Temperature-Controlled, Single-Molecule Kinetic Investigations of Nucleic Acid Folding Motifs

by

David Alexander Nicholson

B.S., Furman University, 2014

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Chemical Physics
2022

Committee Members:
David Nesbitt, Chair
Joseph Falke
Meredith Betterton
Ralph Jimenez
Carl Lineberger

Nicholson, David A. (Ph.D. Chemical Physics)

Temperature-Controlled, Single-Molecule Kinetic Investigations of Nucleic Acid Folding Motifs

Thesis directed by Professor David J. Nesbitt

Nucleic acids serve many biological functions beyond information storage. These functions depend on the ability of a nucleic acid to adopt a specific spatial configuration in a process called folding. The folding of each nucleic acid sequence is unique, but there exist common folding patterns, motifs, that are shared by many nucleic acids. Folding typically occurs through a stepwise process in which motifs in a nucleic acid fold independently of each other before interacting to form higher order structures. The purpose of this dissertation is to provide a detailed account of the kinetic and thermodynamic properties of select motifs to give greater insight into how nucleic acids fold into functional forms.

The primary tool used in this work is single-molecule fluorescence resonance energy transfer (smFRET), which permits the observation of nucleic acid conformational dynamics in real-time and the concomitant determination of the first-order rate constants for folding and unfolding. Furthermore, performing kinetic measurements under controlled temperature enables the evaluation of the thermodynamic properties (e.g., enthalpy and entropy) of the folded and unfolded states, as well as of the transition state along the folding pathway. The nucleic acid motifs surveyed in this work are a DNA hairpin, an RNA tetraloop-tetraloop receptor, and an intramolecular DNA G-quadruplex. Temperature-controlled smFRET experiments on these motifs are used to probe the mechanism of motif folding, with a focus on the nature of charged ligand uptake during folding. Furthermore, the smFRET approach for studying conformational

dynamics has been developed herein to expand the range of measurable rate constants and to obtain higher order thermodynamic information on nucleic acids.

Acknowledgements

Though this dissertation's title page credits me as its sole author, there are in truth countless contributors who made this work possible by their immense support and guidance. I will now do my best to express my gratitude to those people, acknowledging that this limited space permits only abridged summaries of the full contributions of each.

Along the road to graduate school have been numerous educators who have prepared the way for me, and prepared me for the way. I found science to be a dull topic until my 8th grade science teacher Mr. Johnsen described light to the class and sparked a fascination that continues to this day. My high school chemistry instructor Dr. Veronica Price gave me the freedom to explore a topic of my own choosing in an open-ended end-of-year assignment which opened my eyes to the expansiveness of scientific knowledge, and, during the presentation for that assignment, corrected my pronunciation (of "denature") in the most kind and tactful way I have ever seen. My general physics instructor, Mr. Murray, showed me the power of mathematics in the natural sciences, as well as the importance of humor in education.

Numerous educators stand out from my undergraduate studies at Furman University, including Dr. Paul Wagenknecht (inorganic chemistry and photophysics), Dr. Lon Knight (physical chemistry), and Dr. Brian Goess and Dr. Greg Springsteen (bioorganic chemistry). Drs. Goess and Springsteen redefined my understanding of what chemistry education can be, and I hope one day to bring their style to my own classroom. I'd also like to thank several fantastic instructors, Dr. Ken Kolb (sociology) and Dr. Aaron Simmons (philosophy) for pushing my mind to its limits in the search for what truth is, as well as the chaplain Maria Swearingen for our conversations on what it means to be human.

Undergraduate research played a pivotal role in steering me toward majoring in chemistry. For this, I have extreme gratitude for Dr. Sandy Wheeler, who took pity on a freshman that lost the class enrollment lottery to get into General Chemistry II, and then, unprompted, recommended I apply for a summer research fellowship. That fellowship shaped the rest of my undergraduate career, starting by connecting me with Dr. Karen Buchmueller, my first mentor. Dr. Buchmueller instructed me in the fundamentals of laboratory research and gave me career advice during and after my time at Furman University, for which I am extremely grateful. I'd also like to thank my second research advisor, Dr. Jeffrey Petty, who showed me the power of biophysical chemistry and the beauty of single-molecule fluorescence, which had me instantly hooked. How any chemist can resist the allure of studying molecules one at a time is a mystery to me.

I am indebted to many people at the University of Colorado, and I'd like to express my gratitude to those who made my research possible from behind the scenes. Navigating graduate school can be difficult, which is why I'm thankful for the Chemistry graduate program managers I've had, Cora Fagan-Edminster, Otha Barrow, and Matt Bohn, as well as their assistant Kayla Jones. The Molecular Biophysics Training Program has had a tremendous impact on my graduate career, and it wouldn't exist without the tireless efforts of Dr. Joe Falke and Lin Pharris. The members of the JILA computing, electronics, and machine shops, as well as the metrology laboratory, were crucial in my work. I'd especially like to thank Corey Keasling for help with the compute cluster, J.R. Raith for building my data collection computer, David Alchenberger for training me on multiple instruments in the metrology lab, and Calvin Schwadron for teaching (and patiently re-teaching) me how to machine. The JILA administrators also made my life much easier, and I want to give special thanks to Kim Monteleone for making JILA a warmer, more

welcoming place, and Agnieszka Lynch for handling my exceedingly complicated tuition and payroll. I'd also like to thank the JILA supply office, especially Randall Holliness for handling my nucleic acid orders.

Not enough can be said to thank the current and past members of the Nesbitt group who have made enormous contributions to my graduate school experience. In particular, I'd like to thank Abhi Sungupta and Hsuan-Lei Sung for training me on the single-molecule confocal microscope, Jake Pettine for being the resident physicist of the single-molecule laboratory and therefore fielding all my physics questions, Tim Livingston-Large for his generosity with his time and his unmistakable laugh, Brian Riggs for his calming presence and the occasional jam session, Mia Zutz for early sage advice on surviving graduate school, and Dan Lesko for a handsome photo-shopped portrait. Special thanks go to those who worked on the biophysics project before me, including Julie Fiore and Erik Holmstrom, whose invaluable theses were lifesavers in my first year when I felt I was barely staying afloat as I struggled to understand my instrument. Last, I want to express my enormous appreciation for Andrea Marton Menendez who shaped my graduate student career in a way that no one else could have. I have gone to Andrea for advice on nearly every topic in my life, both in and out of the laboratory, about my pipetting or about my baking. I am especially grateful to Andrea for inviting me to "fumble forward" into reimagining what science and academia can be when kindness towards others is taken seriously.

My research advisor, David Nesbitt, is yet another person for whom I cannot fully express the extent of their impact on me. David's graduate students are his highest priority, and he shows this through his generosity with his time. The breadth of David's scientific knowledge is matched by few, and, ever the teacher, he freely shares his wisdom. He communicates with unimpeachable clarity of thought which serves as an example for all his students. Early on,

David recognized my unrefined mathematical passion and helped me to cultivate the “practical” mathematician in me. David taught me the importance of intuition, which I now understand is not divinely inspired guesswork but an engraving of hard-earned experience into the mind. I am inspired by his boundless curiosity, his enthusiasm for science, and his keen ability to evaluate scientific arguments for the “ring of truth.” From David, I have learned to be an ardent critic of my own work and to never be content using a tool that I understand only in part. As my research advisor, David gave me the freedom to pursue my interests at my own pace, and I’ll always be thankful for his mentorship and guidance.

The trials and tribulations of graduate school were made bearable with the help of my friends. I had a wonderful cohort of Chemistry graduate students, including Kristina Vrouwenvelder, Jenny Berry, Kangmin Kim, Shelby Beer, John Malecha, Emily Hopkins, Franklin Maharaj, Austin Cano, Alex Gilligan, Ryan Dill, Izzy Lattke, and Peyton Cline. During the especially challenging first year of difficult classwork and burdensome teaching load, my fellow cohort-mates lifted my spirits through a combination of mind-refreshing hikes, midnight study sessions at Denny’s, and dinner parties full of dancing and joy. My gratitude goes to Steven Sartor for the late-night ping pong and Andrew Kantor for the long-distance chess games. I also indebted to my dance partner Jackie Fleming, who gave me an important lesson in changing my mind. Finally, in the last years of my Ph.D., I’ve had the pleasure of working with and getting to know Bri Dobson and Max O’Connor, who renew my hope for a future that is just and equitable.

I could not be where I am today without the support of my family. My mother Lori has consistently encouraged me to follow my passion and is the highest example I can think of for unconditional love. My father John has always encouraged my curiosity about the world; he

introduced me the acronym “ATP” when I was ten years old, and in that moment started me down the path to a molecular understanding of life. My sister Becca inspired me to take up music, which has played a critical role in maintaining my mental wellbeing during my graduate career. My stepmother Joy is a paragon of generosity whom I can only hope to emulate.

Finally, I want to express my appreciation for my partner and the love of my life, Devon Reynolds. Devon brings unending joy and laughter into my life, and her support has helped me endure the difficult moments of graduate school. When my research feels stagnant or tedious, she expands my perspective and opens my mind to new horizons. When I obsess over an issue in my experiments, she pulls me back into balance by reminding me of the world beyond the lab. When I feel I can do nothing right in the lab, her love envelops me and heals my spirit. For these things and so much more, I am endlessly grateful.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Nucleic acids in biology

Nucleic acids are one of the main categories of polymers in biology, alongside proteins. Nucleic acids come in two flavors: ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). In high school biology classes, the functions of these three polymers (RNA, DNA, and proteins) are laid out in a tidy system of relationships in which DNA transfers information to RNA (transcription) which is then used by the ribosome to synthesize proteins (translation). In this so-called "central dogma" of molecular biology, DNA and RNA are purely information carriers: DNA serves as a long-term memory storage system that keeps a permanent blueprint (a gene) of each protein in the organism, and RNA is a transient messenger which delivers an accurate genetic transcript to the ribosome before being recycled for further messenger duties. In this simplified picture, DNA and RNA exist solely to create proteins, while proteins carry out the real work of the cell, such as catalyzing chemical reactions, orchestrating cellular transport, and providing cellular structure.

Today, DNA and RNA are known to serve cellular functions that go far beyond simply carrying genetic information. A key historical moment in the understanding of nucleic acids was the discovery by Thomas Cech and coworkers in the 1980s of the *tetrahymena* hammerhead intron which catalyzes its own cleavage without any assistance from proteins.[1-2] RNA molecules called ribozymes which perform catalysis are now widely documented across many organisms/domains of life.[3-4] Indeed, every living cell on earth relies on RNA in the ribosome to catalyze peptide bond formation required for protein synthesis.[5] This ability of RNA to act as

both genetic storage and catalyst inspired the development of the "RNA world" hypothesis which postulates that early life relied exclusively on RNA to perform all cellular functions, and that DNA and protein arrived later in evolutionary history.[6] In contemporary biology, RNA has been found to play a profound role in genetic regulation: the determination of whether or not genes are translated into protein.[7] RNA-based genetic regulation occurs through a variety of means. For instance, some organisms use *riboswitches*[8-9] which are short RNA sequences positioned at the start of an RNA transcript and that can bind to specific small molecules which may or may not be present in the cell. Then, depending on whether the small molecule is bound, the riboswitch will block or allow synthesis of the gene. In this way, the riboswitch acts as a sensor that detects a specific metabolite and determines how the cell will respond (e.g., by permitting the synthesis of a protein that metabolizes the riboswitch's target molecule). Other examples of genetic control by RNA include RNA interference,[10] which plays an important role in the eukaryote immune response, and transcript splicing[11] in which non-coding portions of nascent RNA transcripts called introns are removed before translation.

DNA, though not as diverse in biological functionality as RNA, also goes beyond mere information storage. For instance, DNA plays an essential role providing the structure of telomeres and centromeres, which are two large-scale structures in the chromosome.[12-13] Much of the interest in DNA functionality, however, stems from its potential use in biotechnology due to the improved biochemical stability of DNA over RNA. Over the last several decades, researchers have shown that DNA can reproduce most RNA capabilities: *deoxyribozymes* can catalyze reactions and DNA *aptamers* can bind to small molecules with high affinity and specificity.[14-15] Furthermore, the structural capacity of DNA has been exploited to create arbitrarily shaped DNA objects termed "DNA origami."[16-17] Clearly, the functional capability of DNA is firmly

established in the literature, and it is likely that even more nucleic acid functionalities will be discovered as the scientific community continues to explore the 98% of the human genome that does not encode for protein sequences.[18-19]

1.2 Nucleic acid structure

Understanding the many functions of nucleic acids requires first understanding the structure of nucleic acids. As previously stated, nucleic acids are polymers, meaning that they consist of monomers linked together to form a linear chain. In the case of nucleic acids, the monomers are nucleotides, and nucleotides are themselves composed of a ribose or deoxyribose sugar, for DNA or RNA respectively, a phosphate group, and a variable nucleobase (Figure 1.1).[20] In most nucleic acids, there are four possible nucleobases; in DNA, these are adenine, thymine, guanine, and cytosine, denoted as A, T, G, and C, respectively. In RNA, the same nucleobases are present except for thymine (T) which is replaced by the structurally similar uracil (U) that has a hydrogen at the C5 position instead of a methyl group. A, T, G, C, and U are the five so-called *canonical* nucleobases; however, there are several modified versions of these nucleobases that are regularly incorporated into nucleic acids in nature.[21-22] For instance, methylated nucleobases in DNA are part of an extremely common mode of epigenetic regulation.[23-24] The genetic versatility of DNA comes from the ability to freely swap among the 4 nucleobases, which in turn gives each position in a DNA sequence 2 bits of information. This means that the human genome, which is 3 billion nucleotides long, contains slightly less than 1 gigabyte of information (however, not all nucleotides in the human genome are highly conserved, so the true information content is smaller).

There are several important physical properties of nucleic acids that are relevant to this

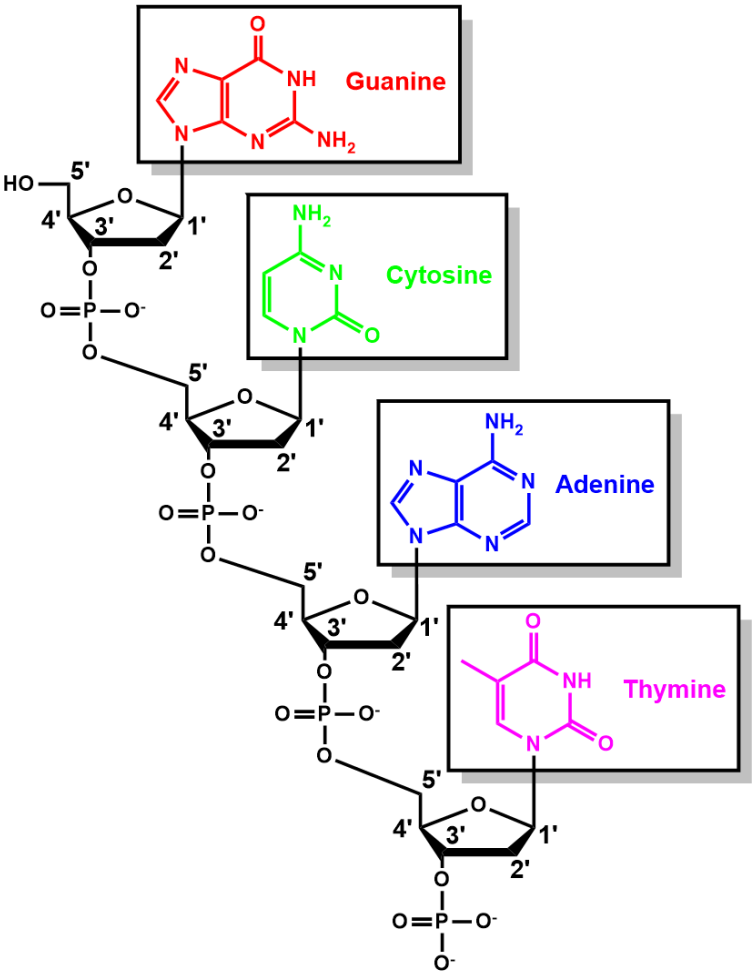

Figure 1.1 Chemical structure of DNA with highlighted nucleobases guanine, cytosine, adenosine, and thymine. In RNA, thymine is replaced uracil, and the 2' position in the sugar ring contains a hydroxyl (-OH) group.

dissertation. First, it is important to note that nucleic acid sequences are *polar*: the two ends of a nucleic acid are chemically distinct, and the nucleotide sequence is therefore directional. Specifically, the phosphodiester linkage between two adjacent nucleotides always, in nature, connects to the 3' carbon of one nucleotide to the 5' carbon of the other (Figure 1.1). Therefore, no matter how long a nucleic acid is, it will always have a 3' terminus and a 5' terminus, and any nucleic acid sequence (e.g., CAG) is ambiguous unless the polarity is specified (e.g., 5'-CAG-3' vs 3'-CAG-5'; it is conventional to write sequences in the 5' to 3' direction, which aligns with the direction of nucleic acid synthesis in nature). This *directionality* is chemically subtle but has

a significant impact on the ability of a nucleic acid to interact with itself or other nucleic acids. The second noteworthy physical property is that nucleic acids are highly electrically charged. Each nucleotide contributes a phosphate group and therefore a charge of $-1\ e$, where e is the elementary charge (1.6×10^{-19} C), and the total nucleic acid will therefore have a large, negative charge distributed evenly along its sugar-phosphate backbone. This net negative charge means that nucleic acids inherently repel one another, and that therefore any inter- or intra-strand interactions, which are the focus of this dissertation, would be impossible were it not for the presence of positive counterions such as K^+, Na^+, and Mg^{2+} which reduce the repulsive forces between phosphate groups through electrostatic screening.[25-26] The importance of nucleic acid-cation interactions cannot be understated and indeed forms a recurring theme in this dissertation.

At the heart of nucleic acid functionality is the ability of two individual strands to bind one another to form a duplex, mediated by a noncovalent interaction called *base pairing*. Base pairing occurs when two nucleobases form hydrogen bonds between them (Figure 1.2A). Due to the structural and geometric requirements of hydrogen bonding, not all possible base pairs have the same strength, with the strongest being the canonical Watson-Crick-Franklin base pairs: C pairing with G, and A pairing with T or U. Other non-canonical base pairs, such as Hoogsteen base pairs and wobble base pairs, are possible and do form in nature, but their stability is far less than that of the canonical CG and A(T/U) base pairs.[27-29] As a result of the specificity of base pairing, two nucleic acid strands are best able to bind to one another and form a duplex, or *hybridize*, if their sequences are *complementary*, meaning that they maximize the number of canonical base pairs formed between them. In addition, the geometry of the canonical base pairs requires the strands to be *antiparallel*, denoting that the direction of the 5' to 3' progression of one strand is the opposite of the other. In this optimal case, every adjacent nucleotide will be

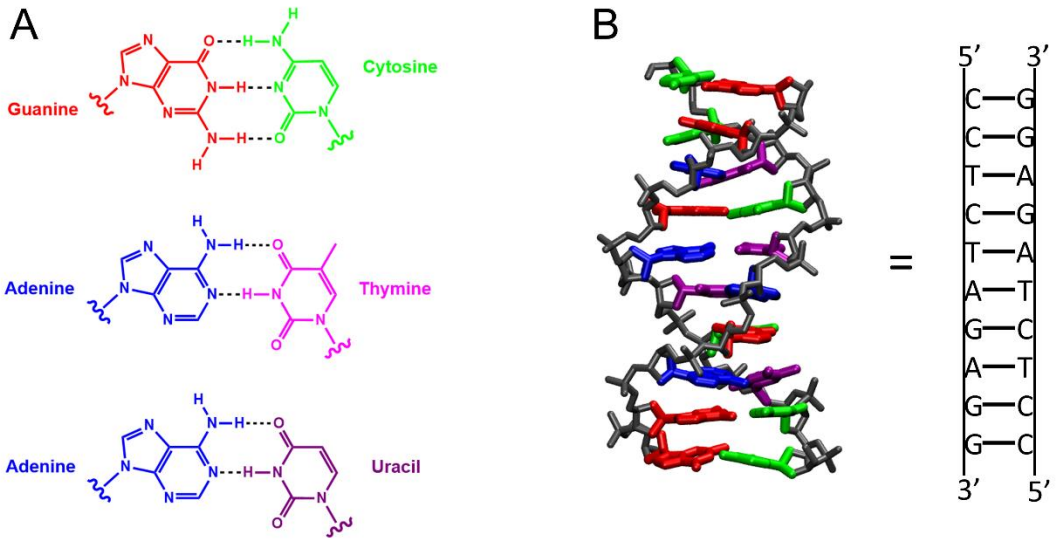

Figure 1.2 Nucleic acid base pairing. (A) Structures of the guanine-cytosine (GC), adenine-thymine (AT), and adenine-uracil (AU) base pairs. (B) Complementary, antiparallel nucleic acids can stack adjacent base pairs to form a double-helix. (PDB record 1ZEW)

part of a base pairing interaction, and the resulting structure is a ladder of base pairs. To minimize the surface area of the hydrophobic nucleobases exposed to water, this ladder then twists, forming the famous double helix (Figure 1.2B). Nucleic acid duplexes can take on several helical geometries with varying degrees of helical twist, including A-form (11 base pairs per turn, 2.6 A rise per base pair, right-handed helix), B-form (10 base pairs per turn, 3.4 A rise per base pair, right-handed helix), and Z-form (12 base pairs per turn, 3.7 A rise per base pair, left-handed helix) geometries.[30] DNA duplexes primarily populate the B-form conformation, whereas RNA duplexes and mixed RNA-DNA duplexes tend toward A-form conformations.

1.3 Intramolecular DNA/RNA folding: Hierarchy and motif

Beyond forming intermolecular duplexes, nucleic acids can also self-interact though intramolecular base-pairing to generate complex, three-dimensional structures capable of various biological functions, examples of which have been described above. The process by which DNA

or RNA forms these structures is known as *folding*,[31] a name which invokes the image of a nucleic acid bending back upon itself to create base pairs, like a rope bending around itself to form a knot. The terminology of folding is borrowed from the protein literature, where the "protein folding problem" has a long history,[32-34] and indeed the framework of nucleic acid folding that I am about to present draws many concepts from the field of protein folding.

First, nucleic acid structure can be divided into three levels: primary, secondary, and tertiary,[20] which are shown schematically in Figure 1.3. The primary structure of a nucleic acid is simply its nucleotide sequence, including any modified nucleobases. The secondary structure is the network of base pairs between nucleotides, which is often represented as a two-dimensional base pair map (e.g., Figure 1.3, middle). Finally, tertiary structure is the full, three-dimensional shape of a nucleic acid as determined by its *tertiary interactions*, which in turn are variously defined as (i) any interactions besides base pairing contributing to the overall structure or (ii) the

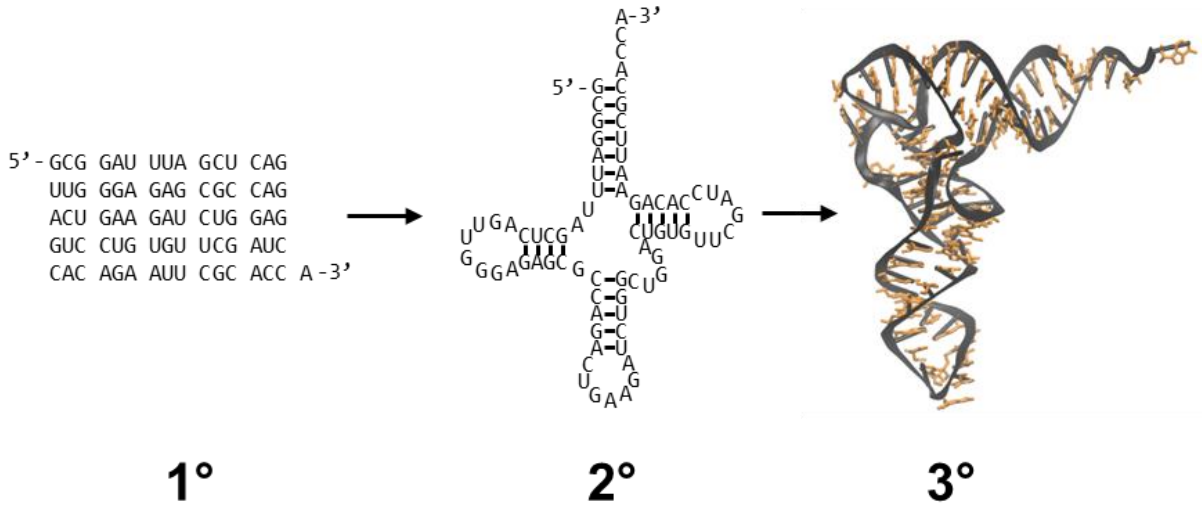

Figure 1.3 The three levels of nucleic acid structure (left to right): primary, secondary, and tertiary. Primary structure is the nucleic acid sequence, secondary structure is the base pairing, and tertiary structure is the 3-dimensional shape. Sequence shown is the yeast phenylalanine tRNA (PDB record 1EHZ).

interactions between secondary structure elements.[*] It is this latter definition that encapsulates the next development of nucleic acid folding theory: the modular nature of folding.

The three nucleic acid structural levels underlie a hierarchical model of nucleic acid folding in which information flow between the levels is one-directional: the primary structure (sequence) determines the secondary structure (hybridized regions) which in turn determine the tertiary structure.[35-37] This hierarchical approach is supported by experimental evidence that, in general, secondary structures form quickly, on a timescale of microseconds to milliseconds, and are extremely stable while tertiary interactions are slow to form (milliseconds to seconds) and much less stable. Therefore, nucleic acid folding is thought of sequentially as a rapid formation of secondary structures which then rearrange as tertiary interactions slowly emerge. Any conformational flexibility stems from the flighty tertiary interactions, while the stable core of the nucleic acid's structure is provided by immutable secondary structural elements. Of course, this rigid presentation of the hierarchical model is an overly simplistic picture. In truth, the interaction between the secondary and tertiary levels has some bidirectional character. For instance, the additional stabilization provided by tertiary interactions may be the energetic tiebreaker between two similarly stable secondary structures. Second, secondary structures are not the permanent, eternal, "rocks" that the hierarchical model supposes them to be. An example of this is shown in Chapter 3, where an intramolecular double helix with eight contiguous base pairs is observed to become completely unfolded multiple times per second. Similarly short or even shorter double helical regions are extremely common in complex nucleic acids, which

[*] Note that these definitions are heuristic in nature and do not form a logically complete set. For instance, how should we classify the kissing loop interaction? It brings together two secondary structure elements (the two stem-loops), so it seems to be a tertiary interaction, and it is usually described in the literature as such. However, the loops themselves interact by base pairing. So, kissing loops also meet the definition of a secondary structure. Another example is the pseudoknot, for which the literature is inconsistent as to whether pseudoknots are secondary or tertiary structures. This inconsistency can be confusing for those new to the field.

implies that the DNA/RNA can experience significant fluctuations in their secondary structure that may interfere or interact with tertiary interactions. Clearly, a static image of secondary structure fails to capture the true dynamical richness of base paired regions in nucleic acids.[†] With those caveats stated, the hierarchical model nevertheless provides a useful framework for understanding nucleic acid folding as a network of secondary structure elements called folding *motifs* that can be reliably predicted and individually characterized.[38-39]

Nucleic acid secondary structure motifs are numerous, and a few that are particularly important to this dissertation are sketched in Figure 1.4. The most basic and ubiquitous structure is the *hairpin* (Figure 1.4A), also known as a *stem-loop*, in which a nucleic acid bends backwards to position two complementary or near-complementary sequences in proximity for base pairing.[40-42] The base-paired region is the *stem*, and the leftover single-stranded region is the *loop*. The stability of a hairpin depends primarily on the number base pairs in the stem, with more base pairs providing more stability. The length of the loop matters as well, with a loop of fewer than three nucleotides being sterically impossible. Some secondary structure features derive from helical regions containing defects: an *internal loop* occurs in a stem with contiguous base pair mismatches, while a *bulge* or *kink* describes a stem where one strand has one or more unpaired, excess nucleotides (Figure 1.4B−D). If a hairpin has a long enough loop, the loop nucleotides can base pair with another portion of the nucleic acid, creating two overlapping stem-loops called a *pseudoknot*.[43] Finally, there are higher order structures that involve more than two strands, such as minor/major groove triplexes, four-way junctions, i-motifs, and G-quadruplexes.[44-49] G-quadruplexes are featured in the final work of this dissertation (Ch. 7) and

[†] In contrast, *protein* secondary structure is reliably quite stable. Protein *tertiary* structure, which are the interactions between amino acid side chains, may be a better comparison to nucleic acid secondary structure, as base pairing is an interaction between nucleobases, which are like the "side chains" of nucleotides Clearly, the parallelism between nucleic acid and protein folding is somewhat stretched.

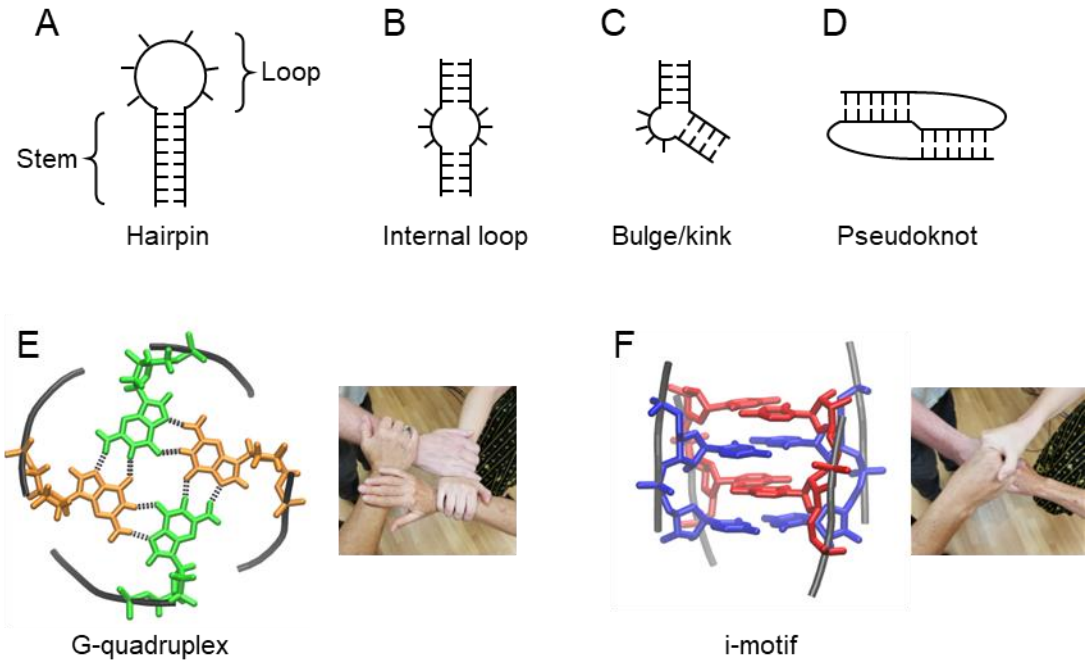

Figure 1.4 Common nucleic acid secondary structure motifs. (A) The hairpin consists of stem and loop, and variations on the stem-loop include (B) the internal loop, (C) the bulge/kink, and (D) the pseudoknot. 4-stranded nucleic acid helices include (E) the G-quadruplex and (F) the i-motif. In the G-quadruplex, 4 guanine residues form a planar "tetrad," similar to planar arrangement of hands in the "wrist-grab star" figure (E, right) in contra dancing, an American folk dance. In contrast, the i-motif consists of cytosine pairs which alternate along the helical axis, which is a configuration similar to the "hands-across star" contra dance figure (F, right).

are characterized by a unique guanine *tetrad* structure with hydrogen bonds occupying both the Watson-Crick-Franklin and Hoogsteen edges of each guanine (Figure 1.4E).

1.4 Nucleic acid folding dynamics

A nucleic acid may have access to many possible structures, or *conformations*, depending on its length and sequence. How then does a newly synthesized RNA molecule "choose" from among its possible conformations? And how quickly does the RNA molecule fold? Does it remain in that configuration indefinitely? These questions of stability and dynamics are addressed in this dissertation using two important theoretical frameworks, statistical thermodynamics and chemical kinetics, which I will now introduce.

1.4.1 The thermodynamics of folding

The field of thermodynamics, with its origins in the flow of heat in engines, may at first glance appear not to have any relationship with nucleic acid folding; however, the principles of classical thermodynamics laid down by Carnot and others can be extended into the molecular realm using the tools of statistical mechanics.[50] At the microscopic level, nucleic acids do in fact exchange minute quantities of heat with their environment during conformational rearrangements due to the making and breaking of bonds—not covalent bonds, but other noncovalent interactions like hydrogen bonding and ion-dipole interactions. These energy changes in turn play a key role in determining which configurations a nucleic acid will adopt.

A central result of statistical thermodynamics is that the states of any system, such as the conformations of a nucleic acid, exist in a dynamic equilibrium. Some states are populated more of the time than others, and the relative population of two states is given by the Boltzmann distribution,

$$\frac{P_2}{P_1} = e^{-\frac{\Delta G^\circ}{k_B T}}. \qquad \text{(Eq. 1.1)}$$

In this equation, P_1 and P_2 are the fraction of molecules in the states labeled 1 and 2, respectively, e is Euler's number, k_B is the Boltzmann constant, T is the absolute temperature, and ΔG° is the Gibbs free energy difference between the two states. This is a remarkably simple result that reduces any arbitrarily complex process down to a single intrinsic quantity: the free energy change ΔG° for the process. Importantly for this work, if an experiment can measure relative populations of states, then ΔG° can be determined. It is possible, therefore, to assign free energies the network of possible nucleic acid configurations and thereby construct the so-called free energy "landscape" for folding (Figure 1.5).[51] Furthermore, the Gibbs free energy change is itself composed of two competing thermodynamic contributions:

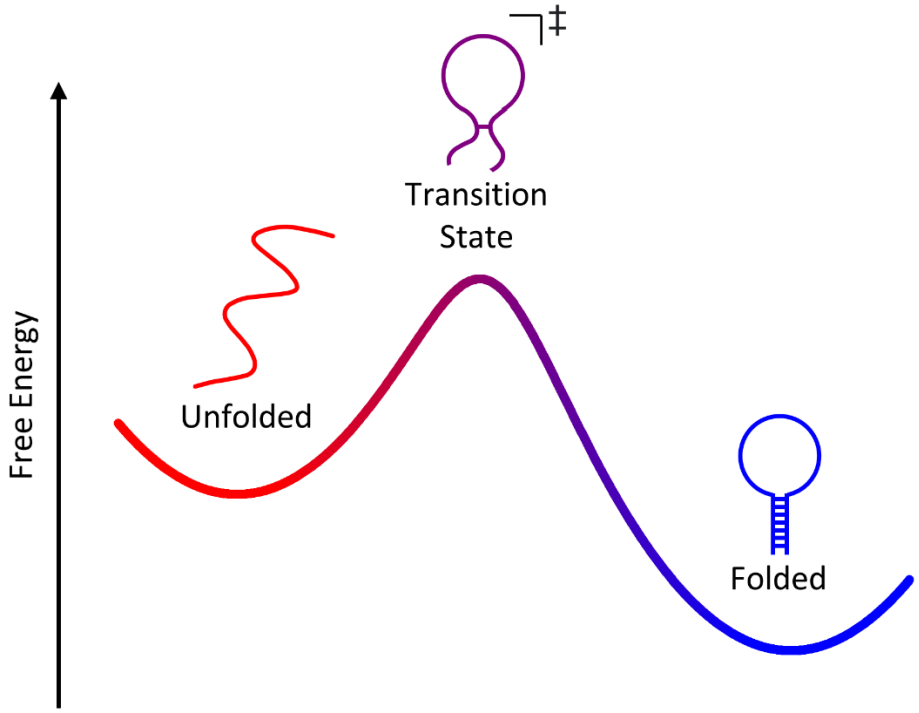

Figure 1.5 Schematic of free energy landscape for a nucleic acid hairpin.

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ, \qquad \text{(Eq. 1.2)}$$

where ΔH° enthalpy, or the heat absorbed, and ΔS° is the change in entropy, which is a measure of system disorder. Nucleic acid folding generally releases heat (ΔH° < 0) and is therefore enthalpically favorable, but the folded nucleic acid is usually conformationally constricted, which is entropically unfavorable (−TΔS° > 0).[52] This competition between enthalpy and entropy is a hallmark of folding[53-54] and is discussed considerably in the results (Chapter 3−7). Notable in Eq. 1.2 is the temperature *T*, which is a parameter that can be controlled in an experiment. In fact, by intentionally changing *T* and recording how ΔG° changes, one can determine the enthalpy ΔH° and the entropy ΔS° for nucleic acid folding, and this strategy is at the heart of much of this dissertation.[55]

1.4.2 The kinetics of folding

Chemical kinetics is the study of chemical reaction rates and, similarly to thermodynamics, may not obviously have connections with nucleic acid conformational dynamics. After all, there are no covalent chemical bonds being broken or formed during a conformational transition. However, just as two chemical reactants must randomly collide with the correct energy and orientation for a reaction, so too must a nucleic acid find the correct conditions for a conformational transition to occur via random collisions with its environment. Both chemical reactions and nucleic acid conformational transitions can be described as *Poisson* processes, meaning that the likelihood of the process happening is independent of time.[56] Therefore, these processes can be described by a single quantity, the *rate constant*, which is the probability of the process occurring per molecule per unit time. For instance, the folding transition from conformation *A* to *B* can be written down schematically as

$$A \xrightarrow{k_{AB}} B, \qquad \text{(Eq. 1.3)}$$

where k_{AB} is the rate constant for the *A* to *B* transition. The result of this kinetics approach is that the enormous complexity of nucleic acid structural dynamics is reduced to a single numerical value for each conformational transition.[57] Of course, the folding rate constant may depend on environmental factors, such as temperature or solute concentrations, and this can be used to make inferences about the *mechanism* of nucleic acid folding. For instance, if a folding rate constant changes upon an increase in the concentration of an ion in solution such as potassium, then it can be inferred that folding involves the binding or releasing of that ion from the nucleic acid. These types of environmental manipulations are used frequently in this work in as a way of proposing possible mechanisms for folding and testing the validity of putative folding mechanisms.

1.5 Single-molecule microscopy of nucleic acid conformational dynamics

Nucleic acids have largely been studied using *ensemble* methods: experiments in which many, many molecules are examined simultaneously. These "bulk" experiments have provided and continue to provide invaluable information on nucleic acid structure and conformational dynamics. For instance, it is now routinely possible to measure the precise position of individual atoms in nucleic acid structures using x-ray crystallography or nuclear magnetic resonance (NRM) spectroscopy[58-59]. The thermodynamics of nucleic acid folding (i.e., enthalpy and entropy) can be determined with high precision either directly by microcalorimetry in isothermal titration calorimetry (ITC) and differential scanning calorimetry (DSC) or indirectly by temperature-controlled ultraviolet-visible (UV-vis) spectroscopy.[60-62] These thermodynamic measurements have been integrated into "nearest neighbor" models for prediction of secondary structure thermodynamics,[52, 63] which in turn forms the basis of numerous software packages that can rapidly predict secondary structures for arbitrary nucleic acid sequences.[64]

In recent decades, it has become possible to observe nucleic acid folding one molecule at a time, and these single-molecule methods allow experimentalists to access information previously unavailable to ensemble measurements.[65-73] First, an ensemble experiment by its nature requires averaging over many molecules, and as a result, information on molecule-to-molecule differences, or heterogeneity, is lost.[74-75] The existence of subpopulations must be inferred out by indirect means. In contrast, observing molecules one at a time permits immediate characterization of sample heterogeneity and the presence of subpopulations. Second, and more importantly for this dissertation, most ensemble methods are incapable of reporting on the timescale for folding. This is because, at equilibrium, the numerous molecules in a bulk measurement are experiencing conformational transitions in an uncoordinated way, with an equal

number of molecules folding and unfolding at any given moment. As a result, the data from an ensemble experiment is typically static and unchanging. The exceptions to this are ensemble experiments that are not carried out at equilibrium, such as the temperature jump and the stopped-flow methods in which the system is rapidly displaced from equilibrium and the timescale of the relaxation to a new equilibrium is recorded.[76-78] However, converting relaxation times into rate constants is only possible under certain assumptions, such as that the folding equilibrium consisting of only two states, or that the dynamics far from equilibrium are representative of those at equilibrium. Single molecule experiments, on the other hand, are able to detect structural fluctuations in a system *at equilibrium* and are therefore well-suited to explore nucleic acid conformational dynamics, even in many-state systems.

1.5.1 Fluorescence resonance energy transfer (FRET)

At the heart of the single molecule method used in this dissertation is the photophysical process of fluorescence[‡] resonance energy transfer (FRET).[79] FRET begins with a "donor" fluorophore, which is a molecule that can absorb light of one color and then emit light of a different color. In photophysical terms, the donor absorbs a photon of light to enter an energetically "excited" state before releasing its excess energy ("relaxing") either (i) as light (fluorescence), (ii) as heat (non-radiative relaxation), or (iii) by transferring its energy to a nearby "acceptor" fluorophore. This last possibility is termed FRET, and the probability that FRET occurs (the "FRET efficiency" or E_{FRET}) depends on how large the rate of energy transfer (k_{FRET}) is compared to the rates for fluorescence (k_f) and non-radiative relaxation (k_{nr}):

[‡] In the literature, the FRET acronym is sometimes spelled out as "Forster resonance energy transfer" after Theodore Forster, the theorist who gave the first derivation of the physics behind FRET. In this dissertation, I prefer "fluorescence" over "Forster" for historical reasons: the FRET acronym was originally defined using "fluorescence," and it was only later that the acronym was redefined to include Forster's name. Additionally, while naming a phenomenon after someone is respectful, it is generally unhelpful to the neophyte, who I think would prefer the more descriptive term "fluorescence" to help them remember that FRET experiments are light-based measurements.

$$E_{FRET} = \frac{k_{FRET}}{k_f + k_{nr} + k_{FRET}}. \quad \text{(Eq. 1.4)}$$

Certain conditions must be met for FRET to occur. First, to conserve energy, the energy gap for donor fluorescence and acceptors absorbance must match. Second, as FRET originates from electric dipole-dipole interactions, the magnitude of k_{FRET} depends on the relative orientation of the donor and acceptor and the distance between them ($k_{FRET} \propto R^{-6}$).[§] Therefore, the E_{FRET} expression can be written as a function of distance:

$$E_{FRET}(R) = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}, \quad \text{(Eq. 1.5)}$$

where R is the donor-acceptor distance, and R_0 is a characteristic length named the Forster distance. R_0 depends on the properties of the donor and acceptor but is usually 1−10 nm, which is serendipitously the same length scale of nucleic acid movement in many conformational changes.[80] From Eq. 1.5, knowledge of E_{FRET} implies knowledge of the donor-acceptor distance, which is why FRET is sometimes referred to as a "nanoscopic ruler." The experimentalist can measure E_{FRET} by examining the fluorescence spectrum: when E_{FRET} is small, fluorescence comes from the donor only; however, when E_{FRET} is large, the donor's fluorescence is quenched, and fluorescence comes from the acceptor instead. Intermediate FRET efficiencies can be determined by the relative amount of donor fluorescence (F_D) and acceptor fluorescence (F_A):[**]

$$E_{FRET} = \frac{F_A}{F_A + F_D}. \quad \text{(Eq. 1.6)}$$

Therefore, by monitoring the fluorescence of the donor and acceptor, on can in principle determine the distance between the two fluorophores, a property which is utilized herein for nucleic acid structural studies.

[§] The full expression for k_{FRET} is derived in Appendix 1.

[**] Eq. 1.6 neglects several important factors that the experimentalist must contend with to accurately determine the FRET efficiency, including donor and acceptor differences in fluorescence quantum yields and detector efficiency. For a more thorough discussion, see Chapter 2.

Using FRET to measure inter-nucleotide distances in DNA and RNA, and by proxy the nucleic acid's conformation, requires specially designed nucleic acid "constructs" (Figure 1.6). First, as nucleic acids are not fluorescent, the nucleic acid of interest must be chemically modified to include a pair of fluorophores to act as FRET donor and acceptor. The selection of the fluorophores is critical, as a good FRET donor-acceptor pair must exhibit several properties, including strong light absorbance, high fluorescent quantum yields (probability of fluorescing versus non-radiatively relaxing), and a large spectral overlap between donor emission and acceptor absorbance.[81] The cyanine-based dyes Cy3 and Cy5 are a popular FRET pair and are used extensively in this work. Equally important are the labeling positions of the donor and acceptor on the nucleic acid, as the distance between the two fluorophores needs to change during the anticipated conformational transition(s). Structural information from the literature (e.g., x-ray crystallography) can assist in determining appropriate labeling positions, but occasionally trial and error is required as FRET efficiency prediction is still a developing field.[82]

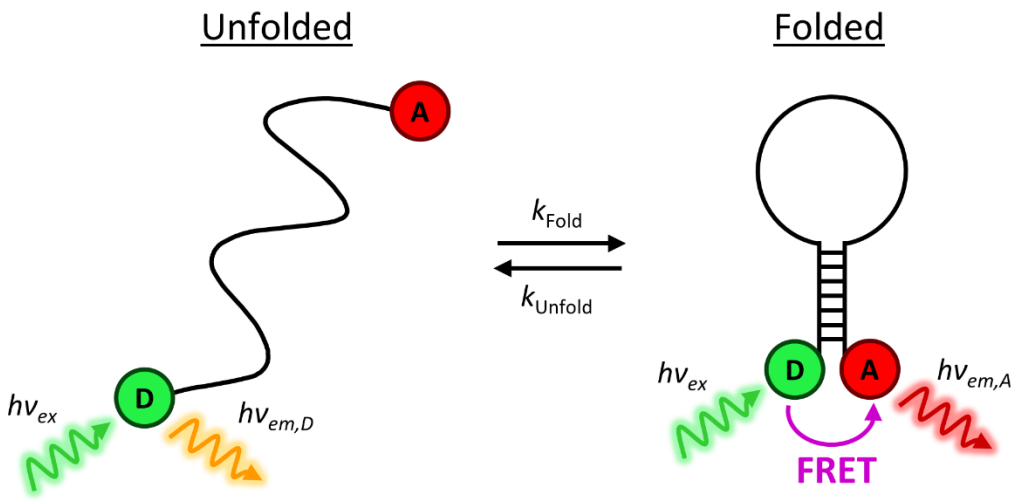

Figure 1.6 Design and use of nucleic acid constructs for conformational detection. The nucleic acid is labeled with donor (D) and acceptor (A) fluorophores. When the nucleic acid is unfolded (left), the donor and acceptor are far apart and excitation of the donor (hv_{ex}) leads only to donor emission ($hv_{em,D}$). However, when the nucleic acid is folded (right), the excited donor can transfer energy to the acceptor via FRET, which results in acceptor emission ($hv_{em,A}$).

Fortunately, the challenge of synthesizing nucleic acid constructs is made simpler through rapid and relatively inexpensive commercial sources that can generate labeled nucleic acids of arbitrary sequence up to a few hundred nucleotides in length ("oligonucleotides") with high purity.

1.5.2 Total internal reflection (TIR) and single-molecule fluorescence microscopy

Essential for performing single-molecule FRET (smFRET) experiments is the ability to measure fluorescence from individual molecules. Traditional, ensemble fluorescence measurements use light to excite many molecules simultaneously and observe the bulk emission. Translating ensemble fluorescence techniques into the single-molecule realm requires severely limiting the excitation of fluorophores so that only one is observed at a time. In this work, this is accomplished using the phenomenon of total internal reflection (TIR).[83]

TIR occurs when light intersects an interface at an angle greater than or equal to the critical angle, which is the incidence angle that produces a transmitted angle of 90° in Snell's law of refraction,

$$n_i \sin(\theta_i) = n_t \sin(\theta_t), \qquad \text{(Eq. 1.7)}$$

where n_i (n_t) and θ_i (θ_t) are the incident (transmitted) index of refraction and angle, respectively. Substituting $\theta_t = \pi/2$, the critical angle θ_c is

$$\sin(\theta_c) = \frac{n_t}{n_i}, \qquad \text{(Eq. 1.8)}$$

which has a solution only if $n_i > n_t$. Beyond this critical angle, no light is transmitted, and the light is said to totally internally reflect. However, the electromagnetic field at the interface cannot discontinuously go to zero, and a more thorough electromagnetic treatment reveals that the forbidden medium contains a so-called "evanescent wave" with an electric field magnitude that decays exponentially with distance from the interface (for more detail, see Chapter 2). The

length scale (1/*e*) of the evanescent decay is the same order of magnitude as the wavelength of the incident light, which in the case of visible light is less than 10^{-6} m. Therefore, while it is true under TIR conditions that no light is transmitted in the "far field," there is a "near field" effect which permits excitation of chromophores near the interface. This fact is exploited in a TIRF microscope by illuminating a glass-water interface under TIR conditions to obtain a fluorescence image with greatly reduced background fluorescence when compared to a typical widefield fluorescence microscope. This low-background, single-molecule fluorescence is extremely dim but can be detected using highly sensitive scientific cameras that record a "movie" documenting the fluorescence of each molecule in its field of view over time (Figure 1.7).

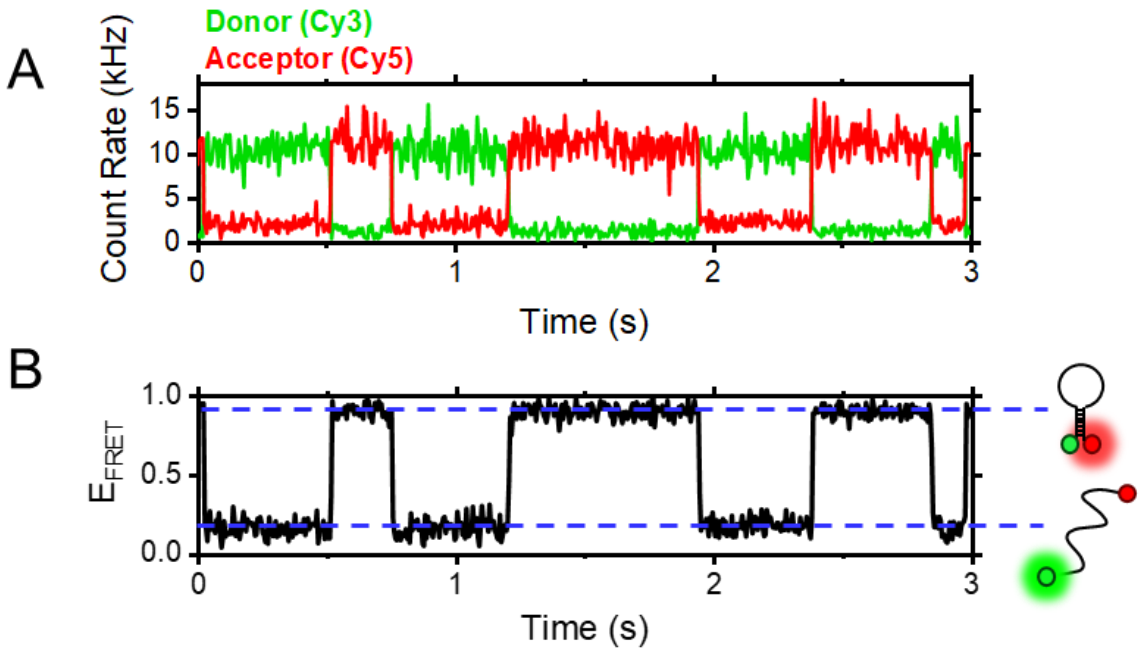

Figure 1.7 (A) Single-molecule fluorescence trajectory from a DNA hairpin labeled with Cy3 and Cy5 fluorophores, obtained at room temperature in buffer with 80 mM Na^+. The fluorescence shows alternative periods of high Cy3 fluorescence (green line) and high Cy5 fluorescence (red line). (B) FRET efficiency (E_{FRET}) trajectory calculated from fluorescence intensities using Eq. 1.6. When Cy3 emission is high, E_{FRET} is low. This implies that the distance between donor and acceptor is large, and the hairpin is in the unfolded state. Conversely, periods of high Cy5 emission correspond to high E_{FRET}, a small donor-acceptor distance, and occupancy of the folded hairpin state.

1.6 Thesis overview

In this work, I use temperature-controlled single-molecule FRET microscopy to investigate the kinetics and thermodynamics of a several fundamental nucleic acid folding motifs. In Chapter 2, I describe the single-molecule FRET microscope in detail as well as sample preparation, data acquisition, and data analysis. In Chapters 3 and 4, I probe the ability of amino acids to act as nucleic acid folding "chaperones", which has implications for the evolution of life.[34, 84] This work is split by structural level: Chapter 3 deals with amino acid effects on secondary structure using a DNA hairpin, while Chapter 4 looks at tertiary interactions via a tetraloop-tetraloop receptor.[85] Amino acids are found to have quite distinct effects on the secondary and tertiary levels, with tertiary structure being far more impacted by amino acids. Furthermore, tertiary structure is sensitive to amino acid *chirality*, and most of Chapter 4 is dedicated to exploring the origin of this chiral sensitivity. Chapters 5 and 6 are devoted to extending the methodology of smFRET measurements. Chapter 5 introduces the use of stroboscopic illumination as a means to extend the upper limit of rate constants determined by smFRET. Chapter 6 pushes the limits of temperature-controlled smFRET experiments to obtain not only the enthalpy and entropy of nucleic acid folding but also the change in *heat capacity* between two nucleic acid conformations.[86] Chapter 7 examines the influence of monovalent ions on the multistate folding dynamics of G-quadruplexes, where K^+ and Na^+ are found to exert control over G-quadruplex polymorphism through distinct mechanisms. Finally, Chapter 8 presents possible directions for future work based on preliminary results from several unfinished projects.

1.7 References

1. Cech, T. R., Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena (Nobel Lecture). *Angew. Chem., Int. Ed. Engl.* **1990,** *29* (7), 759-768.

2. Cech, T. R.; Zaug, A. J.; Grabowski, P. J., In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **1981,** *27* (3, Part 2), 487-496.

3. Müller, S.; Appel, B.; Balke, D.; Hieronymus, R.; Nübel, C., Thirty-five years of research into ribozymes and nucleic acid catalysis: Where do we stand today? *F1000Research* **2016,** *5*, F1000 Faculty Rev-1511.

4. Weinberg, C. E.; Weinberg, Z.; Hammann, C., Novel ribozymes: Discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.* **2019,** *47* (18), 9480-9494.

5. Cech, T. R., The ribosome Is a ribozyme. *Science* **2000,** *289* (5481), 878-879.

6. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* **2002,** *418* (6894), 214-21.

7. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nat. Rev. Genet.* **2014,** *15* (6), 423-37.

8. Sherwood, A. V.; Henkin, T. M., Riboswitch-mediated gene regulation: Novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.* **2016,** *70* (1), 361-374.

9. Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C., Engineering and in vivo applications of riboswitches. *Annu. Rev. Biochem.* **2017,** *86*, 515-539.

10. Agrawal, N.; Dasaradhi, P. V. N.; Mohmmed, A.; Malhotra, P.; Bhatnagar, R. K.; Mukherjee, S. K., RNA interference: Biology, mechanism, and applications. *Microbiol. Mol. Biol. Rev.* **2003,** *67* (4), 657-685.

11. Wilkinson, M. E.; Charenton, C.; Nagai, K., RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **2020,** *89* (1), 359-388.

12. Srinivas, N.; Rachakonda, S.; Kumar, R., Telomeres and telomere length: A general overview. *Cancers (Basel)* **2020,** *12* (3), 558.

13. Plohl, M.; Meštrović, N.; Mravinac, B., Centromere identity from the DNA point of view. *Chromosoma* **2014,** *123* (4), 313-325.

14. Silverman, S. K., Catalytic DNA: Scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.* **2016,** *41* (7), 595-609.

15. Adachi, T.; Nakamura, Y., Aptamers: A review of their chemical properties and modifications for therapeutic application. *Molecules* **2019,** *24* (23), 4229.

16. Dey, S.; Fan, C.; Gothelf, K. V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M. A. D.; Saccà, B.; Simmel, F. C., et al., DNA origami. *Nat. Rev. Methods Primers* **2021,** *1* (1), 13.

17. Yatsunyk, L. A.; Mendoza, O.; Mergny, J.-L., "Nano-oddities": Unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc. Chem. Res.* **2014,** *47* (6), 1836-1844.

18. Makalowski, W., The human genome structure and organization. *Acta Biochim. Pol.* **2001,** *48* (3), 587-98.

19. Palazzo, A. F.; Gregory, T. R., The case for junk DNA. *PLoS Genet.* **2014,** *10* (5), e1004351-e1004351.

20. Minchin, S.; Lodge, J., Understanding biochemistry: Structure and function of nucleic acids. *Essays Biochem.* **2019,** *63* (4), 433-456.

21. Carell, T.; Brandmayr, C.; Hienzsch, A.; Müller, M.; Pearson, D.; Reiter, V.; Thoma, I.; Thumbs, P.; Wagner, M., Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.* **2012,** *51* (29), 7110-7131.

22. Duffy, K.; Arangundy-Franklin, S.; Holliger, P., Modified nucleic acids: Replication, evolution, and next-generation therapeutics. *BMC Biol.* **2020,** *18* (1), 112.

23. Greenberg, M. V. C.; Bourc'his, D., The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **2019,** *20* (10), 590-607.

24. Handy, D. E.; Castro, R.; Loscalzo, J., Epigenetic modifications: Basic mechanisms and role in cardiovascular disease. *Circulation* **2011,** *123* (19), 2145-2156.

25. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **2014,** *83*, 813-41.

26. Pyle, A. M., Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* **2002,** *7* (7-8), 679-90.

27. Nikolova, E. N.; Zhou, H.; Gottardo, F. L.; Alvey, H. S.; Kimsey, I. J.; Al-Hashimi, H. M., A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* **2013,** *99* (12), 955-968.

28. Nagaswamy, U.; Larios-Sanz, M.; Hury, J.; Collins, S.; Zhang, Z.; Zhao, Q.; Fox, G. E., NCIR: A database of non-canonical interactions in known RNA structures. *Nucleic Acids Res.* **2002,** *30* (1), 395-397.

29. Reblova, K.; Spackova, N.; Stefl, R.; Csaszar, K.; Koca, J.; Leontis, N. B.; Sponer, J., Non-Watson-Crick basepairing and hydration in RNA motifs: Molecular dynamics of 5S rRNA loop E. *Biophys. J.* **2003,** *84* (6), 3564-82.

30. Mirkin, S. M., Discovery of alternative DNA structures: A heroic decade (1979-1989). *Front. Biosci.* **2008,** *13*, 1064-71.

31. Tinoco, I.; Bustamante, C., How RNA folds. *J. Mol. Biol.* **1999,** *293* (2), 271-281.

32. Dill, K. A.; MacCallum, J. L., The protein-folding problem, 50 years on. *Science* **2012,** *338* (6110), 1042-1046.

33. Englander, S. W.; Mayne, L., The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (45), 15873-15880.

34. Herschlag, D., RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **1995,** *270*, 20871-20874.

35. Brion, P.; Westhof, E., Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997,** *26*, 113-37.

36. Mustoe, A. M.; Brooks, C. L.; Al-Hashimi, H. M., Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* **2014,** *83* (1), 441-466.

37. Zarrinkar, P. P.; Williamson, J. R., Kinetic intermediates in RNA folding. *Science* **1994,** *265*, 918-924.

38. Leontis, N. B.; Lescoute, A.; Westhof, E., The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **2006,** *16* (3), 279-287.

39. Butcher, S. E.; Pyle, A. M., The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **2011,** *44* (12), 1302-1311.

40. Svoboda, P.; Cara, A. D., Hairpin RNA: A secondary structure of primary importance. *Cell. Mol. Life Sci.* **2006,** *63* (7), 901-908.

41. Tan, Z. J.; Chen, S. J., Salt dependence of nucleic acid hairpin stability. *Biophys. J.* **2008,** *95* (2), 738-52.

42. Kuznetsov, S. V.; Ren, C. C.; Woodson, S. A.; Ansari, A., Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* **2008,** *36* (4), 1098-112.

43. Staple, D. W.; Butcher, S. E., Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **2005,** *3* (6), e213-e213.

44. Chandrasekaran, A. R.; Rusling, D. A., Triplex-forming oligonucleotides: A third strand for DNA nanotechnology. *Nucleic Acids Res.* **2018,** *46* (3), 1021-1037.

45. Lilley, D. M. J., Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* **2001,** *33* (2), 109-159.

46. 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* (16), 8038-8056.

47. Dolinnaya, N. G.; Ogloblina, A. M.; Yakubovskaya, M. G., Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry (Moscow)* **2016,** *81* (13), 1602-1649.

48. Lyu, K.; Chow, E. Y.-C.; Mou, X.; Chan, T.-F.; Kwok, Chun K., RNA G-quadruplexes (rG4s): Genomics and biological functions. *Nucleic Acids Res.* **2021,** *49* (10), 5426-5450.

49. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S., The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020,** *21* (8), 459-474.

50. Garcia, H. G.; Kondev, J.; Orme, N.; Theriot, J. A.; Phillips, R., Thermodynamics of biological processes. In *Methods Enzymol.*, Johnson, M. L.; Holt, J. M.; Ackers, G. K., Eds. Academic Press: 2011; Vol. 492, pp 27-59.

51. Röder, K.; Wales, D. J., The energy landscape perspective: Encoding structure and function for biomolecules. *Front. Mol. Biosci.* **2022,** *9*.

52. SantaLucia, J., Jr.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-40.

53. Fox, J. M.; Zhao, M.; Fink, M. J.; Kang, K.; Whitesides, G. M., The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annu. Rev. Biophys.* **2018,** *47* (1), 223-250.

54. Dunitz, J. D., Win some, lose some: Enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995,** *2* (11), 709-712.

55. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* **2016,** *67*, 441-65.

56. Gillespie, D. T., Stochastic chemical kinetics. In *Handbook of Materials Modeling: Methods*, Yip, S., Ed. Springer Netherlands: Dordrecht, 2005; pp 1735-1752.

57. Chen, S. J., RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **2008,** *37*, 197-214.

58. Egli, M., Diffraction techniques in structural biology. *Curr. Protoc. Nucleic Acid Chem.* **2016,** *65* (1), 7.13.1-7.13.41.

59. Ferentz, A. E.; Wagner, G., NMR spectroscopy: A multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **2000,** *33* (1), 29-65.

60. Jelesarov, I.; Bosshard, H. R., Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **1999,** *12* (1), 3-18.

61. Johnson, C. M., Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.* **2013,** *531* (1-2), 100-109.

62. Lee, H.-T.; Carr, C.; Siebler, H.; Waters, L.; Khutsishvili, I.; Iseka, F.; Domack, B.; Olsen, C. M.; Marky, L. A., A thermodynamic approach for the targeting of nucleic acid structures using their complementary single strands. In *Methods Enzymol.*, Johnson, M. L.; Holt, J. M.; Ackers, G. K., Eds. Academic Press: 2011; Vol. 492, pp 1-26.

63. SantaLucia, J., Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95* (4), 1460-5.

64. Lorenz, R.; Wolfinger, M. T.; Tanzer, A.; Hofacker, I. L., Predicting RNA secondary structures from sequence and probing data. *Methods* **2016,** *103*, 86-98.

65. Elf, J.; Barkefors, I., Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* **2019,** *88* (1), 635-659.

66. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* **2018,** *359* (6373), 288.

67. Shen, H.; Tauzin, L. J.; Baiyasi, R.; Wang, W.; Moringo, N.; Shuang, B.; Landes, C. F., Single particle tracking: From theory to biophysical applications. *Chem. Rev.* **2017,** *117* (11), 7331-7376.

68. Sasmal, D. K.; Pulido, L. E.; Kasal, S.; Huang, J., Single-molecule fluorescence resonance energy transfer in molecular biology. *Nanoscale* **2016,** *8* (48), 19928-19944.

69. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **2015,** *34*, 52-59.

70. Robinson, A.; van Oijen, A. M., Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol.* **2013,** *11* (5), 303-315.

71. Barkai, E.; Garini, Y.; Metzler, R., Strange kinetics of single molecules in living cells. *Phys. Today* **2012,** *65* (8), 29-35.

72. van Oijen, A. M., Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Curr. Opin. Biotechnol.* **2011,** *22* (1), 75-80.

73. Aleman, E. A.; Lamichhane, R.; Rueda, D., Exploring RNA folding one molecule at a time. *Curr. Opin. Chem. Biol.* **2008,** *12* (6), 647-654.

74. Greenfeld, M.; Solomatin, S. V.; Herschlag, D., Removal of covalent heterogeneity reveals simple folding behavior for P4-P6 RNA. *J. Biol. Chem.* **2011,** *286*, 19872-19879.

75. Solomatin, S. V.; Greenfeld, M.; Herschlag, D., Implications of molecular heterogeneity for the cooperativity of biological macromolecules. *Nat. Struct. Mol. Biol.* **2011,** *18* (6), 732-734.

76. Kubelka, J., Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem. Photobiol. Sci.* **2009,** *8* (4), 499-512.

77. Dyer, R. B.; Brauns, E. B., Laser-induced temperature jump infrared measurements of RNA folding. *Methods Enzymol.* **2009,** *469*, 353-72.

78. Zheng, X.; Bi, C.; Li, Z.; Podariu, M.; Hage, D. S., Analytical methods for kinetic studies of biological interactions: A review. *J. Pharm. Biomed. Anal.* **2015,** *113*, 163-180.

79. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* **2008,** *5* (6), 507-16.

80. Preus, S.; Wilhelmsson, L. M., Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem* **2012,** *13* (14), 1990-2001.

81. Ha, T.; Tinnefeld, P., Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* **2012,** *63*, 595-617.

82. Steffen, F. D.; Sigel, R. K. O.; Börner, R., FRETraj: Integrating single-molecule spectroscopy with molecular dynamics. *Bioinformatics* **2021,** *37* (21), 3953-3955.

83. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

84. Rajkowitsch, L.; Chen, D.; Stampfl, S.; Semrad, K.; Waldsich, C.; Mayer, O.; Jantsch, M. F.; Konrat, R.; Blasi, U.; Schroeder, R., RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **2007,** *4* (3), 118-30.

85. Fiore, J. L.; Nesbitt, D. J., An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* **2013,** *46* (3), 223-64.

86. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with nucleic acid folding. *Biopolymers* **2006,** *82* (1), 38-58.

Chapter 2

Methods

2.1 Single-molecule TIRF microscope

Single-molecule measurements requires a specialized microscope capable of recording the minute fluorescence from individual molecules. The single-molecule measurements presented in this dissertation were primarily obtained using a total-internal reflection fluorescence (TIRF) microscope. The exception to this is the first results chapter (Chapter 3) in which a confocal microscope was used while I was building the TIRF microscope. Both the TIRF and confocal microscopes are given brief descriptions in our papers,[1-4] but the details are compact and sparse. Here, I will elaborate on the design and operation of the TIRF microscope. For a more thorough description of the confocal apparatus, see the excellent theses of Dr. Julie Fiore and Dr. Eric Holmstrom.

2.1.1 TIRF vs. confocal microscopy for single-molecule experiments

Before going over the technical details of the TIRF microscope, it is worth noting the differences between TIRF and confocal microscopy for single-molecule measurements, as these are both excellent methods which have complementary uses. These two flavors of single-molecule microscopy are named for the different type of illumination they use: in confocal microscopy, light is focused to a diffraction limited spot in confocal microscopy, while in TIRF, an evanescent field is created through total internal reflection. However, the capabilities and limitations of these methods differ due to the different detector types in each. A single-molecule confocal instrument contains a single-photon counting detector, usually an avalanche photodiode (APD), which responds to single-photon impingement on the detector by emitting an electrical

current pulse for each photon detected. With the appropriate auxiliary equipment, these current pulses can be recorded on a computer as a sequence of individual photon detection events with high absolute time precision (typically ≈ 50 ns due to APD dead time). This excellent time resolution has allowed confocal microscopes to measure dynamics on the microsecond timescale. However, this impressive speed comes at the cost of low throughput: APDs can only be used to observe one molecule at a time. This is a small price to pay when the system of interest has fast dynamics, but for a system with slow dynamics, the experimentalist can be required to spend many hours, or days, to acquire the necessary amount of data for analysis. A TIRF microscope instead uses a charge-coupled device (CCD) which consists of a grid of independent light detectors called pixels. Because each pixel measures light independently, the CCD detector can be used to observe many fluorescent molecules simultaneously, typically 100−1000. This greatly increases the rate of data acquisition over the confocal arrangement. However, digitizing the CCD's many pixels for computer storage is a time-consuming process which can only be done about once per millisecond with state-of-the-art CCD cameras.[*] Any dynamics faster than this "frame rate limit" are lost to the experimentalist. Therefore, the confocal method has better time resolution, while the TIRF method has higher throughput. This tradeoff of time resolution vs. throughput means that neither confocal nor TIRF microscopes are the "silver bullet" which is best for all circumstances, and that they are instead complementary approaches.

2.1.2 TIRF microscope overview: Laser excitation and fluorescence detection

This section will provide an overview of the flow of information in the TIRF microscope, from excitation to detection. Specific details for alignment of optics and system characterization

[*] The ever-evolving nature of CCD technology means that this digitization time will likely be made even shorter in the future. In addition, there are ways to increase a CCD's acquisition rate by decreasing the number of pixels to be digitized (see section 2.1.4).

are provided in Sections 2.1.3−5. The light source in the TIRF microscope is a 532 nm frequency-doubled, continuous wave (CW) Nd:YAG laser with 200 mW power. Light at this wavelength allows for efficient excitation of Cy3 (donor) molecules with little Cy5 (acceptor) excitation ($\sigma_{Cy5}/\sigma_{Cy3}$ = 3% at 532 nm). The raw laser output must be manipulated before going into the microscope, as schematically shown in Figure 2.1A. First, residual 1064 nm light is

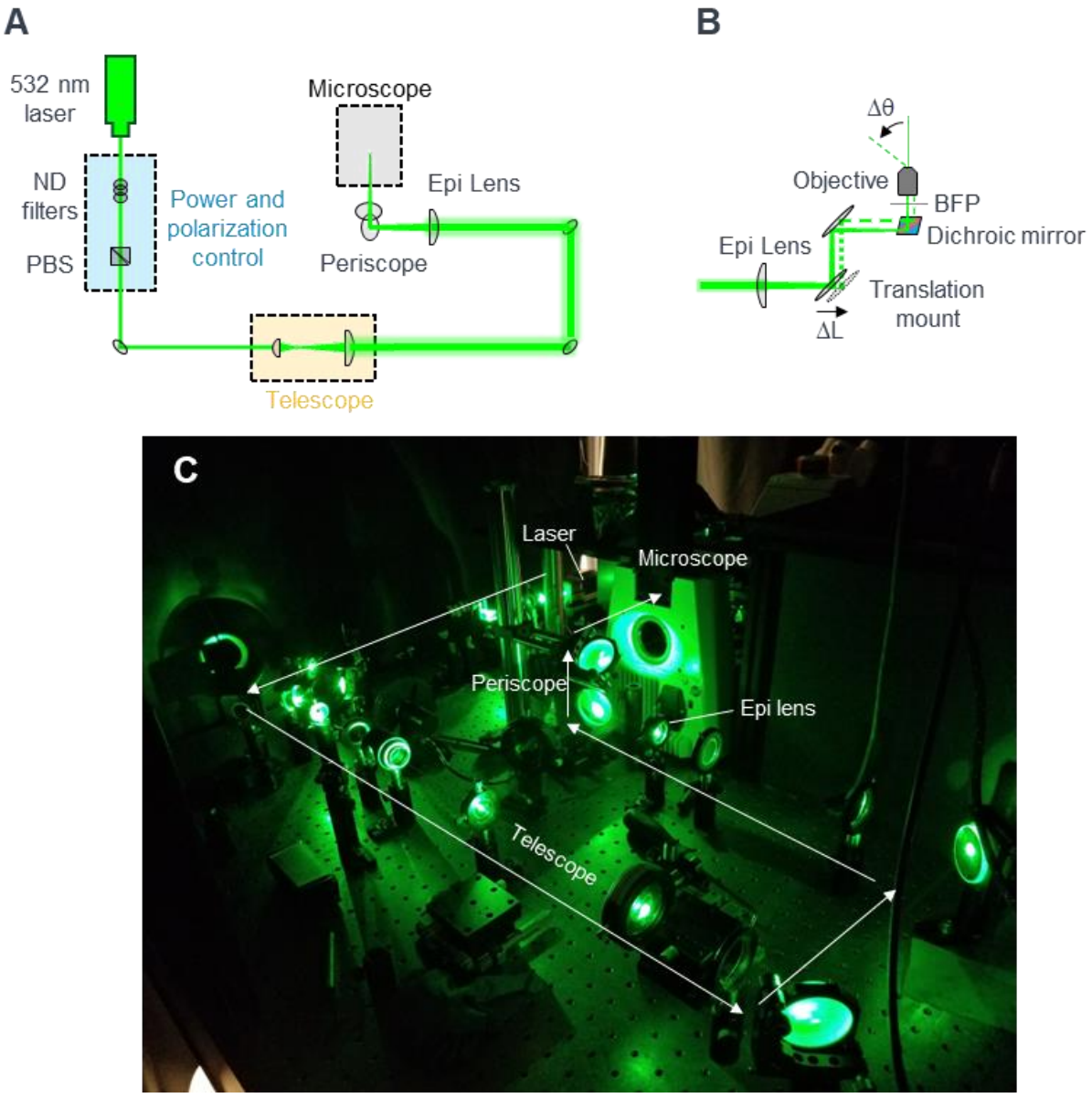

Figure 2.1 (A) Schematic of the TIRF microscope excitation path before entering the microscope. (B) Schematic depicting translation of periscope mirror to deflect laser beam. (C) Photograph of optical arrangement.

removed with a 532 nm bandbass filter with an optical density (OD) of 7, pure linear polarization of the laser output is obtained using a half-wave plate with a polarizing beamsplitter cube, and the beam power is attenuated using neutral density (ND) filters. The beam is then expanded to 1 cm in diameter in a telescope consisting of plano-convex lenses with focal lengths f = 30 mm and f = 300 mm (magnification = 10). The lenses are oriented so that the collimated beam enters or exits the curved face of each lens to minimize spherical aberration. A pair of mirrors in a periscope configuration vertically displace the beam to direct it into the rear port of an inverted microscope (Zeiss Axiovert 135). Inside the microscope, the beam reflects off a long-pass dichroic mirror (λ = 550 nm) before entering a 60X, high numerical aperture (NA = 1.4) oil-immersion microscope objective. To achieve widefield illumination, a f = 400 mm "epifluorescence" lens placed in the beam path before the microscope focuses the beam onto the back focal plane of the microscope objective (Figure 2.1B). The resulting collimated beam transmitted by the objective illuminates a 40 μm diameter spot on the sample. This spot size is determined by the input beam size, the focal length of the epifluorescence lens, and the magnification of microscope objective.

To bring about total internal reflection (TIR), the laser must strike the sample at an oblique angle beyond the critical angle. For experiments performed in aqueous solution with molecules attached to a glass coverslip, the incident medium is glass (n = 1.52) and the transmitted medium is water (n = 1.33); therefore, the critical angle is $\sin^{-1}(1.33/1.52) \approx 61°$. There are two common methods to obtain an angle of incidence which exceeds this critical angle. The first, used in this work, is "through-objective TIRF," in which the excitation laser passes through the microscope objective intentionally off-axis to cause the laser to hit the sample at an angle (Figure 2.1B), just as translating a lens in front of a laser to deflect the beam. This method

is relatively straightforward to implement, as the excitation spot and detection region are automatically aligned. However, there are drawbacks to this method. First, to generate TIR, the microscope objective must have a sufficiently large numerical aperture so that the maximum collection half-angle is greater than the critical angle. This criterion is met if the objective's NA is greater than the index of refraction of the transmitted medium, in this case water (n = 1.33). Such a high numerical aperture can only be obtained using expensive, oil-immersion objectives. Second, the laser inevitably interacts with the objective's internal optics to generate autofluorescence which results in additional background signal. The second method is "prism-based TIRF," in which the laser strikes the sample on the side opposite the microscope objective, where the angle of incidence is controlled by a prism. This method has an improved signal-to-background ratio because the laser never enters the objective, so there is virtually no autofluorescence. However, prism-based TIRF has its own drawbacks, such as needing to align the excitation spot with the objective's field of view, and requiring an objective with a high working distance to focus through the sample to the far interface. We weighed these issues and ultimately decided against the prism-based approach for a practical reason: through-objective TIRF allowed us to leave one side of the sample free to attach a temperature control module (see Section 2.3), which would have been much more awkward to accomplish in a prism-based setup.

