervals from two-state fits (Appendix 6.B).

Sequence	Method	T_m[c] (°C)	ΔH_d° (kJ mol^{-1})	ΔS_d° (J mol^{-1} K^{-1})	ΔG_{d8}°[d] (kJ mol^{-1})
TTg:AA	IR[a]	26 ± 2	91 ± 2	245 ± 6	21.7 ± 1.1
	NMR	28 ± 3	79 ± 5	203 ± 20	21.5 ± 4.5
	ITC[b]	-	83 ± 5	-	23.1 ± 1.3
UUg:AA	IR[a]	24 ± 1	56 ± 1	131 ± 3	19.2 ± 0.6
	ITC[b]	-	51 ± 1	-	23.5 ± 0.4
TTo:AA	IR[a]	5 ± 1	67 ± 1	182 ± 4	16.1 ± 0.8

[a]IR refers to global fits of FTIR and 2D IR temperature series. [b]Isothermal titration calorimetry (ITC) was performed at 8 °C. [c]T_m is defined as the temperature where half of the AA is bound to the template. [d]Free energy change for AA dissociation at 8 °C.

The shape of AA dissociation curves varies between DNA and RNA as well as the type of template. TTg:AA and UUg:AA exhibit similar transition midpoints (T_m), defined as the temperature where half of the total AA concentration is bound to the template, yet UUg:AA shows a broader transition with a dissociation curve slope at T_m of 0.013 K^{-1} relative to 0.021 K^{-1} in TTg:AA. The broader transition width in UUg:AA arises from a 35 kJ mol^{-1} lower dissociation enthalpy (ΔH_d°) and 115 J mol^{-1} K^{-1} lower dissociation entropy (ΔS_d°) relative to TTg:AA. This difference in ΔH_d° is also directly observed through ITC measurements (Table 9.1).[5] Although TTg:AA and UUg:AA exhibit a similar dissociation free energy at 37 °C (ΔG_{d37}°), the larger ΔH_d° and ΔS_d° values lead to greater binding stability for TTg:AA at lower temperatures and UUg:AA at higher temperatures, but we note that ITC gives similar ΔG_d° for each sequence at 8 °C. Thymine contributes greater π-π stacking interactions than uracil,[55-56] which is why AA is the only dinucleotide step with nearest-neighbor ΔH_d° and ΔS_d° parameters that are larger for DNA (33 kJ mol^{-1}; 93 J mol^{-1} K^{-1}) than RNA (28 kJ mol^{-1}; 79 J mol^{-1} K^{-1}).[8-9] This effect contributes to the larger values of ΔH_d° and ΔS_d° in TTg:AA relative to UUg:AA but cannot solely account for the observed magnitude of this difference. The free energy of AA dissociation at 37 °C is roughly 4 kJ mol^{-1} higher in TTg:AA than TTo:AA due to the added stacking stabilization from the 3′-G, and this value is similar to previous measurements of coaxial stacking in 5′-AG-3′ steps of DNA oligonucleotides.[20, 22-23]

9.3.2 Kinetic analysis of AA dissociation

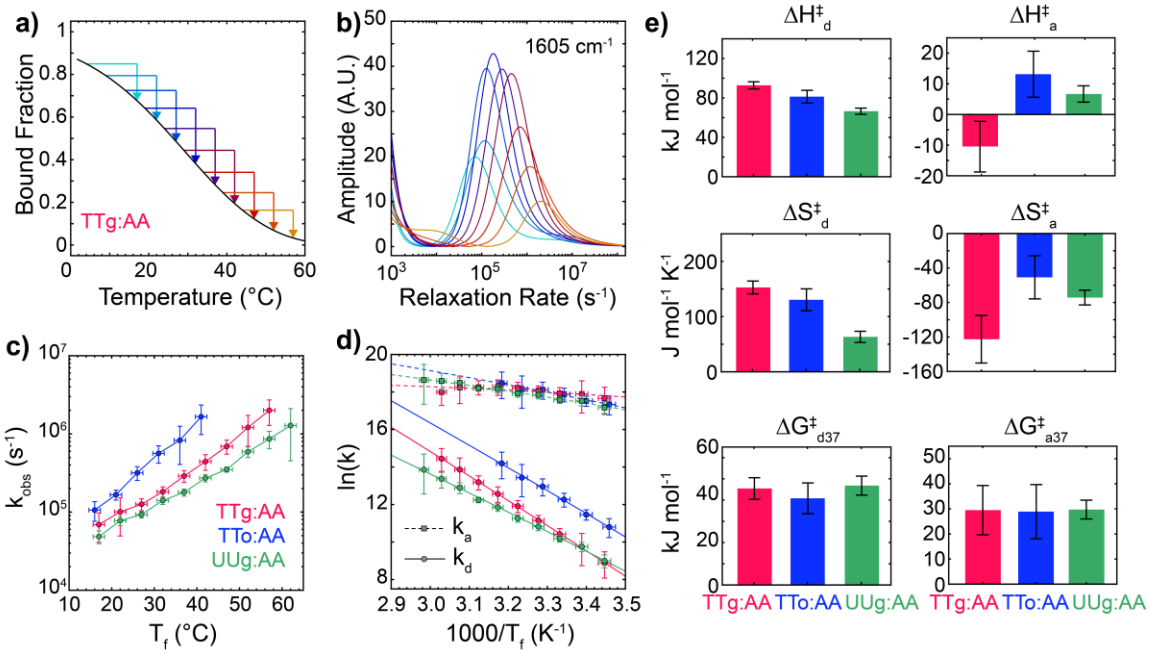

Figure 9.9 Kinetics of AA association and dissociation from gaps and overhangs. (a) Illustration of a series of T-jumps performed from various initial to final temperatures ($T_i \rightarrow T_f$) with ΔT~13 °C for TTg:AA. **(b)** t-HDVE rate distributions of TTg:AA determined from MEM-iLT[6-7] probed at the adenine ring mode (1605 cm^{-1}) for each temperature condition shown in **(a)**. **(c)** Observed relaxation rate (k_{obs}) for AA unbinding as a function of T_f for all sequences. Rates correspond to the amplitude-weighted mean across the adenine ring mode response (1585 to 1610 cm^{-1}) in the rate-domain spectra. Vertical error bars indicate standard deviation over the rate spectra and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. **(d)** Rate constants for AA association (k_a) in M^{-1} s^{-1} and dissociation (k_d) in s^{-1} as a function of T_f determined from a two-state analysis of k_{obs} (eq. 9.2). Data for all sequences is shown in ref 5. **(e)** Enthalpic ($\Delta H_d^{\ddagger}$, $\Delta H_a^{\ddagger}$), entropic ($\Delta S_d^{\ddagger}$, $\Delta S_a^{\ddagger}$), and free energy (37°C, $\Delta G_{d37}^{\ddagger}$, $\Delta G_{a37}^{\ddagger}$) barriers determined from fitting k_d and k_a temperature-trends to a Kramers-like model (eq. 9.3). Error bars indicate 95% confidence intervals propagated from the fits.

The observed relaxation rates for AA dissociation kinetics, k_{obs}, are extracted from T-jump measurements with varying initial temperature across the AA melting transition of each sequence (Figs. 9.7 and 9.9). Due to the differences in the number of relaxation components observed across

a series of temperatures, we found it most robust to extract rates from the rate domain. The time-domain t-HDVE response at each frequency is converted to the rate-domain using an inverse Laplace transform that employs a maximum entropy regularization method (MEM-iLT) described previously.[6-7] Only AA dissociation and G:C fraying are observed below final temperatures T_f = 42 °C for DNA and 52 °C for RNA. At higher temperatures, primer and helper dissociation is observed on 100 μs timescales and is more than 100-fold slower than AA dissociation (Fig. 9.9). The timescale separation between AA dissociation and primer and helper dissociation is sufficient to treat k_{obs} with two-state kinetics (Chapter 3.6).[57] Using TTg:AA as an example,

$$k_{obs} = k_d + k_a\left([\mathrm{TTg}]+[\mathrm{AA}]\right) \tag{9.2}$$

where k_a and k_d are the rate constants for AA association and dissociation. Since $K_d = k_d/k_a$, these rate constants are extracted using the concentrations of template, [TTg], and AA, [AA], determined from the AA dissociation curves in Fig. 9.8. Due to the broad observable temperature range of AA-gap dissociation transitions, T-jump measurements can be performed at multiple T_f values below T_m where k_{obs} is dominated by k_a. Over this temperature range, the dissociation rate follows Arrhenius behavior and increases from 3×10^3 to 2×10^6 s^{-1} while k_a shows just a minor 2-to-5-fold increase. Such contrasting temperature-dependence of k_a and k_d has consistently been observed in studies of short DNA and RNA oligonucleotides (5 – 20 bp),[7, 11-12, 58-59] and has been used to characterize the transition-state between hybridized and dissociated states. We fit the temperature trends of k_a and k_d to a Kramers-like model for diffusive barrier crossing where the rate is inversely proportional to the solvent viscosity, η. Written for the dissociation rate:

$$k_d(T) = \left(c^\circ\right)^{1-m} \frac{k_B T}{h} \frac{\eta_{37}}{\eta(T)} \exp\left(-\frac{\Delta H_d^\ddagger - T\Delta S_d^\ddagger}{RT}\right) \tag{9.3}$$

and the same expression applies for association. Here the free energy barrier ($\Delta G_d^\ddagger$) is written in terms of the enthalpic ($\Delta H_d^\ddagger$) and entropic ($\Delta S_d^\ddagger$) barriers. The temperature-dependent viscosity of D_2O is taken relative to its value at 37 °C (η_{37}/η).[60] Also, c° is the standard state concentration of 1 M, and *m* is the reaction molecularity (*m* = 1 for k_d and *m* = 2 for k_a). We use a simple form of the Eyring pre-factor ($k_B T/h$). This approximation is known to overestimate the frequency of diffusive motion at the barrier for biomolecular folding and binding.[61-62] While this impacts the absolute magnitudes of $\Delta S_d^\ddagger$ and $\Delta G_d^\ddagger$, we assume that the true pre-factor is the same across our sequences such that relative values of $\Delta S_d^\ddagger$ and $\Delta G_d^\ddagger$ are meaningful. Fitting of our data indicates that the free energy barrier for AA dissociation arises from an enthalpic penalty ($\Delta H_d^\ddagger$ > 50 kJ mol^{-1}) that is partially compensated by a positive $\Delta S_d^\ddagger$, whereas the free energy barrier for AA association ($\Delta G_a^\ddagger$) is dominated by an entropic barrier ($\Delta S_a^\ddagger$ < 0).

The kinetics of AA association and dissociation vary substantially depending on the template. $\Delta G_d^\ddagger$ is ~4.5 kJ mol^{-1} smaller for TTo:AA than TTg:AA at 37 °C, which matches the difference in ΔG_{d37}°, suggesting that the 5-fold faster unbinding of AA from an overhang compared to a gap primarily comes from reduced binding stability. The rate of association or $\Delta G_{a37}^\ddagger$ are essentially the same for TTo:AA and TTg:AA but with small differences in its temperature-dependence as quantified by $\Delta H_a^\ddagger$. There are also significant differences between the kinetics of AA dissociation in UUg:AA and TTg:AA. k_d increases more sharply with temperature in

TTg:AA, as quantified by a ~25 kJ mol^{-1} larger $\Delta H_d^{\ddagger}$ relative to UUg:AA, which likely comes from enhanced stacking interactions in the hybridized state of TTg:AA and is consistent with the thermodynamic results presented above. On the other hand, k_a is similar for each sequence across the measured temperature range.

