DNA HAIRPIN FOLDING KINETICS AND MECHANISMS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Michael Ke Zhang
December 2012

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Hideo Mabuchi, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sebastian Doniach

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Daniel Herschlag

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

The nucleic acid hairpin structure, composing of a single-stranded loop and a base-paired stem, plays an important role in DNA replication [BLBM10] and is a building block of the tertiary structure of large RNA molecules[SC06, BB08]. The biological functions of the hairpin depend on what the state the molecule is in. Thus detailed understanding its folding kinetics and mechanisms facilitates investigations on cellular processes that involve the hairpin structure. DNA hairpin folding has been a subject of intense research, but the current literature does not provide a clear understanding of the folding[OJ08]. This is because that the hairpin folding may involve a large ensemble of intermediate states with vastly different characteristic lifetimes. This unexpected complexity requires advanced techniques that can probe folding in both fast (from nanoseconds to milliseconds) and slow time (more than 1 milliseconds) scales, and previous experiments did not have the advanced techniques to fully probe the folding of the hairpin.

The first part of the thesis is a comprehensive study of DNA hairpin folding kinetics. Specifically, we used a multi-faceted and unifying approach that tests each aspect of a general folding model encompassing each of the previously proposed conformational states. Given the complexity in DNA hairpin folding, we have used the following approaches: 1) we extend the temporal range of fluorescence correlation spectroscopy (FCS) by two orders of magnitude to more than 100 mS (for a typical protein diffusing at 50 $\mu m^2/S$). 2) We use both fluorescence quenching and resonance energy transfer to distinguish between proposed structural states; 3) We systematically vary the sequence of the hairpin, both its base pairing and tether regions, to test for the formation of specific intermediate species. 4) We vary ionic conditions

to vary the energy landscape and reveal different species and transitions. 5) We follow the behavior of individual molecules for still longer times by surface-tethering molecules and monitoring their folding by con-focal microscopy. Our results reveal a surprisingly simple behavior of a short DNA hairpin of three base pairs. The DNA transitions between a random coil state and a fully folded hairpin, without significant accumulation ($< 1\%$) of misfolded or partially folded intermediate states. It is also discovered that the folding and diffusion processes are coupled, which could be the main factor contributing to the many discrepancies in the literature. Finally, it is revealed that there are three different salt ranges in which the hairpin behaved quite differently. The cations in the solution modulate both the global conformation of the hairpin and the local formation of the stem.

In the second part of the thesis, I will present an experimental technique that is called tracking-FRET, which allows extended measurement of molecular conformations by FRET or quenching while tracking the freely diffusing molecular complex in the solution. I will present the basics of the tracking, covering necessary details for us to understand later tracking experiments. Theories to model the tracking system dynamics, compute tracking FRET FCS, and extract the folding dynamics are also presented. Finally, I will present two sets of tracking experiments on the same DNA hairpins we studied in the first part of the thesis. In both experiments, there is overwhelming amount of evidence that we have observed folding dynamics, demonstrating that tracking can be a very powerful tool in learning the folding of macromolecules. However the inconsistency between the tracking and solution FCS results suggests that further development of the tracking apparatus is desired.

Acknowledgements

Looking back my six years at Stanford, I am grateful that I was able to be part of the Mabuchi Lab and contribute to science. What I gained, however, is much beyond research: I learned how to approach work and life; it prepared me for the future. Plus, I was married in graduate school - how fortunate I am! Now I would like to take this chance to say thank you...

I thank Hideo for being the best advisor/boss I could ever hope: intelligent, patient, and inspiring. You taught me things from science to wine to English. By the way, thanks for the blue bottle coffee. Now I know it's the best stuff. I am forever grateful and truly in debt. I thank you Dan, for your appreciation of my work and great suggestions in the hairpin project. Your students are incredible.

I thank Ying for love and being the best friend. Thanks for taking good care of me. Thanks for the sandwiches you made for me. Feed me more music (and hug) please?

I thank all of the members in the Mabuchi Lab. I was this undergraduate who actually did some stuff in the lab (I still hold the record of assembling the biggest vacuum chamber), and now I am a graduate! I thank Charles for collaboration and dinning hall. I thank Dr.Mikey - from teaching me how to use an oscilloscope to control theory. I thank Kevin for coaching. Brother, you are the best! I thank Max for collaboration and being my wedding officiant. I thank Orion and Dmitri for chat, Yeong-Dae for cash for food. I thank Joe for jokes on g-chat. I thank Andy for lunch and the thesis. I cannot name everyone here, but hey I have your photo here. You know that I am a true friend.

I thank my not-in-lab friends for support and unconditional friendship. Chinese

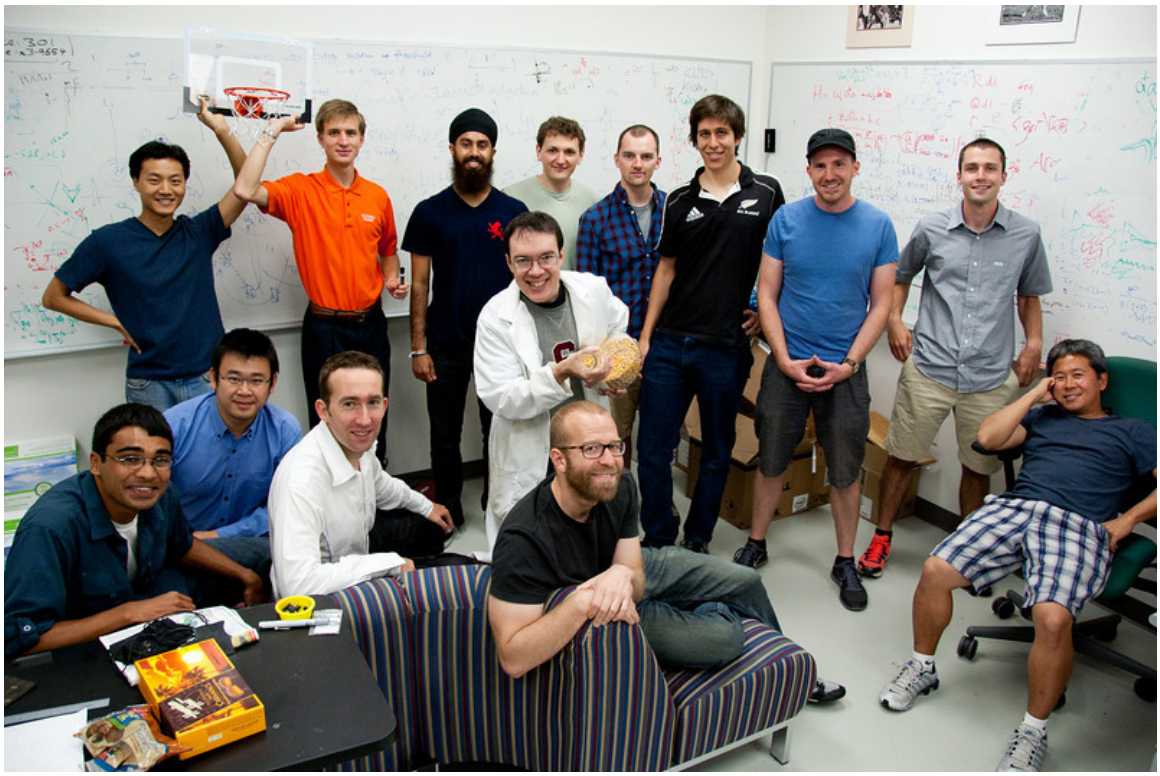

Figure 1: The Mabuchi Group in Summer 2012. This is photographed in the group room in the Nano Center at Stanford. We are missing Jie Wu.

New Year. Trips to Tahoe and Yellowstone. Precious memories. This journey would be much less colorful without you.

I thank all my past mentors and teachers. Eric Black, Ken Libbrecht, Dr.K, Mrs.Chen. You shaped me.

Last but certainly not least, I thank my mom for love, support, and life. I would never forget your persistence and sacrifice for me. I thank my dad for raising me and guiding me through the years. I thank my step-mom for everything she has done for me - washing clothes and cooking etc.

This thesis is for my curiosity in science. I hope I could maintain this curiosity wherever I go and whatever I do.

Contents

List of Tables

List of Figures

1

Motivation, History and Organization

1.1 Motivation

In the context of cellular and molecular biology, single molecule methodologies are interdisciplinary approaches to investigate biomolecules one at a time, obtaining rich information in the folding, assembly, dynamics, and functions of bio-macromolecules. It has found applications from basic science to commercial products such as PacBio single molecule DNA sequencing machines[EFG+09]. In the past two decades, this field has enjoyed incredible development in both the number of systems it investigated and the flourishment of the new experimental techniques and theoretical advancement[JBI+08]. Results documented in this thesis, conducted in the laboratory of Professor Hideo Mabuchi in Applied Physics and Professor Daniel Herschlag in Biochemistry at Stanford, provides a detailed study of the folding of the nucleic acid hairpin, with the intention to develop a new single molecule method that could significantly advance our ability in investigating the folding of macromolecules using fluorescence: tracking-FRET.

Macromolecules such as protein and nucleic acid need to fold into correct, "native" three-dimensional structure in order to perform their intended functions, such

as catalysis and translation. How the correct folding is achieved in the cell is fascinating, because the mind-blowing complexity it involved yet how fast and robust it can be. For example (this is a famous thought experiment coined as Levinthal's paradox[Lev68]), for a 100 monomer polypeptide chain, it can adopt at least 10^{69} conformations, assuming that each peptide bond can adopt to one of the three stable angles. If these conformations are all explored, even rapidly in the rate of 10^{15} per second (which is how fast the bonds vibrate), it would take longer than the entire lifetime of the universe for the peptide chain to find its native state. Yet, in the cell, these folding reactions typically complete in the millisecond to second scale [SR98]. This speed and robustness certainly indicates that these macromolecules do not randomly folds into the native state. It is not a random trial-and-error. There are must be mechanisms that guide the molecules to their final states. Although it is not clear whether proteins fold by going through directed (but unpredictable) search by going downhill on the folding energy landscape or going through defined and predictable intermediate states, it is certain that for complex three-dimensional structures to form, it involves with complex interactions among the monomers (such as forming secondary structures like α-helix), interactions with the solvent and cations, and sometimes even assists with molecular chaperoons[Dob03].

It is important to note that the structure of macromolecules such as protein is not static, rather it is constantly in motion. Sometimes it is in motion simply because of thermal agitation: vibration of the atoms and chemical functional groups occurs at sub nanometer scale and can be as fast as femtoseconds; sometimes it is in motion because that the molecule needs to perform its designed cellular functions: moving of sidechains of a protein due to ligand binding. These motions vary greatly in both length and time scales yet coexist in one molecule. As we will see later, the spread of these length and time scales on one molecule impose stringent requirements on experimental methods to study them. It is also important to note that these physical motions are directly linked to the functionality of the molecules. One example that is directly relevant to our daily lives is the conformation changes of G-protein coupled receptors (GPCRs), which is the target of more than 50% of

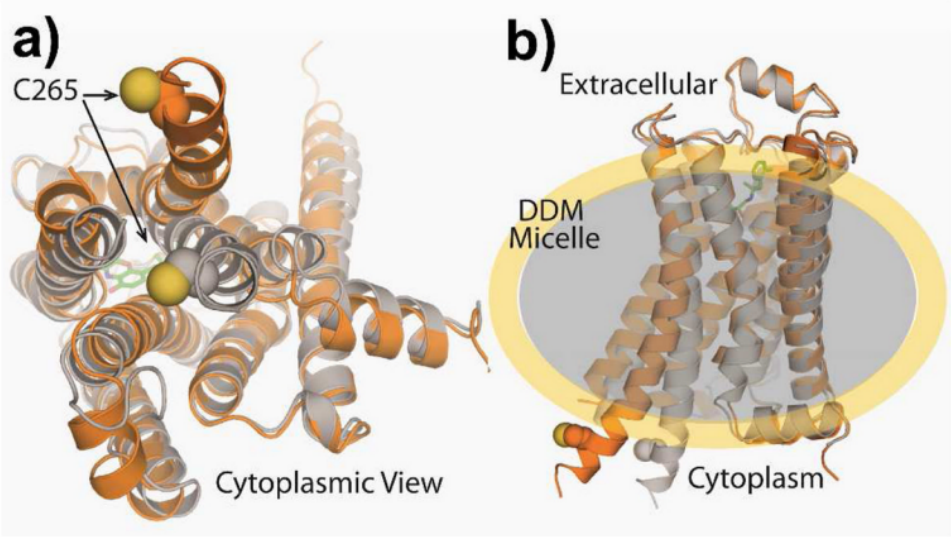

Figure 1.1: Conformational changes of G protein coupled receptors upon ligand binding. This figure is adopted from Ref.[BFY+11]. The orange shows the active state of the molecule while the grey shows the inactive state. The small green molecule at the center is the binding ligand. a) is the viewed from the cytoplasmic side of the cell while b) provides a side view. The particular GPCR shown here is β_2AR. The yellow dots are TMR, which probed the conformational changes of β_2AR at single molecule level.

the modern drags and accounts for top 25% of the top 100 best selling drags worldwide(http://zhanglab.ccmb.med.umich.edu). Upon the binding of the extracellular ligand (which can be a single photon in the retina, odors, hormones, and neurotransimitters), the conformation of this protein changes, activating the intracellular G proteins, which leads to further downstream signal transduction and cellular scale response. Fig.1.1, adopted from Ref.[BFY+11], illustrates this conformation changes of this molecule upon activation of the binding of the ligand. Detailed understanding of this conformational change, such as conformational states involved and the kinetics of folding, has lead to insightful understanding of this family of molecules and new drag discoveries.

There have been much advancement in single molecule experimental methods to probe folding dynamics[DML08], and one method perhaps has enjoyed more popularity than others: fluorescence microscopy. The idea is simple: some properties of

the fluorescence, such as polarization, lifetime, or fluorescing rate, are modulated by the folding of the molecule. For example, in quenching experiment, the rate of fluorescence rate is changed if two points on the molecule are brought into Van del Waals distance[KMS00]. This is useful in monitoring collision rate of pairs of points of interest. Despite its popularity, however, we need to note that there are many serious challenges in this method in probing folding. Some challenges are imposed simply by chemistry, while some challenges are imposed by experimental techniques. We summarize them here to elucidate why tracking is an attracting method in probing folding of bio-macromolecules:

- As mentioned above, folding time scales can be very broad in a molecule at different length scales. This fact has been appreciated for quite some time. What is recently discovered and is less appreciated is that folding can involve with multiple intermediate states or multiple active conformations with lifetimes that are different by several orders of magnitude. As a perfect example, single molecule FRET studies using total internal reflection microscope on Tetrahymena group I ribozyme found that the folding equilibrium constant of this molecule can be different by a factor of 300, suggesting a rugged folding energy landscape that traps the molecule at local energy minimum conformations[SGCH10]. For a good single molecule method to probe folding, it is necessary for this method to have enough temporal resolution and range to measure important features of the dynamics involved.