We obtain through-objective TIR by mounting one of the periscope mirrors onto a micrometer translation stage (Figure 2.1B). As the stage is moved, the beam is translated off the optical axis of the objective which continuously increases the angle of incidence. Sample slides are mounted on a homebuilt translation stage to allow scanning of the sample surface. The stage consists of a pair of translation micrometers, one each for the x and y directions, that are attached to a cantilevered aluminum plate which suspends the sample above the microscope objective.

This design should be improved in future implementations, as the cantilever requires meticulous tuning to align the sample plane with the objective's imaging plane. Fluorescence from the sample is collected by the microscope objective, transmitted through the long-pass dichroic mirror in the microscope, focused by the microscope's internal tube lens, and directed out the side port into the final section of the experiment, the detection tree (Figure 2.2).

The fluorescent image output by the microscope comes to a focus approximately 10 cm beyond the microscope side port. This image is then relayed to a charge-coupled device (CCD) by a f = 75 mm lens that collimates the light and a second f = 400 mm lens which focuses the image onto the camera (to avoid chromatic aberration, all lenses used in the detection tree are achromatic cemented doublets). This lens pair also acts as a telescope to further magnify the

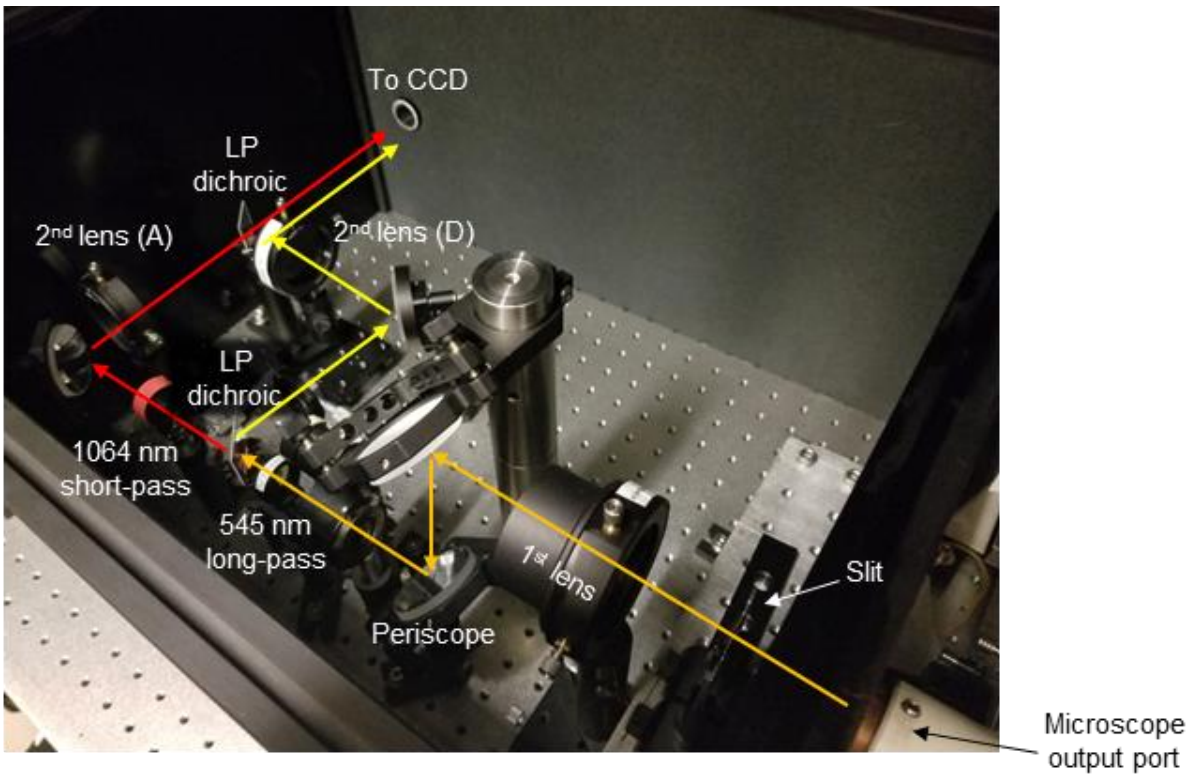

Figure 2.2 Detection tree used to separate donor and acceptor emission and refocus image onto CCD detector. Donor emission is in yellow, acceptor emission is in red, and combined emission before the dichroic is in orange. Two filters remove excitation laser light: a 547 nm long-pass (OD 7) at the excitation wavelength before the dichroic mirror and a 1064 nm short-pass (OD 6) to remove residual un-doubled light from the Nd:YAG after the dichroic mirror. (LP = long-pass).

image. Within this image relay, there are several additional optical elements. First, a periscope lowers the beam to the optical table. Second, a 547 nm long-pass filter (OD 7) removes scattered laser light from the image. Third, and most importantly, a long-pass dichroic mirror (λ = 645 nm) separates the Cy3 and Cy5 emission into independent channels. The two channels are then recombined by a second, identical dichroic mirror before focusing onto the CCD. The splitting and then recombining of the Cy3 and Cy5 emission is used to introduce a translational offset in the two channels. Therefore, two images (donor and acceptor) are projected side-by-side onto the CCD. An alternative method is to use two CCD cameras, one for each channel, but the one-camera approach avoids needing to synchronize the two cameras for data analysis.

One should note two things about this detection setup. First, in this arrangement, there are *two f* = 400 mm lenses, one for each channel. Using two focusing lenses provides even greater mitigation of chromatic aberration by permitting slightly different path lengths for each channel, which ensures that both Cy3 and Cy5 images come to the same focus. Second, it is essential that the donor and acceptor images do not overlap on the CCD. When the images are circular, as is natural when working with a circular laser beam, projecting the images on the CCD in a non-overlapping way results in a significant unused area of the CCD. However, truncating the circular image into a vertical strip permits nearly full use of the CCD's pixels, which in turn allows imaging of more molecules simultaneously (Figure 2.3). We perform this truncation by inserting a slit at the image focus output of the microscope (Figure 2.2).

The CCD used in single-molecule fluorescence measurements must be capable of operating in low-light conditions with rapid readout and low noise. In this experiment, the CCD used for detection is a so-called intensified CCD, or ICCD. In an ICCD, light strikes a multichannel plate (MCP) to generate photoelectrons which are then multiplied ~$10^{5\text{-}6}$ times

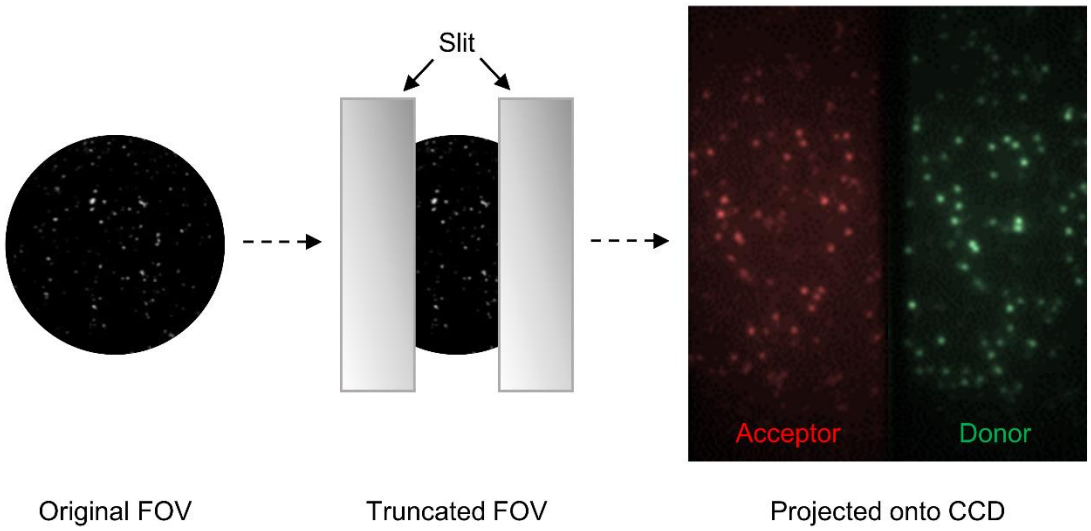

Figure 2.3 Fluorescence image is truncated into strips to maximize used CCD area. Slit is placed at the microscope output tube lens focus (see Figure 2.2).

before being converted back into light using a phosphor screen, and this amplified light is directed by a fiber optic bundle to strike a ¼ megapixel CCD (512 x 512 pixels). The CCD operates in the usual way (more details are given below in Section 2.1.4): photoelectrons are accumulated ("integrated") during the exposure time, after which the pixel grid is shifted column by column into a shift register which delivers each pixel to a fast readout register (5 MHz) for analog-to-digital conversion (ADC) before computer storage. The key principle for CCD operation in low light is the use of signal amplification to bring the signal level (in CCD electrons) above the readout noise level of the ADC. For ICCDs, the MCP plays this role. A popular alternative to ICCDs for smFRET measurements is the electron multiplying CCD (EMCCD), which uses an electron multiplying shift register on the CCD chip to achieve signal amplification. By swapping the MCP for on-chip electron multiplication, the EMCCD architecture has a higher photon detection quantum yield and therefore a higher signal level. However, unlike the ICCD, the EMCCD also amplifies unwanted electrons in the CCD, such as thermally generated "dark electrons," which results in greater noise. As a result, the signal-to-

noise ratio (SNR) of the two camera types is similar. A third camera type gaining interest in the smFRET field is the scientific complementary metal-oxide-semiconductor (sCMOS) camera. Vendors developing sCMOS cameras promise decreased readout noise, higher frame rates, and more pixels than ICCD or EMCCD cameras at comparable prices. This relatively new technology will be one to watch.

2.1.3 Strategy for optical alignment pre- and post-sample

I will now outline the step-by-step process used to align the various optical elements in the TIRF microscope. Alignment is split into pre-sample laser alignment and post-sample emission collection alignment. Laser alignment occurs in three steps: (i) align for confocal illumination, (ii) switch to epifluorescence (widefield) mode, and (iii) generate TIR. For confocal mode, the laser light must be collimated and on-axis when entering the microscope objective. Therefore, I first ensure that the beam is collimated after leaving the expansion telescope (Figure 2.1), which is done simply by temporarily inserting a mirror in the beam path and projecting the beam off the optical table to a far wall (after, of course, taking all safety precautions needed for working with a laser off-table). The beam is well collimated if the beam diameter is unchanged from the table to the wall. Then, to align the beam to the center of the microscope objective, I replace the objective with a long (~0.5 m) lens tube with an iris attached to each end, and I use the two periscope mirrors to align to the two irises through the spot projected onto the ceiling. It is helpful here to use a tight iris early in the beam path to “clamp down” on the beam, which creates a diffraction pattern that is easier to use when determining if the beam is centered on each iris. At this point, the system is aligned to create a diffraction-limited spot (“confocal mode”).

For widefield illumination (epifluorescence), the laser must be focused onto the back focal plane of the objective, which is accomplished by the epi lens (Figure 2.1). First, the spot

location on the ceiling is marked and the lens tube is replaced by the microscope objective. Then, an empty sample chamber (see below) is placed on the microscope stage above the objective, and a flip-mounted epi lens is introduced into the beam path. The epi lens position perpendicular to the beam is adjusted until the ceiling spot is centered on the marked confocal spot location, which indicates that the lens is centered and the beam has not been deflected. The epi lens position along the beam propagation direction is adjusted until the spot size is minimized on the ceiling, as a beam properly focused onto the objective's back focal plane will have the minimum divergence angle. At this point, in epifluorescence mode, the sample is illuminated by a widefield spot that transmits through the sample.

To enter TIR mode, the translation-state mounted periscope mirror is moved to deflect the beam until the critical angle is reached. However, as determining the optimal angle is best done during fluorescence imaging, one should align the fluorescence detection system (described below) in epifluorescence mode first before attempting to enter TIR mode. Optimizing the TIR angle is done by observing how a fluorescence image changes as the angle of incidence is increased: when TIR is obtained, the background fluorescence will sharply diminish and the fluorescence from surface attached molecules will "flash" brighter before becoming progressively dimmer as the critical angle is surpassed. The origin of this flash is described in Section 2.1.5; it suffices here to say that the angle of incidence at which the flash is observed is the optimal angle that maximizes the signal to background ratio.

Alignment of the fluorescence imaging system is done through several levels of increasing refinement. For coarse alignment, it is useful to have a fluorescence source that is visible to the naked eye, such as concentrated (millimolar) rhodamine 6G in ethanol. Better yet is a fluorophore whose emission is split by the dichroic that separates Cy3 and Cy5 emission, such

as 4-dicyanomethylene-2-methyl-6-(4-dimethylaminostyryl)-4H-pyran (DCM). Under strong enough laser illumination (~100 mW), the fluorescence image obtained by the microscope objective from these laser dyes can be seen on white paper in a dark room, which facilitates positioning optical elements to obtain rough alignment. The next step is to turn on the CCD camera and image bright, fluorescent particles larger than the diffraction limit, such as 2 μm fluorophore impregnated beads. These beads are useful because they are extremely bright objects that photobleach very slowly and can thus survive extended imaging. Care should be taken at this step to adjust the laser power so as not to overload the CCD. For 2 μm fluorescent beads, the power level at which CCD saturation occurs is approximately 0.1 mW. First, the acceptor channel is blocked, and a focus is obtained on the donor channel. Then, the donor channel is blocked, and the acceptor channel is brought into focus by translating the acceptor focusing lens (Figure 2.2). One may then switch to imaging photostable objects smaller than the diffraction limit, such as 100 nm diameter fluorescent beads or quantum dots, and further refine the co-alignment of the two channels. At this point, the slit can be introduced at the focal plane of the tube lens output from the microscope. A simple slit construction is two razor blades mounted parallel on optical posts. When positioned correctly, the edges of the slit will appear sharply in focus on the CCD, and the transmitted image will center on the brightest part of the laser illumination spot. Finally, the donor and acceptor channels are simultaneously uncovered, and the offset of the two images is adjusted until the camera field of view is fully occupied. (e.g., Figure 2.3).

2.1.4 Camera sensitivity, data acquisition rate, and image magnification

The ICCD camera used in these experiments reports the light level per pixel in ADUs (analog-to-digital units), which is an arbitrary unit. To convert ADUs per frame into a physically

meaningful quantity, the photon detection rate, we need to know the ADU-to-photon conversion rate κ. For an average number of photons detected by the camera per frame N, the average CCD signal S is

$$S = \kappa N + R + D \quad \text{(Eq. 2.1)}$$

where R is the ADC readout offset, and D is the CCD dark charge. The ADC readout offset R is specific to the ADC and is a constant value for all pixels, measured as R = 140 ADU/pixel. The dark charge D is signal due to thermally generated electron-hole pairs in the CCD pixels. D scales linearly with the time between pixel readouts, and by using an ICCD cooled to −30 °C, the dark current is reduced to D = 4 ADU/s/pixel, which is negligible at all frame rates used in this dissertation. While the average values of R and D can be simply subtracted from each pixel, the noise that R and D contribute cannot and must be accounted for. As N, R, and D are independent, the corresponding uncertainty (standard deviation) of the signal σ_S is the quadrature sum of the uncertainty of each contribution:

$$\sigma_S = \sqrt{\kappa^2 G^2 \sigma_N^2 + \sigma_R^2 + \sigma_D^2} = \sqrt{\kappa^2 G^2 N + \sigma_R^2 + D} \approx \kappa G \sqrt{N} \quad \text{(Eq. 2.2)}$$

Here σ_N, σ_R, and σ_D, are the uncertainties in the photon rate, the readout noise, and the dark charge, respectively, and G is a multiplicative increase in the photon detection uncertainty due to imperfect gain (see Appendix 2), which for the exponential gain in the MCP is approximately $\sqrt{2}$. Photon detection and dark charge accumulation are both Poisson processes whose variances are equal to their averages, which is the basis of the substitutions $\sigma_N = N^{1/2}$ and $\sigma_D = D^{1/2}$. The read noise uncertainty is measured as σ_R = 4 ADU. At high N, Eq. 2.2 reduces to a simple square root relationship between σ_S and N, and the ratio of the CCD variance to the signal level is

$$\frac{\sigma_S^2}{S} = \frac{\kappa^2 G^2 N}{\kappa N} = G^2 \kappa \quad \text{(Eq. 2.3)}$$

Therefore, measuring the ADU/photon conversion rate can be done by taking a CCD movie at a relatively high but non-saturating signal level and computing the uncertainty per pixel squared divide by twice the average per pixel. For the current ICCD camera, κ = 1.9 ADU/photon.

The rate at which CCD frames can be acquired places important limits on the measurable rate constants for this experiment. The frame rate ϕ_{frame} is determined by the per-pixel digitization time τ_{ADC}, the shift register time τ_{SR} which occurs once per digitized row, and the frame shift time τ_{FS}, as well as the number of rows N_x and columns N_y to be digitized:

$$\phi_{frame}^{-1} = N_x N_y \tau_{ADC} + N_x \tau_{SR} + \tau_{FS} \quad \text{(Eq. 2.4)}$$

The measured parameters of the ICCD used in this work are $(\tau_{\text{ADC}})^{-1}$ = 5 MHz, τ_{SR} = 7 μs per row, and τ_{FS} = 5.6 ms. Therefore, the full frame readout time ($N_x = N_y$ = 512 pixels) is 61 ms (ϕ_{frame} = 16 Hz). Higher frame rates can be achieved by "cropping" the field of view to have fewer columns or especially rows, but cropping comes at the cost of decreased throughput, as a smaller imaging area means fewer molecules observed simultaneously. Even if the entire field of view is cropped ($N_x = N_y = 0$), the frame shift must occur once per frame and sets an upper limit of ϕ_{frame} = 180 Hz. Dynamics occurring faster than the frame rate are lost due to averaging, and therefore the ICCD in these studies is most well-suited to study rate constants slower than 10 s^{-1} but can be adjusted as needed and can in principle be used to examine dynamics on the single-digit millisecond timescale.

The net magnification of the CCD image is an important quantity to select for in the experimental design. Due to diffraction, each fluorescent molecule cannot be imaged as a point light source but will instead appear spatially broadened across multiple CCD pixels as a finite-width point spread function (PSF). The PSF depends on the details of the optical system, but a commonly used PSF is the Airy pattern, which is the Fourier transform of a circular aperture.

One may define the minimum resolvable distance as the first zero of the Airy pattern, which is the Rayleigh criterion (r_{RC} = 1.22 λ / 2 *NA*, where λ is the light wavelength and *NA* is the numerical aperture of the microscope objective), or alternatively as the Airy pattern's full width at half max ($FWHM_{Airy}$ = 1.03 λ / 2 *NA*), which is easier to compare to a Gaussian fit ($FWHM_{Gaussian}$ = 2.35 σ, where σ is the radial standard deviation). For Cy3 emission centered near 600 nm, r_{RC} = 260 nm and $FWHM_{Airy}$ = 220 nm. This resolution determines the appropriate pixel resolution d_{pixel}, which is the size of pixels in object space. If d_{pixel} is larger than the PSF width, molecules will appear as single-pixel points, and all width information is lost (for instance, one cannot fit the PSF width to determine if the "molecule" is in fact more than one molecule superimposed on the same pixel). As d_{pixel} is decreased, individual fluorophores will occupy more pixels on the CCD. However, when summing pixels to determine total intensity, each pixel contributes additional read noise, and therefore the signal-to-noise ratio degrades as d_{pixel} decreases. Furthermore, the multiplexing capability of the CCD diminishes as each molecule take up more space. Therefore, an ideal system magnification produces a PSF with a diameter of 2 or 3 pixels. For the current system, the magnification, as measured with a calibration slide of a grid of 5 μm squares (see Figure 2.4A), is 78 nm/pixel. Upon binning pixels into 2×2 or 3×3 "superpixels," the desired 3-pixel wide PSF is obtained (see Figure 2.4B). As of the writing of this dissertation, the system magnification is being altered so as to utilize the full, unbinned 512×512 pixel grid to increase throughput.

2.1.5 System collection efficiency and sources of noise

The fluorescence emitted by molecules in the single-molecule microscope undergo unavoidable losses before being detected by the camera. These losses are quantified by the collection efficiency η, which is the ratio of the detected photon rate ϕ_E to the true fluorophore

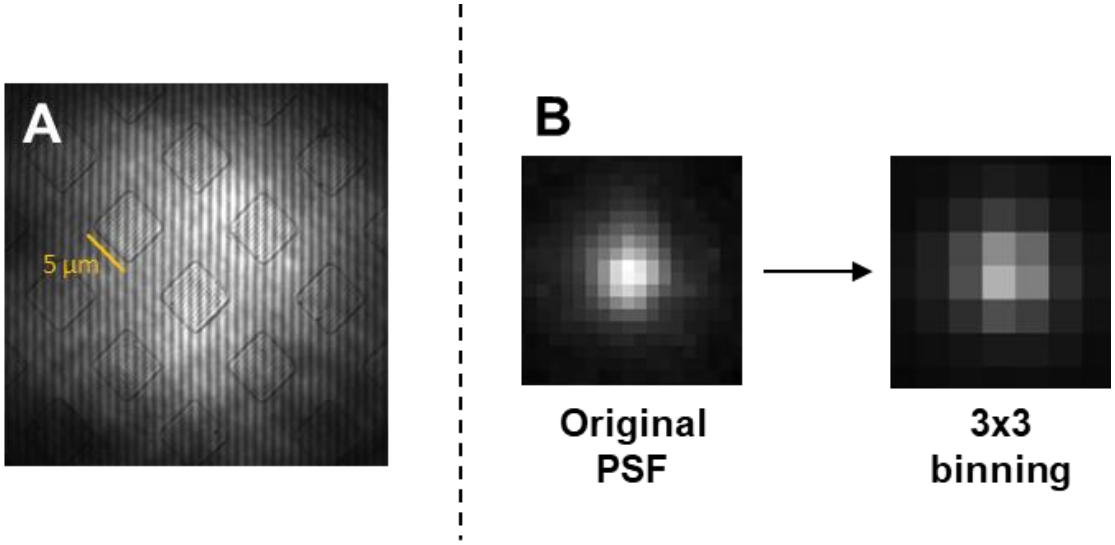

Figure 2.4 (A) Image of calibration grid of squares with side length 5 μm for magnification determination. (B) Use of on-chip pixel binning to decrease system PSF width (image is of a 100 nm diameter quantum dot).

photon emission rate ϕ_{Em} ($\eta = \phi_{Em} / \phi_E$). In the absence of acceptor, the donor emission rate can be calculated as the product of the photon flux density in the TIR evanescent field with the absorbance cross section and fluorescence quantum yield of the donor (Cy3):

$$\phi_{Em} = \frac{I}{h\nu}\sigma_{Cy3}Q_{Cy3} \qquad \text{(Eq. 2.5)}$$

where I is the electromagnetic field intensity (in W/cm^2), $h\upsilon$ is the energy per photon (3.74 x 10^{-19} J at 532 nm), σ_{Cy3} is the absorbance cross section of Cy3 (3.49 x 10^{-16} cm^2) and Q_{Cy3} is the quantum yield of Cy3 fluorescence (≈ 0.25[†]). The intensity I is not the same for all imaged molecules due to the laser's Gaussian beam profile,

$$I(r) = I_0 e^{-2r^2/w^2}, \qquad \text{(Eq. 2.6)}$$

[†]This quantum yield is for Cy3 at room temperature conjugated to a nucleic acid. The quantum yields of Cy3 and Cy5 are greatly enhanced by stacking on nucleobases in DNA and RNA. This has the undesirable effect of stabilizing nucleic acid secondary structure. A purportedly less perturbative FRET pair is Alexa Fluor 555 and 648, which could be used without changing any of the optical components in this work (e.g., dichroic beamsplitters).

where I_0 is the intensity at r = 0 and w is the 1/e electric field radius of the beam, equal to 20 μm for this optical setup.[‡] For this signal calculation, consider a molecule at the center of the beam. The laser power P, which is easy to measure, is related to the peak intensity I_0 by integrating the beam profile:

$$P = \int_0^{2\pi} d\phi \int_0^{\infty} r\, dr\, I(r) = \frac{1}{2} I_0 \pi w^2. \quad \text{(Eq. 2.7)}$$

At a typical operating laser power P = 10 mW, this yields a peak intensity I_0 = 1.6 kW/cm^2. However, this is the intensity of the freely propagating laser, not of the evanescent field that the fluorophores experience, which will now be discussed.

The behavior of light at interfaces is described in general by the Fresnel equations, which show in the case of total internal reflection (TIR) that there is an evanescent wave generated in the low-index medium which oscillates at the same frequency as the incident wave and has a field intensity that decays exponentially with the distance from the interface:[5]

$$I_t(z) = I_0 e^{-z/d}. \quad \text{(Eq. 2.8)}$$

This evanescent wave is characterized by its intensity at the interface I_0 and its penetration depth d. The subscript t stands for the "transmitted" medium, though no light is transmitted in the far field sense. For incident light polarized perpendicular to the plane of incidence (s-polarization), the evanescent field is linearly polarized, while incident light polarized parallel to the plane of incidence (p-polarization) produces an elliptically polarized evanescent field. The penetration depth depends on the laser frequency and angle of incidence:

$$d = \frac{\lambda_0}{4\pi} \frac{1}{\sqrt{n_i^2 \sin^2 \theta_i - n_t^2}} \quad \text{(Eq. 2.9)}$$

[‡] This uneven illumination results in a fluorescence image with a "vignette"—molecules near the edge fluoresce less than those near the center. This has the effect of sampling many different signal-to-noise levels in one movie, which unfortunately means that not all parts of the field of view are equally useable. Adopting "flat field" illumination methods would therefore be a useful improvement in future developments.

where λ_0 is the light wavelength in vacuum, θ_i is the angle of incidence, and n_i and n_t are the indices of refraction for the incident and transmitted media, respectively. For this work, the incident medium is glass (n_i = 1.52) and the transmitted medium is water (n_t = 1.33). The intensity at z = 0 (I_0) depends on the angle of incidence and the laser polarization:

$$I_0^{\perp} = I_i^{\perp} \frac{4\cos^2\theta_i}{1-\left(n_t/n_i\right)^2} \qquad \text{(Eq. 2.10)}$$

$$I_0^{\parallel} = I_i^{\parallel} \frac{4\cos^2\theta_i\left[2\sin^2\theta_i-\left(n_t/n_i\right)^2\right]}{\left(n_t/n_i\right)^4+\sin^2\theta_i-\left(n_t/n_i\right)^2} \qquad \text{(Eq. 2.11)}$$

where the superscript symbols $\perp$ and $\parallel$ refer to s-polarized and p-polarized light, respectively, and $I_i^{\perp,\parallel}$ is the light intensity in the incident medium. The angular dependence of d, $I_0^{\perp}$, and $I_0^{\parallel}$ for 532 nm light at the glass-water interface are shown in Figure 2.5. Note that for incidence

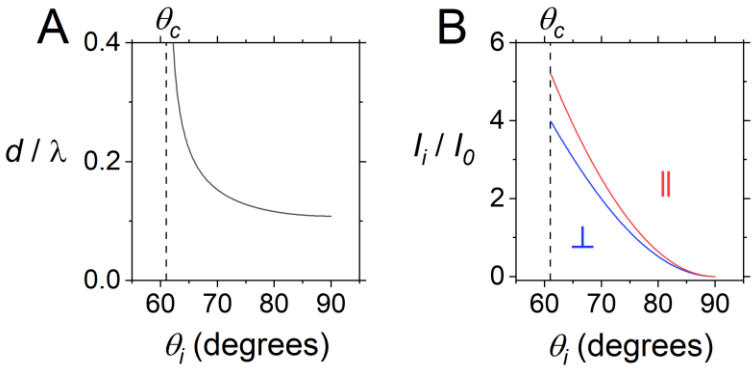

Figure 2.5 Angular dependence of evanescent field penetration depth d and intensity $I_0^{\perp,\parallel}$ for 532 nm light at the water-glass interface (n_t/n_i = 1.33/1.52).

angles θ_i near the critical angle θ_c, the evanescent intensity at the surface is *greater* than the intensity of the incident wave ($I_0^{\perp,\parallel} > I_i^{\perp,\parallel}$); this explains why fluorophores become brighter when TIR is established. The intensity increase is slightly greater for p-polarized light than s-polarized light; however, the microscope in this work is configured to use s-polarization rather than p-polarization to avoid the transmission of imperfectly collimated laser light striking the sample at the Brewster's angle. The angle of incidence in the experiment is difficult to determine precisely

but is narrowly bounded from 61° to 67°. The lower bound is the glass-water critical angle θ_c = 61°, and the upper bound is the maximum acceptance angle of the objective θ_{max} = 67°, which is calculated from the objective's numerical aperture (*NA* = 1.4),

$$NA = n \sin \theta_{max}, \qquad \text{(Eq. 2.12)}$$

where *n* is the index of refraction of glass (*n* = 1.52). A reasonable estimate for θ_i is 64°, for which the penetration depth *d* is 140 nm and the ratio of transmitted to incident surface intensity (I_t / I_i) is 3.3. The fluorophores are tethered to the surface at a distance $z \approx 10$ nm, yielding the fluorophore's true local electric field intensity I = 4.9 W/cm^2. Multiplication of I by σ_{Cy3} gives an expected excitation rate of 4.6 MHz, for an average time between excitations of 220 ns. Since this time is much larger than the fluorescence lifetime of Cy3 ($\tau_f \approx 5$ ns), there is no concern for Cy3 saturation, even up to the present 200 mW laser power limit, and the linear absorbance assumption in Eq. 2.5 is valid. Multiplying by the fluorescence quantum yield gives an emission rate ϕ_{Em} = 1.2 MHz. The measured photon fluorescence rate under these conditions is ϕ_F = 80 kHz, and therefore the system collection efficiency η is 7%.

2.2 Sample cells and sample preparation

Disposable glass flow cells of a cheap and simple "sandwich-style" design act as sample cells which allow for easy optical access and solution exchange. The flow cell, shown in Figure 2.6, consists of #1.5 glass coverslip (CGI Life Sciences) attached to a 75 mm x 25 mm glass slide (VWR) via double-sided tape (3M). It is essential for the coverslip surface to be clean and free of any fluorescent material, so the coverslip is soaked in acetone overnight (>12 hours) before being cleaned in a UV-ozone oven (Jelight Mo. 42). The UV-ozone treatment oxidizes any residual material on the coverslip, and the oxidized products are removed by later wash

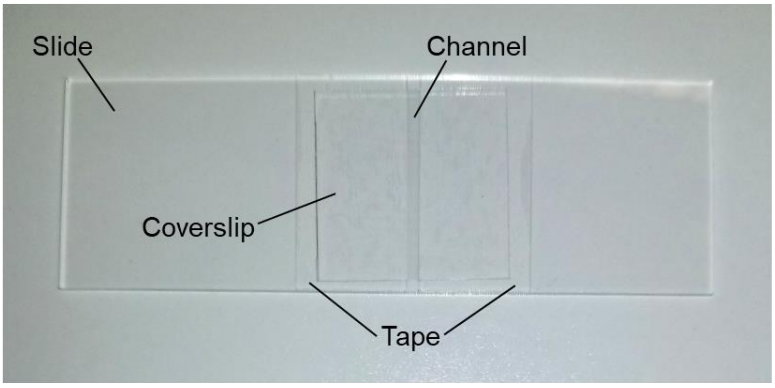

Figure 2.6 "Sandwich" style flow cell for microscopy.

steps. Two parallel strips of double-sided tape are used to create a channel approximately 2 mm wide, 22 mm long, and 0.1 mm deep. This channel can then be used for solution exchange by pipetting solution droplets onto one channel entrance and applying a tissue to the other to draw up solution via capillary action (i.e., wicking). One must be careful when wicking not to introduce bubbles into the channel. Bubbles can become stuck in the channel, which impedes solution flow and limits usable surface area for microscopy. Additionally, even if the bubble clears the channel, it can disturb the surface protein coatings (described next) that permit the tethering of constructs to the surface.

Once a clean glass coverslip incorporated into a flow cell, the coverslip surface is then functionalized to enable tethering of nucleic acid constructs (see Figure 2.7). This is carried out by first coating the surface in bovine serum albumin (BSA, 10 mg/mL) which serves to "block" the glass surface so that the nucleic acids do not adhere to the glass. In addition, 10% of the BSA are labeled with biotin (vitamin B_7), a small molecule that has an extremely high-affinity noncovalent interaction with the protein streptavidin ($K_d \approx 10$ fM, unbinding rate $k_{diss} = 10^{-4}$ s^{-1}). Therefore, treating the BSA-coated surface with streptavidin (0.2 mg/mL) will create BSA-biotin-streptavidin sites. Because streptavidin is a tetrameric protein complex with four binding sites for biotin, each tethered streptavidin is available to bind more biotin molecules. So, a final

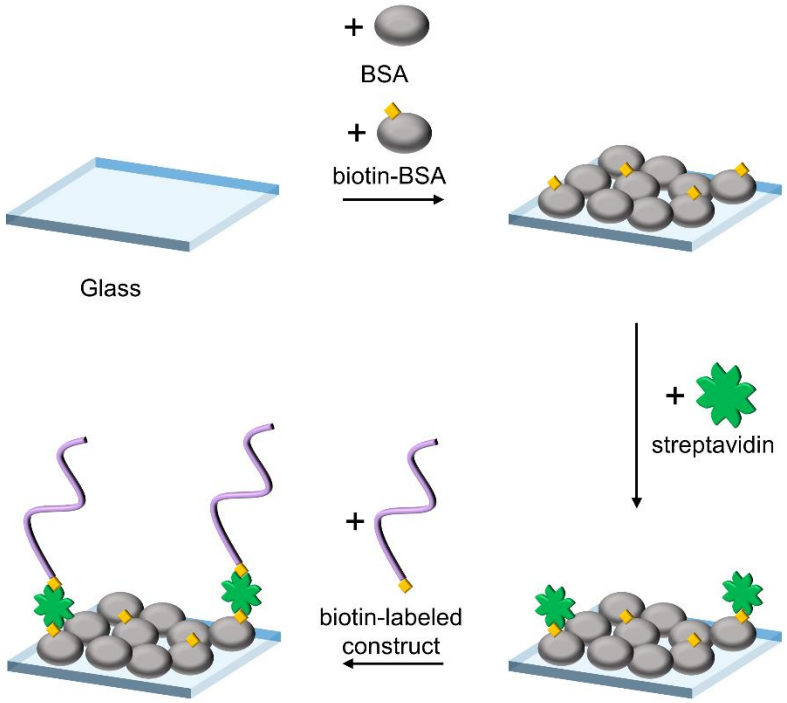

Figure 2.7 Surface functionalization for nucleic acid construct tethering. A clean glass surface is coated in a mixture of bovine serum albumin (BSA) and biotinylated BSA (BSA in gray, biotin in yellow). Streptavidin (green) binds to biotin-BSA sites on the surface, and finally a biotin-labeled construct (purple) is added which attaches to available biotin binding sites in the tetrameric streptavidin complex. The result is a surface decorated in nucleic acid construct with a surface density controlled by the concentration of the proteins and nucleic acid.

wash with a biotin-labeled nucleic acid construct (100 pM) will attach the construct to the streptavidin to complete the surface tethering process. These three solutions (BSA + biotin-BSA, streptavidin, and nucleic acid construct) are each incubated with the surface for 10 minutes in buffer (50 mM hemisodium HEPES, pH 7.6). This incubation time is significantly longer than the ≈ 1 minute required for molecules to diffuse to the surface, as predicted from the Einstein relation for diffusion in 1 dimension, $\Delta x^2 = 2Dt$, where D is the diffusion coefficient ($\approx 10^{-6}$ cm^2/s for a 60 kDa protein in water), Δx is the distance traveled (= 0.1 mm), and t is the diffusion time. The measured construct surface density $\sigma = 1$ molecules/μm^2 is smaller than the theoretical maximum for 100 pM construct of $\sigma = 3$ molecules/μm^2, and the surface density increases if the

construct is incubated for longer, saturating after ≈20 minutes. This clearly indicates the presence of additional kinetic barriers to streptavidin-biotin binding besides simple diffusion. Note that the resulting surface density is highly variable (± 50%) between samples and even between locations on a given sample. Having a reproducible surface density is important for efficient data collection, so it may be worthwhile to explore more stringent coverslip cleaning methods or alternative suppliers for coverslips, double-sided tape, etc.

One issue with the sandwich-style sample holders is sample evaporation. The channel ends are open to atmosphere, and evaporation of the sample becomes visually noticeable after approximately 5 minutes. This problematic for many reasons, including that the loss of solvent will increase the concentration of solutes in the imaging solution during data collection. If data can be collected in 1−2 minutes, such as for relatively rapidly folding constructs, sample evaporation is not a concern. However, for slower nucleic acid folding, or for operation at high temperatures (see below), evaporation poses a serious problem. To halt solution evaporation, the channel ends are sealed with RTV silicone which sets sufficiently 10 minutes after application. The silicone sealant is highly effective and allows a sample to be used for several hours with no noticeable evaporation. It is possible that some of the RTV components, which include acetic acid, may mix with the imaging solution, but a test showed no difference between rate constants measured from samples with or without RTV. Once sealed, the imaging solution in the flow cell can no longer be exchanged, which means that comparing folding dynamics at multiple cosolute concentrations requires creating a sample holder for each concentration. Ideal would be the future development of an alternative flow cell design which has limited evaporative losses without the use of sealants.

2.3 Temperature control

Essential to the work in this dissertation is the ability to perform single-molecule microscopy at a controllable sample temperature. Previously in the group, temperature control on the confocal instrument has been accomplished in two ways: (i) resistively heating a heating stage and objective collar, and (ii) locally heating a column of water around the microscope focus with an infrared laser.[6] Both are capable only of *increasing* the sample temperature above ambient temperature. For the TIRF microscope, I have instead opted to use thermoelectric coolers (TECs) to allow for both cooling and heating of the sample. TECs work via the Peltier effect, whereby a voltage across a thermoelectric material creates a thermal gradient and thereby heat flow across the TEC. By changing the sign of the applied voltage, the direction of heat flow can be switched, which allows the TEC to both heat and cool.

In detail, temperature control is established using an arrangement of two TECs, one resting on top of the sample slide, and one attached to the microscope objective (Figure 2.8). Heating the objective is important, as the objective is in thermal contact with the sample via the immersion oil. Furthermore, the use of two TECs helps eliminate thermal gradients across the

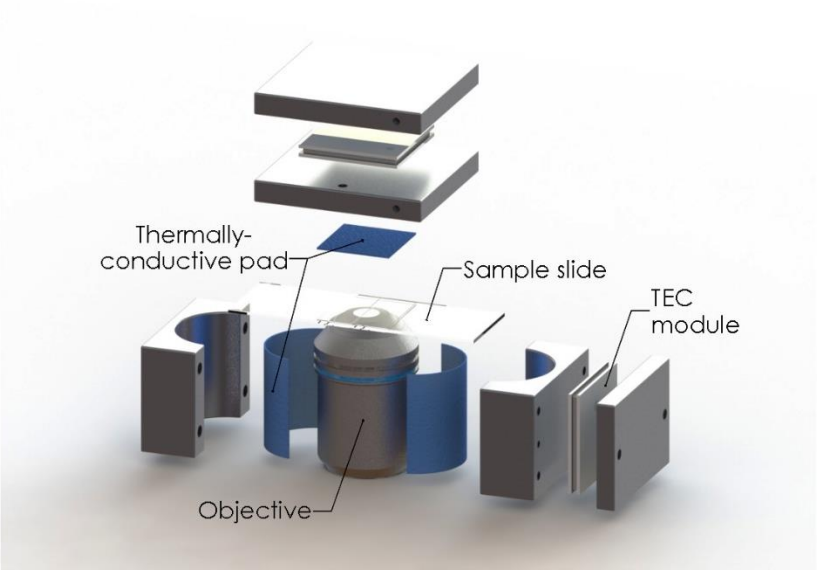

Figure 2.8 Composition of dual TEC temperature controller.

sample. The TECs are constructed from TEC modules (TE Technology HP-127-1.0-0.8) compression-mounted between two home-machined aluminum plates using nylon screws to reduce thermal backflow, and the TECs are coupled to the sample/objective via thermal pads (EN-Laboratories). Temperature stabilization is accomplished via a PI feedback loop. In this loop, the sample temperature is measured by thermistors (Vishay) attached to the aluminum plates, which feed into the error signal of temperature controllers (TE Technology TC-720) that drive the TEC modules.

The rate of cooling that can be achieved by TECs is limited. This is due to the resistive heating of the TEC modules as electrical current is applied as well as the imperfect thermal isolation of the two sides of the TEC devices. The TEC attached to the objective has an especially difficult task as the objective itself is connected to the microscope body, and therefore the entire microscope acts as a thermal reservoir which replaces heat removed by the TEC. To combat these issues, the TECs are cooled by heat sinks containing frozen reservoirs. Using frozen water (ice), the system is capable of cooling to ≈ 10 °C, whereas solid carbon dioxide (dry ice) can cool the sample well below 0 °C.[§]

Tuning of the PI feedback parameters was initially carried out using the Ziegler-Nichols (ZN) method, in which the proportional term P is increased until temperature oscillation occurs (at $\omega = 2.1\ \text{min}^{-1}$ for this system), then the integral term I is selected based on the oscillation period and the P term is reduced by a prescribed amount. However, the ZN parameters were found to be too aggressive. In particular, the large temperature overshoot when changing the set point would occasionally cause the system to become unstable. Therefore, the PI parameters

[§] The ability to operate at subzero temperatures opens up interesting research avenues. For instance, one could examine folding under conditions of freezing point depression, such as in salty water.

were manually decreased until reaching a satisfactory temperature stability (± 0.05 °C) and a modest settling time (< 1 min) without the risk of runaway heating.

2.4 Stroboscopic illumination

Folding rate constants that are faster than 10% of the camera frame rate are subject systematic underestimation in smFRET measurements with constant illumination. The origin of this effect, and the explanation of how a flashing light source (i.e., a strobe) can resolve this issue, are described in detail in Chapter 5. Here, I provide a technical description of how stroboscopic illumination is produced and synchronized with CCD data acquisition. An ideal strobe is a pulse train with one pulse per data acquisition frame and a duty cycle of 10-20%. The first attempt at light pulsing was to send a TTL signal to the Nd:YAG laser driver. However, the resulting laser pulse has a rise time of ~10 ms while lasing was being reestablished, which prevented going to the fastest camera frame rates. The second arrangement was to use an optical chopper wheel driven by a frequency-stabilized feedback controller (ThorLabs). This ensures even intensity through the strobe pulse, and 1 ms pulse times are easily achieved. The strobe pulse must be synchronized to the camera frame acquisition, which is done by sending the reference out signal of the chopper controller to the CCD trigger input. In the CCD control software, *WinView*, the camera's acquisition settings are selected to make the exposure time less than the time between pulses, so that no triggers are missed, and the timing mode is changed from the internally-timed "free-run" mode to "external trigger." Rate constants determined by stroboscopic experiments require a simple correction to account for the unobserved time (the portion of each frame when the laser is off); this correction is described in Chapter 5.

2.5 Data analysis

2.5.1 Extraction of single-molecule trajectories from movies

Once recorded, movies of the construct-decorated surface are processed in order to obtain single-molecule fluorescence trajectories. This analysis is carried out using *OpenTIRF.exe*, a program written by myself in the LabWindows/CVI C-language programming environment (see Figure 2.9). *OpenTIRF* processes each movies through a sequence of steps to extract, for each molecule, the donor and acceptor fluorescence intensity trajectories (D(t) and A(t), in photon counts per unit time) as well as the estimated FRET trajectory (E(t)). I will now describe each of these processing steps.

First, the movies are loaded into memory. Movies are decoded from the SPE file format used by *WinView*, the Princeton Instruments software that controls the CCD camera. The SPE file consists of a header, which contains information about the data collection conditions (number of frames, exposure time, pixel dimensions, etc.), and the body, which contains the 12-bit digitized output per pixel per frame. Because of the information in the header, no information is required from the user to load the movie, unless stroboscopic illumination is used, in which case the frame rate is not saved in the header and must be input manually.

Second, the program identifies the locations of molecules in the donor and acceptor channels. The first step is to average together a range of frames as a representative image to use in particle detection. This averaging is necessary because constructs in folding states with extremal FRET efficiencies ($E_{FRET} \approx 0$ or 1) will only appear in one of the two channels (donor or acceptor). Therefore, using a single frame for particle detection can result in the identification of donors without acceptors, or vice versa. Instead, averaging over the folding equilibration time of the smFRET construct ensures that molecules appear on both donor and acceptor channels.

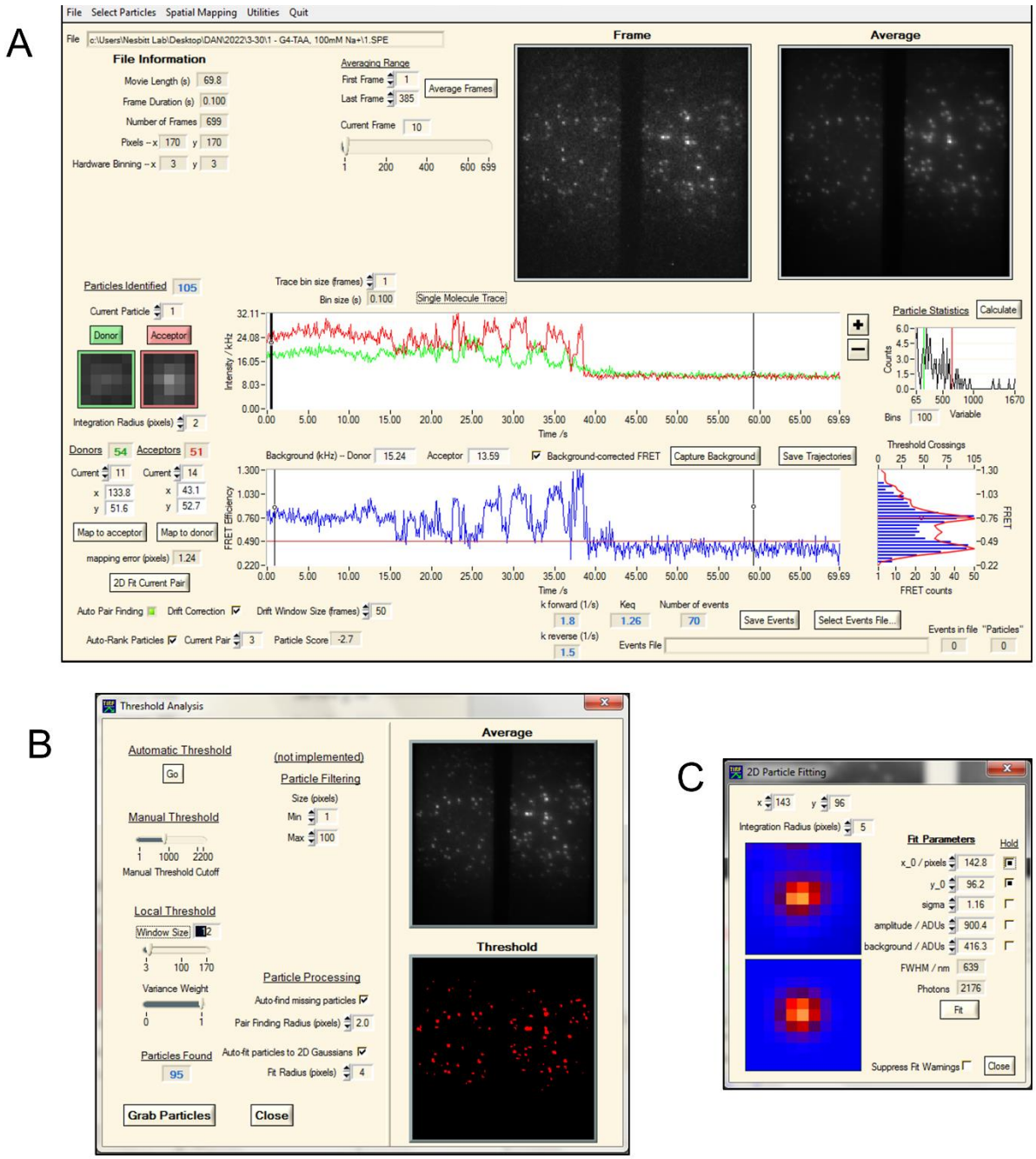

Figure 2.9 Screenshots of *OpenTIRF* program written by the author to extract smFRET trajectories from CCD movies. (A) *OpenTIRF* loads movies with metadata, averages frames for molecule detection, quickly cycles through molecules in a movie, extracts donor and acceptor fluorescence intensities, calculates FRET trajectories, and can threshold FRET trajectories for dwell time analysis. (B) The image thresholding utility automatically finds molecules with several algorithms available. (C) The 2D fitting utility models particle PSFs as 2D Gaussians for position refinement and filtering.

Additionally, this averaging increases the signal to noise ratio, which reduces the likelihood of single bright pixels (due to cosmic rays) being identified as molecules. Next, this averaged image is subjected to a local thresholding algorithm from the LabWindows/CVI machine vision package, and pixel regions above threshold are labeled as molecules (Figure 2.9B; a manual threshold option is available in case the automated thresholding algorithm fails). The center of mass of each above-threshold region is used as a rough estimate of the particle's position. Next, the rough position estimate is refined by fitting each putative molecule to a 2D Gaussian (Figure 2.9C),

$$G_{2D}(x,y) = \frac{A}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[(x-\mu_x)^2 + \left(y-\mu_y\right)^2\right]\right) + B, \qquad \text{(Eq. 2.13)}$$

where *A* is the amplitude, *B* is the background per pixel, σ is the standard deviation (the full width half max is calculated as FWHM ≈ 2.35 σ), and μ_x and μ_y are the Gaussian's x and y centers, respectively. This fit is restricted to the local pixel neighborhood surrounding the molecule to avoid having the fit converge to a nearby brighter molecule. As a result, *B* is a *local* background value, which is useful for computing background-corrected FRET values, described below. This 2D fit is also used to filter out particles whose fit parameters fail to converge or converge to unusual values, such as negative amplitudes, or standard deviations larger than the fit window. Note that this 2D fit function is circularly symmetric, with a single, isotropic width parameter σ. Fitting the PSFs to a full bivariate normal distribution with independent x and y widths and an xy correlation term shows negligible ellipticity (major to minor axis ratio a/b = 1.02 ± 0.06). Therefore, to reduce model complexity and speed up fitting, the simpler, isotropic Gaussian is used instead. However, it may be worthwhile to explore using a bivariate normal fit as a way of obtaining additional parameters to use in particle filtering (e.g., particles that appear elliptical may in fact be two molecules in close proximity). Lastly, particles are assigned as

donors or acceptors by their location on the CCD, i.e., whether they lie on the right or left half, respectively.

With molecule locations determined, the single-molecule fluorescence trajectories can be calculated. There are several options in *OpenTIRF* of how to calculate intensity trajectories. In each case, the fitted local background value is first subtracted from each pixel. The first method is simple pixel summation: pixels within a user-specified radius of the molecule's center position are summed together for each frame ($I(t) = \sum p_i(t)$). This method is robust and fast but requires tuning by the user to find the optimal radius that maximizes the signal to noise ratio. The second option is a weighted pixel sum: each pixel in the molecule's neighborhood is multiplied by the fitted 2D Gaussian value for that pixel and then summed together ($I(t) = \sum p_i(t) \cdot G(x_i,y_i)$). This method gives less weight to pixels far from the molecule center and therefore includes less background noise and can handle overlapping molecules better than simple pixel summation. However, this is a heuristic which should be developed further, preferably under an information-theoretic lens, before being used in literature publications. The final option is to fit the pixel data from each frame to a 2D Gaussian in which only the amplitude is allowed to vary, and the amplitude is recorded as the fluorescence intensity ($I(t) = \iint G(x_i,y_i) = A$). This option suffers from being slow to compute, which is why the analyses in this work have instead utilized simple pixel summation. A thorough comparison of these three methods, and in particular the signal-to-noise ratios of the resulting trajectories, would provide a useful guide as to which method, if any, is superior.

Once donor and acceptor intensity trajectories have been extracted from the movies, the next step is to match donor and acceptor molecules with each other so that the FRET trajectory $E(t)$ can be calculated. This pairing is accomplished with a spatial mapping function $M(\mathbf{r})$ which

maps locations in the donor channel $\mathbf{r}_D$ to the corresponding location in the acceptor channel $\mathbf{r}_A$. For a given donor, the mapping is used to identify the expected location of its acceptor, and the acceptor molecule which is closest to this expected location is chosen as its most likely pair. The discrepancy between the acceptor's predicted and true location, $|M(\mathbf{r}_D) - \mathbf{r}_A|$, is typically less than 2 pixels, and larger values alert the user to a poor match. Initially, $M(\mathbf{r})$ was determined by calibration: 2 known donor-acceptor pairs ("fiduciaries") identified manually are used to determine the parameters of an affine transformation which consists of a translation, rotation, and magnification. However, this requires recalibration each time the optics of the detection tree are aligned, and the rotation and magnification terms were routinely negligible. This inspired the later implementation of a translation-only automatic mapping based on image cross correlation: the donor and acceptor portions of the full CCD image are cross correlated, and the vector that maximizes the value of the cross-correlation is the vector which best translates the donor image onto the acceptor image. The mapping is calculated this way for each movie and requires no calibration but cannot be used if there are rotations or magnifications. In the case of small but nonnegligible rotations or magnifications, a local image registration method could be employed instead.[7]

Occasionally, the translation stage will drift during data collection, which invalidates the molecule locations identified by thresholding. This drift can be corrected for by translating each frame to align with the first frame of the movie. The translation vector is computed using image cross correlation similarly to calculation of the donor-acceptor translation vector. In this case, the frames of the movie are grouped into sections of equal length (usually 50-100 frames per section), and each section is averaged (averaging into sections is done to decrease the length of the calculation). Then the average of each section is cross-correlated with the first section, and

the vector that maximizes the cross-correlation is the drift vector for that section. Reversing this vector yields the drift correction vector. This correction is applied immediately upon loading the movie into memory, before any other processing has occurred.

Finally, with the donor and acceptor matched and their fluorescence intensities extracted from the movie, the FRET efficiency trajectory can be computed. The background-subtracted fluorescence intensities for the donor and acceptor, I_D and I_A, respectively, are

$$I_D = \phi_{ex,D}(1 - E)Q_D\eta \quad \text{(Eq. 2.14)}$$

and

$$I_A = \phi_{ex,D}EQ_A\eta, \quad \text{(Eq. 2.15)}$$

where $\phi_{ex,D}$ is the donor excitation rate, E is the FRET efficiency (the probability of energy transfer from the donor to the acceptor per excitation), Q_D and Q_A are the donor and acceptor fluorescence quantum yields, respectively, and η is the collection efficiency. Solving these expressions for the FRET efficiency yields

$$E = \frac{I_A}{I_A + \frac{Q_A}{Q_D}I_D}. \quad \text{(Eq. 2.16)}$$

If the donor and acceptor quantum yields are equal, then this equation simplifies to the acceptor fraction of emission ($E = I_A / (I_A + I_D)$). Furthermore, the total emission $I_A + I_D$ is a constant. The quantum yield ratio for the Cy3-Cy5 pair is estimated as $Q_A/Q_D = 0.3/0.25 = 1.2$, which is close to achieving constant total signal. The quantum yield ratio can depend on a number of factors, such as fluorophore interactions with nucleic acid, the presence of quenching cosolutes, and variation in temperature, and the constancy of $I_A + I_D$ should be checked by measuring whether the total signal depends on FRET efficiency. It is worth noting that Eqs. 2.14 and 2.15 assume equal donor and acceptor collection efficiencies and no detector crosstalk, which is not guaranteed. Clearly, measuring absolute FRET efficiencies is quite challenging; however, for the

purpose of determining transition rates between FRET states, the naïve FRET efficiency calculation (Eq. 2.16) is sufficient.

Analyzing the molecules in a smFRET movie requires examining each molecule by hand to evaluate if the molecule should be used for data analysis. Valid reasons to reject molecules include (i) rapid photobleaching before dynamics can be observed, (ii) low FRET trajectory signal-to-noise ratio due to molecule positioning on dim outer ring of Gaussian illumination area, (iii) the overlap of two or more molecules which convolutes the FRET trajectory, and (iv) the lack of either donor and acceptor fluorophore on the construct due to incomplete labeling. Additionally, one may observe molecules which do not exhibit the same dynamical behavior as most other molecules in the movie, such as a molecule showing no dynamics at all. This may indicate an incorrectly synthesized construct, or a molecule whose local environment is unusual (e.g., a gap in the BSA layer with an exposed glass surface), though these "exceptional" molecules should not be discarded lightly and may reflect true heterogeneity in the population.

To speed up the tedious process of looking at every molecule in the smFRET movie, *OpenTIRF* has an optional feature to sort molecules by likelihood to be used for analysis using a support vector machine (SVM).[8] SVMs use supervised machine learning to perform binary classification of multidimensional input. This binary classification is done by dividing the multidimensional space with an oriented hyperplane (the "decision boundary"), such that the input is classified as "positive" or "negative" (in this case, "likely to be used for data analysis" or "likely not used," respectively) depending on which side of the hyperplane it lies, with input farther from the hyperplane being more confidently classified. This calculation is carried out by a simple dot product between the input vector $\boldsymbol{v}_i$ and the vector orthogonal to the hyperplane $\boldsymbol{w}_o$ along with an offset b:

$$Score = \vec{v_i} \cdot \vec{w}_0 + b. \quad \text{(Eq. 2.17)}$$

Determining the optimal hyperplane is done using a training set, which *OpenTIRF* accumulates automatically as it is used. The training set includes, for each analyzed molecule, the donor and acceptor brightness and signal-to-noise ratio, the donor-acceptor intensity correlation constant, the donor and acceptor PSF widths, the donor and acceptor decorrelation time, the donor-acceptor spatial mapping error, and of course whether or not the molecule was used for analysis. This training set is fed into MATLAB, where an SVM library is used to efficiently determine $\boldsymbol{w}_o$ and b. This process should be repeated periodically as the training set grows over time. The SVM is not perfectly accurate (the problem is not linearly separable), so each molecule must still be examined by hand, but the user can move more quickly through the molecules at the bottom of the SVM ranking, which makes data analysis faster.

2.5.2 Determination of folding rate constants from trajectories

Nucleic acid conformational dynamics are fundamentally stochastic, and the time required to transition between states is random. If the folding dynamics are memoryless, in other words, the nucleic acid's probability of transitioning between states depends only on its current state, then the conversion between conformations is a Poisson process and can be described using first order kinetics. The transition rate from conformation *A* to conformation *B* is governed by the rate constant $k_{A\to B}$, which is the probability per unit time per molecule that *A* transitions to *B*, and the time rate of change of the concentration of *A* is the sum of all processes creating and depleting *A*:

$$\frac{d[A]}{dt} = \sum_i k_{i\to A}[i] - \sum_i k_{A\to i}[A]. \quad \text{(Eq. 2.18)}$$

Traditional chemical kinetics involves solving these equations, analytically or numerically, for the concentration of a species as a function of time and fitting to experimental results to obtain

the rate constants. Single-molecule kinetics must take a different route, as concentration is an ensemble phenomenon. Two approaches to determining rate constants from single-molecule data are used in this work, dwell time analysis and hidden Markov modeling, which will now be described.

Dwell time analysis involves examining the distribution of time spent in each conformation to determine folding rate constants (Figure 2.10). First, FRET trajectories are split into periods of time spent in each state, or "dwells," by applying a threshold to the trajectory. A high signal to noise ratio is required to avoid random fluctuations in E(t) crossing the threshold and being measured as short dwells. Alternative methods for dwell time extraction exist which are supposedly more robust to FRET noise, such as minimum description length (MDL) analysis,[9-10] but these are beyond the scope of this work. The dwell times from many molecules are then used to construct the complementary cumulative distribution function, also called the survival function, S(t), for dwell times in each state, which is the probability that an observed dwell time in that state is longer than a time t. Given a kinetic model, the dwell time distribution for a given state or manifold of states can be determined by solving the kinetics for a modified kinetic system in which all rate constants generating the state(s) are set to zero. For a

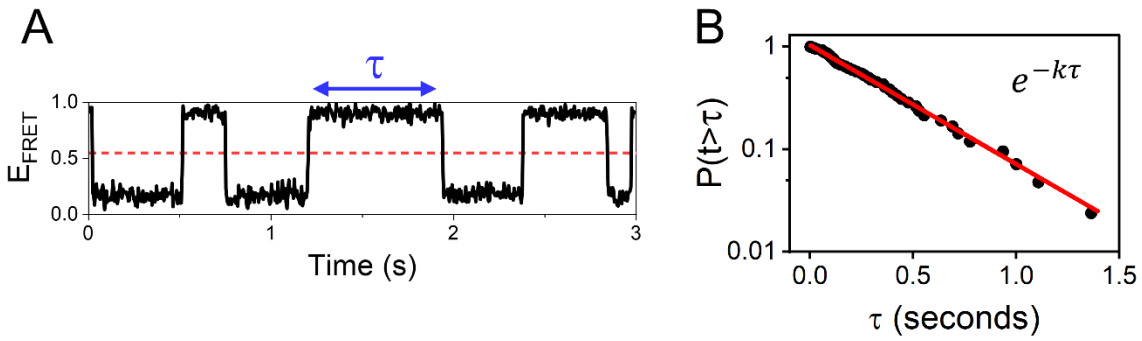

Figure 2.10 Dwell time analysis is carried out by (A) applying a threshold (red line) to the FRET trajectory to determine dwell times $\{\tau\}$ in each state. (B) The survival function S(τ) = P(t>τ) is computed by combining the dwell times from many molecules. For a two-state system, the rate constant k can be determined by fitting S(τ) to an exponential decay.

two-state system ($A \leftrightarrow B$), this is particularly simple and results in an exponential decay:

$$\frac{d[A]}{dt} = -k_{A\to B}[A], \quad \text{(Eq. 2.19)}$$

$$S(t) = \frac{[A](t)}{[A](t=0)} = e^{-k_{A\to B}t}. \quad \text{(Eq. 2.20)}$$

Fitting the dwell time distribution for a state to an exponential decay therefore yields the rate constant $k_{A\to B}$. Uncertainty estimation can be done by splitting the dwell times into N subsets (three or more), fitting each subset, and calculating the standard error of the mean

$$\sigma_{SEM} = \frac{1}{\sqrt{N}} std(k_i), \quad \text{(Eq. 2.21)}$$

where std(k_i) is the standard deviation of the rate constants for the N subsets.

A second approach to rate constant determination, hidden Markov modeling (HMM), describes the data probabilistically and can therefore be used on FRET trajectories with low signal to noise ratio for which dwell time thresholding fails. HMM uses the principal of maximum likelihood: the best model for explaining a set of data is the model for which the probability of observing the data is greatest. In other words, one selects model parameters θ which maximize the probability of observing the data $\{x_i\}$ given those model parameters, $P(\{x_i\}|\,\theta)$, which is also called the likelihood function L. The most important task in maximum likelihood estimation is the computation of the likelihood function, which requires a precise statement of the model being used. In HMM, the model consists of a set of "hidden" states $\{j\}$ (conformations) that cannot be directly observed but are instead inferred from an observable (the FRET efficiency E), with each state having its own "emission" probability distribution $P(E \mid j)$. The system can transition between states in discrete time as governed by a set of transition probabilities per frame $T_{j\to k}$, which can be conveniently arranged as a matrix $\mathbf{T}$. Therefore, the likelihood function L can be computed using the so-called forward algorithm:

$$L = \mathbf{1}^T [\prod_{i\geq 2} \mathbf{O}(E_i)\mathbf{T}\,]\mathbf{O}(E_1)\boldsymbol{p}_{eq}. \quad \text{(Eq. 2.22)}$$

This matrix product is read right to left. First, one determines the equilibrium probability vector $\boldsymbol{p}_{eq}$, which is the unique eigenvector of **T** with eigenvalue 1. One then multiplies this vector by the observation matrix **O**, which is a diagonal matrix whose diagonal elements are the probabilities of observing the i^{th} FRET value E_i for each state (**O** = diag(P(E_i | j)). A simple model for the observation probabilities P(E_i | j) is each state having a Gaussian FRET distribution. Next, multiplication by **T** advances time to the next frame, and the process continues until all frames have been used. Finally, one multiplies by $\mathbf{1}^{\mathrm{T}}$, a row vector of all ones, which serves to sum the elements of the remaining column vector.

As a sample computation, consider a FRET trajectory which begins E = {0.5, 0.65, …} for a two-state system with $\mathbf{T} = \begin{bmatrix} 0.5 & 0.25 \\ 0.5 & 0.75 \end{bmatrix}$ (for which $\boldsymbol{p}_{eq} = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix}$) and Gaussian FRET distributions with centers $\mu_1 = 0.3$ and $\mu_2 = 0.6$ and standard deviations $\sigma_1 = \sigma_2 = 0.15$. The probability density of observing the first FRET value ($E_1 = 0.5$) in state 1 is P(E_1 |1) = $\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{E_1-\mu_1}{\sigma_1}\right)^2} = 1.09$, and likewise is 2.13 for state 2. Therefore, the observation matrix $\mathbf{O}(E_1)$ is $\begin{bmatrix} 1.09 & 0 \\ 0 & 2.13 \end{bmatrix}$. Therefore, the likelihood product after the first frame is $\mathbf{O}(E_1) \cdot \boldsymbol{p}_{eq} \approx \begin{bmatrix} 0.36 \\ 1.42 \end{bmatrix}$. Next, time is advanced to the second frame by multiplication by **T** to give $\mathbf{T} \cdot \mathbf{O}(E_1) \cdot \boldsymbol{p}_{eq} \approx \begin{bmatrix} 0.54 \\ 1.24 \end{bmatrix}$. The observation matrix for the second FRET value ($E_2 = 0.65$) is computed like the first to give $\mathbf{O}(E_2) = \begin{bmatrix} 0.17 & 0 \\ 0 & 2.52 \end{bmatrix}$ which in turn yields $\mathbf{O}(E_2) \cdot \mathbf{T} \cdot \mathbf{O}(E_1) \cdot \boldsymbol{p}_{eq} \approx \begin{bmatrix} 0.09 \\ 3.14 \end{bmatrix}$. This process of alternating multiplication by **T** and **O**(E_i) is continued until the end of the trajectory, at which point L is calculated by summation of the vector elements. For example, the probability density of observing this trajectory after two frames is $L = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0.09 \\ 3.14 \end{bmatrix} = 0.09 + 3.14 = 3.23$.

Note that the forward algorithm calculation almost always results in extremely large or small values for L which risk numerical overflow or underflow, respectively. Therefore, the vector should be periodically renormalized, and the logarithm of L should be computed instead. By describing the first order kinetic system using a rate matrix **K**, where the off-diagonal element K_{ij} is the rate constants for the transition from state j to state i ($k_{j\rightarrow i}$) and the diagonal elements K_{ii} are the negative sum of rate constants leaving that state ($-\sum_i(k_{i\rightarrow j})$), we can calculate **T** as

$$\mathbf{T} = \exp\left(\boldsymbol{K}t_{frame}\right), \qquad \text{(Eq. 2.23)}$$

where t_{frame} is the length of each frame in the trajectory, and the exp() operator is the matrix exponential. It is possible to optimize L using the elements of **T**, as is done in many HMM packages for smFRET analysis, but determining **K** from **T** requires either computing the matrix logarithm of **T**, which can be ill-defined, or using the assumption $\mathbf{T} \approx \mathbf{I} + \mathbf{K}t_{frame}$, which is only true for small t_{frame}. Instead, I perform the optimization directly on the rate constants in **K**, which obviates these issues. The code for this analysis is written in MATLAB; faster implementation in a lower-level language like C would have significant speed improvements.

2.5.3 Dependence of rate constants on temperature and cosolute concentration

A powerful tool for illuminating the mechanism and thermodynamics of nucleic acid folding is the measurement of folding rate constants under a variety of environmental conditions. Two such environmental variables employed heavily in this dissertation are temperature and the concentration of cosolutes such as monovalent cations. By changing the temperature, one can learn about the relative enthalpy H and entropy S (and occasionally heat capacity C_p) of the reactants, products, and the transition state. The calculation of these quantities is done by van't Hoff and Arrhenius analysis (for a thorough description, see Chapter 6). Care must be taken

when interpreting measured thermodynamic changes, as *H* and *S* depend on the nucleic acid as well as its environment (e.g., bound waters and diffuse ions in the nucleic acid's ion atmosphere).

The concentration-dependence of rate constants can elucidate mechanistic details for processes that are otherwise unobservable. For instance, the binding of a ligand to a nucleic acid during folding can occur via two pathways: (i) the ligand can bind to the nucleic acid which induces it to fold ("induced fit"), or (ii) the ligand can stabilize a conformation which the nucleic acid can sample without ligand ("conformational selection"). These mechanisms have distinct kinetic behaviors which can be observed by changing the ligand concentration (for an example, see Chapter 7). Such a kinetic analysis requires solving a particular kinetic model; more general is the model-free measurement of the net stoichiometry of ligand binding using *preferential interaction coefficients*, as described in Appendix 4.

2.6 Complementing experiments with molecular dynamics simulations

The results in this dissertation are primarily derived from single-molecule experimental measurements. However, thanks to the development of easy-to-use computational tools by theorists, it has become increasingly painless[**] for experimentalists to turn to computation to address questions that their experiments cannot answer. Nucleic acids are far too large for quantum mechanical calculations, but they are suitable for molecular dynamics (MD) simulations, in which a classical force field is applied to determine the time-evolution of the atomic positions of the nucleic acid. The principles of MD are described in many excellent

[**] Even with user-friendly software, the learning curve for MD can be steep. Fortunately, a quick internet search reveals a number of online tutorials that can expediate the learning process.

review articles;[11] here, I will describe the process for preparing, running, and analyzing MD calculations as pertains to this thesis.

One begins with an initial structure of the desired nucleic acid. This is usually a crystal structure from the online protein database (PDB), but some simple structures such as hairpins or duplexes can be generated by tertiary structure prediction software. Then, the structure must be prepared for simulation. This includes removing any unnecessary crystalizing agents that may be in the PDB, possibly modifying the identify of cations[††], and perhaps mutating residues if desired. Additionally, the nucleic acid is placed in a box of water molecules,[‡‡] and ions are added to achieve a desired cation concentration.[§§] The next key step is parameterize the simulation, i.e., to determine the forces between each atom. In this work, I use the Amber force field (OL3 for RNA and OL15 for DNA) with the TIP3P rigid water model and Cheatham-Young ion parameters. For other cosolutes, there occasionally exist optimized force fields in the literature or else the general Amber force field (GAFF) can be applied. Manual editing of the starting structure is done in VMD, and parameterization and water box creation are done using LEaP in the AmberTools suite. Finally, the simulation is run in the freely available NAMD program on

[††] Most force fields treat atoms as point charges, and this approach is well-known to poorly describe polarizable molecules, notably multivalent ions. Therefore, it is recommended that any multivalent ions that are not essential to the structure be removed.

[‡‡] The choice of water model is an important decision in the design of an MD calculation. To reduce the computational cost of MD calculations, water is often simplified to a rigid 3-point model with the hydrogens having no van der Waals interactions (e.g., TIP3P). There are numerous water models available (e.g., TIP4P, OPC, OPC3, SPC/Fw), each with their advantages and disadvantages. For instance, currently, the OPC and OPC3 water models best reproduce the behavior of bulk water, at the cost of highly non-physical water geometries.

[§§] There is a common practice of adding only enough ions to neutralize the simulation (e.g., if the nucleic acid has a charge of −30 *e*, then 30 K^+ ions are added). This results in a bulk cation concentration of zero, which far removed from cytoplasmic concentrations. An ionic strength of zero may be fine for proteins with low net charge, but highly charged nucleic acids are known to behave quite unusually at low bulk ion concentrations. Therefore, I strongly recommend simulating at no less than 50 mM ionic strength (i.e., if using KCl, add enough ion pairs so that $[Cl^-] >$ 50 mM).

the JILA compute cluster. Depending on the size of the simulation (typically 30,000-100,000 atoms) and number of compute cores used (8-32), simulations run at 5-25 ns per day.

The final step is to analyze the simulation trajectory. Though the effective computational speed can be increased by running multiple simulations in parallel, it is clearly infeasible to reach timescales long enough to observe nucleic acid folding (typically longer than 10 ms), and even state-of-the-art calculations rarely exceed one millisecond. Therefore, MD calculations must be used to observe other processes, such as the interaction of a nucleic acid with a ligand. There are numerous options available for MD analysis, each tailored to answer different research questions. One especially useful calculation within the reach of microsecond-limited simulations is the measurement of ligand binding free energies. This can be done either by observing ligand binding in an unbiased simulation (i.e., with no added forces), as in Chapter 4, or by enhanced sampling methods (e.g., umbrella sampling, collective variable biasing, and replica exchange), or by alchemical transformations (e.g., free energy perturbation). The latter two options require additional modifications to the system preparation or simulation settings in NAMD. Whatever the purpose of the MD calculation, establishing an analysis plan *before* performing running simulations is an important step to avoid wasted compute time.

2.7 References

1. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B.* **2020,** *124* (51), 11561-11572.

2. Nicholson, D. A.; Nesbitt, D. J., Pushing camera-based single-molecule kinetic measurements to the frame acquisition limit with stroboscopic smFRET. *J. Phys. Chem. B* **2021,** *125* (23), 6080-6089.

3. Hodak, J. H.; Fiore, J. L.; Nesbitt, D. J.; Downey, C. D.; Pardi, A., Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (30), 10505.

4. Fiore, J. L.; Hodak, J. H.; Piestert, O.; Downey, C. D.; Nesbitt, D. J., Monovalent and divalent promoted GAAA tetraloop-receptor tertiary interactions from freely diffusing single-molecule studies. *Biophys. J.* **2008,** *95* (8), 3892-3905.

5. Axelrod, D.; Burghardt, T. P.; Thompson, N. L., Total internal reflection fluorescence. *Ann. Rev. Biophys. Bioeng.* **1984,** *13*, 247-268.

6. 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.

7. Brown, L. G., A survey of image registration techniques. *ACM Comput. Surv.* **1992,** *24* (4), 325–376.

8. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A., A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020,** *408*, 189-215.

9. Shuang, B.; Cooper, D.; Taylor, J. N.; Kisley, L.; Chen, J.; Wang, W.; Li, C. B.; Komatsuzaki, T.; Landes, C. F., Fast step transition and state identification (STaSI) for discrete single-molecule data analysis. *J. Phys. Chem. Lett.* **2014,** *5* (18), 3157-3161.

10. Hu, B.; Rakthanmanon, T.; Hao, Y.; Evans, S.; Lonardi, S.; Keogh, E., Using the minimum description length to discover the intrinsic cardinality and dimensionality of time series. *Data Min. Knowl. Discov.* **2015,** *29* (2), 358-399.

11. Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P., et al., RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chem. Rev.* **2018,** *118* (8), 4177-4338.

Chapter 3

Amino Acid Stabilization of Nucleic Acid Secondary Structure: Kinetic Insights from Single Molecule Studies[*]

3.1 Abstract

Amino-acid and nucleic acid interactions are central in biology and may have played a role in the evolutionary development of protein-based life from an early "RNA Universe." To explore the possible role of single amino acids in promoting nucleic acid folding, single-molecule Förster resonance energy transfer (smFRET) experiments have been implemented with a DNA hairpin construct (7 nucleotide double strand with 40A loop) as a simple model for secondary structure formation. Exposure to positively charged amino acids (arginine and lysine) is found to clearly stabilize secondary structure. Kinetically, each amino acid promotes folding by generating a large increase in the folding rate with little change in the unfolding rate. From analysis as a function of temperature, arginine and lysine are found to significantly increase the overall exothermicity of folding while imposing only a small entropic penalty on the folding process. Detailed investigations into the kinetics and thermodynamics of this amino acid-induced folding stability reveal arginine and lysine to interact with nucleic acids in a manner reminiscent of monovalent cations. Specifically, these observations are interpreted in the context of an ion atmosphere surrounding the nucleic acid, in which amino acid salts stabilize folding qualitatively like small monovalent cations, but also exhibit differences due to the composition of their side chains.