9.4 Structure of AA-gap complexes from MD simulations

The energetics of AA hybridization onto overhangs and gaps differs from free strand hybridization due to a combination of differences in the template and AA-gap complex structural configurations relative to canonical duplexes and single-strands. We characterized the differences in conformation between these species using all-atom MD simulations, finding that most configurational differences are localized near the central AA binding site. As observed previously, gap templates adopt bent configurations in high population which reduces hydration of the gap relative to single strands.[5, 32-33, 63] AA-gap complexes exhibit greater configurational freedom than canonical duplexes (Fig. 9.10), with the largest differences found in backbone and base-stacking structure rather than base-pair geometry. TTg:AA exhibits broadened distributions in shift, slide, and twist parameters at the 5′-GA-3′ and 5′-AG-3′ nicks, whereas the stacking geometry of the between adenine bases is nearly identical to B-DNA (ΔDx, ΔDy < 0.4 Å and $\Delta\Omega$ < 4°). A broader spread of stacking orientations at the nick sites is unsurprising given the greater flexibility in the backbone, particularly for glycosidic bond angles (χ) of the A and G bases on the 3′-side of each nick.[27] The nick sites enable greater local flexibility in the backbone, and in particular for the glycosidic bond angles (χ) of the A and G bases on the 3′-side of each nick. The fluctuations along

slide, twist, shift, and χ coordinates are partially correlated in a given nick site while the structural fluctuations in the 5′-GA-3′ and 5′-AG-3′ nicks appear independent of one another (Fig. 9.10b,c).

The structural differences between UUg:AA and an A-RNA duplex are also minor yet distinct from DNA in a few regards. Overall, UUg:AA exhibits larger changes in stacking orientation (twist, shift) and less change in the backbone torsional angles (χ, δ, γ) relative to DNA. Further, the base-pairing and stacking geometry of the AA differs with broader distributions of shear, stretch, and buckle values, an increase in mean rise from 3.0 to 3.3 Å, and sub-populations with larger shift and twist values. In contrast to DNA, distinct states are observed along twist and shift coordinates for the 5′-GA-3′ and 5′-AG-3′ and the mean values differ from A-RNA by as much as 20° and 1.5 Å, respectively.

During one of two 4 μs trajectories for TTg:AA, we observed AA switch from canonical Watson-Crick-Franklin (WCF) to anti-Hoogsteen (aHG) base pairing geometry. As shown in Fig. 9.10d, the 5′-A flips at 700 ns followed by the 3′-A 200 ns afterward, however, both adenines remain poorly aligned with the gap until reaching the final aHG geometry at 1.3 μs. Our 8 μs of total simulation time for this sequence is insufficient to determine the relative stability of WCF and HG configurations, therefore we applied the On-the-fly Probability Enhanced Sampling (OPES) method to calculate the free energy differences for AA dissociation from a WCF or aHG geometry ($\Delta\Delta F_{WCF-aHG} = \Delta F_{d37,aHG} - \Delta F_{d37,WCF}$).[5] WCF and aHG geometries have nearly the same population and stability at 37 °C with $\Delta\Delta F_{WCF-aHG} \approx 0$ kJ mol^{-1}. Although not previously tested for dinucleotide-gap complexes or aHG base pairing, simulations with the bsc1-AMBER

force field typically capture accurate relative stabilities between WCF and syn-Hoogsteen

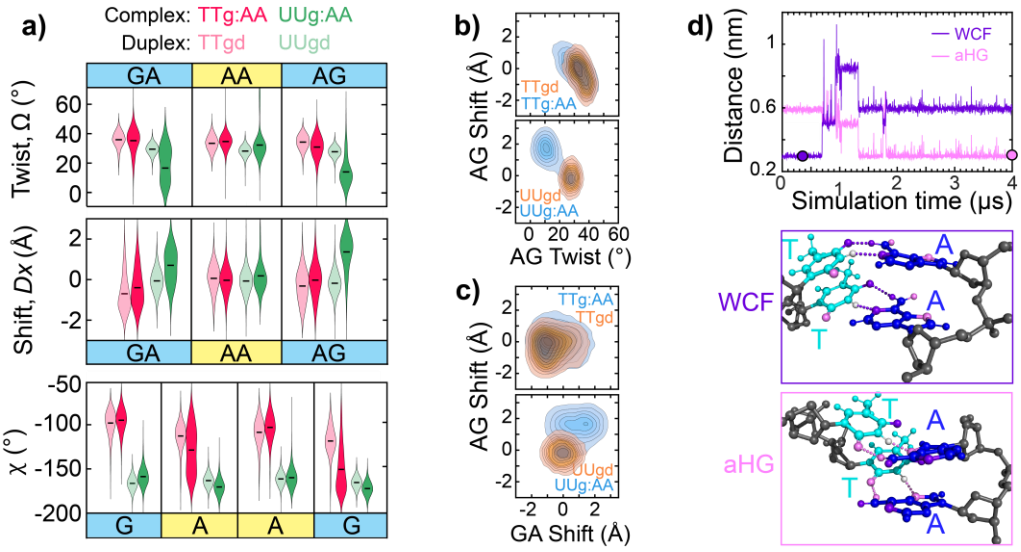

Figure 9.10 Structural properties of AA-gap complexes from MD simulations. (a) Comparison of structural variation between AA-gap complex and canonical duplex. Violin plots show structural distributions for base-twist (Ω) and shift (Dx), and glycosidic bond dihedral angle (χ) for the three central base pair steps of TTgd (red, light), TTg:AA (red, dark), UUgd (green, light), and UUg:AA (green, dark). Black horizontal bars indicate the distribution's median value obtained from a Gaussian kernel estimation using Scott's rule.[4] Distributions were generated from 1.5 μs of simulation for each sequence using the bsc1-AMBER force field for DNA and DES-AMBER force field for RNA. Only Watson-Crick-Franklin (WCF) base pair configurations are included in the distributions. Other structural parameters using both force fields are shown in ref. 47. **(b)** Contour plots showing the correlation of 5′-AG-3′ twist vs. shift. Nine contours with uniform 10% spacing are plotted for each sequence. Diagonally elongated distributions indicate that fluctuations along different base pair coordinates within the same nick site are partially correlated with Pearson correlation coefficients of -0.61 for DNA and -0.56 for RNA. **(c)** Contour plots showing the correlation of 5′-GA-3′ shift vs. 5′-AG-3′ shift. Symmetrical distributions indicate that fluctuations at each nick site are uncorrelated with Pearson correlation coefficients of -0.04 for DNA and 0.05 for RNA. **(d)** 4 μs trajectory of TTg:AA showing the A:T base pairs switching from WCF to an anti-Hoogsteen (aHG) geometry. Trajectories are plotted in terms of the mean WCF (purple) and aHG (pink) distances across each A:T base pair. The WCF distance is defined as the mean of adenine-N6 to thymine-O4 and adenine-N1 to thymine-N3 distances and the aHG distance is the average of adenine-N7 to thymine-N3 and adenine-N6 to thymine-O2 distances. Visualizations of each base pair geometry at time points marked with filled circles indicate WCF distances as purple dashed lines and aHG distances as pink dashed lines.

base-pairing geometries.[64-65] Simulations with the DES-AMBER force field instead suggest that the WCF geometry is more stable ($\Delta\Delta F_{WCF-aHG}$ ~ -8 kJ mol^{-1}). For UUg:AA, WCF and aHG geometries have similar stabilities regardless of the employed force field.[5] The high stability of an aHG geometry may appear unusual, but this has been reported as the dominant base-pairing geometry between AA and TT DNA dinucleotides mediated through self-assembling tripyridyl-triazinine capped hydrophobic cages in aqueous solution, so far the only structural characterization of dinucleotide duplexes in aqueous solution.[66]

9.5 Length scaling of A_n dissociation from templates

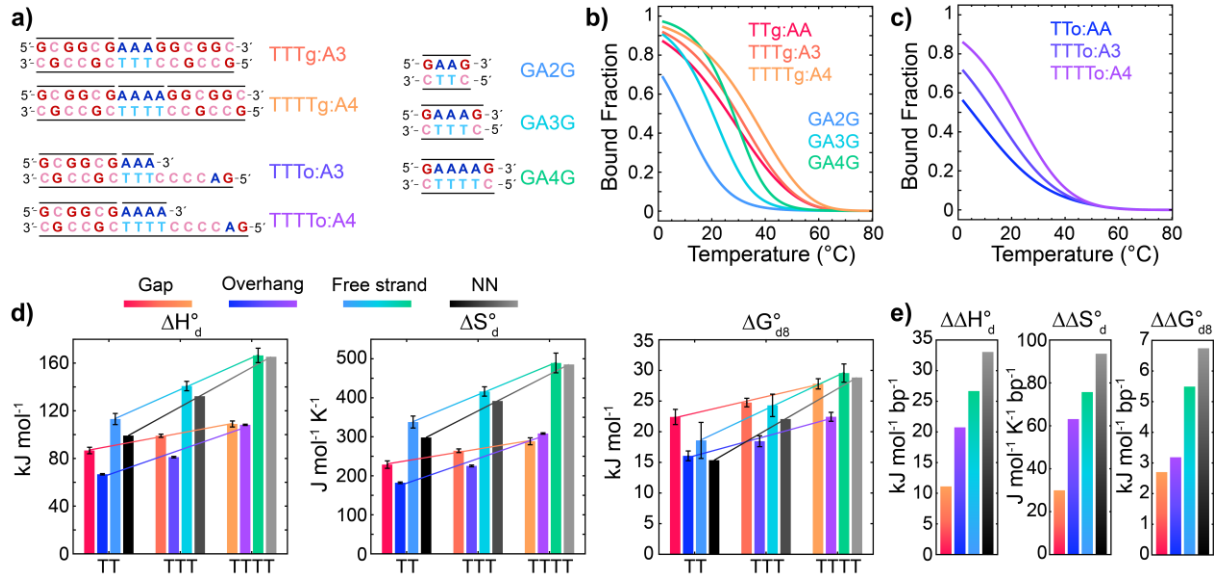

Figure 9.11 Length-dependent dissociation thermodynamics for gaps, overhangs, and free strands. (a) Extended templates for 3 and 4 nucleotide gap (TTTg:A3, TTTTg:A4) and overhang (TTTo:A3, TTTTo:A4) sequences bound to A3 and A4 oligonucleotides. Dissociation of A_n (n = 2, 3, 4) segments from gaps and overhangs are compared with dissociation of GA_nG single-strands from their complement strands. **(b, c)** A_n melting curves for gaps and overhangs extracted from global fitting of FTIR and 2D IR temperature series to three-state sequential model (eqs. 9.1). GA_nG melting curves are shown in **(b)** as extracted from a two-state fit to FTIR temperature series

Figure 9.11 Length-dependent dissociation thermodynamics for gaps, overhangs, and free strands, continued

(Appendix 6.A).[5] **(d)** ΔH_d°, ΔS_d°, and dissociation free energy at 8 °C (ΔG_{d8}°) for all sequences. Red-orange bars correspond to DNA A_n-gap complexes, blue-purple bars correspond to DNA A_n-overhang complexes, light blue-green bars are for GA_nG strands, and black-gray indicate GA_nG dissociation parameters calculated from Santa Lucia's nearest-neighbor (NN) model.[8, 10] Values for gap and GA_nG systems are an average over those determined from IR spectroscopy and ITC[5] while overhang values are only from FTIR and 2D IR global fits. Error bars indicate 95% confidence intervals from fits to thermodynamic models. **(e)** Slopes from linear fits of each thermodynamic parameter across A_n length (solid lines in **(d)**). Data are presented as the change per appended A:T base pair (bp) in the dissociation enthalpy ($\Delta\Delta H_d^\circ$), entropy ($\Delta\Delta S_d^\circ$), and free energy at 8 °C ($\Delta\Delta G_{d8}^\circ$). Although AA-gap complex binding is more stable than GA2G, $\Delta\Delta H_d^\circ$, $\Delta\Delta S_d^\circ$, and $\Delta\Delta G_{d8}^\circ$ are more than 2-fold greater in GA_nG strands than for A_n dissociation from gaps.