- It is difficult to have good spatial resolutions in detecting these folding motions. And without refined resolution, it is impossible to distinguish different conformational states of the molecule. Let us use FRET as example again. The physics of the interactions between the donor and acceptor dictates that the efficiency of the transfer, which is a function of distance between the fluorophores, goes down as $[1 + (d/R_0)^6]^{-1}$, where d is the separation between the donor and acceptor, and R_0 is called Forster radius[RHH08]. FRET is most sensitive to probe distance changes around the Forster Radius values, since the efficiency plateaus at both ends. A survey of the typical FRET pairs reveals

that FRET is effective in monitoring distance changes from 3 to 10 nm. This limits the spatial resolution of this technique, and in order to learn details of the folding, it is necessary to pair FRET with other mechanisms, such as quenching or several FRET pairs in the experiment[TM08].

- Biomolecules are small and fragile. To detect the folding of these molecules, we will need to have labeling and probing techniques not to be intrusive. The necessary labeling procedures in fluorescence microscopy have shown that it can interfere with the folding of molecules and render unexpected effects to the molecules[BMA+10]. Sometimes carefully selected linker chemistry is important in minimizing the interaction between fluorophores and the molecule under investigation.

- Another difficulty is that proteins and nucleic acids are macromolecules that live in aqueous solutions, and are undergoing constant random Brownian motion. These stochastic motions prevent us observing one molecule for a long periods of time (typically $< 1\ mS$), and thus limit our abilities in detection novel features such as heterogeneities in a sample.

Maybe it is a bit cliche, but probing folding of macromolecules is like recording a movie of a beautiful ballet dancer performing on the stage under the spot light. The movie needs to record both slow and fast motions, tiny movements like facial expression to large movement such as grand battement, while the spot light follows the dancer as she moving around the stage. All these should be done without interfering the dancer. The beauty of the dance would be destroyed if one pushes the dancer around or tag a basketball to her.

The current thesis is to study the folding of an important secondary structure in large nucleic acids called hairpin, with all the above considerations taking into account. The folding of hairpins are important to learn because it has many functions such as modulating the gene expression and stabilize the tertiary structures of RNAs[Wad00, SC06]. Importantly, we also aim to develop an experimental method called tracking-FRET: while the molecule is undergoing its random Brownian motion, we follow the molecule and at the same time monitor the FRET signals (sometimes the

donor and acceptor fluorescence are combined to track the molecule). This probing technique has a temporal span from nanoseconds to tens of second, and is not intrusive. Since this is a new technique and there are important and difficult details that need to be figured out, we first used a complimentary and more established method, fluorescence correlation spectroscopy (FCS), to study the folding of the hairpin. The comprehensive results on hairpins serve as reference for our later investigation of tracking-FRET as an method to investigate folding reactions.

1.2 History of Experiments

I started in Hideo's lab around Fall 2007. The first project is to build a single molecule con-focal microscope that has dual channels to examine colocalization of quantum dots covered with two complimentary sequences of DNAs. I built all the apparatus, software, and analysis tools of the microscope and successfully observed colocalization. Later, the association dynamics of these samples are also measured by FCS in the solution. This period I was greatly benefited from the guide and help of Dr. Kevin McHale. Basic alignment techniques and tracking stuff are all due to his credit.

Around 2009 I started working on the hairpin project with collaboration from Max Greenfeld and Sergey Solomatin from the Herschlag lab. I learned from Max and Sergey the basic labeling and purification procedures on nucleic acids. After the initial disappointment of tracking results, I started using FCS to probe the hairpin folding dynamics by using FRET pair Atto425 and Atto532 and attaching the hairpin to a Qdot. It was a very disappointing period: the FRET pair seemed blinking due to the photophysics of Atto425; Qdots polluted the spectrums of the dyes, and it seemed the charges on the Qdot interacted with the negatively charged DNA backbone. We could not distinguish control from sample molecules. No folding dynamics was observed.

After this disappointing results, we decided to use a more established FRET pair, Cy3-Cy5 to monitor the folding. Dye-quencher pair was also introduced (Prof. Chandra was responsible for the dye-quencher experiment before he left about half year later). In terms of experimental techniques, we decided that small steps first

need to be made before we used tracking microscope to examine the folding of the hairpin. After successfully observing some folding dynamics by FCS, we started using bead-hairpin complex in the tracking microscope. Basic experimental techniques and analysis methods were developed during this period. However, the results from FCS and tracking did not match, and results from dye-quencher and FRET did not match. We did not really know why. I spent a large amount of time doing analysis, comparing different strategies in extracting the folding dynamics.

After encouragement from my advisers, I started a new round of experiments by introducing a range of new sequences that could dissect the hairpin folding. I decided to focus on using FCS as the main technique, since both measurement and analysis are easier. The giant beam FCS technique was developed during this period. We found that FCS results of the dye-quencher sample were completely different from last round of results. The new results not only was repeatable but also made intuitive sense - we realized that previous samples perhaps were mishandled. Abasic and FRET samples are also probed. The results from these samples together explained a cohesive of picture of how hairpin folds. New tracking experiments were also done with the success of these experiments. Anti-correlation were observed with FRET samples by five base pair stem hairpins. It is funny to see that the experiments are presented backward in the thesis.

1.3 Organization of the Thesis

This thesis is organized as the follows:

Chapter 2 introduce the the hair structure and its biological functions. We will see that hairpin, although just a secondary structure, has a number of important of functions in the cell. We will then review past studies that investigated the hairpin folding. The purpose of this review is two folds. One is to appreciate the complexity of hairpin folding and summarize all the previously proposed states involved. This will pave the foundation for us to construct the General Folding model for hairpin, which takes into account of all previously proposed state. The second purpose is that by this review, we can learn the pros and cons of many single molecule techniques.

We will see why we are so excited about tracking and its ability in probe folding reactions.

In Chapter 3, we present a comprehensive study of the DNA hairpin folding using FCS. The purpose is to learn as much as we can about the DNA hairpin folding process. We first present enough theory, materials, and experimental setup for the reader to understand later experiments. After these introduction, we will first examine two sources of fluorescence signal variations other than the folding: dye's intrinsic photophysics and hairpin's center-of-mass diffusion. We found our dyes are okay and the hairpin's folding and diffusion is coupled. Next, we present a number of studies dissecting the hairpin folding. First we directly investigated the folding of the three stem base pair hairpin, finding that a simple two-state model can be a valid model in describing the hairpin folding. Then we engineered a number of sequences to probe the significance of stem-loop interactions, finding that these interactions do not play significant role in the hairpin folding process. Next, by taking advantage of the long distance interaction (as long as the contour length of the DNA hairpin chain) of the FRET process, we found that the semi-folded state is not important either in the folding process (represent less than 1% of the whole population). Lastly, by pinning the hairpin on the surface and direct calculations on the solution data, we prove that we have probed the whole hairpin folding.

In Chapter 4, we present a study of the DNA hairpin folding using our tracking apparatus. We will begin with an introduction to the tracking experiment, such as the position sensing technique, and a brief introduction to the setup. We next present some basic theory of the tracking apparatus. Specifically, we will look into the theory related to tracking FRET and how to extract the folding dynamics out of the overall statistics of the fluorescence signal. Effects such as multiple FRET pairs on one tracking construct, imperfect labeling, and crosstalk are also discussed. Then we will present two sets of tracking experiments: one uses dye-quenching system, and one uses the FRET Cy3-Cy5 pair. We will present tracking session examples, and show the dramatic difference between control and sample sequences. We discuss how to extract the folding dynamics. Finally, we will conclude by discussing the discrepancies between our solution and tracking data, elucidating the power of tracking yet the

necessity of further technical development.

In Chapter 5, I will conclude the thesis with discussions on the work left for both DNA hairpin folding studies and tracking as a technique to probe molecular folding at single molecular level. The road is long, but the future is bright. Keep pushing guys!

2

Background and Context

In this chapter, I will provide the necessary background information about the nucleic acid hairpin structure (hairpin) and the context of our kinetic measurements on its folding. First, I first introduce the hairpin structure and its biological functions. Our fundamental understanding of the hairpin structure will later guide us in designing experiments to dissect its folding. We will see that the hairpin, while small, plays critical roles in many important processes such as serving as binding motifs for DNA binding proteins and acts as the terminators for RNA transcription. Next we will provide context for our kinetic measurements by providing a review of the methods and results of experimental studies on the hairpin folding. Many experimental methods have been applied to the study of hairpin folding, such as Fluorescence Correlation Spectroscopy(FCS) [BKL98] and optical tweezers[WBPL+06]. Many folding models have been proposed, but the literature overall displayed inconsistency: while some proposed states are pivotal in the interpretation of the result of one study, it is completely missing from another. By reviewing this literature, 1) it would be clear to see what questions are still not answered despite 30 years of research on this topic; 2) we will formulate a attack plan on our studies of the hairpin folding; 3) we will see the advantages and disadvantages of various experimental methods and allure to why tracking-FRET is uniquely positioned to investigate molecular folding at single-molecule level.

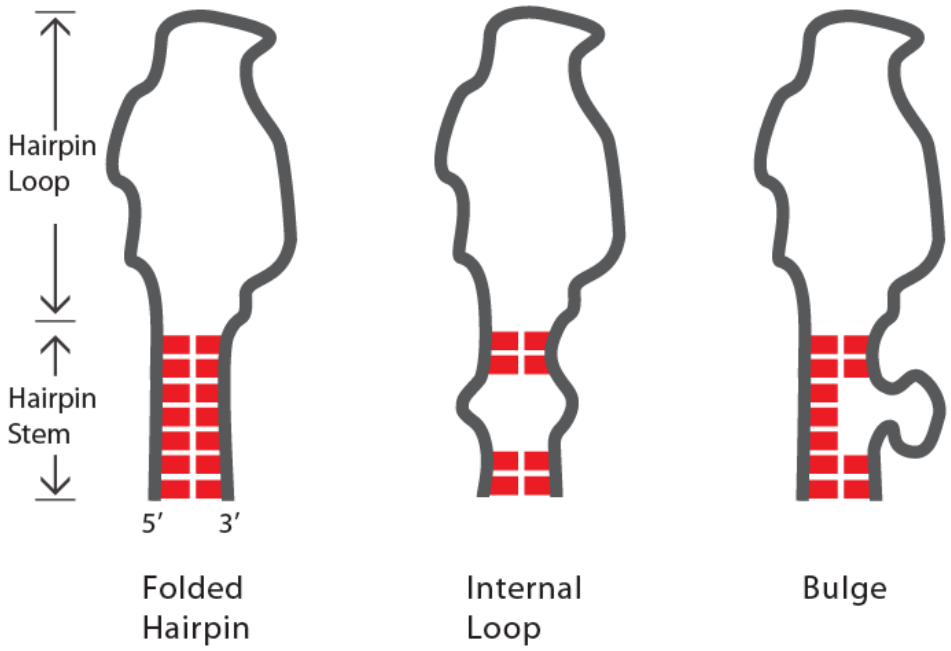

Figure 2.1: The Structure of Nucleic Acid Hairpin. All the non-paired nucleotides are not drawn. This figure is adopted from Bevilacqua et al[BB08]. The fully folded hairpin has all the base pairs in the stem formed; the internal loop and bulge are structure defects of the stem and can be formed when there are mismatches in the stem. These structure diversity contributes to the functional diversity of the hairpins [BB08].

2.1 Background

2.1.1 Structures of Hairpins

The nucleic acid hairpin is formed when a single stranded nucleic acid chain folds back and forms base pairs with itself. Structurally, the folded hairpin has two parts: a double stranded stem and a single stranded loop (Fig.2.1). The stem primarily consists Watson-Crick basepairs formed by complimentary bases. The unfolding of the hairpin structure is directly related to the stability of the stem, which can be calculated by the nearest neighbor rules. This thermodynamical calculation of the stem stability takes into consideration of the direct base pairing as well as the base stacking interactions among the neighboring bases, and it has been proven to be quite

reliable[SH04]. The stem can also have structure defects such as internal loops and bulges[BB08](See Fig.2.1). These structure defects, while destabilizing the hairpin structure, offers unpaired nucleotides and widen the major groove of the stem. Both effects open possibilities for interactions with other structure elements in a large nucleic acid molecule and with proteins.

The loop of the hairpin is single stranded and needs to change its direction for the hairpin structure to form. The flexibility and length of the loop all contribute to the folding rate of the hairpin. Hairpins at high salt in general have higher folding rates, which is attributed to shortened persistence length of the the chain[MRC+04] due to the electrostatic screen of the cations in the solution. It also has been found that poly-A loop folds about three times slower than poly-T loop, because the bases-tacking interactions among the poly-A loop render additional local rigidity[GBKL00]. Certain loop sequences, especially tetranucleotide loop as UNCG, occurs frequently in ribosomal and other RNAs and is exceptionally stable[Var95]. In a 16S RNA, about 70% of the tetraloops are -UNCG- and -GNRA-. Hairpins with these loops have unusual high thermodynamical stability, and can function as nucleation site for RNA folding and protein binding reactions.

It is worth noting that there can be more interactions than the conventional Watson-Crick basepairing in a hairpin. Since all the bases can be hydrogen bond donors and acceptors, non-Watson-Crick base pairing can occur. This is especially true in the case of hairpins, in which the overall structure is not stable and the resulting frequent contacts among the bases can promote non-cannonical basepairing, such as G-T wobble basepairing. In addition to the interaction among the bases, it is possible to have sugar-base and phosphate-base interactions. Although these interaction tends to create less stable conformational states than the fully folded hairpin, they potentially can trap the hairpin into misfolded states and prevent the correct native folding of the hairpin[AKS01, LMSZ08]. This greatly increases the complexity of the folding reaction, and past research has identified some evidence of its existence.

2.1.2 Hairpin has Primary Importance in Biology

> *"The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid."*
>
> – Francis Crick 1970.

The central dogma as stated above is the keystone of modern molecular biology, and it describes the information flow in a biological system. In essence, the genetic information encoded in DNA is transcribed to RNA, which in turn be used in the making of proteins. In addition to this main information flow (DNA to RNA to proteins), there are other important processes involved, such as DNA replication and the gene regulatory processes. The nucleic acid hairpin, although a secondary structure, has primary importance in almost all the processes mentioned above.

Here we list some of the biological functions of the hairpins, categorized by RNA and DNA hairpins. RNA hairpin has many important biological functions[SC06, BB08]. It facilitates the folding of complex molecules such as the ribozyme, terminates the transcription, and plays essential role in translation[HL88]. RNA hairpin is the predominant secondary structure in complex bigger molecules such as a ribozyme. This can be well illustrated by the secondary structure of the 16S rRNA for *E. coli*, a critical component of the ribosome (Fig.2.2 B), second panel.). Out of the 1,541 nucleotides of the 16S rRNA, about 70% folds into 31 hairpin structures[Woe83]. In the presence of divalent metal ions (e.g., Mg^{2+}), these secondary structures interact and form tertiary structures, which is crucial for the molecule's biological functions. The tRNA, part of the molecular machineries in translation, is partially composed by three hairpin structures. The anticodon, which decodes the genetic information on the mRNA molecule, is on the loop of one of the three hairpin structures on the tRNA. The hairpin structure is also responsible for the termination of transcription during the intrinsic termination process in some prokaryotes (Fig.2.2 B, the third panel.) The stability of these hairpin structures causes the polymerase falls out and ends the transcription.