---

[*] This chapter is adapted from: Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J. *J. Phys. Chem. B.* **2018,** *122* (43), 9869-9876.

3.2 Introduction

Interactions between amino acids and nucleic acids are biologically ubiquitous, lying at the core of numerous protein-nucleic acid interactions, such as histone-driven chromatin dynamics[1] and nucleic acid packaging into viral particles.[2-3] Individual amino acids have been found to influence RNA functionality; for instance, many of the riboswitches discovered to date have specific binding sites for amino acids,[4] including glycine,[5-6] lysine[7-8] and arginine.[9] RNA catalysis can also be affected by amino acids, as demonstrated in pioneering work by Yarus and coworkers, in which the self-splicing capability of the *Tetrahymena* ribozyme is inhibited by the presence of L-arginine.[10] Indeed, amino acid influence over RNA catalytic properties has prompted an interesting suggestion for the mechanism by which early RNA-based life[11-12] transitioned into the modern "protein world" via an intermediate step of amino acids acting as RNA-folding chaperones.[13] In this hypothesis, early RNA-based structures which struggle to reach or remain in their biologically competent conformations may have benefited from the presence of amino acids which assist in the folding of RNA. This chaperoning could have begun with abiotically-generated amino acids, with an evolutionary path eventually leading to the biological synthesis of amino acids and small peptides, and ultimately to the protein world of today. This mechanism is speculative yet intriguing, and clearly deserves quantitative evaluation.

Stimulated by the biological implications of amino acid-nucleic acid interactions, crystallographic[14-16] and computational[16-19] studies have been performed, yielding valuable structural information, but the influence of amino acids on nucleic acid folding kinetics is less well-studied. Recently, single-molecule fluorescence microscopy tools have been applied to the study the amino acid-based promotion or inhibition of the ubiquitous tetraloop-tetraloop receptor tertiary interaction.[20] Indeed, amino acids (arginine in particular) were found to *destabilize* the

tertiary folding equilibrium by diminishing folding rates via competitive binding to the GAAA tetraloop. Even the simplest amino acid glycine exhibited a significant decrease in unimolecular unfolding rates for the tetraloop receptor. The present work both builds on and extends these single molecule efforts, focusing on secondary structure rather than tertiary structure. Research into nucleic acid secondary structure is quite mature, complete with apparatus for prediction of hybridization thermodynamics (e.g. nearest-neighbor models[21]), representing an area of nucleic acid behavior that is well-characterized.[22] Exploring the influence of amino acids on secondary structure formation allows us to isolate dynamical effects in a folding process which is better understood than tertiary folding. To perform these experiments, we employ a DNA hairpin as a simple model for nucleic acid secondary structure, which can be interrogated for its conformational status by single-molecule fluorescent resonant energy transfer (FRET) microscopy.[23-25] In contrast to the previous studies on tertiary folding, only the charged amino acids (lysine and arginine, rather than simple zwitterionic species such as glycine) alter the hairpin's folding dynamics. Furthermore, these charged amino acids act to *stabilize* the folded conformation, which is exactly the opposite effect noted for tertiary structures. Implementation of temperature-dependent studies permits folding and transition state free energies to be deconstructed into enthalpic and entropic components.[25] The combination of kinetic and thermodynamic data suggests an ionic screening mechanism in which charged amino acids alleviate the electrostatic cost of nucleic acid folding. This electrostatic interaction is the predominant source of amino acid-based secondary structure stabilization, though the sensitivity of the nucleic acid to the molecular detail of the side chain confers additional, amino acid-specific contributions.

3.3 Methods

The DNA hairpin construct used in these studies has three components (Figure 3.1): (1) a 7 base-pair stem (5'-CTTCAGT-3'), (2) a loop consisting of 40 adenine residues, and (3) an 11 adenine tail at the 3' terminus, end-labeled with biotin for surface attachment. This specific hairpin has no biological significance; rather, the sequence is designed to avoid misfolding or the formation of higher-order structures, as well as to achieve kinetic rates within the current dynamic range of the instrument (< 100 s^{-1}). For spectroscopic readout of conformation, a cyanine-based FRET fluorophore pair is used, consisting of Cy5 at the 5' terminus and internal Cy3 joining the 3' end of the loop with the stem. The complete DNA oligonucleotide (5'-Cy5-CTTCAGT-A_{40}-Cy3-ACTGAAG-A_{11}-Biotin-3') is purchased from Integrated DNA Technologies.

This protocol yields a surface density of ~1/μm^2, which is sufficiently well-resolved for optical microscopy with a high numerical aperture (NA) objective. Measurements are taken in buffer (50 mM HEPES, pH 7.6) with an enzymatic oxygen scavenging system (2 mM TROLOX, 100 nM PCD, 5 mM PCA) to improve fluorophore photostability.[27] All experiments include a

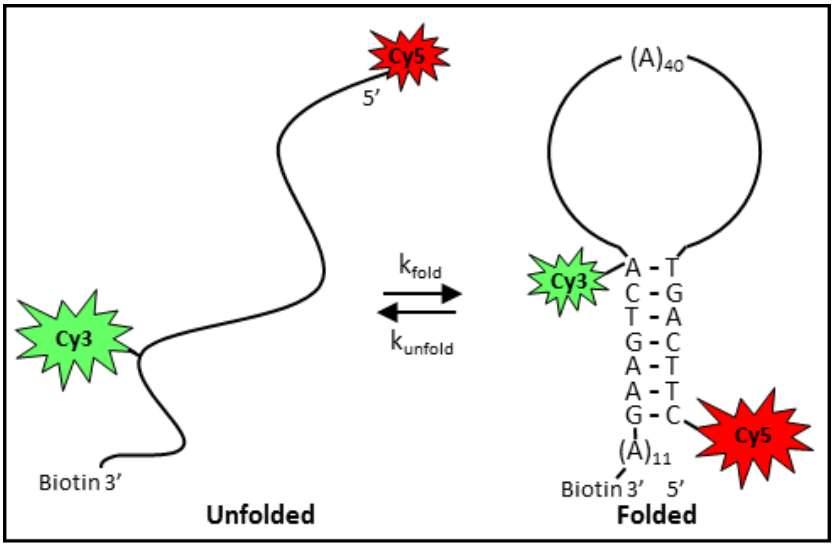

Figure 3.1 DNA hairpin construct used in folding studies.

background of 100 mM monovalent ions (70 mM NaCl and 30 mM KCl for concentration-dependent studies; 100 mM NaCl for temperature-dependent studies). Other solutes are added as described below, using compounds purchased from Sigma.

Single DNA molecules are illuminated in a scanning confocal microscope,[28] as previously described.[20] Briefly, 532 nm laser light is focused through a 1.2 NA water immersion microscope objective to a 270 nm diffraction-limited spot. The spot is scanned over the coverslip surface by a piezoelectric stage to locate DNA molecules. Emission from fluorophores is collected by the microscope objective, split by dichroic filter into donor and acceptor colors, and directed onto single-photon-counting avalanche photodiodes. Concentration-dependent studies are performed at 24 °C, and temperature studies are carried out using a stage heater (HSC60, Instec) and a microscope objective heating collar (Bioptechs) with 0.1 °C temperature accuracy and stability.[29]

3.4 Results

DNA hairpin folding in the absence of amino acids demonstrates that the construct behaves according to expectations in several ways. First, as demonstrated in Figure 3.2a, fluorescence trajectories depict switching between two states, indicating that the hairpin is converting between an unfolded low-FRET state and a folded high-FRET state without significantly sampling additional conformations. Second, the high-FRET state has an average energy transfer efficiency of 0.88, corresponding to an interfluorophore distance of ~38 Å, which agrees well with an estimated distance of 34 Å for a 7 base-pair DNA duplex with B-form dimensions.[30] Third, the fractional residence time in each state can be used to compute a folding equilibrium constant ($K_{eq} = 0.74$), from which the free energy of folding ($\Delta G° = -RT \ln(K)$) is

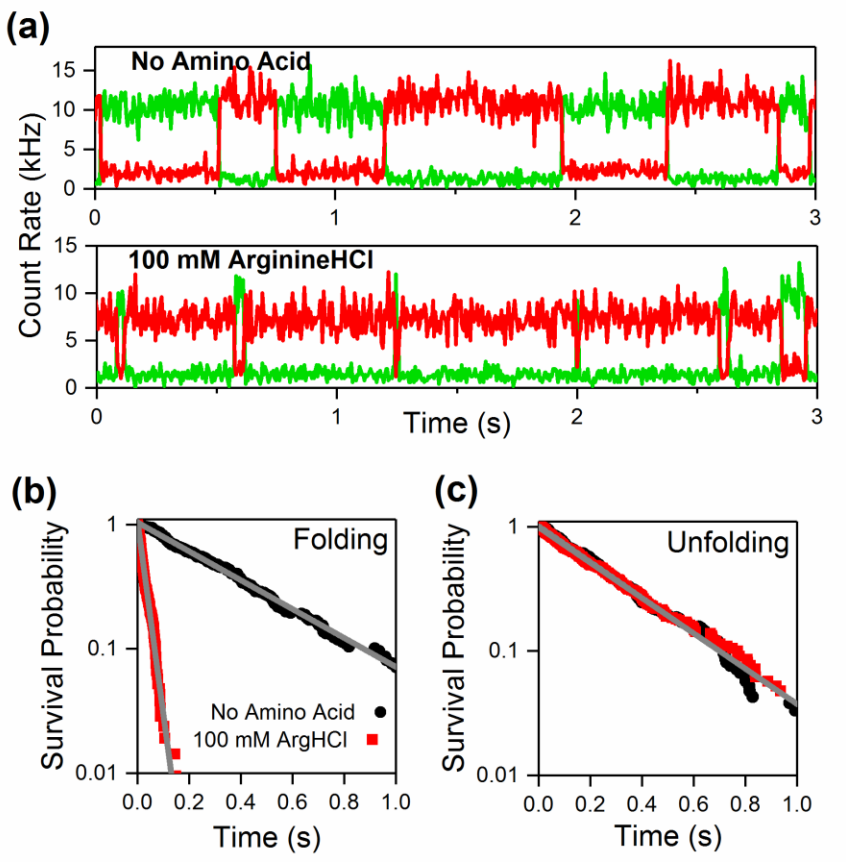

Figure 3.2 (a) Single molecule time traces for hairpin folding with and without 100 mM ArginineHCl, exhibiting two-state behavior. Green and red lines correspond to donor (Cy3) and acceptor (Cy5) fluorescence, respectively. Residence durations are used to construct survival time distributions for **(b)** folding and **(c)** unfolding. Exponential fits are used to extract rate constants. Data are obtained in a background of 70 mM Na^+ and 30 mM K^+.

determined to be 0.75 kJ mol^{-1}. This is comparable to 1.9 kJ mol^{-1} nearest-neighbor model prediction[21] for this hairpin at 24 °C in 100 mM Na^+. Indeed, this ~1 kJ mol^{-1} discrepancy is quite small and attributable to differing conditions in which nearest-neighbor parameters were measured.[31] Fourth, the residence time distributions in each state are well-fit to a first-order kinetic model, as seen in Figure 3.2b, which indicates that folding and unfolding are each governed by a single rate-limiting step, in agreement with literature observations of the folding kinetics of short hairpins.[32-33]

Next, the influence of arginine on hairpin folding is examined. Shown in Figure 3.2a, the addition of 100 mM arginine hydrochloride (ArginineHCl) results in the DNA hairpin spending dramatically less time in the unfolded state while remaining in the folded state for an unchanged amount of time. These qualitative observations can be quantitatively confirmed by analysis of the residence time distributions (Figure 3.2b), in which the folding rate constant is increased by more than an order of magnitude from $2.6 \pm 0.2\ s^{-1}$ in plain buffer to $26.7 \pm 1.5\ s^{-1}$ with 100 mM ArginineHCl. In contrast, the unfolding rate constant is statistically unchanged ($2.3 \pm 0.1\ s^{-1}$ to $2.3 \pm 0.3\ s^{-1}$). By selectively increasing the rate of folding, arginine achieves stabilization of the DNA hairpin.

To explore the mechanism underlying hairpin stabilization, arginine is compared side-by-side with glycine and lysine. Glycine isolates the contribution of the zwitterionic backbone, while lysine represents swapping arginine's guanidinium-based side chain for an ammonium moiety. As seen in Figure 3.3, glycine has little influence on either the folding or unfolding of the hairpin, establishing that the zwitterionic backbone is not sufficient for stabilization of the secondary structure. Lysine hydrochloride (LysineHCl), on the other hand, is able to stabilize the hairpin, again primarily by increasing the folding rate with only a minor slowing of unfolding, in a similar fashion to arginine. However, the magnitude of lysine's influence is not as large: 100 mM ArginineHCl speeds up folding over nine-fold, compared to the four-fold increase in 100 mM LysineHCl. Thus, the positively-charged side chain in lysine is sufficient to stabilize the hairpin, but the guanidinium group in arginine clearly provides an additional stabilizing contribution.

The net positive charge of arginine and lysine at the experimental pH suggests that the mode of interaction with DNA may be primarily electrostatic. To evaluate this possibility,

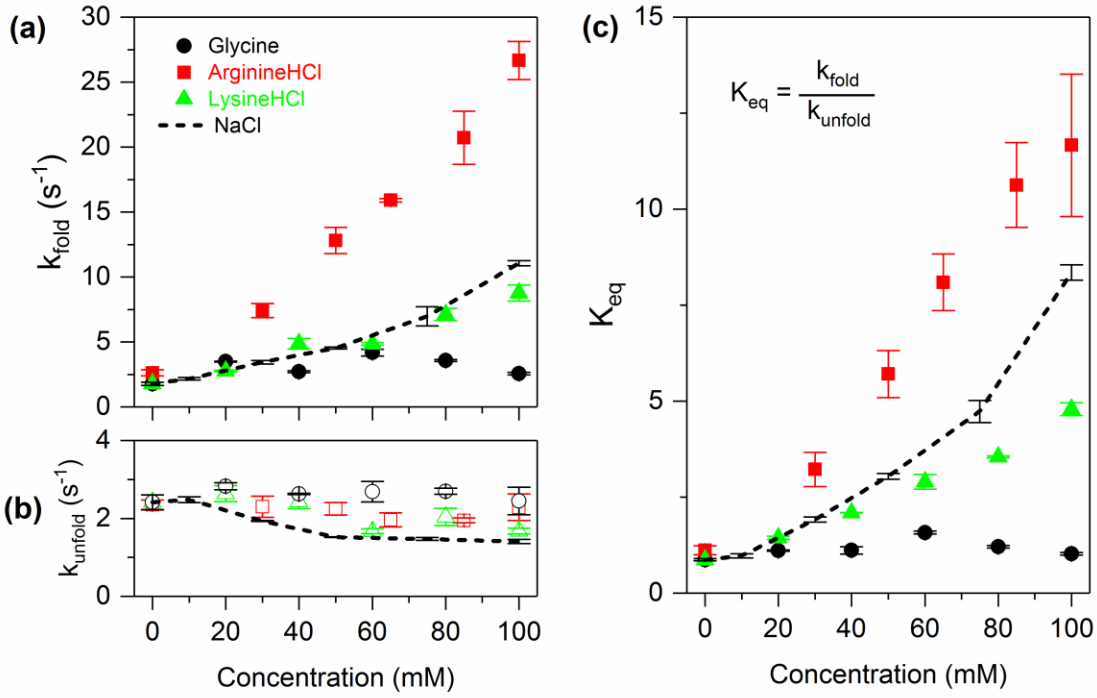

Figure 3.3 Effects of amino acids on DNA hairpin folding. **(a)** The folding rate constant is strongly increased, while **(b)** the unfolding rate constant is slightly decreased. **(c)** The equilibrium constant shows little sensitivity to glycine, while positively-charged amino acids stabilize secondary structure. Sodium chloride (dashed line) is shown for reference.

hairpin folding measurements have been carried out in the presence sodium chloride, a simple monovalent cation (Figure 3.3, dashed line; note that the concentration represents *additional* sodium chloride beyond the 100 mM NaCl present in all samples). Sodium chloride exhibits hairpin stabilization via the same kinetic signature as the amino acid salts, *viz.* increasing the folding rate with a small effect on the unfolding rate. This is in agreement with previous studies of monovalent cation influence on secondary structure formation.[34] However, the net stabilization (η) illustrated by the equilibrium constants in Figure 3.3b was different for each species, resulting in an order of stabilization: $\eta_{ArgHCl} > \eta_{NaCl} > \eta_{LysHCl}$. Though their effects are different in magnitude, all three cations display a similar kind of influence on folding kinetics for

the DNA hairpin, in general corroborating the heuristic connection between amino acid charge and hairpin stabilizing power.

Further insight comes from temperature-dependent measurements, which permit the deconstruction of free energies into entropic and enthalpic contributions. The van't Hoff plot in Figure 3.4a displays the influence of temperature on the DNA hairpin's folding equilibrium constant with and without 100 mM ArginineHCl, LysineHCl, and NaCl over the range of 24–29 °C. Linear fits to these data yield the entropy (ΔS°) and enthalpy (ΔH°) of folding, according to the van't Hoff equation,

$$\ln(K) = -\frac{\Delta G^\circ}{RT} = -\frac{\Delta H^\circ}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^\circ}{R}, \qquad \text{(Eq. 3.1)}$$

where T is the absolute temperature and R is the gas constant. The enthalpies and entropies of folding, derived from these slopes and intercepts respectively, are displayed in Figure 3.4b. Folding of the hairpin without any additional solute (leftmost column) is driven by significant enthalpic release (ΔH° = -138 ± 21 kJ mol^{-1}) offset by a large entropic penalty (ΔS° = -470 ± 70 J mol^{-1} K^{-1}), in agreement with previous observations of hairpin formation.[35-36] Each of the added amino acids *enhances the enthalpic favorability* (average $\Delta\Delta H^\circ$ = -46 ± 23 kJ mol^{-1}; dashed line in upper panel of Figure 3.4b), while also *increasing the entropic penalty* associated with folding (average $\Delta\Delta S^\circ$ = -130 ± 80 J mol^{-1} K^{-1}). The net result between these opposing effects is increased stabilization ($\Delta\Delta G^\circ < 0$); i.e. the enthalpic benefit provided by each of the solutes is stronger than the additional entropic cost incurred. However, any differential effect between these three species is statistically indistinguishable from zero within experimental uncertainty (see Figure 3.4b).

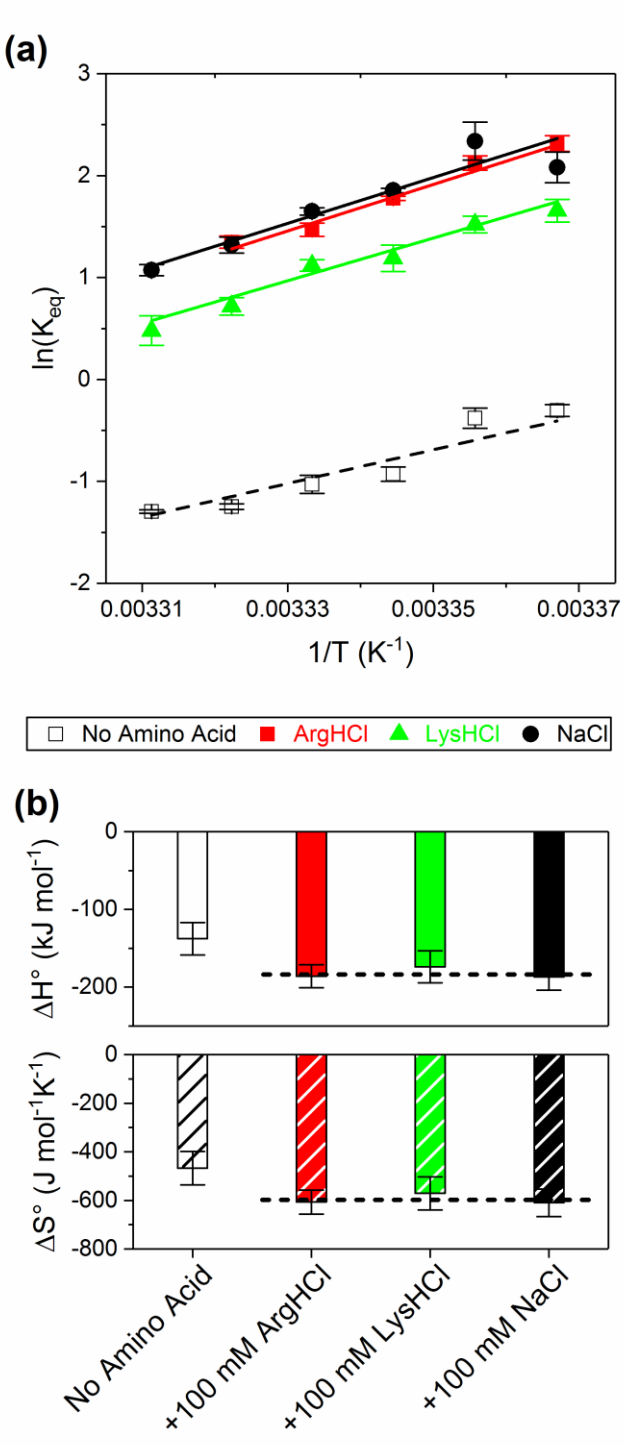

Figure 3.4 (a) van't Hoff plot of the hairpin folding equilibrium's temperature dependence. Colored symbols have 100 mM solute. Black squares have sodium chloride added to make the same ionic strength as colored squares, i.e., 200 mM total [NaCl]. Linear fits are used to extract ΔH and ΔS contributions, shown in **(b)**. Dashed lines indicate average ΔH and ΔS for added solutes.

Similarly, the temperature dependence of the *rate constants* for folding may be used to extract thermodynamic parameters for approaching the folding transition state. To achieve this, the transition state theory expression for the rate constant may be recast into the more convenient, activated complex form:

$$\ln(k) = \ln(\upsilon) + \frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{R}\left(\frac{1}{T}\right), \qquad \text{(Eq. 3.2)}$$

where υ is the attempt frequency along the reaction coordinate, and $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$ are changes in entropy and enthalpy, respectively, required to reach the transition state. In a plot of ln(k) against 1/T, the slope of a linear fit reveals $\Delta H^{\ddagger}$ unambiguously, whereas the intercept reflects information on both $\Delta S^{\ddagger}$ as well as υ. Since the attempt frequency υ is not known *a priori*, the Arrhenius plot does not provide the *absolute* activation entropy. However, if υ is assumed to be constant over the range of conditions in this study, then the solute-induced *change* in the activation entropy ($\Delta\Delta S^{\ddagger}$) can be determined unambiguously. Figure 3.5a reveals activated complex theory Arrhenius plots for the folding rate constant of the DNA hairpin, with the least-squares fitting results summarized in Figure 3.5b. Similar to what was noted in the above van't Hoff analysis, it is difficult to distinguish the more detailed differences between ArginineHCl, LysineHCl, or NaCl. However, all three salts do act in the same general fashion: the enthalpic requirement to reach the transition state is significantly diminished (average $\Delta\Delta H^{\ddagger}$ = -54 ± 15 kJ mol^{-1}), while the transition state becomes entropically less accessible (average $\Delta\Delta S^{\ddagger}$ = -160 ± 50 J mol^{-1} K^{-1}). Notably, the thermodynamic parameters of the folding reaction are similarly impacted by arginine and lysine (Figure 3.5b), which in turn accounts for their comparable influences on the folding kinetics.

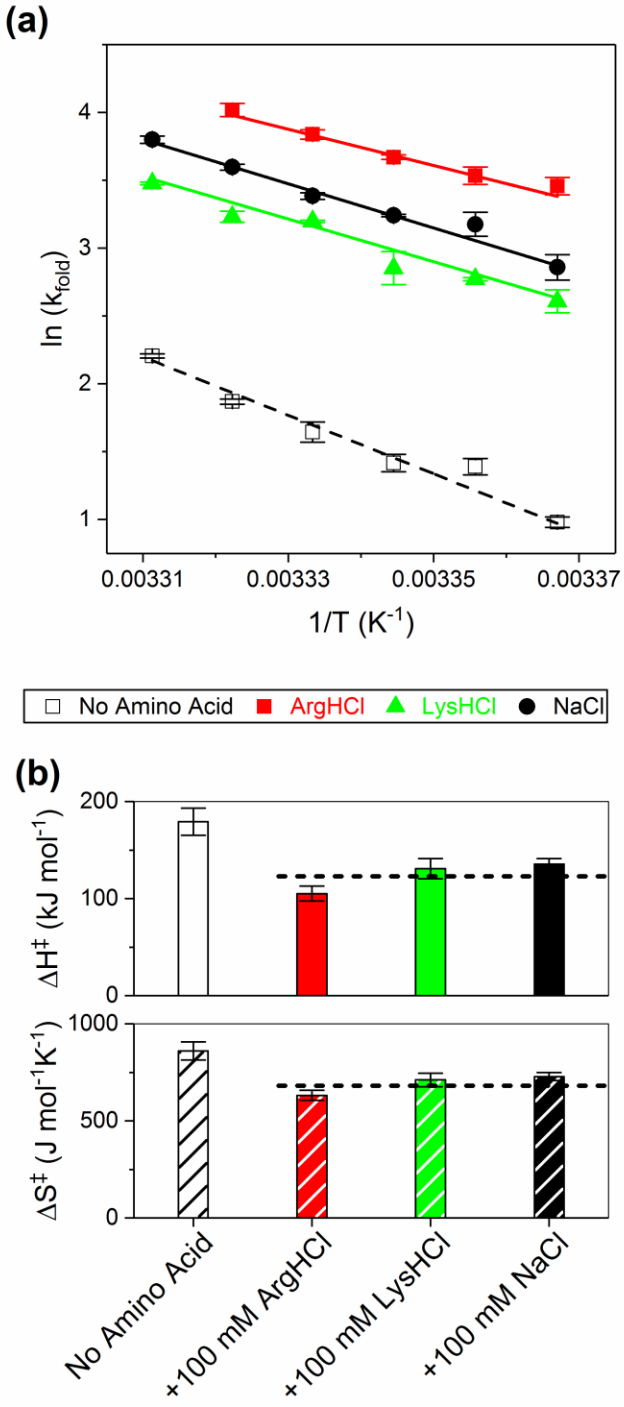

Figure 3.5 (a) Arrhenius plot for hairpin folding rate coefficient as a function of temperature. Linear fits yield activation enthalpies and entropies in **(b)**. Dotted line indicate averaged enthalpies and entropies, not including "no amino acid."

3.5 Discussion

Previous work has shown the ability of amino acids, especially L-arginine, to alter ribozyme activity.[10] The suggestion that the origin of such inhibitory behavior lies in amino acids modification of nucleic acid folding properties has prompted this series of single-molecule investigations into the kinetic behavior of nucleic acids in amino acid environments. Prior efforts examined an isolated tertiary interaction, the tetraloop-tetraloop receptor, revealing the ability of glycine, lysine, and especially arginine to destabilize the tertiary structure, each in an amino acid-specific manner.[20] The current study aims to bring the same analysis to bear on secondary structure formation kinetics, modeled by a DNA hairpin.

First, we have explored the ability of arginine, lysine, and glycine to influence secondary structure. Of these three amino acids, both arginine and lysine are found to stabilize the hairpin formation, whereas glycine has little or no effect. The insensitivity of hairpin folding to the presence of glycine suggest that the amino acid zwitterionic backbone is relatively unimportant in the mechanism by which arginine and lysine promote the folded state. Rather, it is the side chain which determines the effectiveness of the promoting interaction. ArginineHCl and LysineHCl have qualitatively similar influences on the hairpin's folding kinetics (Figure 3.3a) and thermodynamics (Figures 3.4 and 3.5), but the *magnitude* of these effects is different, with arginine offering more pronounced stabilization. That arginine and lysine have a shared kinetic and thermodynamic signature suggests that they also have a shared *mechanism* for stabilization, with the overall effect modulated by subtle differences in residue composition. In particular, arginine and lysine share the structural feature of a charged side chain, suggesting that charge state is the primary source of stabilization. If charge determines stabilizing power, then the addition of other monovalent cations should affect hairpin folding in a similar fashion. Indeed,

sodium chloride titrations produce kinetic and thermodynamic modifications comparable to those of the amino acid salts.

The ability of monovalent cations to promote nucleic acid folding is well known[34, 37] and can be explained by considering the role of the ionic atmosphere in folding.[38] Nucleic acids are highly charged polyanions which recruit electrolytes from solution to form an ionic cloud surrounding the polynucleotide, known as the "ionic atmosphere." A nucleic acid undergoing a conformational change will concomitantly experience a reorganization in its ionic atmosphere.[39-40] Importantly, folding events typically make the nucleic acid more compact, resulting in the net recruitment of ions from the bulk.[34, 41] Consequently, increasing salt concentration serves, by mass action, to shift the equilibrium to favor more compact, folded states. Our kinetic observations can be explained by examining the ionic atmosphere of the transition state. According to a "kinetic zipper" picture,[32, 42] the transition state for the DNA hairpin ought to be structurally similar to the folded state, with the loop already formed (and entropic costs already paid) and held together by a few nascent base pairs. Therefore, the transition state and the folded state will have similar ionic atmospheres, which would explain why the unfolding rate is found to be relatively insensitive to salt levels. On the other hand, the initial folding process requires net uptake of ions to reach the transition state, so the folding rate constant is anticipated to be strongly dependent on salt and/or charged amino acid concentration.

This salt-mediated promotion of secondary structure is partially opposed by an additional entropic cost arising from the salt-dependence of the polynucleotide flexibility. Repulsive forces between phosphates along the nucleic acid backbone tend to straighten the polynucleotide, limiting the conformational phase space sampled by the polymer. Increasing salt levels improve ionic screening, which diminishes phosphate repulsion down the backbone and decreases the

polymer's persistence length.[43-45] In turn, the nucleic acid can sample a greater region of conformational phase space, thereby increasing its entropy. For the DNA hairpin, both the folded and unfolded states enjoy this entropic increase, however the effect is greater in the unfolded state which has a longer single-stranded region with unconstrained ends. Therefore, the folding reaction suffers a net entropic cost due to the increase in bulk salt concentrations. The present temperature-dependent experiments have confirmed the prediction of a salt-induced entropic cost in the folding of the hairpin ($\Delta\Delta S° = -130 \pm 80$ J mol^{-1} K^{-1}, 100 mM added salt). Furthermore, the structural similarity of the folded state and the transition state implies that there should be a similar cost in the *activation* entropy for folding, and indeed the Arrhenius analysis reveals an activation entropy increase of equal magnitude ($\Delta\Delta S^{‡} = -160 \pm 50$ J mol^{-1} K^{-1}). It is worth noting that the salt-dependence of DNA *duplex* formation has been historically observed to be primarily entropic in origin[46] (i.e. $\Delta\Delta S > 0$; $\Delta\Delta H \approx 0$), though recent studies have identified non-negligible enthalpic contributions.[34, 47] That the DNA hairpin in these studies responds to monovalent salt levels in a thermodynamically distinct fashion ($\Delta\Delta S < 0$; $\Delta\Delta H < 0$) is a reminder of the ongoing need for research into the thermodynamic properties of salt-mediated nucleic acid folding (e.g. bimolecular vs. intramolecular processes, long vs. short hairpin loops, long vs. short stems). In particular, we anticipate the entropic effect to be different in hairpins with shorter loops, which are known to be less sensitive to salt overall.[41, 48]

Though arginine and lysine influence hairpin folding in a way that is qualitatively similar to a monovalent cation, the additional stabilizing ability of arginine over lysine suggests that the molecular detail of the side chain further modulates these amino acid-nucleic acid interactions. To isolate the effects of the amino acid side chain without the backbone, we have also explored DNA hairpin folding in the presence of guanidine hydrochloride (GuHCl), which imitates

arginine's guanidinium group, and ammonium chloride, which imitate lysine's ammonium group (Figure 3.6). Both salts are observed to stabilize the hairpin, with GuHCl having a stronger effect than NH_4Cl. This is unsurprising given guanidine's known strong interactions with nucleotides due to hydrogen bonding.[49-50] In addition, GuHCl and NH_4Cl were both found to have greater stabilizing power than their respective amino acid counterparts. Interestingly, the kinetic source of GuHCl's greater stabilization of the hairpin is depression of the unfolding rate constant (3-fold at 100 mM GuHCl) beyond what any of the other salts in this study achieve, so the guanidinium ion stabilizes the folded state relative to the transition state more effectively than any other cation in these experiments.

These observations lead to a simple picture explaining the relative stabilizing power of lysine, arginine, ammonium, and guanidium based on monovalent cation size[41, 51] and hydrogen bonding ability. Lysine interacts with the nucleic acid via a mechanism similar to that of an ammonium cation, but with additional molecular bulk which prevents lysine from approaching the hairpin as closely as ammonium. Lysine therefore cannot contribute to the ionic atmosphere as efficiently as ammonium, resulting in a restricted ability to stabilize secondary structure ($\eta_{\mathrm{LysHCl}} < \eta_{\mathrm{NH_4Cl}}$). Arginine, on the other hand, interacts with the hairpin like guanidinium, which is dominated by the molecule's positive charge but made even stronger due to the hydrogen-bonding abilities of the guanidinium moiety ($\eta_{\mathrm{GuHCl}} > \eta_{\mathrm{NH_4Cl}}$; $\eta_{\mathrm{ArgHCl}} > \eta_{\mathrm{LysHCl}}$). However, arginine suffers from the same steric hindrance as lysine, so arginine cannot stabilize the secondary structure of the nucleic acid as effectively as guanidinium ($\eta_{\mathrm{GuHCl}} > \eta_{\mathrm{ArgHCl}}$).

Several control tests of this model have been carried out. First, we examined the role of amino acid chirality in imparting stability to the hairpin. If the dominant mode of interaction

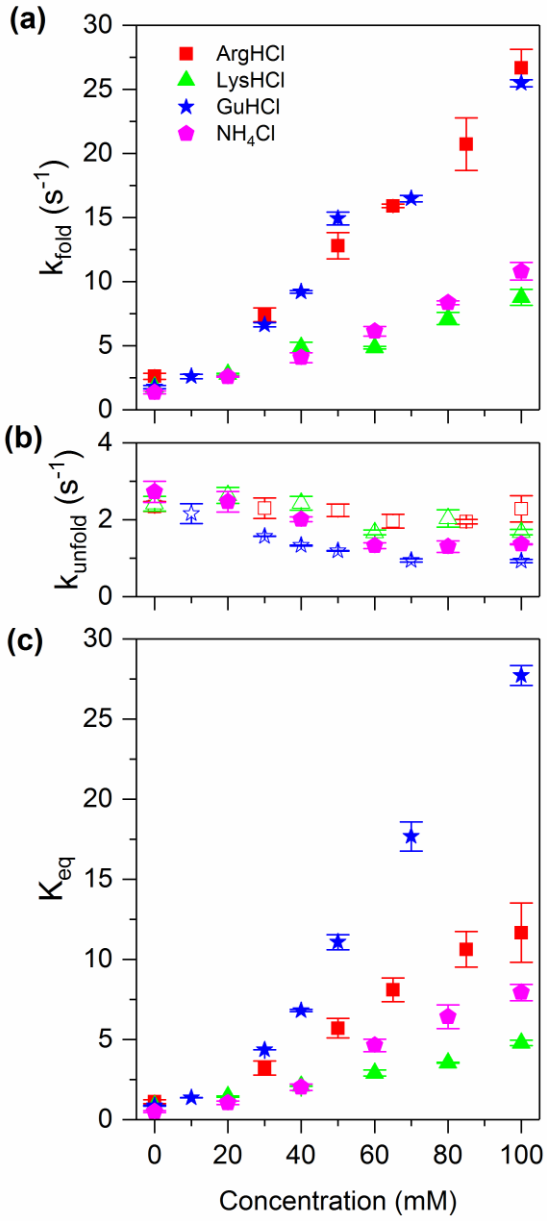

Figure 3.6 Comparison of positively-charged amino acids with their side chain moieties, for **(a)** folding rate constants **(b)** unfolding rate constants, and **(c)** equilibrium constants. ArginineHCl and GuanidineHCl have similar effects on folding kinetics, but GuanidineHCl has a stronger influence on unfolding rates. NH_4Cl has a slightly larger influence on both folding and unfolding over LysineHCl.

between amino acids and nucleic acids is electrostatic, then the folding of the hairpin should be insensitive to the chirality of the amino acid. In addition, the guanidinium group in arginine is achiral, so the hydrogen-bonding ability of arginine should not depend of the chirality of the distant α-carbon. As shown in Figure 3.7, smFRET experiments comparing L- and D-reagents reveal no difference between the enantiomers of ArginineHCl or of LysineHCl, in agreement with the model. Second, we tested to see if *negatively*-charged amino acids can participate in secondary structure stabilization. Since nucleic acids are highly negatively-charged, a negative amino acid like aspartate would not be expected to interact strongly with the hairpin. However, the negative amino acid must be accompanied by a positive counter-ion, which should participate in same stabilizing interactions as other cations. To test this, we compared the impact of sodium aspartate and sodium chloride on the folding kinetics of the DNA hairpin, depicted in Figure 3.8. The data confirm that the effect of sodium aspartate is indistinguishable from that of sodium chloride, supporting our model for nucleic acid-induced secondary structure stabilization.

3.6 Summary and conclusion

In this work, single-molecule analysis has been used to examine the ability of amino acids to influence the formation of secondary structure in nucleic acids. Positively charged amino acids (arginine and lysine) are found to stabilize secondary structure, while the zwitterionic amino acid glycine has no significant effect. The kinetic origin of the stabilization is primarily a large increase in the folding rate constant, with only a minor change in the unfolding rate. Thermodynamic analysis reveals that both arginine and lysine increase the reaction exothermicity ($\Delta\Delta H° < 0$) while also imposing a slight entropic penalty ($\Delta\Delta S° < 0$), yielding a net stabilizing effect ($\Delta\Delta G° < 0$). Furthermore, the barrier to folding is modified, characterized

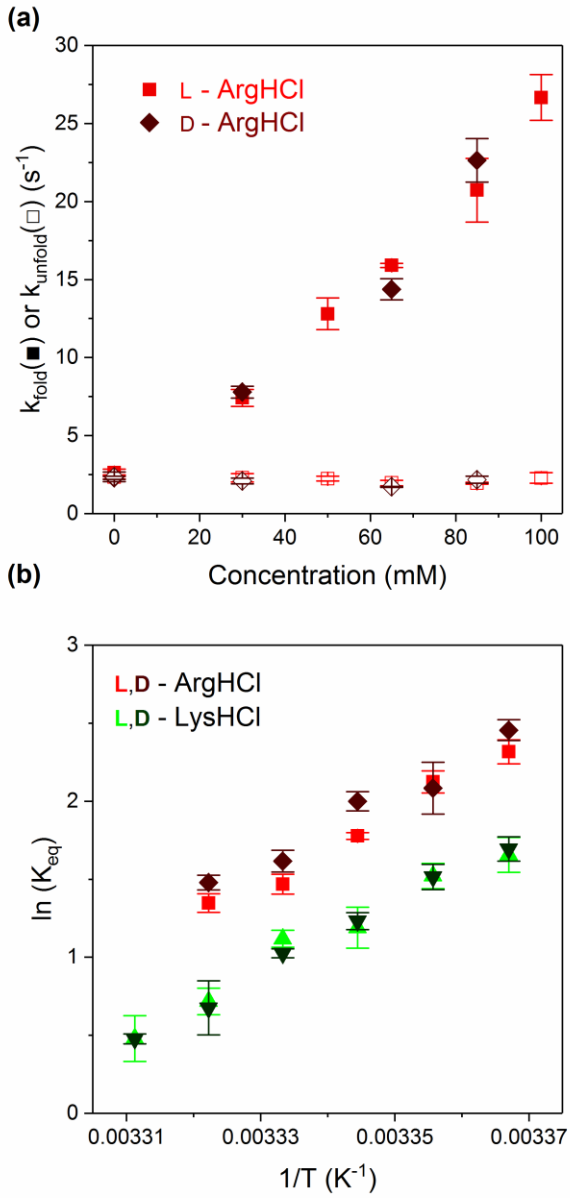

Figure 3.7 Hairpin folding is insensitive to amino acid chirality. **(a)** Concentration-dependence of folding (filled) and unfolding (unfilled) rate constants for enantiomers of ArginineHCl. **(b)** van't Hoff plots for ArginineHCl (red) and LysineHCl (green) enantiomers.

by a transition state which is enthalpically more accessible ($\Delta\Delta H^{\ddagger} < 0$) but entropically less approachable ($\Delta\Delta S^{\ddagger} < 0$). The kinetic and thermodynamic signature of arginine and lysine is shared by other monovalent cations, suggesting that the mechanism of secondary structure stabilization is predominantly electrostatic. A model is proposed to explain these observations,

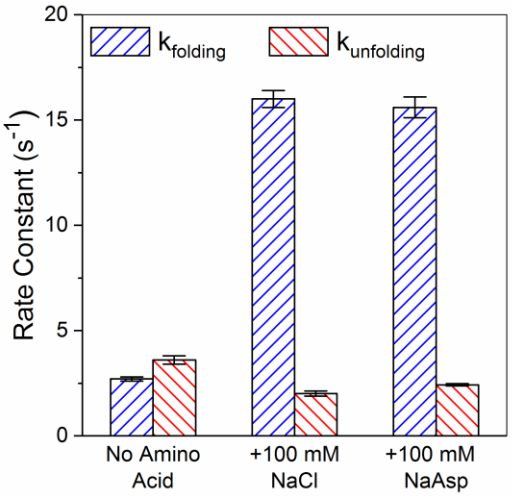

Figure 3.8 Negative amino acids behave according to their positively-charged counterion. Sodium chloride and sodium aspartate have indistinguishable effects on hairpin folding kinetics.

by examining how amino acids contribute to a nucleic acid's ionic atmosphere. In an evolutionary context, significant stabilization effects were observed only at relatively high amino acid concentrations (tens of millimolar), which disfavors secondary structure stabilization as a major driving force for amino acid incorporation into biology. However, secondary structure formation could prove to be far more selectively sensitive to small di- and tripeptides, which could serve as an interesting direction for further experiments.

3.7 References

1. Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. *Cell Res.* **2011**, *21*, 381-395.

2. Knobler, C. M.; Gelbart, W. M., Physical chemistry of DNA viruses. *Annu. Rev. Phys. Chem.* **2009**, *60*, 367-383.

3. Speir, J. A.; Johnson, J. E., Nucleic acid packaging in viruses. *Curr. Opin. Struct. Biol.* **2012**, *22*, 65-71.

4. Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C., Engineering and in vivo applications of riboswitches. *Annu. Rev. Biochem.* **2017**, *86*, 515-539.

5. Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R., A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **2004**, *306*, 275-279.

6. Butler, E. B.; Xiong, Y.; Wang, J.; Strobel, S. A., Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem. Biol.* **2011**, *18*, 293-298.

7. Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R., An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **2003**, *17*, 2688-2697.

8. Serganov, A.; Huang, L.; Patel, D. J., Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* **2008**, *455*, 1263-1267.

9. Yang, Y.; Kochoyan, M.; Burgstaller, P.; Westhof, E.; Famulok, M., Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **1996**, *272*, 1343-1347.

10. Yarus, M., A specific amino acid binding site composed of RNA. *Science* **1988**, *240*, 1751-1758.

11. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, *15*, 423-437.

12. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* **2002**, *418*, 214-221.

13. Herschlag, D., RNA Chaperones and the RNA folding problem. In *J. Biol. Chem.*, 1995; Vol. 270, pp 20871-20874.

14. Hoffman, M. M.; Khrapov, M. A.; Cox, J. C.; Yao, J.; Tong, L.; Ellington, A. D., Aant: The amino acid-nucleotide interaction database. *Nucleic Acids Res.* **2004**, *32*, D174-181.

15. Morozova, N.; Allers, J.; Myers, J.; Shamoo, Y., Protein-RNA interactions: Exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics* **2006**, *22*, 2746-2752.

16. Cheng, A. C.; Chen, W. W.; Fuhrmann, C. N.; Frankel, A. D., Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J. Mol. Biol.* **2003**, *327*, 781-796.

17. Hajnic, M.; Osorio, J. I.; Zagrovic, B., Computational analysis of amino acids and their sidechain analogs in crowded solutions of RNA nucleobases with implications for the mRNA-protein complementarity hypothesis. *Nucleic Acids Res.* **2014**, *42*, 12984-12994.

18. Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D., characterization of nucleobase-amino acid stacking interactions utilized by a DNA repair enzyme. *J. Phys. Chem. B* **2006**, *110*, 19652-19663.

19. de Ruiter, A.; Zagrovic, B., Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. *Nucleic Acids Res.* **2015**, *43*, 708-718.

20. Sengupta, A.; Sung, H. L.; Nesbitt, D. J., Amino acid specific effects on RNA tertiary interactions: Single-molecule kinetic and thermodynamic studies. *J. Phys. Chem. B* **2016**, *120*, 10615-10627.

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

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

23. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* **2008**, *5*, 507-516.

24. Preus, S.; Wilhelmsson, L. M., Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem* **2012**, *13*, 1990-2001.

25. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* **2016**, *67*, 441-465.

26. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on chips II*; Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

27. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **2008**, *94*, 1826-1835.

28. Corle, T. R.; Kino, G. S., *Confocal scanning optical microscopy and related imaging systems*; Academic Press: San Diego, 1996.

29. Fiore, J. L.; Holmstrom, E. D.; Nesbitt, D. J., Entropic origin of Mg2+-facilitated RNA folding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2902-2907.

30. Sinden, R. R.; Pearson, C. E.; Potaman, V. N.; Ussery, D. W., DNA: Structure and function. In *Advances in genome biology*; Verma, R. S., Ed. JAI: 1998; Vol. 5, pp 1-141.

31. Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F., Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15431-15436.

32. Bevilacqua, P. C.; Blose, J. M., Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79-103.

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

34. Holmstrom, E. D.; Fiore, J. L.; Nesbitt, D. J., Thermodynamic origins of monovalent facilitated RNA folding. *Biochemistry* **2012**, *51*, 3732-3743.

35. Rentzeperis, D.; Shikiya, R.; Maiti, S.; Ho, J.; Marky, L. A., Folding of intramolecular DNA hairpin loops: Enthalpy−entropy compensations and hydration contributions. *J. Phys. Chem. B* **2002**, *106*, 9945-9950.

36. Rentzeperis, D.; Alessi, K.; Marky, L. A., Thermodynamics of DNA hairpins: Contribution of loop size to hairpin stability and ethidium binding. *Nucleic Acids Res.* **1993**, *21*, 2683-2689.

37. Auffinger, P.; D'Ascenzo, L.; Ennifar, E., Sodium and potassium interactions with nucleic acids. In *The Alkali Metal Ions: Their Role for Life*; Sigel, A.; Sigel, H.; Sigel, R. K. O., Eds. Springer International Publishing: Cham, 2016; pp 167-201.

38. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813-841.

39. Trachman, R. J., 3rd; Draper, D. E., Divalent ion competition reveals reorganization of an RNA ion atmosphere upon folding. *Nucleic Acids Res.* **2017**, *45*, 4733-4742.

40. Gebala, M.; Giambasu, G. M.; Lipfert, J.; Bisaria, N.; Bonilla, S.; Li, G.; York, D. M.; Herschlag, D., Cation-anion interactions within the nucleic acid ion atmosphere revealed by ion counting. *J. Am. Chem. Soc.* **2015**, *137*, 14705-14715.

41. Vieregg, J.; Cheng, W.; Bustamante, C.; Tinoco, I., Jr., Measurement of the effect of monovalent cations on RNA hairpin stability. *J. Am. Chem. Soc.* **2007**, *129*, 14966-14973.

42. Kuznetsov, S. V.; Ansari, A., A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* **2012**, *102*, 101-111.

43. Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T., Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **2004**, *86*, 2530-2537.

44. 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.

45. Bosco, A.; Camunas-Soler, J.; Ritort, F., Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.* **2014**, *42*, 2064-2074.

46. Privalov, P. L.; Ptitsyn, O. B.; Birshtein, T. M., Determination of stability of DNA double helix in an aqueous medium. *Biopolymers* **1969**, *8*, 559-571.

47. 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.

48. Tan, Z. J.; Chen, S. J., Salt dependence of nucleic acid hairpin stability. *Biophys. J.* **2008**, *95*, 738-752.

49. Huang, P. C., Effect of guanidinium ion on DNA denaturation and renaturation. *Biochem. Biophys. Res. Commun.* **1968**, *33*, 384-390.

50. Blanco, F.; Kelly, B.; Sanchez-Sanz, G.; Trujillo, C.; Alkorta, I.; Elguero, J.; Rozas, I., Non-covalent interactions: Complexes of guanidinium with DNA and RNA nucleobases. *J. Phys. Chem. B* **2013**, *117*, 11608-11616.

51. Lambert, D.; Leipply, D.; Shiman, R.; Draper, D. E., The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.* **2009**, *390*, 791-804.

Chapter 4

Chirality-Dependent Amino Acid Modulation of RNA Folding*

4.1 Abstract

The preponderance of a specific D-or L- chirality in fats, sugars, amino acids, nucleic acids, etc. is ubiquitous in nature, yet the biological origin of such chiral dominance (i.e., with one enantiomer overwhelmingly present) remains an open question. One plausible proposal for the predominance of L-chirality isn amino acids could be through evolutionary templating of chiral RNA-folding via chaperone activity. In order to help evaluate this possibility, single molecule fluorescence experiments have been performed that measure the chiral dependence of chaperone folding dynamics for the simple tetraloop-tetraloop receptor (TL-TLR) tertiary binding motif in the presence of a series of chiral amino acids. Specifically, D- vs. L- arginine are found to accelerate the unfolding of this RNA motif in a chirally selective fashion, with temperature-dependent studies of the kinetics performed to extract free energy, enthalpy, and entropy landscapes for the underlying thermodynamics. Furthermore, all-atom molecular dynamics (MD) simulations are pursued to provide additional physical insight into this chiral sensitivity, which reveal enantiomer-specific sampling of nucleic acid surfaces by D- vs. L-arginine and support a putative mechanism for chirally-specific denaturation of RNA tertiary structure by arginine but not other amino acids.

* This chapter is adapted from: Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J. *J. Phys. Chem. B*. **2020**, *124* (51), 11561-11572.

4.2 Introduction

From the overwhelming abundance of homochiral species in nature, it appears that biology abhors racemic mixtures. The tendency toward homochirality (or chiral purity) is a long appreciated hallmark of biological systems,[1] characterizing the biosynthesis of proteins, nucleic acids, saccharides, and small molecules.[2-3] Consider the case of homochirality in proteins, for which all amino acids more complex than glycine have chiral centers, yet biology almost exclusively favors L-amino acids.[4-5] This remarkable evolution toward homochirality in proteins, for example, demands to be accounted for in any complete model of the origin of life, for which a number of mechanisms have been proposed. One family of theories suggests that homochirality initially developed in nucleic acids during the RNA world[6] and was subsequently transmitted to amino acids. For instance, protein synthesis in a D-ribose RNA world could have an inherent bias toward L-amino acids, as demonstrated in the chiral discrimination of amino acids during uncatalyzed aminoacylation of tRNA, a key step in protein synthesis.[7-8]

An alternative possibility explored herein is that evolutionary selection of amino acid chirality occurred *before* the development of protein synthesis, based instead on the role of amino acids in the RNA world, possibly as nucleic acid-folding chaperones.[9-10] RNA molecules are notoriously inefficient at folding into biologically competent form[11], with slow folding kinetics and therefore a tendency to adopt long-lived misfolded states (with lifetimes of order 1–1000 s)[10]. Modern biology in the current proteomic era addresses such issues through the use of proteins[12] as well as osmolytes[13] as chaperones to guide the pathway toward correct RNA folding. It is thus entirely reasonable to expect that early life would explore similar strategies, utilizing amino acids as small molecule chaperones. Indeed, experiments have recently demonstrated the ability of amino acids to influence RNA folding at secondary[14] and tertiary[15]

structure levels (and at even higher ribozyme[16] levels of complexity) at 10–100 mM concentrations (1mM = 1mmol/L). Although the corresponding affinities are relatively low, such concentrations are well within the limits of osmolyte cellular levels, which can be in excess of 1 M.[17-18] If early life relied on amino acids to modulate the thermodynamic stability and/or kinetics of RNA folding, and if the capacity of amino acids to chaperone RNA folding depends on amino acid chirality, then the resulting evolutionary pressure could have promoted an amino acid synthesis mechanism that favors the now ubiquitous L-amino acid.

This is a bold statement, requiring experimental evidence to evaluate as even a putative evolutionary pathway. Many studies on the chiral dependence of amino acid-nucleic acid interactions have focused on aminoacylation of tRNA structure es, with a sizable body of work established.[7-8, 19-20] Amino acid chirality effects have also been examined in other contexts, such as codon binding,[21-23] histidine selection,[24] chromatographic applications,[25-26] and ribozyme activity.[27] Indeed, the study of Yarus and Majerfeld represents an especially intriguing and relevant case in point, in which they identified an arginine-specific binding pocket in the *Tetrahymena* ribozyme that exhibited a tenfold higher selectivity for L-arginine over D-arginine, with clear catalytic implications. These ensemble studies provided substantial first insights into how nucleic acid structures might sense amino acid chirality. However, few experiments have been performed at the single molecule level, which can in principle provide even deeper levels of physical understanding into a mechanism for chirally sensitive RNA folding.[28]

In a previous work, we reported[14] on single-molecule experiments investigating the effects of amino acids on nucleic acid *secondary* structure folding kinetics, for which no dependence on amino acid chirality was observed. We have also examined[15] how the presence of amino acids can influence the formation of RNA *tertiary* structure, for which significant

chaperone effects were indeed observed but without further exploration into any dependence on amino acid chirality. To test for such chiral chaperone effects, an additional stratum of work is necessary, specifically a single molecule study into chiral amino acid influence on chiral RNA tertiary folding, which represents the focus of the present paper.

Here, we utilize single molecule methods in a search for amino acid chiral dependency on RNA tertiary folding dynamics. Specifically, we explore the ubiquitous 11-nt GAAA tetraloop-tetraloop receptor (TL-TLR) motif[29] as a model tertiary folding system, exploiting single molecule fluorescence resonance energy transfer (smFRET) to probe conformational changes in the RNA construct. Rate constants for TL-TLR folding (k_{fold}) and unfolding (k_{unfold}) are systematically investigated for a series of enantiomeric pairs over a wide range of amino acid concentrations. Interestingly, we observe a strong chiral dependence for D-, L-arginine inhibition of RNA TL-TLR tertiary folding and yet no evidence of a similar chirally sensitive signature for any other chiral species selected from 5 representative classes of soluble amino acids (lysine, histidine, alanine, serine, and proline). In addition, single molecule folding studies of these amino acid-RNA constructs are explored under temperature-controlled conditions, in order to extract both chirally sensitive and chirally insensitive enthalpic and entropic contributions to the folding free energy landscape. Finally, these single molecule experiments are complemented by all-atom molecular dynamics (MD) simulations, in efforts to provide additional insight and microscopic perspective into the chiral dependence of the amino acid assisted folding and unfolding kinetics.

4.3 Methods

4.3.1 RNA FRET construct

To examine the influence of amino acid chirality in modulating RNA tertiary folding, we employ an RNA construct bearing an isolated tetraloop-tetraloop receptor (TL-TLR) folding motif (figure 1), which has been successfully used in previous studies to examine various kinetic and thermodynamic aspects of TL-TLR folding at the single molecule level.[15, 29-30] The construct consists of three strands: (1) a biotinylated DNA oligonucleotide for surface attachment (5'-biotin-CGACTCGTCTCGAG-3'), (2) an RNA sequence labeled with Cyanine 5 (Cy5) at the 5' position (5'-Cy5-GCCGAUAUGGACGACACGCCCUCGAGACGAGUCG-3'), and (3) an RNA strand containing a GAAA tetraloop, an internal hexameric polyethylene glycol (PEG) linkage, and a Cyanine 3 (Cy3) fluorophore at the 5' terminus (5'-Cy3-GGCGAAGCC-PEG$_6$-CGUGUCGUCCUAAGUCGGC-3'). All sequences are commercially synthesized and HPLC-

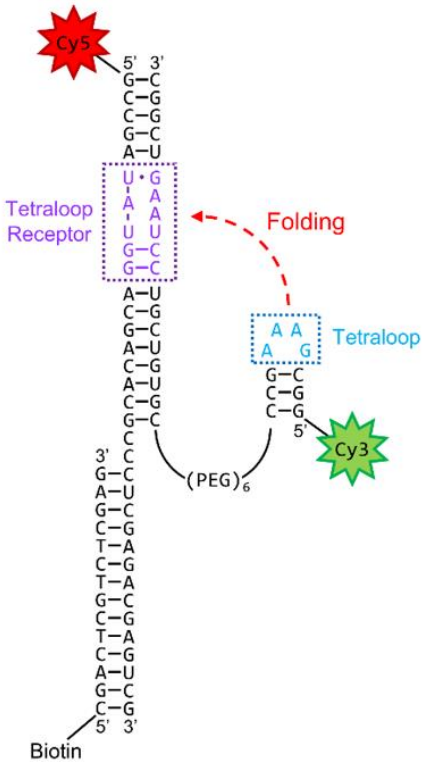

Figure 4.1 Structure of the 3-stranded TL-TLR smFRET construct.

purified (Integrated DNA Technologies), where companies/sources identified herein do not indicate product support. Strands (1) and (2) hybridize to form a stable fourteen base-pair RNA-DNA duplex, while strands (2) and (3) bind to generate the tetraloop receptor (TLR) domain flanked by double helices to help maintain structural stability (see figure 1). The flexible PEG linker in strand (3) largely avoids potential persistence length issues due to base-stacking in an ssRNA linker, which could in turn hinder diffusion of the GAAA tetraloop.[31] The strands are annealed by heating a 10 μmol/L mixture to 85 °C, slowly cooling to room temperature at 1 °C per minute, prior to long term storage at -20 °C. The annealed stock is used without purification, since constructs lacking the biotinylated strand do not attach to the surface, and constructs lacking the Cy3-labeled strand are not excited by the laser ($\sigma_{Cy5}/\sigma_{Cy3}$ = 3% at 532 nm).

4.3.2 Sample preparation

Samples are prepared in a home configured glass flow cell, for which a #1.5 glass coverslip (CGI Life Sciences) is soaked overnight in acetone and cleaned in a UV-ozone oven (Jelight Mo. 42) for 30 minutes. The clean coverslip is attached to a glass slide (VWR) by two parallel strips of double-sided tape, forming a channel approximately 2 mm wide, 22 mm long, and 0.1 mm deep to act as a flow cell for rapid solution exchange. The coverslip surface is labeled with the RNA FRET construct using biotin-streptavidin interactions, as previously described.[15, 32-33] Three solutions, all buffered with 50 mM HEPES at pH 7.6, are sequentially incubated in the sample chamber for 10 minutes in the following order: i) 10 mg/mL bovine serum albumin (BSA) with 1 mg/mL biotinylated-BSA to block and biotinylate the glass surface, ii) 0.2 mg/mL streptavidin to bind the biotin with streptavidin tetramers, and iii) 1 nM biotinylated TL-TLR construct to bind to the streptavidin, resulting in a surface RNA coverage of ~0.1/μm^2 . For each experiment, the sample chamber is washed with buffer prior to filling

with an imaging solution consisting of a buffer (50 mM HEPES, pH 7.6), an oxygen scavenging cocktail[34] (100 nM protocatechuate 3,4-dioxygenase, 5 mM protocatechuic acid in equimolar sodium hydroxide, and 2 mM Trolox), a metal chelating agent (0.1 mM EDTA), monovalent salt (50 mM KCl; total [M^+] = 80 mM), and the desired D/L-amino acid (see results: note that arginine and lysine are added as chloride salts to maintain near neutral pH). HEPES has been purchased from MP Biomedicals, EDTA and L-histidine from Fluka, D-histidine and L-serine from Alfa Aesar, D-alanine and D-proline from Fischer, and with all other compounds obtained from Sigma-Aldrich.

4.3.3 Single-molecule instrumentation

Single-molecule measurements are carried out by total internal reflection fluorescence (TIRF) microscopy (schematically outlined in figure 2), with a similar design to other through-objective TIRF microscopes.[35] Light from a 532 nm diode-pumped solid-state laser (MeshTel GSF32-300PS) is directed into an inverted microscope (Zeiss Axiovert 135) where it is focused onto the back focal plane of a 1.4 NA, oil-immersion microscope objective (Olympus PLAPON60XOSC2; Olympus Type-F immersion oil). To achieve total internal reflection, moving a mirror translates the beam laterally away from the optical axis, thereby increasing angle of incidence at the sample until the critical angle is reached. Fluorescence from the sample is collected by the same objective before passing through a 550 nm LP dichroic mirror (Chroma) to filter out reflected and scattered excitation light, with the transmitted photons collected on a charge-coupled device (CCD) video camera. First, Cy3 and Cy5 fluorescence is separated into two channels by 645 nm LP dichroic mirror (Chroma DRLP), which are recombined by a second dichroic mirror to laterally offset the two images and project them side-by-side onto the CCD. Second, the images are magnified by a 4x telescope, with Cy3 and Cy5 emission focused

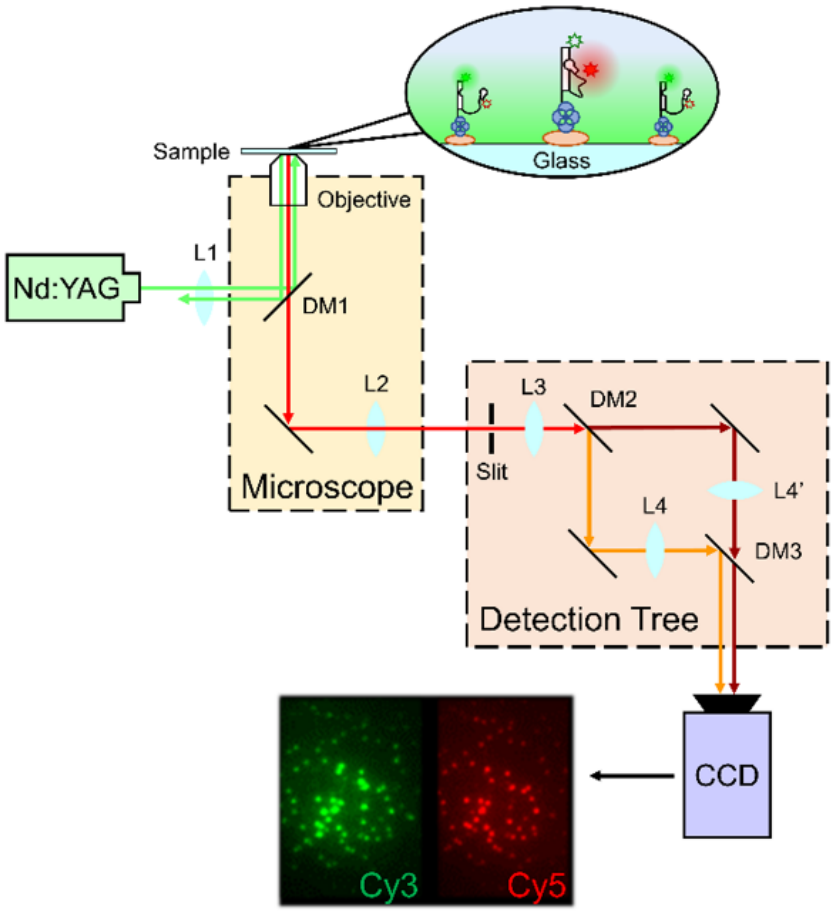

Figure 4.2 (a) TIRF instrument schematic. 532 nm light is focused by lens L1 onto the back focal plane of the microscope objective. Emission collected by the objective passes through dichroic mirror DM1 to remove excitation light and is spatially filtered by a 1.5 mm slit placed at the focus of the tube lens L2. Cy3 and Cy5 emission are separated and recombined as offset images by dichroic mirrors DM2 and DM3. In addition, the images are magnified by 4x telescope (L3: L4/L4') before detection by CCD camera.

independently using separate lenses (L4 and L4') to reduce chromatic aberration. Finally, the emission is transmitted through a ~1.5 mm slit placed at the focus of the microscope's tube lens, truncating the circular image to maximize use of the CCD's square imaging area. Light is detected and recorded as greyscale movies by an intensified CCD camera (Princeton Instruments I-PentaMAX 512-EFT) with on-chip 3x3 pixel binning to increase the acquisition framerate to 40 Hz.

4.3.4 Data analysis

Single-molecule movies are analyzed using custom analysis software written in LabWindows/CVI (National Instruments). The first ~100 frames of each movie are averaged to smooth over conformational fluctuations, and local thresholding is applied to the averaged image to locate particles, whose (x,y) centroid positions are then refined by 2D Gaussian fitting. Summation over a 4 pixel diameter circle yields single particle fluorescence traces as a function of time, which are converted to actual photon count rates with calibration factors obtained from the variance-mean ratio method.[36] Single molecule trajectories are sorted into donors and acceptors by lateral position on the CCD camera, with donor-acceptor pairs identified by calibrated affine mapping.[37] Subsequent FRET data analysis follows protocols we have used in previous studies:[38] donor and acceptor fluorescence traces are combined into single-molecule FRET (smFRET) trajectories, from which state-to-state transition times are extracted by thresholding. Cumulative dwell time distributions are then subjected to single exponential least squares fit analysis, thereby yielding rate constants for both folding and unfolding of the single molecule constructs.

4.3.5 Temperature control

Temperature control is established using dual thermoelectric coolers (TEC), depicted in Figure 4.3, in which two TECs (upper and lower) are used as a distributed source of cooling to minimize thermal gradients across the sample. TEC modules (TE Technology HP-127-1.0-0.8) are compression-mounted between aluminum plates and coupled to the sample via thermal pads (EN-Labs). The upper TEC assembly lies directly on top of the sample slide, while the lower one embraces the microscope objective, which is in thermal contact with the sample through the immersion oil. Precise temperature stabilization (±0.05 °C, τ <1 min response time, Figure 4.3c)

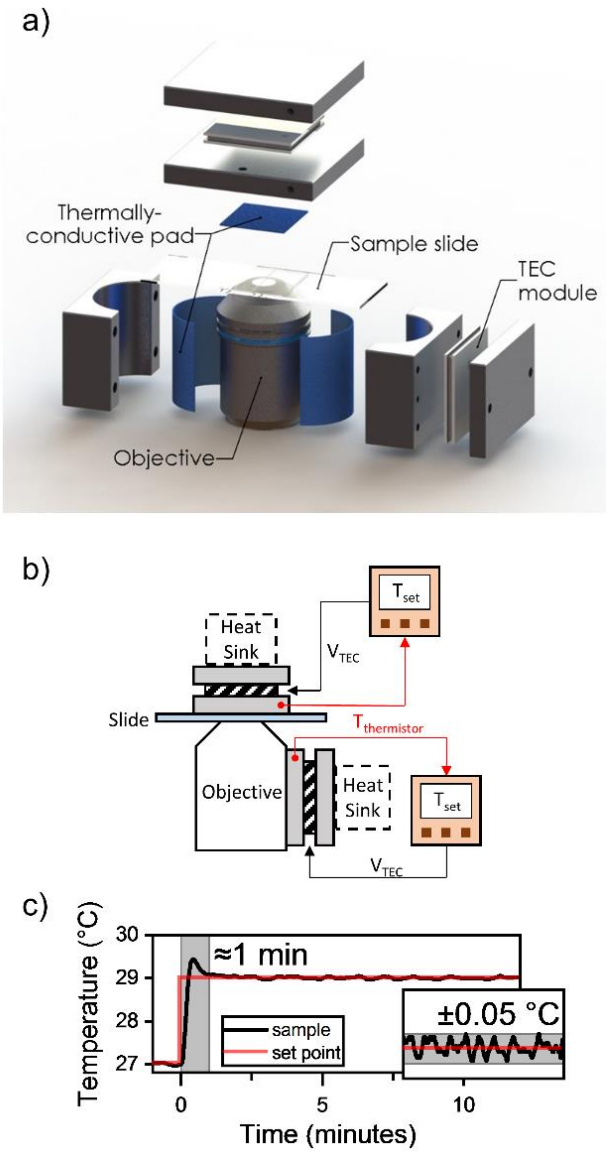

Figure 4.3 Temperature control apparatus. (a) To minimize thermal gradients, the sample is contacted by two TEC assemblies, one directly on the glass slide, and the other mediated by the microscope objective. (b) Schematic of TEC control system. (c) Sample response time and temperature stability (inset) of TEC system.

is achieved via digital servo loop control. Specifically, sample and objective temperatures are measured by thermistors (Vishay) which feed into bidirectional temperature controllers (TE Technology TC-720) that drive the upper and lower TEC modules, respectively. The upper operational temperature of the system (~40 °C) is constrained by softening of optical cement in

the microscope objective at ~45 °C, while the lowest temperature is limited by air cooling efficiency to the heat sink. The coldest temperature data reported herein (~10 °C) required supplemental cooling of the TEC heat sink by contact with an enclosed frozen liquid reservoir (e.g., ice).

4.3.6 Molecular dynamics simulations

To complement these smFRET experimental results, we have in parallel pursued all-atom molecular dynamics (MD) simulations with explicit solvent. Initial structures for the nucleic acid constructs are used directly from the protein database (PDB), with the exception of the docked TL-TLR (PDB 1GID), which is truncated to residues belonging to the tetraloop (TL) and tetraloop receptor (TLR) residue indices (148–155, 220–229, and 245–253) to reduce system size. In the AmberTools19 software suite,[39] the three structures of interest (TL, TLR, TL-TLR) are each solvated in a 15 Å padded TIP3P water box,[40] for which the total charge is neutralized by K^+ ions and with an additional 150 mM KCl to mimic the experimental ionic strength. Amino acids are added as chloride salts using the Visual Molecular Dynamics (VMD)[41] software package, with initial positions selected at random. Dynamics are based on Amber force fields: χ_{OL3} for RNA,[42] Cheatham and Young's parameters for the ions,[43] the TIP3P model for water, and Horn's zwitterionic amino acid parameterization for D- and L- amino acids.[44] Simulations have been performed on the Nanoscale Molecular Dynamics (NAMD)[45] platform using periodic boundary conditions and a 2 fs timestep, made possible by the constraining NH, CH, and OH bonds as rigid. Constant temperature and pressure conditions are maintained by Langevin dynamics at 300 K (1 ps^{-1} damping rate), with a Berendsen barostat at 1 atm (100 fs relaxation time, 4.57 x 10^{-5} bar^{-1} compressibility) and simulation snapshots stored every 10 ps. All other

parameters are initialized at NAMD default values, with VMD software used for simulation analysis and visualization.[46]

4.4 Experimental results

4.4.1 RNA tertiary folding demonstrates a strong chiral sensitivity to arginine

Single molecule experiments have been performed to evaluate the ability of amino acids to promote or inhibit RNA tertiary folding in a chirally sensitive fashion. We select an RNA construct bearing a tetraloop-tetraloop receptor (TL-TLR) motif and a FRET fluorophore pair (Figure 4.1) as a representative tertiary folding system.[29-30] Surface-tethered TL-TLR molecules are observed in a TIRF microscope to monitor and extract the single molecule folding dynamics, with sample smFRET trajectories shown in Figure 4.4a. The trajectories demonstrate clear two-state behavior, with two distinct FRET efficiencies: $E_{low} \approx 0.4$ and $E_{high} \approx 0.7$ representing the TL-TLR unfolded and folded states, respectively.[30] Under control conditions (50 mM HEPES buffer, 80 mM M^+, pH 7.6), the TL-TLR construct favors the folded state ($K_{eq} = 2.9 \pm 0.2$, where the uncertainty represents the standard deviation of the mean (σ_m)). Addition of 300 mM L-arginine (as L-arginine chloride) strongly *destabilizes* the folding equilibrium by nearly twofold ($K_{eq} = 1.51 \pm 0.08$), in agreement with previous observations.[15] More relevantly to the current study, 300 mM D-arginine exerts a 50% greater destabilization effect than L-arginine ($K_{eq} = 0.99 \pm 0.04$). Clearly, folding in the TL-TLR tertiary binding motif is capable of kinetically sensing chirality in amino acids, as well as vice versa.

The robustness of such chiral "anti-chaperone" behavior is next probed by examining the unimolecular folding and unfolding rate constants (k_{fold}, k_{unfold}) over a wide range of arginine concentrations. TL-TLR folding and unfolding rate constants are determined by dwell time

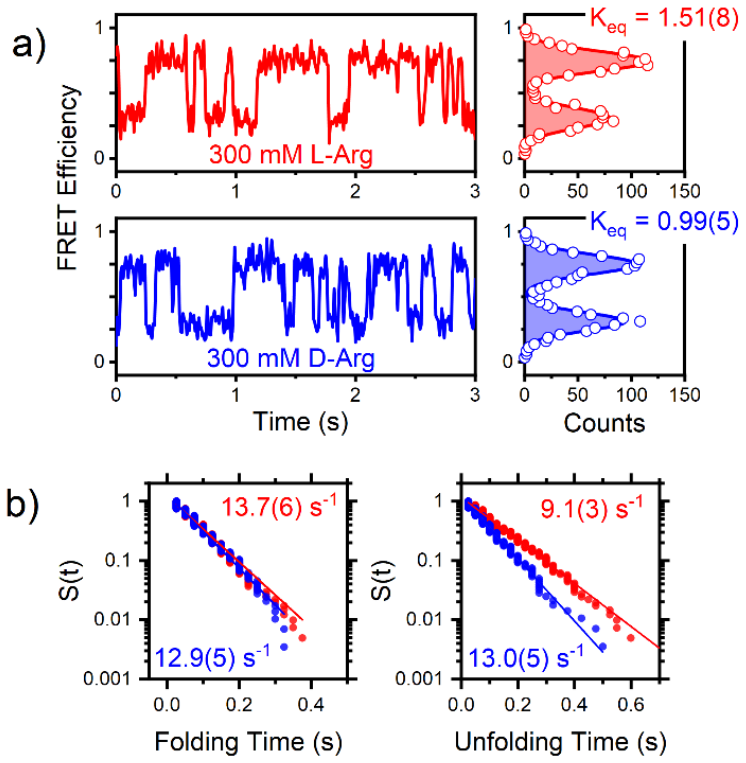

Figure 4.4 (a) Sample single-molecule TIRF trajectories for TL-TLR folding with 300 mM L-arginine (red, top row) and 300 mM D-arginine (blue, bottom row). FRET histograms with double Gaussian fits are shown to the right. (b) Survival functions for TL-TLR dwell times. Lines are single exponential fits, from which rate constants can be extracted. Data are obtained in a background of 50 mM K^+ and 30 mM Na^+.

analysis (Figure 4.4b)[38] and plotted as a function of [arginine] (Figure 4.5). In general, the folding rate constants follow qualitatively similar trends for both arginine chiral enantiomers. Specifically, both the D- and L-*folding* rate constants (k_{fold}) decrease modestly (-30%) and then increase (+30%) with respect to [arginine]. By way of contrast, the D- and L- *unfolding* rate constants (k_{unfold}) monotonically increase by nearly 200-300%, which thereby corresponds to a net *destabilization* of the TL-TLR binding motif with increasing [arginine]. Of interest here is a direct comparison between the two chiral enantiomers, for which the *folding* rate constant reveals remarkably little sensitivity to the substrate chirality. Instead, the chiral differences between

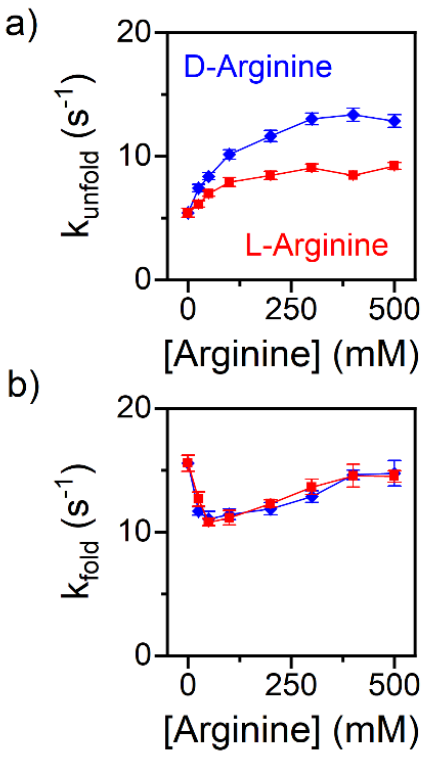

Figure 4.5 Rate constants for TL-TLR folding (a) and unfolding (b) in the presence of L- and D-arginine. Experiments are carried out in buffer (pH 7.6) with 80 mM background [M^+].

enantiomers manifest primarily in the *unfolding* rate constant, for which D-arginine is significantly more effective than L-arginine at enhancing the rate.