To assess the scaling of dissociation thermodynamics and kinetics for 2-to-4 nucleotides sequences, we experimentally examined the dehybridization of AAA (A3) and AAAA (A4) from DNA gaps and overhangs using the extended templates shown in Fig. 9.11a. As expected, longer A_n oligonucleotides bind with greater stability and show a decrease in both k_d and k_a (Figs. 9.11b-d and 9.12). Both gaps and overhangs exhibit an average increase in $\Delta G_{d,8}^\circ$ of 2.7-3.0 kJ mol^{-1} per base pair (bp), yet ΔH_d° and ΔS_d° increase more sharply with length for overhangs and become equivalent to the gap for A4 dissociation. The overhang values for $\Delta\Delta H_d^\circ$, $\Delta\Delta S_d^\circ$, and $\Delta\Delta G_{d,8}^\circ$ are still well below the 33 kJ mol^{-1} bp^{-1}, 94 J mol^{-1} K^{-1} bp^{-1}, and 6.7 kJ mol^{-1} bp^{-1} values predicted for appending AA dinucleotide steps by Santa Lucia's DNA nearest-neighbor (NN) model (Fig. 9.11e).[8]

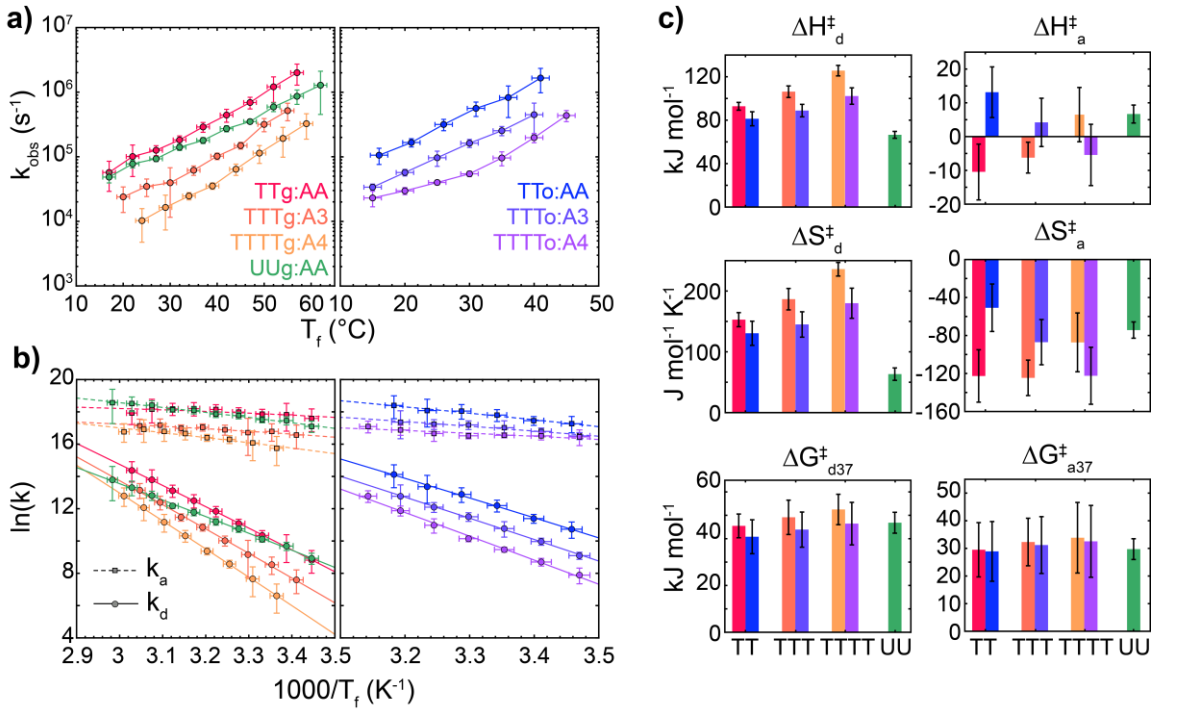

Figure 9.12 Length-scaling of kinetics for A_n association and dissociation from gaps and overhangs. (a) Observed relaxation rate (k_{obs}) for AA dissociation as a function of T_f for all sequences. Rates correspond to the amplitude-weighted mean across the adenine ring mode response (1585 to 1610 cm^{-1}) in the rate-domain spectra determined from MEM-iLT.[6-7] Vertical error bars indicate standard deviation over the rate spectra and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. **(b)** Rate constants for AA association (k_a) and dissociation (k_d) as a function of T_f determined from a two-state analysis of k_{obs} (eq. 9.2). Solid lines correspond to fits to a Kramers-like model (eq. 9.3). **(c)** Enthalpic ($\Delta H^{\ddagger}_{d,}$, $\Delta H^{\ddagger}_{a}$), entropic ($\Delta S^{\ddagger}_{d}$, $\Delta S^{\ddagger}_{a}$), and free energy (37 °C, $\Delta G^{\ddagger}_{d37}$, $\Delta G^{\ddagger}_{a37}$) barriers determined from fits in **(b)**. Error bars indicate 95% confidence intervals propagated from the fits.

To test for the origin of different length-scaling dissociation thermodynamics between overhangs, gaps, and the NN model, we compare to dissociation of oligonucleotides 5′-GAAG-3′ (GA2G), 5′-GAAAG-3′ (GA3G), and 5′-GAAAAG-3′ (GA4G) with their respective complement strands (Figs. 9.11). ΔH°_{d} and ΔS°_{d} are both greater for GA2G than AA-gap dissociation while ΔG°_{d8} is ~4 kJ mol^{-1} lower for GA2G. As a result, the GA2G dissociation transition is much sharper than

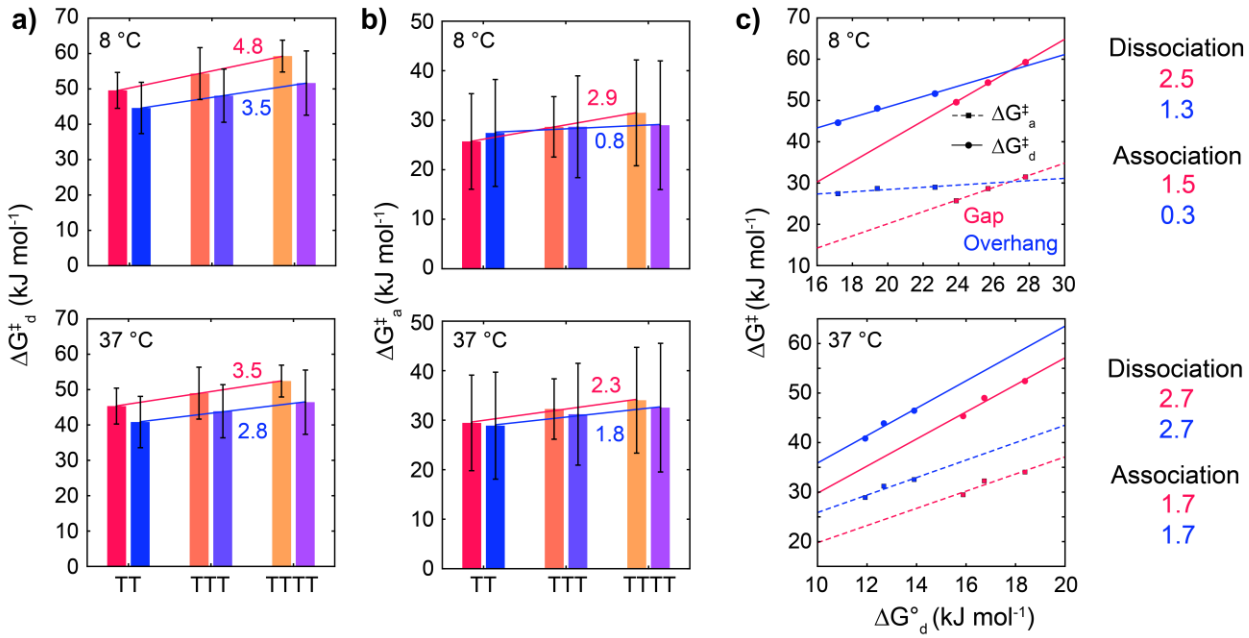

Figure 9.13 Length-dependence of dehybridization and hybridization free-energy barriers. **(a)** $\Delta G_d^{\ddagger}$ at 8 and 37 °C for DNA gap (red-yellow) and overhang (blue-purple) sequences. Red and blue solid lines indicate linear fits to gap and overhang strand length trends, respectively. The slope of gap and overhang fit lines are listed above and below each line, respectively. Error bars are determined from 95% confidence intervals of fits to eq. 9.3. **(b)** Same plots for $\Delta G_a^{\ddagger}$ at 8 and 37 °C. **(c)** Correlation plots between $\Delta G_d^{\ddagger}$, $\Delta G_a^{\ddagger}$ and ΔG_d° at 8 and 37 °C for DNA gap and overhang length series. Solid and dashed lines indicate linear fits with respective slopes reported on the right.

for AA-gap dissociation yet with T_m shifted ~15 °C lower in temperature. Values of $\Delta\Delta H_d^{\circ}$, $\Delta\Delta S_d^{\circ}$, and $\Delta\Delta G_{d8}^{\circ}$ for GA$_n$G are more than 2-fold greater than for gap sequences and are much closer the NN model values (Fig. 9.11e). This leads to nearly equivalent ΔG_{d8}° for TTTTg:A4 and GA4G, and binding between free single-strands may progressively become more stable than for binding of longer oligonucleotides onto gaps and overhangs. Additionally, we find that length-scaling of the dissociation free-energy barrier ($\Delta\Delta G_d^{\ddagger}$) is ~2.5-fold greater than $\Delta\Delta G_d^{\circ}$ for gaps (Fig. 9.13), which is a much larger difference than previously reported for dehybridization of free single-strands.[58, 67] This sharp dependence of $\Delta G_d^{\ddagger}$ (and k_d) on length is important for predicting

dissociation kinetics in template systems and also indicates a length-dependence in the free energies of the association transition state and/or the unbound state.[68-69]

Based on previous studies of gap and overhang structure,[28-33] it is possible that the single-stranded binding region in gaps and overhangs becomes less ordered as the length increases from 2 to 4 nucleotides. However, our all-atom MD simulations indicate essentially no difference in the structural distribution of the canonical unbent T_n configurations in gaps, but instead highly bent configurations and alternative stacking geometries are observed for certain gap lengths.[5] AMBER force fields, including bsc1-AMBER, tend to overestimate base-stacking stability in single-strand DNA as they are parameterized for duplex DNA,[70-71] and therefore these results should be interpreted with caution.

9.6 Concluding remarks

We demonstrate an approach for measuring temperature-dependent binding thermodynamics and kinetics of nucleic acids as short as dinucleotides onto gaps and overhangs using IR spectroscopy. By determining accurate melting curves, our method extracts enthalpic and entropic contributions to binding that are typically missing in previous titration studies. Further, to our knowledge we provide the first measurements of dinucleotide association and dissociation timescales (0.5 – 20 μs) through the application of T-jump IR spectroscopy. Although our study is limited to pure A_n segments, we identify striking energetic differences between cases of association onto gaps or overhangs with free-strand association that have not been observed previously.

Most notable is that association of AA onto a gap with G:C base pairs on each side is energetically more favorable than association between 5′-GAAG-3′ and 5′-CTTC-3′ ($\Delta\Delta G^{\circ}_{d8}$ = 4 kJ mol^{-1}, Fig. 7d). This distinction may arise from energetic differences in either the bound or dissociated states or a combination of both through factors such as single-strand stacking, optimized coaxial stacking, or non-canonical base pairing. Most association measurements, including the approach in this work, directly probe energetic differences between the bound and dissociated states but not energetic differences between the respective bound or dissociated states of different samples. As a result, we cannot directly dissect the enhanced stability of AA-template complexes but rather infer how the energy of the associated and dissociated states may change between gap, overhang, and free single-strand scenarios.