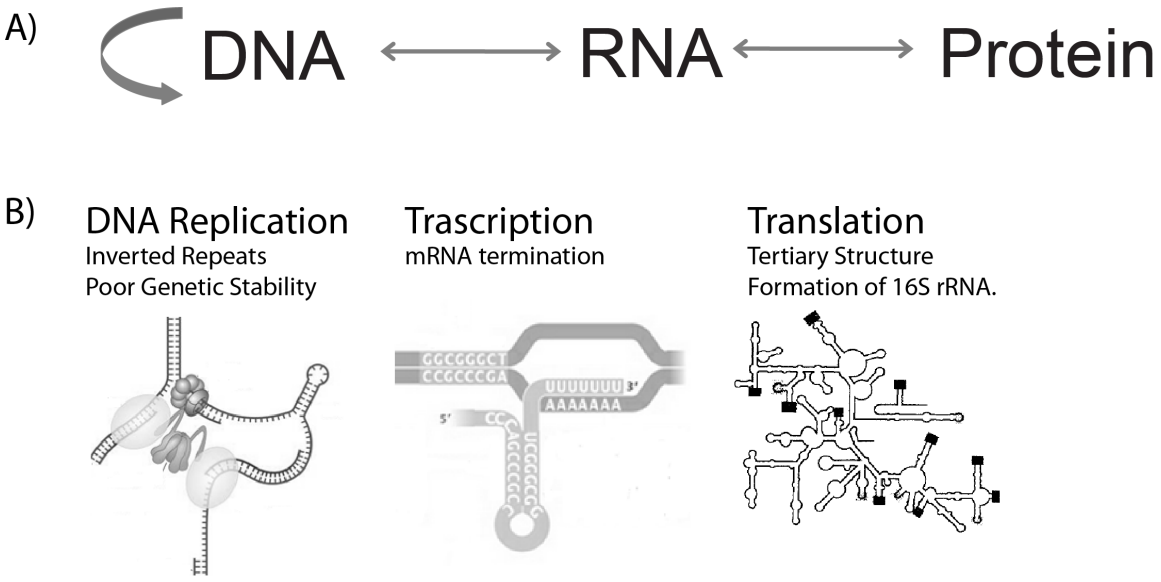

Figure 2.2: Hairpin's biological importance. A) The central dogma of molecular biology [Cri70]. B) Examples of how hairpin participate in various important processes of life. From left to right: first panel illustrates the formation of DNA hairpin during the replication process, which leads to deletion in many species[BLBM10]; second panel illustrates the hairpin as the terminator for mRNA synthesis[WvH95] ; the third panel illustrate the prevalence of hairpin structure in a ribosomal RNA, the 16S subunit in E. Coli[Woe83, Var95].

While RNA hairpin's functions have been long appreciated, DNA hairpin's biological functions are less appreciated[BLBM10]. This certainly have to do with the fact that RNA in most instances are single stranded and therefore secondary structures like the stem-loop are more prevalent. DNA, on the other hand, mostly presents itself in the double-stranded form in the cell. However, cellular processes such as DNA replication and transcription will locally open the double stranded chain, which allows secondary structures such as the hairpins to form[Wad00]. For example, during the replication process, the antiparallel structure of the DNA requires that one stranded must be synthesized in fragments, which leaves the DNA as single stranded. In E.Coli, the length of the ssDNA on the lagging strand template can be as long as 1,000 to 2,000 ($http://en.wikipedia.org/wiki/Okazaki\_fragments$) nucleotides. If there are inverted repeats on the chain, then hairpin structures can be formed. For many species, these nucleotides in the hairpin structure will be skipped by the DNA polymerase and not copied. This explains why long inverted-repeats have poor genetic stability[BLBM10]. On the other hand, the potential to form hairpin like structures, such as the cruciform, can modulate the super coiling of the DNA. Since the super-coiling of DNA is important to many DNA interaction protein involved in gene regulation, such as promoters, DNA hairpin modulates gene expression. Finally, DNA hairpins serve as binding motifs for many DNA interaction proteins and are the essential structures for molecular beacons, which are used in many applications involving nucleic acid detection and quantification[TBRB03].

In essence, hairpins are fundamental structures and their dynamics are integral and basic to many biological processes as mentioned above. The biological functions of the hairpin depend on which conformational state the hairpin is in, thus a detailed understanding of the folding dynamics could facilitate our understanding of its biological functions.

2.2 Context of Our Experiments

2.2.1 DNA Hairpin Folding Literature Review

In this section, I will give a mini review of past studies on DNA hairpin folding, which serves as the context for our experiments. For a list of papers, sequences investigated, experimental methods used, and some key insights such as the folding model proposed, please refer to Table.2.1. Hairpin folding has been a subject of intense research over the past three decays. Many experimental techniques, such as FRET[WYBK01], quenching, fluorescence correlation spectroscopy, temperature jump[MPK+06, MWWZ07], optical tweezers[WBPL+06, WGGB08], have been applied to measurement of its folding kinetics. Theoretical calculations and simulations are also performed[ZC06, ZC02, SRNP03, SRP05]. These studies focused on measuring the time scales of folding, how chain composition and experimental conditions can affect the folding[KSBA01, KRWA08]; from these experimental observations, these studies also emphasized on learning thermal dynamics of folding and on elucidating folding models that can capture the essential features of experimental observations[SH04]. We will first give an overview of such past experimental studies. While covering basics findings, this mini-review gives special attention to evidences that could elucidate a hairpin folding model: that is how the hairpin transit from a flexible coil to a fully folded hairpin structure with all the base pairs on the stem formed. At the end of this review, it is clear that despite many intriguing findings in the past studies, there is still much to learn and to clarify because of the unexpected complexity of hairpin folding.

FCS has been a popular tool in studying the DNA hairpin folding[BKL98, WYBK00, WYBK01, KDNS06]. The first hairpin FCS study [BKL98] investigated the folding of five base-pair stem hairpins (CCCAA-T12,16,21,30-TTGGG) in solutions containing 200 mM NaCl. It was concluded in the study that the hairpin folding process could be described by a Markovian two-state process, in which the hairpin conformation transits between a flexible coil and a fully-folded hairpin structure. It is a very clean folding pathway. The unfolding time scale for the 5 base-pair stem hairpin is in the order of 10 uS, and it is a function of the stem's properties (i.e, length and sequence)

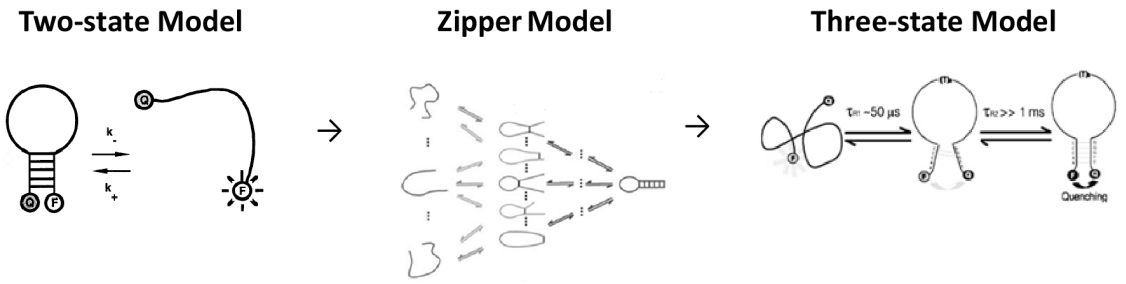

Figure 2.3: Hairpin folding models as proposed by the literature. From left to right: a two state model in which the hairpin transits from fully folded to fully open state (this figure is adopt from [BKL98]); zipper model, in which misfolded states can trap the hairpin before the hairpin zipping to form the fully folded state (this figure is adopt from [AKS01]); three-state model, in which a semifolded state is first formed before the hairpin can fold into the full hairpin structure(this figure is adopt from [JVO06]).

but not those of the loop. It is evident from this study that FCS can be one effective technique to measure hairpin folding, but we need to be careful with our interpretation of the results as the temporal range of the FCS is limited by how much time the free diffusing molecules stay in the laser focus. Indeed, later experiments using advanced probing techniques found much longer-lived states using almost the same sequences.

The second wave of hairpin experiments used laser induced temperature-jump spectroscopy (T-jump) as their main technique. Ansari and coworkers conducted T-jump experiments on a hairpin with five base-pair in the stem (5'-GGATAA-T4-TTATCC-3') in solutions containing 100 mM NaCl[AKS01, AK05, KRWA08]. The Markovian two-state model failed to describe critical features of the kinetic data in the T-jump experiment. A new folding model named the Configuration Diffusion Model was proposed. This new model states that before the hairpin could correctly fold to the native hairpin structure, transient misfolded states were first formed through stem-loop interactions by canonical and/or non-canonical base pairing. The hairpin needs to explore all the conformational space before folds into the correct native structure. Ansari's results also revealed that the folding of this five base-pair hairpin is fast - the measured folding and unfolding time constants are in the order of 10 μS,

consistent with the results from [BKL98].

The very recent hairpin studies were fueled by new experimental techniques that could probe slow dynamics (slower than $1mS$). These new experimental techniques include: optical trapping[GWAB05, WBPL+06, WGGB08], surface attachment single molecule FRET[GGL+01] and dual-beam FCS[JVO06, JIS+08]. Block and his colleagues applied optical trapping to measure the folding dynamics of relatively long hairpin sequences (stem length ranging from 6 to 30 base-pairs)[WBPL+06, WGGB08]. It is a surprise that a two-state folding behavior was observed on these long sequences, because 1) almost every hairpin studies after [BKL98] found that a two-state folding model is not sufficient in describing the folding reaction, 2) one would expect that there would be more points along the long chain to form hydrogen bonds and thus there would be higher chance to trap the chain to misfolded states in longer chains, and therefore a two-state folding model is especially unlikely being the correct model for longer chains. Another surprising finding in this study is that the hairpin unfolding time constants measured by Block et al is different from that of Ansari et al by four orders of magnitude despite the fact that they had almost the same hairpin sequences[OJ08].

These differences in the literature might be attributed to different dynamical range of the measurement methods, and the fact that hairpin folding is much more complex than we expected[OJ08]. T-jump experiments can only measure folding dynamics faster than the typical relaxation time, which is in the order of 10 μS; whereas optical trapping, because of its mechanical nature, can only measure dynamics slower than 1 mS. It is possible that the hairpin folding is so complex, involving many different folding states with vastly different stabilities, that these studies measured only different parts of the reaction. Recognizing the need for techniques that have broad temporal range to study the hairpin folding, Van Orden and his colleagues applied a technique named dual-beam FCS to measure both fast (in the order of μS) and slow (in the order of mS) processes in hairpin folding. They investigated the same five base-pair stem hairpin studied in [BKL98]. Their data and subsequent control experiments found that both fast (50 μS) and slow ($> 1\ mS$) folding processes exist on the same molecule. They proposed a three-state folding model: before the

chain could fold to the fully-folded hairpin, a partial-folded stem is first formed. This was the first time that enough care had been given to make sure the entire folding reaction was probed, and a much more complicated folding pathway was found.

2.2.2 Much Still Needs to be Done in the Request of Understanding DNA Hairpin Folding

Although the past literature has shed light in how hairpin folds and provided both experimental and theoretical fundamentals of how to investigate hairpin folding, there are many aspects of the folding are not clear nor investigated thoroughly. First of all, it is not clear what conformational states participated in the folding pathway and how significant each state is. According to the literature, at least four groups of conformations may participate in the folding. They are:

a) the open state, in which the chain is a flexible random coil,
b) the misfolded state, which traps the hairpin from forming the native hairpin structure,
c) the semi-folded state, which is a group of conformations formed by partial base-pairing of the stem
d) the fully folded state, in which all the base pairs on the stem are formed.

It is not clear whether these states really exist or participated in the folding process, because there is apparent discrepancies in the literature: while the misfolded state is pivotal in one study[AKS01], it is completely missing in another[JVO06], and vise versa. Whether there is a general folding model for hairpins with different sequence is another question, but it is apparent that for almost the same sequence, two different study would report fundamental folding parameters that are different by four orders of magnitude. As stated before, this perhaps has to do with the fact that past studies might not have the advanced technology required to probe hairpin folding. As a consequence, depending on what conformation state each experimental technique is best to find, we have seen different folding models at different studies.

Often these different models are not inclusive or in fact sometimes in direct conflict with one another.

Second, how cations affect the DNA melting process has long been appreciated, and there has been some theoretical development[TC08b, TC08a]. However, not many experiments has been focused in learning the cation effects on the context of DNA hairpin folding kinetics. Since in hairpin folding, one is dealing with both ssDNA and dsDNA, cations can modulate the DNA hairpin folding process in three very different aspects, depending on whether the molecule is single stranded or double stranded. First, cations can modulate the flexibility and electrostatic interaction of the ssDNA chain. Thus it can change the global conformation and modulate the end-to-end collision rate. Second, the same screening effects of the cations could allow the ends of the chain approaching together and have higher probability of forming the stem. Third, once the stem is formed, cations help to stabilize double-stranded DNA, which is another physical process. So far, there has been no DNA hairpin folding studies that investigated thoroughly how cations modulate various pathways of the folding reaction. Finally, a critical process involved in the hairpin folding process is the end-to-end collision of the chain. There are some experimental studies of this ssDNA chain conformational dynamics[MP09, WN03, MRC+04, QYL+10], but we argue that since this process is so integral to the hairpin folding (it is the pre-cursor for the chain to form hairpins), we need to design experiments that directly study this intramolecular process together with the hairpin folding (stem formation/detachment) under the exact same experimental methods and conditions[WF74]. So far there has not been any study that looked at this intramolecular process together in the context of the hairpin folding.

Ref	Sequence	Method	State(s) lifetime	Model
[KDNS06]	$C-T^3-G$	FCS	2˜5μS	disorder
[KDNS06]	$C-T^4-G$	FCS	2 μS	disorder
[KDNS06]	$CC-T^3-GG$	FCS	3 μS	disorder
[KDNS06]	$CC-T^4-GG$	FCS	2 μS	disorder
[JVO06]	$AACC-T^{21}-GGTT$	FCS	84 & 393 μS	three-state
[BKL98]	$CCCAA-T^{12,16,21,30}-TTGGG$	FCS	10 μS	two-state
[GBKL00]	$CCCAA-T^{8)arrow30}TTGGG$	FCS	˜10 μS	disorder
[GBKL00]	$CCCAA-T^{8)arrow30}TTGGG$	FCS	˜100 μS	disorder
[WYBK00]	$CCCAA-A^{30}-TTGGG$	FCS	0.7 mS and 0.5 mS	disorder
[JIS+08]	$AACCC-T^{21}-GGGTT$	two beam FCS	50 μS & > 1mS	three-state
[WBPL+06]	$GAGCTA-T^4-TAGCTC$	Optical Trapping	100 mS	two-state
[AKS01]	$GGATAA-T^4-TTATCC$	T-jump	10 μS	zipper
[GGL+01]	$CTCTTCA-A^{13}-TGAAGAG$	Surface TIR	100 mS	two-state
[AK05]	$CGGATAA-T^8-TTATCCG$	T-jump	100 μS	zipper
[WBPL+06]	$GAGTCCTA-T^4-TAGGACTC$	Optical Trapping	1 S	two-state
[GGL+01]	$CTCTTCAGT-A^{13}-ACTGAAGAG$	Surface. TIR	125 mS	two-state

Table 2.1: Experimental hairpin literature results. Rows are grouped by the number of bp in the stem region.