4.4.2 Thermodynamic studies of chirally sensitive D-,L-arginine assisted unfolding

To probe the thermodynamics of RNA unfolding by the enantiomers of arginine, we turn now to temperature-dependent folding studies. Changes in temperature modify the free energy of folding by the Gibbs expression

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -RT\ln(K_{eq}), \qquad \text{(Eq. 4.1)}$$

where ΔG° is the folding free energy change, ΔH° is the enthalpy of folding, T is the absolute temperature, ΔS° is the entropy of folding, R is the gas constant, and K_{eq} is the folding equilibrium constant. Rearranged into the standard van't Hoff form,

$$\ln(K_{eq}) = -\frac{\Delta H^{\circ}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\circ}}{R}, \quad \text{(Eq. 4.2)}$$

a plot of $\ln(K_{eq})$ vs. 1/T can be used to recover the enthalpy and entropy of folding via linear least squares analysis. Similarly, an Arrhenius-type expression can be derived from Kramers' theory for the temperature dependence of rate constants,[47-48]

$$\ln(k_{\text{fold/unfold}}) = -\frac{\Delta H^{\ddagger}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\ddagger}}{R} + \ln(\kappa \cdot \nu), \quad \text{(Eq. 4.3)}$$

where k is the rate constant, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are the *activation* enthalpy and entropy, respectively, ν is the attempt frequency along the reaction coordinate, and κ is the transmission coefficient accounting for deviation from transition state (TS) theory[49] due to diffusive barrier recrossing. The prefactor product $\kappa \cdot \nu$ depends on the exact details of the folding energy landscape, but an estimate of $\kappa \cdot \nu \sim 10^6$ s^{-1} is plausible from previous studies of RNA and protein folding.[48, 50] Furthermore, the $\Delta S^{\ddagger}$ thermodynamic entropy change parameters obtained from a TS analysis depend only logarithmically on the choice of $\kappa \cdot \nu$, while the changes in these TS entropies $\Delta\Delta S^{\ddagger}$ are rigorously independent of the choice of $\kappa \cdot \nu$. Nevertheless, caution is warranted when interpreting the absolute intercept values in these Arrhenius plots, as they represent a combination of activation entropy and prefactor contributions.

The results from temperature-dependent experiments (11–27 °C) at a fixed concentration of 100 mM L- and D-arginine are displayed in Figure 4.6. Consistent with concentration studies, D-arginine is found to decrease TL-TLR stability more than L-arginine (Figure 4.6a), and the kinetic basis of this chiral-specific destabilization is primarily due to differences in unfolding rather than folding (Figure 4.6b). Linear regression is applied, and extracted enthalpies and entropies are reported in Table 4.1. Visual inspection of folding rate constant data (Figure 4.6b, left panel; see zoomed insert) reveals small but nevertheless clear differences in slopes between enantiomers, indicating that the enthalpic cost to reach the transition state during folding is

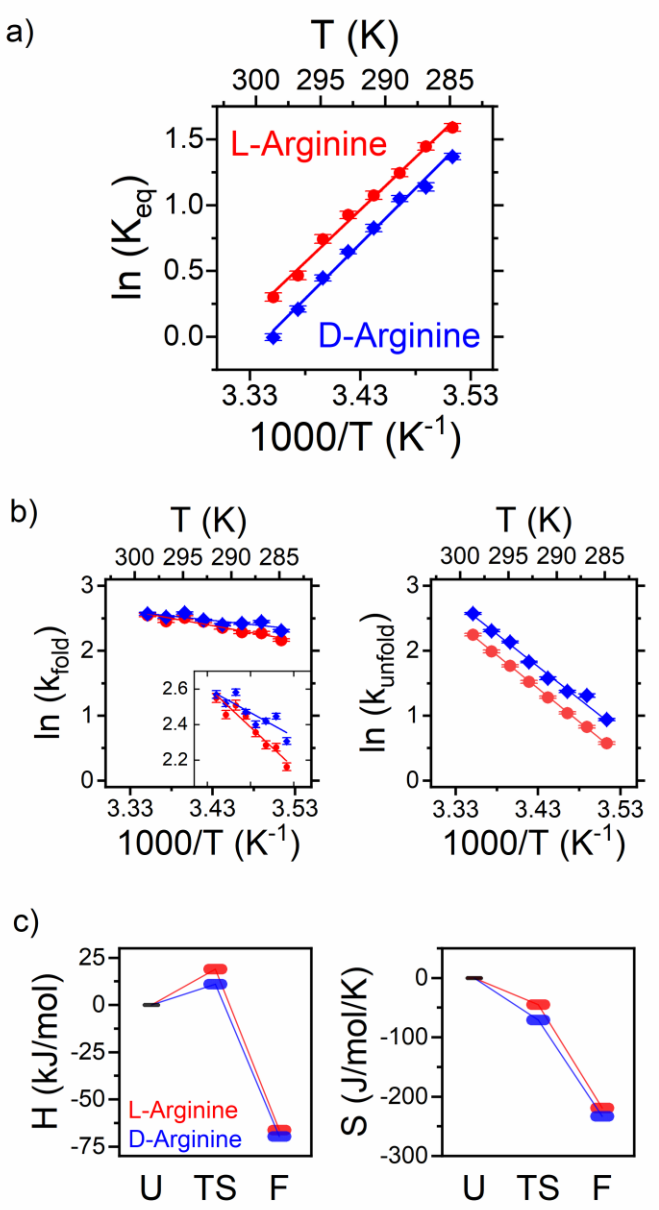

Figure 4.6 Temperature dependent studies of folding and unfolding thermodynamics. (a) van't Hoff plot and (b) Arrhenius plots for TL-TLR folding with 100 mM L- and D-arginine. (c) Fitted values for enthalpies (left) and entropies (right) of the folded state (F) and the transition state (TS) relative to the unfolded state (U).

greater for L-arginine ($\Delta H^{\ddagger}_{fold}$ 18.9 ± 2.2 kJ/mol for L-arginine and 11.1 ± 2.8 kJ/mol for D-arginine). This is intriguing, as no chiral behavior in the folding rate constant was observed in

	$\Delta H°$ (kJ/mol)	$\Delta H^{\ddagger}_{fold}$ (kJ/mol)	$\Delta H^{\ddagger}_{unfold}$ (kJ/mol)
L-arginine	-66.2(2.2)	18.9(2.2)	85(1)
D-arginine	-69.7(2.6)	11.1(2.8)	83(3)
	$\Delta\Delta H°$ (kJ/mol)	$\Delta\Delta H^{\ddagger}_{fold}$ (kJ/mol)	$\Delta\Delta H^{\ddagger}_{unfold}$ (kJ/mol)
Difference	3(3)	8(4)	2(3)
	$\Delta S°$ (J/mol K)	$\Delta S^{\ddagger}_{fold}$ (J/mol K)	$\Delta S^{\ddagger}_{unfold}$ (J/mol K)
L-arginine	-219(8)	-45(8)	174(3)
D-arginine	-233(9)	-71(9)	169(10)
	$\Delta\Delta S°$ (J/mol K)	$\Delta\Delta S^{\ddagger}_{fold}$ (J/mol K)	$\Delta\Delta S^{\ddagger}_{unfold}$ (J/mol K)
Difference	14(12)	26(12)	5(10)

Table 4.1. Thermodynamic values from van't Hoff and Arrhenius analyses for TL-TLR folding in 100 mmol/L L- vs. D-arginine.

room temperature experiments (Figure 4.5). The explanation for this is that the difference in activation enthalpy is counteracted by a difference in activation entropy ($\Delta\Delta S^{\ddagger}_{fold}$ = -26 ± 12 J/molK), which results in a negligibly small net free energy difference at room temperature ($\Delta\Delta G^{\ddagger}_{fold} = \Delta\Delta H^{\ddagger}_{fold} - T\Delta\Delta S^{\ddagger}_{fold}$ = 0.1 ± 5 kJ/mol). Similarly, the unfolding rate constant (Figure 4.6b, right panel) appears to exhibit indistinguishable slopes but a clear vertical offset, corresponding to more purely entropic rather than enthalpic contributions to free energies differences between the two enantiomers. However, the data are taken only over a relatively small (< 10%) absolute temperature range, with extrapolation resulting in significant correlation between enthalpic (slope) and entropic (intercept) parameters. After careful inclusion of correlated error propagation, the fitted unfolding activation entropies prove to be statistically indistinguishable (174 ± 3 J/mol K for L-arginine and 169 ± 10 J/mol K for D-arginine). Similarly, the van't Hoff thermodynamic parameters (ΔH^0, ΔS^0) are also found to be identical

within experimental error for both 100 mM L- and D-arginine. In order to eliminate the possibility that such agreement is accidental for the chosen arginine concentration, temperature-dependent experiments were also performed at 200 mM arginine, but again exhibit no chiral sensitivity (outside of 1σ uncertainties) in the least squares fits (Figure 4.7). In summary, temperature-dependent measurements reveal a clear difference in *folding* activation enthalpies in the presence of L- vs. D-arginine, but the thermodynamic basis of chiral differences in *unfolding* activation remains smaller than our experimental resolution.

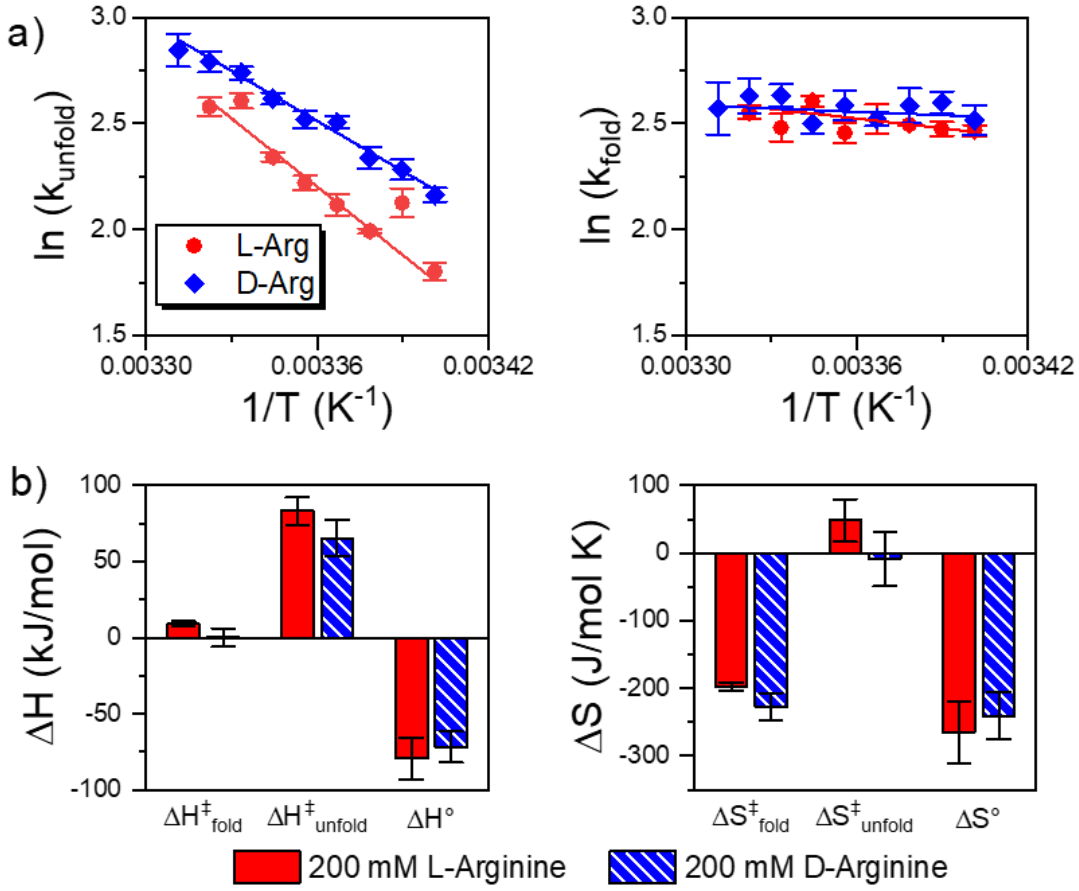

Figure 4.7 Temperature-dependent experiments for 200 mM L- and D-arginine. (a) Arrhenius plots for unfolding (left) and folding (right) rate constants. (b) Enthalpies (ΔH) and entropies (ΔS) parameters from linear fits. Both enantiomers of arginine yield the same fit results, within error.

4.4.3 In search of other amino acid chiral sensitivities in RNA folding/unfolding

To provide further insight into the structural origin of such chiral sensitivity, we have extended our kinetic analysis to additional amino acids (Figure 4.8). By determining which other amino acids, if any, interact with the TL-TLR construct with chiral specificity, we might glean key molecular characteristics that induce enantiomer-dependent amino acid chaperoning of RNA tertiary folding, namely whether amino acid side chains are hydrophilic or hydrophobic, charged or uncharged, saturated or unsaturated, etc. Previous work determined that the TL-TLR fold necessitated amino acid concentrations of order 100 mM to observe significant changes in the folding equilibrium,[15] which restricts the present slate of amino acids based on solubility in water.[51] Nevertheless, the set of five soluble amino acids explored herein (lysine, histidine, alanine, serine, and proline) bear side chains with a diverse range of physicochemical properties which covers much of "amino acid space." By exposing the RNA tertiary fold to this representative suite of five amino acids and their enantiomers, we hope to uncover general structural qualities which give rise to the chiral behavior observed above for arginine.

As a first candidate we consider lysine, which, like arginine, also has a long side chain with a net positive charge, but with an ammonium moiety substituting for the guanidinium group. Despite these physical similarities, little to no chiral dependence between the folding and unfolding rate constants is observed for lysine (see Figure 4.8, top row). We next examine histidine, whose side chain contains a conjugated hydrocarbon network similar to arginine, but which is predominantly unprotonated (~2% cationic) under our experimental conditions (pH = 7.6). Again, despite these qualitative similarities, the results reveal no distinction between L- and D-histidine behavior. We additionally investigated amino acids with other side chain types: alanine, with a small and hydrophobic side chain; serine, with a polar hydroxymethyl side chain,

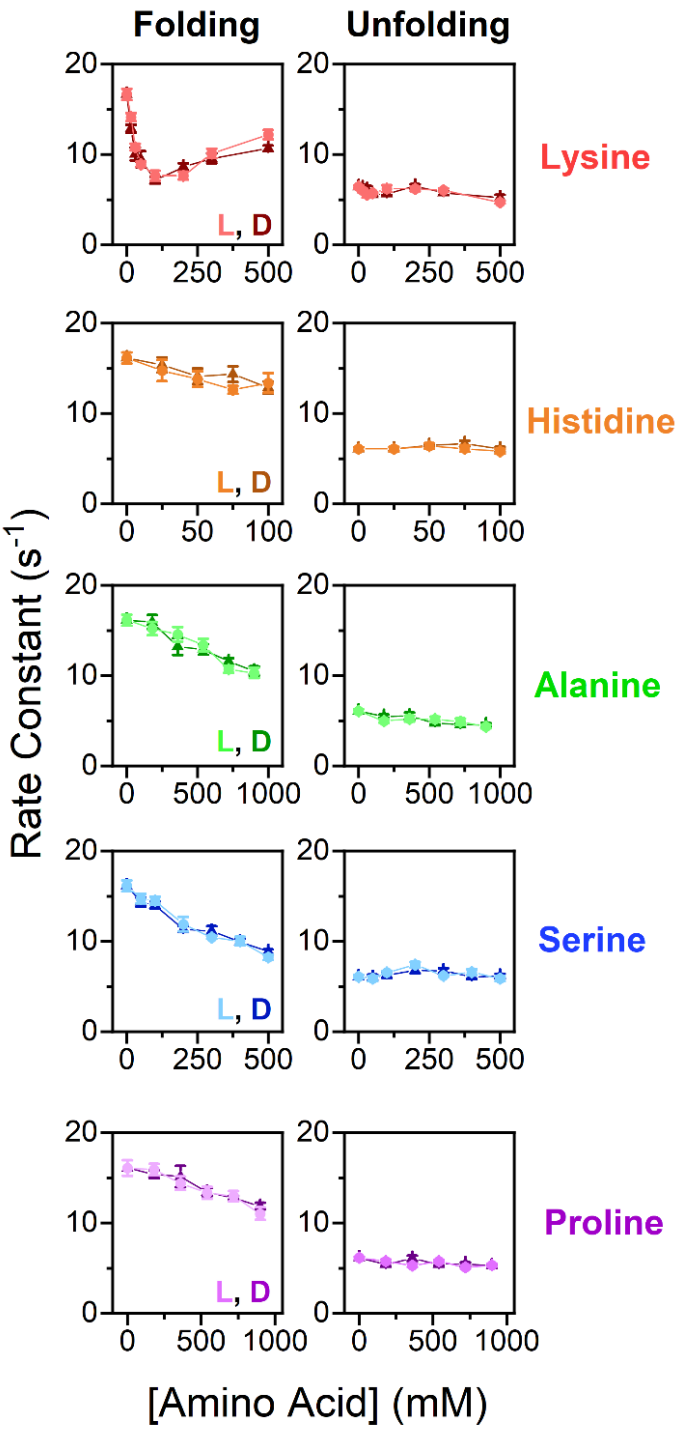

Figure 4.8 Rate constants for TL-TLR folding (left) and unfolding (right) in the presence of L-amino acids (light symbols) and their D-enantiomers (dark symbols). Concentration ranges for each amino acid are constrained by solubility.

and proline, whose amine group is incorporated into an aliphatic heterocycle. The TL-TLR response to each of these amino acids proves to be independent of chirality (Figure 4.8, rows 3-5). Indeed, among all of the amino acids tested, arginine and arginine alone induces chiral-specific modulation of tertiary folding rates in the TL-TLR construct.

Nevertheless, the different classes of amino acids tested do provide some insight into the salient aspects of arginine's effect on the TL-TLR. Firstly, all amino acids, regardless of side chain, can down regulate the folding rate. One simple explanation for this is that all amino acids engage in nonspecific binding to the surface of the nucleic acid, such that the *decrease* in surface area upon TL-TLR folding would be thermodynamically disfavored. However, note that not all amino acids have the same effectiveness in this regard; 100 mM histidine, for example, achieves the same degree of slowdown as 500 mM alanine, a five-fold greater sensitivity. Second, only the positively charged amino acids can, at higher concentrations, *increase* the folding rate. This behavior likely derives from simple ionic shielding of the nucleic acid phosphate backbone by electrolytes, similar to addition of monovalent atomic cations, whereby decreased repulsion between phosphate groups lowers the barrier to folding and increases the folding rate constant.[30, 52-53] Finally, and most importantly, arginine is unique in its ability to speed up the rate of *unfolding*. Furthermore, the chiral specific response to arginine is exclusively contained in the unfolding rate constant behavior. Structurally, these data would suggest that arginine's guanidinium-based side chain is responsible its chiral specificity. This is quite curious, however, as the guanidinium group itself is achiral and is 6–7 covalent C–C bonds removed from arginine's chiral center. This immediately raises many intriguing questions, such as how does arginine's chiral carbon affect the binding mode of the arginine molecule, and why does this not occur in the other tested amino acids, which undoubtedly also bind to RNA?

4.5 Molecular dynamics simulations

To provide insight into the mechanistic origin of such a chiral sensitivity to amino acid assisted RNA folding/unfolding, we have additionally performed computational modeling of the interactions between the TL-TLR single molecule construct and the two chiral enantiomers of arginine. At the outset, we first considered a "molecular docking" strategy,[54] for which a variety of candidate ligand-macromolecule binding "poses" are heuristically generated and scored to predict binding affinities to specific binding pockets. However, the use of molecular docking to predict relative binding free energies of enantiomeric pairs has been shown to be insufficiently reliable.[55] Therefore, the molecular docking approach was not selected for these studies. Instead, we choose to employ all-atom, explicit solvent molecular dynamics (MD)[56-57] as our computational strategy for assessing amino acid binding to the TL-TLR fold, exploiting well tested computational tools such as Nanoscale Molecular Dynamics (NAMD) and Visual Molecular Dynamics (VMD) for MD simulation as parallel *in silico* experiments. The key limitation of such calculations is the simulation timescale, which for the present construct sizes is typically < 10 μs, though longer run times can in principle be accessed with enhanced sampling methods[56, 58] or coarse-grained models.[59-60] However, as the TL-TLR construct used in these studies folds on the 10–100 *millisecond* timescale, efforts to simulate the full equilibrium dynamics of TL-TLR tertiary folding in real time would be computationally prohibitive. We instead take a conceptually simpler approach, performing three separate shorter timescale simulations with which to evaluate arginine distributions in equilibrium with each of the TL-TLR components. Specifically, we run three NPT simulations in parallel for 100 ns, probing L- and D-arginine attachment onto (i) the GAAA tetraloop (initial structure: PDB 1ZIF), (ii) the 11-nt tetraloop receptor (initial structure: PDB 1TLR), and (iii) the docked tetraloop-tetraloop

receptor contained in the P4-P6 domain of the *Tetrahymena thermophilia* intron (initial structure: PDB 1GID, truncated to TL-TLR residues; see methods). Each of the three initial TL, TLR, and TL-TLR structures is placed in a water box with periodic boundary conditions and appropriate potassium and chloride ion densities to mimic ionic strength values used experimentally, plus 100 mM of either L- or D-arginine (Figure 4.9; see methods for details). In effect, this piecewise

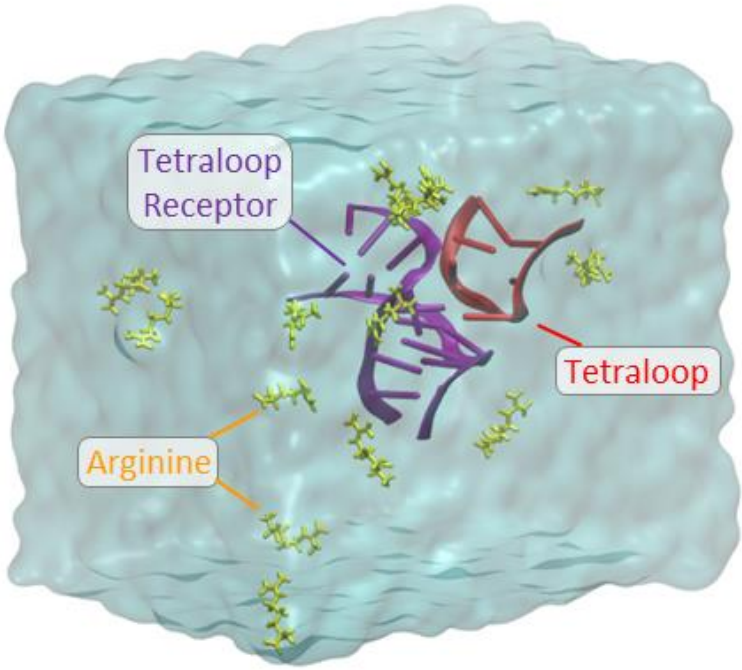

Figure 4.9 All-atom, explicit solvent simulation of arginine condensation onto the TL-TLR system. Simulations consist of nucleic acid (docked TL-TLR shown; TL in red, TLR in purple) solvated in TIP3P water with 100 mM L/D-arginine (yellow).

strategy serves to model arginine interaction with the TL-TLR in the vicinity of its folding and unfolding free energy minima, which circumvents the need for ultralong simulation timescales. The corresponding penalty is being unable to observe the folding transition state, which limits

dynamical interpretation. Nevertheless, these MD simulations offer first insights into the thermodynamics of the overall folding process.

As the first step in our analysis, we evaluate the interaction energy between the three nucleic acid constructs and their surrounding arginine ions. Free energies corresponding to individual snapshots 200 ps apart for each of the MD trajectories are calculated by the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method.[61] In MMPBSA, the binding free energy ΔG_{bind} is computed by combining the change in three energy contributions upon ligand binding:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv,PB} + \Delta G_{solv,SA}. \quad \text{(Eq. 4.4)}$$

In this expression, E_{MM} is molecular mechanics force field energy, $G_{solv,PB}$ is the solvation free energy under a Poisson-Boltzmann treatment, and $G_{solv,SA}$ is a semiempirical term accounting for the free energy of the solute-solvent interface and assumed proportional to surface area. One important caveat to the use of MMPBSA is that the entropic contribution to the binding free energy is incomplete. Specifically, the ligand-receptor conformational entropy change (ΔS_{conf}) is unaccounted for and cannot be rigorously incorporated. While there are a number of approximate methods for calculating ΔS_{conf}, in practice such computations fail to improve the accuracy and thus are not included in our estimates of ΔG_{bind}.[61] Despite such potential concerns, MMPBSA has previously been shown to accurately reproduce experimental relative binding free energies of enantiomer pairs,[62-63] motivating its use here.

These MMPBSA computations reveal two important features for arginine binding onto the single molecule nucleic acid construct. First, as depicted in Figure 4.10, the MMPBSA energies exhibit an initial, fast relaxation of t_{relax} ~10 ns, with an additional longer timescale relaxation which goes beyond our simulation time limits. The fast relaxation agrees reasonably

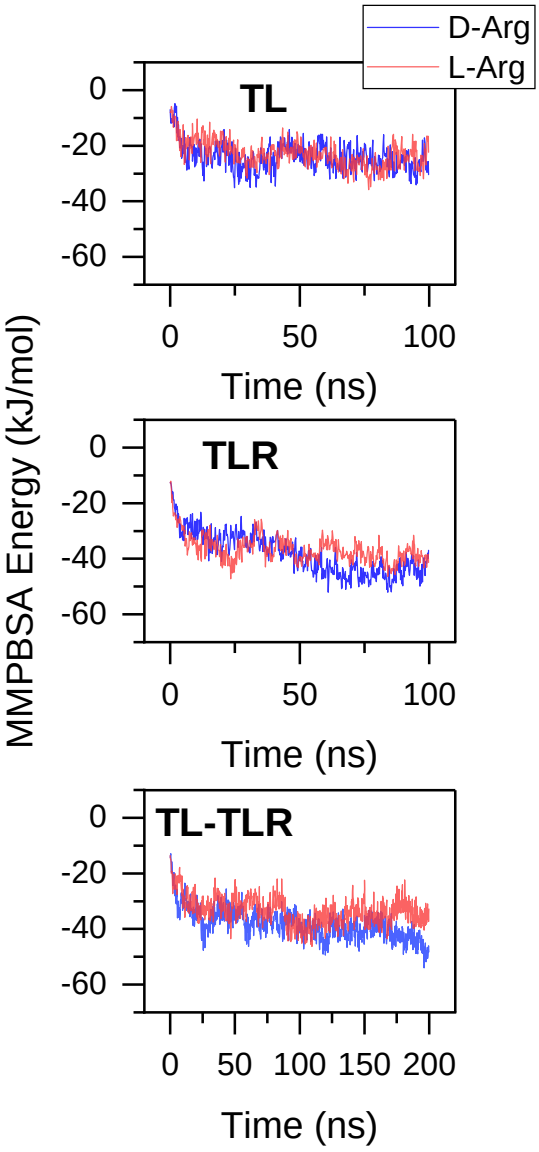

Figure 4.10 MMPBSA energies of arginine binding to RNA computed frame-by-frame for MD trajectories. Binding is with GAAA tetraloop (TL, top), 11-nt tetraloop receptor (TLR, middle), and docked tetraloop-tetraloop receptor (TL-TLR, bottom).

well with the predicted diffusion times for arginine to reach to the surface of the nucleic acid ($\tau_{diffusion} \sim R^2/D \sim 13$ ns), based on the typical distance between the arginine and nucleic acid (R ~ 30 Å) and the diffusion coefficient of arginine in water (D = 80 μm^2/s).[64] Energetic relaxation

appears to converge most quickly for the smallest simulated system, the lone tetraloop (TL). On the other hand, the TLR and docked TL-TLR constructs do not reach full equilibration after 100 ns, with arginine molecules continuing to explore a variety of binding poses, some more favorable than others. Even if we provide an additional +100 ns to the simulation, the TL-TLR construct still does not reach the equilibrium ergodic limit. It is not surprising that the binding free energies are not yet fully converged in ~ 200 ns, as small molecule docking to specific binding sites can take multiple microseconds to occur.[56-57]

Secondly, and of primary importance, the MD simulations reveal that MMPBSA energies of arginine interactions with nucleic acids to be both D-, L-enantiomer- and folding state-specific. In particular, the MMPBSA energies computed for L- and D-arginine interactions with i) TL, ii) TLR, and iii) the docked TL-TLR are plotted as a histogram in Figure 4.11. The MD simulations indicate only marginal differences in the binding affinities of L- and D-arginine to the TL ($\Delta\Delta G_{bind} = \Delta G_{bind,D\text{-}Arg} - \Delta G_{bind,L\text{-}Arg} = -1.3 \pm 1.9$ kJ/mol) and TLR ($\Delta\Delta G_{bind} = +1.6 \pm 4.1$ kJ/mol). The sum in the two chiral free energy differences is therefore $\Delta\Delta G_{bind,unfolded} = 0.3 \pm 4.5$, which is zero within the simulation uncertainties. However, for the docked TL-TLR simulation, there is a much larger difference in binding energies for L- and D-arginine, favoring D-arginine by $\Delta\Delta G_{bind,folded} = -5.3 \pm 3.4$ kJ/mol.

To better appreciate both the statistical significance and uncertainties of these chirally sensitive results, we have explored identical MD simulations with D/L-lysine in place of D/L-arginine (Figure 4.11, right panels). In particular, recall that lysine experimentally exhibits no chirally specific behavior in excess of the uncertainty limit (Figure 4.8), from which we would expect to see statistically vanishing chiral sensitivity in the molecular dynamics simulations. In terms of absolute magnitudes, we find that lysine displays ~2-fold smaller binding free energies

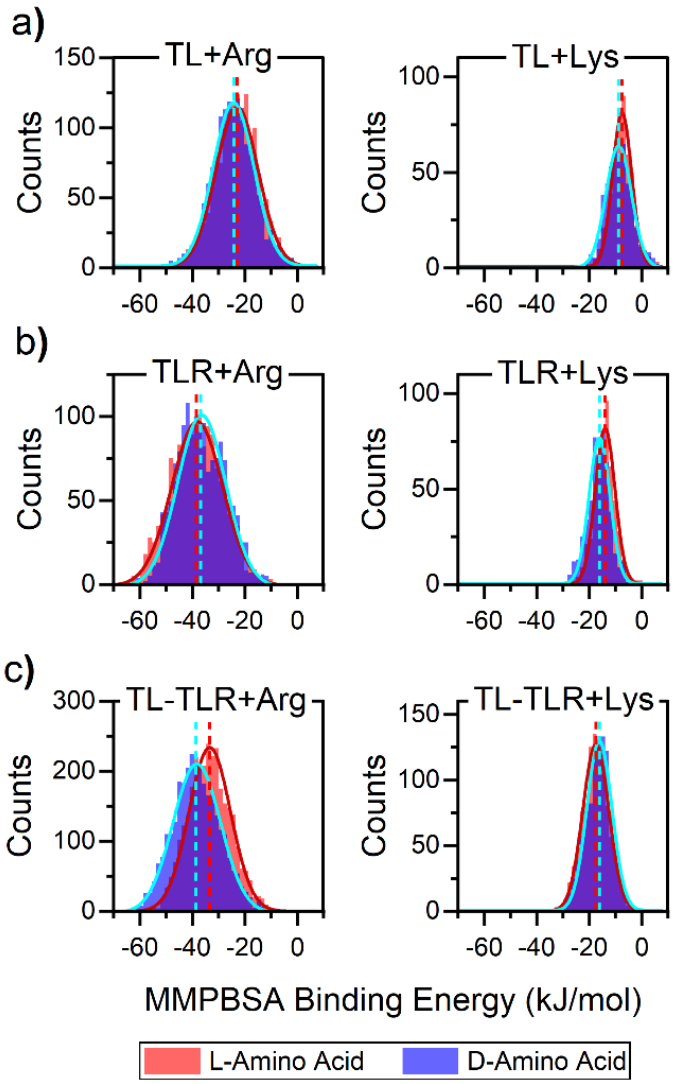

Figure 4.11 Histogram of MMPBSA-computed binding free energies for L- and D- enantiomers of arginine (left) and lysine (right) with (a) the undocked TL, (b) the undocked TLR, and (c) the docked TL-TLR. Solid curves are Gaussian fits, and vertical dashed lines are histogram centers.

than arginine, with chirally dependent differences only on the order of 1-2 kJ/mol ($\Delta\Delta G_{bind}$(TL) = +1.3 kJ/mol, $\Delta\Delta G_{bind}$(TLR) = +2.1 kJ/mol, $\Delta\Delta G_{bind}$(TL-TLR) = −1.4 kJ/mol). Indeed, from these values we can estimate roughly that the computational measurement uncertainties are on the same 1-2 kJ/mol order. Most importantly, we can conclude that arginine binding to the unfolded construct components (TL, TLR) is chirally insensitive within our MMPBSA

uncertainties, whereas arginine binding to the fully folded TL-TLR shows a distinct preference for D- vs L-arginine.

We can take these MD simulations one step further in an effort to identify the spatial region of preferred attachment for chiral D- vs L-arginine. To do this, 3D probability distributions of the arginine atoms $p(x,y,z)$ are computed by aligning MD trajectories and binning arginine atom locations using a 0.5 Å grid spacing. In the equilibrium thermodynamic limit, these probabilities can be converted into a 3D free energy distribution $\Delta G(x,y,z)$ via:[65]

$$\Delta G(x,y,z) = -k_B T \ln\left(\frac{p(x,y,z)}{p_0}\right)$$

where p_0 is the bulk concentration at the periphery of the simulation cell, as determined by the asymptotic value of the arginine-RNA radial distribution function (Figure 4.12). By way of example, the resulting two free energy isosurfaces for L- and D-arginine binding to the docked TL-TLR construct are plotted in Figure 4.13, corresponding to a relatively strong binding slice at

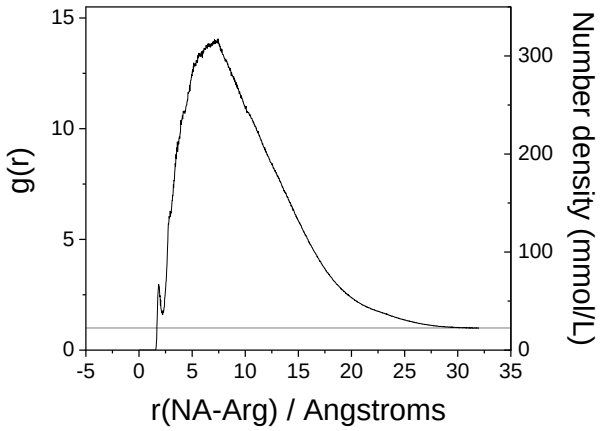

Figure 4.12 Representative radial distribution function g(r) of the distance between nucleic acid and arginine heavy atoms for the TL system. Large enrichment at medium distances is observed, corresponding to arginine cation recruitment into the ion atmosphere of the nucleic acid. For other systems, similar local concentrations of 10–25x the bulk concentration are observed. Asymptotic values of the number density (right axis) are used to compute the effective bulk concentration.

ΔG = −12 kJ/mol and therefore revealing the most significantly occupied binding pockets (We note that many weaker and stronger binding locations exist across the full surface of the nucleic acid. For a full 3D view, including isosurfaces for other binding pocket depths and nucleic acids, see Movies S1−S3 at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c07420?goto=supporting-info). Interestingly, most of these binding pockets are shared relatively equally by both L- and D-arginine. However, there is at least one notably large region at the cleft interface between the TL and TLR constructs to which D-arginine binds with significantly greater propensity than L-arginine, as identified with an arrow in Figure 4.13. Most importantly, MD simulations for the undocked TL and TLR, by way of contrast, show no binding pockets with chiral preference for L- or D-arginine (as evident in Movies S1 and S2). Furthermore, for parallel MD simulations with lysine binding, no chirally-specific binding pockets are observed for any of the TL, TLR,

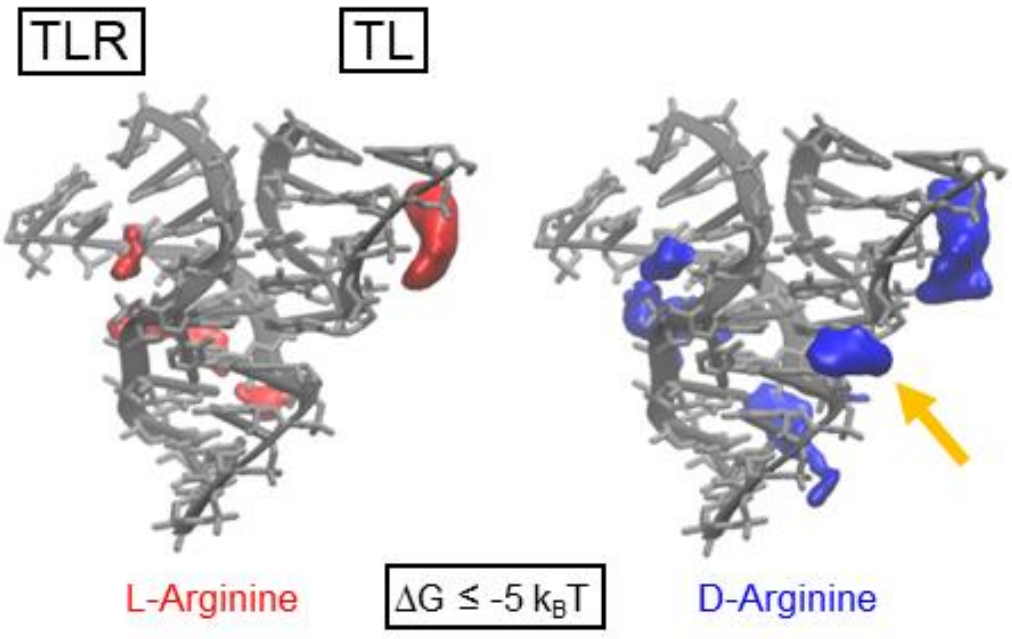

Figure 4.13 Free energy isosurfaces (ΔG ≤ -5 k_BT) for arginine binding to the docked TL-TLR structure. Arrow indicates the presence of an arginine-binding pocket preferential for the D- vs L-enantiomer at the interface of TL and TLR. For animated perspectives and additional free energy isosurfaces, go to https://pubs.acs.org/doi/10.1021/acs.jpcb.0c07420?goto=supporting-info.

and TL-TLR constructs simulated systems (Movies S4−S6), consistent with the complete lack of chiral sensitivity observed experimentally in amino acid induced folding/unfolding. Thus, although these simulations are run only for 100–200 ns and therefore do not fully sample the true equilibrium state, these MD trajectories do already reveal chiral enantiomer-specific binding interactions on the sub-microsecond timescale, qualitatively consistent with our single molecule experimental results. It is of course possible that even longer MD simulations would provide additional examples of stereospecific binding of arginine to this TL-TLR tertiary folding motif.

4.6 Discussion

This paper speculates that homochirality may have been transmitted to amino acids from RNA through chiral differences in amino acid-based modification of RNA folding behavior. To help assess such a conjecture, we have used single-molecule FRET measurements coupled with all-atom MD simulations, focusing on the 11-nt GAAA tetraloop-tetraloop receptor as a model RNA tertiary fold. The TL-TLR folding/unfolding of this tertiary motif is found to clearly depend on amino acid chirality for arginine (Figure 4.5), with only negligible sensitivities exhibited for each of the other classes (hydrophobic/hydrophilic, charged/uncharged, saturated/unsaturated) of the 6 amino acids investigated (Figure 4.8). Outside of simple empirical observation, the question remains: how does the chiral nature of D- vs L-arginine influence its interaction with the TL-TLR tertiary binding motif, and why might structurally similar amino acids not also exhibit a similarly strong chiral effect? We examine the evidence for relevant clues.

Firstly, the concentration-dependent data (Figures 4.5 and 4.7) reveal that arginine is unique among the 6 amino acids tested, specifically impacting the rate constant for TL-TLR

unfolding and indeed, buried therein lies all chiral dependence. Each of the other amino acids tested modify the folding rates but have only minimal effect on the unfolding rates. Lysine even shares arginine's nonmonotonic influence on the folding rate constant, likely a positive charge effect, but only in the presence of arginine does the TL-TLR *unfolding* rate constant change significantly, specifically increasing from $k_{unfold} = 5.4 \pm 0.4\ s^{-1}$ to $8.0 \pm 0.4\ s^{-1}$ with 100 mM L-arginine. Furthermore, such clear TL-TLR sensitivity to arginine chirality is manifested exclusively in the unfolding process, with D-arginine accelerating the unfolding rate constant by 50% more than L-arginine at 100 mM. This suggests a connection between arginine's special ability to enhance TL-TLR unfolding and its chiral-specific influence. Molecular dynamics (MD) simulations strongly support this connection, as any differential L- vs. D-arginine binding to the TL and TLR constructs is only observed in the folded (TL-TLR) state, whereas the unfolded state would appear to bind both enantiomers of arginine equivalently.

Secondly, results from temperature-dependent experiments reveal small but significant differences between the folding activation enthalpy of the TL-TLR motif with L- vs. D-arginine, with $\Delta\Delta H^{\ddagger}_{fold}$ (D- vs. L-) = 8 ± 4 kJ/mol in favor of the D-enantiomer. The differential binding of arginine to the folding, unfolding, and transitions states result in changes in the relative free energies of these states, with greater binding leading to greater free energy stabilization. Therefore, the chiral difference in activation enthalpy could either be due to preferential binding of D-arginine to the transition state of the TL-TLR, or superior L-arginine binding to the unfolded state, or a combination of the two. As the MMPBSA calculations suggest that each enantiomer of arginine binds equivalently to the separated TL and TLR constructs, the more consistent explanation would seem to be that D-arginine binds to the TL-TLR folding transition state more favorably than L-arginine, thereby reducing the activation enthalpy for folding . It is worth noting

that this reduced activation enthalpy is almost exactly compensated by an additional entropy cost ($\Delta\Delta S^{\ddagger}_{fold} = -26 \pm 12$ J/molK, raising the free energy barrier for D-arginine), presumably due to the loss of translational and conformational entropy with one or more arginine molecules binding. This results in vanishingly small differences in activation free energies ($\Delta\Delta G^{\ddagger}_{fold} = \Delta\Delta H^{\ddagger}_{fold} - T\Delta\Delta S^{\ddagger}_{fold}$) at room temperature, and therefore the *folding* rate constants for L- and D-arginine become largely indistinguishable. Less can be inferred about the unfolding process from temperature-dependent experiments, as the measured thermodynamic parameters for L- and D-arginine are within experimental uncertainty ($\Delta\Delta H^{\ddagger}_{unfold} < 3$ kJ/mol, $\Delta\Delta S^{\ddagger}_{unfold} < 10$ J/molK). That the thermodynamic differences are modest is not surprising; D-arginine's 30% faster unfolding rate constant k_{unfold} at 100 mM would from transition state theory require a difference in activation free energy, $\Delta\Delta G^{\ddagger} = 0.6$ kJ/mol, which is relatively small compared to the thermal energy product $k_BT = 2.4$ kJ/mol.

Our MD simulations have revealed that TL and TLR docking creates an interstitial cleft to which D-arginine binds more effectively than L-arginine. However, it must be the case that these simulations cannot fully explore all binding interactions, as a prediction of greater D-arginine stabilization of the folded state from MD would be in direct contradiction with the experimental results, for which we find that destabilization of the TL-TLR is found to be greater for D-arginine than L-arginine. One possible reason for such a discrepancy is the limited simulation time (100 ns) explored, which prevents the system from surmounting large kinetic barriers, for instance arginine displacement of a tightly bound potassium ion or large structural rearrangements in the nucleic acid. Presumably, at longer timescales an additional binding pocket preferential to L-arginine would emerge, and therefore long simulations and/or enhanced sampling techniques are necessary to resolve the ultimate cause of greater destabilization of the

TL-TLR fold by D-arginine. Nevertheless, it is interesting that even relatively weak transient binding can display strong chiral preferences, an observation which we might expect to be generalizable to other RNA-ligand systems. Furthermore, the fact that the chirality-sensitive binding pocket is localized at the interface of the TL and TLR suggests that RNA tertiary interfaces in general might hold the conformational key to generating such enantiospecific behavior.

MD simulations suggest that arginine's positive charge in combination with its sidechain's especially strong interactions with nucleic acids is responsible for its special ability to modulate TL-TLR folding in a chiral specific fashion. Due to the cationic nature of the arginine residue, it readily participates in the anionic atmosphere of the nucleic acid.[66] Indeed, the arginine-nucleic acid radial distribution function (Figure 4.12) show that arginine concentration near nucleic acids is enhanced by 10-25 fold over the bulk concentration. One might therefore expect such large local concentrations to enable arginine to sample weak, chirality-specific binding sites inaccessible to neutral or negatively-charged amino acids. With its positively charged amide side chain, lysine would also be expected to experience a similar increase in local ion atmosphere near the polyanionic backbone. However, the present MMPBSA calculations suggest and are at least consistent with the lysine side chain having a greatly reduced ability for binding to RNA than the guanidinium cation group in arginine.

The primary motivation of this study is to investigate a potential mechanism by which nucleic acid chirality was imprinted upon amino acids, and a few remarks on the evolutionary implications of these results are appropriate. In the hypothesis considered here, the role of amino acids in early biology was to promote correct RNA folding, perhaps by destabilizing misfolded states. This is similar to the role of modern chaperone proteins, which unfold misfolded proteins

to allow them a second chance to fold correctly. L-amino acids may have acted as superior chaperones to D-amino acids, leading to increased fitness for cells containing an enantiomeric excess of L-amino acids, which would drive evolution toward homochiral synthesis of L-amino acids. In our experiments, only one amino acid, arginine, is found to exhibit chiral-specific chaperone properties, while all other tested amino acids modify RNA folding independent of chirality. This casts doubt on the chaperone-homochirality hypothesis, as it seems unlikely that a single amino acid's interaction with RNA would drive synthesis of all amino acids to L-chirality. Furthermore, arginine is usually considered to have been a late arrival to the amino acid alphabet, based on its low prebiotic availability.[67] However, prebiotic availability may not be an appropriate criterion for the development of the genetic code, and recently Blanco et al. have highlighted arginine as a likely candidate for an early amino acid based on RNA-protein binding.[68] If arginine or another biophysically-related amino acid was highly enriched in ancient biology, then the chiral sensitivity observed in these studies may have been significant enough to steer evolution toward L-amino acids.

4.7 Conclusion

Single molecule experiments have been used to explore the potential sensitivity of RNA tertiary folding rate constants and equilibria to amino acid chirality. Of the 4 classes of amino acids studied, only arginine exhibits any chiral-specific influence on the RNA folding equilibrium for the TL-TLR. Both enantiomers of arginine increase the unfolding rates and thereby destabilize the RNA, but the non-natural enantiomer (D-arginine) is found to be more strongly-destabilizing, by up to a factor of 50% difference in equilibrium constant at 300 mM. Kinetically the source of this effect is exclusively rooted in changes the unfolding rate constant, with essentially no

measurable difference in the corresponding folding rate constant with D- vs L-arginine chirality. From temperature dependent studies of the rate constants, these chiral sensitivities can be traced back to simple thermodynamical variables, specifically with a reduced activation enthalpy for folding ($\Delta\Delta H^{\ddagger} = 8 \pm 4$ kJ/mol) in the presence of D-arginine. In order to obtain additional preliminary insights into the mechanism for such chiral sensitivity, we have explored molecular dynamics (MD) simulations for the folding and unfolding events with NAMD and VMD computer platforms, specifically using 100 ns long trajectories to discover chirally sensitive free energy differences in D- vs L- arginine binding to the TL, TLR and TL-TLR constructs. Interestingly, the results indicate a complete lack of chiral sensitivity to the unfolded TL and TLR species, but with a clear differential chiral effect on D- vs. L-arginine binding to the fully folded TL-TLR construct. Furthermore, the probability distributions from the MD simulations have been used to generate a 3D free energy landscape for binding of D- and L-arginine to the folded TL-TLR tertiary motif, revealing that RNA-RNA tertiary interfaces may provide a more general source of chirality-sensing binding pockets for enhanced D- vs L-arginine attachment. These preliminary computational results corroborate many but not all of the chiral experimental findings, offering first insights in support of a putative mechanism for RNA chirality influencing and being influenced by associated amino acid chirality.

4.8 Supporting information

Animations of full 3D amino acid binding distributions S1−S6 are available online at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c07420?goto=supporting-info.

4.9 References

1. Blackmond, D. G., The origin of biological homochirality. *Cold Spring Harbor Perspect. Biol.* **2010,** *2* (5), a002147-a002147.
2. Zask, A.; Ellestad, G., Biomimetic syntheses of racemic natural products. *Chirality* **2018,** *30* (2), 157-164.
3. Finefield, J. M.; Sherman, D. H.; Kreitman, M.; Williams, R. M., Enantiomeric natural products: occurrence and biogenesis. *Angew. Chem., Int. Ed. Engl.* **2012,** *51* (20), 4802-4836.
4. Banik, S. D.; Nandi, N., Chirality and protein biosynthesis. In *Biochirality: Origins, Evolution and Molecular Recognition*, Cintas, P., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 255-305.
5. Genchi, G., An overview on D-amino acids. *Amino Acids* **2017,** *49* (9), 1521-1533.
6. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* **2002,** *418* (6894), 214-21.
7. Tamura, K.; Schimmel, P., Chiral-selective aminoacylation of an RNA minihelix. *Science (New York, N.Y.)* **2004,** *305* (5688), 1253.
8. Tamura, K., Molecular basis for chiral selection in RNA aminoacylation. *Int. J. Mol. Sci.* **2011,** *12*, 4745-4757.
9. Herschlag, D., RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **1995,** *270*, 20871-20874.
10. Woodson, S. A., Taming free energy landscapes with RNA chaperones. *RNA Biol.* **2010,** *7* (6), 677-86.
11. Chen, S. J., RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **2008,** *37*, 197-214.
12. Saibil, H., Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell. Biol.* **2013,** *14* (10), 630-642.
13. Arakawa, T.; Timasheff, S. N., The stabilization of proteins by osmolytes. *Biophys. J.* **1985,** *47* (3), 411-414.
14. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: kinetic insights from single-molecule studies. *J. Phys. Chem. B.* **2018,** *122* (43), 9869-9876.
15. Sengupta, A.; Sung, H. L.; Nesbitt, D. J., Amino acid specific effects on rna tertiary interactions: single-molecule kinetic and thermodynamic studies. *J. Phys. Chem. B.* **2016,** *120*, 10615-10627.

16. Yarus, M., A specific amino acid binding site composed of RNA. *Science (New York, N.Y.)* **1988,** *240* (4860), 1751-8.

17. Burg, M. B.; Ferraris, J. D., Intracellular organic osmolytes: function and regulation. *J. Biol. Chem.* **2008,** *283* (12), 7309-7313.

18. Khan, S. H.; Ahmad, N.; Ahmad, F.; Kumar, R., Naturally occurring organic osmolytes: from cell physiology to disease prevention. *IUBMB Life* **2010,** *62* (12), 891-895.

19. Ando, T.; Takahashi, S.; Tamura, K., Principles of chemical geometry underlying chiral selectivity in RNA minihelix aminoacylation. *Nucleic Acids Res.* **2018,** *46* (21), 11144-11152.

20. Tamura, K., RNA-directed molecular asymmetry of amino acids. *Viva Origino* **2010,** *38* (4), 18-22.

21. Illangasekare, M.; Yarus, M., Phenylalanine-binding RNAs and genetic code evolution. *J. Mol. Evol.* **2002,** *54* (3), 298-311.

22. Yarus, M., Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. *J. Mol. Evol.* **1998,** *47* (1), 109-117.

23. Yarus, M., RNA-ligand chemistry: a testable source for the genetic code. *RNA* **2000,** *6* (4), 475-484.

24. Illangasekare, M.; Turk, R.; Peterson, G. C.; Lladser, M.; Yarus, M., Chiral histidine selection by D-ribose RNA. *RNA* **2010,** *16* (12), 2370-2383.

25. Tohala, L.; Oukacine, F.; Ravelet, C.; Peyrin, E., Chiral resolution capabilities of DNA oligonucleotides. *Anal. Chem.* **2015,** *87* (11), 5491-5495.

26. Sivaleela, T.; Kumar, M. R.; Prabhakar, S.; Bhaskar, G.; Vairamani, M., Chiral discrimination of α-amino acids by DNA tetranucleotides under electrospray ionization conditions. *Rapid Commun. Mass Spectrom.* **2008,** *22* (2), 204-210.

27. Yarus, M.; Majerfeld, I., Co-optimization of ribozyme substrate stacking and L-arginine binding. *J. Mol. Biol.* **1992,** *225* (4), 945-9.

28. Holmstrom, E. D.; Nesbitt, D. J., Biophysical Insights from Temperature-Dependent Single-Molecule Forster Resonance Energy Transfer. *Annu. Rev. Phys. Chem.* **2016,** *67*, 441-65.

29. Fiore, J. L.; Nesbitt, D. J., An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* **2013,** *46* (3), 223-64.

30. Downey, C. D.; Fiore, J. L.; Stoddard, C. D.; Hodak, J. H.; Nesbitt, D. J.; Pardi, A., Metal ion dependence, thermodynamics, and kinetics for intramolecular docking of a GAAA tetraloop and receptor connected by a flexible linker. *Biochemistry* **2006,** *45*, 3664-3673.

31. Seol, Y.; Skinner, G. M.; Visscher, K.; Buhot, A.; Halperin, A., Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys. Rev. Lett.* **2007,** *98* (15), 158103.

32. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* **2008,** *5* (6), 507-16.

33. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on Chips II*, Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

34. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **2008,** *94* (5), 1826-35.

35. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

36. Mullikin, J. C.; Vliet, L. J. v.; Netten, H.; Boddeke, F. R.; Feltz, G. v. d.; Young, I. T., *Methods for CCD camera characterization*. SPIE: 1994; Vol. 2173.

37. Cohen, E. A. K.; Ober, R. J., Image registration error analysis with applications in single molecule microscopy. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012; pp 996-999.

38. Hodak, J. H.; Fiore, J. L.; Nesbitt, D. J.; Downey, C. D.; Pardi, A., Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (30), 10505.

39. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden,; R.E. Duke, D. G., G. Giambasu, et al. *AMBER 2019*, University of California, San Francisco, 2019.

40. Mark, P.; Nilsson, L., Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A.* **2001,** *105* (43), 9954-9960.

41. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graphics* **1996,** *14* (1), 33-38.

42. Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P., Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011,** *7* (9), 2886-2902.

43. Joung, I. S.; Cheatham, T. E., 3rd, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* **2008,** *112* (30), 9020-41.

44. Horn, A. H., A consistent force field parameter set for zwitterionic amino acid residues. *J. Mol. Model.* **2014,** *20* (11), 2478.

45. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.

46. Stone, J. E. An efficient library for parallel ray tracing and animation. Master's Thesis, University of Missouri, Rolla, 1998.

47. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Tests of Kramers' theory at the single-molecule level: evidence for folding of an isolated RNA tertiary interaction at the viscous speed limit. *J. Phys. Chem. B.* **2018,** *122* (38), 8796-8804.

48. Hori, N.; Denesyuk, N. A.; Thirumalai, D., Frictional effects on RNA folding: speed limit and Kramers turnover. *J. Phys. Chem. B.* **2018,** *122* (49), 11279-11288.

49. Laidler, K. J.; King, M. C., Development of transition-state theory. *J. Phys. Chem.* **1983,** *87* (15), 2657-2664.

50. Best, R. B.; Hummer, G., Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys. Rev. Lett.* **2006,** *96* (22), 228104.

51. Amend, J. P.; Helgeson, H. C., Solubilities of the common L-α-amino acids as a function of temperature and solution pH. *Pure Appl. Chem.* **1997,** *69* (5), 935-942.

52. Bisaria, N.; Herschlag, D., Probing the kinetic and thermodynamic consequences of the tetraloop/tetraloop receptor monovalent ion-binding site in P4-P6 RNA by smFRET. *Biochem. Soc. Trans.* **2015,** *43* (2), 172-8.

53. Holmstrom, E. D.; Fiore, J. L.; Nesbitt, D. J., Thermodynamic origins of monovalent facilitated RNA folding. *Biochemistry* **2012,** *51* (18), 3732-43.

54. Meng, X. Y.; Zhang, H. X.; Mezei, M.; Cui, M., Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput.-Aided Drug Des.* **2011,** *7* (2), 146-57.

55. Ramírez, D.; Caballero, J., Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *Int. J. Mol. Sci.* **2016,** *17* (4), 525.

56. Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; Walter, N. G.; Otyepka, M., RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **2018,** *118* (8), 4177-4338.

57. Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kuhrova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M., Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J. Phys. Chem. Lett.* **2014,** *5* (10), 1771-82.

58. Mlynsky, V.; Bussi, G., Exploring RNA structure and dynamics through enhanced sampling simulations. *Curr. Opin. Struct. Biol.* **2018,** *49*, 63-71.

59. Hyeon, C.; Thirumalai, D., Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011,** *2*, 487.

60. Jonikas, M. A.; Radmer, R. J.; Laederach, A.; Das, R.; Pearlman, S.; Herschlag, D.; Altman, R. B., Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* **2009,** *15* (2), 189-199.

61. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* **2015,** *10* (5), 449-461.

62. Kocakaya, S. Ö.; Turgut, Y.; Pirinççioglu, N., Enantiomeric discrimination of chiral organic salts by chiral aza-15-crown-5 ether with C1 symmetry: experimental and theoretical approaches. *J. Mol. Model.* **2015,** *21* (3), 55.

63. Choi, Y.; Jung, S., Molecular dynamics (MD) simulations for the prediction of chiral discrimination of N-acetylphenylalanine enantiomers by cyclomaltoheptaose (β-cyclodextrin, β-CD) based on the MM–PBSA (molecular mechanics–Poisson–Boltzmann surface area) approach. *Carbohydr. Res.* **2004,** *339* (11), 1961-1966.

64. Wu, Y.; Ma, P.; Liu, Y.; Li, S., Diffusion coefficients of L-proline, L-threonine and L-arginine in aqueous solutions at 25°C. *Fluid Phase Equilib.* **2001,** *186* (1), 27-38.

65. Chen, A. A.; Draper, D. E.; Pappu, R. V., Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. *J. Mol. Biol.* **2009,** *390* (4), 805-819.

66. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **2014,** *83*, 813-41.

67. Trifonov, E. N., Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **2000,** *261* (1), 139-51.

68. Blanco, C.; Bayas, M.; Yan, F.; Chen, I. A., Analysis of evolutionarily independent protein-RNA complexes yields a criterion to evaluate the relevance of prebiotic scenarios. *Curr. Biol.* **2018,** *28* (4), 526-537 e5.

Chapter 5

Pushing Camera-Based Single Molecule Kinetic Measurements to the Frame Acquisition Limit with Stroboscopic smFRET[*]

5.1 Abstract

Single-molecule fluorescence resonance energy transfer (smFRET) experiments permit detailed examination of microscopic dynamics. However, kinetic rate constants determined by smFRET are susceptible to systematic underestimation when the rate constants are comparable to the data acquisition rate. We demonstrate how such systematic errors in camera-based TIRF experiments can be greatly reduced by using stroboscopic illumination/detection, allowing accurate rate constant determination up to the data sampling rate and yielding an order of magnitude increase in dynamic range. Implementation of these stroboscopic smFRET ideas is straightforward and the stroboscopically obtained data are compatible with multiple trajectory analysis methods, including dwell time analysis and hidden Markov modeling. Such stroboscopic methods therefore offer a remarkably simple yet valuable addition to the smFRET toolkit, requiring only relatively trivial modification to the normal data collection and analysis procedures.

5.2 Introduction

Single-molecule microscopy is a powerful tool for examining kinetic systems at otherwise inaccessible levels of detail. From the folding of biopolymers,[1-2] to single enzyme catalysis,[3-4] to single DNA replication[5] and transcription,[6] both *in vitro* and *in vivo*,[7-9] as well as

---

[*] This chapter is adapted from: Nicholson, D. A.; Nesbitt, D. J. *J. Phys. Chem. B* **2021,** *125* (23), 6080-6089.

nonbiological applications,[10-11] the ability to probe at the sub-ensemble level clarifies underlying mechanisms in ways that are impossible in bulk studies. Especially powerful in the field of biophysics has been the use of single-molecule fluorescence resonance energy transfer (smFRET),[12-14] in which spatial motion on biologically relevant length scales (1–10 nm) can be converted into a colorimetric ratio, the FRET efficiency (E_{FRET}). By monitoring the time-dependence of E_{FRET} for surface-tethered molecules, structural rearrangements can be observed in real time. Stochastic state-to-state hopping in E_{FRET} trajectories can then be analyzed to extract dynamical information, e.g., the number of thermally accessible states and the rate constants for interconversion between those states.

smFRET experiments perform well when rate constants are slow compared to the data acquisition rate, but rate constants that are comparable to or exceed the sampling rate pose additional challenges. In particular, these faster rate constants are susceptible to systematic underestimation, with bias becoming significant even for as little as $k \geq 10\%$ of the sampling rate.[15] This systematic error/bias arises from well-known distortion of the FRET distributions by "camera averaging" or "camera blurring."[16-17] Similar to "motional blur" in conventional still photography, the multiple FRET states become temporally averaged over (i.e., "blur") with increasingly larger exposure times, introducing artifactual density in the FRET histograms and intercalation between the true FRET values. These blurred FRET values present problems for accurate kinetic analysis of the data and can even incorrectly suggest the existence of additional, nonphysical states.[15]

The simplest way to avoid such bias is obviously to increase the acquisition rate, which for camera-based detection, options include cropping the field of view or performing on-chip pixel binning. To push hardware limitations even further, Tang et al. have achieved increased

effective frame rates through clever use of sparse imaging and a galvo-mirror.[18] However, each of these options requires sacrificing some multiplexing capacity of the experiment, which is indeed a primary benefit of widefield methods over non-multiplexed detection such as scanning confocal microscopy.[19-20] Instead, it would be more desirable to mitigate camera averaging effects while still maintaining the highest possible parallel throughput of data.

More sophisticated smFRET data analysis methods incorporate such camera averaging effects directly. For instance, much work has been put into modeling the resulting distortions in these FRET distributions,[21] as in probability distribution analysis (PDA).[22-24] A more agnostic approach is taken by Chen et al.[25] in which the experimentally-determined (underestimated) rate constants are least squares compared with those derived from simulated data subject to the same camera averaging artifact. The parameters of the kinetic model are then iteratively varied until the simulated results match the experiment. While effective, this simulate-compare-iterate approach requires accurate modeling of the experimental system and is therefore prone to introducing additional systematic errors.

In the present work, we propose a relatively simple experimental solution based on stroboscopic illumination to eliminate camera averaging effects. Stroboscopes have been used before in single-molecule microscopy, especially in single-particle tracking, where a flashing light source can reduce diffusional blurring to increase accuracy in position determination.[26-28] Our interest in the strobe is not to reduce *motional* blur, but rather to reduce *temporal* blur. Our development builds on the work of Farooq et al.,[29] who used stroboscopic imaging to mimic the fluorescence burst data from confocal diffusing studies and then analyzed the FRET histogram using PDA. Here, we show that stroboscopic data can be analyzed as trajectories, which has the benefit of utilizing the full information content of the smFRET data and permits rate constant

determination up to the data acquisition rate. Under typical smFRET experimental conditions, this translates into roughly an order of magnitude enhancement in data collection bandwidth for trajectory-based analysis, free from systematic underestimation of rate constants.

5.3 Methodology

5.3.1 Theory of single-molecule kinetics

The kinetic systems studied in single-molecule experiments can be abstractly represented as a finite set of discrete states $\{S_i\}$ which undergo state-to-state transitions governed by first order kinetics. Specifically, the state transitions are considered Poisson processes, with k_{ij} as the unimolecular rate constants for state S_j converting to state S_i. The rate of change of the population in state S_i is the sum of the total loss rate and the total production rate:

$$\frac{d}{dt}S_i = -\sum_{j\neq i} k_{ji}S_i + \sum_{j\neq i} k_{ij}S_j. \qquad \text{(Eq. 5.1)}$$

One may compactly express the total rate of change of all states in matrix-vector form by forming the state vector $\mathbf{S} = \{S_1, S_2, \ldots\}$ and the rate matrix $\mathbf{K}$ (with off-diagonal elements $K_{ij} = k_{ij}$ and on-diagonal elements $K_{ii} = -\sum_{j\neq i} k_{ji}$) which satisfy the master equation

$$\frac{d}{dt}\mathbf{S} = \mathbf{KS}. \qquad \text{(Eq. 5.2)}$$

From Eq. 5.1 it is clear that the sum of any column of K vanishes, which is equivalent to conservation of molecular number. The formal solution to this system of differential equations is

$$\mathbf{S}(t) = e^{\mathbf{K}t}\mathbf{S}_0 = \mathbf{T}(t)\mathbf{S}_0 \qquad \text{(Eq. 5.3)}$$

where $\mathbf{S}_0$ is the initial state distribution at $t = 0$ and $\mathbf{T}(t)$ is the time evolution operator. In Eq. 5.3, the matrix exponential operator is evaluated by Taylor series expansion as $e^{\mathbf{K}t} = (\mathbf{I} + \mathbf{K}t + (\mathbf{K}t)^2/2 + (\mathbf{K}t)^3/(3!) + \ldots)$, where $\mathbf{K}t$ simply multiplies each element by t and $\mathbf{I}$ is the identity matrix.

To connect this formalism, which is based on ensemble chemical kinetics, to the dynamics of single-molecule systems, $\mathbf{T}(t)$ is interpreted as the matrix of transition probabilities at lag time t. Specifically, the matrix element $T_{ji}(t\text{-}t')$ is the probability that the system will be in state j at time t after being in state i a time t' earlier. Note that this does not specify which path the system took to reach the final state j. Indeed, $T_{ji}(t)$ includes all possible trajectories the system could take to go from i to j (e.g., $i \rightarrow j$, $i \rightarrow k \rightarrow j$, $i \rightarrow j \rightarrow i \rightarrow j$). For stroboscopic experiments this has important implications, discussed below, in which the experimentalist is "blind" for some fraction of the observation time window.

5.3.2 Photon-by-photon trajectory simulation

Single-molecule trajectories are simulated using a photon-by-photon approach, similar to that of Szabo and Gopich.[30] This is carried out in two steps: i) simulation of the state of the molecular system as a function of time S(t), followed by ii) simulation of the fluorescently emitted and detected photons. First, the initial state S_0 (e.g., 0 or 1 for $i=1, n$) for the molecule is randomly selected from an equilibrium probability vector $\mathbf{p}_{eq}$, corresponding to the unique eigenvector of the rate matrix $\mathbf{K}$ with an eigenvalue of zero. Time is then iteratively forward-propagated by randomly choosing a dwell time for the current state i, which is exponentially distributed with a time constant associated with the total loss rate from that state, i.e., $-1/K_{ii}$. The next state index is then randomly selected, with a probability weighted by the branching ratio into the selected target state j, $P(i \rightarrow j) = K_{ji}/\sum_{j \neq i} K_{ji}$, with this exponential time propagation continued until the desired total simulation time is achieved.

After the state trajectory $S(t)$ is generated, a sequence of individual photon detection events is produced. Similar to the case for state transitions, photon detection is assumed to be a Poisson process, which is accurate for fluorophore excitation rate small compared to the

fluorescence rate $I_{fluor} = 1/\tau_{fluor}$, where τ_{fluor} is the fluorescence lifetime. In this limit, single photons will arrive exponentially distributed in time as determined by the photon detection rate k_{photon}. In general, k_{photon} can be a function of the conformational state of the molecule, for example, due to a FRET pair with different quantum yields. However, many FRET dye pairs, in particular the Cy3-Cy5 pair used in the present experiments, have very similar quantum yields, which motivates treating k_{photon} as constant, though deviations could be easily incorporated into the kinetic model. The color of the emitted photon is probabilistically distributed based on the conformation (i.e., FRET state) of the molecule at the time of the excitation event. Specifically, the photon is labelled as an "acceptor" with probability E_i or a "donor" with probability $(1-E_i)$, where E_i is the FRET efficiency of the molecule in state i. As with the state trajectories, the process of exponential time-jumping and assignment of photon color is continued until the full desired simulation window is achieved. Lastly, photons are binned at the frame time Δt_{frame}, to generate donor and acceptor average intensities as a function of time, $D(t)$ and $A(t)$ respectively. Note that all noise considered in the modeled data arises exclusively from quantum fluctuations ("shot noise") in the photon counting process. While other sources such as dark count noise and read noise[31] are also present in camera-based smFRET, the experimental conditions we are interested place the system well within the limit where shot noise dominates over all other sources of noise. This simulation method can be extended to include background donor and acceptor average countrates (B_D and B_A); however, the effect of adding background photons can be equivalently achieved by (1) shifting the FRET efficiencies and (2) increasing the relative noise by decreasing the photon detection rate, and therefore for simplicity we have set the background intensities to zero in these studies. MATLAB code for trajectory simulation and analysis (see below) are available freely upon request.

5.3.3 Single molecule microscopy

smFRET experiments are performed on a DNA hairpin by total internal reflection fluorescence (TIRF) microscopy, as described previously.[32] Briefly, a 7 bp DNA hairpin[33] with a 40-adenine loop is biotinylated at the 3' end and attached to a glass surface via streptavidin-biotin binding.[34] The hairpin is labeled with Cy3 and Cy5 for FRET-based conformational detection. The DNA construct is imaged in 50 mM HEPES buffer (pH = 7.6) with 70 mM total monovalent cations (K^+ and Na^+) and a PCA-PCD-TROLOX oxygen scavenging and triplet quenching cocktail for enhanced fluorophore photostability.[35-36] A diode-driven Nd:YAG laser illuminates the surface-attached DNA construct in a through-objective TIRF configuration.[37] Fluorescence is separated by a dichroic mirror into donor (Cy3) and acceptor (Cy5) channels, which are each focused onto one half of an intensified charge-coupled device (ICCD) array. To achieve stroboscopic (gated) illumination, the diode light output is modulated by a variable duty cycle square wave current profile generated in LabVIEW (NI, Austin, TX), which also triggers the ICCD to initiate frame acquisition. Movies are analyzed using homebuilt software programmed in LabWindows/CVI to extract single molecule trajectories. Particles are located by brightness thresholding and then sorted into donors and acceptors and paired based on relative location. Finally, integration inside a 2-pixel radius around particle centers then generates donor and acceptor intensity trajectories, $D(t)$ and $A(t)$.

5.3.4 Single molecule trajectory analysis

5.3.4.1 Dwell time analysis

In the simplest mode of analysis, time dependent FRET trajectories are computed from the binned donor ($D(t)$) and acceptor ($A(t)$) intensities by $\mathrm{FRET}(t) = A(t) / (A(t) + D(t))$, where we assume any corrections for differential donor vs. acceptor quantum yields to be negligible.

Simple two-state thresholding at the arithmetic mean of low and high FRET values is applied to the FRET trajectories to determine the state function, $S(t)$, with the temporal duration between threshold crossings yielding a histogram of dwell times $N(\Delta t_{dwell})$. This dwell time histogram is then re-expressed as the "survival probability" $P(\Delta t_{dwell})$ for a given conformational state lasting longer than Δt_{dwell}, with single exponential fits for folded state dwell times (Δt_{fold}) yielding the unfolding rate constant k_{unfold} (and vice versa).

5.3.4.2 Hidden Markov modeling

As a second analysis scheme, Hidden Markov modeling (HMM) is performed by computing the likelihood function (L) of observing an experimental FRET trajectory given a string of model parameters, and then finding the string of parameters that maximizes L. For a FRET trajectory E_n of temporal length $N\Delta t_{frame}$, the scalar probability L is computed as

$$L = \mathbf{1}^{\mathrm{T}} * [\prod_{n=2}^{N} \mathbf{O}(\mathrm{E}_n) * e^{\mathbf{K}*\Delta t_{\mathrm{frame}}}] * \mathbf{O}(\mathrm{E}_1) * \mathbf{p}_{\mathrm{eq}}. \qquad \text{(Eq. 5.4)}$$

This equation is read right to left, where $\mathbf{p}_{eq}$ is the equilibrium probability vector, $\mathbf{O}(E_n)$ is the diagonal matrix of observation probabilities for the n^{th} observed FRET value, $\mathbf{K}$ is the rate matrix, Δt_{frame} is the time between frames, and $\mathbf{1}^T$ is the row vector (1, 1, …) of length equal to the number of states in the system. Diagonal elements of the observation matrix (O_{ii}) are the probabilities of observing a FRET value in the state i, which we model as Gaussian variables, i.e. $O_{ii}(E) \propto \exp(-(E-E_i)^2/2\sigma_i^2)$ with center E_i and width σ_I. Gradient descent is used to determine the set of parameters (k_{ij}, E_i, σ_i) which maximize L, where L is periodically renormalized to avoid instabilities due to numerical underflow.[38]. Note that this probability is maximized by optimization of the rate matrix elements K_{ij} rather than the transition probabilities T_{ij}, which avoids inaccuracies and instabilities due to computation of the matrix logarithm $\mathbf{K} = \ln(\mathbf{T})/\Delta t_{frame}$.

5.3.4.3 Time correlation function analysis

In the third kinetic analysis approach, time correlation functions (TCFs) are computed for the four possible combinations of donor and acceptor intensities: $\langle D(t)D(t+\tau)\rangle$, $\langle D(t)A(t+\tau)\rangle$, $\langle A(t)D(t+\tau)\rangle$, $\langle A(t)A(t+\tau)\rangle$. For example, $\langle D(t)A(t+\tau)\rangle$ reflects the donor-acceptor cross-correlation function, given by

$$\langle D(t)A(t+\tau)\rangle = \sum_i P_i\, D_i \left(\sum_f P_{i\to f}(\tau)\, A_f\right). \qquad \text{(Eq. 5.5)}$$

In Eq. 5.5, P_i is the equilibrium probability of initial state i, D_i the average donor intensity in state i, $P_{i\to f}(\tau)$ is the transition probability from i to f in lag time τ, A_f is the average acceptor intensity in the final state f, and the summation is over all initial and final states. Additional contributions to the TCF from fast dynamics (e.g. , fluorophore blinking) are not included, as the time resolution of these CCD measurements is typically in the 10s of milliseconds domain, far longer than the timescales of such photophysical behavior in Cy3 and Cy5[36] (in effect, these contributions are statically incorporated into A_i and D_i). We may therefore generalize to any combination of donor and acceptor TCFs by writing Eq. 5.5 in matrix form as

$$\frac{\langle C_1(t)C_2(t+\tau)\rangle}{I^2} = \mathbf{1}^{\mathrm{T}} * \mathbf{E}(C_2) * \mathbf{T}(\tau) * \mathbf{E}(C_1) * \mathbf{p}_{\mathrm{eq}} \qquad \text{(Eq. 5.6)}$$

where the channel C_k is now labelled by either donor D or acceptor A, $\mathbf{p}_{\mathrm{eq}}$ is the equilibrium probability vector, $\mathbf{E}(C_k)$ is a diagonal matrix whose diagonal elements are FRET efficiencies in each conformational state (i.e., E_i if $C_k = A$ or $(1\text{-}E_i)$ for $C_k = D$), $\mathbf{T}(\tau)$ is the transition probability matrix at lag time τ equal to $\exp(\mathbf{K}\tau)$, and $\mathbf{1}^{\mathrm{T}}$ is the row vector (1, 1, …) of length equal to the number of states in the system. Here we have normalized the TCFs by the square of the total intensity ($I^2 = (D_i + A_i)^2$), which for comparable quantum yields is only weakly dependent on the system state, and have exploited the simple definition of the FRET efficiency as $E_i = A_i / (A_i + D_i)$. The four TCF's are evaluated at nonnegative integer multiples of the frame time $\Delta t_{\mathrm{frame}}$ and

simultaneously fit to equation 6 by minimizing the sum of the squares of the residuals in MATLAB.