A clear distinction between gap and free single-strand binding scenarios is present in both the enthalpic and entropic contributions to association. ΔH°_{d} and ΔS°_{d} are lower for the gap, and therefore the extra stability of TTg:AA relative to GA2G arises from an entropic benefit. Enhanced binding stability of dinucleotides onto gaps and overhangs may arise from greater base-pair ordering and constraints in water dynamics within the single-strand binding region and/or increased configurational flexibility in the AA-template complexes relative to canonical B-DNA, which we observe in MD simulations near the nick sites (Figs. 9.10).[5] The former effect decreases the entropy of the dissociated state while the latter increases that of the hybridized state, each of which will decrease ΔS°_{d} relative to free stand hybridization. The same effect that may increase rigidity and stacking interactions in the gap single-strand will also reduce the enthalpy of the dissociated state. Another reduction in enthalpy may arise from bending about the single-stranded

segment in gap templates as observed in our MD simulations and reported in previous studies.[31, 33-34, 72-73] Lastly, our simulations of TTg:AA suggest that AA can bind in WCF or aHG geometries with significant population, although the relative stability depends on the choice of force field.[5] aHG configurations are likely only relevant to binding of pure A:T or A:U dinucleotides or G:G mismatch pairing where the aHG configuration is a stable option. To date, the only crystallographic studies of dinucleotide binding to a gap or overhang template are for G:C base pairing of modified GG RNA dinucleotides,[74-76] a scenario where aHG configurations are highly improbable. Unfortunately, we cannot indicate nor disprove the presence of A:T aHG base pairing using the IR and NMR measurements presented in this work, and further structural characterization is necessary to clarify the role of aHG base pairing.

The energetics of dinucleotide association onto gaps and overhangs differs from free-strand hybridization, and these effects stem from underlying dynamics of association and dissociation. Our T-jump IR measurements of DNA and RNA AA association and dissociation kinetics provide a first step toward understanding the structural dynamics. We determine a k_a value of ~5 × 10^7 M^{-1} s^{-1} that corresponds to a time constant of 20 μs at 1 mM oligonucleotide concentration. This k_a value is nearly 100-fold slower than an estimated diffusion-limited rate constant (k_{diff}) of 5 × 10^9 M^{-1} s^{-1} calculated using the translational diffusion coefficients of AA and TTg determined from DOSY NMR measurements and estimated radii of gyration from experimental length-scaling of ssDNA and worm-like chain model of duplex DNA.[5, 77-78] Although faster than association between oligonucleotides of 5 – 20 nucleotide lengths, the 100-fold difference between AA association and diffusion-limited rates suggest that numerous unsuccessful collisions and/or

significant structural rearrangement occur prior to successful binding. Association rate constants for longer oligonucleotides are also consistently reported to be 100-1000-fold smaller than k_{diff},[58] suggesting that each system's hybridization transition-state and encounter complex may share certain characteristics. Even for longer oligonucleotides these properties are still poorly understood but are thought to involve some degree of single-strand ordering, a small number of base-pair contacts, and potentially rearrangement of water molecules and counterions. In the case of dinucleotide association to gaps and overhangs, water and counterion rearrangement is still necessary, and the ordering of the binding region of the gap or overhang contains a large penalty. The time ordering of these events are unclear for each system and must be investigated further to develop a molecular picture for the dynamics of short oligonucleotide hybridization.

We have primarily focused on short DNA oligonucleotide association, and it is important to assess whether similar characteristics are found for short RNA oligonucleotides, which are more relevant in the context of non-enzymatic replication.[79] Our results show large differences in energetics and kinetics of AA dissociation from DNA and RNA gaps. ΔH_d° and ΔS_d° values for UUg:AA are approximately half of those predicted by NN models and obtained for TTg:AA.[5] ΔH_d° and ΔS_d° are likely higher in DNA for multiple reasons. The AA dinucleotide in particular contains enhanced stacking arising from the C5-methyl groups of thymine.[55-56, 80] Another contribution may come from the wider breadth of stacking configurations observed in UUg:AA relative to UUgd. However, RNA duplexes are generally more thermodynamically stable than an equivalent sequence DNA duplex due to multiple effects from the C2′-OH groups. Differences in base stacking, hydrogen bonding between C2′-OH groups and water molecules, and greater uptake and

ordering of cations and water molecules are all effects that contribute to larger ΔH_d°, ΔS_d°, and ΔG_d° in A-RNA than B-DNA.[81-82] Therefore, we predict that binding of other RNA dinucleotide sequences will generally be more stable than the equivalent DNA sequence.

Gaps and overhangs are expected to have a qualitatively similar thermodynamic and kinetic impact on DNA and RNA hybridization, but the relative magnitude may generally differ. For instance, single- and double-strand RNA are shown to have a longer persistence length than DNA and the change in strand rigidity upon introducing an overhang or gap may differ as well as the propensity for bending of gapped RNA.[35, 83] MD simulations of UUg:AA and UUgd indicate that the distribution of stacking geometries at nick sites between primer and helper segments and AA differ from A-RNA while the backbone geometry is relatively unchanged, yet the opposite is found for DNA (Fig. 9.10a). Such differences in geometry may influence the enthalpy of the AA-gap complex while the enhanced configurational flexibility increases its entropy. Together, the combination of differences between DNA and RNA templates and dinucleotide-template complexes may promote significant contrast in their association energetics and kinetics across variations in dinucleotide sequence and template construction.

Our experimental and computational results indicate multiple structural, energetic, and kinetic factors that differentiate hybridization between short nucleic acids and templates with hybridization between free single-strands, and these differences may play an important role in processes such as non-enzymatic replication. In addition to various structural and chemical factors, the efficiency of this process depends upon the binding probability, dissociation rate of the bound complex, and structural dynamics of the complex.[76, 84-86] The structural dynamics of the gap

template and bound complex suggested by our experimental and computational results are likely to influence the efficiency of non-enzymatic extension. Further, we measure an AA dissociation rate from UUg:AA that is orders of magnitude higher than the catalytic rate (k_{cat}) of non-enzymatic primer extension for a 2-aminoimidaloze-bridged (2AI) dinucleotide in an equivalent sequence context,[85] indicating that on average the 2AI dinucleotide binds and dissociates from the gap numerous times before extension occurs. Although measured under different solution conditions, the dissociation constant of UUg:AA at 25 °C (K_d = 1.05 mM) is similar to the Michaelis constant of the 2AI system (K_m = 0.64 mM) extracted from a Michaelis-Menten analysis of the extension rate.[85] These observations support the use of a Michaelis-Menten kinetic model for non-enzymatic extension of this sequence and template, and our study demonstrates a method for direct measurement of energetics and kinetics for dinucleotide binding that is applicable to a wide range of oligonucleotide sequences and template designs.

9.7 Acknowledgements

I thank Michael Jones and Smayan Khanna for carrying out the MD simulations in this chapter, Aleksandar Radakovic for synthesizing and purifying the RNA samples, and Joseph Sachleben for help with NMR measurements. I also thank Yumin Lee, Marco Todisco, and Jakob Schauss for helpful discussions and feedback on the manuscript that covers the material in this chapter.

9.8 References

1. Yin, Y.; Zhao, X. S., Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* **2011,** *44*, 1172-1181.

2. Wetmur, J. G., Hybridization and renaturation kinetics of nucleic acids. *Annu. Rev. Biophys.* **1976,** *5*, 337-361.

3. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.

4. Scott, D. W., On optimal and data-based histograms. *Biometrika* **1979,** *66*, 605-610.

5. Ashwood, B.; Jones, M. S.; Radakovic, A.; Khanna, S.; Lee, Y.; Sachleben, J. R.; Szostak, J. W.; Ferguson, A. L.; Tokmakoff, A., Thermodynamics and kinetics of DNA and RNA dinucleotide hybridization to gaps and overhangs. *Biophys. J.* **2023,** *122*, 3323-3339.

6. Kumar, A. T.; Zhu, L.; Christian, J.; Demidov, A. A.; Champion, P. M., On the rate distribution analysis of kinetic data using the maximum entropy method: Applications to myoglobin relaxation on the nanosecond and femtosecond timescales. *J. Phys. Chem. B* **2001,** *105*, 7847-7856.

7. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.

8. SantaLucia Jr, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci.* **1998,** *95*, 1460-1465.

9. Xia, T.; SantaLucia Jr, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H., Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with Watson− Crick base pairs. *Biochem.* **1998,** *37*, 14719-14735.

10. Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A., Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* **2008,** *47*, 5336-5353.

11. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.

12. Pörschke, D.; Eigen, M., Co-operative non-enzymatic base recognition III. Kinetics of the helix—coil transition of the oligoribouridylic· oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **1971,** *62*, 361-381.

13. Simmel, F. C.; Yurke, B.; Singh, H. R., Principles and applications of nucleic acid strand displacement reactions. *Chem. Rev.* **2019,** *119*, 6326-6369.

14. Wei, X.; Nangreave, J.; Liu, Y., Uncovering the self-assembly of DNA nanostructures by thermodynamics and kinetics. *Acc. Chem. Res.* **2014,** *47*, 1861-1870.

15. Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S., Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **1998,** *391*, 304-307.

16. Steitz, T. A., A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **2008,** *9*, 242-253.

17. Zhou, L.; Ding, D.; Szostak, J. W., The virtual circular genome model for primordial RNA replication. *RNA* **2021,** *27*, 1-11.

18. Duzdevich, D.; Carr, C. E.; Ding, D.; Zhang, S. J.; Walton, T. S.; Szostak, J. W., Competition between bridged dinucleotides and activated mononucleotides determines the error frequency of nonenzymatic RNA primer extension. *Nucleic Acids Res.* **2021,** *49*, 3681-3691.

19. Leveau, G.; Pfeffer, D.; Altaner, B.; Kervio, E.; Welsch, F.; Gerland, U.; Richert, C., Enzyme-Free Copying of 12 Bases of RNA with Dinucleotides. *Angew. Chem. Int. Ed.* **2022**, e202203067.

20. Pyshnyi, D. V.; Goldberg, E. L.; Ivanova, E. M., Efficiency of coaxial stacking depends on the DNA duplex structure. *J. Biomol. Struct. Dyn.* **2003,** *21*, 459-467.

21. Lane, M. J.; Paner, T.; Kashin, I.; Faldasz, B. D.; Li, B.; Gallo, F. J.; Benight, A. S., The thermodynamic advantage of DNA oligonucleotide 'stacking hybridization'reactions: energetics of a DNA nick. *Nucleic Acids Res.* **1997,** *25*, 611-616.

22. Vasiliskov, V. A.; Prokopenko, D. V.; Mirzabekov, A. D., Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips. *Nucleic Acids Res.* **2001,** *29*, 2303-2313.

23. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D., Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006,** *34*, 564-574.

24. Walter, A. E.; Turner, D. H.; Kim, J.; Lyttle, M. H.; Müller, P.; Mathews, D. H.; Zuker, M., Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci.* **1994,** *91*, 9218-9222.

25. Walter, A. E.; Turner, D. H., Sequence dependence of stability for coaxial stacking of RNA helixes with Watson-Crick base paired interfaces. *Biochem.* **1994,** *33*, 12715-12719.

26. Parker, T. M.; Hohenstein, E. G.; Parrish, R. M.; Hud, N. V.; Sherrill, C. D., Quantum-mechanical analysis of the energetic contributions to π stacking in nucleic acids versus rise, twist, and slide. *J. Am. Chem. Soc.* **2013,** *135*, 1306-1316.

27. Häse, F.; Zacharias, M., Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Res.* **2016,** *44*, 7100-7108.

28. Israels, B.; Albrecht, C. S.; Dang, A.; Barney, M.; von Hippel, P. H.; Marcus, A. H., Submillisecond Conformational Transitions of Short Single-Stranded DNA Lattices by Photon Correlation Single-Molecule Förster Resonance Energy Transfer. *J. Phys. Chem. B* **2021,** *125*, 9426-9440.