3

Dissect DNA Hairpin Folding by FCS

3.1 Abstract

Our understanding of hairpin folding could be advanced from a comprehensive investigation that rigorously examines the existence of previously proposed states (such as misfoldeds), their stability, and how they contributed to the folding pathway. In this chapter of the *thesis*, using fluorescence correlation spectroscopy (FCS), we conducted kinetic studies on a three-basepair (bp) stem DNA hairpin with over ten control sequences in a wide range of sodium chloride (NaCl) concentrations. To setup a frame work for our investigation, we first included the conformational states previously proposed and constructed a very General Folding Model (GFM) for the hairpin, see Fig.3.1. The GFM effectively allowed all logical transitions among the states. Our plan is to examine each folding branch of the GFM one at a time. For example, we examined the misfolded hairpin to open hairpin by a sequence that promotes the stem-loop interaction - the loop is a poly-thymine chain, but the stem is replaced by an adenine. The native base-pairing between thymine and adenine promoted the stem-loop interaction and this sequence allowed us to study this process in the absence of other component processes such as the stem formation/detachment.
We hope to answer these questions by our study:

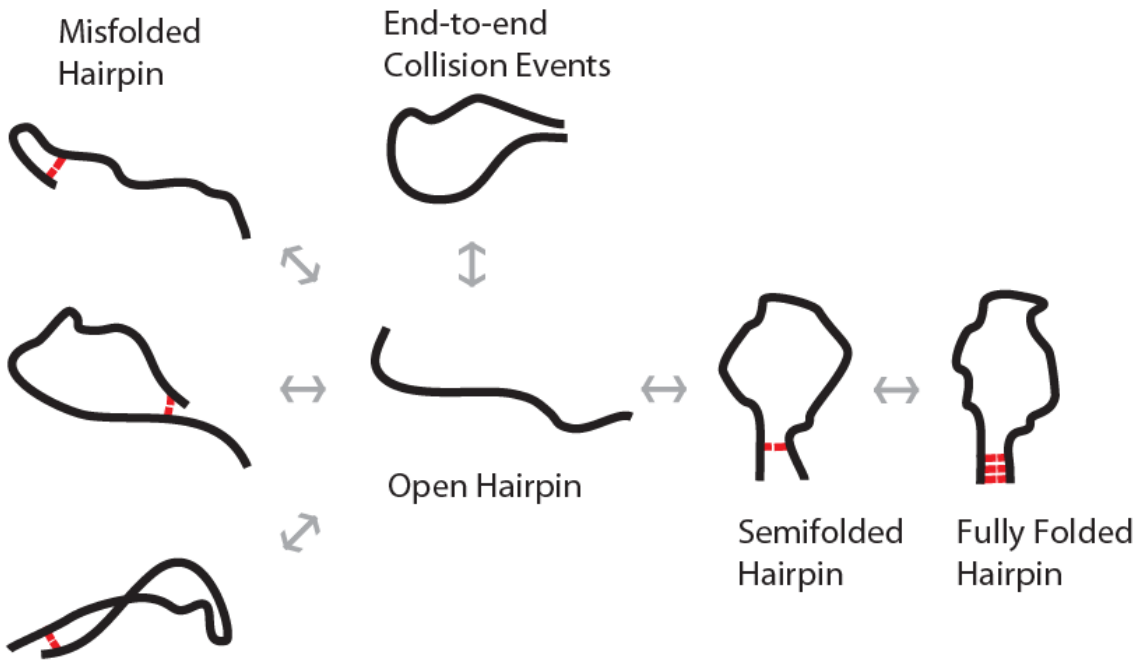

Figure 3.1: General Folding Model of hairpin folding. This model incorporates previously proposed states involved in the hairpin folding, such as the misfolded hairpin and the semifolded hairpin. The end-to-end collision event is also considered in this model because it can present itself in the measurement signals (fluorescence signals) in the experiments. We designed sequences to isolate and examine branches. This advances our understanding of the hairpin folding, because it determines the existence of previously proposed states, their stability, and how they contributed to the folding pathway.

- What conformational states play critical roles in hairpin folding?
- What are the time constants for these critical states?
- How do cations affect these processes?
- Can we conclude a model that explains how hairpin folds and unfolds?
- By this comprehensive study, can we learn what caused discrepancies in the literature?

3.2 Theory of FCS in Probing Molecular Folding Reactions

In this section, I will present basic FCS theories for the study of hairpin folding. FCS has been shown as a powerful tool in monitoring diffusion, folding, and chemical reactions [KB02]. There are many theoretical development in the context of folding (particularly in the case of FRET), but I am just not impressed with the current state of the literature on this front[WYBK00, TL07, Lev10]. Here are some of my own calculations.

3.2.1 Markov Model as a General Description of the Hairpin Folding.

General Case

Despite of its popular usage[BKL98, KB02, WGGB08] in modeling DNA hairpin folding reaction, Markov Model of the hairpin folding is a dramatically simplified model for the hairpin folding reaction, assuming that we can approximate the continuous folding reaction to a series of "states" of the hairpin. In the Markov Model, there is a transition matrix governing the probabilities of transition among the states, and these probabilities only do not depend on the system's histories. If we let Q be the transition rate matrix, the probabilities for the hairpin to jump from initial state i to state j after a time delay τ is:

$$Tran(\tau) = e^{Q\tau}$$

If we continuously monitoring the fluorescence of a single hairpin construct (without any other source of fluorescence intensity modulation such as diffusion or tracking) , the autocorrelation function of its intensity is given by:

$$< B(t)B(t+\tau) > \quad = \quad \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \cdot Trans_{ij}(\tau) \cdot QB_i \cdot QB_j$$

where π_i is the average fraction of time the molecule is at state i, $Trans_{ij}(\tau)$ is the

probability for the molecule to jump from state i to state j after waiting τ, QB_i is the dye intensity at state i, and N is the total number of states in the model. Therefore the FCS for Markov Model in general can be written as:

$$\begin{aligned} g_{HMM}(\tau) &= \frac{< B(t)B(t+\tau) >}{< B(t) >^2} - 1 \\ &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \cdot Trans_{ij}(\tau) \cdot QB_i \cdot QB_j}{(\sum_{i=1}^{K} \pi_i QB_i)^2} - 1 \end{aligned}$$

The explicit expression for the above results is complicated for models with more than two states, but it can be easily calculated using computation software such as Matlab.

It should also be noted that the number of free parameters defines the FCS curves for a HMM model scales as $N^2 - 1$, where N is the number of states. Once the model has more than three parameters, since it has so many degree of freedom, it can almost fit to any experimental curves.

Two-State Case

Let's consider a two-state system with a fluorescently low state and a high state. Define transition rate matrix and equilibrium levels as:

$$Q = \begin{pmatrix} -k_{lh} & k_{lh} \\ k_{hl} & -k_{hl} \end{pmatrix}$$

$$(\pi_l, \pi_h) = (\frac{k_{hl}}{k_{lh} + k_{hl}}, \frac{k_{lh}}{k_{lh} + k_{hl}})$$

where k_{lh} is the transition rate of the jumps from low to high state. It is trivial

to calculate $Trans$:

$$\begin{aligned} Trans(\tau) =& e^{Q\tau} \\ =& \begin{pmatrix} \pi_l + \pi_h e^{-k\tau} & \pi_h(1-e^{-k\tau}) \\ \pi_l(1-e^{-k\tau}) & \pi_h + \pi_l e^{-k\tau} \end{pmatrix} \end{aligned} \tag{3.1}$$

where

$$k = k_{lh} + k_{hl}$$

Therefore FCS for two-state Markov Model is

$$g_{HMM-2DD}(\tau) = \frac{K(1-Q)^2}{(1+KQ)^2} Exp[-k\tau] \tag{3.2}$$

$$g_{HMM-2AD}(\tau) = -\frac{K(Q_D-1)(Q_A-1)}{(1+KQ_D)(Q_A+K)} Exp[-k\tau] \tag{3.3}$$

in which we defined the equilibrium constants of the reaction K as $K = \frac{\pi_h}{\pi_l}$, and the ratio of fluorescence rates $Q = \frac{Q_h}{Q_l}$. Note that the acceptor and donor cross-correlation is negative, reflecting the anticorrelation nature between the two signals. It should be emphasized that the cross-correlation in theory is the best to analyze the folding dynamics, as uncorrelated noise sources (such as dye blinking) in the donor and acceptor channels are all averaged out in their anti-correlations (see measured anti-correlations in Fig.4.24).

3.2.2 Stationary FCS or sFCS to Probe Folding Dynamics

Ideal Situations

The most familiar and often used experimental conditions for FCS is perhaps the so-called stationary FCS, in which a stationary laser beam with confocal collection optics were used to probe a pico-molar to nano-molar concentration of solutions. We use the word "stationary" in order to distinguish the tracking case where our laser beams is actively moving around to follow the molecule. The following calculations

is well documented in many early papers of FCS and reviews, but we repeat the calculations here as we need to be very careful with many assumptions used in these theoretical calculations - we found many times that these assumptions were not valid in the complex environment of the experiments.

We first assume a linear dependence of the fluorescence to the laser intensity, which is only true if we operated well below the saturation regime of the dyes:

$$F(\vec{r},t) = B(t)E(\vec{r})$$

Here $B(t)$ is the brightness of the particle as seen by our detectors, which is a function of time since we assume it is modulated by the dynamics of interest. $E(\vec{r})$ is a convolution of the three dimensional Gaussian profile of a excitation laser beam and the collection volume of the downstream optics. The total fluorescence signals we observe is:

$$I(t) = \int_{\Omega} B(t)E(\vec{r})C(\vec{r},t)dV$$

$C(\vec{r},t)$ is the concentration of the molecules and is a both a function of time and space due to diffusion. Ω indicates this integral is performed over the entire solution space.

The definition of FCS function is given by:

$$\begin{aligned} g_2(\tau) =& \frac{< I(t)I(t+\tau) >}{< I(t) >^2} - 1 \\ =& \frac{< \delta I(t)\delta I(t+\tau) >}{< I(t) >^2} \end{aligned} \tag{3.4}$$

where $\delta I(t) = I(t) - < I(t) >$.

The average of $I(t)$ is given by:

$$< I(t) >= \int_{\Omega} < B(t) > E(\vec{r}) < C(\vec{r},t) > dV$$

$< C(\vec{r},t) >$ is the average concentration of the solution. The average of $< I(t) >$ can only be calculated by the above equation if the diffusive properties of the molecule and folding dynamics are not coupled. However, if the folding of the molecule involves a large conformational change of the molecule, such as in the case of the hairpin, one cannot assume there will be no coupling between the diffusive motion and folding motion. In this "ideal" situation of sFCS experiments, we will assume the decoupling between the diffusion and folding is valid.

Next,

$$\begin{aligned} < I(t) >^2 &= < B(t) >^2 < C >^2 \int\int_\Omega dV dV' E(\vec{r}) E(\vec{r'}) \\ &= < B(t) >^2 < C >^2 \frac{\pi^3 r_o^4 z_o^2}{8} \\ &= < B(t) >^2 < C >^2 \frac{V_{eff}^2}{8} \end{aligned} \tag{3.5}$$

In calculating the FCS function, we need to calculate the autocorrelation function of $\delta I(t)$ (again here we assumed the decoupling between folding dynamics and diffusive dynamics.):

$$\begin{aligned} < \delta I(t)\delta I(t+\tau) > &= < \int_\Omega B(t)E(\vec{r})\delta C(\vec{r},t)dV \cdot \int_{\Omega'} B(t+\tau)E(\vec{r'})\delta C(\vec{r'},t+\tau)dV' > \\ &= < B(t)B(t+\tau) > * DIFF(\tau) \end{aligned}$$

where

$$DIFF(\tau) = \int\int_\Omega dV dV' E(\vec{r})E(\vec{r'}) < \delta C(\vec{r},t)\delta C(\vec{r'},t+\tau) >$$

We know diffusion without interaction gives (Einstein's 1905 paper Eq.36):

$$< \delta C(\vec{r},t)\delta C(\vec{r'},t+\tau) > = < C > [4\pi D\tau]^{-\frac{3}{2}} Exp[-\frac{(\vec{r}-\vec{r'})^2}{4D\tau}]$$

Utilizing

$$\int_{-\infty}^{+\infty} Exp[-ax^2+bx]dx = Exp[-\frac{b^2}{4a}]\sqrt{\frac{\pi}{a}}$$

We have

$$\int\int_{\Omega} dV dV^{'} E(\vec{r})E(\vec{r'}) < \delta C(\vec{r},t)\delta C(r',\vec{t+\tau}) > = \frac{< C >}{8}\pi^{\frac{3}{2}} r_o^2 z_o \frac{1}{1+\frac{\tau}{\tau_D}} \frac{1}{\sqrt{1+(\frac{r_o}{z_o})^2\frac{\tau}{\tau_D}}}$$
$$= \frac{< C >}{8} V_{eff} Diff(\tau)$$

Where

$$Diff(\tau) = \frac{1}{1+\frac{\tau}{\tau_D}} \frac{1}{\sqrt{1+(\frac{r_o}{z_o})^2\frac{\tau}{\tau_D}}}$$

Here we defined τ_D as the typical time for the molecule to transit through the laser focus ($\tau_D = r_0^2/4D$), r_0 and z_0 are the waists of the gaussian profile the laser and collection space in xy and z direction respectively.

In probing molecular dynamics, it is worth nothing that τ_D is the slowest dynamics one can probe using FCS. Thus to probe potential slow dynamics in the biological system, it is important to expand the laser size as much as you can. Expanding the laser waist size, however, will require you to have more power to achieve the same intensity, which will induce more Raman Scattering and reduce the signal-to-noise ratio. Thus it is a balance to increase the laser size without averaging forever in the experiment.

Finally, we have:

$$< I(t)I(t+\tau) > = < B(t)B(t+\tau) > * DIFF(\tau)$$
$$= < B(t)B(t+\tau) > \frac{< C >}{8} V_{eff} Diff(\tau) \quad (3.6)$$

Combining Eq.3.4,3.5 and 3.6, we have:

$$\begin{aligned} g_2(\tau) =& \frac{<\delta I(t)\delta I(t+\tau)>}{<I(t)>^2} \\ =& \frac{<B(t)B(t+\tau)>}{<B(t)>^2}[\frac{1}{V_{eff}<C>}Diff(\tau)] \quad (3.7) \\ =& (DNA(\tau)+1)\cdot[\frac{1}{N}Diff(\tau)] \quad (3.8) \end{aligned}$$

$DNA(\tau)$ here represents the DNA hairpin folding dynamics we are interested in - it can be FCS functions of the donor and acceptor dye signals, or the FCS functions of the two dyes.