5.3.4.4 Probability distribution analysis

In probability distribution analysis (PDA), the observed FRET histogram is compared to a simulated FRET histogram, which depends on model parameters such as rate constants and state FRET efficiencies, as well as the exposure time of each frame. The model parameters are then optimized to achieve the best match between observed and simulated FRET histograms.

To simulate FRET histograms, we use code from Farooq and Hohlbein, who have described the method in detail. In brief, the FRET histogram of a 2-state system is modeled using 6 parameters: forward and reverse rate constants (k_{12} and k_{21}), the FRET efficiencies of the states (E_1 and E_2), and an excess FRET width for each state (σ_1 and σ_2) which accounts for FRET broadening due to contributions other than shot noise. Then, for each frame in the observed smFRET data, a simulated FRET value is generated as follows: first, the amount of time spent in each state is randomly drawn from a Monte Carlo simulation using the rate constants; second, the number of acceptor and donor photons from each state are drawn from binomial distributions, where the total number of simulated photons is equal to the number of measured photons in the observed frame, and the FRET value of each state is drawn from a normal distribution (e.g. the FRET probability density in state 1 is $N(E_1, \sigma_1^2)$); finally the FRET efficiency, $E_{FRET} = A/(A+D)$ is computed. Once all frames have been simulated, a FRET histogram is generated with bin size $\Delta E_{FRET} = 0.01$. The process is then repeated K times (here, $K = 8$) and the K simulated FRET histograms are averaged. A goodness-of-fit statistic is then calculated to compare the simulated and observed FRET histograms:

$$\chi_r^2 = \frac{1}{n-m}\sum_{i=1}^{n}\frac{[\mathrm{Freq}_{\mathrm{obs}}(E_i)-\mathrm{Freq}_{\mathrm{sim}}(E_i)]^2}{\mathrm{Freq}_{\mathrm{obs}}(E_i)} \qquad \text{(Eq. 5.7)}$$

where n is the number of nonzero bins, m is the number of model parameters, $\mathrm{Freq}_{\mathrm{obs}}(E_i)$ and $\mathrm{Freq}_{\mathrm{sim}}(E_i)$ are the frequency of FRET values in bin E_i for the observed and simulated data sets, respectively.

The best choice of model parameters (k_{12}, k_{21}, E_1, E_2, σ_1, σ_2) that will be that which minimizes χ_r^2. A common method of minimization is gradient descent; however, since the simulated FRET histogram is random, χ_r^2 is itself a random variable. Because of this, determining the gradient using finite-differences is prone to errors, and we found gradient descent to be challenging. Instead, we opted for a derivative-free grid-based approach, in which each parameter θ_i is sequentially optimized one at a time by sampling a set of points centered on the current best parameter value $\theta_{i,\mathrm{best}}$ (20 points evenly distributed from $\theta_{i,\mathrm{best}} - 50\%$ to $\theta_{i,\mathrm{best}} + 50\%$). If a parameter value is found that decreases the χ_r^2 statistic below the current lowest χ_r^2, then that parameter value becomes the new estimate $\theta_{i,\mathrm{best}}$. The process is then repeated until no parameter changes by more than 0.1% in an iteration, with a minimum of 12 iterations. We performed minimization on $K_{\mathrm{eq}} = k_{12} / k_{21}$ and $\tau = (k_{12} + k_{21})^{-1}$ rather than k_{12} and k_{21}, as we found working in this transformed parameter space to increase the rate of convergence. This optimization method is admittedly inefficient but circumvents the issues of deterministic gradient-descent approach; a promising alternative to improve efficiency is stochastic approximation gradient descent.

5.4 Simulation results

5.4.1 Fast rate constants are systematically underestimated in smFRET experiments

To establish the effects of bin time on smFRET-derived rate constants, we begin by analyzing our simulation (“synthetic”) single-molecule fluorescence trajectories. The primary

benefit of such an approach is that the actual simulation parameters are known precisely, allowing the computation of absolute and asymmetrical errors. Additionally, use of the simulations permit rapid testing of these kinetic analysis methods over the full parameter space, including modifying parameters not typically under experimental control, such as rate constants. There are numerous protocols for modeling single-molecule trajectories;[30-31] we opted for a photon-by-photon approach, in which both state-to-state transitions and photon detection events are treated as continuous-time processes, with the simulated photon arrivals binned by time/color into frames (Δt_{frame}) to generate discrete-time fluorescence trajectories of a typical CCD-based smFRET experiment. In doing so, we make no assumptions as to how photons are distributed per time bin, nor are requirements imposed that a molecule remain in a given state for the entire bin. Since camera averaging is a result of mapping from continuous-time to discrete-time processes, this photon-by-photon approach is ideal for capturing and modeling such artifacts.

We introduce our model for camera averaging by considering the two-state system in Figure 5.1a, which interconverts between two well-resolved FRET states (E_{low} = 0.2, E_{high} = 0.8) with forward rate k_F = 20.0 s^{-1} and reverse rate k_R = 10.0 s^{-1}, and achieves a total fluorescence signal-to-noise ratio (SNR = I/σ_I, where σ_I is the standard deviation in the brightness I) of 10:1 (100 photons per frame). Simulated trajectories of 200 s in length are binned with Δt_{frame} from 5 ms to 150 ms and are subjected to conventional dwell time analysis[39] to determine the apparent rate constants plotted in Figure 5.1b. Both k_F and k_R are accurately determined at short bin times but are systematically underestimated as these rate constants become comparable to the frame rate ($k_{frame} = 1/\Delta t_{frame}$). As Figure 5.1b clearly highlights, the magnitude of systematic rate constant errors smoothly increase in transition to the fast rate constant regime (k_F,$k_R \approx k_{frame}$).

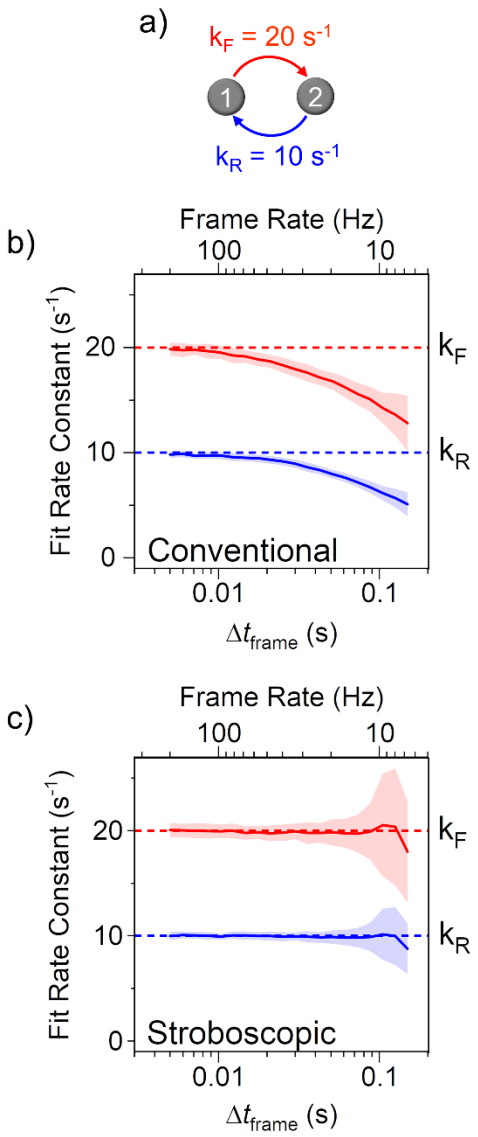

Figure 5.1 Systematic underestimation of rate constants in a model system. (a) Markov schematic of two-state system governed by rates k_F = 20.0 s^{-1} and k_R = 10.0 s^{-1}. Simulated trajectories are analyzed by dwell time analysis to extract rate constants as a function of camera frame length for (b) continuous illumination and (c) stroboscopic illumination (20% duty cycle). Shaded regions represent ±1σ uncertainty bands. Simulation conditions: 2000 s, E_{low} = 0.2, E_{high} = 0.8, 100 photons per frame (SNR = 10), repeated for each Δt_{frame} until standard error of the mean $\sigma_{SEM} = \sigma/\sqrt{N}$ reached 0.5%.

The obvious source of this systematic rate constant underestimation is due to errors in the transformation between continuous-time dynamics and discrete-time measurements. The true

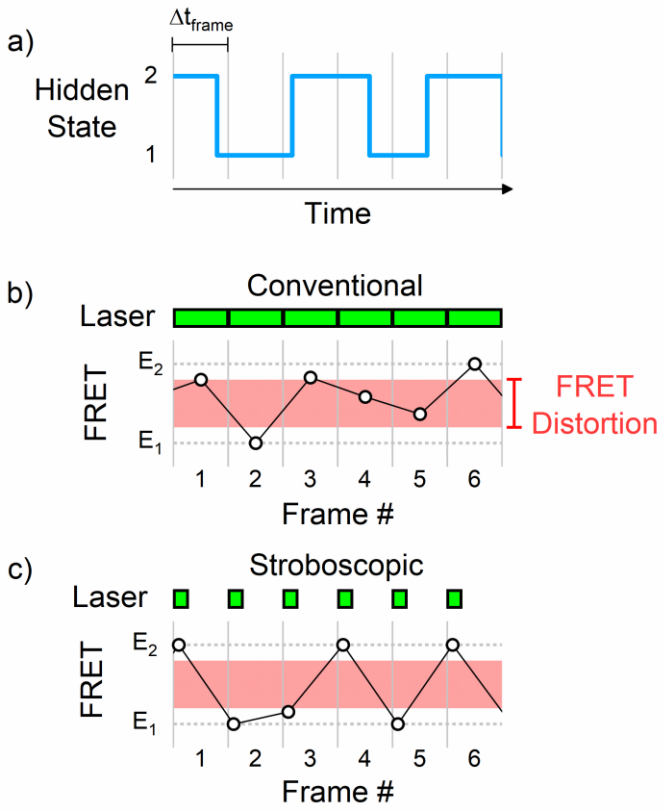

Figure 5.2 Mapping of continuous-time dynamics onto discrete-time results in blurred FRET states. (a) Sample true (hidden) state trajectory over six camera frames. (b) and (c) Ideal, noiseless FRET values averaged over time window during which laser is on (green blocks) for (b) continuous illumination and (c) stroboscopic illumination.

state of a system is a continuous function of time (Figure 5.2a), which is converted into a sequence of discrete frames by temporal binning (Figure 5.2b). The ideal, noiseless FRET value equals the time-weighted average of all FRET values visited in Δt_{frame}, with frames consisting of a single state *i* exhibiting a constant E_i FRET value. Conversely, frames for which the system has traversed multiple states in Δt_{frame} present an average FRET value contaminated by each of the individual FRET values. As a result of this "FRET averaging," [16-17] accurately assigning which state or combination of states the system occupies in each frame becomes difficult. However, FRET averaging is not a major concern as long as frames with more than one state occur infrequently. This will be true if the sampling frame rate is much faster than the rates of

interconversion between states ($-K_{ii} << k_{frame}$). Indeed, the measured rate constants for the synthetic data agree well with the extracted rate constants for sufficiently fast ($k_{frame} > 10\ k_F, k_R$) frame rates (Figure 5.1b). However, when the frame rate is within an order of magnitude of k_F or k_R, the probability of a frame containing multiple states results in many frames with incorrect state assignments. Since the theoretical basis of most analyses of single-molecule trajectories assumes single-state occupancy per frame, faster rate constants lead to a greater fraction of frames that are unmodeled. In addition to making accurate state assignment difficult, fast rate constants and/or slow frame rates result in fundamentally "missed dynamics." Short excursions to a state lasting less than one frame will only result in a slightly altered FRET efficiency, which will be misinterpreted as if no transition had occurred at all. The number of threshold-crossing transitions in a given smFRET trajectory will be lower than expected, resulting again in systematic underestimation of the rate constants. We note that simply discarding the shortest dwell times (< 1−2 frames), as is sometimes used to account for spurious short dwell times,[39] is insufficient to correct the underestimation of rate constants (Figure 5.3).

5.4.2 Stroboscopic illumination eliminates rate constant underestimation

To help resolve this issue, we propose an extremely simple solution which uses stroboscopic illumination to effectively address both FRET- averaging and missed dynamics as fundamental causes of rate constant underestimation. In a stroboscopic smFRET experiment, the excitation time window is compacted into a small fraction of the full frame. This reduction in duty cycle can be accomplished in a variety of ways, including modulating the excitation laser current, gating the CCD detection, or simply using an analog optical chopper wheel. Reducing the time per frame that the molecule is interrogated increases the probability that the molecule

occupies only one state for the duration of the observation pulse (Figure 5.2c). For example, in a two-state system with $k_{frame} \approx k_F, k_R$, the probability of remaining in a single state for an entire frame is on average only 1/e = 37%. The same system with a 10% duty cycle strobe will have a

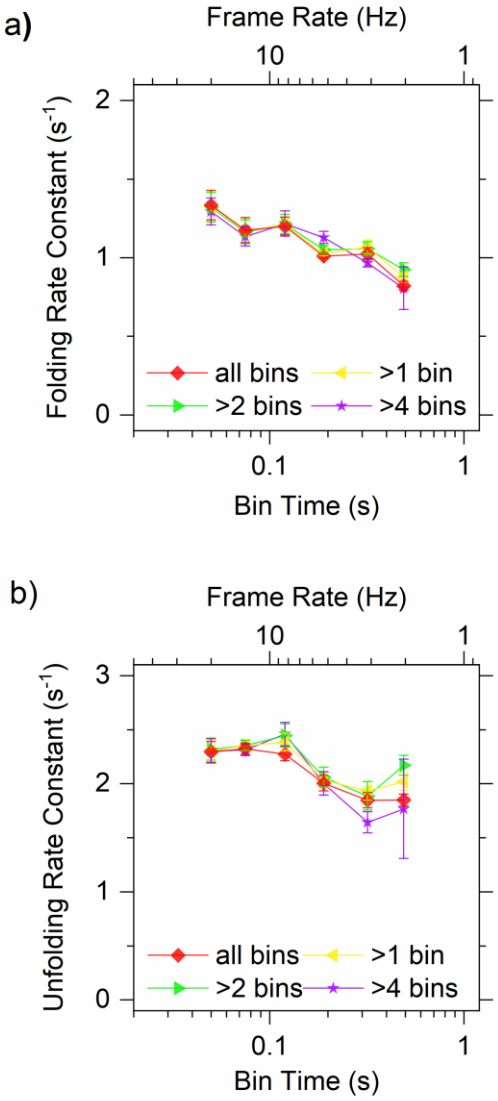

Figure 5.3 Experimental rate constants determined by dwell time distributions after discarding the shortest dwell times. Fitting the full dwell time distribution (red) produces similar dependence on choice of bin time as when fitting only dwell times lasting longer than 1 (yellow), 2 (green), or 4 (purple) frames. The analyzed data is the same non-stroboscopic data set as in Figure 5.8 in the main text.

survival probability $\exp(-k\Delta t_{\text{strobe}}) = 90\%$, which makes state assignment more accurate and lowers the probability of both FRET averaging and missed dynamics.

The experimental "cost" of implementing stroboscopic methods is that one is blind to the molecular state behavior during the increasing window duration when the illumination is off. We address this with the second component of the proposed technique, which involves analytical correction to the measured rate constants accounting for missed transitions. This correction incurs negligible computational cost and in fact becomes exact in the limit of vanishing strobe pulse width. In the interest of space, we briefly motivate and describe how to calculate the correction, with a complete derivation to be found in the Appendix. The first correction is specific to dwell time analysis, in which each state's survival function is fit to an exponential distribution to obtain an apparent rate constant k_i^{app}. For fast frame rates, k_i^{app} is approximately the total rate constant for leaving state i, but more generally k_i^{app} is related to the single-frame self-transition probability T_{ii} by

$$T_{ii} = e^{-k_i^{\text{app}} * \Delta t_{\text{frame}}} \qquad \text{(Eq. 5.8)}$$

One first computes the diagonal elements T_{ii} of the transition probability matrix **T**, from which the full matrix **T** can be constructed. This is particularly trivial for a two-state system, as each of the columns of **T** must sum to 1. For more complex systems, the empirical branching ratios are required, as addressed in the Appendix. The second step in the correction applies to any method that measures the transition probability matrix **T** (e.g., hidden Markov Modeling), in which the rate constant matrix **K** is determined from **T** by computing the matrix logarithm

$$\mathbf{K} = \frac{\ln(\mathbf{T})}{\Delta t_{\text{frame}}}. \qquad \text{(Eq. 5.9)}$$

which yields the desired corrected rate constants as matrix elements K_{ij}. Note that the common approximation $\mathbf{K} = (1/\Delta t_{\text{frame}})(\mathbf{T} - \mathbf{I})$ is only accurate in the slow rate constant regime ($-K_{ii}\Delta t_{\text{frame}}$

<<1), with the full matrix logarithm required when measuring rate constants comparable to the data acquisition rate. Fortunately, there are a number of computationally efficient algorithms available for performing this matrix logarithm.[40-41]

Combining stroboscopic illumination with the matrix logarithm correction abrogates the systematic underestimation of rate constants in our simulated smFRET data, as seen in Figure 5.1c. The forward and reverse rate constants determined by dwell time analysis under a 20% duty cycle strobe recapitulate the simulated values quite well and are independent of the choice of bin time. Neither the strobe nor analytical correction alone is sufficient to completely remove systematic error, with only the combination correctly addressing the underlying complications of state-averaging and missed dynamics. Additional random (nonsystematic) errors are also observed at large frame times (shaded uncertainty bands in Figure 5.1c; note that these uncertainties are standard deviations σ, whereas the standard error of the mean $\sigma_{SEM} = \sigma/\sqrt{N}$ can be decreased to arbitrary precision by repeating the measurement *N* times and averaging). In this regime, the average dwell time is shorter than a single frame, and the dwell time distributions cover only a few frames in time, which increases uncertainty in fitting. Said differently, the information content of the trajectory diminishes as frame-to-frame correlations become sufficiently weak. This loss of correlation generates an upper limit on the rate constants measurable with this combination of techniques for a given frame time, as will be empirically verified in Sec III.D.

5.4.3 Camera averaging artifacts cannot be resolved by modified experimental conditions

One is tempted to think it is possible to reduce or eliminate such underestimation of rate constants by control of experimental parameters (e.g., laser intensity). However, this is not the case. To explicitly address this possibility, we have performed data simulations over a wide

variety of "experimental" conditions, as summarized in Figure 5.4a-e and which reveal no impact on the rate constant deviations. By way of example, the results shown in Figure 5.4a represent simulations under identical conditions as in Figure 5.1, but with a 1:1 rather than 2:1 ratio of forward:reverse rate constants. Notice that the rate constants bias becomes significant (i.e., > 5%, as indicated by the yellow banded region) at $k\Delta t_{\text{frame}} \geq 0.2$, in agreement with previous observations.[15] As a second example, we can increase the signal to noise ratio (SNR) by 4x (Figure 5.4b), which improves the accuracy of state identification, but has little effect on avoiding rate constant underestimation error. This is because binning-induced broadening of FRET values is fundamentally not related to shot noise on the photon stream but rather on the state-transition dynamics. Thirdly, we can increase the total duration of the FRET trajectories (e.g., by obtaining data from more molecules), which reduces statistical noise in the measured rate constants, but once again has little effect on the systematic bias (Figure 5.4c). As a fourth example, we might hope to modify the 2-state FRET efficiencies to be better resolved (Figure 5.4d), which could be experimentally achieved by redesigning the single-molecule construct to optimize placement of the fluorophores. However, an increase in ΔE_{FRET} from 0.6 to 1.0 leaves the rate constant underestimation errors remarkably similar. Finally, increasing the equilibrium constant from K_{eq} = 1 to 4 (in Figure 5.4e by reducing the reverse rate k_R by 4x) leads to only small differences in the errors for k_F and k_R, with the average magnitude of the bias unchanged. In summary, systematic underestimation of fast rate constants by camera averaging and missing dynamics proves to be remarkably insensitive to the choice of parameters potentially under experimental control.

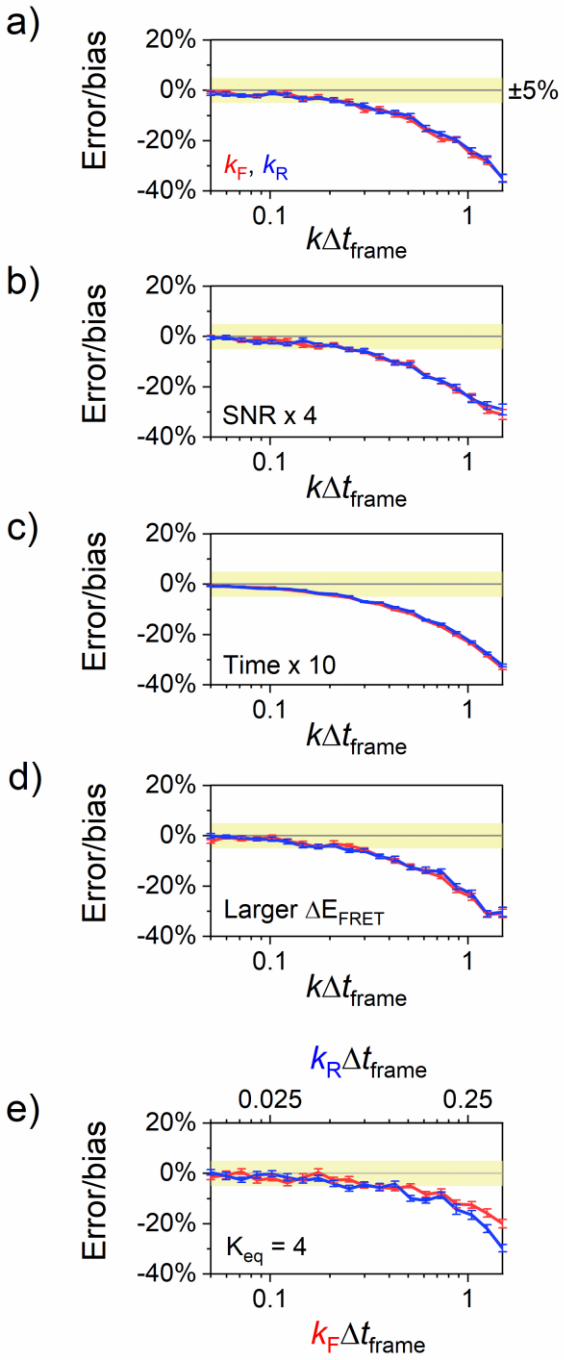

Figure 5.4 Camera averaging artifacts are robust to experimental parameters. (a) Bias as percent error in rate constants for simulated two-state system with equal forward and backward rate constants. Simulation conditions: 2000 s, E_{low} = 0.2, E_{high} = 0.8, k_F = k_R, 100 photons per frame (SNR = 10), repeated 24 times to determine uncertainties (standard error of the mean). For reference, the yellow region demarks a ±5% bias. (b–e) Same as (a) but simulation conditions are modified as follows: (b) 4x greater signal to noise ratio, (c) 10x more simulated time, (d) increased FRET separation from ΔE_{FRET} = E_{high} – E_{low} = 0.6 to 1.0, and (e) increased equilibrium constant from K_{eq} = 1 to 4 by decreasing k_R.

5.4.4 Optimal excitation duty cycle and maximum measurable rate constants

We have shown that introducing a low duty cycle (i.e., stroboscopic) light source can help eliminate binning-related artifacts in smFRET studies, but this obviously can result in loss of signal, signal to noise ratio, and reduction in performance. To compensate for loss of signal, the experimentalist may wish to increase laser power, but this also has obvious limits due to photophysics (e.g., photobleaching, triplet state formation) and nonlinearity in fluorophore brightness when the time interval between photoexcitation events approaches the fluorescence lifetime. Therefore, it is a useful exercise to identify the largest duty cycle that still reduces systematic errors to acceptable levels. To determine this optimal strobe duty cycle, we have simulated smFRET data over a range of duty cycles, rate constants, and frame acquisition rates (Figures 5.5 and 5.6). For a fixed $k\Delta t_{\text{frame}}$ (Figure 5.5a), decrease in strobe duty cycle monotonically reduces systematic error. However, under real world experimental conditions, single-molecule rate constants might have typical statistical uncertainties of several percent or larger; hence, reduction of these systematic errors to < 3% is usually unwarranted. For rate constants equal to the frame rate ($k\Delta t_{\text{frame}} = 1$), we observe in Figure 5.5a that a strobe duty cycle of 10–20% already has reduced systematic error in the extracted rates to < 3%. Obviously for rate constants lower than the frame rate ($k\Delta t_{\text{frame}} < 1$), one achieves this 3% error limit target even more quickly with reduction in duty cycle. Alternatively summarized, below a 10% duty cycle, one is discarding signal for little gain in extracted rate constant accuracy.

The corresponding upper limit in the measured rate constants is determined by the data acquisition rate. Trajectory-based analyses of rate constants rely on frame-to-frame correlations, and these methods fail when kinetic relaxation occurs on the timescale of the frame rate. Indeed, this mode of failure is readily observable in the shaded uncertainty regions in the right-hand side

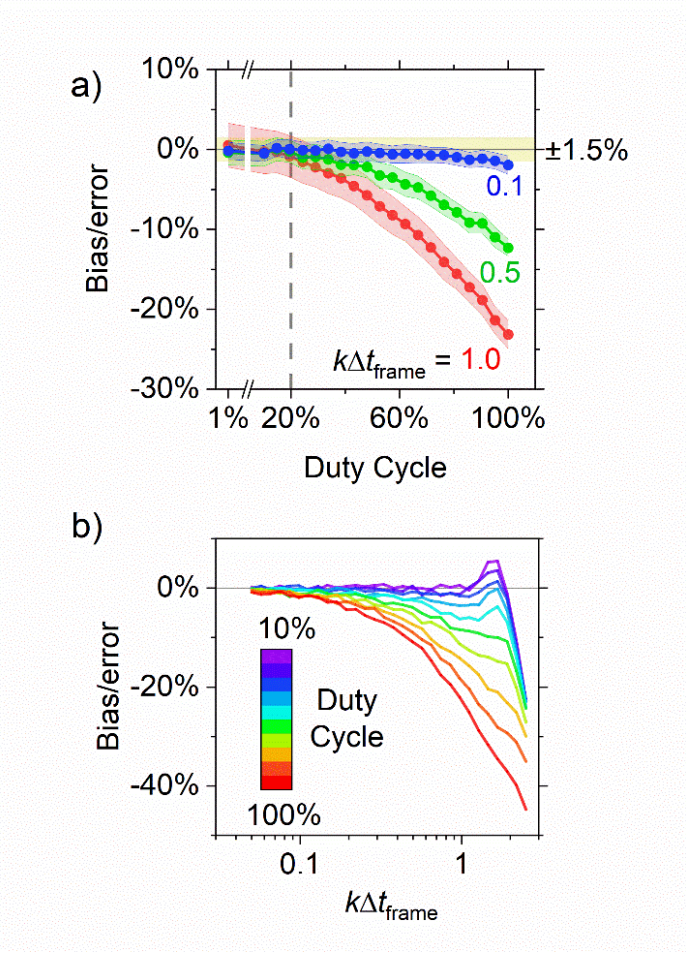

Figure 5.5 Determining the optimal strobe duty cycle. (a) Rate constant bias as a function of duty cycle, for fixed $k\Delta t_{frame}$ (1.0, red; 0.5, green; 0.1, blue). Duty cycle of 100% corresponds to no strobe. Shaded bands indicate uncertainties (±1σ). Bias band of ±1.5% highlighted in yellow, and recommended duty cycle (< 20%) indicated by vertical dashed line. (b) Rate constant bias at fixed duty cycles. The quality of results decreases for $k\Delta t_{frame} > 1$. Simulation conditions: 8000 s, $k_F = k_R$, $E_{low} = 0.2$, $E_{high} = 0.8$, 100 photons per frame (SNR = 10). For 2D heatmaps of bias and uncertainty for combinations of duty cycle and framerate, see Figure 5.6.

of Figure 5.1c, for which even the stroboscopic data produce results with high levels of statistical noise when the rate constants exceed ≈ 1.5 k_{frame}. Simply stated, this is due to low information content per frame when the system decorrelation is fast. In this limit, individual data points in the FRET trajectory therefore become uncorrelated and can provide only non-dynamical information (e.g., equilibrium constants). This same behavior can also be noted in Figure 5.5b, where even

for low duty cycle data acquisition the quality of results decays quickly at k_F,k_R > k_{frame}. In summary, stroboscopic smFRET methods can help rescue rate constants up to the frame rate, but faster rate constants require alternative approaches, such as stroboscopic probability distribution analysis discussed below.[29]

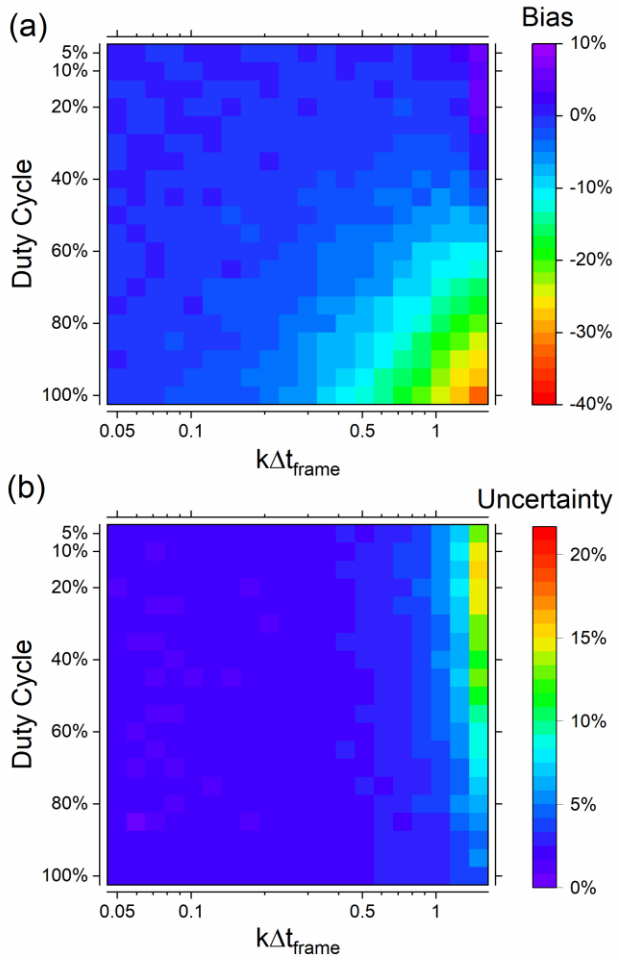

Figure 5.6 (a) Bias and (b) uncertainty (standard deviation) of rate constants determined by dwell time analysis for a variety of rate constants and stroboscopic duty cycles, where a duty cycle of 100% is equivalent to no strobe. Of interest are the regions of bias near zero (blue in (a)) and uncertainty near zero (blue in (b)). Uncertainties for all simulations increase rapidly for $k\Delta t > 1$. Simulation conditions: 2000 s, 100 photons per bin (SNR = 10), $k_F = k_R$, $E_{low} = 0.2$, $E_{high} = 0.8$, repeated 48 times to determine uncertainty.

5.5 Experimental tests

To demonstrate that the above results and predictions based on simulated data are experimentally valid, we have performed a series of smFRET experiments measuring kinetic rate constants with and without stroboscopic illumination (Figures 5.7 and 5.8). Specifically, we choose to examine a very simple single-molecule construct consisting of a DNA hairpin with a 7

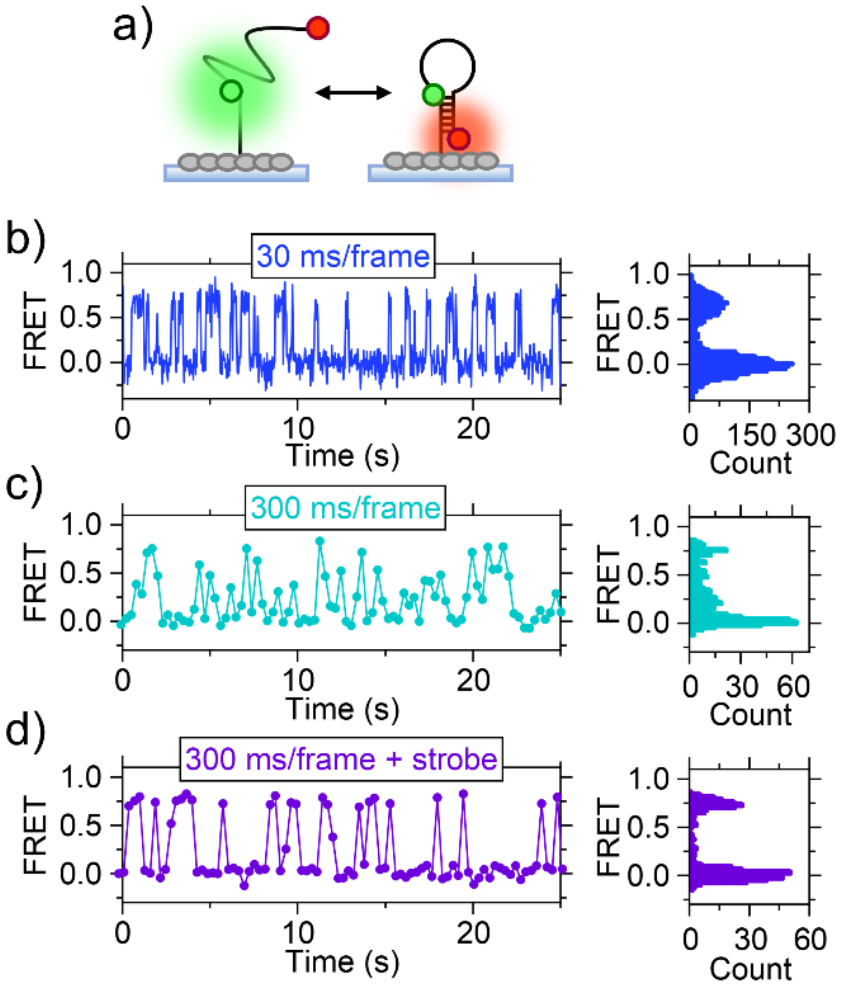

Figure 5.7 Experimental evaluation of stroboscopic smFRET. (a) Cartoon depiction of surface-tethered, two-state DNA hairpin investigated in these studies. (b) Sample smFRET trajectory (left) taken at a fast frame rate (30 ms frame^{-1}) compared to underlying dynamics. FRET histogram (right) exhibits two well-resolved peaks. (c) Same as (b), but for a slow frame rate (300 ms frame^{-1}). Many frames contain intermediate FRET values due to camera averaging. (d) Same as (c) but using stroboscopic illumination at 20% duty cycle. Data are obtained at room temperature in 40 mM K^+ and 30 mM Na^+.

base-pair stem and 40-adenine loop which we have studied previously (Figure 5.7a).[33] This construct exhibits well-behaved two-state single-exponential kinetics (with $k_{fold} = 1.2\ s^{-1}$, $k_{unfold} = 2.2\ s^{-1}$) and therefore represents a useful model system with which to test these stroboscopic analysis methods. The data collection is performed on a total internal reflection fluorescence (TIRF)[37] microscope apparatus with charge-coupled device (CCD) camera detection, and the incident laser power is increased or decreased to maintain a constant signal-to-noise ratio (SNR) ≈ 7 at all acquisition rates. As shown in Figure 5.7b at fast frame rate, smFRET trajectories under continuous illumination exhibit switching between two states, with the integrated FRET histogram (right panel) clearly depicting two distinct FRET peaks. On the other hand, continuous illumination at this slower frame rate (Figure 5.7c) yields a notably smeared smFRET histogram, with spurious density appearing between the two FRET states. By way of contrast, however, the use of a 10% duty cycle strobe at this same slow frame rate (Figure 5.7d) removes camera averaging artifacts from the FRET histogram (far right), therefore restoring distinct two-state behavior.

Furthermore, the rate constants from the smFRET trajectories were measured by dwell time analysis over a systematic range of data acquisition bandwidths (40 Hz to 1.3 Hz, Figure 5.8). Without stroboscopic illumination, the extracted rate constants exhibit the behavior characteristic of our simulated results, i.e., with the measured rate constants (filled squares) systematically

dependent on frame rate and underestimated with increasing Δt_{frame}. Inclusion of a 10% duty cycle strobe (open diamonds) completely removes this dependence on frame rate, in agreement with our simulations. Indeed, the kinetic measurements remain faithful even up to $k\Delta t_{frame} = 1.7$, which agrees with our upper limit of $k\Delta t_{frame} \approx 1.5$ predicted above.

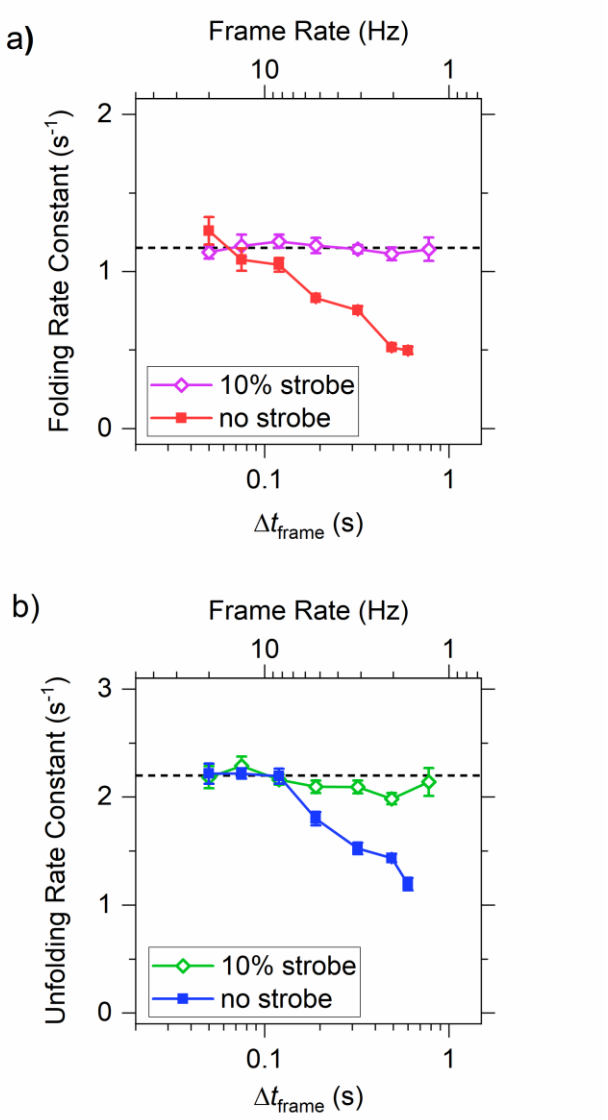

Figure 5.8 Experimental validation of stroboscopic method by measuring DNA hairpin folding dynamics over a range of camera frame rates. (a) Folding rate constants. Without stroboscopic illumination (red), the measured folding rate constant is a function of frame rate. Application of a 10% duty cycle strobe (purple) removes the dependence. (b) Same as (a), but for the unfolding rate constants. Each data point represents an analysis of 2000-3000 observed dwell times from ~50 molecules. Error bars are standard errors of the mean as determined by bootstrapping.

5.6 Discussion

smFRET has proven itself to be an invaluable technique for measuring the detailed dynamics of biological processes.[13-14] However, this work demonstrates that rate constants

determined in such smFRET experiments can be prone to systematic underestimation, particularly when the conformational state-to-state transition rates are comparable to the data acquisition rate ($k\Delta t_{\text{frame}} \gtrapprox 0.1$). Herein, we have presented an extremely simple "stroboscopic" method for eliminating such "camera averaging" artifacts by restricting fluorescence collection to a reduced but contiguous portion of each time bin, for instance, by gating the excitation light source. We have demonstrated that the use of stroboscopic data collection, together with a simple mathematical correction, can accurately recover rate constant information up to the data acquisition rate, validating these methods on both synthetic (Figure 5.1) and experimental (Figure 5.8) data. As a result, stroboscopic smFRET methods in principle extend the upper limit of measurable rate constants by up to an order of magnitude over that of conventional smFRET without resorting to reducing the instrument throughput by cropping the field of view to increase frame rates.

We can extend the use of stroboscopic smFRET one step further. The issue of systematic rate constant underestimation due to time binning is ubiquitous and generates similar constraints for more sophisticated smFRET trajectory analysis methods such as hidden Markov modeling (HMM)[39, 42-44] and time correlation function (TCF) fitting.[45-46] Consequently, each of these and other analysis methods might also benefit from stroboscopic data collection. To explore this in more detail, we have analyzed simulated smFRET trajectories using each of three methods i) dwell time analysis, ii) HMM, and iii) TCF fitting, under both continuous and stroboscopic illumination conditions (Figure 5.9). As clearly evident in Figure 5.9b–c (filled symbols), neither HMM nor TCF fitting escapes this fundamental issue of underestimating fast rate constants. This is quite simply because both methods make the assumption, as in dwell time analysis, that the system occupies a single state throughout each observation point. This assumption becomes

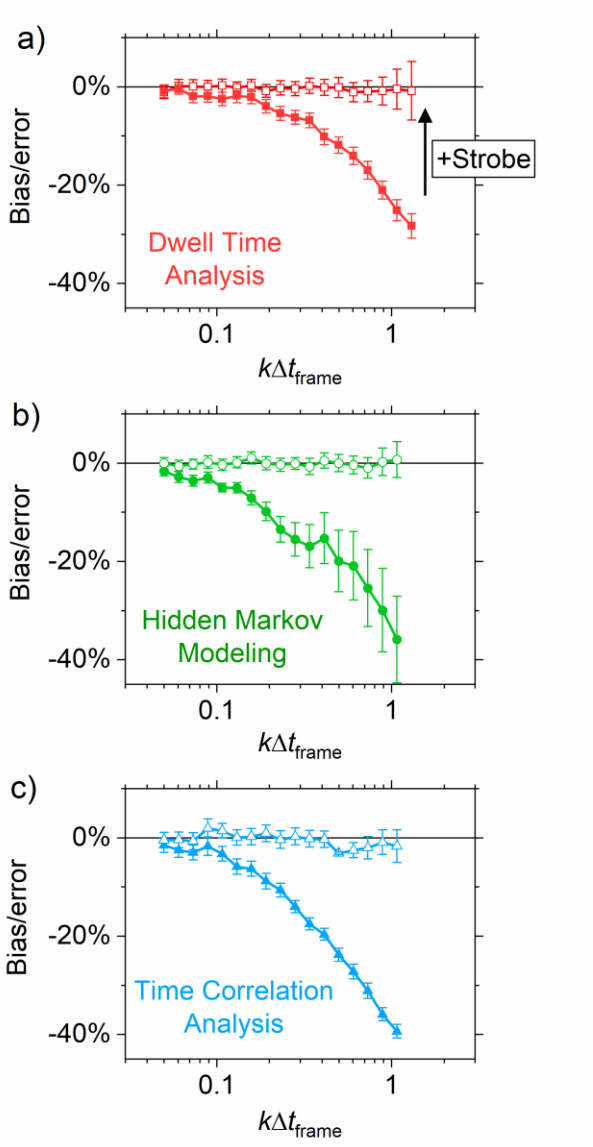

Figure 5.9 Systematic underestimation of rate constants is a shared feature of multiple smFRET analysis methods. Simulated conditions are the same as in Figure 5.4a. (a) Dwell time analysis. (b) Hidden Markov modeling (HMM). (c) Time correlation function (TCF) fitting. Each analysis is performed on data simulated under full-frame illumination (solid symbols) as well as 25% duty cycle stroboscopic illumination (open symbols). Error bars are ±1σ.

more nearly correct when fractional duty cycle "stroboscopic" illumination is applied, with all methods accurately estimate the rate constant for a 10% duty cycle excitation (Figure 5.9, open symbols). Consequently, the data in Figure 5.9 demonstrate that stroboscopic smFRET improves the accuracy of rate constant determination by other analysis methods besides dwell time

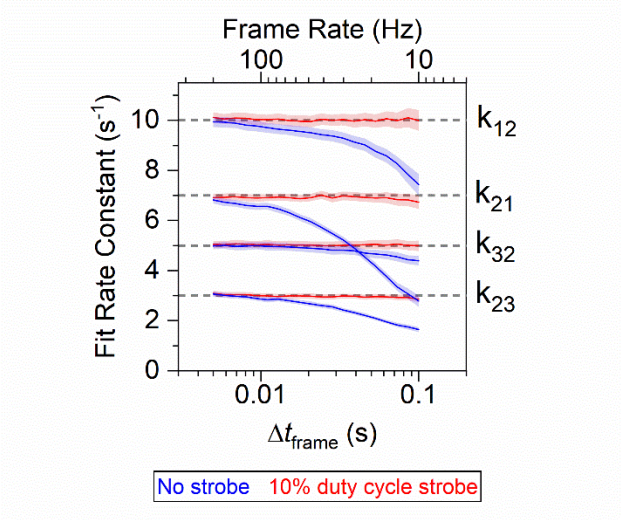

Figure 5.10 Rate constants determined by hidden Markov modeling of simulated 3-state system ($1 \leftrightarrow 2 \leftrightarrow 3$), with (red) and without (blue) stroboscopic excitation. Shaded regions are 1-σ uncertainties. Under continuous illumination, rate constants are systematically underestimated at slow frame rates (large Δt_{frame}), whereas stroboscopic data yields accurate rate constants across all frame rates. Simulation conditions: 2000 s, 300 photons per bin, $k_{12} = 10\ s^{-1}$, $k_{21} = 7\ s^{-1}$, $k_{23} = 3\ s^{-1}$, $k_{32} = 5\ s^{-1}$, $E_1 = 0.2$, $E_2 = 0.5$, $E_3 = 0.8$, repeated 192 times to determine uncertainties.

analysis. In addition, the use of HMM enables an especially facile extension of the stroboscopic method to systems of more than 2 states, as demonstrated in Figure 5.10.

Stroboscopic smFRET requires greater laser power than conventional smFRET to obtain the same signal level, which can potentially exacerbate issues of fluorophore saturation and photobleaching. To examine this effect, we collected stroboscopic data with a 20% duty cycle at 100 frames per second (Figure 5.11), which is approximately the full-frame capture rate of current CCD technology. Working at our maximum available laser fluence (≈50 mW over a 20 μm diameter TIR spot), we observed a 50% reduction in signal-to-noise ratio compared to non-stroboscopic data at the same conditions (SNR = 3 vs. 6; Figure 5.11a). Despite the loss of SNR, the quality of the data is sufficient for dwell time analysis, and the extracted rate constants are equivalent to those obtained without stroboscopic illumination (Figure 5.11b–c). These results

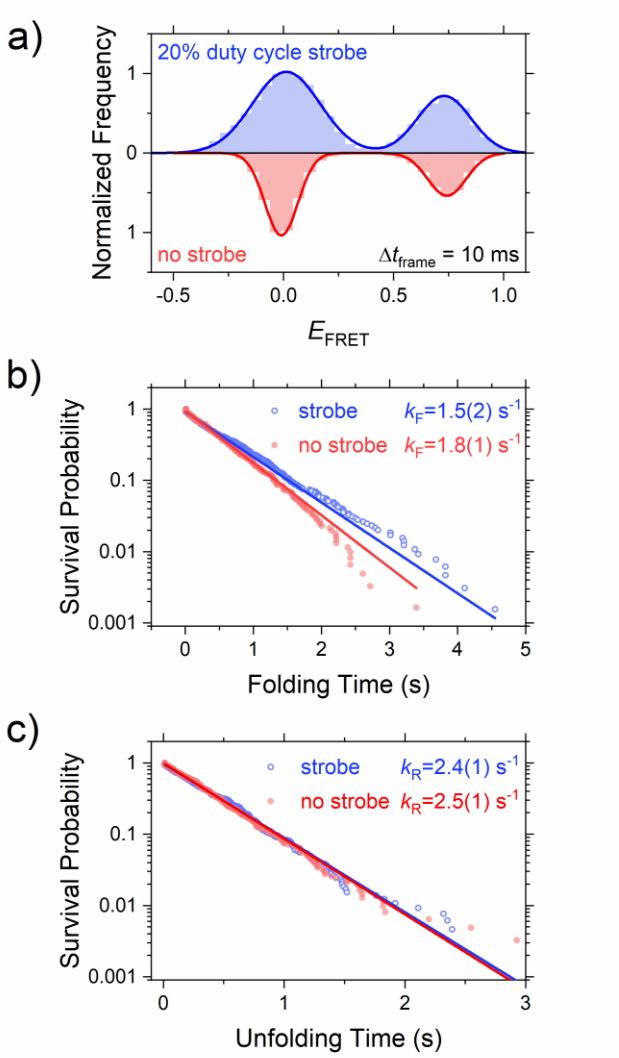

Figure 5.11 Experimental measurement of DNA hairpin folding at Δt_{frame} = 10 ms (100 Hz frame rate). (a) Normalized FRET histogram of data acquired at the same laser power (50 mW) with 20% duty cycle strobe (blue) and without strobe (red). Though the FRET widths are larger for the strobe data due to the short exposure time ($\Delta t_{exposure}$ = 2 ms), the FRET peaks are resolved sufficiently to permit dwell time analysis, shown in (b) and (c) for folding and unfolding times, respectively. The extracted rate constant for folding (k_F) and unfolding (k_U) are statistically equivalent for the case of data acquired with strobe (blue) and without strobe (red). Experimental conditions: 75 mM total M^+ (K^+ and Na^+), 20.0 °C.

show that, with the help of oxygen removal[35] and triplet quenching,[36] fluorophore photophysics do not impede stroboscopic smFRET operation at acquisition rates up to 100 Hz.

This article builds off the work of Farooq and Hohlbein[29] who first demonstrated the use of stroboscopic smFRET to measure fast rate constants. However, instead of analyzing stroboscopic data by trajectory-based analyses, as in these studies, Farooq and Hohlbein used probability distribution analysis (PDA).[22-24] In PDA, data points are binned into a FRET histogram, which is then fit to a model that includes kinetic parameters. Therefore, PDA treats each observed FRET value as an independent measurement. In contrast, trajectory-based analysis (e.g., dwell time analysis, HMM, TCF fitting) make use of the correlation between data points, which has the potential advantage of utilizing a greater portion of the information content of the data. Therefore, intuitively, the trajectory-based method presented in this paper should yield rate constants with smaller uncertainties than those determined by PDA. Indeed, when simulated data are analyzed by both PDA and dwell time analysis, the results from dwell time analysis have ≈50% smaller uncertainties (Figure 5.12). This reduction in uncertainty may make the present method more attractive to some smFRET researchers, especially those who already rely on trajectory-based analysis and do not wish to switch to PDA, and entirely different form of analysis. However, as shown by Farooq and Hohlbein, the combination of stroboscopic smFRET and PDA has in principle access to faster dynamics than stroboscopic trajectory analysis, as stroboscopic PDA is limited by the exposure time rather than the frame duration. Due to this combination of factors, we consider the trajectory-based approach and the PDA-based approach to be complementary methods for analyzing stroboscopic smFRET data.

We emphasize that stroboscopic smFRET is quite straightforward to incorporate into any existing smFRET experiment, with stroboscopic illumination implementable at relatively low cost. Furthermore, stroboscopic data does not require adopting new analysis methods, as it is fully compatible with conventional Dwell Time, Hidden Markov Model (HMM), and Time

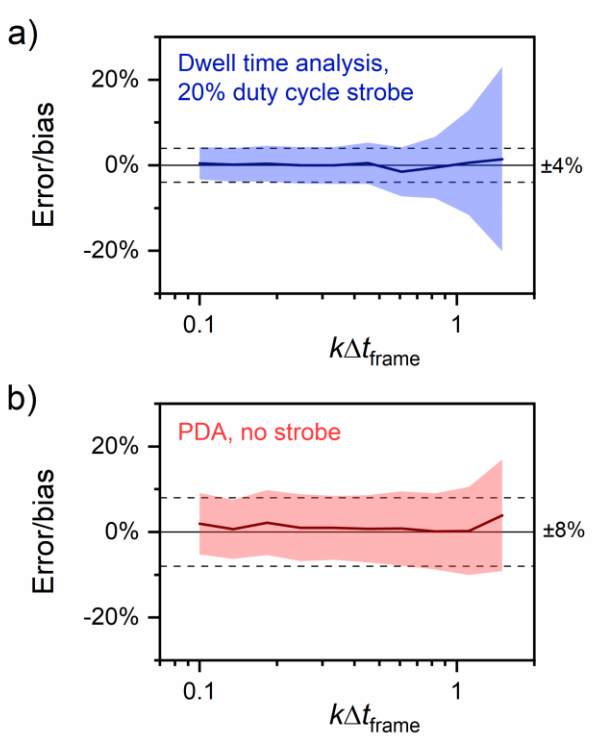

Figure 5.12 Analysis of simulated data by trajectory-based approach (dwell time analysis) and FRET histogram fitting (PDA). Under the simulated conditions (2000 s, 100 photons per bin, $k_F = k_R$, $E_{low} = 0.2$, $E_{high} = 0.8$), dwell time analysis (a) yields accurate rate constants up to the frame rate with an uncertainty (shaded band; standard deviation) of ≈4% which grows rapidly beyond $k\Delta t_{frame} \approx 1$. (b) PDA accurately determines rate constants, with uncertainties of ≈8% that are largely independent of $k\Delta t_{frame}$.

Correlation Function (TCF) analysis. Therefore, any smFRET laboratory can readily incorporate stroboscopic methods while largely maintaining the current experimental arrangement and analysis. In the interest of completeness and kinetic rigor, we recommend reporting frame acquisition rates along with published rate constants, as well as closely monitoring the ratio of smFRET-measured rate constants to this acquisition rate. If any rate constants exceed 10–20% of the frame rate, then FRET averaging is a significant concern, for which application of stroboscopic methods offers a reduction in rate constant systematic errors down to < 3% level. Finally, while this paper has focused on using smFRET to determine conformational dynamics,

such a stroboscopic approach should also be extendable to any experimental method based on a fluorescence measurement of discrete state sampling, such as transitions between diffusional states in single molecule diffusion studies[7] or binding processes studied by protein induced fluorescence enhancement (PIFE).[47]

5.7 Conclusion

We have developed, tested, and presented a stroboscopic solution to the underestimation of rate constants in time-binned smFRET experiments. Our work demonstrates that kinetic rate constants are significantly *underestimated* when comparable to the sampling rate upper limit, with > 5% systematic errors in the measured rate constant if within 10% of the ultimate frame acquisition limit. These deviations cannot be avoided by modifying experimental parameters such as light intensity or duration of trajectory. The core cause of the artifact is the mapping of continuous time dynamics onto a discrete time domain of binned data, resulting in blurring of FRET states and missed dynamics. These fundamental issues can be addressed through the combined use of i) stroboscopic illumination and ii) analytic mathematical correction to the rate constants. Stroboscopic smFRET is thereby capable of measuring rate constants up to at least the data acquisition rate, which for a 5% threshold of measurement accuracy amounts to a tenfold increase in dynamic range. This allows one to measure single molecule rate constants up to the frame acquisition rate limit without compromising any other multiplexing capacity of the measurement.

5.8 Acknowledgements

We would like to thank Johannes Hohlbein for providing code for PDA calculations.

5.9 References

1. Aleman, E. A.; Lamichhane, R.; Rueda, D., Exploring RNA folding one molecule at a time. *Curr. Opin. Chem. Biol.* **2008,** *12* (6), 647-654.

2. Buchner, G. S.; Murphy, R. D.; Buchete, N.-V.; Kubelka, J., Dynamics of protein folding: Probing the kinetic network of folding-unfolding transitions with experiment and theory. *Biochim. Biophys. Acta, Proteins Proteomics* **2011,** *1814* (8), 1001-1020.

3. Grima, R.; Walter, N. G.; Schnell, S., Single-molecule enzymology à la Michaelis-Menten. *FEBS J.* **2014,** *281* (2), 518-530.

4. van Oijen, A. M., Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Curr. Opin. Biotechnol.* **2011,** *22* (1), 75-80.

5. Stratmann, S. A.; van, O. A. M., DNA replication at the single-molecule level. *Chem. Soc. Rev.* **2014,** *43* (4), 1201-20.

6. Friedman, L. J.; Gelles, J., Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods* **2015,** *86*, 27-36.

7. Elf, J.; Barkefors, I., Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* **2019,** *88* (1), 636-659.

8. Barkai, E.; Garini, Y.; Metzler, R., Strange kinetics of single molecules in living cells. *Phys. Today* **2012,** *65* (8), 29-35.

9. Goryaynov, A.; Ma, J.; Yang, W., Single-molecule studies of nucleocytoplasmic transport: from one dimension to three dimensions. *Integr. Biol.* **2012,** *4* (1), 10-21.

10. Guan, J.; Jia, C.; Li, Y.; Liu, Z.; Wang, J.; Yang, Z.; Gu, C.; Su, D.; Houk, K. N.; Zhang, D., et al., Direct single-molecule dynamic detection of chemical reactions. *Sci. Adv.* **2018,** *4* (2), eaar2177.

11. Tachikawa, T.; Majima, T., Single-molecule, single-particle approaches for exploring the structure and kinetics of nanocatalysts. *Langmuir* **2012,** *28* (24), 8933-8943.

12. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* **2008,** *5* (6), 507-16.

13. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Foerster resonance energy transfer. *Science* **2018,** *359* (6373), 288.

14. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **2015,** *34*, 52-59.

15. Kinz-Thompson, C. D.; Bailey, N. A.; Gonzalez, R. L., Precisely and accurately inferring single-molecule rate constants. In *Methods in Enzymology*, Spies, M.; Chemla, Y. R., Eds. Academic Press: 2016; Vol. 581, pp 187-225.

16. Bronson, J. E.; Fei, J.; Hofman, J. M.; Gonzalez, R. L., Jr.; Wiggins, C. H., Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **2009,** *97* (12), 3196-205.

17. Hadzic, M.; Borner, R.; Konig, S. L. B.; Kowerko, D.; Sigel, R. K. O., Reliable state identification and state transition detection in fluorescence intensity-based single-molecule Forster resonance energy-transfer data. *J. Phys. Chem. B* **2018,** *122* (23), 6134-6147.

18. Tang, J.; Sun, Y.; Pang, S.; Han, K. Y., Spatially encoded fast single-molecule fluorescence spectroscopy with full field-of-view. *Sci. Rep.* **2017,** *7* (1), 10945.

19. Corle, T. R.; Kino, G. S., *Confocal scanning optical microscopy and related imaging systems*. Academic Press: San Diego, 1996.

20. Tan, Y. W.; Hanson, J. A.; Chu, J. W.; Yang, H., Confocal single-molecule FRET for protein conformational dynamics. *Methods Mol. Biol.* **2014,** *1084*, 51-62.

21. Gopich, I.; Szabo, A., Theory of photon statistics in single-molecule Forster resonance energy transfer. *J. Chem. Phys.* **2005,** *122* (1), 14707.

22. Schrangl, L.; Göhring, J.; Schütz, G. J., Kinetic analysis of single molecule FRET transitions without trajectories. *J. Chem. Phys.* **2018,** *148* (12), 123328.

23. Santoso, Y.; Torella, J. P.; Kapanidis, A. N., Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem* **2010,** *11* (10), 2209-19.

24. Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A., Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B* **2010,** *114* (23), 7983-95.

25. Chen, J.; Pyle, J. R.; Sy Piecco, K. W.; Kolomeisky, A. B.; Landes, C. F., A two-step method for smFRET data analysis. *J. Phys. Chem. B* **2016,** *120* (29), 7128-32.

26. Elf, J.; Li, G.-W.; Xie, X. S., Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **2007,** *316* (5828), 1191.

27. Hansen, A. S.; Woringer, M.; Grimm, J. B.; Lavis, L. D.; Tjian, R.; Darzacq, X., Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **2018,** *7*, e33125.

28. Fontana, M.; Fijen, C.; Lemay, S. G.; Mathwig, K.; Hohlbein, J., High-throughput, non-equilibrium studies of single biomolecules using glass-made nanofluidic devices. *Lab Chip* **2018,** *19* (1), 79-86.

29. Farooq, S.; Hohlbein, J., Camera-based single-molecule FRET detection with improved time resolution. *Phys. Chem. Chem. Phys.* **2015,** *17* (41), 27862-72.

30. Gopich, I. V.; Szabo, A., Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B* **2009,** *113* (31), 10965-73.

31. Borner, R.; Kowerko, D.; Hadzic, M.; Konig, S. L. B.; Ritter, M.; Sigel, R. K. O., Simulations of camera-based single-molecule fluorescence experiments. *PLoS One* **2018,** *13* (4), e0195277.

32. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B* **2020,** *124* (51), 11561-11572.

33. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: kinetic insights from single-molecule studies. *J. Phys. Chem. B* **2018,** *122* (43), 9869-9876.

34. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on Chips II*, Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

35. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **2008,** *94* (5), 1826-35.

36. Ha, T.; Tinnefeld, P., Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* **2012,** *63* (1), 595-617.

37. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

38. Stigler, J.; Rief, M., Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *ChemPhysChem* **2012,** *13* (4), 1079-1086.

39. Blanco, M.; Walter, N. G., Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol.* **2010,** *472*, 153-178.

40. Al-Mohy, A. H.; Higham, N. J., Improved inverse scaling and squaring algorithms for the matrix logarithm. *SIAM J. Sci. Comput.* **2012,** *34* (4), C153-C169.

41. Cheng, S. H.; Higham, N. J.; Kenney, C. S.; Laub, A. J., Approximating the logarithm of a matrix to specified accuracy. *SIAM J. Matrix Anal. Appl.* **2001,** *22* (4), 1112-1125.

42. Chodera, J. D.; Noe, F., Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25*, 135-144.

43. Lee, T. H., Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden markov models. *J. Phys. Chem. B* **2009,** *113* (33), 11535-42.

44. McKinney, S. A.; Joo, C.; Ha, T., Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **2006,** *91* (5), 1941-51.

45. Lee, T. H.; Lapidus, L. J.; Zhao, W.; Travers, K. J.; Herschlag, D.; Chu, S., Measuring the folding transition time of single RNA molecules. *Biophys. J.* **2007,** *92* (9), 3275-83.

46. Lu, H. P., Sizing up single-molecule enzymatic conformational dynamics. *Chem. Soc. Rev.* **2014,** *43* (4), 1118-43.

47. Lerner, E.; Ploetz, E.; Hohlbein, J.; Cordes, T.; Weiss, S., A quantitative theoretical framework for protein-induced fluorescence enhancement-Forster-type resonance energy transfer (PIFE-FRET). *J. Phys. Chem. B* **2016,** *120* (26), 6401-10.

Chapter 6

Measuring Excess Heat Capacities of DNA Folding at the Single Molecule Level*

6.1 Abstract

Measurements of the thermodynamic properties of biomolecular folding ($\Delta G°$, $\Delta H°$, $\Delta S°$, etc.) provide a wealth of information on the folding process and have long played a central role in biophysical investigations. In particular, the excess heat capacity of folding (ΔC_P) is particularly crucial, as typically measured in bulk ensemble studies by differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). Here, we report the first measurements of ΔC_P at the single molecule level using single-molecule fluorescence resonance energy transfer (smFRET), as well as the very first measurements of the heat capacity change associated with achieving the transition state ($\Delta C^{\ddagger}_P$) for nucleic acid folding. The DNA hairpin used in these studies exhibits an excess heat capacity for hybridization ($\Delta C_P = -340 \pm 60$ J/mol/K per base pair) consistent with the range of literature expectations ($\Delta C_P = -100$ to -420 J/mol/K per base pair). Furthermore, the measured activation heat capacities ($\Delta C^{\ddagger}_P$) for such hairpin unfolding are consistent with a folding transition state containing few fully formed base pairs, in agreement with prevailing models of DNA hybridization.

6.2 Introduction

The thermodynamics of nucleic acid folding/hybridization are of primary interest to the biophysics community.[1-4] It is widely understood that the temperature stability profile of a DNA duplex requires knowledge of the folding enthalpy $\Delta H°$ and entropy $\Delta S°$, as, for instance, often

---

* This chapter is adapted from: Nicholson, D. A.; Jia, B.; Nesbitt, D. J. *J. Phys. Chem. B* **2021**, *125*, 34, 9719–9726.

predicted from nearest-neighbor models.[5-6] Less well investigated are *heat capacities* and in particular heat capacity *differences* between hybridized and unhybridized states, the "excess" heat capacity (ΔC_P) for a given conformational transition,[2, 7-13] which directly impacts any predictions for temperature-dependence in $\Delta H°$ and $\Delta S°$. As a result, the neglect of such excess heat capacity effects results in erroneous extrapolation of nucleic acid folding to conditions other than the reference temperature at which thermodynamic parameters have been determined. Such errors can arise, for example, in the prediction of polymerase chain reaction (PCR) primer stabilities, due to requisite operation at much higher than body temperatures.[7] As a second example, attempts to unify disparate sets of nearest neighbor model parameters for nucleic acid stabilities have proven challenging due to the need for comparison/extrapolation between different temperature conditions.[13] Even beyond such technological concerns, a knowledge of excess heat capacities offers first insights into the microscopic dynamics of solvent restructuring during the folding/hybridization event that are inaccessible from measurements of $\Delta H°$ and $\Delta S°$ alone.[2, 7, 9-10, 14]

The existence of such finite excess heat capacities ($\Delta C_P \neq 0$) in nucleic acid folding is well-established, with the majority of ΔC_P measurements obtained from precision calorimetric methods under bulk ensemble conditions. Specifically, isothermal titration calorimetry (ITC) represents a precision tool for measuring ΔC_P in systems of medium to high affinity bimolecular association,[2, 12, 15-17] with differential scanning calorimetry (DSC) able to sensitively report on ΔC_P even for weak association processes as well as unimolecular folding dynamics.[17-20] UV absorption optical detection methods offer yet another alternative, whereby hypochromic shifts in DNA absorbance can be used to determine folding equilibrium constants (K_{eq}), which are subsequently analyzed via temperature dependent van't Hoff theory in order to extract ΔC_P.[2]

Each of these approaches offers a powerful and sensitive window into heat capacity differences associated with biomolecular conformational change, though up until now constrained to equilibrium bulk ensemble conditions.

The last two decades have witnessed considerable interest in extending thermodynamic study of variables such as $\Delta G°$, $\Delta H°$, and $\Delta S°$ based on single-molecule methods,[21-26] which offer a unique opportunity to probe thermodynamic "landscapes" down at the ultimate single-molecule sensitivity limit. Application of such measurement techniques to ΔC_P, however, have proven far more challenging and been notably absent in the single-molecule literature. In particular, temperature-dependent single-molecule folding studies are now relatively straightforward[1, 27] but to date have focused exclusively on measurements of folding free energies, enthalpies, and entropies rather than excess heat capacities. Indeed, despite an initial report by Williams et al. in 2001 establishing the feasibility of determining ΔC_P via optical tweezers,[28] twenty years have passed without the publication of any articles utilizing single molecule methods to extract ΔC_P.

In this work, we revisit the prospect for determination of ΔC_P at the single molecule level, based on high precision temperature dependent studies with single-molecule fluorescence resonance energy transfer (smFRET).[23, 26] We acknowledge at the outset that ITC and DSC measurements under bulk conditions offer greater sensitivity and precision; our goal is simply to establish smFRET measurement of ΔC_P as a viable, alternative tool in the single-molecule toolbox. Although this proves more experimentally challenging than conventional van't Hoff temperature measurements of $\Delta H°$ and $\Delta S°$, we note that such an approach nevertheless does offer unique new insights into oligomer hybridization dynamics. In particular, the single molecule FRET methods described herein provide additional access to folding *kinetics*

unavailable to ITC and DSC approaches, from which we can report first measurements of excess heat capacities involved in accessing *the transition state (i.e.,* $\Delta C^{\ddagger}_{P}$) for hybridization in nucleic acid oligomers.

6.3 Methods

The smFRET construct used in these studies is a DNA hairpin consisting of a 7-base pair stem and a 40-(dA) adenosine loop, which we have previously characterized.[29-30] As depicted schematically in Figure 6.1, the construct is fluorescently labeled with Cy3 and Cy5 for FRET measurements, and a biotin moiety is attached to the 3' end for surface tethering. The full DNA sequence (5'-Cy5-CTTCAGT-A_{40}-Cy3-ACTGAAG-A_{11}-biotin-3') is purchased in high performance liquid chromatography (HPLC) purified form from Integrated DNA Technologies. (Any company names listed herein are in the interest of completeness, not as endorsement of vendor).

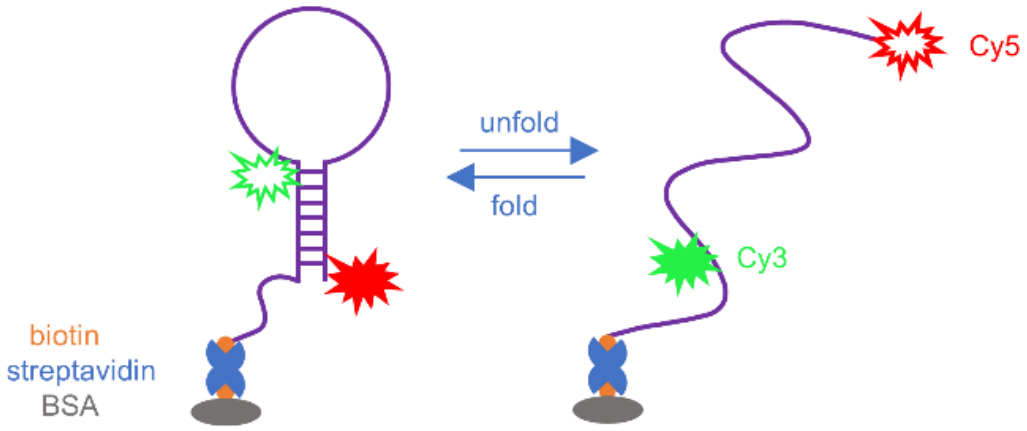

Figure 6.1 Schematic of DNA hairpin construct used in these studies.

Glass surfaces are cleaned and decorated with the smFRET construct in accordance with previously described protocols.[29, 31] Simply summarized, glass coverslips are soaked in acetone and treated in a UV-ozone cleaner for 30 minutes. The cleaned coverslips are then exposed, in sequence, to i) 10%-biotinylated bovine serum albumin (10 mg/mL), ii) streptavidin (0.2 mg/mL), and finally iii) the DNA construct (100 pM; 1 pM = 1 pmol/L). The resulting sample has a surface density of ~1 construct per 10 μm^2, yielding approximately 100 molecules in the microscope field of view. Molecules are observed in buffer (50 mM HEPES, pH 7.6) with 120 mM monovalent (K^+) ions. To decrease the rate of fluorophore photobleaching, an enzymatic oxygen scavenging cocktail[32] (100 nM protocatechuate 3,4-dioxygenase, 5 mM protocatechuic acid, and 2 mM Trolox) is also present.

Single-molecule fluorescence measurements are taken on a through-objective total internal reflection fluorescence (TIRF) microscope.[33,31] Briefly, laser light at 532 nm is focused onto the back focal plane of a high numerical aperture oil-immersion microscope objective, with a pair of mirrors translating the beam off axis to increase the angle of incidence at the sample until total internal reflection is achieved. Widefield fluorescence is collected by the objective and separated by dichroic mirror into Cy3 and Cy5 emission channels, which are each focused onto a charge-coupled device (CCD) video camera. To increase accuracy when measuring folding rate constants comparable to the CCD frame rate, we exploit stroboscopic excitation methods by modulating the laser to excite the sample for the first 20% of each frame.[34] Single molecule FRET trajectories are extracted from CCD movies by[31] i) applying a brightness threshold to locate particles, ii) summing pixels in a circular neighborhood around each particle to obtain fluorescence rates as a function of time, iii) subtracting local background, iv) spatially pairing Cy3 and Cy5 particles by a calibrated affine map,[35] and v) computing the FRET efficiency,

FRET(t) = I_{Cy5}(t)/(I_{Cy3}(t)+I_{Cy5}(t)). FRET trajectories are analyzed by dwell time analysis with correction for stroboscopic excitation[36] to yield folding and unfolding rate constants (k_F and k_U) as well as equilibrium constants ($K_{eq} = k_F/k_U$). Agreement between K_{eq} values measured by ratios of rate constants to those determined from integrated fractional time spent in the unfolded (T_U) vs folded (T_F) state ($K_{eq} = T_F/T_U$) is routinely at the < 2% level.

Sample temperature is controlled using thermoelectric cooling/heating modules under servo loop control, as previously established.[31] To prevent thermal gradients, the sample is heated from above by a module in direct contact with the sample and from below by a module attached to the microscope objective, which maintains thermal contact with the sample by the immersion oil. The system temperature is measured by thermistors, which we have calibrated by *in situ* measurements of the sample temperature using a thin wire thermocouple. The thermistor readings are used as input to the thermoelectric modules for computer-based feedback stabilization with a 0.015 Hz bandwidth, resulting in 0.1 °C temperature accuracy and stability on the 1-minute time scale.

6.4 Results

6.4.1 Heat capacity in van't Hoff analysis

Measurement of the equilibrium constant at multiple temperatures provides insight into the thermodynamic parameters (enthalpy/entropy) of the process. Specifically, the van't Hoff equation[2] describes the dependence of the equilibrium constant K_{eq} as a function of overall $\Delta H°$ and entropy $\Delta S°$:

$$\ln(K_{eq}) = -\frac{\Delta H°(T)}{R}\left(\frac{1}{T}\right) + \frac{\Delta S°(T)}{R} \qquad \text{(Eq. 6.1)}$$

where T is the absolute temperature and R is the gas constant. Notably, if $\Delta H°$ and $\Delta S°$ are independent of temperature, then the resulting van't Hoff plot of $\ln(K_{eq})$ vs. $1/T$ is predicted to be linear, with slope $-\Delta H°/R$ and intercept $\Delta S°/R$.

If products and reactants have different heat capacities C_P, then the overall reaction will have a nonzero excess heat capacity $\Delta C_P = C_P$ (products) – C_P (reactants). The excess heat capacity is related to $\Delta H°$ and $\Delta S°$ by the fundamental thermodynamic expressions[2]

$$\Delta C_P = \frac{\partial}{\partial T}(\Delta H°)_P \qquad \text{(Eq. 6.2)}$$

$$\Delta C_P = T\frac{\partial}{\partial T}(\Delta S°)_P \qquad \text{(Eq. 6.3)}$$

If ΔC_P is approximated as constant over the measured temperature range, we may integrate these equations and substitute the resulting temperature-dependent $\Delta H°(T)$ and $\Delta S°(T)$ into Eq. 6.1 to produce the ΔC_P-modified van't Hoff equation,

$$\ln(K_{\mathrm{eq}}) = -\frac{\Delta H^{o}_{ref}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{o}_{ref}}{R} - \frac{\Delta C_P}{R}\left[\left(\frac{T-T_{ref}}{T}\right) + \ln\left(\frac{T_{ref}}{T}\right)\right] \qquad \text{(Eq. 6.4)}$$

where T_{ref} is an arbitrary reference temperature and $\Delta H°_{ref}$ and $\Delta S°_{ref}$ are $\Delta H°(T)$ and $\Delta S°(T)$ evaluated at T_{ref}, respectively. Most importantly, the additional logarithmic term introduces non-linearity into the single molecule van't Hoff plot, which provides a simple experimental diagnostic for nonzero ΔC_P.