29. Qu, P.; Yang, X.; Li, X.; Zhou, X.; Zhao, X. S., Direct measurement of the rates and barriers on forward and reverse diffusions of intramolecular collision in overhang oligonucleotides. *J. Phys. Chem. B* **2010,** *114*, 8235-8243.

30. Rivetti, C.; Walker, C.; Bustamante, C., Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.* **1998,** *280*, 41-59.

31. Lin, S.; Horning, D. P.; Szostak, J. W.; Chaput, J. C., Conformational analysis of DNA repair intermediates by time-resolved fluorescence spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 9585-9587.

32. Kashida, H.; Kurihara, A.; Kawai, H.; Asanuma, H., Orientation-dependent FRET system reveals differences in structures and flexibilities of nicked and gapped DNA duplexes. *Nucleic Acids Res.* **2017,** *45*, e105-e105.

33. Roll, C.; Ketterlé, C.; Faibis, V.; Fazakerley, G. V.; Boulard, Y., Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochem.* **1998,** *37*, 4059-4070.

34. Craggs, T. D.; Sustarsic, M.; Plochowietz, A.; Mosayebi, M.; Kaju, H.; Cuthbert, A.; Hohlbein, J.; Domicevica, L.; Biggin, P. C.; Doye, J. P. K., Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase. *Nucleic Acids Res.* **2019,** *47*, 10788-10800.

35. Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L., Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci.* **2012,** *109*, 799-804.

36. Dans, P. D.; Balaceanu, A.; Pasi, M.; Patelli, A. S.; Petkevičiūtė, D.; Walther, J.; Hospital, A.; Bayarri, G.; Lavery, R.; Maddocks, J. H., The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules. *Nucleic Acids Res.* **2019,** *47*, 11090-11102.

37. Harris, P. D.; Hamdan, S. M.; Habuchi, S., Relative contributions of base stacking and electrostatic repulsion on DNA nicks and gaps. *J. Phys. Chem. B* **2020,** *124*, 10663-10672.

38. Kervio, E.; Claasen, B.; Steiner, U. E.; Richert, C., The strength of the template effect attracting nucleotides to naked DNA. *Nucleic Acids Res.* **2014,** *42*, 7409-7420.

39. Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W., Uncovering the thermodynamics of monomer binding for RNA replication. *J. Am. Chem. Soc.* **2015,** *137*, 6373-6382.

40. Tam, C. P.; Fahrenbach, A. C.; Björkbom, A.; Prywes, N.; Izgu, E. C.; Szostak, J. W., Downstream oligonucleotides strongly enhance the affinity of GMP to RNA primer–template complexes. *J. Am. Chem. Soc.* **2017,** *139*, 571-574.

41. Kashida, H.; Kokubo, Y.; Makino, K.; Asanuma, H., Selective binding of nucleosides to gapped DNA duplex revealed by orientation and distance dependence of FRET. *Org. Biomol. Chem.* **2019,** *17*, 6786-6789.

42. Spasic, A.; Berger, K. D.; Chen, J. L.; Seetin, M. G.; Turner, D. H.; Mathews, D. H., Improving RNA nearest neighbor parameters for helices by going beyond the two-state model. *Nucleic Acids Res.* **2018,** *46*, 4883-4892.

43. Salditt, A.; Keil, L. M.; Horning, D. P.; Mast, C. B.; Joyce, G. F.; Braun, D., Thermal habitat for RNA amplification and accumulation. *Phys. Rev. Lett.* **2020,** *125*, 048104.

44. Mast, C. B.; Schink, S.; Gerland, U.; Braun, D., Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci.* **2013,** *110*, 8030-8035.

45. Ianeselli, A.; Mast, C. B.; Braun, D., Periodic melting of oligonucleotides by oscillating salt concentrations triggered by microscale water cycles inside heated rock pores. *Angew. Chem. Int. Ed.* **2019,** *131*, 13289-13294.

46. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.

47. Banyay, M.; Sarkar, M.; Gräslund, A., A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **2003,** *104*, 477-488.

48. Peng, C. S.; Jones, K. C.; Tokmakoff, A., Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 15650-15660.

49. Lee, C.; Cho, M., Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **2006,** *125*, 114509.

50. Braunlin, W. H.; Bloomfield, V. A., Proton NMR study of the base-pairing reactions of d (GGAATTCC): salt effects on the equilibria and kinetics of strand association. *Biochem.* **1991,** *30*, 754-758.

51. Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A., Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* **2007,** *78*, 063101.

52. Jones, K. C.; Ganim, Z.; Peng, C. S.; Tokmakoff, A., Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* **2012,** *29*, 118-129.

53. Ashwood, B.; Jones, M. S.; Ferguson, A. L.; Tokmakoff, A., Disruption of Energetic and Dynamic Base Pairing Cooperativity in DNA Duplexes by an Abasic Site. *Proc. Natl. Acad. Sci.* **2022,** *10.1073/pnas.2219124120*.

54. Jones, M. S.; Ashwood, B.; Tokmakoff, A.; Ferguson, A. L., Determining Sequence-Dependent DNA Oligonucleotide Hybridization and Dehybridization Mechanisms Using Coarse-Grained Molecular Simulation, Markov State Models, and Infrared Spectroscopy. *J. Am. Chem. Soc.* **2021,** *143*, 17395-17411.

55. Wang, S.; Kool, E. T., Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2'-hydroxyl effects. *Biochem.* **1995,** *34*, 4125-4132.

56. Huang, Y.; Weng, X.; Russu, I. M., Enhanced base-pair opening in the adenine tract of a RNA double helix. *Biochem.* **2011,** *50*, 1857-1863.

57. Bernasconi, C., *Relaxation kinetics*. Academic Press: 1976.

58. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.

59. Chen, C.; Wang, W.; Wang, Z.; Wei, F.; Zhao, X. S., Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* **2007,** *35*, 2875-2884.

60. Cho, C.; Urquidi, J.; Singh, S.; Robinson, G. W., Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **1999,** *103*, 1991-1994.

61. Hori, N.; Denesyuk, N. A.; Thirumalai, D., Frictional effects on RNA folding: Speed limit and Kramers turnover. *J. Phys. Chem. B* **2018,** *122*, 11279-11288.

62. Neupane, K.; Hoffer, N. Q.; Woodside, M., Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* **2018,** *121*, 018102.

63. Fontana, F.; Bellini, T.; Todisco, M., Liquid Crystal Ordering in DNA Double Helices with Backbone Discontinuities. *Macromolecules* **2022,** *55*, 5946-5953.

64. Stone, S. E.; Ray, D.; Andricioaei, I., Force Field Dependent DNA Breathing Dynamics: A Case Study of Hoogsteen Base Pairing in A6-DNA. *J. Chem. Inf. Model.* **2022,** *62*, 6749-6761.

65. Pérez de Alba Ortíz, A.; Vreede, J.; Ensing, B., Sequence dependence of transient Hoogsteen base pairing in DNA. *PLOS Comput. Biol.* **2022,** *18*, e1010113.

66. Sawada, T.; Yoshizawa, M.; Sato, S.; Fujita, M., Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts. *Nat. Chem.* **2009,** *1*, 53-56.

67. Rejali, N. A.; Ye, F. D.; Zuiter, A. M.; Keller, C. C.; Wittwer, C. T., Nearest-neighbour transition-state analysis for nucleic acid kinetics. *Nucleic Acids Res.* **2021,** *49*, 4574-4585.

68. Szabo, A., Kinetics of hemoglobin and transition state theory. *Proc. Natl. Acad. Sci.* **1978,** *75*, 2108-2111.

69. Eaton, W. A.; Henry, E. R.; Hofrichter, J., Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin. *Proc. Natl. Acad. Sci.* **1991,** *88*, 4472-4475.

70. Oweida, T. J.; Kim, H. S.; Donald, J. M.; Singh, A.; Yingling, Y. G., Assessment of AMBER force fields for simulations of ssDNA. *J. Chem. Theory Comput.* **2021,** *17*, 1208-1217.

71. Tucker, M. R.; Piana, S.; Tan, D.; LeVine, M. V.; Shaw, D. E., Development of Force Field Parameters for the Simulation of Single-and Double-Stranded DNA Molecules and DNA–Protein Complexes. *J. Phys. Chem. B* **2022,** *126*, 4442-4457.

72. Roll, C.; Ketterlé, C.; Fazakerley, G. V.; Boulard, Y., Solution structures of a duplex containing an adenine opposite a gap (absence of one nucleotide) An NMR study and molecular dynamic simulations with explicit water molecules. *Eur. Biophys. J.* **1999,** *264*, 120-131.

73. Guo, H.; Tullius, T. D., Gapped DNA is anisotropically bent. *Proc. Natl. Acad. Sci.* **2003,** *100*, 3743-3747.

74. Zhang, W.; Tam, C. P.; Walton, T.; Fahrenbach, A. C.; Birrane, G.; Szostak, J. W., Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc. Natl. Acad. Sci.* **2017,** *114*, 7659-7664.

75. Zhang, W.; Tam, C. P.; Zhou, L.; Oh, S. S.; Wang, J.; Szostak, J. W., Structural rationale for the enhanced catalysis of nonenzymatic RNA primer extension by a downstream oligonucleotide. *J. Am. Chem. Soc.* **2018,** *140*, 2829-2840.

76. Zhang, W.; Walton, T.; Li, L.; Szostak, J. W., Crystallographic observation of nonenzymatic RNA primer extension. *Elife* **2018,** *7*, e36422.

77. Sim, A. Y.; Lipfert, J.; Herschlag, D.; Doniach, S., Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E* **2012,** *86*, 021901.

78. Chen, Z. Y.; Noolandi, J., Renormalization-group scaling theory for flexible and wormlike polymer chains. *J. Chem. Phys.* **1992,** *96*, 1540-1548.

79. Giurgiu, C.; Fang, Z.; Aitken, H. R.; Kim, S. C.; Pazienza, L.; Mittal, S.; Szostak, J. W., Structure–Activity Relationships in Nonenzymatic Template-Directed RNA Synthesis. *Angew. Chem. Int. Ed.* **2021,** *133*, 23107-23114.

80. Conte, M. R.; Conn, G. L.; Brown, T.; Lane, A. N., Conformational properties and thermodynamics of the RNA duplex r (CGCAAAUUUGCG) 2: comparison with the DNA analogue d (CGCAAATTTGCG) 2. *Nucleic Acids Res.* **1997,** *25*, 2627-2634.

81. Cruz-León, S.; Schwierz, N., RNA Captures More Cations than DNA: Insights from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2022,** *126*, 8646-8654.

82. Auffinger, P.; Westhof, E., Water and ion binding around r (UpA) 12and d (TpA) 12Oligomers-comparison with RNA and DNA (CpG) 12 duplexes. *J. Mol. Biol.* **2001,** *305*, 1057-1072.

83. Abels, J.; Moreno-Herrero, F.; Van der Heijden, T.; Dekker, C.; Dekker, N. H., Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* **2005,** *88*, 2737-2744.

84. Šulc, P.; Ouldridge, T. E.; Romano, F.; Doye, J. P.; Louis, A. A., Modelling toehold-mediated RNA strand displacement. *Biophys. J.* **2015,** *108*, 1238-1247.

85. Ding, D.; Zhou, L.; Giurgiu, C.; Szostak, J. W., Kinetic explanations for the sequence biases observed in the nonenzymatic copying of RNA templates. *Nucleic Acids Res.* **2022,** *50*, 35-45.

86. Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **2013,** *41*, 10641-10658.

87. Stootman, F. H.; Fisher, D. M.; Rodger, A.; Aldrich-Wright, J. R., Improved curve fitting procedures to determine equilibrium binding constants. *Analyst* **2006,** *131*, 1145-1151.

88. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with DNA duplex formation: salt-and sequence-dependent effects. *Biochem.* **2006,** *45*, 604-616.

89. Marky, L. A.; Breslauer, K. J., Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **1987,** *26*, 1601-1620.

90. Böttcher, A.; Kowerko, D.; Sigel, R. K., Explicit analytic equations for multimolecular thermal melting curves. *Biophys. Chem.* **2015,** *202*, 32-39.