Eq.3.7 basically says that, in theory, the total FCS curve is a product of diffusive contribution and folding contribution - if the diffusive dynamics and folding dynamics is not coupled. A typical sFCS experiments to probe the folding dynamics will involve a sample that contains a donor dye whose intensity is quenched or transferred to a quencher/acceptor dye and is modulated by the conformational dynamics, and a control sample with the same type of donor dye but no quencher or acceptor to measure the diffusive contribution of the FCS curves. Experiments with control sample is of critical importance, since we use it to measure or eliminate other sources of intensity fluctuation such as dye blinking and sample aggregation. Finally, in the analysis step, constants will be multiplied to the two FCS curves to match the diffusive dynamics at the long time scales in order to normalize the FCS curves (assuming the $DNA(\tau)$ has decayed to zero at these time scales), which in general have different amplitudes due to their different concentrations. This is necessary as precise concentration matching is almost impossible to achieve due to pipetting errors and sample sticking. Finally, division of the normalized sample FCS to the control FCS with appropriate adding or subtracting constants yields the folding dynamics contributed FCS.

It is true that the simple connection between the total measurable FCS and folding dynamics as described by Eq.3.7 allows us to probe the folding dynamics, but it should be noted that for folding dynamics slower than the typical time of molecular

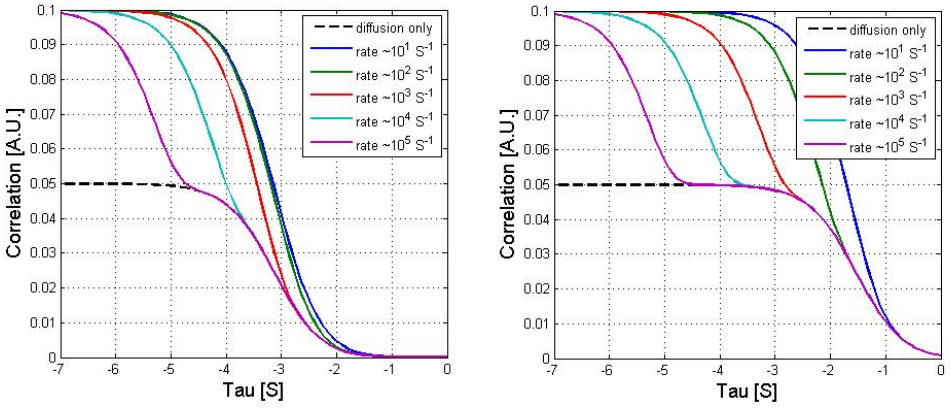

Figure 3.2: Limitations of openloop FCS to probe folding dynamics. In both panels, the dashed black curve is the diffusive contribution to the FCS curve, and the colored FCS curves are products of diffusive contribution with a two-state contribution with dye brightness ratio being infinity. The left panel has diffusion coefficient of 140 $\mu m^2/S$. With laser size being a gaussian with waist size of 0.65 μm in xy and 2.5 μm in z, the typical diffusion time is less than 1 mS. Therefore dynamics slower than 10^3 S^{-1} will not be measured. The right panel has particles with diffusion coefficient of 3.5 $\mu m^2/S$. In this case, dynamics faster than10^2 S^{-1} can be measured.

transit the folding dynamics will not be well resolved or even measured. Fig.3.2 illustrates this limitation of openloop FCS. Note the blue and green curves were not resolved at fast diffusive case. It is therefore crucial to check with other experimental evidences or literature results in order to not make mistakes in naively believing the folding dynamics obtained from openloop FCS is from the full folding reaction.

Imperfect Labeling and Crosstalk Contaminates sFCS Curves

There two other assumptions in Eq.3.7, making the equation too ideal. One assumption is that the sample is perfectly labeled - all the samples are labeled by the same dyes, and they are all functional. In reality, however, the acceptor dye can be bleached or inactive, and as a result some of the donor signal will not contain the folding dynamics. Another assumption is that there will be no crosstalk between the donor and acceptor channels. The following results are intended to take account the

imperfect labeling and crosstalking. All the calculation details has been omitted, but I will briefly discuss implications of imperfect labeling and crosstalk.

Donor channel autocorrelation function for sFCS is:

$$\begin{aligned} g_{2dd}(\tau) &= \{[\frac{1+L}{NL}][\frac{(1-e)L}{(1-e)L+1}]^2[DNA(\tau)+1] \\ &+ [\frac{1+L}{N}][\frac{1}{(1-e)L+1}]^2\}Diff(\tau) \end{aligned} \tag{3.9}$$

where $L = \bar{C}1/\bar{C}2$ ($\bar{C}1$ is the average concentration of the construct with functional acceptor/quencher), $e = 1 - B\bar{D}1/B\bar{D}2$ (eg, time averaged FRET efficiency), and N is the average total number of molecules inside the laser focus. We can vary L or e to check Eq.3.9's validity. For example, when L goes to infinity, that is when we have a pure FRET labeled sample, we simplify Eq.3.9 to the product of autocorrelation functions of diffusion and FRET dynamics. When L goes to zero, we have the pure diffusion. Thus at both limit of L, Eq.3.9 makes perfect sense. In addition to check Eq.3.9 to both limits of L, we can also fix both L and N but ask the question that how the autocorrelation function changes as a function of e, which represents the experimental conditions in our hairpin studies when salt level were increased. In the simplest case, when e is set to zero or one, we will reduce Eq.3.9 to the pure diffusion case. And when we have e somewhere between zero and one, we will have a blend of the FCS curves. In sum, Eq.3.9 makes intuitive sense.

Acceptor channel autocorrelation function for sFCS is:

$$\begin{aligned} g_{2aa} &= \frac{1+L}{NL} \cdot [1 + c(\frac{1-e}{e}) + \frac{c}{Le}]^{-2}(1-2c) \\ &\times [DNA(\tau)+1]Diff(\tau) \end{aligned} \tag{3.10}$$

c is the fraction of crosstalk from the donor channel to the acceptor channel. Note in calculating this result, we omitted terms that are in the order of c^2. The nice thing about Eq.3.10 is that it is always a product of diffusion dynamics and molecular dynamics.

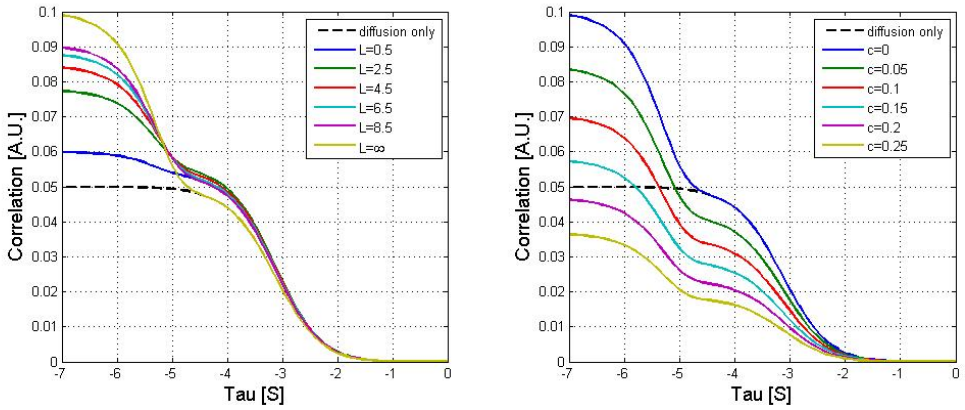

Figure 3.3: Imperfect labeling and crosstalk contaminate the measured sFCS curves. The left panel is the donor channel sFCS curves (Eq.3.9). The number of molecules that are FRET labeled are increased from half of the donor-only labeled molecules in the blue curve to no donor-only labeled molecules in the yellow curve. The right panel is the acceptor channel sFCS curves (Eq.3.10) wtih perfect labeling sample. Crosstalk levels from the donor channel is increased from the blue curve to the yellow curve. Additional parameters used in plotting the curves: $N = 20$, $e = 0.5$, $D = 140\ \mu m^2/S$, laser waist sizes: $[0.65, 0.65, 2.5]\mu m$.

Crosscorrelation between the donor and acceptor channels is:

$$g_{2ad}(\tau) \quad = \quad F1 \cdot [DNA(\tau) + 1] Diff(\tau) + F2 \cdot Diff(\tau)$$

where

$$\begin{aligned} F1 &= \frac{1+L}{NL}\{\frac{-e^2L^2 + cL^2(1-e)^2}{[L(1-e)+1][eL+cL(1-e)+c]}\} \\ F2 &= \frac{c(1+L)}{N}\{\frac{1}{[L(1-e)+1][eL+cL(1-e)+c]}\} \end{aligned}$$

Fig.3.3 illustrates the imperfect labeling and crosstalk's effects on the measured sFCS curves. Imperfect labeling adversely affects the donor and crosscorrelation sFCS by reducing the apparent amplitudes contributed by the folding reaction and

the number of molecules of in the solution. Crosstalk's adverse effects on the acceptor channel are only limited to multiplying a constant to the sFCS curves, which can be easily fixed in the data analysis step by normalizing the sFCS curves.

Background Noise Diminishes FCS Curves

Lastly, we need to consider the background from the leaking laser and Raman scattering. Here the word background is used for any signal other than the system of interest. Leaking laser background can be well controlled by appropriate shielding and filters, whereas photos from Raman scattering can only be reduced by using low Raman scattering materials in our optics. The relationship between measured FCS with the dynamics contributed FCS is given by the following equation:

$$g_{2dd}(\tau) = \frac{Dynamics\_g2(\tau)}{(1 + 1/SN)^2} \tag{3.11}$$

where SN is the signal-to-noise ratio of our fluorescence signal. Keeping the noise level low helps maintaining the contrast of FCS curves.

3.3 DNA Constructs, Dyes, Labeling and Purification

In this section, we introduce the DNA constructs, dyes we used, and labeling and purification procedures.

3.3.1 DNA Constructs

In general, our hairpin constructs are assembled by hybridizing a 26 nucleotides (nt) single stranded DNA (ssDNA) and a 47nt ssDNA together (Fig.3.4A). We will call the 26nt ssDNA as the Short Piece, and the 47nt ssDNA as the Long Piece. The Short Piece is fully complementary to one end of the Long Piece. The assembled DNA constructs have a 26 basepair (bp) dsDNA tether and a 21 nt ssDNA region at the 5' end. The sequence of the Short Piece is 5'- AGA TGC TAC CGT TCA GAG ATT ATA TT -3'. We varied the sequence of the 21nt ssDNA region to make hairpins of different sequences, which are shown in Table.3.3.1.

The abasic samples have the ssDNA part of the Long Piece is purchased from Pan Facilities at Stanford (pan.stanford.edu). All other DNA samples were purchased from Integrated DNA Technologies (Coralville, IA.).

The hybridization is typically done by mixing a few μl of $10\mu M$ samples together in 150 mM NaCl solution. It was found that 1:1 mixing ratio (concentration of the oligos was measured by NanoDrop beforehand) produced samples that had good hybridization percentage.

The Atto700-dG quencher-dye system utilizes photon induced electron transferring (PET) from the quencher to the dye. This process requires the distance between the quencher and the dye to be within the Van der Waals contact. Thus the dye-quencher system can monitor local collision event of the dye and the quencher, but will not be sensitive to other conformational changes. The rate of the PET was measured to be as high as $10^9\ S^{-1}$, fast enough for us to measure many processes involved with hairpin folding[KMS00, HKPS03].

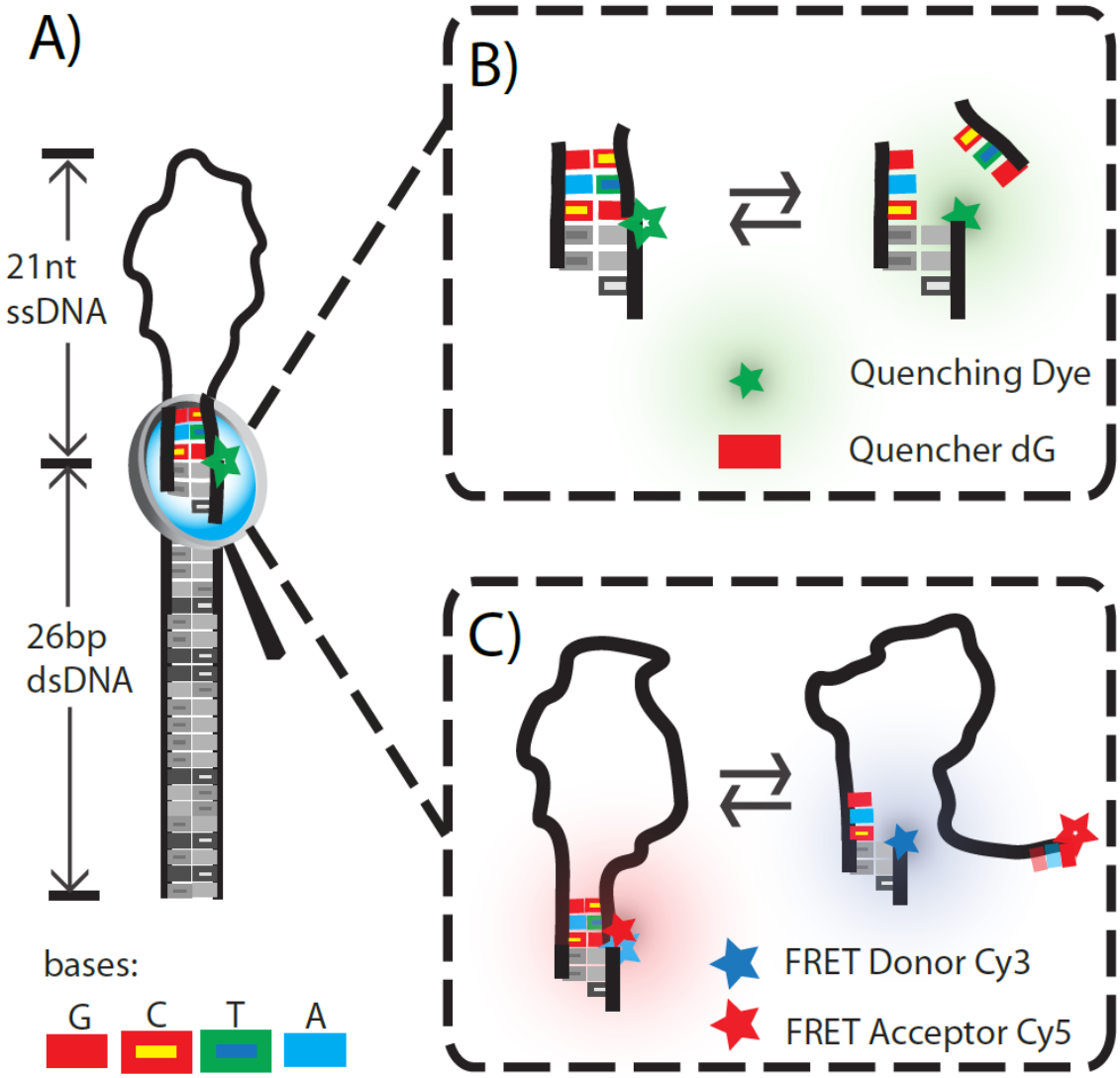

Figure 3.4: A schematic drawing of the three-basepair stem hairpin (HP3) investigated in this study with two probing dye systems. A). The hairpin construct is assembled by hybridizing a short, 26 nt, ssDNA and a long, 47 nt, ssDNA together. The last 21 nt of the long strand is the hairpin. The short ssDNA has a dye labeled at its 3' end, while the long ssDNA has either a quencher (dG) or a acceptor labeled at its 5' end. B), Atto700 with dG form a dye-quencher pair, which is sensitive to local ($< 1nm$) folding motions; C), Cy3-Cy5 forms a FRET pair, with its most sensitive range near its Forster radius (˜5.4nm).