6.4.2 smFRET measurements reveal nonzero ΔC_P in DNA folding

In order to test these predictions, we have measured the folding kinetics of a FRET-labeled DNA hairpin (see Figure 6.1) as a function of temperature. Across all temperatures (13–35.5 °C), smFRET trajectories exhibit clear switching between two distinct FRET states (see Figure 6.2), representing unhybridized ($E_{FRET} \approx 0$) and hybridized ($E_{FRET} \approx 0.8$) conformations.

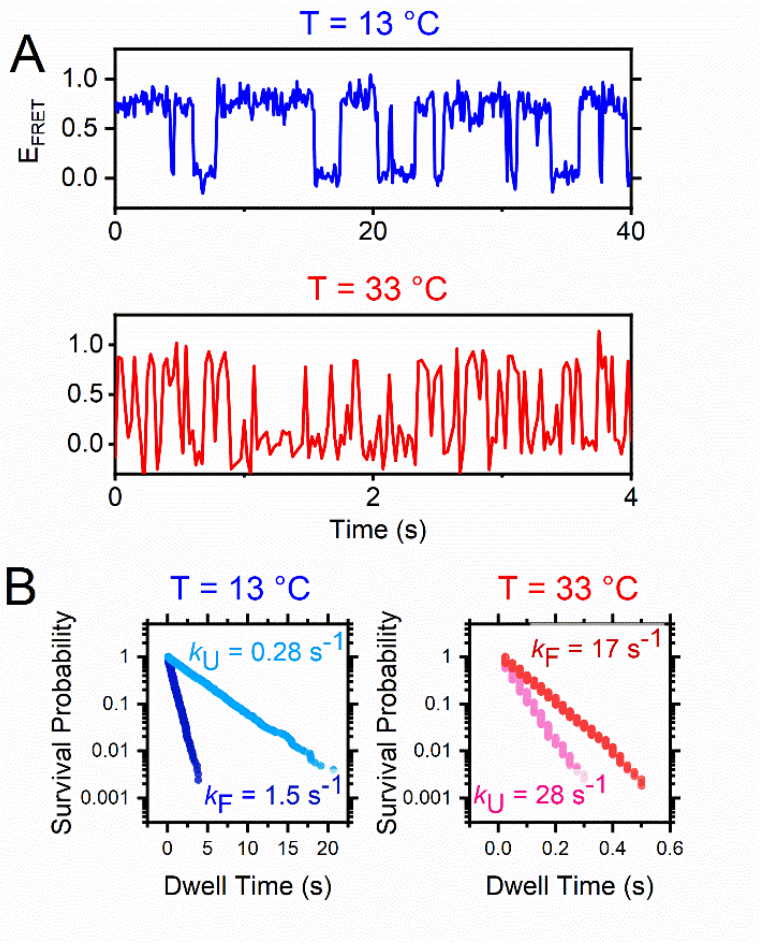

Figure 6.2 Sample single-molecule FRET time trajectories (A) and dwell time distributions (B) for DNA hairpin folding at 13 °C and 33 °C. All data are obtained in 90 mM K^+ and 30 mM Na^+.

As expected for an exothermic folding process ($\Delta H^0 < 0$), the population of the folded state decreases with increasing temperature relative to the unfolded state, corresponding to heat induced denaturation or "melting" of the DNA hairpin.[3, 6] In order to extract the folding/unfolding dynamics in more quantitative detail, we perform dwell time analysis on the smFRET trajectories.[37] At each temperature, the folding and unfolding cumulative dwell time distributions are well-fit by single exponential decays, indicative of simple two-state kinetics governed by the unimolecular rate constants k_F and k_U. From these rate constants, we can directly compute the folding equilibrium constant $K_{eq} = k_F / k_U$ as a function of temperature.

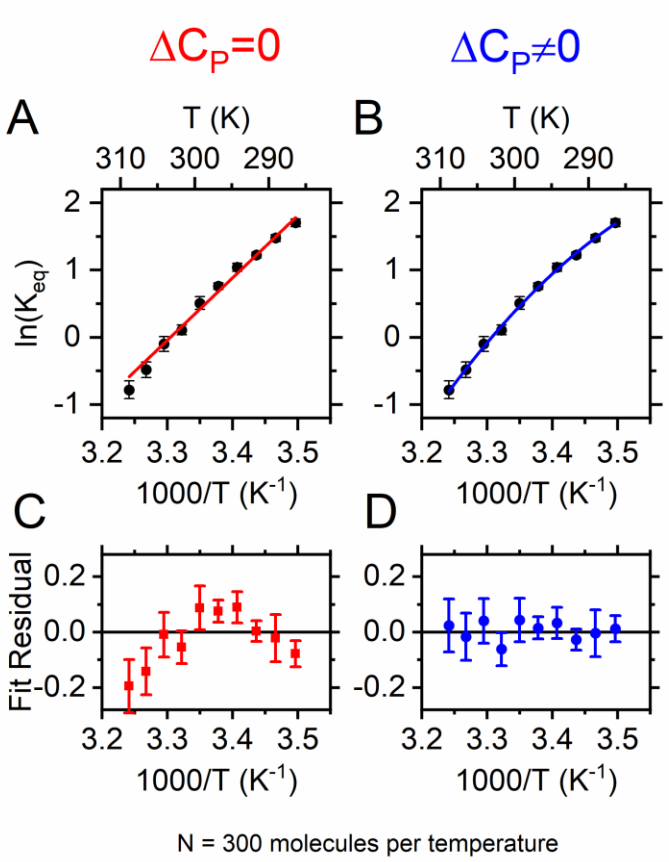

Figure 6.3 van't Hoff analysis of smFRET-derived equilibrium constants with and without modifications for nonzero excess heat capacity (ΔC_P). (A) Linear van't Hoff fit. (B) Fit with ΔC_P term to account for curvature. Fit residuals for the linear fit (C) show systematic deviations away from linearity, while residuals for ΔC_P -modified fit (D) exhibit no temperature dependence.

The temperature-dependence of K_{eq} for the DNA hairpin is presented as a van't Hoff plot in Figure 6.3. By way of a first order treatment, we have analyzed these data with a simple linear fit (Figure 6.3A, red line), which is equivalent to setting $\Delta C_P = 0$ in Eq. 6.4. From the fit residuals (Figure 6.3C), however, the data clearly deviate systematically from such a linear model, with the fit overestimating K_{eq} at both the lowest and highest temperatures. To account for such curvature in the van't Hoff plot, we therefore fit the data to the modified van't Hoff equation (Eq. 6.4), permitting $\Delta C_P \neq 0$. The residuals now exhibit little to no systematic temperature dependence, suggesting a more satisfactory model fit. From a non-linear weighted

least squares fit, we infer the excess heat capacity to be $\Delta C_P = -2.5 \pm 0.4$ kJ/mol/K, where the uncertainty represents 1σ standard deviation. The small fractional uncertainty (< 20%) in ΔC_P provides additional support for data consistency with *nonzero excess heat capacity* ($\Delta C_P \neq 0$) upon DNA oligo hybridization.

6.4.3 Arrhenius analysis of rate constants: Transition state excess heat capacity

Temperature dependent equilibrium constants provide clear evidence for the influence of finite excess heat capacity ($\Delta C_P \neq 0$) on the overall DNA hybridization event, as observed herein at the single molecule level. However, there is also additional thermodynamic information encoded in the *kinetics* of such single molecule data, which is not readily accessible in bulk ITC or DSC studies. Specifically, we can exploit Arrhenius plots of ln(k) vs. 1/T to further obtain thermodynamic information about the *transition state* for DNA hybridization. For example, the kinetic data is often fit to the standard Eyring transition state theory (TST) result[38-39]

$$\ln(k) = -\frac{\Delta H^{\ddagger}(T)}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\ddagger}(T)}{R} + \ln(\nu^{\ddagger}) \qquad \text{(Eq. 6.5)}$$

where k is the folding or unfolding rate constant, $\Delta H^{\ddagger}$ is the activation enthalpy, $\Delta S^{\ddagger}$ is the activation entropy, and $\nu^{\ddagger}$ is the attempt frequency for the system to reach the transition state. Similar to the van't Hoff analysis (Eq. 6.2−6.4), we can extend our TST expression in Eq. 6.5 to include non-zero excess heat capacity between transition state and the reactants, yielding:

$$\ln(k) = -\frac{\Delta H^{\ddagger}_{ref}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\ddagger}_{ref}}{R} - \frac{\Delta C_P^{\ddagger}}{R}\left[\left(\frac{T - T_{ref}}{T}\right) + \ln\left(\frac{T_{ref}}{T}\right)\right] + \ln(\nu^{\ddagger}) \qquad \text{(Eq. 6.6)}$$

where $\Delta H^{\ddagger}_{ref}$ and $\Delta S^{\ddagger}_{ref}$ are $\Delta H^{\ddagger}(T)$ and $\Delta S^{\ddagger}(T)$ evaluated at the reference temperature T_{ref}, and $\Delta C^{\ddagger}_P$ is the excess activation heat capacity. Just as in our van't Hoff analysis, a nonzero $\Delta C^{\ddagger}_P$ results in curvature in the Arrhenius plot, which for $\Delta C^{\ddagger}_P = 0$ would be perfectly linear. We note

that $\Delta S^{\ddagger}_{ref}$ and $\ln(\nu^{\ddagger})$ cannot be determined independently from such an analysis, since both parameters are perfectly correlated and simply account for a vertical offset. However, such potential ambiguity in $\Delta S^{\ddagger}_{ref}$ and $\ln(\nu^{\ddagger})$ has no effect on $\Delta C^{\ddagger}_{P}$, which is a function only of *curvature* in the temperature dependent Arrhenius plot. We also note that the applicability of TST to nucleic acid folding is complicated by the presence of multiple folding pathways, as is posited for instance in the kinetic zipper model of DNA hybridization[40-42] (see Discussion). In such a scenario, the measured $\Delta H^{\ddagger}_{ref}$, $\Delta S^{\ddagger}_{ref}$, and $\Delta C^{\ddagger}_{P}$ are instead interpreted as an average over the multiple folding pathways.

The temperature-dependent unfolding kinetic measurements for the DNA hairpin are reported in Figure 6.4A, where the black squares represent experimental data and the solid red

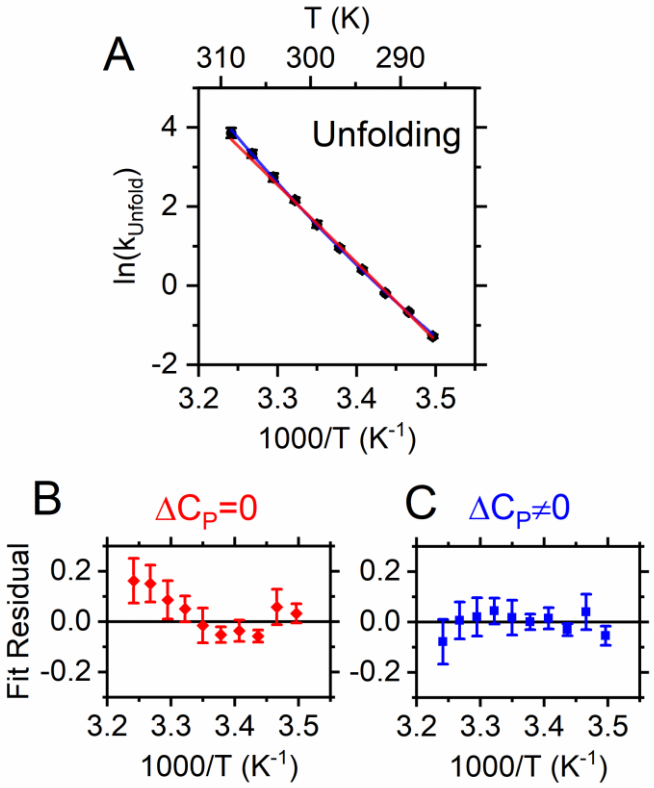

Figure 6.4 Arrhenius analysis of unfolding rate constant k_U. (A) Data are fit to linear (red) and nonlinear (blue) models. Linear fit residuals (B) exhibit systematic errors compared to nonlinear fit residuals (C), in support of non-negligible ΔC_P.

and black lines reflect the linear ($\Delta C^{\ddagger}_P = 0$) and higher order, non-linear approximations ($\Delta C^{\ddagger}_P \neq 0$), respectively. The rate constants increase by more than two orders of magnitude over the temperature range tested, from 0.28 ± 0.01 s^{-1} at 13 °C to 48 ± 4 s^{-1} at 35.5 °C. We have fit these data with both a $\Delta C^{\ddagger}_P = 0$ two-parameter fit (Eq. 6.5) and variable $\Delta C^{\ddagger}_P \neq 0$ three-parameter fit (Eq. 6.6), with differences between the two fits highlighted by examining the residuals (Figure 6.4B−C). The residuals for the $\Delta C^{\ddagger}_P = 0$ fit clearly show systematic temperature-dependent deviations, with the linear fit underestimating measured rate constants at both low and high temperatures. Allowance for a nonzero $\Delta C^{\ddagger}_P$ results in significant improvement, though not as dramatically as for the van't Hoff analysis (Figure 6.3C−D). To help evaluate the quality of fit (and suitability of the underlying model), we have examined the reduced χ^2 statistic, which is the average of the square of the residuals divided by their variance ($\chi^2 = \langle (y_i - y_{i,fit})^2/\sigma_i^2 \rangle$), which should be ≈ 1 for a physically correct model.[43] For the linear fit ($\Delta C^{\ddagger}_P = 0$), we find $\chi^2 = 2.7$, while for the non-linear fit ($\Delta C^{\ddagger}_P \neq 0$), $\chi^2 = 0.98$, which signals clear justification for the higher-level analysis. In summary, the non-linear least squares fits yield $\Delta C^{\ddagger}_P = 2.3 \pm 0.7$ kJ/mol/K for the excess *activation heat capacity* between the i) fully folded and ii) transition state for unfolding of the smFRET construct.

A similar analysis can of course be performed on k_F, though the curvature in such an Arrhenius plot (see Figure 6.5) is much reduced, with a non-zero value of $\Delta C^{\ddagger}_P$ on the threshold of our experimental resolution. Inspection of the fit residuals shows negligible visible improvement in the residuals upon introducing a nonzero $\Delta C^{\ddagger}_P$ term. Indeed, this is confirmed by the χ^2 values for the folding rate constant statistics, which are close to unity for both linear (χ^2 =

1.7) and nonlinear (χ^2 = 1.3) fits. Nevertheless, an error-propagated estimate of $\Delta C^{\ddagger}_P$, −0.7 ± 0.4 kJ/mol/K, for unfolding kinetic data appears finite, though zero within 95% (2σ) uncertainty. The results from all fits of the equilibrium and rate constants to van't Hoff and Arrhenius models with finite excess heat capacity are summarized in Table 6.1.

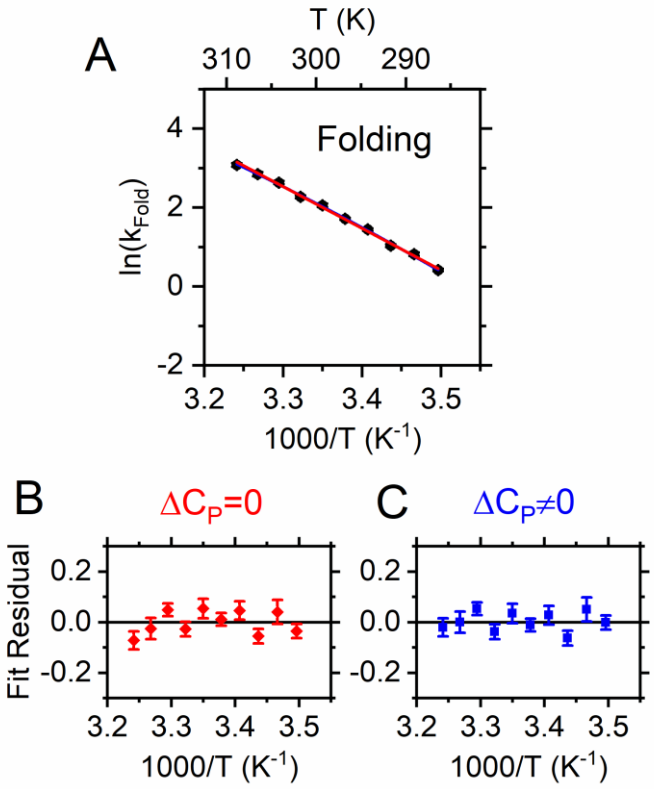

Figure 6.5 Arrhenius analysis of folding rate constant k_F. (A) Linear (red) and nonlinear (blue) fits for k_F are visually indistinguishable. (B) Residuals of the linear fit are unstructured, and (C) introduction of a nonzero ΔC_P in fitting has little effect on residual errors.

	$\Delta H°_{ref}$ (kJ/mol)	$\Delta S°_{ref}$ (J/mol K)	ΔC_P (kJ/mol K)
	-88(2)	-294(8)	-2.5(4)
	$\Delta H^{\ddagger}_{ref}$ (kJ/mol)	$\Delta S^{\ddagger}_{ref}$ (J/mol K)	$\Delta C^{\ddagger}_P$ (kJ/mol K)
Folding	85(2)	57(7)	-0.7(4)
Unfolding	176(4)	357(13)	2.3(7)

Table 6.1. Fit Results of Temperature-Dependent Measurements. The reference temperature is T_{ref} = 300 K and the attempt frequency is $\nu^{\ddagger}$ = 6 x 10^{12} s^{-1}. See text for details.

6.5 Discussion

In this work, we have explored the capacity of single molecule FRET microscopy for measuring excess heat capacities in biomolecular folding. As a proof-of-concept demonstration, we have measured the folding dynamics of a DNA hairpin test system[29-30] (Figure 6.1) from 13 °C to 35.5 °C on a single-molecule TIRF microscope. Deviations from linear behavior in van't Hoff plots of K_{eq} and Arrhenius plots of k_U (Figs. 3 and 4), both of which indicate clear effects due to non-zero excess heat capacities. From the curvature in these plots, we have extracted differential heat capacities for the overall folding reaction (ΔC_P) and for the approach to the transition state ($\Delta C^{\ddagger}_P$) from the reactants and products (Table 6.1).

The ability to perform such measurements for excess heat capacity ΔC_P requires the combination of a large temperature range and precise rate constant measurements. For example, maximal deviations from linearity in the Arrhenius plot for the unfolding rate constant k_U (Figure 6.4) are only of order ≈ 0.1 logarithmic units, which in turn demands <10% fractional uncertainties to observe. We achieve this level of precision by acquiring 20,000 dwell times from 300 molecules for each of 10 different temperatures. Obviously, a high-throughput method such as widefield TIRF microscopy greatly facilitates this level of data acquisition in comparison to confocal microscopy.[25] Such demands on precision can be relaxed if the temperature range is expanded, as the deviations from linearity grow quadratically with ΔT. However, the experimentally available range of temperatures is restricted by the stability of the folding, since smFRET kinetic measurements are in practice typically limited[37] to systems with $0.1 \leq K_{eq} \leq 10$. It is worth noting that this restriction only applies when measuring equilibrium constants via $K_{eq} = k_F/k_U$, with the dynamic range of measurable K_{eq} values 1−2 orders of magnitude greater if one instead integrates FRET histograms to determine population ratios. The measurement of rate

constants is clearly worth the additional effort, however, since such kinetic data provide the additional capacity to measure excess activation heat capacities *for accessing the transition state*, which represents one unique advantage of using single molecule methods over ITC or DSC.

Excess heat capacities in protein and nucleic acid folding are phenomenologically linked to changes in surface area during folding.[2, 7, 9-10, 12, 14] The surface area effect has been ascribed to solvent interactions with the nucleic acid, including the hydrophobic effect, perturbations of intramolecular vibrations, and hydrogen bonding.[13] Proteins undergo significant surface area loss during folding, whereas the higher charge densities of polyanionic nucleic acids prevents such compactification. Thus, nucleic acids tend to have smaller specific excess heat capacities than proteins.[13] As a consequence, excess heat capacities in nucleic acids have been more challenging to measure than for proteins and were indeed once thought to be identically zero.[13] Nonzero ΔC_P values for nucleic acid folding are now well-established,[7, 44-45] with burial of hydrophobic nucleobases during base pairing considered the primary contributor.[2]

The current single molecule data for excess heat capacities of nucleic acid folding can be usefully compared with results from bulk calorimetric studies. The specific DNA hairpin construct explored in these studies experiences a *decrease* in heat capacity upon folding (ΔC_P = −2.4 ± 0.4 kJ/mol/K), which is consistent with the general literature consensus of $\Delta C_P < 0$ due to a decrease in accessible surface area. Since the surface area loss is proportional to the number of base pairs formed, and since ΔC_P values are at least thought to have only a minor sensitivity to DNA sequence,[12, 20] it is convenient to report ΔC_P values per base pair, which for our 7 base pair hairpin is −340 ± 60 J/mol/K/bp (bp = base pair). Previous literature values for ΔC_P from bulk studies exhibit a rather broad range from −100 to −420 J/mol/K/bp with a recent metastudy[7] recommending −130 J/mol/K/bp, against which the current single molecule measurement is high

but nevertheless qualitatively consistent. Indeed, the 3-fold dynamic range of these literature results is an indication of the significant experimental challenges involved even in bulk studies of DNA folding as well as additional sensitivity to the nature of the DNA construct.

As a particularly important conformational distinction, most bulk studies have explored duplex formation for *bimolecular* rather than *unimolecular* folding, as is the case for the 7 bp DNA hairpin construct connected by a 40-adenine loop utilized in the current single molecule efforts. Indeed, if we further limit the previous literature comparison only to measurements on unimolecular hairpins, the reported ΔC_P values grow significantly to −300 J/mol/K/bp, which is now in even better agreement with our single molecule results.[46-48] Other aspects of single molecule experimental design can influence duplex stability, including surface tethering[49] and fluorophore incorporation,[50] which may limit comparison with bulk studies of freely-diffusing, unlabeled oligonucleotides. Finally, the polyadenosine (poly-dA) loop in the hairpin is known to form single-stranded base-stacked structures[51] which are likely disrupted during folding, leading to an increase in surface area and a concomitant increase in C_P. Therefore, loops sequences that are less prone to base-stacking, such as polythymidine (poly-dT), may have a more negative ΔC_P for folding. These comparisons serve to further highlight the level of complexity in such measurements, with expectations for a possible dependence on loop sequence and GC/AT composition in our smFRET construct.

In addition to the overall heat capacity change during folding (ΔC_P), we have measured the heat capacity change required to reach the folding transition state ($\Delta C^{\ddagger}_P$). While activation heat capacities have been reported for proteins,[52] this work provides, to the best of our knowledge, the first such measurement of $\Delta C^{\ddagger}_P$ for nucleic acid folding. As one simple check, we confirm that the overall folding heat capacity ($\Delta C_P = -2.5 \pm 0.4$ kJ/mol/K) is within

uncertainty of the sum of the activation heat capacities ($\Delta C^{\ddagger}_{P,Fold}$ + -$\Delta C^{\ddagger}_{P,Unfold}$ = −3.0 ± 0.8 kJ/mol/K), as required since C_P is a state function. We can combine this information in ΔC_P and $\Delta C^{\ddagger}_P$ to construct a profile of the system heat capacity "landscape" along the reaction coordinate (Figure 6.6). If negative excess heat capacities in nucleic acid folding arise primarily from the loss of accessible surface area during base pair formation, as is the literature consensus,[2, 13] then one consistent expectation is that the heat capacity of the transition state should lie between the folded and unfolded values, since the transition state cannot form more base pairs than the folded state nor less base pairs than the unfolded state. Indeed, the experimental ΔC_P landscape profile reveals an intermediate transition state heat capacity between that of the unfolded and folded states, in good agreement with this expectation. Interestingly, the transition state excess heat capacity $C^{\ddagger}_P$ appears to be closer to that of the completely unhybridized unfolded state. If one

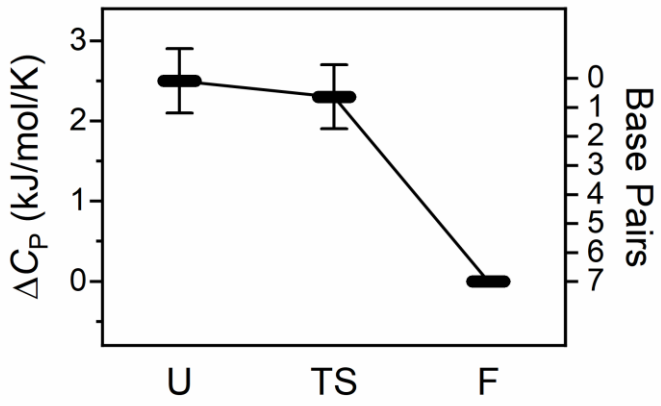

Figure 6.6 Excess heat capacity "landscape" changes along the folding reaction coordinate, relative to the folded state. U = unfolded state, TS = transition state, F = folded state. Note the monotonic decrease in this excess heat capacity upon hybridization of the 7 bp DNA stem and the relatively small activation heat capacity in approaching the transition state from the unfolded direction. The right axis indicates the number of base pairs formed at each state, based on a simple linear model for heat capacity proportional to solvent accessible surface area. Such behavior suggests minimal base pairing at the transition state, which is entirely consistent with "zippering" models for DNA hybridization.[41, 59] Error bars reflect one standard error of the mean.

assumes linearity in these excess heat capacities with hybridization sequence length, this would suggest that relatively few base pairs ($n^{\ddagger}_{bp} = 1 \pm 1$) are fully formed in the transition state, certainly including zero within error.

The conclusion that few base pairs are formed at the DNA hairpin's folding transition state is in good agreement with the current kinetic zipper model of nucleic acid hybridization.[40-42, 53-54] In the zipper model, the free energy barrier for folding is primarily entropic, as the two DNA strands must "pre-align" in a correct orientation before forming one or more key base pair contact(s).[55] This nucleation site then catalyzes the formation of base pairs at adjacent sites in an enthalpically-downhill cascade to the folded state. The number of base pairs in the transition state is a subject of ongoing investigation, with evidence ranging for a transition state containing a single base pair,[56] a minimum of 2 base pairs,[41] or the absence of any nascent base pairs at all.[57-58] Clearly, more work is warranted, and while the uncertainty in our results is too large to resolve these possibilities ($n^{\ddagger}_{bp} = 1 \pm 1$), our heat capacity-based measurement provides further confirmation of the kinetic zipper prediction of an "early" transition state for hybridization; i.e., a transition state having few if any fully formed base pairs.

6.6 Summary and conclusion

In this report, we demonstrate the use of smFRET to measure excess and activation heat capacities in biomolecular folding. By combining a large temperature range with high level data collection statistics, our experiment achieves the necessary precision to resolve nonzero heat capacities in the folding of a DNA hairpin. The measured value of ΔC_P (340 ± 60 J/mol/K per base pair) is in good agreement with literature calorimetric results on similar constructs, which provides confirming evidence for the accuracy of our method. Furthermore, the present single

molecule studies provide kinetic evidence for finite excess heat capacities upon approach to the transition state ($\Delta C^{\ddagger}_{P}$), which provides novel confirmation that the transition state contains only few fully formed base pairs ($n^{\ddagger}_{bp} = 1 \pm 1$) and support of the "kinetic zipper" model for DNA hybridization kinetics. In summary, heat capacity measurements at the single molecule level offer a useful complement to calorimetric bulk methods and are uniquely capable of inspecting heat capacities of transition states.

6.7 References

1. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* **2016,** *67*, 441-65.

2. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with nucleic acid folding. *Biopolymers* **2006,** *82* (1), 38-58.

3. SantaLucia, J., Jr.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-40.

4. Tsourkas, A.; Behlke, M. A.; Rose, S. D.; Bao, G., Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Res.* **2003,** *31*, 1319-1330.

5. Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F., Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (35), 15431-6.

6. SantaLucia, J., Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (4), 1460-5.

7. Dragan, A.; Privalov, P.; Crane-Robinson, C., Thermodynamics of DNA: Heat capacity changes on duplex unfolding. *Eur. Biophys. J.* **2019,** *48* (8), 773-779.

8. Gallagher, K.; Sharp, K., Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* **1998,** *75* (2), 769-776.

9. Hadzi, S.; Lah, J., Origin of heat capacity increment in DNA folding: The hydration effect. *Biochim. Biophys. Acta, Gen. Subj.* **2021,** *1865* (1), 129774.

10. Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; Record, M. T., Jr., 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. *Biochemistry* **1999,** *38* (26), 8409-22.

11. Hughesman, C. B.; Turner, R. F. B.; Haynes, C., Correcting for heat capacity and 5'-TA type terminal nearest neighbors improves prediction of DNA melting temperatures using nearest-neighbor thermodynamic models. *Biochemistry* **2011,** *50* (13), 2642-2649.

12. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with DNA duplex formation: Salt- and sequence-dependent effects. *Biochemistry* **2006,** *45* (2), 604-616.

13. Volker, J.; Plum, G. E.; Breslauer, K. J., Heat capacity changes (ΔC_p) for interconversions between differentially-ordered DNA states within physiological temperature domains: Implications for biological regulatory switches. *J. Phys. Chem. B* **2020,** *124* (27), 5614-5625.

14. Cooper, A., Heat capacity effects in protein folding and ligand binding: A re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.* **2005,** *115* (2), 89-97.

15. Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q., Insights into protein-ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **2016,** *17* (2).

16. Majhi, P. R.; Qi, J.; Tang, C.-F.; Shafer, R. H., Heat capacity changes associated with guanine quadruplex formation: An isothermal titration calorimetry study. *Biopolymers* **2008,** *89* (4), 302-309.

17. Jelesarov, I.; Crane-Robinson, C.; Privalov, P. L., The energetics of HMG box interactions with DNA: Thermodynamic description of the target DNA duplexes. *J. Mol. Biol.* **1999,** *294* (4), 981-95.

18. Jelesarov, I.; Bosshard, H. R., Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **1999,** *12* (1), 3-18.

19. Johnson, C. M., Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.* **2013,** *531* (1-2), 100-109.

20. Chalikian, T. V.; Volker, J.; Plum, G. E.; Breslauer, K. J., A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (14), 7853-8.

21. Aleman, E. A.; Lamichhane, R.; Rueda, D., Exploring RNA folding one molecule at a time. *Curr. Opin. Chem. Biol.* **2008,** *12* (6), 647-654.

22. Elf, J.; Barkefors, I., Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* **2019,** *88* (1), 635-659.

23. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **2015,** *34*, 52-59.

24. Bizzarri, A. R.; Cannistraro, S., The application of atomic force spectroscopy to the study of biological complexes undergoing a biorecognition process. *Chem. Soc. Rev.* **2010,** *39* (2), 734-749.

25. Tan, Y. W.; Hanson, J. A.; Chu, J. W.; Yang, H., Confocal single-molecule FRET for protein conformational dynamics. *Methods. Mol. Biol.* **2014,** *1084*, 51-62.

26. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Foerster resonance energy transfer. *Science* **2018,** *359* (6373), 288.

27. Baker, M. A.; Inoue, Y.; Takeda, K.; Ishijima, A.; Berry, R. M., Two methods of temperature control for single-molecule measurements. *Eur. Biophys. J.* **2011,** *40* (5), 651-60.

28. Williams, M. C.; Wenner, J. R.; Rouzina, I.; Bloomfield, V. A., Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching. *Biophys. J.* **2001,** *80* (4), 1932-9.

29. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: Kinetic insights from single-molecule studies. *J. Phys. Chem. B* **2018,** *122* (43), 9869-9876.

30. Sung, H. L.; Nesbitt, D. J., Single-molecule kinetic studies of DNA hybridization under extreme pressures. *Phys. Chem. Chem. Phys.* **2020,** *22* (41), 23491-23501.

31. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B* **2020,** *124* (51), 11561-11572.

32. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **2008,** *94* (5), 1826-35.

33. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

34. Nicholson, D. A.; Nesbitt, D. J., Pushing camera-based single molecule kinetic measurements to the frame acquisition limit with stroboscopic smFRET. *J. Phys. Chem. B* **2021,** *125* (23), 6080-6089.

35. Cohen, E. A. K.; Ober, R. J., Image registration error analysis with applications in single molecule microscopy. In *Proc. I. S. Biomed. Imag.*, 2012; pp 996-999.

36. Nicholson, D. A.; Nesbitt, D. J., Pushing camera-based single molecule kinetic measurements to the frame acquisition limit with stroboscopic smFRET. *J. Phys. Chem. B* **2021,** *in press*.

37. Kinz-Thompson, C. D.; Bailey, N. A.; Gonzalez, R. L., Precisely and accurately inferring single-molecule rate constants. In *Methods in Enzymology*, Spies, M.; Chemla, Y. R., Eds. Academic Press: 2016; Vol. 581, pp 187-225.

38. Pechukas, P., Transition state theory. *Annu. Rev. Phys. Chem.* **1981,** *32* (1), 159-177.

39. Laidler, K. J.; King, M. C., Development of transition-state theory. *J. Phys. Chem.* **1983,** *87* (15), 2657-2664.

40. Kuznetsov, S. V.; Ansari, A., A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* **2012,** *102* (1), 101-11.

41. Ouldridge, T. E.; Sulc, P.; Romano, F.; Doye, J. P.; Louis, A. A., DNA hybridization kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **2013,** *41* (19), 8886-95.

42. Yin, Y.; Zhao, X. S., Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* **2011,** *44*, 1172-1181.

43. Taylor, J. R., *An introduction to error analysis : The study of uncertainties in physical measurements*. University Science Books: 1982.

44. Rouzina, I.; Bloomfield, V. A., Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.* **1999,** *77* (6), 3242-51.

45. Rouzina, I.; Bloomfield, V. A., Heat capacity effects on the melting of DNA. 2. Analysis of nearest-neighbor base pair effects. *Biophys. J.* **1999,** *77* (6), 3252-5.

46. Völker, J.; Makube, N.; Plum, G. E.; Klump, H. H.; Breslauer, K. J., Conformational energetics of stable and metastable states formed by DNA triplet repeat oligonucleotides: Implications for triplet expansion diseases. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (23), 14700.

47. Bourdélat-Parks, B. N.; Wartell, R. M., Thermodynamic stability of DNA tandem mismatches. *Biochemistry* **2004,** *43* (30), 9918-9925.

48. Amrane, S.; Saccà, B.; Mills, M.; Chauhan, M.; Klump, H. H.; Mergny, J.-L., Length-dependent energetics of $(CTG)_n$ and $(CAG)_n$ trinucleotide repeats. *Nucleic Acids Res.* **2005,** *33* (13), 4065-4077.

49. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* **2008,** *5* (6), 507-16.

50. Moreira, B. G.; You, Y.; Owczarzy, R., Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: Predictive thermodynamic model. *Biophys. Chem.* **2015,** *198*, 36-44.

51. McIntosh, Dustin B.; Duggan, G.; Gouil, Q.; Saleh, Omar A., Sequence-dependent elasticity and electrostatics of single-stranded DNA: Signatures of base-stacking. *Biophys. J.* **2014,** *106* (3), 659-666.

52. Oliveberg, M.; Tan, Y. J.; Fersht, A. R., Negative activation enthalpies in the kinetics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92* (19), 8926-8929.

53. 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* (50), 16947-16952.

54. He, G.; Li, J.; Ci, H.; Qi, C.; Guo, X., Direct measurement of single-molecule DNA hybridization dynamics with single-base resolution. *Angew. Chem., Int. Ed.* **2016,** *55* (31), 9036-9040.

55. 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.

56. Strunz, T.; Oroszlan, K.; Schafer, R.; Guntherodt, H. J., Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (20), 11277-82.

57. Whitley, K. D.; Comstock, M. J.; Chemla, Y. R., Elasticity of the transition state for oligonucleotide hybridization. *Nucleic Acids Res.* **2017,** *45* (2), 547-555.

58. Ho, D.; Zimmermann, J. L.; Dehmelt, F. A.; Steinbach, U.; Erdmann, M.; Severin, P.; Falter, K.; Gaub, H. E., Force-driven separation of short double-stranded DNA. *Biophys. J.* **2009,** *97* (12), 3158-67.

59. Kuznetsov, S. V.; Ansari, A., A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* **2012,** *102* (1), 101-11.

Chapter 7

Kinetic and thermodynamic control of G-quadruplex polymorphism by Na^+ and K^+ cations*

7.1 Abstract

G-quadruplexes (G4s) are ubiquitous nucleic acid folding motifs that exhibit structural diversity dependent on cationic conditions. In this work, we exploit temperature-controlled single-molecule fluorescence resonance energy transfer (smFRET) to elucidate the kinetic and thermodynamic mechanisms by which monovalent cations (K^+, Na^+) impact folding topologies for a simple G-quadruplex sequence (5'-GGG-$(TAAGGG)_3$-3'). Kinetic measurements indicate that Na^+ and K^+ influence G4 formation in two distinctly different ways: the presence of Na^+ modestly enhances a parallel/hybrid G4 topology through an induced-fit (IF) mechanism with low affinity (K_d = 228 ± 26 mM), while K^+ drives G4 into an antiparallel topology via a conformational selection (CS) mechanism with a much higher affinity (K_d = 1.9 ± 0.2 mM). Furthermore, temperature dependent studies of the equilibrium folding ratio reveal distinctly opposing thermodynamic signatures for binding of K^+ ($\Delta H°_{bind} > 0$, $\Delta S°_{bind} > 0$) vs. Na^+ ($\Delta H°_{bind} < 0$, $\Delta S°_{bind} < 0$) to G4, which further illuminates the diversity of possible pathways for monovalent facilitation of G-quadruplex folding.

7.2 Introduction

G-quadruplexes (G4s) represent ubiquitous nucleic acid secondary folding motifs that can form in G-rich DNA and RNA.[1] Stable under both *in vitro* and *in vivo* conditions, G4s play

* Submitted for publication.

important roles in numerous biological contexts, including replication, transcription, and chromosomal stabilization.[2] Recently, G4s have drawn particular attention as potential drug targets for antiviral and anticancer therapies,[3-6] while G4 folds have been incorporated as modular components in DNA nanostructures.[7-8] Predictive understanding and control of G4s in these diverse biological and biotechnological areas has proven elusive and requires a comprehensive knowledge of the underlying biophysical issues regarding structure, kinetics, and thermodynamics for nucleic acid folding.

In contrast with canonical nucleic acid duplexes, G4s represent a highly structurally diverse secondary folding motif.[9-10] The signature element of G4 structure is the presence of G-tetrads, planar arrangements of four noncanonically base-paired (and non-contiguous) guanine residues, which then stack ($n \approx 3, 4$) to stabilize multiple G4 planes through adjacent π-π interactions. The guanine residues in a G4 can come from multiple nucleic acids as part of a higher order molecular complex, or from a single contiguous nucleic acid, in which case the G4 is referred to as a unimolecular fold. In such a unimolecular fold G4, sets of 4 guanine residues, called G-tracts, are spaced by loop sequences (usually 2-3 residues) to sterically allow the G-tracts to interact in a common plane. These loops permit adjacent G-tracts to have the same or opposite polarity (5’ to 3’ orientation), which allows G4s to fold into one of many topologies, each defined by a specific combination of the G-tract polarities.[11] G4 topologies are therefore classified as parallel (all strands share the same polarity), antiparallel (strands alternate polarities) or hybrid. Whether a G4 folds into a single topology or multiple topologies (polymorphism) depends on multiple factors, including loop sequence.[12-13] Beyond the multiple possible folded state topologies, G4 structural diversity is further increased by the presence of

partially folded intermediates such as G-triplexes (G3) or G-duplexes (G2) which can also be significantly populated and interconvert.[14-15]

G4 folding is accompanied by obligate binding of cations to the central channel formed by the stacked G-tetrads.[16] This binding site is capable of binding polyvalent cations but greatly favors monovalent cations, especially potassium (K^+). Importantly, the structure adopted by a G4 is influenced by the identity of its bound cation(s). For example, the human telomere G4 sequence folds into a parallel or hybrid topology in K^+ solution and an antiparallel topology in Na^+ solution.[17-18] The polymorphism in G4 sequences is clearly regulated by cationic interactions, which has been a subject of much research activity. The influence of ions on G4 folding has been studied extensively by bulk methods; [10, 12-13, 19-25] however, the multi-state polymorphism of G4s can make such ensemble results difficult to interpret. Single-molecule experiments, on the other hand, are more ideally suited to resolve state-specific properties in multistate folding systems.[26-27] In particular, single-molecule methods have been used to examine many aspects of G4 folding[28] including G4 folding kinetics[29-33] and temperature-dependence of the folding equilibrium constants.[34] Interestingly, temperature dependent studies of the folding kinetics, which in principle can provide first access to thermodynamic information on transition states for G4 formation, have yet to be reported in single-molecule experiments.

As a focus of the present work, we use temperature-dependent single-molecule microscopy to investigate the influence of monovalent cations (K^+ and Na^+) on the kinetics and thermodynamics of G4 folding. Specifically, we study the folding of the 5'-GGG-$(TAAGGG)_3$-3' sequence, which is known to form a polymorphic G4.[32] We examine this system using fluorescence resonance energy transfer (FRET) to monitor the dynamics of G4 folding with a single-molecule FRET (smFRET) microscope. Transitions between the polymorphic network of

G4 states are observed in real time, from which we determine the equilibrium constants and rate constants for these transitions. By analyzing the dependence of such folding dynamics on cation concentration and temperature, we obtain first information on the kinetic mechanism for polymorphism control by K^+ and Na^+ as well as the thermodynamics of cation binding to G4.

7.3 Methods

7.3.1 G-quadruplex FRET construct

The single-molecule construct used in these studies consists of two annealed DNA oligomers. The first oligomer is a Cy3-labeled DNA G-quadruplex fold (5'-GGG-$(TAAGGG)_3$-3') connected by 8 thymine residues to an 18-nucleobase duplex-forming domain (full sequence: 5'-TGGCGACGGCAGCGAGGC-T_8- GGG-$(TAAGGG)_3$-Cy3-3'). The second oligomer is complementary to the duplex-forming domain and is labeled with Cy5 and biotin (5'-Cy5-GCCTCGCTGCCGTCGCCA-Biotin-3') for surface tethering by biotin-streptavidin interactions. Both oligomers are purchased in HPLC-purified form from Integrated DNA Technologies and used as is. (Company names listed herein are in the interest of completeness and do not reflect endorsement of a particular vendor.) The oligomers are annealed by mixing 10 μM aliquots (1 M = 1 mol/L) and heating to 85 °C to disrupt intramolecular base-pairing before slowly cooling to room temperature at 1 °C/min. The annealed construct is used without purification, as constructs lacking the biotinylated strand do not adhere to the surface and constructs lacking the Cy3-labeled strand are inefficiently excited by the laser ($\sigma_{Cy3}/\sigma_{Cy5}$ = 3% at 532 nm).

7.3.2 Sample preparation and single-molecule microscopy

Sample preparation and single-molecule microscopy are carried out according to previously described protocols.[35-36] In brief, a glass coverslip is cleaned by 24 hr soaking in

acetone followed by a 30 min treatment with UV-generated ozone. The coverslip surface is incorporated into a flow cell, where it is functionalized by sequential 10-minute exposure to buffered solutions of (i) 10 mg/mL bovine serum albumin (BSA) with 1 mg/mL biotinylated-BSA, (ii) 0.2 mg/mL streptavidin, and (iii) 150 pM smFRET construct. This process produces a surface decorated with approximately one smFRET construct per 10 μm^2. The surface-tethered constructs are then imaged in 50 mM HEPES buffer (pH 7.6) with an oxygen scavenging cocktail to extend observation time (100 nM protocatechuate 3,4-dioxygenase, 5 mM protocatechuic acid) and 2 mM Trolox to increase fluorophore brightness by quenching of fluorophore triplets to the ground state.[37-38] Sodium and potassium levels are controlled by titration with NaCl and KCl solutions, respectively.

The prepared sample is observed on a though-objective total internal reflection fluorescence (TIRF) microscope,[39] which permits 532 nm excitation of Cy3 molecules within ≈100 nm of the surface. Fluorescence from the smFRET constructs is collected and separated by dichroic mirror into Cy3 and Cy5 channels before being directed onto an intensified charge-coupled device (CCD) camera operating at 10 frames per second. Movies are analyzed using software written in LabWindows which identifies particle locations by thresholding and extracts raw donor and acceptor fluorescence trajectories by integration over the local pixel neighborhood for each frame. Background fluorescence levels are determined by fits to a delocalized 2D Gaussian, which permits calculation of background-corrected donor and acceptor trajectories, D(t) and A(t), with FRET trajectories determined from $E_{FRET}(t) = A(t) / (D(t) + A(t))$. Temperature control is achieved through a thermoelectric element in thermal contact with the sample and capable of heating and cooling with 0.1 °C stability, as previously described.[36]

7.3.3 Hidden Markov modeling

Rate constants for conformational transitions are determined by hidden Markov modeling of FRET trajectories. The model consists of *n* states, each with a Gaussian-distributed FRET observable, $P(E_i, \sigma_i) \propto \exp[-(E-E_i)/2\sigma_i^2]$, where E_i and σ_i are the FRET center and width of state *i*, respectively. Transitions between states occur according to first-order kinetics as contained in the rate matrix **K**, where K_{ij} is the rate constant for the transition from state *j* to state *i*, and $K_{ii} = -\Sigma_{j\neq i} K_{ij}$. The transition probability per frame is calculated via the matrix exponential $\mathbf{T} = \exp(\mathbf{K}t_{frame})$, where t_{frame} is the experimental time between adjacent frames (0.1 s). The likelihood function for a single trajectory is therefore described by the matrix product

$$L = \mathbf{1}^T[\textstyle\prod_{i\geq 2} \mathbf{O}(E_i)\mathbf{T}]\mathbf{O}(E_1)\boldsymbol{p}_{eq}, \qquad \text{(Eq. 7.1)}$$

where $\boldsymbol{p}_{eq}$ is the equilibrium probability vector ($\mathbf{K}\cdot\boldsymbol{p}_{eq} = 0$), E_i is the i^{th} observed FRET value, $\mathbf{O}(E)$ is an $n \times n$ diagonal matrix with diagonal entries $O_{ii}(E)$ equal to the probability density of observing the FRET value *E* in the i^{th} state, $\mathbf{1}^T$ is the row vector (1, 1, …) of length equal to the number of states *n*, and the product is taken over all frames in the trajectory except the first. For multiple trajectories, the total likelihood function is the product of the single-trajectory likelihood functions ($L_{total} = \prod_i L_i$). The maximum likelihood estimate of K_{ij}, E_i, and σ_i is obtained by performing gradient ascent on L_{total} using MATLAB, and bootstrapping analysis is performed to estimate uncertainties.[40]

7.3.4 Three-state Hill fit

Population versus concentration data for the G4 are fit to a Hill model. The standard Hill equation describes ligand binding in systems with two distinguishable receptor configurations: a ligand-bound state and a ligand-free state.[41] However, the G4 construct in this work has three observable states, so we have instead used a modified Hill analysis which includes three states. The binding model consists of two competing ligand association processes which convert a

receptor *R* into one of two distinguishable ligand-bound species B_1 or B_2, each with a unique stoichiometry (n_1 and n_2) and dissociation constant ($K_{d,1}$ and $K_{d,2}$):

1: $R \xrightarrow{+n_1 L} B_1$ $\qquad \left(K_{d,1}\right)^{n_1} = \frac{[R]}{[B_1]}[L]^{n_1}$ (Eq. 7.2)

2: $R \xrightarrow{+n_2 L} B_2$ $\qquad \left(K_{d,2}\right)^{n_2} = \frac{[R]}{[B_2]}[L]^{n_2}.$ (Eq. 7.3)

The above dissociation constant expressions can be readily manipulated to yield the fractional populations in each state as a function of ligand concentration:

$$\theta_R([L]) = \frac{[R]}{[R]+[B_1]+[B_2]} = \frac{1}{1+\left(\frac{[L]}{K_{d,1}}\right)^{n_1}+\left(\frac{[L]}{K_{d,2}}\right)^{n_2}} \quad \text{(Eq. 7.4)}$$

$$\theta_{B_1}([L]) = \frac{[B_1]}{[R]+[B_1]+[B_2]} = \frac{\left(\frac{[L]}{K_{d,1}}\right)^{n_1}}{1+\left(\frac{[L]}{K_{d,1}}\right)^{n_1}+\left(\frac{[L]}{K_{d,2}}\right)^{n_2}} \quad \text{(Eq. 7.5)}$$

$$\theta_{B_2}([L]) = \frac{[B_2]}{[R]+[B_1]+[B_2]} = \frac{\left(\frac{[L]}{K_{d,2}}\right)^{n_2}}{1+\left(\frac{[L]}{K_{d,1}}\right)^{n_1}+\left(\frac{[L]}{K_{d,2}}\right)^{n_2}}. \quad \text{(Eq. 7.6)}$$

This model is then applied to data for sodium and potassium titrations of the G-quadruplex, where $L = Na^+$ or K^+, *R* is the low FRET state, B_1 is the middle FRET state, and B_2 is the high FRET state. Dissociation constants ($K_{d,i}$) and apparent stoichiometries (n_i) are determined by simultaneously fitting all three populations to this model using the weighted, nonlinear least-squares regression tools available in OriginPro.

7.3.5 van't Hoff and Arrhenius analyses

Temperature-dependent equilibrium constants determined from hidden Markov modeling are then subjected to van't Hoff analysis,[42-43] whereby the Gibbs free energy

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -RT \ln\left(K_{eq}\right) \quad \text{(Eq. 7.7)}$$

is rewritten in the van't Hoff form

$$\ln\left(K_{eq}\right) = \frac{-\Delta H^{\circ}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\circ}}{R}, \qquad \text{(Eq. 7.8)}$$

with K_{eq} as the equilibrium constant, R the gas constant (8.314 J/molK), T the absolute temperature, ΔH° the reaction enthalpy, and ΔS° the reaction entropy. In accord with such a van't Hoff model, we assume ΔH° and ΔS° to be approximately constant over the temperature range explored, with a linear fit of $\ln(K_{eq})$ vs. 1/T to extract ΔH° and ΔS° via the slope and intercept, respectively. In analogous fashion, the temperature-dependent rate constants can be analyzed in the context of Kramer's theory[44-46] to yield an Arrhenius-type expression

$$\ln\left(k/\upsilon\right) = \frac{-\Delta H^{\ddagger}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\ddagger}}{R}, \qquad \text{(Eq. 7.9)}$$

where k is the rate constant, υ is the attempt frequency along the reaction coordinate, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are the activation enthalpy and entropy, respectively. The attempt frequency can be approximately estimated as 1×10^{6} s^{-1} from previous RNA folding studies,[44-45] but its actual value is dependent on the structure of the free energy landscape. As a consequence, the activation entropy $\Delta S^{\ddagger}$ is not determined absolutely, though any reported differential changes in this entropy ($\Delta\Delta S^{\ddagger}$), for instance, due to a change in ionic conditions, are rigorously independent of the choice of υ.

7.3.6 Prediction of FRET efficiencies

FRET efficiencies of the proposed G-quadruplex structures are predicted by combining structural data on G-quadruplex folds with the worm-like chain (WLC) model[47] applied to single-stranded regions of the DNA construct. The WLC result for the average square end-to-end distance $\langle R^2 \rangle$ of a polymer of contour length L and persistence length L_0 is

$$\langle R^2 \rangle = 2L_0 L\left[1 - \frac{L_0}{L}\left(1 - e^{-L/L_0}\right)\right]. \qquad \text{(Eq. 7.10)}$$

Literature-reported values for single-stranded polyT at 100 mM ionic strength[48-49] vary from L = 5−7 Å per nucleotide (nt) and L_0 = 15−25 Å. If we assume average values of $\langle L \rangle$ = 6 ± 1 Å/nt

and L_0 = 20 ± 5 Å, the dT_8 spacer in the single-molecule construct would be expected to exhibit a root-mean-square (RMS) end-to-end distance of R_{rms} = 35 ± 5 Å.

We then consider states with a fully folded G-quadruplex, for which there are multiple folding topologies. For a G-quadruplex topology in which the first and last guanine residue are part of the same G-tetrad, as in an antiparallel G-quadruplex fold, the Cy3 and Cy5 fluorophores will be positioned especially close to each other (here labeled the *proximal* configuration). For such a proximal topology, the average Cy3-Cy5 distance is approximately the length of the dT_8 ssDNA spacer in the construct (R_{rms} = 35 ± 5 Å), yielding an expected FRET efficiency of $E_{proximal}$ = 0.89. On the other hand, a topology in which the first and last guanine residues are on opposing sides of the G-quadruplex, such as a parallel topology or a hybrid topology, the Cy3 and Cy5 are further separated by the G-quadruplex fold (labeled *distal* topology). Based on the crystal structure of a similar 3-tetrad G-quadruplex sequence as reference (PDB 1K8P),[17] the G-quadruplex fold in the smFRET construct may be approximated as a rigid cylinder of diameter 18 Å and length 7 Å, and adding the cylinder length to the R_{rms} of dT_8 yields E_{distal} = 0.75. We approximate the fully unfolded GQ as having the same persistence length as polythymine to give E_{FRET} = 0.07, with the accuracy of this assumption limited by sequence dependence of the ssDNA persistence lengths.

7.4 Results

7.4.1 G-quadruplex smFRET construct exhibits multistate folding

To probe the thermodynamics of G-quadruplex folding at the single-molecule level, we have designed a fluorophore-labeled DNA construct for FRET microscopy based on the smFRET construct used by Tippana et al. (Figure 7.1A).[32] The folding domain of the construct

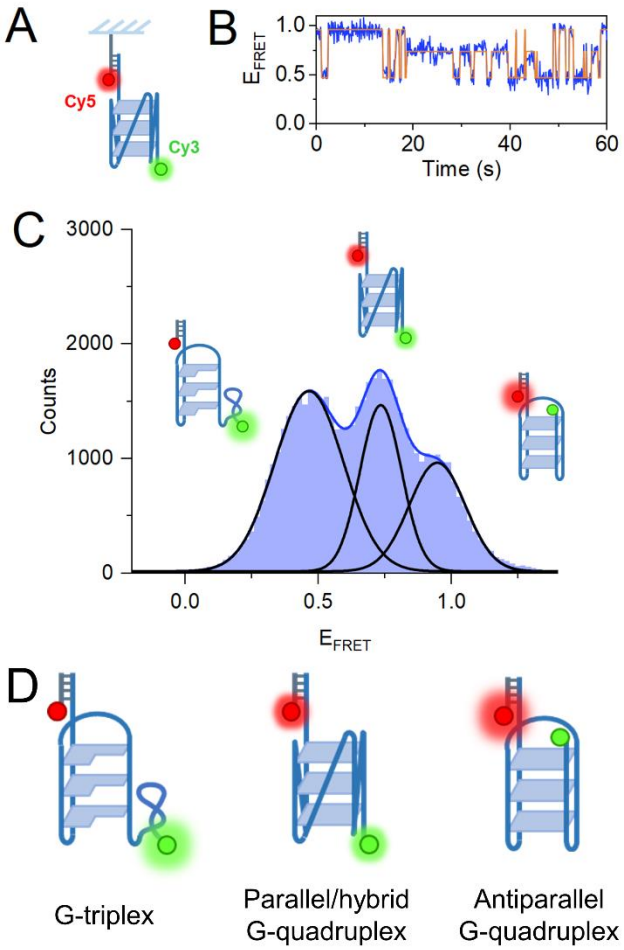

Figure 7.1 (A) G-quadruplex smFRET construct design, labeled with Cy3 and Cy5 for FRET detection of conformational dynamics. (B) Sample smFRET trajectory taken in buffer with 100 mM Na^+. Fitted state trajectory from hidden Markov modeling with three states overlaid in orange. (C) FRET histogram (blue) aggregated from N = 83 trajectories. Gaussian fits (black) are used to the FRET efficiencies and relative abundance of the three states. (D) Cartoons indicate putative structures of the three states: a G-triplex (E_{FRET} = 0.44) and two G-quadruplex topologies, one with Cy3 and Cy5 in a distal arrangement (E_{FRET} = 0.73), and one in a proximal arrangement (E_{FRET} = 0.97). See text for details.

consists of a repeated six-nucleotide sequence (5'-TAAGGG-3') which forms a G-quadruplex with three G-tetrads. The construct is labeled with the Cy3-Cy5 FRET pair[38] such that folding of the G-quadruplex brings Cy3 and Cy5 in closer proximity, increasing the energy transfer efficiency (E_{FRET}) between the fluorophores. The full construct is surface-tethered by biotin-streptavidin interactions[50] to enable observation of the G-quadruplex in a single-molecule fluorescence microscope. Temporal trajectories in 100 mM Na^+ buffer reveal multiple, distinct

E_{FRET} states which interconvert/equilibrate on the 100 ms to second timescale (Figure 7.1B). A histogram of E_{FRET} values gathered from $N = 83$ molecules is well-fit as a sum of three Gaussian distributions ($E_{low} = 0.44$, $E_{middle} = 0.73$, $E_{high} = 0.97$), which corresponds to three construct conformations with resolved interfluorophore distances (Figure 7.1C). This observation of conformational polymorphism is consistent the work of Tippana et. al on the same G4-forming sequence which also exhibited the formation of three FRET states.[32]

Structural assignment of the three observed FRET states is performed by comparison to predicted E_{FRET} values for the smFRET construct under several different G-quadruplex folding topologies. Our approach is to combine structural data on G-quadruplex folding with a worm-like chain (WLC) treatment of single-stranded regions of the construct (see Methods). This model is simple but provides a reasonable starting point for consideration of possible construct structures, which could be refined by methods such as molecular dynamics simulations.[51] The G-quadruplex in the construct can fold to place the Cy3 and Cy5 fluorophores on the same side of the G-quadruplex (a *proximal* configuration, as in an antiparallel topology) or on opposite sides of the G-quadruplex (a *distal* configuration, as in a parallel or hybrid topology), leading to two distinguishably different FRET efficiencies ($E_{proximal} = 0.89$, $E_{distal} = 0.75$). We next consider contributions from the fully unfolded G-quadruplex construct, in which the Cy3 and Cy5 are separated by a 29-nucleotide ssDNA region, but the calculated FRET efficiency ($E_{unfolded} = 0.10$) is much smaller than the lowest FRET states experimentally observed ($E_{low} = 0.44$). Therefore, we instead consider a partially (un)folded construct in which one of the G-tracts does not participate in the G-quadruplex, also called a G-triplex (G3). Previous work by Hou et al. demonstrates that G3 structures are populated, metastable folding intermediates in smFRET constructs containing G-quadruplex sequences.[15] We calculate a FRET efficiency $E_{FRET} = 0.40$

or 0.46 for the G3 in a proximal or distal topology, respectively. Therefore, we tentatively assign the E_{low} state as G3 (*T*), the E_{middle} state as parallel/hybrid G4 (*P*), and the E_{high} state as antiparallel G4 (*A*) and include these three putative structures in Fig. 1D.

7.4.2 Sodium ions drive the folding equilibrium to the parallel/hybrid topology

We first explore G-quadruplex folding as a function of Na^+ concentration at ambient temperature (23 °C), which from the histograms in Figure 7.2A reveal an overall shift from the G3 state to the parallel/hybrid and antiparallel G4 topologies with increasing [Na^+]. Integration over these histogram peaks yields the subpopulations over 50−300 mM Na^+ concentration (Fig. 2B), which can then be fit globally to a 3-state Hill model (Eqs. 7.4−7.6) that explicitly includes all 3 interconversion routes. The least squares fits to such a multistate model are shown as solid lines in Fig. 2B, yielding the stoichiometry for Na^+ binding (Δn) and associated dissociation constants (K_d) for each interconversion pathway, as summarized in Table 7.1. Interestingly, the G3 state gains only one Na^+ ion when folding into either G4 state ($\Delta n_{T\rightarrow P} \approx \Delta n_{T\rightarrow A} \approx 0.87 \pm 0.09$), though Na^+ has a 2-fold stronger affinity for parallel/hybrid G4 (K_d = 132 ± 9 mM) than for antiparallel G4 (K_d = 228 ± 26 mM). Conversely, there is no difference within experimental uncertainty in the number of Na^+ bound to the two G4 conformations. With the G4 construct

		Na^+	K^+
$E_{low} \rightarrow E_{middle}$	Δn	0.87 ± 0.09	-0.05 ± 0.05
	K_d (mM)	132 ± 9	—
$E_{low} \rightarrow E_{high}$	Δn	0.87 ± 0.09	0.62 ± 0.05
	K_d (mM)	228 ± 26	1.9 ± 0.2
$E_{middle} \rightarrow E_{high}$	Δn	0.00 ± 0.09	0.67 ± 0.07
	K_d (mM)	—	2.9 ± 0.3

Table 7.1 Results of fitting G-quadruplex population data vs. [Na^+] and [K^+] to 3-state Hill model. Uncertainties are reported as standard errors of the mean. For processes with Hill coefficients (Δn) equal to zero within uncertainty, dissociation constants (K_d) are not reported.

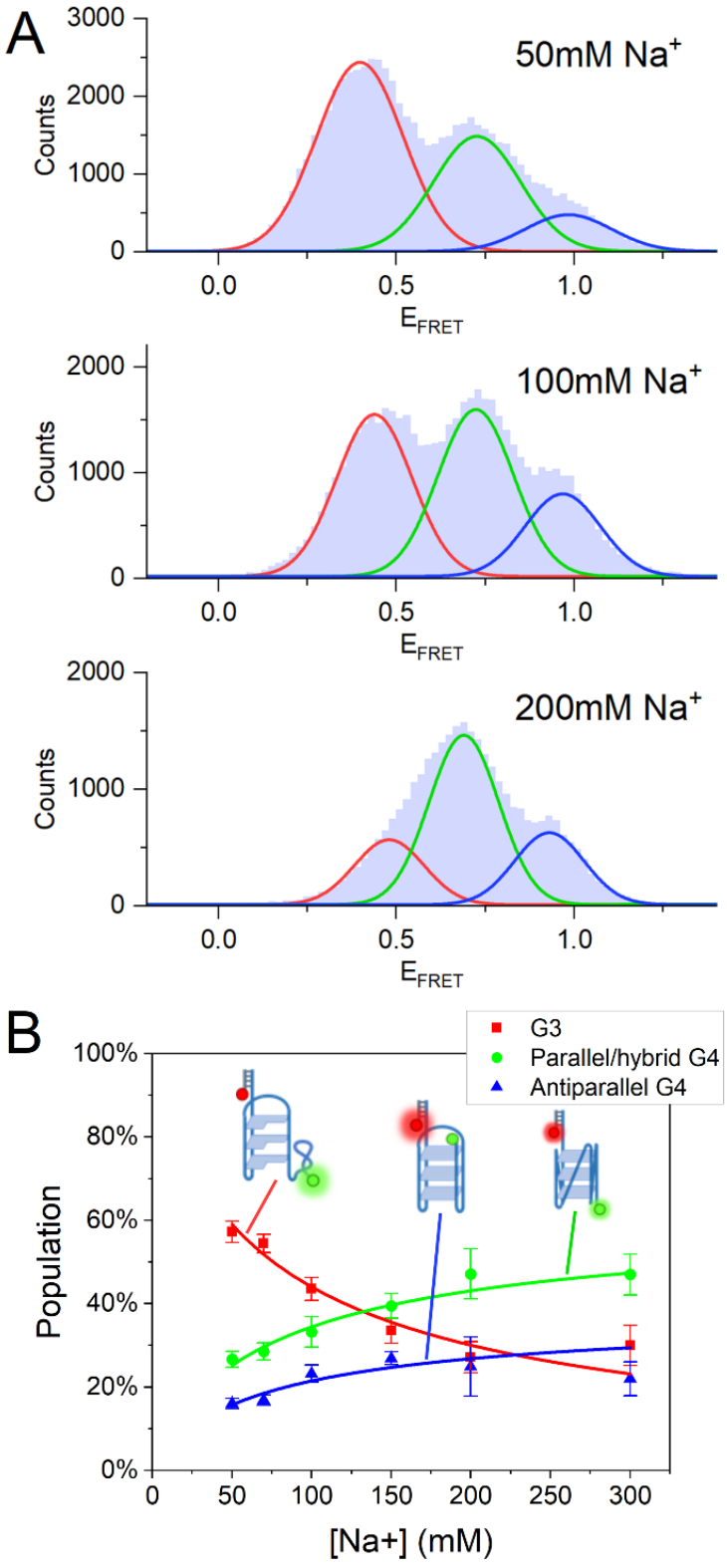

Figure 7.2 Influence of Na^+ on G4 folding equilibrium. (A) FRET histograms at 50, 100, and 200 mM Na^+. (B) Population of the three G4 conformational states in [Na^+] = 50−300 mM. Fits to a 3-state Hill model shown as solid curves.

populating both the parallel/hybrid and antiparallel topologies even at [Na^+] = 300 mM, it is clear that Na^+ has only a modest impact on G4 polymorphism.

Additional mechanistic insight can be obtained from kinetic analysis of the Na^+-dependence of G-quadruplex folding events in the time domain. Rate constants for folding are determined by hidden Markov modeling (HMM) of smFRET trajectories[52-54] using a 3-state model which includes all 6 rate constants describing forward and reverse interconversion between each FRET pair (Figure 7.3A). The rate constants for G4 folding as a function of [Na^+] = 50−300 mM are shown in Figure 7.3B. As a rigorous check on the validity of these extracted rate constants, we can calculate the free energy sum around the three-state cycle,

$$\Delta G_{cycle} = -kT \ln\left(\frac{k_{T\to P}k_{P\to A}k_{A\to T}}{k_{P\to T}k_{A\to P}k_{T\to A}}\right), \qquad \text{(Eq. 7.11)}$$

at each Na^+ concentration and plot the values in Figure 7.3C. By detailed balance, the free energy sum around any cycle of states must equal zero within uncertainty, and indeed, the calculated $\Delta G_{cycle}/kT$ values vanish at all Na^+ concentrations. In Figure 7.3B, the rate constants for the conversion of G3 to and from G4 conformations have the largest magnitude, while transitions directly between the parallel/hybrid and antiparallel G4 topologies are a minor, though not forbidden, kinetic process. Dependence on [Na^+] is most evident in the "forward" folding rate constants, $k_{T\to P}$ and $k_{T\to A}$, while the "reverse" unfolding rate constants, $k_{P\to T}$ and $k_{A\to T}$, appear less impacted by Na^+ concentration.

To additionally explore the effects of Na^+ cation on G-quadruplex folding, we have further developed an explicit 6-state kinetic model for Na^+ binding to the G-quadruplex (Figure 7.3D). The model consists of three DNA conformations, each with ligand-free and ligand-bound forms. In accordance with 1:1 Na^+ binding stoichiometry from the Hill fit of the population data (Table 7.1), the kinetic model includes a single Na^+ ion in each ligand-bound state. Ligand

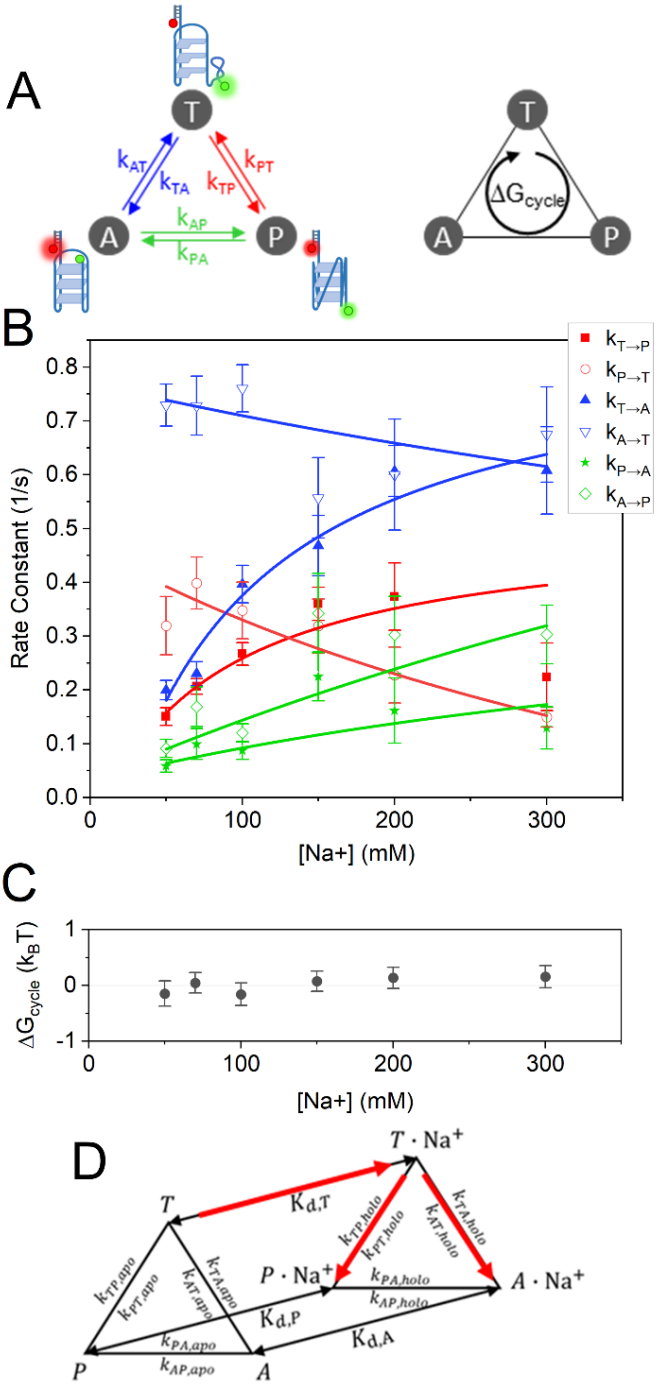

Figure 7.3 Influence of Na^+ on G4 folding kinetics. (A) Left schematic: structure of hidden Markov model (HMM) used to fit smFRET trajectories to obtain G4 folding rate constants. Right schematic: the orientation of the 3-state cycle used to calculate the free energy change around the loop, $\Delta G_{cycle} = \Delta G_{T\rightarrow P} + \Delta G_{P\rightarrow A} + \Delta G_{A\rightarrow T}$. (B) The rate constants for G4 folding in 50−300 mM Na^+. Fits to the 6-state Na^+-binding model depicted in (D) are shown as solid curves. (C) Plot of ΔG_{cycle} at each [Na^+]. To satisfy detailed balance, ΔG_{cycle} should equal zero. (D) 6-state binding model for Na^+ influence on G4 folding kinetics. The dominant pathways through this network are shown as red arrows, which coincide with a "bind-then-fold" mechanism.

binding is assumed to happen rapidly compared to G4 conformational dynamics, which allows all ligand-binding steps to be treated as a quasi-equilibrium governed by a single binding constant

$$K_{d,i} = \frac{[C_i]}{[C_i \cdot L]}[L], \quad \text{(Eq. 7.12)}$$

where *i* labels the conformational state (triplex *T*, parallel/hybrid G4 *P*, or antiparallel G4 *A*), C_i is the ligand-free conformation, $C_i \cdot L$ is the ligand-bound conformation, *L* is the ligand (in this case, Na^+), and square brackets indicate concentration of the enclosed species. The fraction of ligand-bound molecules is therefore

$$\theta_{bound,i} = \frac{[C_i \cdot L]}{[C_i]+[C_i \cdot L]} = \frac{[L]}{[L]+K_{d,i}}. \quad \text{(Eq. 7.13)}$$

Transitions between conformations are governed by first order rate constants depending on ligand occupancy:

$$C_i \xrightarrow{k_{ij,apo}} C_j \quad \text{(Eq. 7.14)}$$

$$C_i \cdot L \xrightarrow{k_{ij,holo}} C_j \cdot L, \quad \text{(Eq. 7.15)}$$

where $k_{ij,apo}$ and $k_{ij,holo}$ are rate constants for transition from state *i* to state *j* in the ligand-free ("*apo*") and ligand-bound ("*holo*") forms, respectively. The observed rate constants $k_{ij,obs}$ for transitions between conformational states are population-weighted averages of the *apo* and *holo* rate constants:

$$k_{ij,obs} = k_{ij,apo}\left(1 - \frac{[L]}{[L]+K_{d,i}}\right) + k_{ij,holo}\left(\frac{[L]}{[L]+K_{d,i}}\right). \quad \text{(Eq. 7.16)}$$

For the 6-state kinetic model in Figure 7.3D, Eq. 7.16 forms a system of 6 equations, where the number of free parameters in the model can be reduced from 15 to 11 by imposing detailed balance constraints ($\Delta G_{cycle} = 0$) on all cycles. The resulting constrained equations are then globally fit to the kinetic data in Figure 7.3C, as summarized in solid line fits.

The kinetic behavior G-quadruplex folding reveals that Na^+ drives folding by binding to the G3 state which subsequently folds into either G4 conformation in a "bind-then-fold" mechanism (Figure 7.3D, pathway highlighted in red).[55-57] There are several key indicators in support of this mechanism: first, the forward folding rate constants $k_{T\rightarrow P}$ and $k_{T\rightarrow A}$ vanish at low Na^+ (Figure 7.3B, filled red and blue symbols). Folding, evidently, is impossible without Na^+ ($k_{TP,apo} = k_{TA,apo} = 0\ s^{-1}$). Additionally, $k_{T\rightarrow P}$ and $k_{T\rightarrow A}$ saturate at high [Na^+] ($k_{TP,holo} = 0.7 \pm 0.3\ s^{-1}$, $k_{TA,holo} = 1.0 \pm 0.4\ s^{-1}$), which indicates that binding of Na^+ to G3 is driven to completion. However, there is a clear albeit weaker dependence of the unfolding rate constants $k_{P\rightarrow T}$ and $k_{A\rightarrow T}$ on [Na^+], so other pathways besides the bind-then-fold pathway are permitted to a lesser degree.