91. van Stokkum, I. H.; Larsen, D. S.; Van Grondelle, R., Global and target analysis of time-resolved spectra. *Biochim. Biophys. Acta* **2004,** *1657*, 82-104.

92. Mullen, K. M.; Vengris, M.; van Stokkum, I. H., Algorithms for separable nonlinear least squares with application to modelling time-resolved spectra. *J. Glob. Optim.* **2007,** *38*, 201-213.

93. Lee, Y.; Gorka, M.; Golbeck, J. H.; Anna, J. M., Ultrafast energy transfer involving the red chlorophylls of cyanobacterial photosystem I probed through two-dimensional electronic spectroscopy. *J. Am. Chem. Soc.* **2018,** *140*, 11631-11638.

Appendix 9.A FTIR-monitored titrations of A_n binding at 1 °C

Short oligonucleotides (A_n, with n = 2,3,4) were titrated against gap and overhang templates in order to determine the fraction of bound A_n in a 1:1 mixture used for temperature-dependent measurements. Titrations were performed at 1 °C, where the fraction of bound primer and helper is assumed to be ~1, and A_n association can be modeled as a two-state equilibrium.

$$X + A_n \rightleftharpoons X:A_n \tag{9.A1}$$

Here X is the template and X:A_n is the bound A_n-template complex. Upon addition of A_n, a reduction in absorbance is observed near 1665 cm^{-1} with a corresponding gain centered near 1695 cm^{-1} (Fig. 9.A1). These changes come from thymine (T) and guanine (G) carbonyl bands and are consistent with an both an increase A:T base pairing and stacking of guanine nucleotides.[46] The second component from singular value decomposition (SVD) component over the frequency window from 1650 to 1720 cm^{-1} was used to describe the concentration of X:A_n and was fit to a general two-state binding expression (eq. 9.A2).[87]

$$[\mathrm{X{:}A_n}] = 0.5\left(K_\mathrm{d} + c_{\mathrm{A_n}} + c_\mathrm{Temp} - \sqrt{\left(K_\mathrm{d} + c_{\mathrm{A_n}} + c_\mathrm{Temp}\right)^2 - 4c_{\mathrm{A_n}} c_\mathrm{Temp}} \right) \tag{9.A2}$$

$$\text{where } c_{\mathrm{A_n}} = [\mathrm{X{:}A_n}] + [\mathrm{A_n}] \text{ and } c_\mathrm{Temp} = [\mathrm{X{:}A_n}] + [\mathrm{X}]$$

An initial measurement was made with $[A_n] = 0$, so the minimum of the 2nd SVD component was set to 0 prior to fitting. In practice, an amplitude scaling factor and K_d were used as fit parameters and the normalized fit was used to represent fraction of bound $X{:}A_n$.

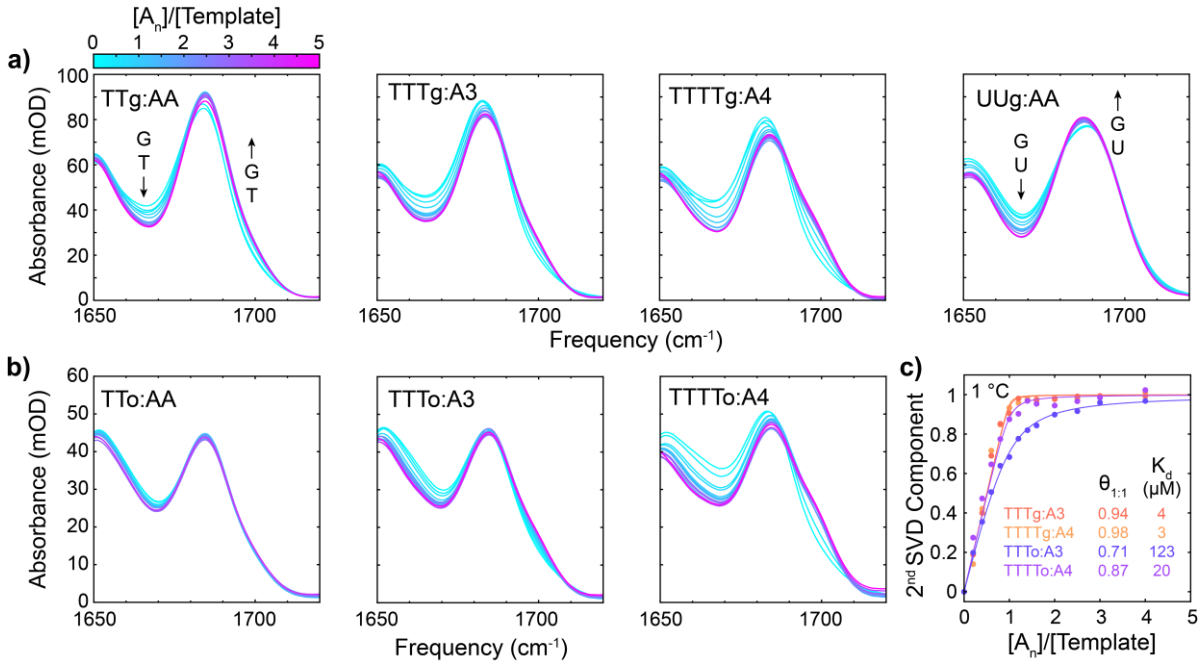

Figure 9.A1 FTIR-monitored titrations of A_n binding at 1 °C. (a) FTIR spectra as a function of molar ratio between A_n and their gap template. The template is at a concentration of 1 mM for all spectra. **(b)** FTIR spectra for titrations with overhang templates. **(c)** Titration curves for TTTg:A3, TTTTg:A4, TTTo:A3, and TTTTo:A4 determined from the second component of singular value decomposition (SVD) of FTIR-monitored titration data. SVD was performed over the 1650 to 1720 cm^{-1} spectral range. Solid lines correspond to fits to a two-state binding model (eq. 9.A2). Fraction of X bound to A_n at a 1:1 molar ratio ($\theta_{1:1}$) and dissociation constants (K_d) extracted from the fits are listed.

Table 9.A1 Thermodynamic parameters for A_n dissociation at 1 °C determined from FTIR-monitored titrations[a,c]

Complex	K_d (μM)	$\theta_{1:1}$ [b]
TTg:AA	12 ± 10	0.89 ± 0.06
TTTg:A3	4 ± 4	0.94 ± 0.07
TTTTg:A4	3 ± 3	0.98 ± 0.08
UUg:AA	147 ± 40	0.70 ± 0.07
TTo:AA	330 ± 150	0.54 ± 0.18
TTTo:A3	123 ± 38	0.71 ± 0.08
TTTTo:A4	20 ± 16	0.87 ± 0.11

[a]All values determined in deuterated pH* 6.8 400 mM SPB buffer with c_{Temp} = 1 mM. [b]Fraction of X bound to A_n at a 1:1 molar ratio. [c]Errors determined from 95% confidence intervals of titration curve fit parameters.

Appendix 9.B Three-state modeling of temperature series data

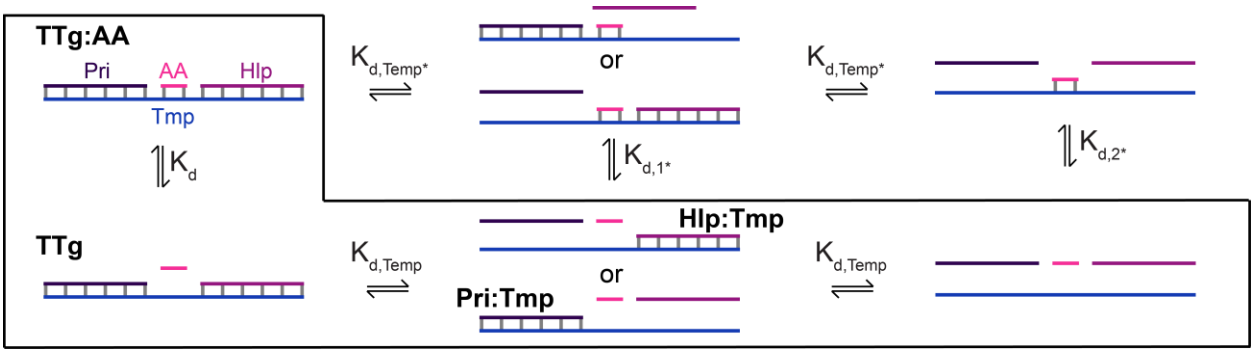

Figure 9.B1 Schematic of coupled binding equilibria for AA-template complexes. Example for TTg:AA. Direct binding of AA with the template strand (Tmp), primer:template complex, and helper:template complex is negligible due to the much greater binding stability of the primer (Pri) and helper (Hlp) relative to AA. The full unbinding of the complex proceeds first by AA dissociation (bottom-left) followed by dissociation of Pri and Hlp at higher temperatures.

The A_n-template complex (X:A_n) formed between a short oligonucleotide (A_n) and the gap or overhang template (X) exhibits two dissociation equilibria as a function of temperature. In general, there may be multiple thermodynamic paths for unbinding where either A_n, primer, or

helper strands dissociate first (Fig. 9.B1). For all complexes studied in this work, $K_d >> K_{d,Temp^*}$ & $K_{d,Temp2^*}$ over the measured temperature range due to use of G:C-rich 6-mer primer (Pri) and helper (Hlp) strands that bind to the template strand (Tmp) much more strongly than A_n. For the same reason, $K_{d,1^*}$ & $K_{d,2^*} >> K_{d,Temp1}$ & $K_{d,Temp2}$, and A_n binding to Tmp will be negligible population over the measured temperature range. Additionally, primer and helper strands have nearly identical binding stability such that $K_{d,Temp1} \approx K_{d,Temp2} \approx K_{d,Temp1^*} \approx K_{d,Temp2^*} = K_{d,Temp}$. Due to these conditions, we can treat the dissociation thermodynamics using just K_d & $K_{d,Temp}$ equilibria.

$$\mathrm{X{:}A}_n \overset{K_\mathrm{d}}{\rightleftharpoons} \mathrm{X} + \mathrm{A}_n \tag{9.B1a}$$

$$\mathrm{X} \overset{K_{d,\mathrm{Temp}}}{\rightleftharpoons} \mathrm{Pri{:}Tmp} + \mathrm{Hlp} \overset{K_{d,\mathrm{Temp}}}{\rightleftharpoons} \mathrm{Pri} + \mathrm{Hlp} + \mathrm{Tmp} \quad \text{(Gap)} \tag{9.B1b}$$

$$\mathrm{X} \overset{K_{d,\mathrm{Temp}}}{\rightleftharpoons} \mathrm{Hlp{:}Tmp} + \mathrm{Pri} \overset{K_{d,\mathrm{Temp}}}{\rightleftharpoons} \mathrm{Pri} + \mathrm{Hlp} + \mathrm{Tmp} \quad \text{(Gap)} \tag{9.B1c}$$

$$\mathrm{X} \overset{K_{\mathrm{d,Temp}}}{\rightleftharpoons} \mathrm{Pri} + \mathrm{Tmp} \quad \text{(Overhang)} \tag{9.B1d}$$

9.B.1 Independent model of dissociation

The A_n and primer and helper melting curves are typically well-separated in temperature ($K_d >> K_{d,Temp}$) for the sequences studied in this work. Therefore, it may be adequate to treat these processes as independent equilibria with the following equilibrium constants.