Name	Sequence	Labeling	Purpose
HP3-noTether	3'-CAG-T^{15}-CTG-5	FRET	to test diffusion folding coupling
ALLT-noTether	3'-T^{21}-5'	FRET, Cy3	to serve as control for HP3-noTether
HP3	3'-CAG-T^{15}-CTG-5'	Atto, FRET	to learn the hairpin folding
ALLT	3'-T^{21}-5'	Atto, Cy3	to learn processes other than folding
T20G	3'-T^{20}-G-5'	Atto	to learn ssDNA end-to-end collision
GTA	3'-A-T^{19}-G-5'	Atto	to check wobble basepair formation
GTC	3'-C-T^{19}-G-5'	Atto	to examine if one basepair could hold the chain
T4C	3-$(CTTTT)^3$-G-5'	Atto	to check wobble basepairs
HP3-CTL	3'-CAG-T^{15}-TTT-5'	Atto	to check the source of quenching
AllX	3'-X^{21}-5'	Atto	abasic hairpin
X20G	3'-X^{20}-G-5'	Atto	abasic hairpin to learn abasic chain end-to-end collisions
HP3X	3'-CAG-X^{15}-CTG-5'	Atto	to check stem-loop interaction

Table 3.1: Hairpins probed in this study.

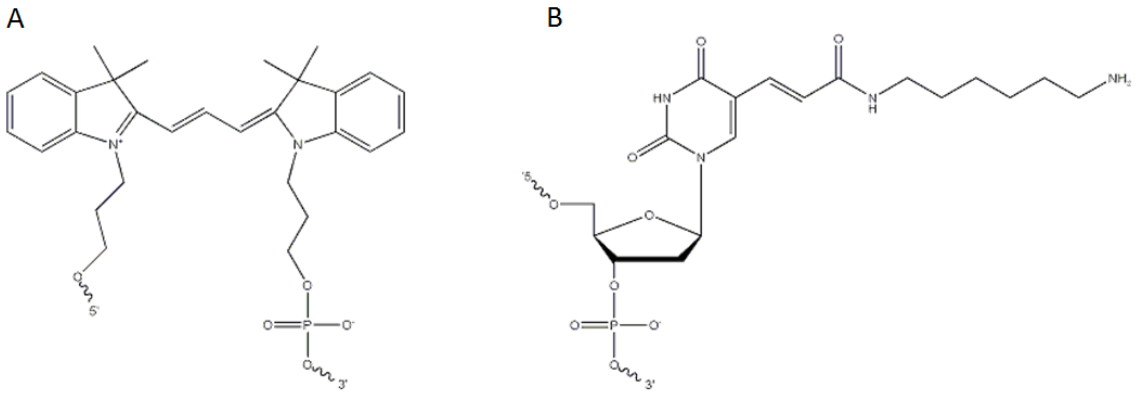

Figure 3.5: Dye linker structures for FRET and Quenching probing systems(Photo credit: www.idtdna.com). A. Linker Structure for the labeling of Cyanine dyes. The Cyanine dyes were coupled last in the synthesis process to the 5' using an amedite chemistry in the same way a standard base would be coupled during the synthesis process. B. Linker Structure for Atto700 dyes. We used C6 linker to tie the Atto700 to the last T residue on the Short Piece.

3.3.2 Dyes and Linker Chemistry

We used amino-modified DNA and NHS derivatives of dyes to label the DNA. Atto700 (Atto Tech, Germany) and Cy3 (GE Healthcare Life Sciences, USA) were labeled at the 3' end of the Short Piece. As shown in Fig.3.5, the Cy dyes was coupled last in the synthesis process to the 5' using an amedite chemistry in the same way a standard based would be coupled during synthesis. C6 linker was used to tie the Atto700 to the last T residue on the Short Piece.

It is worthwhile to notice that all our conclusions of this chapter of the thesis is based on experiments with these linkers. Even though our results are self-consistent, we need to be aware that our results have not been checked by different types of linkers[WSG+08]. This is one important future step to take because it has been shown that linker chemistry can add non-trivial dynamics of the fluorescence signal and thus the end results of a fluorescence experiment.

3.3.3 Our Practical Experience with Dyes

In this section, some practical experience with the dyes are presented. The purpose is to serve as a guide for future experiments. Dye is one of the most important components in a fluorescent experiment. We want to have dyes that can be honest reporters of the molecular folding. Thus the perfect dye would be:

- It has no blinking at all the temporal range.
- It can have infinite signal-to-noise ratio and last forever.
- It has extremely narrow spectral range (a few tens of nanometers.).
- It is insensitive to its chemical and physical environment but highly sensitive to the folding of the molecule.
- It is the same from molecule to molecule: for example, the brightness of individual molecule is a constant.

In reality, certainly we do not have any of the above. Dyes are major frustrations[RHH08]. What we typically see is:

- It usually has laser power dependent blinking, and it highly depends on the redox environment[VCF+09].
- Most dyes can only emit 10^5 photons and last less than 10 seconds at 10 kHz fluorescence rate without any chemical enhancement. The signal-to-noise ratio, depends on the exact experiment setup, the optical material used in the setup, and how hard the dye is excited. The setup we had typically gave us S/N less than 3 at reasonable laser power values (around tens of μW.)
- It is sensitive to its chemical and physical environment[RMH06].
- Dye's photophysics is different from molecule to molecule. For example, in this study, the variation of quantum yield of Cy3 and TMR are studied and found that it can affect the computation of inter-dye distance[SEM05].

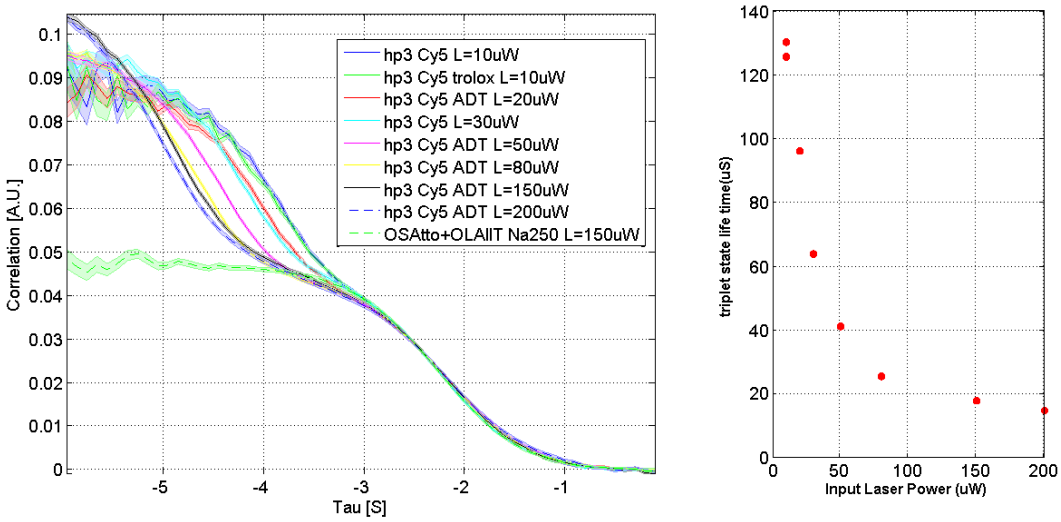

Figure 3.6: The laser power dependent blinking of Cy5. Left, the normalized FCS data. Right, the extracted triplet state life time vs laser power. The consistent relaxation at 10^{-2} seconds is from the diffusion process. The faster, well-resolved relaxation is due to Cy5 blinking, most likely from the isomerization and back-isomerization of the dye. Atto700, on the other hand, revealed no blinking at this time regime, but as we will show later, it blinks at a much slower rate than FCS could resolve.

For example, in Fig. 3.6, we show the laser power dependent blinking of Cy5. The FCS curves of the Cy5 labeled HP3 at a wide range of laser powers showed one relaxation at 10^{-2} second due to the center-of-mass diffusion, and another well-resolved relaxation that is laser power dependent blinking. The origin of this blinking is most likely due to the isomerization and back-isomerization of the Cy5. We used a widely-used blinking model, the triplet state model proposed by [WMR95] to analyze our data. It was found that as the laser power is increased, the triplet state lifetime decreased.

Cy3B (GE Healthcare Life Sciences, USA) typically can only last a couple of seconds at 20 kHz. But with the right oxygen scavenging system and the right redox conditions[AMP08], it can consistently last for more than 30 seconds and fluoresce at 20 kHz. Trolox, a water-soluble analog of vitamin E, has shown to be effective in quenching the presence of triplet state of the Cy3 an Cy5 system we used in our

experiments.

Our experience with Atto425 is that it is very dim and the blue excitation laser can induce a large amount of Ramon Scattering. Atto532 with Atto700 is a good tracking and probing dye pair. As we will see in Chapter 4, there is minimum amount of cross talk between the Atto532 an Atto700.

Qdots are major disappointments. Be careful with them because they blink, also even the specs claim that they have narrow spectrum, it can have some strange spectrum pollution to the dye channel. Finally, qdots can be excited by a very broad spectrum, which sometime would require alternative excitation of the laser. And this alternative excitation of the laser beam can render difficulties in analyzing data.

In our experience, Cy3-Cy5 pair is not the best FRET pair. Cy5 can blink, as shown in Fig.3.6. Cy3 at the same time, can interact with DNA (our own data, not shown).

3.3.4 Labeling and Purification

In this section we outline the major steps and results in labeling and purifying our samples. We learned this technique from Max Greenfeld and Sergey Solomatin in the Herschlag Group in the Biochemistry Department at Stanford, and the detailed procedures are published here:[SH09]. We use one example in which we labeled two 26nt oligos (Old Short and New Short) with Cy3, Cy3B and Atto700 dyes to illustrate these steps.

Briefly, the major steps for labeling and purification are:

1. Ethnol precipitation of the DNA. This step will remove residue amines and to ensure high efficiency in the labeling reaction.

2. Test labeling the DNA. This step is the labeling reaction at a small scale. It can ensure all the components in a reaction function the way it supposed to be.

3. Run analytical gel to check the efficiency of the labeling reaction.

4. Large scale labeling and purification. This is the final step after we made sure the reaction will work.

Now we present the detail of the procedures as well as the results.

Ethnol Precipitation of the DNA

In order to have high labeling efficiencies, it is first important to remove any residue amines, which might exist in the commercial products from the synthesize steps. Ethnol precipitation of the DNA is a good way to remove these residues. The procedures are:

1. Spin the tube, this is just to make sure the DNA won't fly out when you open the lid.

2. Make 3M Na Acetate. MW is 82.03. So add 246mg in 1ml water to make 3M of Na Acetate.

3. Now add 200ul water and 20ul 3M sodium acetate to the DNA tube. The purpose of Na Acetate is to add in cations to allow the DNA to precipitate.

4. Add 3 volume of 100% cold ethanol to the DNA solution. Put the solution in the -80 $^oC$for 3 hours or -20 ^oC freezer overnight.

5. Centrifuge the solution at ¿ 12,000 RPM for 30 minutes. Use centrifuges in a refrigerated room to keep the temperature low.

6. Carefully examine the tube - a white pellet of DNA should be visible at the bottom of the tube. Pipette out the supernatant without disturbing the pellet. Keep the supernatant just in case large quantity of DNA is in there.

7. Add 50ul 70% ethnol and centrifuge again for 20 min. This step is to remove salt.

8. Pipette out the supernatant without disturbing the pellet. Cover the tube and let dry.

Test labeling Reaction

It is a good idea to first perform the test labeling reaction to make sure everything works before dumping all the material to the large scale labeling reaction. The reaction can fail as the active functional groups of the labeling reactants are very volatile and can be inactive if stored for long periods of time. One can also miscalculate concentrations and such. Thus it is necessary to first perform a test reaction prior the large scale reaction.

Before the reaction, we need to pre-run a polyacrylamide gel electrophoresis (PAGE). The exact materials or procedures might vary for different lab, but here is what we have used in the Herschlag Lab:

1. Get some gel plates and put them on small boxes to level off the ground.
2. Wash the plates with ethnol.
3. Put spacers between the plates and then use tape to tape them over. The spacer is about 1mm thick for the small scale testing reaction. Also put a pair of the clips at the end of the plates (the end that oppose the opening) in order to form a shape that is welcome to liquid filling.
4. make the 50mL of gel solution by mixing: 37.5ml of 20% and 12.5ml of 0% gel solution. You also need to make 10% by weight APS (ammonium persulfate), and add in about 500 ul of 10% APS. Finally, add in 50ul of TEMED (N, N, N', N'-tetramethylethylenediamine). You need to do this quickly since the gel can be polymerized very quickly (in minutes).
5. Carefully pour the gel solution to the plate. Knock out any bubble. Insert the well-plate at the opening side.
6. Clamp the last two clips at the end near the entrance. Mark the wells. Then you can just wait for the polymerization complete.
7. Once the polymerization is done, hold the gel plates on the top firmly and then peel off the tape. Razor blade the extra gel on the sides, and also you need to cut the extra well gel by the short plate edge.

8. Take a gel box, and then put the gel plates with the short one facing the top well. Get a green spacer, (the one that has a shape that matches the opening of the top well.) and two black spacers. Clamp the gel plates to the top well, with the green spacer and black little spacer in between. You should also push the the black spacer down in order to form good seal between the plate and the well. Remember to use a second pair of clips to clamp down the plates as well.

9. Get the 10X solution. Fill 100ml into the cylinder, and then add 900ml of water. Mix it by flipping it up and down, and then pour the solution to the top well as well as the bottom well of the gel box.

10. Finally, plug in the electrodes: positive on the bottom and negative on the top. The total amount of power should be around 40 to 50 watts. You need to check the gel every 15 to 20 min to see if there is any thing strange. Pay attention to see if there is any leak.

11. To prepare the sample for the gel, add in 4ul of loading buffer to the sample.

12. Wash the gel channels using a syringe. Then load in the solution by the really long flat head tip. You will have to check the gel constantly just to make sure they will not overrun.

Before the labeling reaction, one needs to measure the concentration of the oligonucleotide, and use high concentration of the oligonucleotide (5 mM in this example). The labeling reaction is done by first mixing the the aqueous reaction components - oligonucleotide, water, and phosphate buffer, then adding the dye in DMSO buffer, and incubating at 37 ^oC for about one hour. It is important to work fast to avoid any unnecessary hydrolysis of the DHS. As an example, the following chart shows what we have used in our labeling reaction. The amount of DNA are varied for comparison.