7.4.3 K^+ has a much higher affinity than Na^+ for the G-quadruplex and exclusively promotes the antiparallel topology

Formation of G-quadruplexes is known to be particularly sensitive to K^+, which we next explore to measure K^+ vs Na^+ specific effects on G4 folding (Figure 7.4). Interestingly, K^+ drives the conformational equilibrium of the G4 smFRET construct exclusively toward the antiparallel G4 topology (Figure 7.4A), in essence by depleting both the G3 and parallel/hybrid G4 states. Furthermore, this dramatic increase in folding occurs at only a few mM K^+ concentrations, in contrast to the 100-fold higher concentrations of Na^+ required (Figure 7.2B). Clearly, the G4 construct has a much higher affinity for K^+ than Na^+ but only in achieving the antiparallel G4 conformation. Fitting the full suite of K+ dependent population data to a 3-state Hill model quantitatively confirms these qualitative observations, as summarized in Table 7.1. In this case, no K^+ uptake occurs during the transition from the triplex to the parallel/hybrid G4 ($\Delta n_{T\rightarrow P} = -0.05 \pm 0.05$), which indicates that K^+ does not participate in this process. Meanwhile, approximately one K^+ binds to during folding into the antiparallel G4 ($\Delta n_{T\rightarrow A} = 0.62 \pm 0.05$)

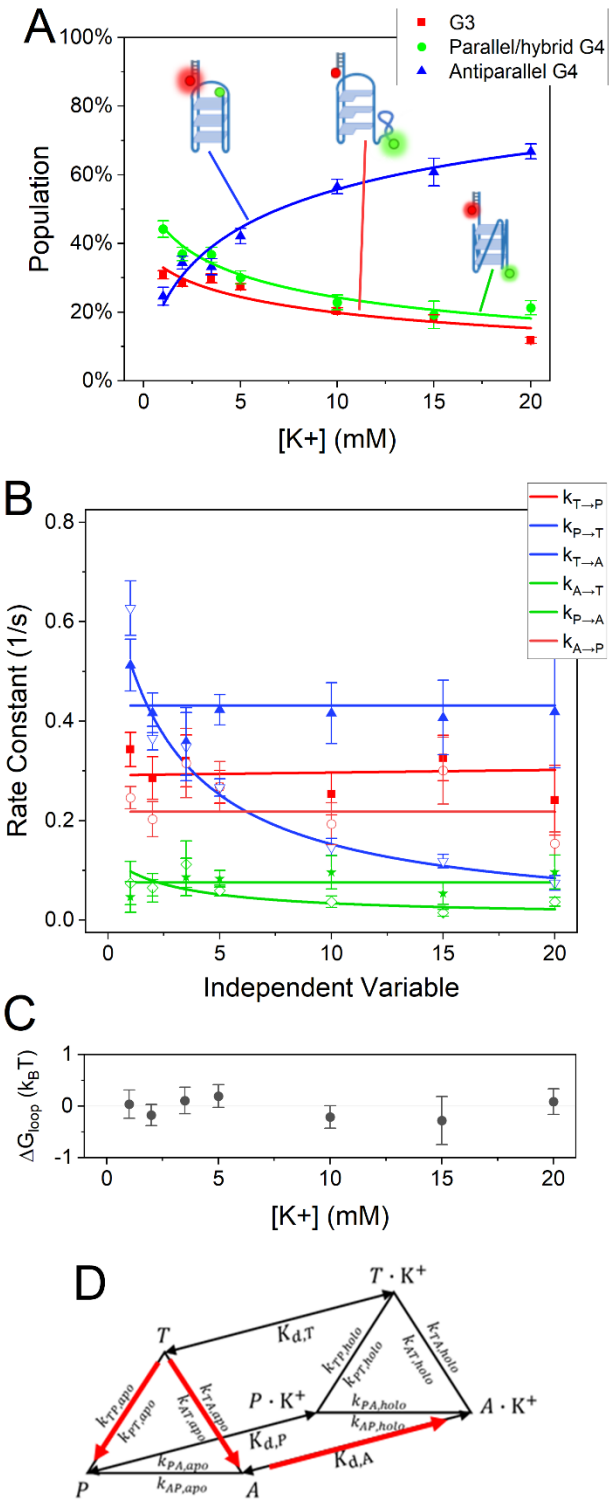

Figure 7.4 Influence of K^+ on G4 folding. State populations (A) and folding rate constants (B) determined in 100 mM Na^+ with 0–20 mM K^+. Data are fit to a 3-state Hill model in (A) and a 6-state ligand binding model in (B). (C) Calculation of ΔG_{cycle} at each K^+ concentration to check whether the rate constants satisfy detailed balance. (D) 6-state model of K^+ binding to G4. Red arrws indicate the highest flux pathways ("fold-then-bind" mechanism).

with a binding affinity (K_d = 1.9 ± 0.2 mM) that is 2 orders of magnitude greater than Na^+ in achieving the same conformational state (K_d = 228 ± 26 mM). That the Hill coefficient is less than unity indicates that the antiparallel G4 can form in the absence of K^+ cation, consistent with the observation of antiparallel G4 in Na^+-only buffer (Figure 7.2). By driving the G4 construct exclusively to the antiparallel topology, K^+ exerts much stronger control of G4 polymorphism than Na^+.

Similar to analysis of the Na^+ kinetics, smFRET trajectories are subjected to hidden Markov modeling to extract rate constants as a function of [K^+] (Figure 7.4B). As before, we can rigorously confirm these kinetic results by verifying that the rate constants explicitly obey detailed balance (ΔG_{cycle} = 0) at all K^+ concentrations (Figure 7.4C). Five of the six folding rate constants show little to no dependence on K^+ concentration, while the rate constant for unfolding from the antiparallel G4 to G3 ($k_{A\rightarrow T}$) is strongly suppressed by the presence of K^+. To explore the mechanistic implications of these data, the folding rate constants are fit to the 6-state kinetic model (Eq. 7.16) and overlaid on Figure 7.4B as solid lines. The fit correctly predicts that $k_{A\rightarrow T}$ goes to zero in the high [K^+] limit ($k_{AT,holo}$ = 0.01 ± 0.03 s^{-1}), consistent with a conformational selection (CS, "fold-then-bind") model in which the *apo* DNA pre-folds to a transient antiparallel G4 state before binding K^+ with high affinity ($K_{d,A}$ = 2.6 ± 1.2 mM) to stabilize the antiparallel configuration. This K^+-induced folding pathway (highlighted in Figure 7.4D) contrasts strongly from the induced fit (IF, "bind-then-fold") mechanism seen for Na^+-stabilized formation of the G4 quadruplex and originates from a much stronger and topology-specific binding interaction.

7.4.4 Temperature-dependence of G-quadruplex folding

Temperature-controlled single-molecule experiments can be used to deconstruct the overall free energy of G-quadruplex folding into enthalpic and entropic contributions.[43] The

temperature-dependence of the folding rate constants and equilibrium constants in 100 mM Na^+ over a temperature range of 13−29 °C is captured in Figure 7.5. We note that this is, to our knowledge, the first use of smFRET to examine temperature-dependent kinetics of a system with more than two conformations. We observe first that the majority of folding events start or end in the G3 state (Fig. 5A, B, left), with folding/unfolding between the parallel/hybrid G4 and antiparallel G4 being a minor kinetic process at all temperatures (Fig. 5C, left). Second, the equilibrium constants for folding from the G3 state to both G4 states ($K_{eq,TP}$ and $K_{eq,TA}$) decrease at high temperature, which shows that the overall folding process is enthalpically favorable. The kinetic origin of this behavior can be understood by examining the corresponding rate constants:

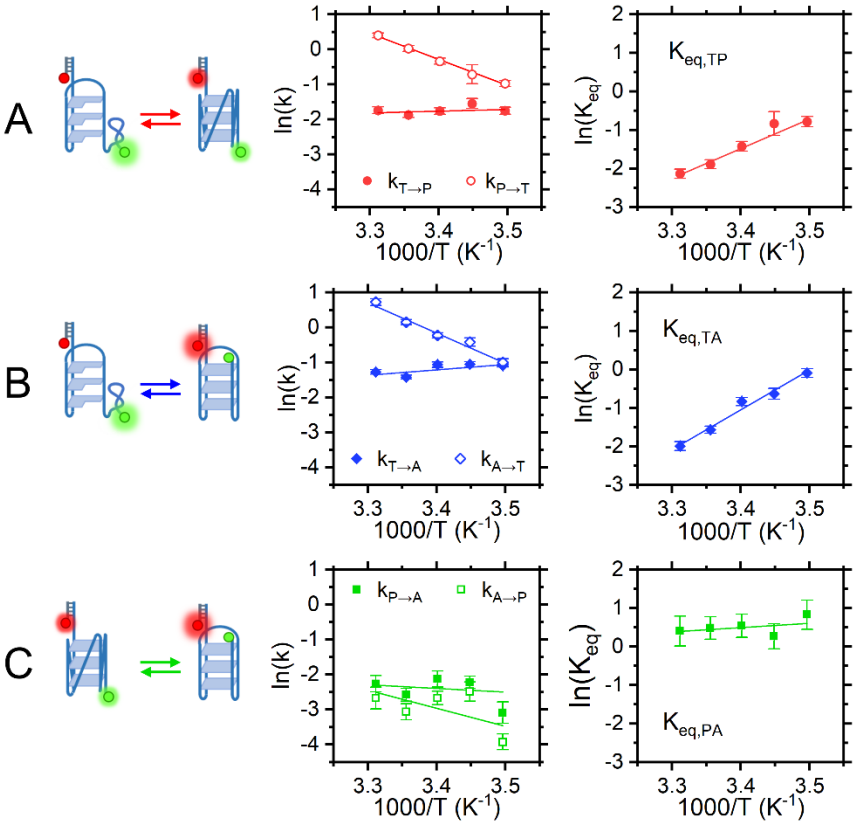

Figure 7.5 Temperature-dependence of G4 folding rate constants (left panels) and equilibrium constants (right panels) in 100 mM Na^+ for (A) G3 → parallel/hybrid G4, (B) G3 → antiparallel G4, and (C) parallel/hybrid G4 → antiparallel G4. Lines indicate Arrhenius and van't Hoff fits for rate constants and equilibrium constants, respectively, which are used to extract the enthalpy and entropy change for each process.

the forward rate constants ($k_{T\rightarrow P}$ and $k_{T\rightarrow A}$) are largely temperature insensitive, indicating negligible activation enthalpy, while the reverse rate constants ($k_{P\rightarrow T}$ and $k_{A\rightarrow T}$) are strongly increased at high temperature, indicating an enthalpic barrier for unfolding.

A more quantitative analysis is carried out by fitting rate and equilibrium constants to Arrhenius and van't Hoff models, respectively, to determine overall and transition state enthalpies (ΔH^0, $\Delta H^{\ddagger}$, from slopes) and entropies (ΔS^0, $\Delta S^{\ddagger}$, from intercepts) from the linear fits. The thermodynamic values obtained from such 3-state fits can in turn be used to construct 3D folding energy landscapes (see Fig. 6) which depicts the enthalpy, entropy, and free energy for each of the G-quadruplex conformations and the intervening transition states between them. The enthalpic landscape (Figure 7.6A) shows that both folding processes (G3→ parallel/hybrid G4 and G3 → antiparallel G4) are enthalpically favorable ($\Delta H°_{TP} = -66 \pm 7$ kJ/mol, $\Delta H°_{TA} = -84 \pm 9$ kJ/mol) with a negligible barrier ($\Delta H^{\ddagger}_{TP} = -4 \pm 6$ kJ/mol, $\Delta H^{\ddagger}_{TA} = -13 \pm 6$ kJ/mol). Meanwhile, the enthalpy landscape (Figure 7.6B) indicates that folding is entropically costly ($\Delta S°_{TP} = -240 \pm 20$ J/mol K, $\Delta S°_{TA} = -300 \pm 30$ J/mol K), where the entropic penalty is partially paid upon reaching the transition state ($\Delta S^{\ddagger}_{TP} = -140 \pm 20$ J/mol K, $\Delta S^{\ddagger}_{TA} = -170 \pm 20$ J/mol K). The overall free energy changes for folding (Figure 7.6C) are small in comparison to the individual contributions from enthalpy (ΔH) and entropy ($-T\Delta S$) which nearly cancel, an example of enthalpy-entropy compensation.[58-59]

7.4.5 Thermodynamics of G-quadruplex folding at elevated monovalent cation concentrations

We have repeated these temperature-dependent studies of the G-quadruplex under multiple cation conditions to probe the effects of Na^+ and K^+ cations on the folding thermodynamics. The temperature dependence for G-quadruplex folding under ambient (100

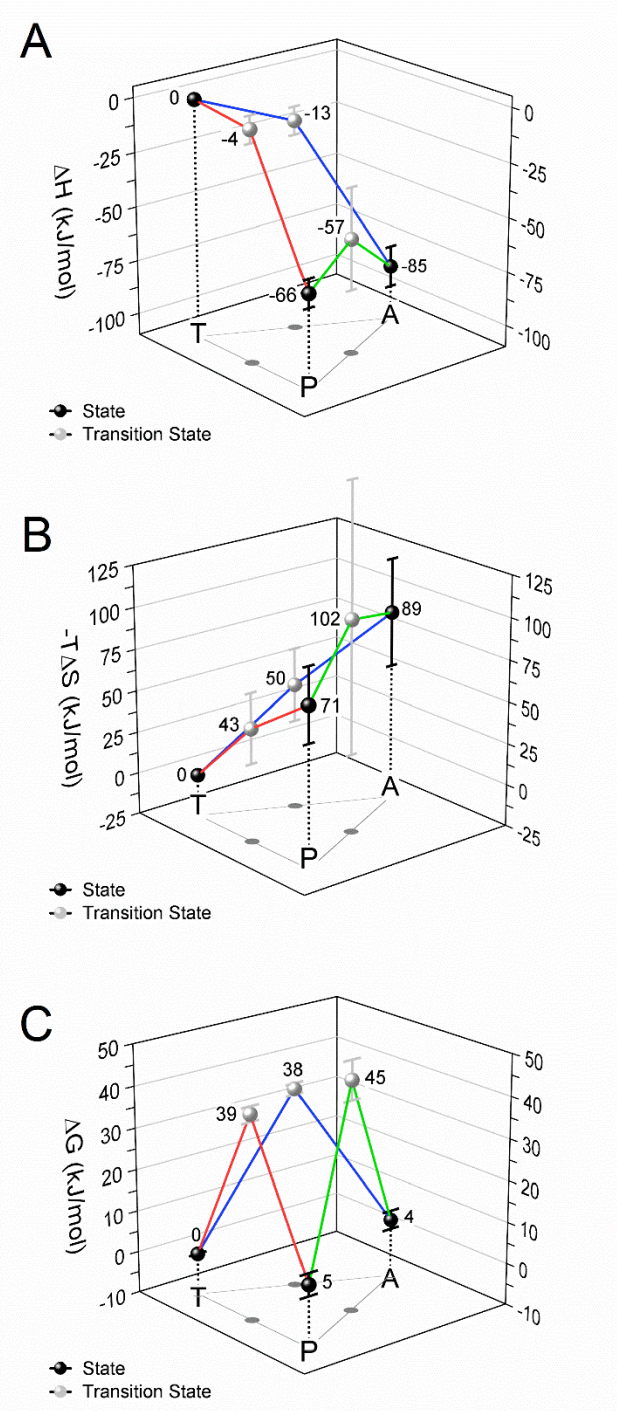

Figure 7.6 Folding landscape for G4 in 100 mM Na^+ as determined by van't Hoff and Arrhenius analyses. (A) Enthalpy of each stable conformation (black points) and transition state (grey points) defined relative to the G3 state. (B) same as (A), but for the enthalpic contribution to the free energy ($-T\Delta S$) evaluated at 300 K. (C) same as (A), but for the free energy ($\Delta G = \Delta H - T\Delta S$) evaluated at 300 K.

mM) Na^+, high (200 mM) Na^+, and 100 mM Na^+ with 4 mM K^+ are shown in Figure 7.7, where the latter two data sets can be compared to the first to isolate impact due to Na^+ and K^+, respectively. Arrhenius and van't Hoff analyses are performed on the datasets, and the fitted entropic and enthalpic parameters for the three cationic conditions (100 mM Na^+, 200 mM Na^+, and 100 mM Na^+ with 4 mM K^+) are shown side-by-side in Table 7.2.

The dependence of folding enthalpies and entropies on cation concentration can be used to determine the thermodynamic signature of Na^+ and K^+ binding. For example, consider the change in the folding enthalpy of the G3 to antiparallel G4 transition upon an increase in [Na^+] from 100 mM to 200 mM, defined as

$$\Delta\Delta H°_{TA}(Na^+) \equiv \Delta H°_{TA}(200\ \text{mM}\ Na^+) - \Delta H°_{TA}(100\ \text{mM}\ Na^+). \qquad \text{(Eq. 7.17)}$$

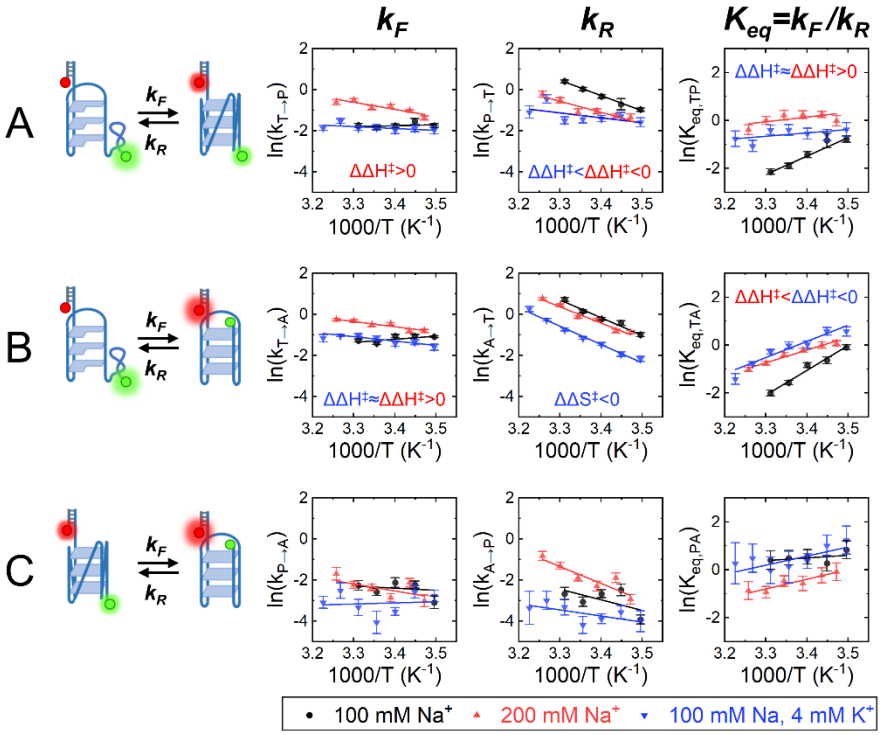

Figure 7.7 Temperature-response of G4 folding under multiple cationic conditions (100 mM Na^+ in black, 200 mM Na^+ in red, and 100 mM Na^+ with 4 mM K^+ in blue) for the three G4 dynamical processes: (A) G3 → parallel/hybrid G4, (B) G3 → antiparallel G4, and (C) parallel/hybrid G4 → antiparallel G4. Arrhenius plots for the forward and reverse rate constants are shown in the left and middle panels, respectively, while the van't Hoff plot for the equilibrium constants are in the right panel.

Our results show that this quantity is positive in value ($\Delta\Delta H°_{TA}(Na^+)$ = 40 ± 11 kJ/mol). Na^+ interacts with the G4 construct through a bind-then-fold mechanism, for which $\Delta\Delta H°_{TA}(Na^+)$ is opposite in sign to the enthalpy of Na^+ binding, $\Delta H°_{bind}(Na^+)$, as depicted in Figure 7.8A. Therefore, since $\Delta\Delta H°_{TA}(Na^+) > 0$, we infer that Na^+ binding is enthalpically favorable ($\Delta H°_{bind}(Na^+) < 0$). We also observe that the Na^+-induced change in folding entropy for this process is positive ($\Delta\Delta S°_{TA}(Na^+)$ = 140 ± 40 J/mol K), so by a similar logic Na^+ binding must

ΔH (kJ/mol)		100 mM Na^+	200 mM Na^+	100 mM Na^+, 4 mM K^+
G3 → G4 (parallel/hybrid)	$\Delta H°_{TP}$	-66 ± 7	-16 ± 9*	-12 ± 14*
	$\Delta H^{‡}_{TP}$	-4 ± 6	29 ± 6*	8 ± 6
	$\Delta H^{‡}_{PT}$	62 ± 2	45 ± 7*	20 ± 13*
G3 → G4 (antiparallel)	$\Delta H°_{TA}$	-85 ± 9	-44 ± 7*	-56 ± 6*
	$\Delta H^{‡}_{TA}$	-13 ± 6	21 ± 4*	18 ± 4*
	$\Delta H^{‡}_{AT}$	72 ± 7	66 ± 6	74 ± 4
G4 (parallel/hybrid) → G4 (antiparallel)	$\Delta H°_{PA}$	-34 ± 40	-40 ± 25	-29 ± 24
	$\Delta H^{‡}_{PA}$	9 ± 23	28 ± 19	-5 ± 18
	$\Delta H^{‡}_{AP}$	43 ± 33	68 ± 15	25 ± 16

ΔS (J/mol K)		100 mM Na^+	200 mM Na^+	100 mM Na^+, 4 mM K^+
G3 → G4 (parallel/hybrid)	$\Delta S°_{TP}$	-240 ± 20	-50 ± 30*	-50 ± 50*
	$\Delta S^{‡}_{TP}$	-140 ± 20	-30 ± 20*	-100 ± 20
	$\Delta S^{‡}_{PT}$	90 ± 10	30 ± 20*	-60 ± 40*
G3 → G4 (antiparallel)	$\Delta S°_{TA}$	-300 ± 30	-150 ± 20*	-190 ± 20*
	$\Delta S^{‡}_{TA}$	-170 ± 20	-50 ± 10*	-70 ± 10*
	$\Delta S^{‡}_{AT}$	130 ± 20	100 ± 20	120 ± 10
G4 (parallel/hybrid) → G4 (antiparallel)	$\Delta S°_{PA}$	-110 ± 140	-140 ± 80	-90 ± 80
	$\Delta S^{‡}_{PA}$	-100 ± 80	-40 ± 70	-160 ± 60
	$\Delta S^{‡}_{AP}$	10 ± 110	100 ± 50	-60 ± 50

Table 7.2 Thermodynamic parameters for G-quadruplex folding determined by van't Hoff and Arrhenius analysis of temperature dependent kinetic measurements at multiple cation concentrations (Figure 7.7). Uncertainties are reported as standard errors of the mean. Asterisks indicate statistical significance (evaluated at the p = 0.01 level) of parameters when compared to the 100 mM Na^+ case.

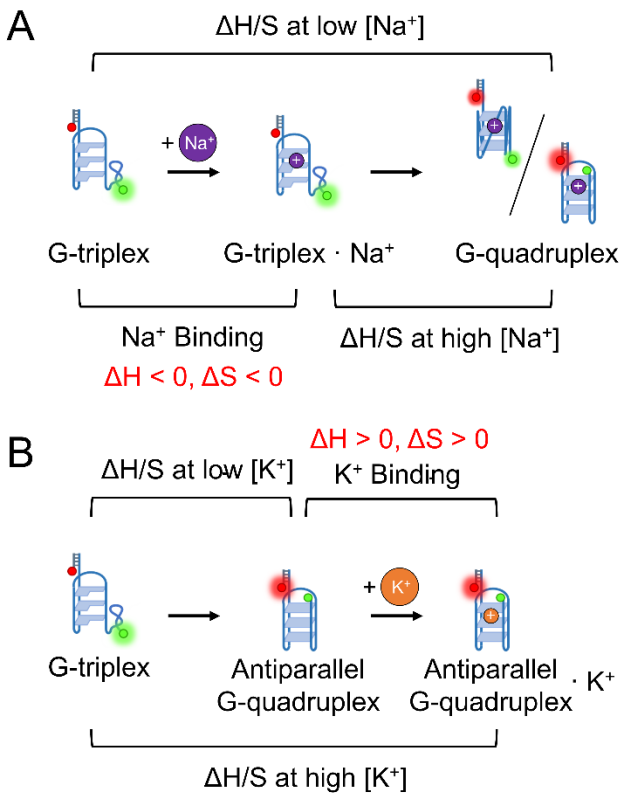

Figure 7.8 Determination of cation binding thermodynamics. (A) For Na^+, which follows a bind-then-fold model, the measured enthalpy/entropy ΔH/S at low [Na^+] contains contributions from both Na^+ binding ($\Delta H/S_{bind}$) and the conformational change ($\Delta H/S_{conf}$). However, at high [Na^+], the G3 conformation saturates with Na^+, and ΔH/S reflects only $\Delta H/S_{conf}$. Therefore, $\Delta\Delta H/S = \Delta H/S(\text{high } [Na^+]) - \Delta H/S(\text{low } [Na^+])$ has the *opposite* sign of $\Delta H/S_{bind}$. (B) For K+, which operates by a fold-then-bind mechanism, the differential enthalpy/entropy $\Delta\Delta H/S = \Delta H/S(\text{high } [K^+]) - \Delta H/S(\text{low } [K^+])$ has the *same* sign of $\Delta H/S_{bind}$.

result in a loss of entropy ($\Delta S°_{bind}(Na^+) < 0$). Together, $\Delta H°_{bind}(Na^+)$ and $\Delta S°_{bind}(Na^+)$ indicate that Na^+ binding is characterized by the formation of energetically stable contacts with G3 at the entropic cost of restricting the translational freedom of Na^+ and/or the conformational flexibility of the triplex.

Similarly, the thermodynamics of K^+ binding to the antiparallel G4 can be determined by examining the differential folding enthalpy ($\Delta\Delta H°_{TA}(K^+)$) and entropy ($\Delta\Delta S°_{TA}(K^+)$) upon addition of K^+, in this case defined as

$$\Delta\Delta H°_{TA}(K^+) \equiv \Delta H°_{TA}(100 \text{ mM } Na^+, 4 \text{ mM } K^+) - \Delta H°_{TA}(100 \text{ mM } Na^+) \quad \text{(Eq. 7.18)}$$

and

$$\Delta\Delta S°_{TA}(K^+) \equiv \Delta S°_{TA}(100\text{ mM Na}^+, 4\text{ mM K}^+) - \Delta S°_{TA}(100\text{ mM Na}^+). \quad \text{(Eq. 7.19)}$$

In contrast to Na^+, K^+ interacts with the G-quadruplex through a fold-then-bind mechanism, and as a result $\Delta\Delta H°_{TA}(K^+)$ and $\Delta\Delta S°_{TA}(K^+)$ have the *same* sign as $\Delta H°_{bind}(K^+)$ and $\Delta S°_{bind}(K^+)$, respectively (Figure 7.8B). The observed $\Delta\Delta H°_{TA}(K^+)$ and $\Delta\Delta S°_{TA}(K^+)$ are both positive, which indicates that K^+ binding is enthalpically unfavorable ($\Delta H°_{bind}(K^+) > 0$) and entropically favorable ($\Delta S°_{bind}(K^+) > 0$). Interestingly, this thermodynamic signature is exactly the opposite of that for Na^+-binding, a fact which reinforces the wholly distinct nature of Na^+ and K^+ binding to the G-quadruplex construct.

7.5 Discussion

In this work, we have used the 5'-GGG-$(TAAGGG)_3$-3' DNA G-quadruplex (TAA-G4) as a model system for understanding the thermodynamic influence of monovalent cations Na^+ and K^+ on G-quadruplex folding. The TAA-G4 is incorporated into a DNA construct designed for single-molecule FRET (smFRET) microscopy. The observed smFRET data reveal that the TAA-G4 populates three interconverting FRET states (Figure 7.1) in agreement with previous single-molecule measurements by Tippana et al. on a similar TAA-G4 construct.[32] From the smFRET trajectories, equilibrium constants and rate constants for transitions between the three states are measured, and the dependence of these quantities on cation concentration and temperature have been obtained.

We used simple structure prediction to assign structures to the FRET states of the TAA-G4 construct. The middle and high FRET state efficiencies ($E_{middle} = 0.73$, $E_{high} = 0.97$) are consistent with folded G-quadruplexes with the fluorophores positioned on the opposite side

(distal) or the same side (proximal) as the G-quadruplex fold, respectively. The proximal configuration would be observed in a G-quadruplex with antiparallel topology (chair or basket), while the distal configuration corresponds to the parallel and hybrid topologies which are indistinguishable in our experiment. For an assignment of the low FRET state (E_{low} = 0.44), we considered but ruled out a fully unfolded G-quadruplex for two reasons. First, the predicted FRET efficiency is too low ($E_{unfolded} \approx 0.10$). Second, from the crystal structures of similar G-quadruplexes containing 3 G-tetrads,[10, 16-17] we expect the folded G-quadruplex to bind two monovalent cations, while the fully unfolded state should bind none. Therefore, the binding stoichiometry should be two ions; however, the measured Na^+ and K^+ binding stoichiometries (Figure 7.2B and Figure 7.4A) are more consistent with single ion binding. Therefore, we instead propose a G-triplex (G3) structure as a partially folded species whose predicted FRET efficiency ($E_{triplex}$ = 0.40–0.46) is much closer to E_{low}. Furthermore, G3 structures have been reported to bind to a single monovalent cation,[14] which would produce the correct binding stoichiometry. Nevertheless, alternative structural assignments are certainly plausible due to the uncertainty in our simple E_{FRET} calculations, especially for less-structured conformations such as the G3 state for which inaccuracies in the worm-like chain model will be magnified.

Monovalent cations drive the TAA-G4 folding equilibrium from G3 into G4 states in a cation-specific fashion. K^+ exclusively promotes the antiparallel configuration (Figure 7.4A), while Na^+ prefers the parallel/hybrid G4 with some sampling of the antiparallel G4 (Figure 7.2), i.e., the TAA-G4 is slightly polymorphic in Na^+ solution. This result is surprising in light of studies showing that G-quadruplexes adopt parallel or hybrid topologies in K^+ solution and antiparallel topologies in Na^+ solution.[10, 16] As stated above, it is possible that the FRET efficiencies have been incorrectly predicted by our simple structural model; however, it is

unlikely that the *order* of the FRET efficiencies is incorrect (i.e. $E_{antiparallel} > E_{parallel/hybrid}$). Instead, this disagreement is more likely due to the exquisite sensitivity of G-quadruplex folding equilibria to their nucleic acid context, i.e., loop sequence, adjacency to single-stranded or double-stranded nucleic acids, composition (RNA vs. DNA), and strand polarity (see ref. 15 for an especially dramatic demonstration of this sensitivity). Here, we are using the relatively unstudied TAA loop sequence which is especially polymorphic,[32] and we have introduced an ssDNA spacer between the TAA-G4 and the hybridizing sequence for surface tethering, unlike that used in the Tippana et al. construct upon which ours is based. Therefore, the structural impact of K^+ and Na^+ on the TAA-G4 construct may differ from that of other G-quadruplex constructs in the literature.

Measurement of concentration-dependent kinetics (Figures 7.3C and 7.4C) reveal that Na^+ and K^+ promote G-quadruplex folding through distinct kinetic mechanisms. Na^+ binds to the G-triplex before the ligand-bound complex folds into one of the G-quadruplex configurations (Figure 7.3D, a "bind-then-fold" mechanism). In contrast, the TAA-G4 sequence must first fold into the antiparallel G4 topology before K^+ will bind to form a stabilized complex (Figure 7.4D, a "fold-then-bind" mechanism). That these two monovalent cations act on the TAA-G4 in such different mechanisms reinforces how sensitive G-quadruplexes are to their environment, or in this case, cation size. We present a simplified, combined kinetic model of Na^+- and K^+-induced folding of G-quadruplexes in Figure 7.9. In the model, Na^+ binding is a prerequisite for triplex to quadruplex transitions. Then, the antiparallel configuration of the G4 can subsequently bind a K^+ ion, perhaps by replacement of Na^+ (as shown in the figure) or by attachment to an alternative binding site. It is possible and even likely that the identified Na^+-binding site has a similar affinity for other monovalent cations, including K^+, but the much higher affinity of the K^+-

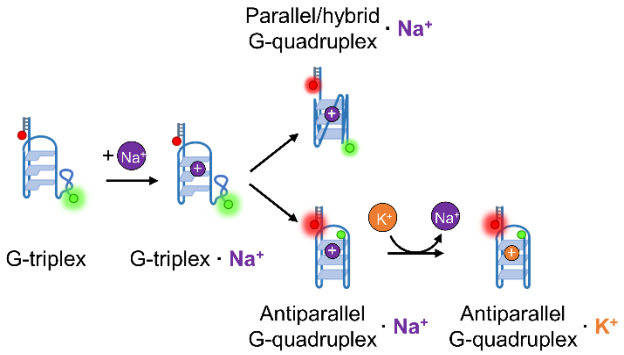

Figure 7.9 Proposed scheme of Na^+ and K^+ binding to TAA-GQ.

specific binding site means that our K^+ titrations saturate before the low affinity binding to the triplex can be observed.

To learn about the influence of Na^+ and K^+ on the thermodynamics of TAA-G4 folding, we used temperature-dependent experiments to obtain the folding enthalpies and entropies at multiple cation concentration conditions. By considering the change in thermodynamic parameters after an increase in $[Na^+]$ or $[K^+]$, we can infer the thermodynamics of cation binding. Interestingly, the folding thermodynamics of the TAA-G4 construct respond similarly to the addition of either 100 mM Na^+ or 4 mM K^+, as summarized in Table 7.2. However, because Na^+ and K^+ interact with the TAA-G4 through different mechanisms (bind-then-fold vs. fold-then-bind), we conclude that thermodynamic signature of Na^+ binding is the opposite of that for K^+ binding (Figure 7.8). Specifically, Na^+ binds to the G3 conformation driven by enthalpically favorable interactions ($\Delta H°_{bind}(Na^+) < 0$) that outweigh a loss of entropy ($\Delta S°_{bind}(Na^+) < 0$) which is perhaps due to G-triplex conformational restriction upon ion binding. In contrast, K^+ binds to the antiparallel G4 topology in an *entropically* driven process ($\Delta S°_{bind}(K^+) > 0$) that incurs an enthalpic cost ($\Delta H°_{bind}(K^+) > 0$). One explanation of this thermodynamic signature is that K^+ replaces the poorly fitting Na^+ in the binding pocket to form a more compact G-quadruplex which results in the expulsion of surface bound water molecules. Such a release of water molecules would be entropically favorable but energetically costly due to breaking

hydrogen bonds between the water molecules and the nucleic acid. This hypothesis of G-quadruplex compaction and water release upon K^+ binding is generally supported by literature on ion dependence of G-quadruplex folding volumes,[60-61] though one study reported that K^+ produces *larger* G4s than Na^+.[21] Clearly, further results are needed to clarify this issue, such as could be obtained by pressure-dependent single-molecule folding experiments.[62-63]

7.6 Summary and Conclusion

In this work, we have explored the influence of monovalent cations on G-quadruplex folding through single-molecule, temperature-controlled kinetic measurements. To our knowledge, this represents the first use of smFRET to measure the temperature-dependent folding kinetics on a system with more than two FRET conformations. We observe 3-state G-quadruplex folding, which we ascribe to conformational transitions between a G-triplex and two G-quadruplex topologies. Concentration-dependent measurements show that the Na^+ binds to the G-triplex and remains bound while G-quadruplex topologies are sampled. In contrast, K^+ binds exclusively to the antiparallel G-quadruplex conformation and has a binding affinity approximately 100 times larger than that of Na^+. The thermodynamics of Na^+ binding are enthalpically-driven ($\Delta H°_{bind}(Na^+) < 0$, $\Delta S°_{bind}(Na^+) < 0$), while entropy is the driving force of K^+ binding ($\Delta H°_{bind}(K^+) > 0$, $\Delta S°_{bind}(K^+) > 0$), likely due to G4 compaction and concomitant water release upon K^+ binding.

7.7 References

1. Dolinnaya, N. G.; Ogloblina, A. M.; Yakubovskaya, M. G., Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry (Moscow)* **2016,** *81* (13), 1602-1649.

2. Rhodes, D.; Lipps, H. J., G-quadruplexes and their regulatory roles in biology. *Nuclear Acids Res.* **2015,** *43* (18), 8627-8637.

3. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X., Targeting RNA G-quadruplex in SARS-CoV-2: A promising therapeutic target for COVID-19? *Angew. Chem., Int. Ed.* **2021,** *60* (1), 432-438.

4. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S., The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020,** *21* (8), 459-474.

5. Neidle, S., Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.* **2016,** *59* (13), 5987-6011.

6. Banerjee, N.; Panda, S.; Chatterjee, S., Frontiers in G-Quadruplex therapeutics in cancer: Selection of small molecules, peptides and aptamers. *Chem. Biol. Drug Des.* **2022,** *99* (1), 1-31.

7. Yatsunyk, L. A.; Mendoza, O.; Mergny, J.-L., "Nano-oddities": Unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc. Chem. Res.* **2014,** *47* (6), 1836-1844.

8. Rajendran, A.; Endo, M.; Hidaka, K.; Lan Thao Tran, P.; Mergny, J.-L.; Sugiyama, H., Controlling the stoichiometry and strand polarity of a tetramolecular G-quadruplex structure by using a DNA origami frame. *Nucleic Acids Res.* **2013,** *41* (18), 8738-8747.

9. Harkness, R. W.; Mittermaier, A. K., G-quadruplex dynamics. *Biochim. Biophys. Acta, Proteins Proteomics* **2017,** *1865* (11, Part B), 1544-1554.

10. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34* (19), 5402-5415.

11. Ma, Y.; Iida, K.; Nagasawa, K., Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem. Biophys. Res. Commun.* **2020,** *531* (1), 3-17.

12. Grün, J. T.; Schwalbe, H., Folding dynamics of polymorphic G-quadruplex structures. *Biopolymers* **2022,** *113* (1), e23477.

13. Nguyen, T. Q. N.; Lim, K. W.; Phan, A. T., Folding kinetics of G-quadruplexes: Duplex stem loops drive and accelerate G-quadruplex folding. *J. Phys. Chem. B.* **2020,** *124* (25), 5122-5130.

14. Cerofolini, L.; Amato, J.; Giachetti, A.; Limongelli, V.; Novellino, E.; Parrinello, M.; Fragai, M.; Randazzo, A.; Luchinat, C., G-triplex structure and formation propensity. *Nucleic Acids Res.* **2014,** *42* (21), 13393-404.

15. Hou, X.-M.; Fu, Y.-B.; Wu, W.-Q.; Wang, L.; Teng, F.-Y.; Xie, P.; Wang, P.-Y.; Xi, X.-G., Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* **2017,** *45* (19), 11401-11412.

16. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S., Metal cations in G-quadruplex folding and stability. *Front. Chem.* **2016,** *4* (38), 38.

17. Parkinson, G. N.; Lee, M. P. H.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-880.

18. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the human telomere in K+ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006,** *128* (30), 9963-9970.

19. Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O., Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008,** *36* (17), 5482-5515.

20. Chalikian, T. V.; Liu, L.; Macgregor, J. R. B., Duplex-tetraplex equilibria in guanine- and cytosine-rich DNA. *Biophys. Chem.* **2020,** *267*, 106473.

21. Li, Y. Y.; Dubins, D. N.; Le, D. M. N. T.; Leung, K.; Macgregor, R. B., The role of loops and cation on the volume of unfolding of G-quadruplexes related to HTel. *Biophys. Chem.* **2017,** *231*, 55-63.

22. Majhi, P. R.; Qi, J.; Tang, C.-F.; Shafer, R. H., Heat capacity changes associated with guanine quadruplex formation: An isothermal titration calorimetry study. *Biopolymers* **2008,** *89* (4), 302-309.

23. Green, J. J.; Ying, L.; Klenerman, D.; Balasubramanian, S., Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap. *J. Am. Chem. Soc.* **2003,** *125* (13), 3763-7.

24. 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* (29), 8088-8098.

25. Hatzakis, E.; Okamoto, K.; Yang, D., Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: Effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes. *Biochemistry* **2010,** *49* (43), 9152-9160.

26. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **2015,** *34*, 52-59.

27. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* **2018,** *359* (6373), 288.

28. Maleki, P.; Budhathoki, J. B.; Roy, W. A.; Balci, H., A practical guide to studying G-quadruplex structures using single-molecule FRET. *Mol. Genet. Genomics* **2017,** *292* (3), 483-498.

29. Budhathoki, Jagat B.; Maleki, P.; Roy, William A.; Janscak, P.; Yodh, Jaya G.; Balci, H., A comparative study of G-quadruplex unfolding and DNA reeling activities of human RECQ5 helicase. *Biophys. J.* **2016,** *110* (12), 2585-2596.

30. Aznauryan, M.; Sondergaard, S.; Noer, S. L.; Schiott, B.; Birkedal, V., A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Res.* **2016,** *44* (22), 11024-11032.

31. Noer, S. L.; Preus, S.; Gudnason, D.; Aznauryan, M.; Mergny, J. L.; Birkedal, V., Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na+ solutions by single molecule FRET microscopy. *Nucleic Acids Res.* **2016,** *44* (1), 464-71.

32. Tippana, R.; Xiao, W.; Myong, S., G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **2014,** *42* (12), 8106-8114.

33. Jansson, L. I.; Hentschel, J.; Parks, J. W.; Chang, T. R.; Lu, C.; Baral, R.; Bagshaw, C. R.; Stone, M. D., Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116* (19), 9350-9359.

34. Arns, L.; Knop, J.-M.; Patra, S.; Anders, C.; Winter, R., Single-molecule insights into the temperature and pressure dependent conformational dynamics of nucleic acids in the presence of crowders and osmolytes. *Biophys. Chem.* **2019,** *251*, 106190.

35. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: Kinetic insights from single-molecule studies. *J. Phys. Chem. B.* **2018,** *122* (43), 9869-9876.

36. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B.* **2020,** *124* (51), 11561-11572.

37. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **2008,** *94* (5), 1826-35.

38. Ha, T.; Tinnefeld, P., Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* **2012,** *63*, 595-617.

39. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

40. Hanson, S. M.; Ekins, S.; Chodera, J. D., Modeling error in experimental assays using the bootstrap principle: Understanding discrepancies between assays using different dispensing technologies. *J. Comput.-Aided Mol. Des.* **2015,** *29* (12), 1073-1086.

41. Leipply, D.; Draper, D. E., Dependence of RNA tertiary structural stability on Mg^{2+} concentration: Interpretation of the Hill equation and coefficient. *Biochemistry* **2010,** *49* (9), 1843-1853.

42. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with nucleic acid folding. *Biopolymers* **2006,** *82* (1), 38-58.

43. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* **2016,** *67*, 441-65.

44. Hori, N.; Denesyuk, N. A.; Thirumalai, D., Frictional effects on RNA folding: Speed limit and Kramers turnover. *J. Phys. Chem. B* **2018,** *122* (49), 11279-11288.

45. 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* (1), 018101.

46. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Tests of Kramers' Theory at the Single-Molecule Level: Evidence for Folding of an Isolated RNA Tertiary Interaction at the Viscous Speed Limit. *J. Phys. Chem. B* **2018,** *122* (38), 8796-8804.

47. Slater, G. W.; Gratton, Y.; Kenward, M.; McCormick, L.; Tessier, F., Deformation, stretching, and relaxation of single-polymer chains: Fundamentals and examples. *Soft Mater.* **2004,** *2* (2-3), 155-182.

48. Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T., Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **2004,** *86* (4), 2530-7.

49. 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* (3), 799-804.

50. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on Chips II*, Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

51. Steffen, F. D.; Sigel, R. K. O.; Börner, R., FRETraj: Integrating single-molecule spectroscopy with molecular dynamics. *Bioinformatics* **2021,** *37* (21), 3953-3955.

52. McKinney, S. A.; Joo, C.; Ha, T., Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **2006,** *91* (5), 1941-51.

53. Lee, T. H., Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *J. Phys. Chem. B.* **2009,** *113* (33), 11535-42.

54. Chodera, J. D.; Noe, F., Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25*, 135-144.

55. Leulliot, N.; Varani, G., Current topics in RNA−protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001,** *40* (27), 7947-7956.

56. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009,** *5* (11), 789-796.

57. Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q., Insights into protein-ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **2016,** *17* (2).

58. Dunitz, J. D., Win some, lose some: Enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995,** *2* (11), 709-712.

59. Fox, J. M.; Zhao, M.; Fink, M. J.; Kang, K.; Whitesides, G. M., The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annu. Rev. Biophys.* **2018,** *47* (1), 223-250.

60. Hellman, L. M.; Rodgers, D. W.; Fried, M. G., Phenomenological partial-specific volumes for G-quadruplex DNAs. *Eur. Biophys. J.* **2010,** *39* (3), 389-396.

61. Knop, J.-M.; Patra, S.; Harish, B.; Royer, C. A.; Winter, R., The deep sea osmolyte trimethylamine N-oxide and macromolecular crowders rescue the antiparallel conformation of the human telomeric G-quadruplex from urea and pressure stress. *Chem. Eur. J.* **2018,** *24* (54), 14346-14351.

62. Sung, H. L.; Nesbitt, D. J., Single-molecule kinetic studies of DNA hybridization under extreme pressures. *Phys. Chem. Chem. Phys.* **2020,** *22* (41), 23491-23501.

63. Sung, H. L.; Nesbitt, D. J., DNA hairpin hybridization under extreme pressures: A single-molecule FRET study. *J. Phys. Chem. B* **2020,** *124* (1), 110-120.

Chapter 8

Postmortem

8.1 Introduction

In professional chess, at the end of a game between two masters, there is a tradition called the "postmortem" analysis. The two opponents, after hours of exhausting mental effort spent in adversarial silence, come together for a friendly discussion of what happened during the game. They talk about the mistakes they made, the places they could have improved, the critical moments that shaped the outcome. In fact, most of what they discuss were moves *not* played in the game. They look at the hypothetical directions, the "what-if's" of where the game could have gone had they made different choices.

So far, this dissertation has only included research which 'worked', the projects which produced results that could pass the inspection of peer review and merited presentation to the scientific community. In this concluding chapter, I will perform a 'postmortem' analysis of my graduate career, exploring the "what-if's" of my research. These are the efforts which never made it to the literature, the partially completed projects with inconclusive or confounding results, and the possible research paths that I did not have the time to travel along. I write this in the hope of inspiring some future researcher to take up these lost projects, someone with the expertise which I lacked, or the insight which eluded me, or simply the time which I ran out of.

8.2 Determination of folding volume change with molecule dynamics (MD)

During my time in the Nesbitt group, Hsuan-Lei Sung, a graduate student, implemented pressure-controlled single-molecule FRET measurements. Nucleic acids can experience a change in volume during folding, which affects the Gibbs free energy via pressure volume work:

$$\frac{\partial(\Delta G°)}{\partial P} = -\frac{\Delta V°}{RT}$$

Where ΔG° is the folding free energy, P is the pressure, $\Delta V°$ is the folding volume, R is the gas constant, and T is the absolute temperature. These folding volumes are typically small (on the order of the volume of one water molecule), so the contribution to ΔG° at atmospheric pressure is usually negligible. However, at elevated pressures, such as those in the deep ocean, the pressure-volume work is significant. By changing the pressure in smFRET experiments, Hsuan-Lei Sung has been able to measure the volume change for various nucleic acids.[1-2] One of these is the "40-adenine hairpin" used in Chapter 3, for which $\Delta V°$ was measured as +23 ± 2 mL/mol. The positive sign indicates that the folded state of the hairpin is *larger* than the unfolded state, a somewhat counterintuitive result which is nevertheless generally true for nucleic acids as well as proteins.

After using molecular dynamics (MD) to study the chirality-dependent binding of amino acids to nucleic acids (Chapter 4), and inspired by Hsuan-Lei's results, I set out to measure $\Delta V°$ for the 40-adenine hairpin using MD. Rather than measure pressure dependence of the folding free energy, which is inaccessible to short time MD simulations, my strategy was to use the fact that volume is a readily available quantity in MD to instead measure the absolute volume of the folded and unfolded configurations and take the difference. Specifically, I first created the folded 40-adenine hairpin structure using the RNAComposer webserver, converted the structure into DNA, and solvated the system in TIP3P water and 100 mM KCl. I then created an unfolded structure by generating the linear sequence in the Ambertools program LEaP, collapsing the structure in 1 M salt for 30 ns, then stripping away the solvent and resolvating so that the folded and unfolded simulations had exactly the same number of water molecules, ions, etc. (Figure 8.1). These simulations were then run in NAMD at constant pressure P = 1 atm, which means

that the volume of the unit cell for the periodic simulation is changed to maintain a constant pressure. The unit cell volume versus time trajectory output by NAMD were then used to calculate an average system volume. To estimate uncertainty in the volume, the autocorrelation function of the volume trajectory was calculated to determine the volume decorrelation time τ, and the standard error of the mean σ_{SEM} was calculated from the raw standard deviation in the volume σ_V as $\sigma_{SEM} = \sigma_V/\sqrt{t/\tau}$, where t is the total simulation time. The calculations were also run 5x in parallel, and the uncertainties derived using the 5 independent simulations were in agreement with σ_{SEM}. Usually, ~10-25 ns of total simulation time were sufficient to reach ±5 mL/mol uncertainty, which is approximately ⅓ the volume of a single water molecule.

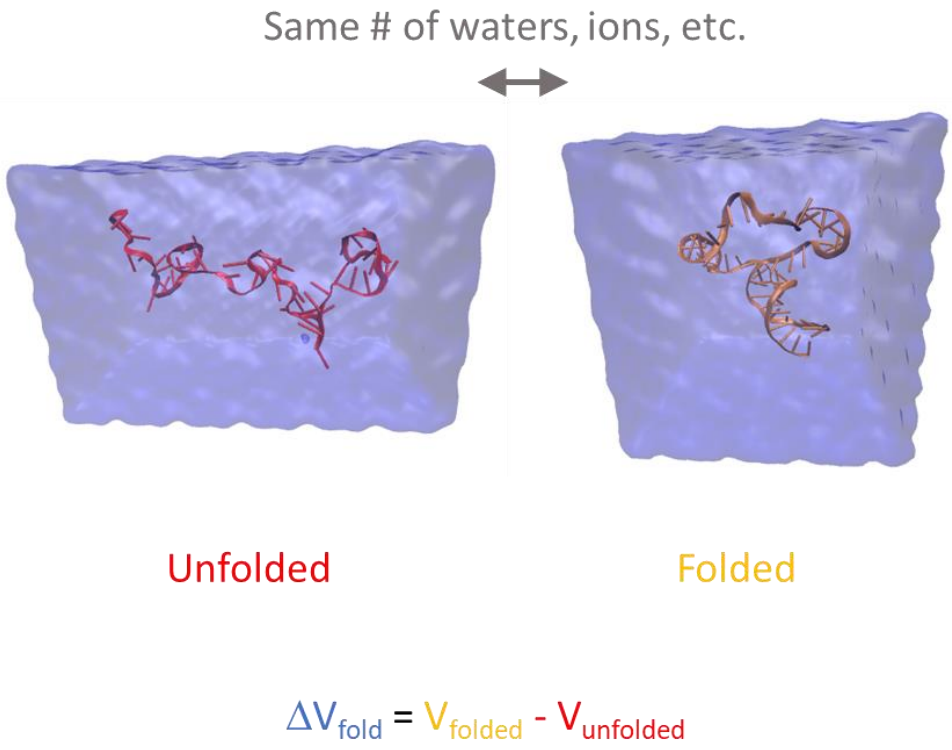

Figure 8.1 Schematic of strategy for determining nucleic acid folding volume in MD.

The results of the MD calculations yielded a volume change $\Delta V° = -73 \pm 5$ mL/mol, which is similar in magnitude to the experimental result but opposite in sign (i.e., in MD the

folded state is larger than the unfolded state). This is a *qualitatively* incorrect result which suggests something quite wrong with the calculation. My advisor and I were particularly suspicious of the water model, TIP3P, which is a 3-point rigid model. Interaction with solvent is thought to be key in folding volume changes, so an accurate water model may be important. Unfortunately, despite trying many different water models, including other 3-point models, 4-point rigid models, and a flexible water model, every simulation showed a negative value for $\Delta V°$ (Table 8.l). Concerned that the long 40-adening loop in the hairpin might be poorly converged in MD, I switched to a short 4-adenine loop, but the negative sign remained ($\Delta V° = -35 \pm 4$ mL/mol). Finally, I tried the calculation at multiple pressures, for which we expected $\Delta V°$ to be constant, but it was not (Figure 8.2). No closer to understanding these confounding results, I abandoned this effort. Future attempts may find it fruitful to look at alternative folding systems, more refined water models including polarizable models, or extending simulation time to more thoroughly investigate the convergence properties of these calculations.

	water Model	folded(A^3)	SEM	unfolded(A^3)	SEM	Δ(A^3)	SEM
3-point	SPC/E	431284	21	431385	17	-101	27
	OPC3	459842	8	459955	22	-113	23
	TIP3P	436985	5	437081	5	-96	8
4-point	OPC	434364	12	434386	17	-22	20
	TIP4P	424701	14	424753	16	-52	22
Flexible	SPC/Fw	427327	17	427427	16	-100	23
	experiment					+30	10

Table 8.1 Folding volume changes for 40-adenine hairpin for multiple water models in MD.

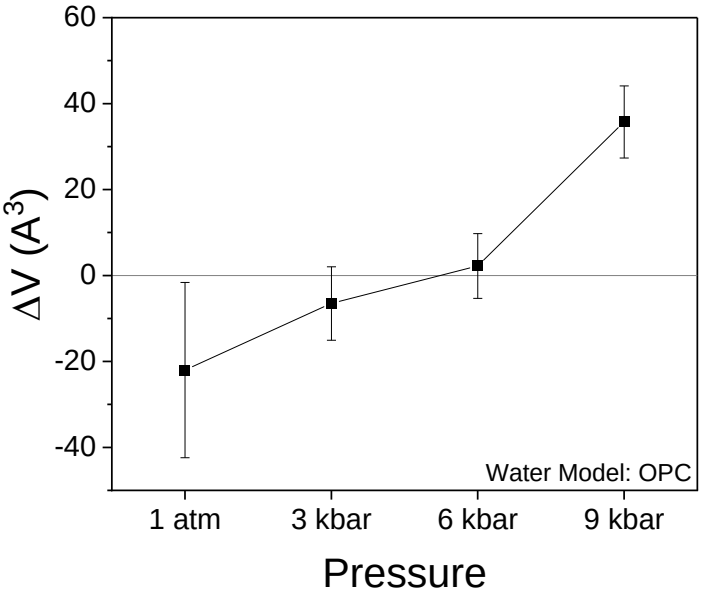

Figure 8.2 40-adenine hairpin volume change at different simulation pressures.

8.3 Cold shock protein and mRNA

One under-utilized element of the TIRF system is its ability to cool samples below room temperature. Cooling was used in this thesis as simply a tool for extending the temperature range for van't Hoff and Arrhenius analysis, but there is some interesting biology to explore near the freezing point of water. A particularly intriguing and relatively simple area to explore is that of cold shock proteins (CSPs).[3-4] When exposed to low temperatures, a cell's mRNA transcripts may form unwanted secondary structure elements that are too stable to permit ribosome processing. CSPs are nucleic acid folding chaperones which are released to *destabilize* those unwanted secondary structures.

The proposed experiment is quite simple: observe secondary structure formation kinetics, perhaps for a DNA hairpin, using smFRET at low temperatures with and without the presence of CSPs. Despite its simplicity, this experiment would provide a deeper mechanistic understanding of how exactly CSPs modify RNA folding dynamics than has previously been available. I was

prevented from attempting this experiment due to the difficulty in obtaining purified CSPs, which to my knowledge are not commercially available. Collaboration with researchers who are experienced in protein synthesis may be necessary. Fortunately, most CSPs are relatively short proteins with fewer than 100 amino acids.

An alternative but related experiment is measuring the folding dynamics of the mRNA that encodes for the CSP protein. It has been proposed[5] that this mRNA has evolved to avoid forming secondary structure even at low temperatures, which solves the problem of the CSP protein itself not being able to be synthesized at low temperatures due to secondary structure formation in its mRNA transcript. Again, this is an experiment I have not performed, which would require in vitro RNA ligation to use the full mRNA transcript, or else selection of a subsequence of interest which is short enough to be commercially synthesized without the need for ligation. Additionally, relevant FRET pair labeling locations would need to be determined.

8.4 Modulation of RNA folding dynamics by small peptides

As seen in Chapters 3 and 4, even single amino acids can affect RNA folding behavior. In those chapters, this result was used to propose that, in the history of evolution, the first role of amino acids in biology was to serve as RNA folding chaperones. The next logical step in this research is to explore whether the evolution of peptide synthesis could have been driven by the ability of peptides to act as even better RNA folding chaperones. My preliminary work here is limited to glycine-based peptides of increasing length (Gly_2 and Gly_3), and my results showed that diglycine and triglycine have approximately the same influence on folding as glycine. Going further in this area is challenging, as the chemical space of peptides increases in size exponentially with the peptide length, and exhaustively exploring even the dipeptides would be extremely challenging. However, there are certain amino acids which are believed to be older

member of the codon code than others, and these amino acids are good candidates for exploration. In particular, the four amino acids glycine, alanine, valine, and aspartic acid would be a good choice for a subspace of 16 possible dipeptides to explore. Furthermore, beyond an interest in the evolution of life, there are simple peptides which are of interest for their RNA-binding abilities that may have clinical uses, such as stop codon binders and the HIV TAR binder.[6-7]

8.5 Surface tethered kinetic studies of Brome mosaic virus

A less ambitious project could be the re-initiation of the Nesbitt group effort to study the kinetics of the Brome mosaic virus (BMV) in single-molecule microscopy. A previous postdoctoral researcher, Dr. Mario Vieweger, created a smFRET construct for diffusing studies on the confocal microscope.[8] From this construct, three FRET conformations were observed, and their population dependence on salt-levels was measured. However, no kinetic data was obtained, as surface tethering was not performed. I restarted this project by tethering an old BMV construct to a surface to try to observe the three FRET states at E = 0.1, 0.3, and 0.8. However, my trajectories showed little population of the high FRET state at all salt conditions used (Figure 8.3). Furthermore, the high FRET states were inconsistent from molecule to molecule. This suggests that the RNA sample has degraded and is no longer kinetically homogenous. Obtaining a new construct would be recommended; however, this construct is longer than is commercially available to synthesize and therefore requires a ligation design.

8.6 RNA conformational dynamics in liquid RNA droplets

When stressed, some cells halt mRNA translation through the creation of liquid RNA droplets.[9] These RNA droplets, or membraneless organelles, or stress granules, are composed of negatively charged RNA and positively charged proteins and they behave similarly to liquid

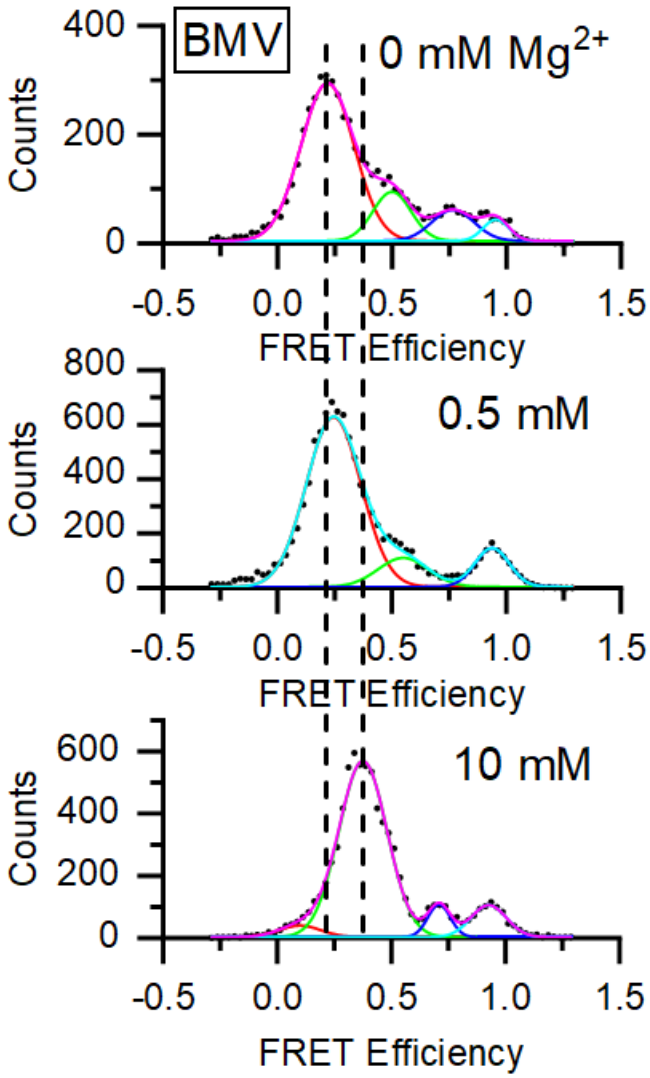

Figure 8.3 FRET histograms of the BMV construct from surface tethered measurements in 75 mM monovalent cation. Two dominant FRET states were observed at E = 0.2 and E = 0.4 (vertical dashed lines), with high magnesium driving the equilibrium to the high FRET state. However, there is no highly populated high FRET state as observed in Vieweger's diffusion studies (expected population at 0.5 mM Mg^{2+} = 50%). Instead, many low population FRET states at high E are observed.

crystals. To my knowledge, there has not been a single-molecule FRET study of RNA conformational dynamics inside RNA droplets, and such a work could probe the way that RNA molecules organize themselves in membraneless organelles. My graduate career ended before I could explore this possibility, but in principle, the experiment is fairly simple: follow developed procedures for generating RNA droplets *in vitro* and dope the droplets with a small concentration

of FRET-labeled RNA molecules for single-molecule microscopy. There are certainly challenges to this work, such as the possibility of a continuum of FRET states or extremely rapid structural fluctuations. However, CU has a number of experts on RNA droplets who could be excellent collaborators, such as Dr. Roy Parker.

8.7 Summary of work and open research questions

We may also find possible future research endeavors by examining the results found in Chapters 3−7. Each chapter has more detail in its discussion and summary sections, but I will now compile the major findings and what future research questions those findings raise.

In Chapters 3 and 4, I examine the effect of amino acids on nucleic acid folding. These series of experiments are inspired by a proposed evolutionary scenario in which the earliest role of amino acids was not incorporation into peptides but rather acting as RNA folding chaperones. In my experiments, I found that secondary structure responds only minimally to amino acids (Chapter 3), while in contrast tertiary structure is far more sensitive to amino acids, even capable of discerning amino acid chirality via specific binding of select amino acids (Chapter 4). As stated in section 8.4, the clear next step is to explore dipeptides and tripeptides, as an even stronger RNA chaperoning by small peptides, if observed, could be the evolutionary pressure responsible for the development of peptide synthesis. Additionally, the work in Chapter 4 was performed on a single tertiary fold, a tetraloop-tetraloop receptor, but other tertiary structures should be explored to evaluate the generality of strong amino acid sensitivity.

Chapters 5 and 6 contain efforts to extend the methodological capabilities of single-molecule FRET. Chapter 5 explores the origin of an artifact called “camera blurring” which occurs when FRET dynamics happen on the same timescale as the FRET data acquisition rate. Use of stroboscopic illumination was found to eliminate the artifact and thereby increase the

upper limit on measurable rate constants in smFRET by a factor of 5−10. In Chapter 6, I use temperature to probe thermodynamics of nucleic acid folding and obtain not only the usual enthalpy and entropy of folding but also the *heat capacity* ΔC_P of folding. This heat capacity measurement was used to obtain novel structural information about the folding transition state of a DNA hairpin. The ideal next step is to find additional nucleic acid folding systems which are well suited for heat capacity measurements. This is a challenging task, as heat capacities in folding are typically small due to the small surface area changes in nucleic acid folding (as compared to, say, surface area changes in globular protein folding). Perhaps a good choice would be a folding system which encapsulates a ligand during folding, such a riboswitch.

In Chapter 7, I turn to the polymorphic folding of G-quadruplexes. Rather than a system with 2 folding states, as is found in most smFRET kinetic studies, this work examines a G-quadruplex sequence which has 3 observable FRET states. Assigning a physical structure to these three states proved quite challenging, and utilization of higher quality FRET prediction calculations is recommended for future efforts. In this chapter, I performed the first temperature-dependent smFRET kinetic measurements of G-quadruplex folding. From my results, I deduced that the tested G-quadruplex sequence bound to both Na^+ and K^+, but that the two cations exhibited entirely different kinetic mechanisms, affinities, and thermodynamics of binding. The examined G-quadruplex sequence was selected because of its known propensity for polymorphic folding; the clear next step is to examine if the observed behaviors depend on G-quadruplex sequence, such as the number of G-tetrads or the loop length/sequence. Of particular interest is the human G-quadruplex sequence, which differs from the sequence used here by a T→A mutation in the loop sequence. Another sequence of interest is the putative G-quadruplex

structure found in the SARS-CoV-2 viral genome which is a possible drug target of high relevance to the current COVID-19 pandemic.[10]

8.8 References

1. Sung, H. L.; Nesbitt, D. J., Single-molecule kinetic studies of DNA hybridization under extreme pressures. *Phys. Chem. Chem. Phys.* **2020,** *22* (41), 23491-23501.

2. Sung, H. L.; Nesbitt, D. J., DNA hairpin hybridization under extreme pressures: A single-molecule FRET study. *J. Phys. Chem. B* **2020,** *124* (1), 110-120.

3. Rennella, E.; Sára, T.; Juen, M.; Wunderlich, C.; Imbert, L.; Solyom, Z.; Favier, A.; Ayala, I.; Weinhäupl, K.; Schanda, P., et al., RNA binding and chaperone activity of the E. coli cold-shock protein CspA. *Nucleic Acids Res.* **2017,** *45* (7), 4255-4268.

4. Phadtare, S.; Severinov, K., RNA remodeling and gene regulation by cold shock proteins. *RNA Biol.* **2010,** *7* (6), 788-795.

5. Keto-Timonen, R.; Hietala, N.; Palonen, E.; Hakakorpi, A.; Lindstrom, M.; Korkeala, H., Cold shock proteins: A minireview with special emphasis on Csp-family of enteropathogenic Yersinia. *Front. Microbiol.* **2016,** *7*, 1151.

6. Xie, B.; Calabro, V.; Wainberg, M. A.; Frankel, A. D., Selection of TAR RNA-binding chameleon peptides by using a retroviral replication system. *J. Virol.* **2004,** *78* (3), 1456-63.

7. Ito, K.; Uno, M.; Nakamura, Y., A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **2000,** *403* (6770), 680-4.

8. Vieweger, M.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule FRET reveals three conformations for the TLS domain of brome mosaic virus genome. *Biophys. J.* **2015,** *109*, 2625-2636.

9. Rhine, K.; Vidaurre, V.; Myong, S., RNA droplets. *Annu. Rev. Biophys.* **2020,** *49* (1), 247-265.

10. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X., Targeting RNA G-quadruplex in SARS-CoV-2: A promising therapeutic target for COVID-19? *Angew. Chem. Int. Ed.* **2021,** *60* (1), 432-438.

Bibliography

Chapter 1 Introduction

1. Cech, T. R., Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena (Nobel Lecture). *Angew. Chem., Int. Ed. Engl.* 1990, *29* (7), 759-768.
2. Cech, T. R.; Zaug, A. J.; Grabowski, P. J., In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* 1981, *27* (3, Part 2), 487-496.
3. Müller, S.; Appel, B.; Balke, D.; Hieronymus, R.; Nübel, C., Thirty-five years of research into ribozymes and nucleic acid catalysis: Where do we stand today? *F1000Research* 2016, *5*, F1000 Faculty Rev-1511.
4. Weinberg, C. E.; Weinberg, Z.; Hammann, C., Novel ribozymes: Discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.* 2019, *47* (18), 9480-9494.
5. Cech, T. R., The ribosome Is a ribozyme. *Science* 2000, *289* (5481), 878-879.
6. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* 2002, *418* (6894), 214-21.
7. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nat. Rev. Genet.* 2014, *15* (6), 423-37.
8. Sherwood, A. V.; Henkin, T. M., Riboswitch-mediated gene regulation: Novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.* 2016, *70* (1), 361-374.
9. Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C., Engineering and in vivo applications of riboswitches. *Annu. Rev. Biochem.* 2017, *86*, 515-539.
10. Agrawal, N.; Dasaradhi, P. V. N.; Mohmmed, A.; Malhotra, P.; Bhatnagar, R. K.; Mukherjee, S. K., RNA interference: Biology, mechanism, and applications. *Microbiol. Mol. Biol. Rev.* 2003, *67* (4), 657-685.
11. Wilkinson, M. E.; Charenton, C.; Nagai, K., RNA splicing by the spliceosome. *Annu. Rev. Biochem.* 2020, *89* (1), 359-388.
12. Srinivas, N.; Rachakonda, S.; Kumar, R., Telomeres and telomere length: A general overview. *Cancers (Basel)* 2020, *12* (3), 558.
13. Plohl, M.; Meštrović, N.; Mravinac, B., Centromere identity from the DNA point of view. *Chromosoma* 2014, *123* (4), 313-325.
14. Silverman, S. K., Catalytic DNA: Scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.* 2016, *41* (7), 595-609.

15. Adachi, T.; Nakamura, Y., Aptamers: A review of their chemical properties and modifications for therapeutic application. *Molecules* 2019, *24* (23), 4229.

16. Dey, S.; Fan, C.; Gothelf, K. V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M. A. D.; Saccà, B.; Simmel, F. C., et al., DNA origami. *Nat. Rev. Methods Primers* 2021, *1* (1), 13.

17. Yatsunyk, L. A.; Mendoza, O.; Mergny, J.-L., "Nano-oddities": Unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc. Chem. Res.* 2014, *47* (6), 1836-1844.

18. Makalowski, W., The human genome structure and organization. *Acta Biochim. Pol.* 2001, *48* (3), 587-98.

19. Palazzo, A. F.; Gregory, T. R., The case for junk DNA. *PLoS Genet.* 2014, *10* (5), e1004351-e1004351.

20. Minchin, S.; Lodge, J., Understanding biochemistry: Structure and function of nucleic acids. *Essays Biochem.* 2019, *63* (4), 433-456.

21. Carell, T.; Brandmayr, C.; Hienzsch, A.; Müller, M.; Pearson, D.; Reiter, V.; Thoma, I.; Thumbs, P.; Wagner, M., Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.* 2012, *51* (29), 7110-7131.

22. Duffy, K.; Arangundy-Franklin, S.; Holliger, P., Modified nucleic acids: Replication, evolution, and next-generation therapeutics. *BMC Biol.* 2020, *18* (1), 112.

23. Greenberg, M. V. C.; Bourc'his, D., The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* 2019, *20* (10), 590-607.

24. Handy, D. E.; Castro, R.; Loscalzo, J., Epigenetic modifications: Basic mechanisms and role in cardiovascular disease. *Circulation* 2011, *123* (19), 2145-2156.

25. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* 2014, *83*, 813-41.

26. Pyle, A. M., Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* 2002, *7* (7-8), 679-90.

27. Nikolova, E. N.; Zhou, H.; Gottardo, F. L.; Alvey, H. S.; Kimsey, I. J.; Al-Hashimi, H. M., A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* 2013, *99* (12), 955-968.

28. Nagaswamy, U.; Larios-Sanz, M.; Hury, J.; Collins, S.; Zhang, Z.; Zhao, Q.; Fox, G. E., NCIR: A database of non-canonical interactions in known RNA structures. *Nucleic Acids Res.* 2002, *30* (1), 395-397.

29. Reblova, K.; Spackova, N.; Stefl, R.; Csaszar, K.; Koca, J.; Leontis, N. B.; Sponer, J., Non-Watson-Crick basepairing and hydration in RNA motifs: Molecular dynamics of 5S rRNA loop E. *Biophys. J.* 2003, *84* (6), 3564-82.

30. Mirkin, S. M., Discovery of alternative DNA structures: A heroic decade (1979-1989). *Front. Biosci.* 2008, *13*, 1064-71.

31. Tinoco, I.; Bustamante, C., How RNA folds. *J. Mol. Biol.* 1999, *293* (2), 271-281.

32. Dill, K. A.; MacCallum, J. L., The protein-folding problem, 50 years on. *Science* 2012, *338* (6110), 1042-1046.

33. Englander, S. W.; Mayne, L., The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* 2014, *111* (45), 15873-15880.

34. Herschlag, D., RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 1995, *270*, 20871-20874.

35. Brion, P.; Westhof, E., Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* 1997, *26*, 113-37.

36. Mustoe, A. M.; Brooks, C. L.; Al-Hashimi, H. M., Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* 2014, *83* (1), 441-466.

37. Zarrinkar, P. P.; Williamson, J. R., Kinetic intermediates in RNA folding. *Science* 1994, *265*, 918-924.

38. Leontis, N. B.; Lescoute, A.; Westhof, E., The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* 2006, *16* (3), 279-287.

39. Butcher, S. E.; Pyle, A. M., The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* 2011, *44* (12), 1302-1311.

40. Svoboda, P.; Cara, A. D., Hairpin RNA: A secondary structure of primary importance. *Cell. Mol. Life Sci.* 2006, *63* (7), 901-908.

41. Tan, Z. J.; Chen, S. J., Salt dependence of nucleic acid hairpin stability. *Biophys. J.* 2008, *95* (2), 738-52.

42. Kuznetsov, S. V.; Ren, C. C.; Woodson, S. A.; Ansari, A., Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* 2008, *36* (4), 1098-112.

43. Staple, D. W.; Butcher, S. E., Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* 2005, *3* (6), e213-e213.

44. Chandrasekaran, A. R.; Rusling, D. A., Triplex-forming oligonucleotides: A third strand for DNA nanotechnology. *Nucleic Acids Res.* 2018, *46* (3), 1021-1037.

45. Lilley, D. M. J., Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* 2001, *33* (2), 109-159.

46. 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* (16), 8038-8056.

47. Dolinnaya, N. G.; Ogloblina, A. M.; Yakubovskaya, M. G., Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry (Moscow)* 2016, *81* (13), 1602-1649.

48. Lyu, K.; Chow, E. Y.-C.; Mou, X.; Chan, T.-F.; Kwok, Chun K., RNA G-quadruplexes (rG4s): Genomics and biological functions. *Nucleic Acids Res.* 2021, *49* (10), 5426-5450.

49. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S., The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 2020, *21* (8), 459-474.

50. Garcia, H. G.; Kondev, J.; Orme, N.; Theriot, J. A.; Phillips, R., Thermodynamics of biological processes. In *Methods Enzymol.*, Johnson, M. L.; Holt, J. M.; Ackers, G. K., Eds. Academic Press: 2011; Vol. 492, pp 27-59.

51. Röder, K.; Wales, D. J., The energy landscape perspective: Encoding structure and function for biomolecules. *Front. Mol. Biosci.* 2022, *9*.

52. SantaLucia, J., Jr.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 2004, *33*, 415-40.

53. Fox, J. M.; Zhao, M.; Fink, M. J.; Kang, K.; Whitesides, G. M., The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annu. Rev. Biophys.* 2018, *47* (1), 223-250.

54. Dunitz, J. D., Win some, lose some: Enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* 1995, *2* (11), 709-712.

55. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* 2016, *67*, 441-65.

56. Gillespie, D. T., Stochastic chemical kinetics. In *Handbook of Materials Modeling: Methods*, Yip, S., Ed. Springer Netherlands: Dordrecht, 2005; pp 1735-1752.

57. Chen, S. J., RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* 2008, *37*, 197-214.

58. Egli, M., Diffraction techniques in structural biology. *Curr. Protoc. Nucleic Acid Chem.* 2016, *65* (1), 7.13.1-7.13.41.

59. Ferentz, A. E.; Wagner, G., NMR spectroscopy: A multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* 2000, *33* (1), 29-65.

60. Jelesarov, I.; Bosshard, H. R., Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* 1999, *12* (1), 3-18.

61. Johnson, C. M., Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.* 2013, *531* (1-2), 100-109.

62. Lee, H.-T.; Carr, C.; Siebler, H.; Waters, L.; Khutsishvili, I.; Iseka, F.; Domack, B.; Olsen, C. M.; Marky, L. A., A thermodynamic approach for the targeting of nucleic acid structures using their complementary single strands. In *Methods Enzymol.*, Johnson, M. L.; Holt, J. M.; Ackers, G. K., Eds. Academic Press: 2011; Vol. 492, pp 1-26.

63. SantaLucia, J., Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* 1998, *95* (4), 1460-5.

64. Lorenz, R.; Wolfinger, M. T.; Tanzer, A.; Hofacker, I. L., Predicting RNA secondary structures from sequence and probing data. *Methods* 2016, *103*, 86-98.

65. Elf, J.; Barkefors, I., Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* 2019, *88* (1), 635-659.

66. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* 2018, *359* (6373), 288.

67. Shen, H.; Tauzin, L. J.; Baiyasi, R.; Wang, W.; Moringo, N.; Shuang, B.; Landes, C. F., Single particle tracking: From theory to biophysical applications. *Chem. Rev.* 2017, *117* (11), 7331-7376.

68. Sasmal, D. K.; Pulido, L. E.; Kasal, S.; Huang, J., Single-molecule fluorescence resonance energy transfer in molecular biology. *Nanoscale* 2016, *8* (48), 19928-19944.

69. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* 2015, *34*, 52-59.

70. Robinson, A.; van Oijen, A. M., Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol.* 2013, *11* (5), 303-315.

71. Barkai, E.; Garini, Y.; Metzler, R., Strange kinetics of single molecules in living cells. *Phys. Today* 2012, *65* (8), 29-35.

72. van Oijen, A. M., Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Curr. Opin. Biotechnol.* 2011, *22* (1), 75-80.

73. Aleman, E. A.; Lamichhane, R.; Rueda, D., Exploring RNA folding one molecule at a time. *Curr. Opin. Chem. Biol.* 2008, *12* (6), 647-654.

74. Kubelka, J., Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem. Photobiol. Sci.* 2009, *8* (4), 499-512.

75. Dyer, R. B.; Brauns, E. B., Laser-induced temperature jump infrared measurements of RNA folding. *Methods Enzymol.* 2009, *469*, 353-72.