$$K_\mathrm{d} = \frac{[\mathrm{X}][\mathrm{A}_n]}{[\mathrm{X{:}A}_n]} \tag{9.B2a}$$

$$K_{\mathrm{d,Temp}} = \frac{[\mathrm{Pri{:}Tmp}][\mathrm{Hlp}]}{[\mathrm{X}]} = \frac{[\mathrm{Pri}][\mathrm{Tmp}]}{[\mathrm{Pri{:}Tmp}]} = \frac{[\mathrm{Pri}][\mathrm{Hlp{:}Tmp}]}{[\mathrm{X}]} = \frac{[\mathrm{Hlp}][\mathrm{Tmp}]}{[\mathrm{Hlp{:}Tmp}]} \text{ (Gap)} \quad (9.B2b)$$

$$K_{\mathrm{d,Temp}} = \frac{[\mathrm{Pri}][\mathrm{Tmp}]}{[\mathrm{X}]} \text{ (Overhang)} \quad (9.B2c)$$

A_n dissociation is described using a two-state model for non-self-complementary oligonucleotides where K_d is related to the fraction of bound A_n (θ_{A_n}).

$$K_{\mathrm{d}} = \frac{\left(1-\theta_{\mathrm{A}_n}\right)^2 c_{\mathrm{A}_n}}{2\theta_{\mathrm{A}_n}} \quad (9.B3)$$

$$\text{where, } \theta_{\mathrm{A}_n} = \frac{2[\mathrm{X{:}A}_n]}{c_1} \quad \text{and } c_1 = 2[\mathrm{X{:}A}_n] + [\mathrm{X}] + [\mathrm{A}_n]$$

Eq. 9.B3 assumes a 1:1 molar ratio between X and A_n, which is most valid at low temperature when [X] >> [Pri], [Hlp], [Tmp]. Then, eq. 9.B3 can be arranged to express θ_{A_n} in terms of K_d.

$$\theta_{\mathrm{A}_n} = 1 + \frac{K_{\mathrm{d}} - \sqrt{K_{\mathrm{d}}^2 + 2c_1 K_{\mathrm{d}}}}{c_1} \quad (9.B4)$$

The temperature-dependence of θ_{A_n} and K_d are determined by the enthalpy (ΔH_d°) and entropy (ΔS_d°) for dissociation of A_n.

$$K_{\mathrm{d}}(T) = \exp\left[-\frac{\Delta H_d^\circ}{RT} + \frac{\Delta S_d^\circ}{R}\right] \quad (9.B5)$$

The dissociation temperature (T_m), which is defined as the temperature where θ_{A_n}= 0.5, can be re-cast in terms of ΔH_d° and ΔS_d°.

$$T_m = \frac{\Delta H_d^\circ}{\Delta S_d^\circ - R\ln\left(c_1/4\right)} \tag{9.B6}$$

We neglect the temperature-dependence of ΔH_d° and ΔS_d°, which is set by the change in heat capacity (ΔC_p) between states. ΔC_p has previously been measured for dehybridization of DNA oligonucleotides to give a length-scaling of ~0.18 kJ mol^{-1} K^{-1} bp^{-1} but it also depends on sequence, ionic strength, and temperature.[88] It is also unclear whether coaxial stacking and changes in gap or overhang conformation following A_n unbinding will significantly change ΔC_p.

If the template is an overhang (eq. 9.B1c) rather than gap, then primer unbinding can be treated identically to dissociation of A_n.

$$\theta_{\text{Temp}} = 1 + \frac{K_{\text{d,Temp}} - \sqrt{K_{\text{d,Temp}}^2 + 2c_2 K_{\text{d,Temp}}}}{c_2} \tag{9.B7}$$

$$\text{where, } c_2 = [\text{X}] + [\text{Pri}] + [\text{Tmp}]$$

$$T_{m,\text{Temp}} = \frac{\Delta H_{d,\text{Temp}}^\circ}{\Delta S_{d,\text{Temp}}^\circ - R\ln\left(c_2/4\right)} \tag{9.B8}$$

Primer and helper dissociation from a gap template involves a network of bimolecular equilibria with nearly identical unbinding constants represented by $K_{d,Temp}$. Unbinding of primer and helper strands may be cooperative or non-cooperative. For completely non-cooperative unbinding, where unbinding of the primer and helper from the template strand are independent of one another, the

total fraction of bound primer and helper ($\theta_{Temp,NC}$) may be approximated as the product of each binding fraction.

$$\theta_{\mathrm{Temp,NC}} = \theta_{\mathrm{Temp}}^2 \tag{9.B9}$$

In the opposite limit of high cooperativity, the primer and helper may unbind from the template strand in an all-or-nothing fashion. In our case, [Tmp]=[Pri]=[Hlp], leading to:[89]

$$K_{\mathrm{d,TempC}} = \frac{\left(1-\theta_{\mathrm{Temp,C}}\right)^3 c_2^2}{9\theta_{\mathrm{Temp,C}}} \tag{9.B10}$$

$$\text{where, } \theta_{\mathrm{Temp,C}} = \frac{3[\mathrm{X}]}{c_2} \text{ and } c_2 = 3[\mathrm{X}] + [\mathrm{Pri}] + [\mathrm{Hlp}] + [\mathrm{Tmp}]$$

Arranging in terms of $\theta_{Temp,C}$ gives a cubic polynomial:

$$c_2^2\theta_{\mathrm{Temp,C}}^3 - 3c_2^2\theta_{\mathrm{Temp,C}}^2 + \left[3c_2^2 + 9K_{\mathrm{d,TempC}}\right]\theta_{\mathrm{Temp,C}} - c_2^2 = 0 \tag{9.B11}$$

$\theta_{Temp,C}$ may be solved for by reducing eq. 9.B11 to a depressed cubic polynomial and applying Cardano's method and Vieta's substitution.[90]

$$\theta_{\mathrm{Temp,C}} = \sqrt[3]{W} - \frac{p}{3\sqrt[3]{W}} \tag{9.B12a}$$

$$W = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \tag{9.B12b}$$

$$\text{where, } p = \frac{9K_{\mathrm{d,TempC}}}{c_2^2} \text{ and } q = -\frac{9K_{\mathrm{d,TempC}}}{c_2^2}$$

As for the bimolecular equilibria, the melting temperature ($T_{m,Temp,C}$) is defined as the temperature where $\theta_{Temp,C}$ is 0.5.

$$T_{m,\mathrm{TempC}} = \frac{\Delta H^{\circ}_{d,\mathrm{TempC}}}{\Delta S^{\circ}_{d,\mathrm{TempC}} - R\ln\left(c_2^2/36\right)} \tag{9.B13}$$

The non-cooperative approximation (eq. 9.B9) is likely suitable for the sequences studied in this work because the binding regions for primer and helper strands are separated by at least a two nucleotide gap. We also find that the primer and helper melting profile is identical for gap sizes of one to four nucleotides (Fig. 9.B3), supporting that coaxial stacking interactions between primer and helper strands are negligible.

9.B.2 Sequential model of dissociation

In reality, the K_d and $K_{d,Temp}$ equilibria are coupled to a degree depending on the sequence of each oligonucleotide. Coaxial stacking of the primer and helper with A_n stabilizes both X:A_n as well as binding of the primer and helper to the template strand, and A_n unbinding or primer and helper unbinding will reduce the binding stability of the other. Therefore, we can treat the processes in eqs. 9.B1abc and eqs. 9.B1a,d as sequential equilibria.

For sequential bimolecular equilibria, the concentration of each species follows a 4th order polynomial, and we find that the solution to this polynomial is not stable over the full temperature range that covers both melting transitions. Therefore we approximate $K_{d,Temp}$ as a unimolecular equilibrium between X and a "dissociated" template (dX).

$$\mathrm{X{:}A_n} \underset{}{\overset{K_\mathrm{d}}{\rightleftharpoons}} \mathrm{X} + \mathrm{A_n} \underset{}{\overset{K_\mathrm{d,Temp}}{\rightleftharpoons}} \mathrm{dX} + \mathrm{A_n} \tag{9.B14}$$

where each equilibrium constant may be expressed using a van't Hoff form as in eq. 9.B5.

$$K_{\mathrm{d}}(T)=\frac{[\mathrm{X}][\mathrm{A_n}]}{[\mathrm{X{:}A_n}]}=\exp\left[-\frac{\Delta H_d^{\circ}}{RT}+\frac{\Delta S_d^{\circ}}{R}\right] \tag{9.B15a}$$

$$K_{\mathrm{d,Temp}}(T)=\frac{[\mathrm{dX}]}{[\mathrm{X}]}=\exp\left[-\frac{\Delta H_{d,Temp}^{\circ}}{RT}+\frac{\Delta S_{d,Temp}^{\circ}}{R}\right] \tag{9.B15b}$$

The concentrations of each species can be determined in terms of the equilibrium constants as well as the total concentration of template strand (c_{Temp}) and $\mathrm{A_n}$ (c_{A_n}), which are known values and independently controlled in experiments.

$$c_{\mathrm{A_n}}=[\mathrm{X{:}A_n}]+[\mathrm{A}_n] \tag{9.B16a}$$

$$c_{\mathrm{Temp}}=[\mathrm{X{:}A_n}]+[\mathrm{X}]+[\mathrm{dX}] \tag{9.B16b}$$

By re-arranging eq. 9.B16b in terms of [dX] and using eqs. 9.B15ab, we can express [dX]:

$$[\mathrm{dX}]=c_{\mathrm{Temp}}-\frac{[\mathrm{dX}]}{K_{\mathrm{d,Temp}}}-\frac{[\mathrm{dX}][\mathrm{A_n}]}{K_{\mathrm{d}}K_{\mathrm{d,Temp}}} \tag{9.B17}$$

$[\mathrm{A_n}]$ is determined from equating $[\mathrm{X{:}A_n}]$ from eqs. 9.B16ab and using eq. 9.B15a.

$$[\mathrm{A}_n]=c_{\mathrm{A}_n}-c_{\mathrm{Temp}}+\frac{[\mathrm{dX}]}{K_{\mathrm{d,Temp}}}+[\mathrm{dX}] \tag{9.B18}$$

Substituting eq. 9.B18 into eq. 9.B17 and simplifying leads to a quadratic polynomial in terms of [dX].

$$\left(\frac{1+K_{\mathrm{d,Temp}}}{K_{\mathrm{d}}K_{\mathrm{d,Temp}}^{2}}\right)[\mathrm{dX}]^{2}+\left(1+\frac{K_{\mathrm{d}}+c_{\mathrm{A}_n}-c_{\mathrm{Temp}}}{K_{\mathrm{d}}K_{\mathrm{d,Temp}}}\right)[\mathrm{dX}]-c_{\mathrm{Temp}}=0 \tag{9.B19a}$$

where,

$$[\mathrm{dX}]=\frac{-K_{\mathrm{d,Temp}}\left(K_{\mathrm{d,Temp}}K_{\mathrm{d}}+K_{\mathrm{d}}-c_{\mathrm{Temp}}+c_{\mathrm{A}_n}\right)+K_{\mathrm{d,Temp}}\sqrt{\left(K_{\mathrm{d}}K_{\mathrm{d,Temp}}+K_{\mathrm{d}}+c_{\mathrm{A}_n}-c_{\mathrm{Temp}}\right)^{2}+4c_{\mathrm{Temp}}\left(1+K_{\mathrm{d,Temp}}\right)}}{2\left(1+K_{\mathrm{d,Temp}}\right)} \tag{9.B19b}$$

The fractions of intact X (θ_{Temp}) and X:A$_n$ (θ_{A_n}) are computed separately for each equilibrium and under the condition $c_{A_n} = c_{Temp}$, which is used for all temperature-dependent FTIR and NMR measurements.

$$\theta_{\mathrm{Temp}}=\frac{[\mathrm{X}]}{c_{\mathrm{Temp}}}=1+\frac{K_{\mathrm{d,Temp}}\left(K_{\mathrm{d,Temp}}K_{\mathrm{d}}+K_{\mathrm{d}}\right)-K_{\mathrm{d,Temp}}\sqrt{\left(K_{\mathrm{d}}K_{\mathrm{d,Temp}}+K_{\mathrm{d}}\right)^{2}+4c_{\mathrm{Temp}}\left(1+K_{\mathrm{d,Temp}}\right)}}{2\left(1+K_{\mathrm{d,Temp}}\right)} \tag{9.B20a}$$

$$\theta_{\mathrm{A}_n}=\frac{[\mathrm{X{:}A}_n]}{c_{\mathrm{A}_n}}=\frac{c_{\mathrm{Temp}}^{2}\left(1-\theta_{\mathrm{Temp}}\right)^{2}}{c_{\mathrm{A}_n}K_{\mathrm{d}}K_{\mathrm{d,Temp}}\left(K_{\mathrm{d,Temp}}+c_{\mathrm{Temp}}\left(1-\theta_{\mathrm{Temp}}\right)\right)} \tag{9.B20b}$$