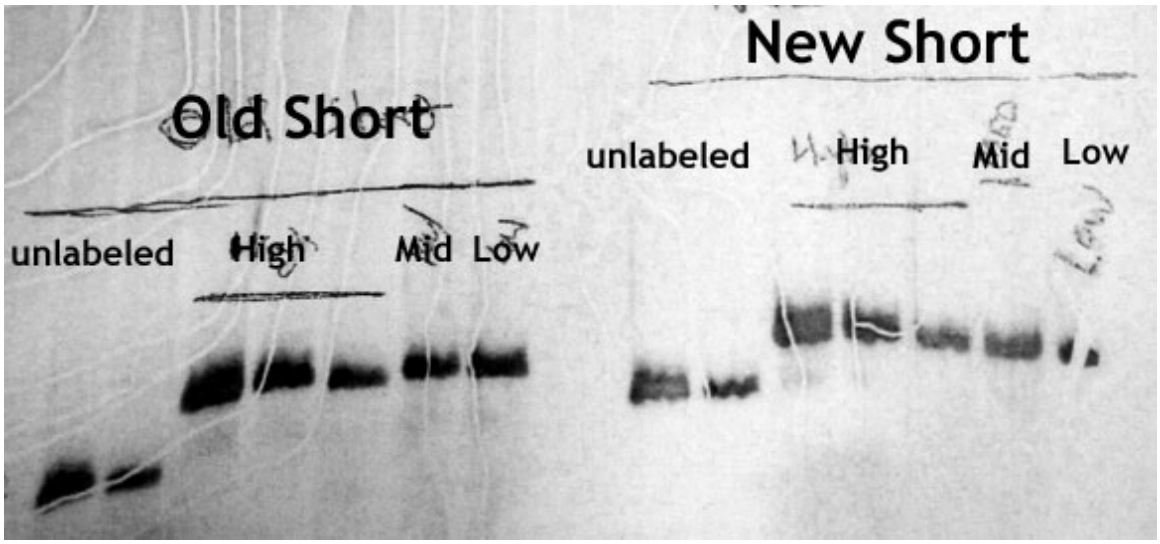

Figure 3.7: Test labeling results.

DNA Sample	High	Mid	Low		High	Mid	Low
Old Short 1X (5.3 mM)	0.5 μl						
Old Short 10X (0.53 mM)		0.8 μl	0.2 μl				
New Short 1X (5.6 mM)					0.5 μl		
New Short 10X (0.56 mM)						0.8 μl	0.2 μl
Phosphate Buffer 500mM pH 8.7	0.2 μl	0.2 μl	0.2 μl		0.2 μl	0.2 μl	0.2 μl
water	0.3 μl	0μl	0.6 μl		0.3 μl	0 μl	0.6 μl
Cy3 Dye + DMSO (7 ul)	1 μl	1 μl	1 μl		1 μl	1 μl	1 μl
Total	2 μl	2 μl	2 μl		2 μl	2 μl	2 μl

At the end of the incubation, add in suitable amount of water so that all the tubes have the same concentration of DNA. In our case, add 148 μl water in sample 1 and 4, 22 μl water in sample 2 and 5, and 4 μl water in sample 3 and 6. Finally pipette out about 250 ng of DNA to tubes with gel loading buffer and run the gel prepared. After stain the gel, we see this:

Note that the unlabeled samples goes further than the labeled ones. Since there is very little amount of unlabeled samples at the labeled channel, our labeling reaction was quite successful.

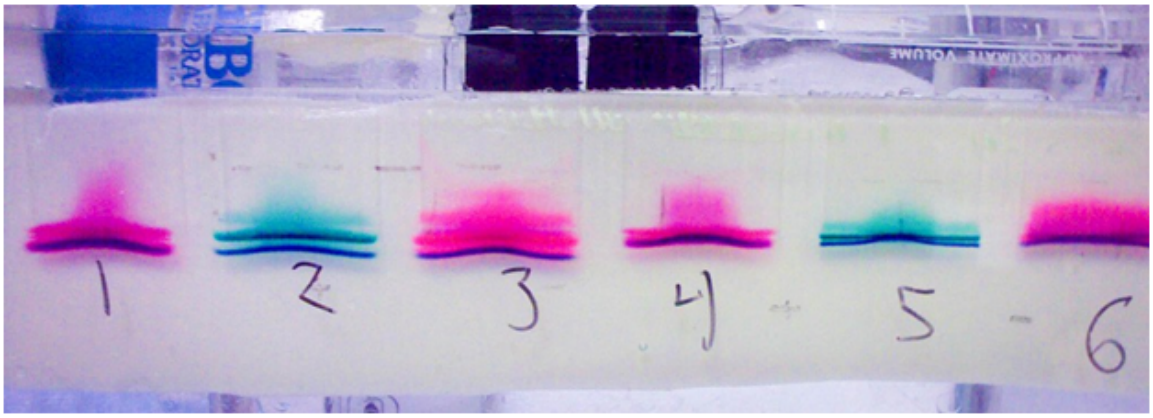

Figure 3.8: Purification gel. From 1, 2, 3, they are Cy3, Atto700, and Cy3B. 1 to 3 is the old short strand, and 4 to 6 is the new short strand. The smeared color is likely to be dye aggregates while the concentrated line is the labeled DNA.

Large Scale Labeling and Purification

Next we can perform a large scale labeling reaction and purifying the samples. The procedures are very similar to what we described above. You can also run a round of ethnol precipitation before running the purification gel. Here is one photo we have for the purification gel. From 1, 2, 3, they are Cy3, Atto700, and Cy3B. 1 to 3 is the old short strand, and 4 to 6 is the new short strand.

Once the purification gel is done, a clear band of the labeled material is visible. Pop one of the glass side open, cover it with saran wrap, put on a piece of white paper, and then use fresh blades to cut the band that contains the DNA. Finally, put the gel pieces in a tube. Next, use a pipette tip to crush the gel and spin the gel pieces to break them up. Then put some (300 μl) TEN buffer into the tubes, and place the tubes on the dry ice. When everything is completely frozen, take the tube out and unfrozen it. You want to freeze and defrost the tubes at least for three times. Then put them in 4 ^oC refrigerator overnight. At the end of this procedure, one should expect the labeled DNA is in the solution. Pipette out the solution (careful not to pipette out small gel pieces), and run a round of ethnol precipitation to concentrate it. Finally perform a final round of testing gel to check the purity of the sample.

Solution Conditions and Measurement Channel Construction

With a dilution of four, TE buffer (Invitrogen) was used as the buffer in the openloop FCS measurements in order to protect the DNA. Various amount of 1M NaCl (Sigma) was added in the TE buffer to the desired NaCl concentration.

Chemical systems can be powerful tools in improving dye stability. 1mM of trolox was used in FCS measurements of Cy3-Cy5 labeled hairpin, as trolox has shown to suppress the blinking of the acceptor[RMH06]. In experiments with Atto dyes, we did not add any oxygen scavenging or triplet trenching chemical (such as ascorbic acid) in the solution as they would quench the Atto700 fluorescence entirely or cause Atto700 blinking through photoninduced electron transferring (PET) [KMS00, HKPS03, KDNS06]. It has been shown that by balancing the amount of reductant or oxidant, the oxazine dyes can become single molecule switches, which may find applications in super resolution imaging[VCF+09, OSG12].

Two pieces of cover slips (VWR micro cover glass) were sandwiched by two pieces of doubled sided tape (3M) to form a 5mm wide, 20mm long, and about $100\mu m$ deep channel as our solution space. One percent casein solution (Sigma) was applied to the channel for ˜5 minutes. Then the channel was washed three times using 25 μl of measurement solutions before the final measurement solution with DNA samples was loaded. This procedure greatly reduced hairpin absorptions to the glass surface.

3.4 FCS Setup and Measurement Procedures

In this section, we briefly describe the setup and procedures we take in performing the FCS measurements.

3.4.1 FCS setup

Standard FCS measurements were performed on our in-house custom-built fluorescence microscope, shown in Fig.3.9. This setup is part of the tracking setup that we have built [MM09]. The FCS measurements are done with expanded laser beams with maximum waist size around $10\mu m$ (For dye-quencher, we used a 634nm laser with waist size ~$2\mu m$. For FRET probing system, we used both $1\mu m$ and $10\mu m$ laser beams). Since the typical diffusion time scales as waist square, this greatly improved the upper time scales of FCS: from less than 1 mS to more than 20 mS. We used pinhole size of ~ $300\mu m$. Fluorescence after the pinhole was spectrally separated by a dichroic mirror (Chroma 625 DCXR) and detected by photon avalanche photodiodes (APDs. Model: Perkin Elmer SPCM-AQR 13/15) with bandpass filters of 585/40 (Chroma) and 700/50 (Chroma). In all the FCS experiments, fluorescence signals were recorded by pairs of APDs in Hanbury-Brown-Twiss configuration in order to get around the after-pulsing effects of our detectors and to probe sub-microseconds dynamics. Laser profiles were obtained by scanning a bead (Bang's Lab, USA) on the coverslip, similar to our previous experiments[MM09]. Laser power were locked at constant level using home-made locking-servos and set well below the saturation level of the dyes. A typical FCS experiments used peak laser power intensity of $0.3kW/cm^2$.

3.4.2 Measurement Procedure and Data Analysis

Once the setup and channel is ready (make sure there is good alignment, adjust to the right laser power level, and measurement channel has been coated by casein), we typically take 60 ten seconds FCS measurement. For each measurement, we use

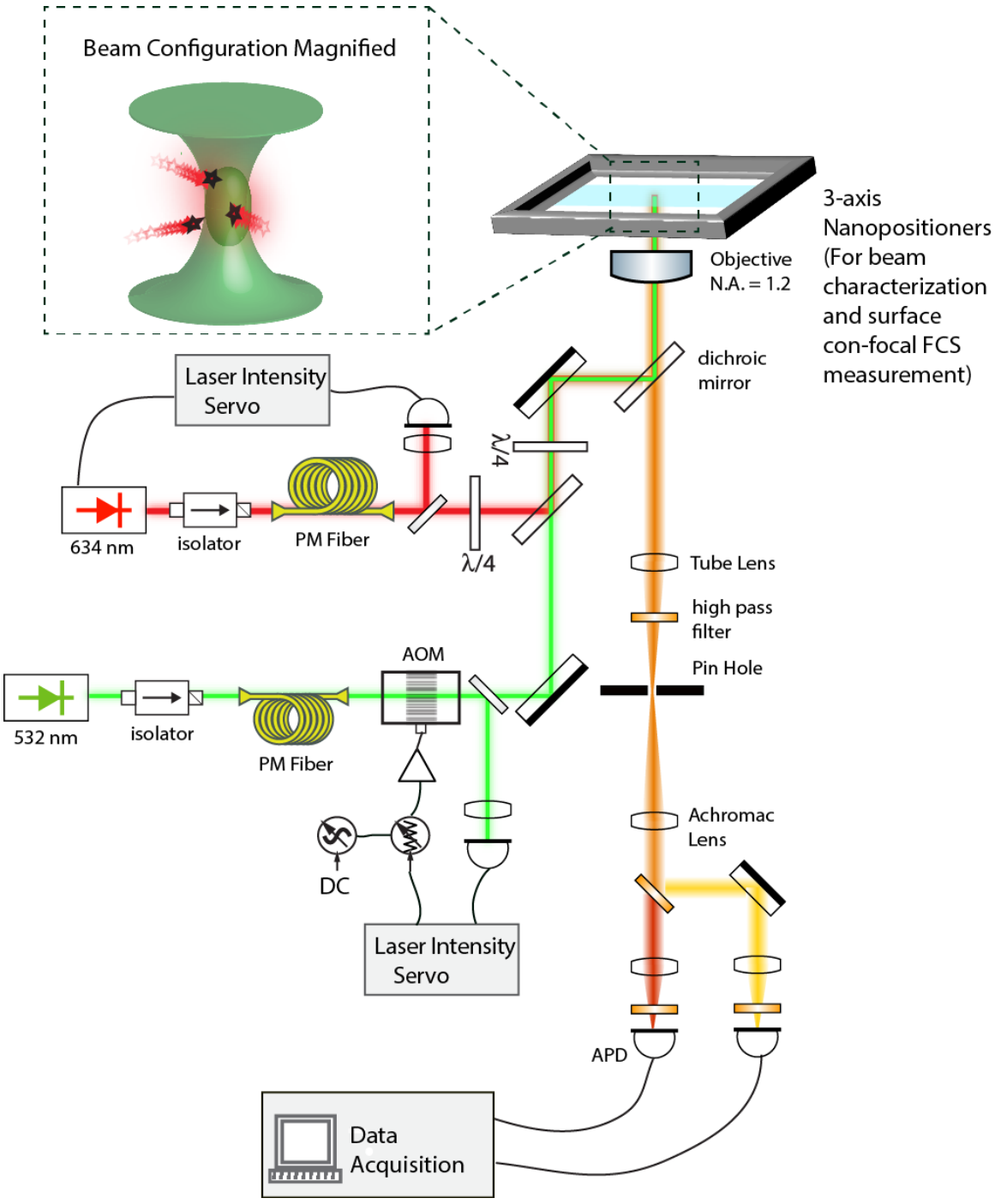

Figure 3.9: The schematic of the FCS setup (part of the tracking setup (shown in 4.2 and [MM09]). The beam size is controlled by a telescope (not shown) on the laser beam path before the microscope, which is quiet useful in aligning a big beam. For details, please see text.

all the photon-arriving events and calculate the individual FCS curves using an algorithm described in[LFH06]. The 60 individual FCS curves are then averaged and the standard deviation of the mean is calculated as the measurement error. See next section for examples!

3.5 Characterizing the dye systems, experimental apparatus, and center-of-mass diffusion of the hairpin

We probe the folding reaction by monitoring the fluorescence signal. However, besides folding, probing dye's intrinsic blinking and the center of mass diffusion can also modulate the fluorescence signal. Therefore we must first perform control experiments to characterize them before our measurement of the hairpin folding kinetics. At the same time, these experiments allow us to understand what temporal range our FCS apparatus can probe. This understanding helps us in improving the apparatus.

3.5.1 Probing Dye Photophysics

The two major dye systems we have used in our studies is 1) the Atto700-dG, the dye-quencher pair, and 2) Cy3-Cy5, the FRET pair. In this section we will present experimental results and analysis in characterizing their photophysics. Both systems have been characterized thoroughly in the literature. Specifically, Sauer and co-workers have pioneered our understanding and usage of the ordinary oxazine dyes. They have shown that the photon induced intramolecular electron transferring (PET) can occurs when the dye contacts with guanosine, which can be used in monitoring conformational changes of DNA hairpins[KMS00, KDNS06]. They have also presented elaborated studies in which they demonstrated that by carefully crafting the redox environment, one can independently control the on- and off-state lifetimes of the blinking behavior of the dye[VCF+09]. This remarkable discovery allows oxazine dyes to be widely applied in various super-resolution imaging techniques[OSG12]. Cy3-Cy5 FRET pair, on the other hand, has been extensively used in many folding studies[HTL+99].