76. Zheng, X.; Bi, C.; Li, Z.; Podariu, M.; Hage, D. S., Analytical methods for kinetic studies of biological interactions: A review. *J. Pharm. Biomed. Anal.* 2015, *113*, 163-180.

77. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* 2008, *5* (6), 507-16.

78. Preus, S.; Wilhelmsson, L. M., Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem* 2012, *13* (14), 1990-2001.

79. Ha, T.; Tinnefeld, P., Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* 2012, *63*, 595-617.

80. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

81. Rajkowitsch, L.; Chen, D.; Stampfl, S.; Semrad, K.; Waldsich, C.; Mayer, O.; Jantsch, M. F.; Konrat, R.; Blasi, U.; Schroeder, R., RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* 2007, *4* (3), 118-30.

82. Fiore, J. L.; Nesbitt, D. J., An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* 2013, *46* (3), 223-64.

83. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with nucleic acid folding. *Biopolymers* 2006, *82* (1), 38-58.

Chapter 2 Methods

1. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B.* 2020, *124* (51), 11561-11572.

2. Nicholson, D. A.; Nesbitt, D. J., Pushing camera-based single-molecule kinetic measurements to the frame acquisition limit with stroboscopic smFRET. *J. Phys. Chem. B* 2021, *125* (23), 6080-6089.

3. Hodak, J. H.; Fiore, J. L.; Nesbitt, D. J.; Downey, C. D.; Pardi, A., Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* 2005, *102* (30), 10505.

4. Fiore, J. L.; Hodak, J. H.; Piestert, O.; Downey, C. D.; Nesbitt, D. J., Monovalent and divalent promoted GAAA tetraloop-receptor tertiary interactions from freely diffusing single-molecule studies. *Biophys. J.* 2008, *95* (8), 3892-3905.

5. Axelrod, D.; Burghardt, T. P.; Thompson, N. L., Total internal reflection fluorescence. *Ann. Rev. Biophys. Bioeng.* 1984, *13*, 247-268.

6. 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.

7. Brown, L. G., A survey of image registration techniques. *ACM Comput. Surv.* 1992, *24* (4), 325–376.

8. Shuang, B.; Cooper, D.; Taylor, J. N.; Kisley, L.; Chen, J.; Wang, W.; Li, C. B.; Komatsuzaki, T.; Landes, C. F., Fast step transition and state identification (STaSI) for discrete single-molecule data analysis. *J. Phys. Chem. Lett.* 2014, *5* (18), 3157-3161.

9. Hu, B.; Rakthanmanon, T.; Hao, Y.; Evans, S.; Lonardi, S.; Keogh, E., Using the minimum description length to discover the intrinsic cardinality and dimensionality of time series. *Data Min. Knowl. Discov.* 2015, *29* (2), 358-399.

10. Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P., et al., RNA structural dynamics as captured by

molecular simulations: A comprehensive overview. *Chem. Rev.* 2018, *118* (8), 4177-4338.

Chapter 3 Amino Acid Stabilization of Nucleic Acid Secondary Structure: Kinetic Insights from Single Molecule Studies

1. Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. *Cell Res.* 2011, *21*, 381-395.

2. Knobler, C. M.; Gelbart, W. M., Physical chemistry of DNA viruses. *Annu. Rev. Phys. Chem.* 2009, *60*, 367-383.

3. Speir, J. A.; Johnson, J. E., Nucleic acid packaging in viruses. *Curr. Opin. Struct. Biol.* 2012, *22*, 65-71.

4. Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C., Engineering and in vivo applications of riboswitches. *Annu. Rev. Biochem.* 2017, *86*, 515-539.

5. Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R., A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 2004, *306*, 275-279.

6. Butler, E. B.; Xiong, Y.; Wang, J.; Strobel, S. A., Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem. Biol.* 2011, *18*, 293-298.

7. Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R., An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* 2003, *17*, 2688-2697.

8. Serganov, A.; Huang, L.; Patel, D. J., Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* 2008, *455*, 1263-1267.

9. Yang, Y.; Kochoyan, M.; Burgstaller, P.; Westhof, E.; Famulok, M., Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* 1996, *272*, 1343-1347.

10. Yarus, M., A specific amino acid binding site composed of RNA. *Science* 1988, *240*, 1751-1758.

11. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nat. Rev. Genet.* 2014, *15*, 423-437.

12. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* 2002, *418*, 214-221.

13. Herschlag, D., RNA Chaperones and the RNA folding problem. In *J. Biol. Chem.*, 1995; Vol. 270, pp 20871-20874.

14. Hoffman, M. M.; Khrapov, M. A.; Cox, J. C.; Yao, J.; Tong, L.; Ellington, A. D., Aant: The amino acid-nucleotide interaction database. *Nucleic Acids Res.* 2004, *32*, D174-181.

15. Morozova, N.; Allers, J.; Myers, J.; Shamoo, Y., Protein-RNA interactions: Exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics* 2006, *22*, 2746-2752.

16. Cheng, A. C.; Chen, W. W.; Fuhrmann, C. N.; Frankel, A. D., Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J. Mol. Biol.* 2003, *327*, 781-796.

17. Hajnic, M.; Osorio, J. I.; Zagrovic, B., Computational analysis of amino acids and their sidechain analogs in crowded solutions of RNA nucleobases with implications for the mRNA-protein complementarity hypothesis. *Nucleic Acids Res.* 2014, *42*, 12984-12994.

18. Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D., characterization of nucleobase-amino acid stacking interactions utilized by a DNA repair enzyme. *J. Phys. Chem. B* 2006, *110*, 19652-19663.

19. de Ruiter, A.; Zagrovic, B., Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. *Nucleic Acids Res.* 2015, *43*, 708-718.

20. Sengupta, A.; Sung, H. L.; Nesbitt, D. J., Amino acid specific effects on RNA tertiary interactions: Single-molecule kinetic and thermodynamic studies. *J. Phys. Chem. B* 2016, *120*, 10615-10627.

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

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

23. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* 2008, *5*, 507-516.

24. Preus, S.; Wilhelmsson, L. M., Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem* 2012, *13*, 1990-2001.

25. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* 2016, *67*, 441-465.

26. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on chips II*; Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

27. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* 2008, *94*, 1826-1835.

28. Corle, T. R.; Kino, G. S., *Confocal scanning optical microscopy and related imaging systems*; Academic Press: San Diego, 1996.

29. Fiore, J. L.; Holmstrom, E. D.; Nesbitt, D. J., Entropic origin of Mg2+-facilitated RNA folding. *Proc. Natl. Acad. Sci. U.S.A.* 2012, *109*, 2902-2907.

30. Sinden, R. R.; Pearson, C. E.; Potaman, V. N.; Ussery, D. W., DNA: Structure and function. In *Advances in genome biology*; Verma, R. S., Ed. JAI: 1998; Vol. 5, pp 1-141.

31. Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F., Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U.S.A.* 2010, *107*, 15431-15436.

32. Bevilacqua, P. C.; Blose, J. M., Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem.* 2008, *59*, 79-103.

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

34. Holmstrom, E. D.; Fiore, J. L.; Nesbitt, D. J., Thermodynamic origins of monovalent facilitated RNA folding. *Biochemistry* 2012, *51*, 3732-3743.

35. Rentzeperis, D.; Shikiya, R.; Maiti, S.; Ho, J.; Marky, L. A., Folding of intramolecular DNA hairpin loops: Enthalpy−entropy compensations and hydration contributions. *J. Phys. Chem. B* 2002, *106*, 9945-9950.

36. Rentzeperis, D.; Alessi, K.; Marky, L. A., Thermodynamics of DNA hairpins: Contribution of loop size to hairpin stability and ethidium binding. *Nucleic Acids Res.* 1993, *21*, 2683-2689.

37. Auffinger, P.; D'Ascenzo, L.; Ennifar, E., Sodium and potassium interactions with nucleic acids. In *The Alkali Metal Ions: Their Role for Life*; Sigel, A.; Sigel, H.; Sigel, R. K. O., Eds. Springer International Publishing: Cham, 2016; pp 167-201.

38. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* 2014, *83*, 813-841.

39. Trachman, R. J., 3rd; Draper, D. E., Divalent ion competition reveals reorganization of an RNA ion atmosphere upon folding. *Nucleic Acids Res.* 2017, *45*, 4733-4742.

40. Gebala, M.; Giambasu, G. M.; Lipfert, J.; Bisaria, N.; Bonilla, S.; Li, G.; York, D. M.; Herschlag, D., Cation-anion interactions within the nucleic acid ion atmosphere revealed by ion counting. *J. Am. Chem. Soc.* 2015, *137*, 14705-14715.

41. Vieregg, J.; Cheng, W.; Bustamante, C.; Tinoco, I., Jr., Measurement of the effect of monovalent cations on RNA hairpin stability. *J. Am. Chem. Soc.* 2007, *129*, 14966-14973.

42. Kuznetsov, S. V.; Ansari, A., A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* 2012, *102*, 101-111.

43. Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T., Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* 2004, *86*, 2530-2537.

44. 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.

45. Bosco, A.; Camunas-Soler, J.; Ritort, F., Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.* 2014, *42*, 2064-2074.

46. Privalov, P. L.; Ptitsyn, O. B.; Birshtein, T. M., Determination of stability of DNA double helix in an aqueous medium. *Biopolymers* 1969, *8*, 559-571.

47. 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.

48. Tan, Z. J.; Chen, S. J., Salt dependence of nucleic acid hairpin stability. *Biophys. J.* 2008, *95*, 738-752.

49. Huang, P. C., Effect of guanidinium ion on DNA denaturation and renaturation. *Biochem. Biophys. Res. Commun.* 1968, *33*, 384-390.

50. Blanco, F.; Kelly, B.; Sanchez-Sanz, G.; Trujillo, C.; Alkorta, I.; Elguero, J.; Rozas, I., Non-covalent interactions: Complexes of guanidinium with DNA and RNA nucleobases. *J. Phys. Chem. B* 2013, *117*, 11608-11616.

51. Lambert, D.; Leipply, D.; Shiman, R.; Draper, D. E., The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.* 2009, *390*, 791-804.

Chapter 4 Chirality-Dependent Amino Acid Modulation of RNA Folding

1. Blackmond, D. G., The origin of biological homochirality. *Cold Spring Harbor Perspect. Biol.* 2010, *2* (5), a002147-a002147.

2. Zask, A.; Ellestad, G., Biomimetic syntheses of racemic natural products. *Chirality* 2018, *30* (2), 157-164.

3. Finefield, J. M.; Sherman, D. H.; Kreitman, M.; Williams, R. M., Enantiomeric natural products: occurrence and biogenesis. *Angew. Chem., Int. Ed. Engl.* 2012, *51* (20), 4802-4836.

4. Banik, S. D.; Nandi, N., Chirality and protein biosynthesis. In *Biochirality: Origins, Evolution and Molecular Recognition*, Cintas, P., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 255-305.

5. Genchi, G., An overview on D-amino acids. *Amino Acids* 2017, *49* (9), 1521-1533.

6. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* 2002, *418* (6894), 214-21.

7. Tamura, K.; Schimmel, P., Chiral-selective aminoacylation of an RNA minihelix. *Science (New York, N.Y.)* 2004, *305* (5688), 1253.

8. Tamura, K., Molecular basis for chiral selection in RNA aminoacylation. *Int. J. Mol. Sci.* 2011, *12*, 4745-4757.

9. Herschlag, D., RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 1995, *270*, 20871-20874.

10. Woodson, S. A., Taming free energy landscapes with RNA chaperones. *RNA Biol.* 2010, *7* (6), 677-86.

11. Chen, S. J., RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* 2008, *37*, 197-214.

12. Saibil, H., Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell. Biol.* 2013, *14* (10), 630-642.

13. Arakawa, T.; Timasheff, S. N., The stabilization of proteins by osmolytes. *Biophys. J.* 1985, *47* (3), 411-414.

14. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: kinetic insights from single-molecule studies. *J. Phys. Chem. B.* 2018, *122* (43), 9869-9876.

15. Sengupta, A.; Sung, H. L.; Nesbitt, D. J., Amino acid specific effects on rna tertiary interactions: single-molecule kinetic and thermodynamic studies. *J. Phys. Chem. B.* 2016, *120*, 10615-10627.

16. Yarus, M., A specific amino acid binding site composed of RNA. *Science (New York, N.Y.)* 1988, *240* (4860), 1751-8.

17. Burg, M. B.; Ferraris, J. D., Intracellular organic osmolytes: function and regulation. *J. Biol. Chem.* 2008, *283* (12), 7309-7313.

18. Khan, S. H.; Ahmad, N.; Ahmad, F.; Kumar, R., Naturally occurring organic osmolytes: from cell physiology to disease prevention. *IUBMB Life* 2010, *62* (12), 891-895.

19. Ando, T.; Takahashi, S.; Tamura, K., Principles of chemical geometry underlying chiral selectivity in RNA minihelix aminoacylation. *Nucleic Acids Res.* 2018, *46* (21), 11144-11152.

20. Tamura, K., RNA-directed molecular asymmetry of amino acids. *Viva Origino* 2010, *38* (4), 18-22.

21. Illangasekare, M.; Yarus, M., Phenylalanine-binding RNAs and genetic code evolution. *J. Mol. Evol.* 2002, *54* (3), 298-311.

22. Yarus, M., Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. *J. Mol. Evol.* 1998, *47* (1), 109-117.

23. Yarus, M., RNA-ligand chemistry: a testable source for the genetic code. *RNA* 2000, *6* (4), 475-484.

24. Illangasekare, M.; Turk, R.; Peterson, G. C.; Lladser, M.; Yarus, M., Chiral histidine selection by D-ribose RNA. *RNA* 2010, *16* (12), 2370-2383.

25. Tohala, L.; Oukacine, F.; Ravelet, C.; Peyrin, E., Chiral resolution capabilities of DNA oligonucleotides. *Anal. Chem.* 2015, *87* (11), 5491-5495.

26. Sivaleela, T.; Kumar, M. R.; Prabhakar, S.; Bhaskar, G.; Vairamani, M., Chiral discrimination of α-amino acids by DNA tetranucleotides under electrospray ionization conditions. *Rapid Commun. Mass Spectrom.* 2008, *22* (2), 204-210.

27. Yarus, M.; Majerfeld, I., Co-optimization of ribozyme substrate stacking and L-arginine binding. *J. Mol. Biol.* 1992, *225* (4), 945-9.

28. Holmstrom, E. D.; Nesbitt, D. J., Biophysical Insights from Temperature-Dependent Single-Molecule Forster Resonance Energy Transfer. *Annu. Rev. Phys. Chem.* 2016, *67*, 441-65.

29. Fiore, J. L.; Nesbitt, D. J., An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* 2013, *46* (3), 223-64.

30. Downey, C. D.; Fiore, J. L.; Stoddard, C. D.; Hodak, J. H.; Nesbitt, D. J.; Pardi, A., Metal ion dependence, thermodynamics, and kinetics for intramolecular docking of a GAAA tetraloop and receptor connected by a flexible linker. *Biochemistry* 2006, *45*, 3664-3673.

31. Seol, Y.; Skinner, G. M.; Visscher, K.; Buhot, A.; Halperin, A., Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys. Rev. Lett.* 2007, *98* (15), 158103.

32. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* 2008, *5* (6), 507-16.

33. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on Chips II*, Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

34. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* 2008, *94* (5), 1826-35.

35. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

36. Mullikin, J. C.; Vliet, L. J. v.; Netten, H.; Boddeke, F. R.; Feltz, G. v. d.; Young, I. T., *Methods for CCD camera characterization*. SPIE: 1994; Vol. 2173.

37. Cohen, E. A. K.; Ober, R. J., Image registration error analysis with applications in single molecule microscopy. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012; pp 996-999.

38. Hodak, J. H.; Fiore, J. L.; Nesbitt, D. J.; Downey, C. D.; Pardi, A., Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* 2005, *102* (30), 10505.

39. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden,; R.E. Duke, D. G., G. Giambasu, et al. *AMBER 2019*, University of California, San Francisco, 2019.

40. Mark, P.; Nilsson, L., Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A.* 2001, *105* (43), 9954-9960.

41. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graphics* 1996, *14* (1), 33-38.

42. Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P., Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* 2011, *7* (9), 2886-2902.

43. Joung, I. S.; Cheatham, T. E., 3rd, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* 2008, *112* (30), 9020-41.

44. Horn, A. H., A consistent force field parameter set for zwitterionic amino acid residues. *J. Mol. Model.* 2014, *20* (11), 2478.

45. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 2005, *26* (16), 1781-1802.

46. Stone, J. E. An efficient library for parallel ray tracing and animation. Master's Thesis, University of Missouri, Rolla, 1998.

47. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Tests of Kramers' theory at the single-molecule level: evidence for folding of an isolated RNA tertiary interaction at the viscous speed limit. *J. Phys. Chem. B.* 2018, *122* (38), 8796-8804.

48. Hori, N.; Denesyuk, N. A.; Thirumalai, D., Frictional effects on RNA folding: speed limit and Kramers turnover. *J. Phys. Chem. B.* 2018, *122* (49), 11279-11288.

49. Laidler, K. J.; King, M. C., Development of transition-state theory. *J. Phys. Chem.* 1983, *87* (15), 2657-2664.

50. Best, R. B.; Hummer, G., Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys. Rev. Lett.* 2006, *96* (22), 228104.

51. Amend, J. P.; Helgeson, H. C., Solubilities of the common L-α-amino acids as a function of temperature and solution pH. *Pure Appl. Chem.* 1997, *69* (5), 935-942.

52. Bisaria, N.; Herschlag, D., Probing the kinetic and thermodynamic consequences of the tetraloop/tetraloop receptor monovalent ion-binding site in P4-P6 RNA by smFRET. *Biochem. Soc. Trans.* 2015, *43* (2), 172-8.

53. Holmstrom, E. D.; Fiore, J. L.; Nesbitt, D. J., Thermodynamic origins of monovalent facilitated RNA folding. *Biochemistry* 2012, *51* (18), 3732-43.

54. Meng, X. Y.; Zhang, H. X.; Mezei, M.; Cui, M., Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput.-Aided Drug Des.* 2011, *7* (2), 146-57.

55. Ramírez, D.; Caballero, J., Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *Int. J. Mol. Sci.* 2016, *17* (4), 525.

56. Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; Walter, N. G.; Otyepka, M., RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* 2018, *118* (8), 4177-4338.

57. Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kuhrova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M., Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J. Phys. Chem. Lett.* 2014, *5* (10), 1771-82.

58. Mlynsky, V.; Bussi, G., Exploring RNA structure and dynamics through enhanced sampling simulations. *Curr. Opin. Struct. Biol.* 2018, *49*, 63-71.

59. Hyeon, C.; Thirumalai, D., Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* 2011, *2*, 487.

60. Jonikas, M. A.; Radmer, R. J.; Laederach, A.; Das, R.; Pearlman, S.; Herschlag, D.; Altman, R. B., Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 2009, *15* (2), 189-199.

61. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* 2015, *10* (5), 449-461.

62. Kocakaya, S. Ö.; Turgut, Y.; Pirinççioglu, N., Enantiomeric discrimination of chiral organic salts by chiral aza-15-crown-5 ether with C1 symmetry: experimental and theoretical approaches. *J. Mol. Model.* 2015, *21* (3), 55.

63. Choi, Y.; Jung, S., Molecular dynamics (MD) simulations for the prediction of chiral discrimination of N-acetylphenylalanine enantiomers by cyclomaltoheptaose (β-cyclodextrin, β-CD) based on the MM–PBSA (molecular mechanics–Poisson–Boltzmann surface area) approach. *Carbohydr. Res.* 2004, *339* (11), 1961-1966.

64. Wu, Y.; Ma, P.; Liu, Y.; Li, S., Diffusion coefficients of L-proline, L-threonine and L-arginine in aqueous solutions at 25°C. *Fluid Phase Equilib.* 2001, *186* (1), 27-38.

65. Chen, A. A.; Draper, D. E.; Pappu, R. V., Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. *J. Mol. Biol.* 2009, *390* (4), 805-819.

66. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* 2014, *83*, 813-41.

67. Trifonov, E. N., Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 2000, *261* (1), 139-51.

68. Blanco, C.; Bayas, M.; Yan, F.; Chen, I. A., Analysis of evolutionarily independent protein-RNA complexes yields a criterion to evaluate the relevance of prebiotic scenarios. *Curr. Biol.* 2018, *28* (4), 526-537 e5.

Chapter 5 Pushing Camera-Based Single Molecule Kinetic Measurements to the Frame Acquisition Limit with Stroboscopic smFRET

1. Aleman, E. A.; Lamichhane, R.; Rueda, D., Exploring RNA folding one molecule at a time. *Curr. Opin. Chem. Biol.* 2008, *12* (6), 647-654.

2. Buchner, G. S.; Murphy, R. D.; Buchete, N.-V.; Kubelka, J., Dynamics of protein folding: Probing the kinetic network of folding-unfolding transitions with experiment and theory. *Biochim. Biophys. Acta, Proteins Proteomics* 2011, *1814* (8), 1001-1020.

3. Grima, R.; Walter, N. G.; Schnell, S., Single-molecule enzymology à la Michaelis-Menten. *FEBS J.* 2014, *281* (2), 518-530.

4. van Oijen, A. M., Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Curr. Opin. Biotechnol.* 2011, *22* (1), 75-80.

5. Stratmann, S. A.; van, O. A. M., DNA replication at the single-molecule level. *Chem. Soc. Rev.* 2014, *43* (4), 1201-20.

6. Friedman, L. J.; Gelles, J., Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods* 2015, *86*, 27-36.

7. Elf, J.; Barkefors, I., Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* 2019, *88* (1), 636-659.

8. Barkai, E.; Garini, Y.; Metzler, R., Strange kinetics of single molecules in living cells. *Phys. Today* 2012, *65* (8), 29-35.

9. Goryaynov, A.; Ma, J.; Yang, W., Single-molecule studies of nucleocytoplasmic transport: from one dimension to three dimensions. *Integr. Biol.* 2012, *4* (1), 10-21.

10. Guan, J.; Jia, C.; Li, Y.; Liu, Z.; Wang, J.; Yang, Z.; Gu, C.; Su, D.; Houk, K. N.; Zhang, D., et al., Direct single-molecule dynamic detection of chemical reactions. *Sci. Adv.* 2018, *4* (2), eaar2177.

11. Tachikawa, T.; Majima, T., Single-molecule, single-particle approaches for exploring the structure and kinetics of nanocatalysts. *Langmuir* 2012, *28* (24), 8933-8943.

12. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* 2008, *5* (6), 507-16.

13. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Foerster resonance energy transfer. *Science* 2018, *359* (6373), 288.

14. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* 2015, *34*, 52-59.

15. Kinz-Thompson, C. D.; Bailey, N. A.; Gonzalez, R. L., Precisely and accurately inferring single-molecule rate constants. In *Methods in Enzymology*, Spies, M.; Chemla, Y. R., Eds. Academic Press: 2016; Vol. 581, pp 187-225.

16. Bronson, J. E.; Fei, J.; Hofman, J. M.; Gonzalez, R. L., Jr.; Wiggins, C. H., Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* 2009, *97* (12), 3196-205.

17. Hadzic, M.; Borner, R.; Konig, S. L. B.; Kowerko, D.; Sigel, R. K. O., Reliable state identification and state transition detection in fluorescence intensity-based single-molecule Forster resonance energy-transfer data. *J. Phys. Chem. B* 2018, *122* (23), 6134-6147.

18. Tang, J.; Sun, Y.; Pang, S.; Han, K. Y., Spatially encoded fast single-molecule fluorescence spectroscopy with full field-of-view. *Sci. Rep.* 2017, *7* (1), 10945.

19. Corle, T. R.; Kino, G. S., *Confocal scanning optical microscopy and related imaging systems*. Academic Press: San Diego, 1996.

20. Tan, Y. W.; Hanson, J. A.; Chu, J. W.; Yang, H., Confocal single-molecule FRET for protein conformational dynamics. *Methods Mol. Biol.* 2014, *1084*, 51-62.

21. Gopich, I.; Szabo, A., Theory of photon statistics in single-molecule Forster resonance energy transfer. *J. Chem. Phys.* 2005, *122* (1), 14707.

22. Schrangl, L.; Göhring, J.; Schütz, G. J., Kinetic analysis of single molecule FRET transitions without trajectories. *J. Chem. Phys.* 2018, *148* (12), 123328.

23. Santoso, Y.; Torella, J. P.; Kapanidis, A. N., Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem* 2010, *11* (10), 2209-19.

24. Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A., Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B* 2010, *114* (23), 7983-95.

25. Chen, J.; Pyle, J. R.; Sy Piecco, K. W.; Kolomeisky, A. B.; Landes, C. F., A two-step method for smFRET data analysis. *J. Phys. Chem. B* 2016, *120* (29), 7128-32.

26. Elf, J.; Li, G.-W.; Xie, X. S., Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 2007, *316* (5828), 1191.

27. Hansen, A. S.; Woringer, M.; Grimm, J. B.; Lavis, L. D.; Tjian, R.; Darzacq, X., Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* 2018, *7*, e33125.

28. Fontana, M.; Fijen, C.; Lemay, S. G.; Mathwig, K.; Hohlbein, J., High-throughput, non-equilibrium studies of single biomolecules using glass-made nanofluidic devices. *Lab Chip* 2018, *19* (1), 79-86.

29. Farooq, S.; Hohlbein, J., Camera-based single-molecule FRET detection with improved time resolution. *Phys. Chem. Chem. Phys.* 2015, *17* (41), 27862-72.

30. Gopich, I. V.; Szabo, A., Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B* 2009, *113* (31), 10965-73.

31. Borner, R.; Kowerko, D.; Hadzic, M.; Konig, S. L. B.; Ritter, M.; Sigel, R. K. O., Simulations of camera-based single-molecule fluorescence experiments. *PLoS One* 2018, *13* (4), e0195277.

32. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B* 2020, *124* (51), 11561-11572.

33. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: kinetic insights from single-molecule studies. *J. Phys. Chem. B* 2018, *122* (43), 9869-9876.

34. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on Chips II*, Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

35. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* 2008, *94* (5), 1826-35.

36. Ha, T.; Tinnefeld, P., Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* 2012, *63* (1), 595-617.

37. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

38. Stigler, J.; Rief, M., Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *ChemPhysChem* 2012, *13* (4), 1079-1086.

39. Blanco, M.; Walter, N. G., Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol.* 2010, *472*, 153-178.

40. Al-Mohy, A. H.; Higham, N. J., Improved inverse scaling and squaring algorithms for the matrix logarithm. *SIAM J. Sci. Comput.* 2012, *34* (4), C153-C169.

41. Cheng, S. H.; Higham, N. J.; Kenney, C. S.; Laub, A. J., Approximating the logarithm of a matrix to specified accuracy. *SIAM J. Matrix Anal. Appl.* 2001, *22* (4), 1112-1125.

42. Chodera, J. D.; Noe, F., Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* 2014, *25*, 135-144.

43. Lee, T. H., Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden markov models. *J. Phys. Chem. B* 2009, *113* (33), 11535-42.

44. McKinney, S. A.; Joo, C.; Ha, T., Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* 2006, *91* (5), 1941-51.

45. Lee, T. H.; Lapidus, L. J.; Zhao, W.; Travers, K. J.; Herschlag, D.; Chu, S., Measuring the folding transition time of single RNA molecules. *Biophys. J.* 2007, *92* (9), 3275-83.

46. Lu, H. P., Sizing up single-molecule enzymatic conformational dynamics. *Chem. Soc. Rev.* 2014, *43* (4), 1118-43.

47. Lerner, E.; Ploetz, E.; Hohlbein, J.; Cordes, T.; Weiss, S., A quantitative theoretical framework for protein-induced fluorescence enhancement-Forster-type resonance energy transfer (PIFE-FRET). *J. Phys. Chem. B* 2016, *120* (26), 6401-10.

Chapter 6 Measuring Excess Heat Capacities of DNA Folding at the Single Molecule Level

1. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* 2016, *67*, 441-65.

2. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with nucleic acid folding. *Biopolymers* 2006, *82* (1), 38-58.

3. SantaLucia, J., Jr.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 2004, *33*, 415-40.

4. Tsourkas, A.; Behlke, M. A.; Rose, S. D.; Bao, G., Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Res.* 2003, *31*, 1319-1330.

5. Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F., Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U. S. A.* 2010, *107* (35), 15431-6.

6. SantaLucia, J., Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* 1998, *95* (4), 1460-5.

7. Dragan, A.; Privalov, P.; Crane-Robinson, C., Thermodynamics of DNA: Heat capacity changes on duplex unfolding. *Eur. Biophys. J.* 2019, *48* (8), 773-779.

8. Gallagher, K.; Sharp, K., Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* 1998, *75* (2), 769-776.

9. Hadzi, S.; Lah, J., Origin of heat capacity increment in DNA folding: The hydration effect. *Biochim. Biophys. Acta, Gen. Subj.* 2021, *1865* (1), 129774.

10. Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; Record, M. T., Jr., 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. *Biochemistry* 1999, *38* (26), 8409-22.

11. Hughesman, C. B.; Turner, R. F. B.; Haynes, C., Correcting for heat capacity and 5'-TA type terminal nearest neighbors improves prediction of DNA melting temperatures using nearest-neighbor thermodynamic models. *Biochemistry* 2011, *50* (13), 2642-2649.

12. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with DNA duplex formation: Salt- and sequence-dependent effects. *Biochemistry* 2006, *45* (2), 604-616.

13. Volker, J.; Plum, G. E.; Breslauer, K. J., Heat capacity changes (ΔC_p) for interconversions between differentially-ordered DNA states within physiological temperature domains: Implications for biological regulatory switches. *J. Phys. Chem. B* 2020, *124* (27), 5614-5625.

14. Cooper, A., Heat capacity effects in protein folding and ligand binding: A re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.* 2005, *115* (2), 89-97.

15. Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q., Insights into protein-ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* 2016, *17* (2).

16. Majhi, P. R.; Qi, J.; Tang, C.-F.; Shafer, R. H., Heat capacity changes associated with guanine quadruplex formation: An isothermal titration calorimetry study. *Biopolymers* 2008, *89* (4), 302-309.

17. Jelesarov, I.; Crane-Robinson, C.; Privalov, P. L., The energetics of HMG box interactions with DNA: Thermodynamic description of the target DNA duplexes. *J. Mol. Biol.* 1999, *294* (4), 981-95.

18. Jelesarov, I.; Bosshard, H. R., Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* 1999, *12* (1), 3-18.

19. Johnson, C. M., Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.* 2013, *531* (1-2), 100-109.

20. Chalikian, T. V.; Volker, J.; Plum, G. E.; Breslauer, K. J., A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. U. S. A.* 1999, *96* (14), 7853-8.

21. Aleman, E. A.; Lamichhane, R.; Rueda, D., Exploring RNA folding one molecule at a time. *Curr. Opin. Chem. Biol.* 2008, *12* (6), 647-654.

22. Elf, J.; Barkefors, I., Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* 2019, *88* (1), 635-659.

23. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* 2015, *34*, 52-59.

24. Bizzarri, A. R.; Cannistraro, S., The application of atomic force spectroscopy to the study of biological complexes undergoing a biorecognition process. *Chem. Soc. Rev.* 2010, *39* (2), 734-749.

25. Tan, Y. W.; Hanson, J. A.; Chu, J. W.; Yang, H., Confocal single-molecule FRET for protein conformational dynamics. *Methods. Mol. Biol.* 2014, *1084*, 51-62.

26. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Foerster resonance energy transfer. *Science* 2018, *359* (6373), 288.

27. Baker, M. A.; Inoue, Y.; Takeda, K.; Ishijima, A.; Berry, R. M., Two methods of temperature control for single-molecule measurements. *Eur. Biophys. J.* 2011, *40* (5), 651-60.

28. Williams, M. C.; Wenner, J. R.; Rouzina, I.; Bloomfield, V. A., Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching. *Biophys. J.* 2001, *80* (4), 1932-9.

29. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: Kinetic insights from single-molecule studies. *J. Phys. Chem. B* 2018, *122* (43), 9869-9876.

30. Sung, H. L.; Nesbitt, D. J., Single-molecule kinetic studies of DNA hybridization under extreme pressures. *Phys. Chem. Chem. Phys.* 2020, *22* (41), 23491-23501.

31. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B* 2020, *124* (51), 11561-11572.

32. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* 2008, *94* (5), 1826-35.

33. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

34. Nicholson, D. A.; Nesbitt, D. J., Pushing camera-based single molecule kinetic measurements to the frame acquisition limit with stroboscopic smFRET. *J. Phys. Chem. B* 2021, *125* (23), 6080-6089.

35. Cohen, E. A. K.; Ober, R. J., Image registration error analysis with applications in single molecule microscopy. In *Proc. I. S. Biomed. Imag.*, 2012; pp 996-999.

36. Nicholson, D. A.; Nesbitt, D. J., Pushing camera-based single molecule kinetic measurements to the frame acquisition limit with stroboscopic smFRET. *J. Phys. Chem. B* 2021, *in press*.

37. Kinz-Thompson, C. D.; Bailey, N. A.; Gonzalez, R. L., Precisely and accurately inferring single-molecule rate constants. In *Methods in Enzymology*, Spies, M.; Chemla, Y. R., Eds. Academic Press: 2016; Vol. 581, pp 187-225.

38. Pechukas, P., Transition state theory. *Annu. Rev. Phys. Chem.* 1981, *32* (1), 159-177.

39. Laidler, K. J.; King, M. C., Development of transition-state theory. *J. Phys. Chem.* 1983, *87* (15), 2657-2664.

40. Kuznetsov, S. V.; Ansari, A., A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* 2012, *102* (1), 101-11.

41. Ouldridge, T. E.; Sulc, P.; Romano, F.; Doye, J. P.; Louis, A. A., DNA hybridization kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* 2013, *41* (19), 8886-95.

42. Yin, Y.; Zhao, X. S., Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* 2011, *44*, 1172-1181.

43. Taylor, J. R., *An introduction to error analysis : The study of uncertainties in physical measurements*. University Science Books: 1982.

44. Rouzina, I.; Bloomfield, V. A., Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.* 1999, *77* (6), 3242-51.

45. Rouzina, I.; Bloomfield, V. A., Heat capacity effects on the melting of DNA. 2. Analysis of nearest-neighbor base pair effects. *Biophys. J.* 1999, *77* (6), 3252-5.

46. Völker, J.; Makube, N.; Plum, G. E.; Klump, H. H.; Breslauer, K. J., Conformational energetics of stable and metastable states formed by DNA triplet repeat oligonucleotides: Implications for triplet expansion diseases. *Proc. Natl. Acad. Sci. U. S. A.* 2002, *99* (23), 14700.

47. Bourdélat-Parks, B. N.; Wartell, R. M., Thermodynamic stability of DNA tandem mismatches. *Biochemistry* 2004, *43* (30), 9918-9925.

48. Amrane, S.; Saccà, B.; Mills, M.; Chauhan, M.; Klump, H. H.; Mergny, J.-L., Length-dependent energetics of $(CTG)_n$ and $(CAG)_n$ trinucleotide repeats. *Nucleic Acids Res.* 2005, *33* (13), 4065-4077.

49. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat. Methods* 2008, *5* (6), 507-16.

50. Moreira, B. G.; You, Y.; Owczarzy, R., Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: Predictive thermodynamic model. *Biophys. Chem.* 2015, *198*, 36-44.

51. McIntosh, Dustin B.; Duggan, G.; Gouil, Q.; Saleh, Omar A., Sequence-dependent elasticity and electrostatics of single-stranded DNA: Signatures of base-stacking. *Biophys. J.* 2014, *106* (3), 659-666.

52. Oliveberg, M.; Tan, Y. J.; Fersht, A. R., Negative activation enthalpies in the kinetics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 1995, *92* (19), 8926-8929.

53. 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* (50), 16947-16952.

54. He, G.; Li, J.; Ci, H.; Qi, C.; Guo, X., Direct measurement of single-molecule DNA hybridization dynamics with single-base resolution. *Angew. Chem., Int. Ed.* 2016, *55* (31), 9036-9040.

55. 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.

56. Strunz, T.; Oroszlan, K.; Schafer, R.; Guntherodt, H. J., Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* 1999, *96* (20), 11277-82.

57. Whitley, K. D.; Comstock, M. J.; Chemla, Y. R., Elasticity of the transition state for oligonucleotide hybridization. *Nucleic Acids Res.* 2017, *45* (2), 547-555.

58. Ho, D.; Zimmermann, J. L.; Dehmelt, F. A.; Steinbach, U.; Erdmann, M.; Severin, P.; Falter, K.; Gaub, H. E., Force-driven separation of short double-stranded DNA. *Biophys. J.* 2009, *97* (12), 3158-67.

59. Kuznetsov, S. V.; Ansari, A., A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* 2012, *102* (1), 101-11.

Chapter 7 Kinetic and thermodynamic control of G-quadruplex polymorphism by Na^+ and K^+ cations

1. Dolinnaya, N. G.; Ogloblina, A. M.; Yakubovskaya, M. G., Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry (Moscow)* **2016,** *81* (13), 1602-1649.

2. Rhodes, D.; Lipps, H. J., G-quadruplexes and their regulatory roles in biology. *Nuclear Acids Res.* **2015,** *43* (18), 8627-8637.

3. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X., Targeting RNA G-quadruplex in SARS-CoV-2: A promising therapeutic target for COVID-19? *Angew. Chem., Int. Ed.* **2021,** *60* (1), 432-438.

4. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S., The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020,** *21* (8), 459-474.

5. Neidle, S., Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.* **2016,** *59* (13), 5987-6011.

6. Banerjee, N.; Panda, S.; Chatterjee, S., Frontiers in G-Quadruplex therapeutics in cancer: Selection of small molecules, peptides and aptamers. *Chem. Biol. Drug Des.* **2022,** *99* (1), 1-31.

7. Yatsunyk, L. A.; Mendoza, O.; Mergny, J.-L., "Nano-oddities": Unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc. Chem. Res.* **2014,** *47* (6), 1836-1844.

8. Rajendran, A.; Endo, M.; Hidaka, K.; Lan Thao Tran, P.; Mergny, J.-L.; Sugiyama, H., Controlling the stoichiometry and strand polarity of a tetramolecular G-quadruplex structure by using a DNA origami frame. *Nucleic Acids Res.* **2013,** *41* (18), 8738-8747.

9. Harkness, R. W.; Mittermaier, A. K., G-quadruplex dynamics. *Biochim. Biophys. Acta, Proteins Proteomics* **2017,** *1865* (11, Part B), 1544-1554.

10. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34* (19), 5402-5415.

11. Ma, Y.; Iida, K.; Nagasawa, K., Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem. Biophys. Res. Commun.* **2020,** *531* (1), 3-17.

12. Grün, J. T.; Schwalbe, H., Folding dynamics of polymorphic G-quadruplex structures. *Biopolymers* **2022,** *113* (1), e23477.

13. Nguyen, T. Q. N.; Lim, K. W.; Phan, A. T., Folding kinetics of G-quadruplexes: Duplex stem loops drive and accelerate G-quadruplex folding. *J. Phys. Chem. B.* **2020,** *124* (25), 5122-5130.

14. Cerofolini, L.; Amato, J.; Giachetti, A.; Limongelli, V.; Novellino, E.; Parrinello, M.; Fragai, M.; Randazzo, A.; Luchinat, C., G-triplex structure and formation propensity. *Nucleic Acids Res.* **2014,** *42* (21), 13393-404.

15. Hou, X.-M.; Fu, Y.-B.; Wu, W.-Q.; Wang, L.; Teng, F.-Y.; Xie, P.; Wang, P.-Y.; Xi, X.-G., Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* **2017,** *45* (19), 11401-11412.

16. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S., Metal cations in G-quadruplex folding and stability. *Front. Chem.* **2016,** *4* (38), 38.

17. Parkinson, G. N.; Lee, M. P. H.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-880.

18. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the human telomere in K+ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006,** *128* (30), 9963-9970.

19. Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O., Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008,** *36* (17), 5482-5515.

20. Chalikian, T. V.; Liu, L.; Macgregor, J. R. B., Duplex-tetraplex equilibria in guanine- and cytosine-rich DNA. *Biophys. Chem.* **2020,** *267*, 106473.

21. Li, Y. Y.; Dubins, D. N.; Le, D. M. N. T.; Leung, K.; Macgregor, R. B., The role of loops and cation on the volume of unfolding of G-quadruplexes related to HTel. *Biophys. Chem.* **2017,** *231*, 55-63.

22. Majhi, P. R.; Qi, J.; Tang, C.-F.; Shafer, R. H., Heat capacity changes associated with guanine quadruplex formation: An isothermal titration calorimetry study. *Biopolymers* **2008,** *89* (4), 302-309.

23. Green, J. J.; Ying, L.; Klenerman, D.; Balasubramanian, S., Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap. *J. Am. Chem. Soc.* **2003,** *125* (13), 3763-7.

24. 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* (29), 8088-8098.

25. Hatzakis, E.; Okamoto, K.; Yang, D., Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: Effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes. *Biochemistry* **2010,** *49* (43), 9152-9160.

26. Sustarsic, M.; Kapanidis, A. N., Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **2015,** *34*, 52-59.

27. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* **2018,** *359* (6373), 288.

28. Maleki, P.; Budhathoki, J. B.; Roy, W. A.; Balci, H., A practical guide to studying G-quadruplex structures using single-molecule FRET. *Mol. Genet. Genomics* **2017,** *292* (3), 483-498.

29. Budhathoki, Jagat B.; Maleki, P.; Roy, William A.; Janscak, P.; Yodh, Jaya G.; Balci, H., A comparative study of G-quadruplex unfolding and DNA reeling activities of human RECQ5 helicase. *Biophys. J.* **2016,** *110* (12), 2585-2596.

30. Aznauryan, M.; Sondergaard, S.; Noer, S. L.; Schiott, B.; Birkedal, V., A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Res.* **2016,** *44* (22), 11024-11032.

31. Noer, S. L.; Preus, S.; Gudnason, D.; Aznauryan, M.; Mergny, J. L.; Birkedal, V., Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na+ solutions by single molecule FRET microscopy. *Nucleic Acids Res.* **2016,** *44* (1), 464-71.

32. Tippana, R.; Xiao, W.; Myong, S., G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **2014,** *42* (12), 8106-8114.

33. Jansson, L. I.; Hentschel, J.; Parks, J. W.; Chang, T. R.; Lu, C.; Baral, R.; Bagshaw, C. R.; Stone, M. D., Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116* (19), 9350-9359.

34. Arns, L.; Knop, J.-M.; Patra, S.; Anders, C.; Winter, R., Single-molecule insights into the temperature and pressure dependent conformational dynamics of nucleic acids in the presence of crowders and osmolytes. *Biophys. Chem.* **2019,** *251*, 106190.

35. Nicholson, D. A.; Sengupta, A.; Sung, H.-L.; Nesbitt, D. J., Amino acid stabilization of nucleic acid secondary structure: Kinetic insights from single-molecule studies. *J. Phys. Chem. B.* **2018,** *122* (43), 9869-9876.

36. Nicholson, D. A.; Sengupta, A.; Nesbitt, D. J., Chirality-dependent amino acid modulation of RNA folding. *J. Phys. Chem. B.* **2020,** *124* (51), 11561-11572.

37. Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **2008,** *94* (5), 1826-35.

38. Ha, T.; Tinnefeld, P., Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* **2012,** *63*, 595-617.

39. Axelrod, D., Total internal reflection fluorescence microscopy. In *Methods in Cell Biology*, 1 ed.; Elsevier Inc.: 2008; Vol. 89, pp 169-221.

40. Hanson, S. M.; Ekins, S.; Chodera, J. D., Modeling error in experimental assays using the bootstrap principle: Understanding discrepancies between assays using different dispensing technologies. *J. Comput.-Aided Mol. Des.* **2015,** *29* (12), 1073-1086.

41. Leipply, D.; Draper, D. E., Dependence of RNA tertiary structural stability on Mg^{2+} concentration: Interpretation of the Hill equation and coefficient. *Biochemistry* **2010,** *49* (9), 1843-1853.

42. Mikulecky, P. J.; Feig, A. L., Heat capacity changes associated with nucleic acid folding. *Biopolymers* **2006,** *82* (1), 38-58.

43. Holmstrom, E. D.; Nesbitt, D. J., Biophysical insights from temperature-dependent single-molecule Forster resonance energy transfer. *Annu. Rev. Phys. Chem.* **2016,** *67*, 441-65.

44. Hori, N.; Denesyuk, N. A.; Thirumalai, D., Frictional effects on RNA folding: Speed limit and Kramers turnover. *J. Phys. Chem. B* **2018,** *122* (49), 11279-11288.

45. 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* (1), 018101.

46. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Tests of Kramers' Theory at the Single-Molecule Level: Evidence for Folding of an Isolated RNA Tertiary Interaction at the Viscous Speed Limit. *J. Phys. Chem. B* **2018,** *122* (38), 8796-8804.

47. Slater, G. W.; Gratton, Y.; Kenward, M.; McCormick, L.; Tessier, F., Deformation, stretching, and relaxation of single-polymer chains: Fundamentals and examples. *Soft Mater.* **2004,** *2* (2-3), 155-182.

48. Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T., Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **2004,** *86* (4), 2530-7.

49. 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* (3), 799-804.

50. Smith, C. L.; Milea, J. S.; Nguyen, G. H., Immobilization of nucleic acids using biotin-strept(avidin) systems. In *Immobilisation of DNA on Chips II*, Wittmann, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 63-90.

51. Steffen, F. D.; Sigel, R. K. O.; Börner, R., FRETraj: Integrating single-molecule spectroscopy with molecular dynamics. *Bioinformatics* **2021,** *37* (21), 3953-3955.

52. McKinney, S. A.; Joo, C.; Ha, T., Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **2006,** *91* (5), 1941-51.

53. Lee, T. H., Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *J. Phys. Chem. B.* **2009,** *113* (33), 11535-42.

54. Chodera, J. D.; Noe, F., Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25*, 135-144.

55. Leulliot, N.; Varani, G., Current topics in RNA−protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001,** *40* (27), 7947-7956.

56. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009,** *5* (11), 789-796.

57. Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q., Insights into protein-ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **2016,** *17* (2).

58. Dunitz, J. D., Win some, lose some: Enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995,** *2* (11), 709-712.

59. Fox, J. M.; Zhao, M.; Fink, M. J.; Kang, K.; Whitesides, G. M., The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annu. Rev. Biophys.* **2018,** *47* (1), 223-250.

60. Hellman, L. M.; Rodgers, D. W.; Fried, M. G., Phenomenological partial-specific volumes for G-quadruplex DNAs. *Eur. Biophys. J.* **2010,** *39* (3), 389-396.

61. Knop, J.-M.; Patra, S.; Harish, B.; Royer, C. A.; Winter, R., The deep sea osmolyte trimethylamine N-oxide and macromolecular crowders rescue the antiparallel conformation of the human telomeric G-quadruplex from urea and pressure stress. *Chem. Eur. J.* **2018,** *24* (54), 14346-14351.

62. Sung, H. L.; Nesbitt, D. J., Single-molecule kinetic studies of DNA hybridization under extreme pressures. *Phys. Chem. Chem. Phys.* **2020,** *22* (41), 23491-23501.

63. Sung, H. L.; Nesbitt, D. J., DNA hairpin hybridization under extreme pressures: A single-molecule FRET study. *J. Phys. Chem. B* **2020,** *124* (1), 110-120.

Chapter 8 Postmortem

1. Sung, H. L.; Nesbitt, D. J., Single-molecule kinetic studies of DNA hybridization under extreme pressures. *Phys. Chem. Chem. Phys.* **2020,** *22* (41), 23491-23501.

2. Sung, H. L.; Nesbitt, D. J., DNA hairpin hybridization under extreme pressures: A single-molecule FRET study. *J. Phys. Chem. B* **2020,** *124* (1), 110-120.

3. Rennella, E.; Sára, T.; Juen, M.; Wunderlich, C.; Imbert, L.; Solyom, Z.; Favier, A.; Ayala, I.; Weinhäupl, K.; Schanda, P., et al., RNA binding and chaperone activity of the E. coli cold-shock protein CspA. *Nucleic Acids Res.* **2017,** *45* (7), 4255-4268.

4. Phadtare, S.; Severinov, K., RNA remodeling and gene regulation by cold shock proteins. *RNA Biol.* **2010,** *7* (6), 788-795.

5. Keto-Timonen, R.; Hietala, N.; Palonen, E.; Hakakorpi, A.; Lindstrom, M.; Korkeala, H., Cold shock proteins: A minireview with special emphasis on Csp-family of enteropathogenic Yersinia. *Front. Microbiol.* **2016,** *7*, 1151.

6. Xie, B.; Calabro, V.; Wainberg, M. A.; Frankel, A. D., Selection of TAR RNA-binding chameleon peptides by using a retroviral replication system. *J. Virol.* **2004,** *78* (3), 1456-63.

7. Ito, K.; Uno, M.; Nakamura, Y., A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **2000,** *403* (6770), 680-4.

8. Vieweger, M.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule FRET reveals three conformations for the TLS domain of brome mosaic virus genome. *Biophys. J.* **2015,** *109*, 2625-2636.

9. Rhine, K.; Vidaurre, V.; Myong, S., RNA droplets. *Annu. Rev. Biophys.* **2020,** *49* (1), 247-265.

10. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X., Targeting RNA G-quadruplex in SARS-CoV-2: A promising therapeutic target for COVID-19? *Angew. Chem. Int. Ed.* **2021,** *60* (1), 432-438.

Appendix 3 Analytic Correction to Rate Constants Determined by Dwell Time Analysis

1. Shuang, B.; Cooper, D.; Taylor, J. N.; Kisley, L.; Chen, J.; Wang, W.; Li, C. B.; Komatsuzaki, T.; Landes, C. F., Fast step transition and state identification (STaSI) for discrete single-molecule data analysis. *J. Phys. Chem. Lett.* 2014, *5* (18), 3157-3161.

Appendix 1

The rate constant for fluorescence resonance energy transfer (FRET)

In this appendix, I will provide a derivation of the rate constant for FRET, which is

$$k_{FRET} = \frac{9\times10^3 \ln(10) k_{rad,D} \kappa^2}{128\pi^5 n^4 r_{DA}^6 N_A} J, \quad \text{(Eq. A1.1)}$$

where $k_{rad,D}$ is the radiative rate constant of the donor excited state, κ^2 is the donor-acceptor orientation factor, J is the donor-acceptor spectral overlap integral, n is the index of refraction, r_{DA} is the distance between the donor and acceptor, and N_A is Avogadro's number. The expression for κ^2 is

$$\kappa^2 = (\cos\theta_{AD} - 3\cos\theta_D \cos\theta_A)^2, \quad \text{(Eq. A1.2)}$$

where θ_{AD} is the angle between the donor and acceptor transition dipole moments, θ_D (θ_A) is the angle between the donor (acceptor) transition dipole moment and the displacement vector from the donor to acceptor. The overlap integral J is

$$J = \int_0^\infty d\lambda \, \overline{F}_D(\lambda) \varepsilon_A(\lambda) \lambda^4, \quad \text{(Eq. A1.3)}$$

Where $\overline{F}_D(\lambda)$ is the donor emission spectrum normalized to an area of 1, and $\varepsilon_A(\lambda)$ is the acceptor absorbance molar extinction coefficient. In this appendix, I will derive the FRET rate constant expression using a semiclassical approach based on Fermi's golden rule.

Fermi's golden rule states that the transition rate $\Gamma_{i\to f}$ for transitions due to a perturbing quantum Hamiltonian $\widehat{H}$ between states i and f is

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar^2} \left|\langle f \left| \widehat{H} \right| i \rangle\right|^2, \quad \text{(Eq. A1.4)}$$

where $\hbar$ is the reduced Planck's constant ($\hbar = h/2\pi$), $|i\rangle$ and $|f\rangle$ are the initial and final quantum states, respectively, and $\widehat{H}$ is the coupling Hamiltonian. In FRET, the initial state $|i\rangle$ is the direct

product of the donor excited state |D*⟩ and acceptor ground state |A⟩ and the final state |*f*⟩ is the direct project of the donor ground state |D⟩ and acceptor excited state |*A⟩:

$$|i\rangle = |D^*\rangle \times |A\rangle \qquad \text{(Eq. A1.5)}$$

$$|f\rangle = |D\rangle \times |A^*\rangle \qquad \text{(Eq. A1.6)}$$

The coupling Hamiltonian is the donor-acceptor dipole-dipole interaction, i.e., the energy of the dipole moment of the acceptor in the electric field generated by the donor dipole:

$$\widehat{H} = \hat{V}_{dipole-dipole} = -\widehat{\boldsymbol{\mu}}_A \cdot \boldsymbol{E}_D = -\widehat{\boldsymbol{\mu}}_A \cdot \left(\frac{1}{4\pi\varepsilon_0 n^2}\right)\left(\frac{1}{r_{DA}^3}\right)[3(\widehat{\boldsymbol{\mu}}_D \cdot \hat{r}_{DA})\hat{r}_{DA} - \widehat{\boldsymbol{\mu}}_D]$$

$$= \hat{\mu}_A\hat{\mu}_D\left(\frac{1}{4\pi\varepsilon_0 n^2}\right)\left(\frac{1}{r_{DA}^3}\right)[\underbrace{\cos\theta_{AD} - 3\cos\theta_D\cos\theta_A}_{\kappa}], \qquad \text{(Eq. A1.7)}$$

where $\hat{\mu}_D$ and $\hat{\mu}_A$ are the dipole moment operators for the donor and acceptor, respectively, ε_0 is the vacuum permittivity, and *n* is the index of refraction of the medium.* Evaluating the matrix element $\widehat{H}_{i,f}$ and taking the square modulus yields

$$\left|\langle f|\widehat{H}|i\rangle\right|^2 = \mu_A^2\mu_D^2\left(\frac{1}{4\pi\varepsilon_0 n^2}\right)^2\left(\frac{1}{r_{DA}^6}\right)\kappa^2. \qquad \text{(Eq. A1.8)}$$

Next, we treat the rate constant for FRET as the contribution of wavelength-specific transition rates: $k_{FRET} = \int_0^\infty d\lambda\, \Gamma(\lambda)$, where $\Gamma(\lambda)$ is the hypothetical transition rate if donor and acceptor electronic transitions occur at a single wavelength λ. So far, by substituting A1.8 into A1.4, we have

$$\Gamma(\lambda) = \left(\frac{2\pi}{\hbar^2}\right)\left(\frac{1}{4\pi\varepsilon_0}\right)^2\left(\frac{\kappa^2}{n^4 r_{DA}^6}\right)\mu_A^2\mu_D^2. \qquad \text{(Eq. A1.9)}$$

* Normally, one would write $4\pi\varepsilon_0 n^2$ as $4\pi\varepsilon(\omega)$, where $\varepsilon(\omega)$ is the (frequency-dependent) dielectric constant of the medium. However, the symbol $\varepsilon(\omega)$ is already used to describe the extinction coefficient, so I have replaced $\varepsilon(\omega)$ by $\varepsilon_0 n^2$ (which one would normally do at the end of the derivation), using the fact that the relative dielectric constant, $\varepsilon_R(\omega) \equiv \varepsilon(\omega)/\varepsilon_0$, is approximately the square of the index of refraction *n* of the medium at the optical frequency of the transition.

We now replace the transition dipole moments with experimentally accessible quantities using several well-known expressions:[†]

$$\mu_D^2 = \left(\frac{3\varepsilon_0 \hbar c^3}{2\omega^3}\right) k_{rad,D} \overline{F}_D(\lambda), \text{ and} \qquad \text{(Eq. A1.10)}$$

$$\mu_A^2 = \left(\frac{3\varepsilon_0 \hbar c}{\pi\omega}\right) \sigma_A(\lambda), \qquad \text{(Eq. A1.11)}$$

where c is the speed of light in vacuum, ω is the angular frequency, $k_{rad,D}$ is the donor radiative rate constant, $\overline{F}_D(\lambda)$ is the fluorescence spectrum of the donor, normalized so that $\int_0^\infty d\lambda\, \overline{F}_D(\lambda) = 1$, and $\sigma_A(\lambda)$ is the acceptor absorbance cross section. To convert the absorbance cross section (in cm^2) into a molar extinction coefficient $\varepsilon_A(\lambda)$ (in $M^{-1}cm^{-1}$), we use the relation

$$\sigma_A(\lambda) = \frac{10^3 \ln(10)}{N_A} \varepsilon_A(\lambda), \qquad \text{(Eq. A1.12)}$$

where N_A is Avogadro's number. Substituting Eqs. A1.10–12 into Eq. A1.9, we obtain

$$\Gamma(\lambda) = \left(\frac{9\times10^3 \ln(10)}{8\pi N_A}\right)\left(\frac{\kappa^2 k_{rad,D}}{n^4 r_{DA}^6}\right)\left(\frac{\lambda^4}{(2\pi)^4}\right) \overline{F}_D(\lambda)\varepsilon_A(\lambda), \qquad \text{(Eq. A1.13)}$$

where I have used the fact that $\left(\frac{c}{\omega}\right)^4 = \left(\frac{\lambda}{2\pi}\right)^4$. Finally, integrate over all wavelengths to produce

$$k_{FRET} = \left(\frac{9\times10^3 \ln(10)}{128\pi^5 N_A}\right)\left(\frac{\kappa^2 k_{rad,D}}{n^4 r_{DA}^6}\right) \int_0^\infty d\lambda\, \overline{F}_D(\lambda)\varepsilon_A(\lambda)\lambda^4, \qquad \text{(Eq. A1.14)}$$

which is the desired result.

Note that the key terms in k_{FRET} come from the physics of the dipole-dipole interaction: the r_{DA}^{-6} factor is the square of the r^{-3} decay of the dipole field, the κ^2 factor is the square of the dipole-dipole coupling projection, n^4 is the square of the relative dielectric constant at optical frequency, and finally $k_{rad,D}$ and $\varepsilon_A(\lambda)$ are related to the donor and acceptor transition dipole moments, respectively.

[†] Hilborn, R. C., Einstein coefficients, cross sections, f values, dipole moments, and all that. *Am. J. Phys.* **1982,** *50* (11), 982-986.

Appendix 2

Signal-to-noise ratio (S/N) degradation due to noisy amplifiers

---

An ideal amplifier multiplies an input signal by a fixed factor *G*, which is the gain. This increases this signal but also equally the noise, and the net result is that the signal-to-noise ratio (S/N) of the input signal is unchanged by the amplifier. However, a nonideal amplifier has a gain factor *G* which is variable and as a result introduces a loss of S/N. In this appendix, I will evaluate the S/N for nonideal amplification of an input signal with Poisson noise ("shot noise"). This model is relevant to this thesis as it describes how light is amplified by the intensifier in the charge-coupled device (CCD).

Consider a Poisson random variable *N*, i.e., *N* has an expectation value $\langle N \rangle$ and standard deviation $\sqrt{\langle N \rangle}$. *N* determines the number of times that a random variable *G* (the single-count gain) is added to itself to produce the output signal *S*:

$$S = G_1 + G_2 + \cdots + G_N = \sum_{i=1}^{N} G_i. \qquad \text{(Eq. A2.1)}$$

In other words, for *N* detected counts, the amplified output is the sum of *N* independent and identically distributed random variables G_i. This is known as a *compound Poisson process*. The gain distribution *G* can be any distribution with finite expectation value $\langle G \rangle$ and variance σ_G^2.

First, compute the average signal $\langle S \rangle$. The conditional expectation of *S* given *n* counts is

$$\langle S|n \rangle = \langle \sum_{i=1}^{n} G_i \rangle = \sum_{i=1}^{n} \langle G_i \rangle = n \langle G \rangle, \qquad \text{(Eq. A2.2)}$$

where the expectation value can be exchanged with the summation because *n* is not a random variable. Next, use the conditional expectation $\langle S|n \rangle$ to calculate the expectation value of *S*:

$$\begin{aligned} \langle S \rangle &= \sum_{n=0}^{\infty} \langle S|n \rangle P_N(n) = \sum_{n=0}^{\infty} n \langle G \rangle P_N(n) \\ &= \langle G \rangle \sum_{n=0}^{\infty} n P_N(n) = \langle G \rangle \langle N \rangle. \end{aligned} \qquad \text{(Eq. A2.3)}$$

This is an intuitive result: the average signal is the average number of counts multiplied by the average gain per count.

Now, calculate the noise in the signal. This is done using the law of total variance,

$$\sigma_S^2 = \langle \sigma_{S|n}^2 \rangle + \sigma_{\langle S|n \rangle}^2, \quad \text{(Eq. A2.4)}$$

which states that the total variance σ_S^2 is equal to the expectation of the conditional variance $\langle \sigma_{S|n}^2 \rangle$ plus the variance of the conditional expectation $\sigma_{\langle S|n \rangle}^2$. The conditional variance is

$$\sigma_{S|n}^2 = n\, \sigma_G^2, \quad \text{(Eq. A2.5)}$$

Because $S|n$ is the sum of *n* independent copies of *G* each with variance σ_G^2. The expectation of the conditional variance is therefore (by a calculation similar to that for $\langle S \rangle$ above)

$$\langle \sigma_{S|n}^2 \rangle = \sigma_G^2 \langle N \rangle. \quad \text{(Eq. A2.6)}$$

The variance of the conditional expectation $\sigma_{\langle S|n \rangle}^2$ is calculated from the definition of variance and the fact that $\langle S|n \rangle = n\langle G \rangle$ from Eq. A2.2:

$$\sigma_{\langle S|n \rangle}^2 = \sigma_{n\langle G \rangle}^2 = \langle (n\langle G \rangle - \langle n\langle G \rangle \rangle)^2 \rangle = \langle (n\langle G \rangle - \langle G \rangle \langle N \rangle)^2 \rangle$$

$$= \langle G \rangle^2 \underbrace{\langle (n - \langle N \rangle)^2 \rangle}_{\equiv \sigma_N^2} = \langle G \rangle^2 \langle N \rangle. \quad \text{(Eq. A2.7)}$$

Therefore, the variance in *S* is

$$\sigma_S^2 = \sigma_G^2 \langle N \rangle + \langle G \rangle^2 \langle N \rangle = \langle N \rangle (\langle G \rangle^2 + \sigma_G^2), \quad \text{(Eq. A2.8)}$$

Finally, compute the signal-to-noise ratio S/N of the amplified signal,

$$\frac{\langle S \rangle}{\sigma_S} = \frac{\langle G \rangle \langle N \rangle}{\sqrt{\langle N \rangle \left(\langle G \rangle^2 + \sigma_G^2 \right)}} = \sqrt{N} \frac{1}{\sqrt{1 + \left(\frac{\sigma_G}{\langle G \rangle} \right)^2}}, \quad \text{(Eq. A2.9)}$$

and compare this to the S/N of the Poisson input signal,

$$\frac{\langle N \rangle}{\sigma_N} = \sqrt{N}. \quad \text{(Eq. A2.10)}$$

The output S/N is equal to the input S/N multiplied by a factor $\left(\sqrt{1+\left(\frac{\sigma_G}{\langle G\rangle}\right)^2}\right)^{-1}$, which is the gain noise factor. This value is always less than or equal to 1, meaning that the amplifier can only decrease the S/N. For an ideal amplifier, σ_G = 0 and the gain noise factor is equal to 1. However, many gain sources such as photomultiplier tubes and multichannel plates have exponential gain distributions for which $\sigma_G = \langle G\rangle$. In this case, the gain noise factor is $1/\sqrt{2}$, or approximately a 30% loss in signal-to-noise ratio.

Appendix 3

Analytic correction to rate constants determined by dwell time analysis

---

Consider a kinetic system governed by first order rate processes with rate matrix **K**. For state *i*, the survival function $S_i(n)$ is the probability that the system is observed in state *i* for more than *n* consecutive frames, $P(n_i > n)$. If the underlying dynamics are Markovian, then this probability can be expanded in a product of single frame survival probabilities

$$S_i(n) = P(n_i > n) = [P_{i\to i}(\Delta t_{\text{frame}})]^{\text{n}}, \qquad \text{(Eq. A3.1)}$$

where $P_{i\to i}(\Delta t_{\text{frame}})$ is the probability that the system is observed to be in state *i* in one frame and again in state *i* one frame later. If we rearrange Eq. A3.1 and change variables from number of frames to time ($n = t/\Delta t_{\text{frame}}$), the survival function becomes

$$S_i(t) = e^{\left(\frac{1}{\Delta t_{\text{frame}}}\right)\ln\left(P_{i\to i}(\Delta t_{\text{frame}})\right)t} = e^{-k_i^{\text{app}}*t}, \qquad \text{(Eq. A3.2)}$$

which decays with apparent rate constant k_i^{app}.

The single frame self-transition probability $P_{i\to i}(\Delta t_{\text{frame}})$ is the diagonal element T_{ii} of the transition matrix **T** evaluated at lag time Δt_{frame}:

$$P_{i\to i}(\Delta t_{\text{frame}}) = \mathbf{T}_{\text{ii}}(\Delta t_{\text{frame}}) = (e^{\mathbf{K}\Delta t_{\text{frame}}})_{\text{ii}}. \qquad \text{(Eq. A3.3)}$$

Combining Eqs. A3.2 and A3.3, we arrive at the following relationship between k_i^{app} and T_{ii}:

$$T_{\text{ii}} = e^{-k_i^{\text{app}}*\Delta t_{\text{frame}}}. \qquad \text{(Eq. A3.4)}$$

Armed with the diagonal matrix elements T_{ii}, we now construct the full transition matrix **T**. Since the columns of **T** must each sum to one, this is simple for a two-state system: $T_{\text{ji}} = 1\text{-}T_{\text{ii}}$. For a system of more than two states, one must use the empirical branching ratios η_{ji} for transitions from state *i* to each other state *j*, in which case the off-diagonal elements are computed as $T_{\text{ji}} = \eta_{\text{ji}}\ (1\text{-}T_{\text{ii}})$. Determination of the branching ratios requires a more sophisticated

change point analysis than the simple thresholding method used here (e.g. STaSI).[1] After computing **T**, the final step is to recover **K** using the matrix logarithm,

$$\mathbf{K} = \frac{\ln(\mathbf{T})}{\Delta t_{\text{frame}}}, \qquad \text{(Eq. A3.5)}$$

which completes our derivation for the analytic correction. Note that for small frame times, the transition matrix can be well approximated by $\mathbf{T} \approx \mathbf{I} + \mathbf{K}\Delta t_{\text{frame}}$. In this case, the single frame transition probabilities simplify to $T_{\text{ii}} = 1 - \sum_{j \neq i} k_{ij}$, and the apparent rate constant k_i^{app} will be the sum of the rate constants leaving state *i*. Therefore, in the slow rate constant regime ($-K_{\text{ii}} << k_{\text{frame}}$), the apparent rate constant will accurately reproduce the true rate of depletion of each state, and no correction is necessary.

Finally, it may be of interest to some to instead consider the histogram of dwell times $H_i(n)$, which consists of the number of dwell times in state *i* of length *n*. Dividing $H_i(n)$ by the total number of dwell times yields the probability mass function $P(n_i = n)$, which represents the probability of a dwell time in state *i* lasting exactly *n* bins. Similarly to our above treatment of $S_i(n)$, $P(n_i = n)$ can be expanded into a product of single frame self-transition probabilities as $T_{\text{ii}}^{(n-1)}(1-T_{\text{ii}})$, where the factor $(1-T_{\text{ii}})$ accounts for the termination of the dwell period. Therefore, the histogram $H_i(n)$ will also decay exponentially according to the same apparent rate constant k_i^{app}, with the remainder of the derivation identical as for analysis of the survival function $S_i(n)$.

References

1. Shuang, B.; Cooper, D.; Taylor, J. N.; Kisley, L.; Chen, J.; Wang, W.; Li, C. B.; Komatsuzaki, T.; Landes, C. F., Fast step transition and state identification (STaSI) for discrete single-molecule data analysis. *J. Phys. Chem. Lett.* **2014,** *5* (18), 3157-3161.

Appendix 4

Model-free, graphical analysis of ligand binding stoichiometry using preferential interaction coefficients

---

Nucleic acid folding is often accompanied by the binding or release of ligands, which may be ions, small molecules, or proteins. By measuring the dependence of the folding equilibrium constant K_{eq} on the concentration of a binding partner, an experimentalist can learn about the strength (e.g., dissociation constant K_d) and stoichiometry (i.e., number of bound ligands) of the underlying binding process. Often, this analysis is carried out by fitting a curve to a plot of K_{eq} versus the concentration of ligand $[L]$. The functional form for the fitting curve depends on which binding model is selected to represent the binding interaction. For instance, one may use a "bind-then-fold" model,

$$U + nL \rightarrow U \cdot L_n \rightarrow F \cdot L_n, \qquad \text{(Eq. A4.1)}$$

or a "fold-then-bind" model,

$$U + nL \rightarrow F + nL \rightarrow F \cdot L_n. \qquad \text{(Eq. A4.2)}$$

Here, U is the unfolded state, F is the folded state, L is the ligand, n is the ligand binding stoichiometry, and $U{\cdot}L_n$ and $F{\cdot}L_n$ are the ligand-bound unfolded and folded states, respectively. These two models have different K_{eq} behavior at $[L] = 0$ and $[L] = \infty$ and can therefore be experimentally distinguished. However, these are only two of infinitely many possible binding models, and other models may have similar dependences on concentration. Therefore, while some mechanistic information can be inferred from model-dependent fitting, resolving the true binding model with certainty is extremely difficult.

The method of *preferential interaction coefficients* is a way to determine ligand binding stoichiometry using only the slope on a log-log plot of K_{eq} vs. $[L]$. In this method, one sacrifices mechanistic detail (i.e., the exact set of states in the binding model) to instead obtain an averaged ligand binding stoichiometry, i.e., the difference in the average number of ligands bound to the unfolded and folded states, respectively. This method relies on the key result that

$$\Delta n = \frac{d \ln K_{eq}}{d \ln [L]}, \qquad \text{(Eq. A4.3)}$$

which states that the average ligand binding stoichiometry (also called the preferential interaction coefficient), Δn, is the logarithmic derivative of the folding equilibrium constant, K_{eq}, with respect to the logarithm of ligand concentration, $[L]$. This powerful expression is often stated without proof, and the proofs given in the literature usually invoke thermodynamic arguments which are highly abstract and, as a result, unfortunately difficult to understand for those not well-versed in thermodynamics. In this appendix, I will provide an alternative derivation of this result which relies only on the law of mass action, and which therefore will be accessible to anyone who has taken a university-level course in general chemistry.

Consider the binding model shown in Figure A4.1, which includes every possible binding stoichiometry between a ligand *L* and two receptor configurations *U* and *F*. This is the most general possible binding model, and therefore any result derived from this model will be true of any binding model. Each process in this network must satisfy a corresponding equilibrium

$$\begin{array}{ccccccc} U & \leftrightarrow & U \cdot L & \leftrightarrow & U \cdot L_2 & \leftrightarrow & U \cdot L_3 \leftrightarrow \cdots \\ \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\ \mathrm{F} & \leftrightarrow & F \cdot L & \leftrightarrow & F \cdot L_2 & \leftrightarrow & F \cdot L_3 \leftrightarrow \cdots \end{array}$$

Figure A4.1 General binding model of ligand *L* binding to *U* and *F* receptor configurations. Note that free ligands are omitted from this scheme for clarity.

constant expression. Let $K_{n,U}$ be the equilibrium constant for the binding of n ligands to the unfolded state ($U + nL \rightarrow U{\cdot}L_n$). From the law of mass action, write

$$K_{n,U} = \frac{[U{\cdot}L_n]}{[U][L]^n}. \qquad \text{(Eq. A4.4)}$$

This expression assumes ideal (dilute) conditions. To be rigorously correct, one should replace concentrations with thermodynamic activities in this expression and all that follow, but the mathematics remains the same. These ligand-binding equilibrium constants $K_{n,U}$ fully characterize the distribution of ligand-bound states in the U manifold of states. Similarly, the F manifold of states is characterized by another infinite set of equilibrium constants $K_{n,F}$ for ligand binding to the folded configuration. Finally, for each ligand-bound unfolded state $U{\cdot}L_n$, there is a folding process $U{\cdot}L_n \rightarrow F{\cdot}L_n$ which has its own equilibrium constant. However, there is only one degree of freedom for these equilibrium constants, as it suffices to specify one of these equilibrium constants before the remainder are determined by the fact that the product of equilibrium constants around any cycle of states must equal 1, due to free energy being a state function. For the purpose of this appendix, the one specified equilibrium constant is the no-ligand ("*apo*") folding equilibrium constant

$$K_{apo} = [F] / [U]. \qquad \text{(Eq. A4.5)}$$

Note that this is *not* equal to the empirically observed folding equilibrium constant,

$$K_{eq} = [F_{total}] / [U_{total}], \qquad \text{(Eq. A4.6)}$$

which is the ratio of total folded population (i.e., $[F_{total}] = [F] + [F{\cdot}L] + [F{\cdot}L_2] + \ldots$) to the total unfolded population (i.e., $[U_{total}] = [U] + [U{\cdot}L] + [U{\cdot}L_2] + \ldots$).

Begin by computing the average number of bound ligands in the manifold of unfolded states:

$$\langle n \rangle_U = P(U) * 0 + P(U \cdot L) * 1 + P(U \cdot L_2) * 2 + \cdots$$

$$= \sum_{n=1}^{\infty} P(U \cdot L_n) * n. \qquad \text{(Eq. A4.7)}$$

This is the average over each state $U{\cdot}L_n$ of the number of ligands bound to that state n weighted by the probability of being in that state $P(U{\cdot}L_n)$. Now using the fact that

$$P(U \cdot L_n) = \frac{[U{\cdot}L_n]}{[U_{total}]}, \qquad \text{(Eq. A4.8)}$$

along with the equilibrium constant expressions (Eq. A4.4), write

$$\langle n \rangle_U = \frac{0[U]+1[U{\cdot}L]+2[U{\cdot}L_2]+\cdots}{[U]+[U{\cdot}L]+[U{\cdot}L_2]+\cdots}$$

$$= \frac{0+K_{1,U}[L]+2K_{2,U}[L]^2+\cdots}{1+K_{1,U}[L]+K_{2,U}[L]^2+\cdots} = \frac{\sum_{n=0}^{\infty} nK_{n,U}[L]^n}{1+\sum_{n=1}^{\infty} K_{n,U}[L]^n}. \qquad \text{(Eq. A4.9)}$$

Now utilize the fact that the numerator is the derivative of the denominator with respect to the logarithm of $[L]$, and that the denominator is equal to $[U_{total}]/[U]$:

$$\langle n \rangle_U = \frac{d/_{d\ln[L]}\left(1+\sum_{n=1}^{\infty} K_{n,U}L^n\right)}{1+\sum_{n=1}^{\infty} K_{n,U}L^n}$$

$$= \frac{d}{d\ln[L]} \ln\left(1+\sum_{n=1}^{\infty} K_{n,U}L^n\right) = \frac{d\ln\left(\frac{[U_{total}]}{[U]}\right)}{d\ln[L]}. \qquad \text{(Eq. A4.10)}$$

A similar expression can be derived for the folded state manifold.

The derivation is finished by calculating the difference of these two binding averages and simplify to get the desired result:

$$\Delta n \equiv \langle n \rangle_F - \langle n \rangle_U = \frac{d}{d\ln[L]}\left[ln\left(\frac{[F_{total}]}{[F]}\right) - ln\left(\frac{[U_{total}]}{[U]}\right)\right]$$

$$= \frac{d}{d\ ln[L]}\left[ln\left(\frac{[F_{total}]}{[U_{total}]}\right) - ln\left(\frac{[F]}{[U]}\right)\right]. \qquad \text{(Eq. A4.11)}$$

Here, recognize $[F_{total}]$ / $[U_{total}]$ as the observed equilibrium constant K_{eq}, and $[F]$ / $[U]$ as the equilibrium constant for folding in the *apo* state K_{apo} which is independent of ligand concentration. Therefore,

$$\Delta n = \frac{d}{d\ ln[L]}\left[ln\,K_{eq} - ln\,K_{apo}\right] = \frac{d\ln K_{eq}}{d\ln[L]}, \qquad \text{(Eq. A4.12)}$$

which completes the derivation. Note that Δn is itself a function of $[L]$ and is not guaranteed to be an integer.