The melting temperature for AA dissociation, T_m, is defined as the temperature where half of the A$_n$ is bound to the template.

$$\Delta G_d^{\circ}=-RT_m\ln\left(c_{\mathrm{A}_n}/4\right) \tag{9.B21}$$

9.C Global fitting of FTIR and 2D IR temperature series

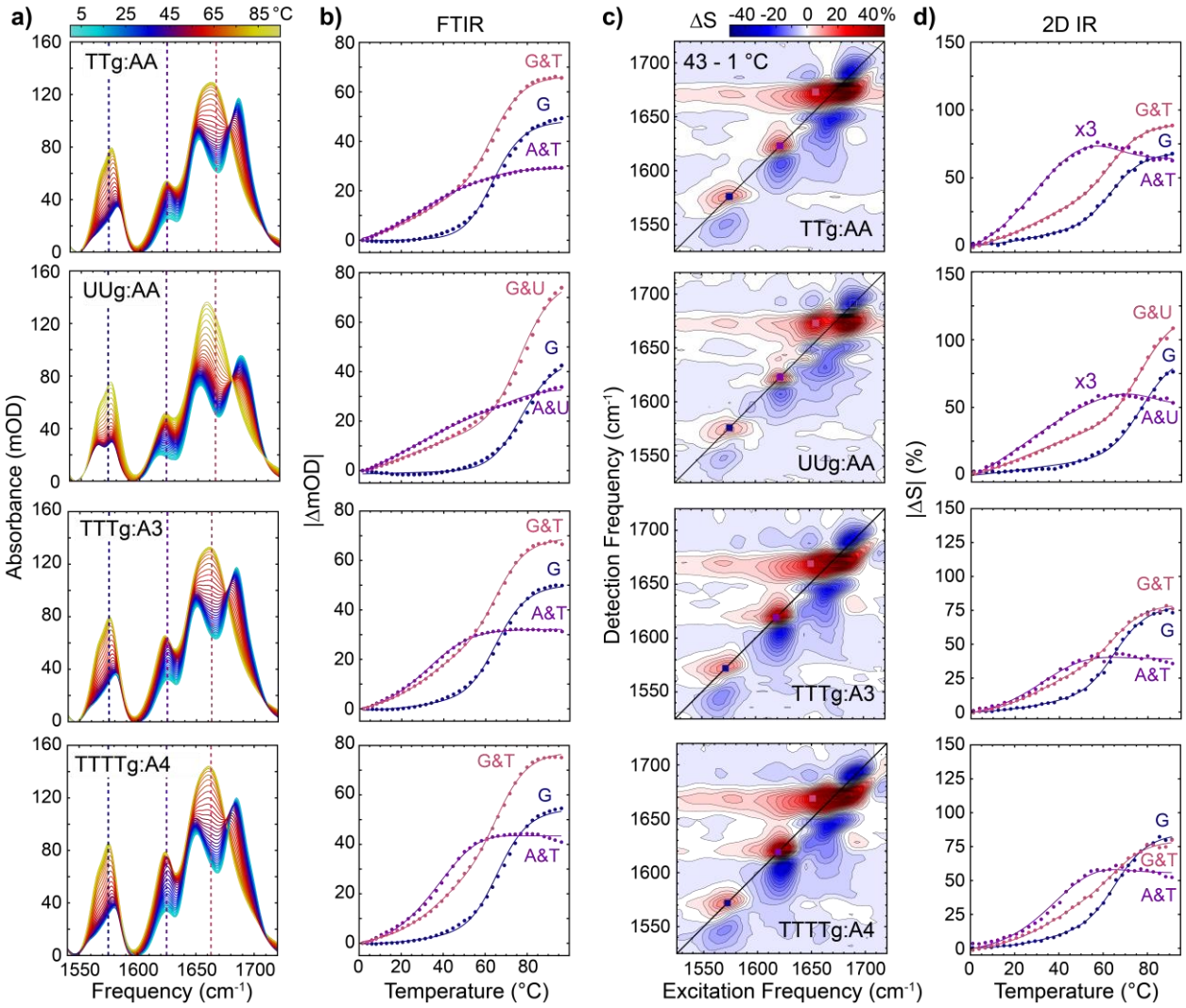

Figure 9.C1 Global fitting of temperature-dependent IR spectra for gap sequences. (a) FTIR spectra of TTg:AA, UUg:AA, TTTg:A3 and TTTTg:A4 from 1 to 96 °C in ~2.6 °C steps. **(b)** Temperature-dependent change in absorption relative to 1 °C for select frequencies. Frequencies are indicated in **(a)** with vertical dashed lines. **(c)** 2D IR difference spectra between 43 and 1 °C. Spectra are plotted in units of percentage change relative to maximum value of the 1 °C spectrum (ΔS). 25 contours with uniform 2% spacing are plotted for each sequence. **(d)** Temperature-dependent spectral change relative to the spectrum at 1 °C at select frequencies indicated by colored squares on the 2D IR spectra in **(c)**. Solid lines in **(b)** and **(d)** correspond to global fits to a three-state sequential model (eqs. 9.B20a,b).

Global fitting is applied to describe multiple melting components present in the FTIR and 2D IR temperature series (Figs. 9.2 and 9.C1). Global fitting describes the temperature-dependence at each frequency with a set of shared fitting parameters, which in this case are $\Delta H^{\circ}_{d,i}$ and $\Delta S^{\circ}_{d,i}$ for *i* components. The frequency- and temperature-dependent absorption data, $A(\omega,T)$, is described by the sum of amplitude-weighted melting components θ_{A_n} and θ_{Temp} plus an offset *c*.

$$A(\omega,T) = C(T)E(\omega)^t = \begin{bmatrix} \theta_{A_n}(T) \\ \theta_{\mathrm{Temp}}(T) \\ 1 \end{bmatrix} \begin{bmatrix} a(\omega) \\ b(\omega) \\ c(\omega) \end{bmatrix}^t \tag{9.C1}$$

Equation 9.C1 is expressed in a general form where θ_{A_n} and θ_{Temp} can be independent (eqs. 9.B4, 9.B9, and 9.B12) or dependent through a sequential model (eqs. 9.B20ab). The component spectra *a* and *b* correspond to difference FTIR and 2D IR spectra associated with A_n unbinding and primer and helper unbinding, respectively (Figs. 9.4 and 9.C2). The value of θ_{A_n} at 1 °C determined from FTIR titrations (Figs. 9.1 and 9.A1) was included in the fitting to constrain θ_{A_n}. This constraint is unnecessary for many sequences, but was particularly important to extract accurate melting curves for UUg:AA and TTo:AA.[5]

As is often done for global fitting of time-dependent spectra,[91-92] gradient-based alternating least squares is applied so that the objective function (*F*) to be optimized during fitting only contains the parameters of C.

$$F = \left\| \left(I - C(T,n)C(T,n)^{+} \right) \Psi \right\|_2 \tag{9.C2}$$

$$\text{where, } E(\omega)^T = C^+(T)\Psi$$

C^+ indicates the Moore-Penrose pseudoinverse of C. 2D IR spectra are reshaped from an m x n matrix, where m and n are the number of ω_1 and ω_3 points, respectively, to a vector of length m*n for fitting as described in previous application of global lifetime fitting to 2D spectra.[93] After fitting, the extracted component spectra are reshaped into an m × n matrix.

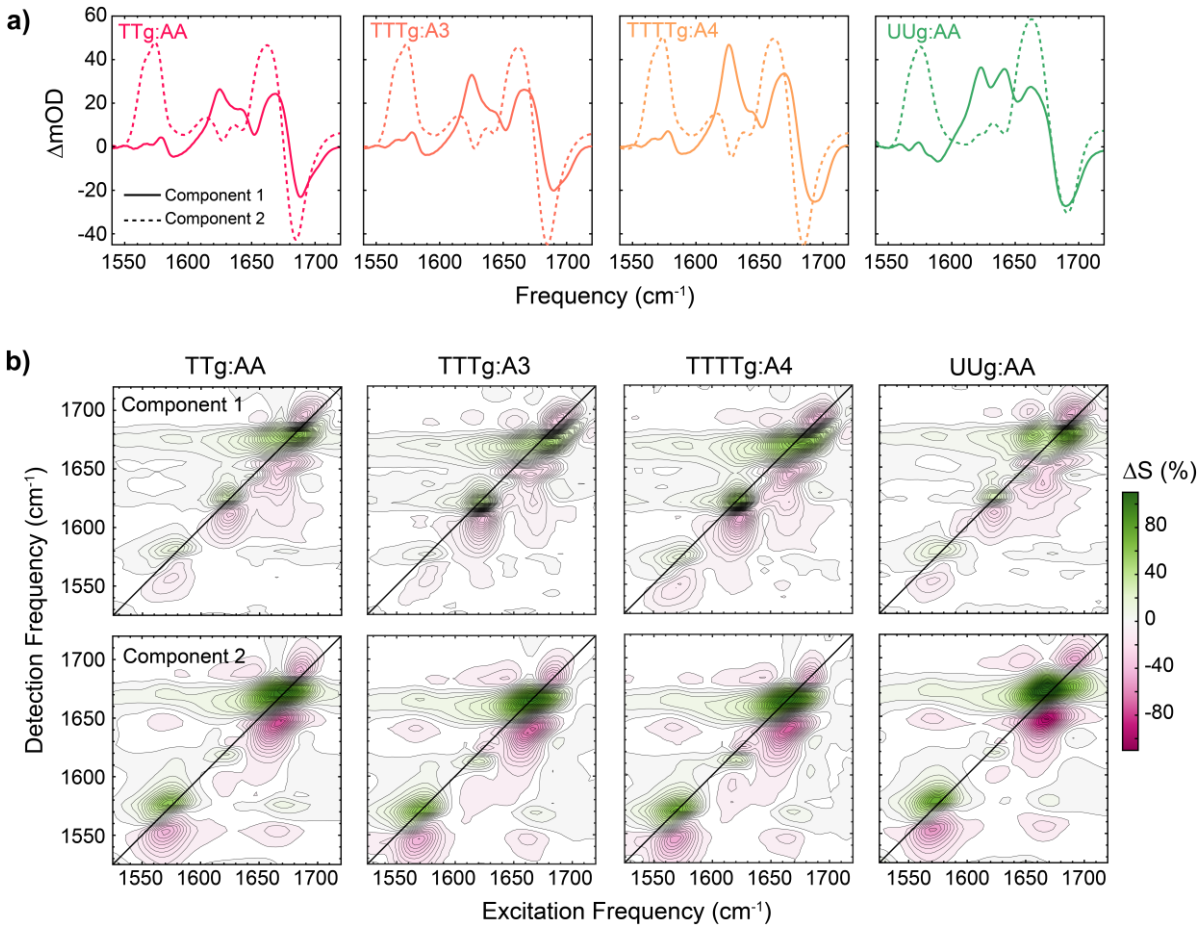

Figure 9.C2 FTIR and 2D IR component spectra from global fitting of A_n-gap complex temperature series. (a) (solid lines) first and (dashed lines) second FTIR spectral components extracted from global fitting of A_n-gap complex FTIR and 2D IR temperature series (Figs. 9.2 & 9.C1) to a three-state sequential model (eq. 9.B20a,b). **(b)** (top) first and (bottom) second 2D IR spectral components plotted in terms of signal percent change (ΔS). 25 contours with uniform 2% spacing are plotted for each sequence.