We mainly used FCS to check whether the dyes blink. Eq.3.7 states that if there is no extra dynamics but the diffusion, the measured FCS curve should be satisfactorily

described by the pure diffusion FCS given by Eq.3.2.2:

$$\begin{aligned} g_2(\tau) =& \frac{< \delta I(t)\delta I(t+\tau) >}{< I(t) >^2} \\ =& (DNA(\tau)+1)\cdot[\frac{1}{N}Diff(\tau)] \end{aligned} \tag{3.12}$$

where,

$$Diff(\tau) = \frac{1}{1+\frac{\tau}{\tau_D}}\frac{1}{\sqrt{1+(\frac{r_o}{z_o})^2\frac{\tau}{\tau_D}}}$$

This means that if our measured dynamics cannot be described by $Diff(\tau)$, it certainly has more dynamics other than the simple diffusion process. On the other hand, if it can be described by $Diff(\tau)$, it can still have dynamics that are either too long for FCS to measure or behave in a way that the autocorrelation functions of the dynamics is independent of the time delay τ. The latter effect will not affect our measurement of the folding behavior anyway, as long as that the two dynamics are not correlated. If the photophysics of the dye is too slow for FCS to measure, then we need to resort to other methods.

Fig 3.10 displays the FCS signals as well as a fit to the standard diffusion model in our measurement solution. Testing the residues against a null hypothesis that the data is generated from a Gaussian distribution with zero mean, we cannot reject the standard diffusion model for this data set. This is one evidence that our Atto700 dye does not have significant blinking in the diffusion time regime for this particular experiment (from 10^{-7} S to 10^{-4} S).

Keep in mind that the blinking behavior of the dye is highly influenced by the solution condition. Depending on the oxygen levels and redox condition, the same Atto700 dye will have very different blinking statistics. For example, addition of ascorbic acid in the solution can switch the dye into dark state and only spikes of fluorescence (< 10 mS) were observed.

We next measured the Atto700 dynamics with a DNA sequence that has no quencher. Data is very similar to Fig.3.10, indicating that Atto700 does not interact with other bases or the backbone of the DNA.

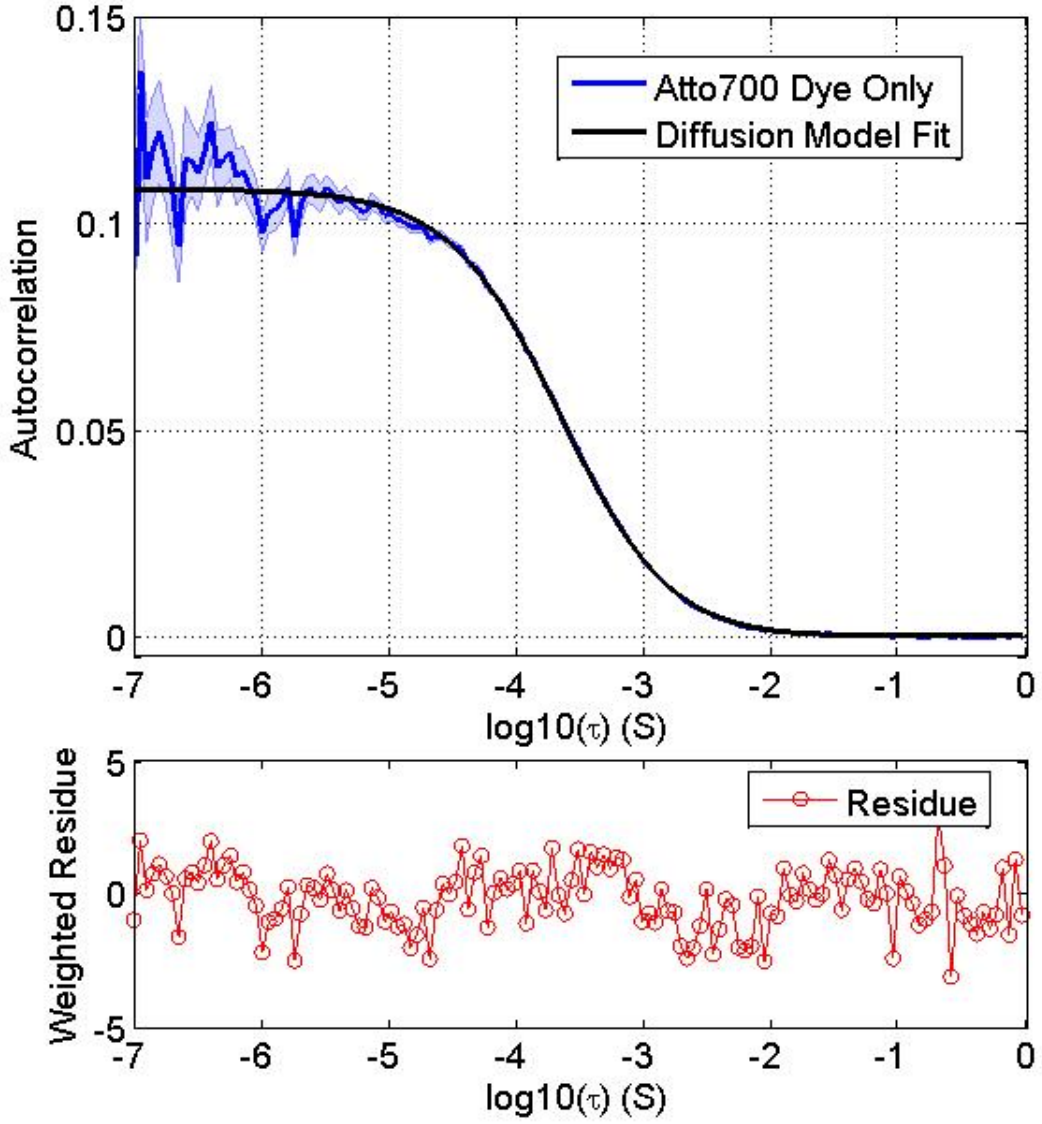

Figure 3.10: Atto700 dye photophysics probed by FCS. FCS signals (blue trace) of Atto700 in our measurement solution and a standard diffusion fit (black trace) to the FCS signals. Residues are displayed in the panel below. Testing the residues, we cannot reject the standard diffusion model to describe the dynamics observed. This measurement suggests that Atto700 does not have intrinsic photophysics in the time regime that FCS could probe.

Our FCS measurement cannot observe dynamics that happens slower than the typical diffusion time. To see if there is any slow blinking, we tethered ALLT sample on functionalized glass surface and used a con-focal microscope to observe one molecule a time. The tethering is done by utilizing the strong binding between biotin and streptavidin. What we observed is displayed in Fig.3.11. Strikingly, we see apparent blinking of the Atto700 dye. With filtering, we determined that on average, the on-state life time is 7.2 seconds and the off-state life time is 0.2 seconds. This dynamics is too slow to be observed by FCS.

These measurements also reveal the brightness of the dye, which later is a crucial parameter for us to determine the transition rates of the folding reaction. We will discuss about this in detail in the following sections.

Similar experiments to test whether Cy3 blinks at the time scales that our FCS apparatus could detect reveal that there is no significant blinking behavior. Because of the resemblance of the data, we will not repeat presenting it here.

3.5.2 Improving the Temporal Detection Range of FCS

At the same time, we need to understand what temporal range our FCS apparatus can probe. For fast dynamics, one can always increase the data-taking time to accumulate more statistics to increase accuracy at that time regime. Given our typical data acquisition time (10 minutes), dye brightness and noise level, we can detect microseconds dynamics if they present more than 1% of the entire states population. The accuracy of our detection goes down for faster dynamics but can be compensated by increasing the detection time or laser power.

The slowest dynamics can be observed by FCS experiments is determined by the time that the molecule stays in the laser focus. Given the molecular species and solution condition, this upper bound is ultimately limited by the signal-to-noise ratio of the fluorescence signal, which is proportional to the brightness of the dye and inverse proportional to the area of the laser. Past FCS experiments always used diffraction limited laser beam since the dyes were dim. Thanks to recent advancement in engineering better dyes, it is possible to extend the temporal upper bound of FCS

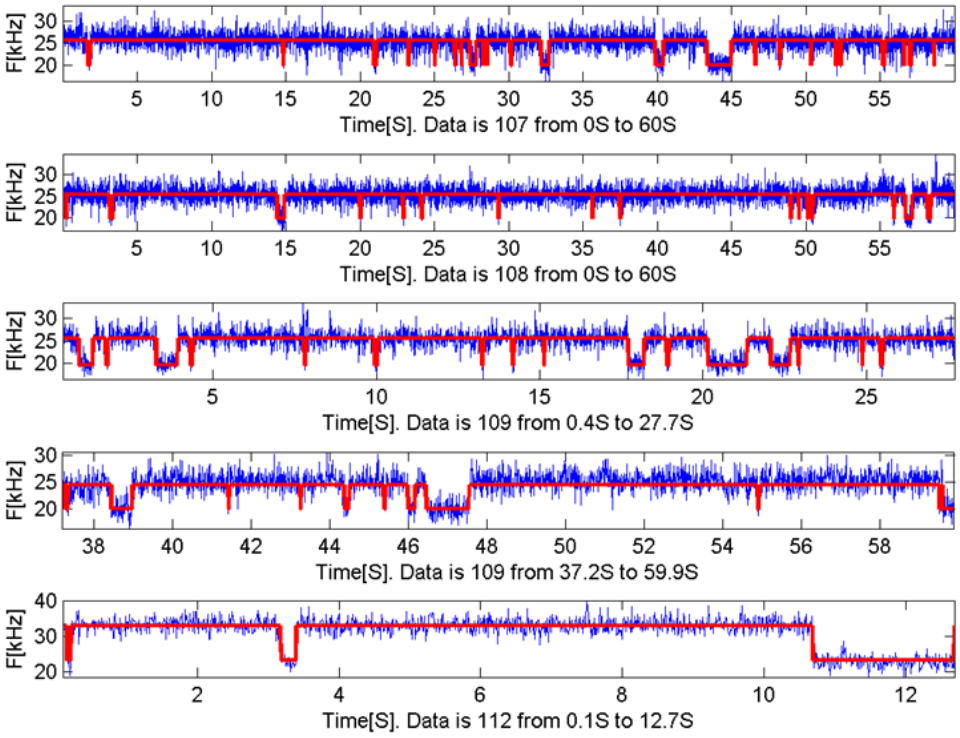

Figure 3.11: Single molecule surface measurement reveals that Atto700 blinks at time scales that could not be detected by previous FCS measurement. Blue traces are fluorescence of single Atto700 dyes labeled on DNA construct ALLT, which is tethered to glass surface by biotin-strepavidin link. Red curves are the state estimation based on a maximum-likelihood estimator. On average, the on-state lifetime is 7.2 seconds and the off-state lifetime is 0.2 seconds.

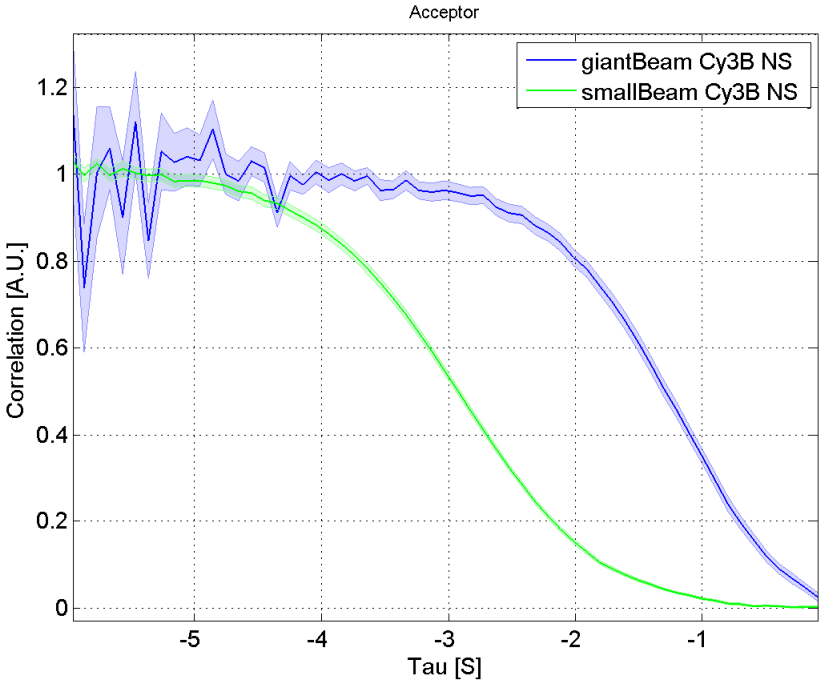

Figure 3.12: Expanding the temporal range of FCS measurement. With bright dyes such as Cy3B, it is possible to use beam size (waist size) as large as $10\mu m$ to perform FCS experiment. In this example, folding dynamics up to 40 mS can be observed on a Cy3B labeled 26 nt ssDNA by FCS. Comparing a $0.4\mu m$ size beam, this is almost two orders of magnitude improvement in terms of temporal range that can be probed by FCS. Note that the current sample is a very small biological molecule: 26 nt ssDNA. Its diffusion coefficient is around 140 $\mu m^2/S$. A typical 60 kDa protein would have diffusion coefficient around 60 $\mu m^2/S$, which means that we can observe around 100 mS folding dynamics by FCS alone!

to more than 40mS by bright dyes such as Cy3B (sample is a 26 ssDNA). As shown in Fig.3.12, the $10\mu m$ waist size beam improved the slowest temporal range of the FCS by almost two orders of magnitude when compared to the results obtained by a $0.4\mu m$ beam. This greatly extends the detection range of FCS in probing molecular folding dynamics.

3.5.3 Folding and Diffusion of the Hairpin is Coupled.

The General Model is about the intramolecular folding, which seems to have little to do with the global diffusion of the molecule. However, the formation of stem limits what conformations the hairpin could adopt, which in turn can affect the diffusion. If this is the case, the coupling of these two processes could contribute to the functional form of the FCS signal. Since the functional form of the FCS signal is what we use to extrapolate the folding model of the hairpin, it is important for us to investigate the possibility of this coupling.

We first measured diffusion coefficient of the three base pairs hairpin. Diffusion coefficient of (HP3 - Cy3) increased 30 to 40% as salt concentration is elevated. (Fig 3.13A, C) Although we know that the hairpin stays at its folded form more at higher concentrations of salt, this observation along cannot conclude that the formation of the stem is the main cause of the increased diffusion. This is because that our construct has two parts: the dsDNA tether and the ssDNA hairpin. We will need to first examine which part is responsible for this increased diffusion. Diffusion coefficient measurements on dsDNA tether samples did not reveal dependence of diffusion coefficient on salt concentrations (Fig.3.13), confirming literature's result[SBRN06]. This suggests that the ssDNA part is responsible for this diffusion coefficient increase for the overall construct. Diffusion coefficient measurement on ALLT (no tether) shows one step increase at 5mM NaCl (Fig.3.13D). The poly-thymine chain diffuses about 20% faster at higher salt concentration (also confirming literature's result [MN08]). Diffusion coefficient measurement on HP3 (no tether) shows two step increase at 5mM and

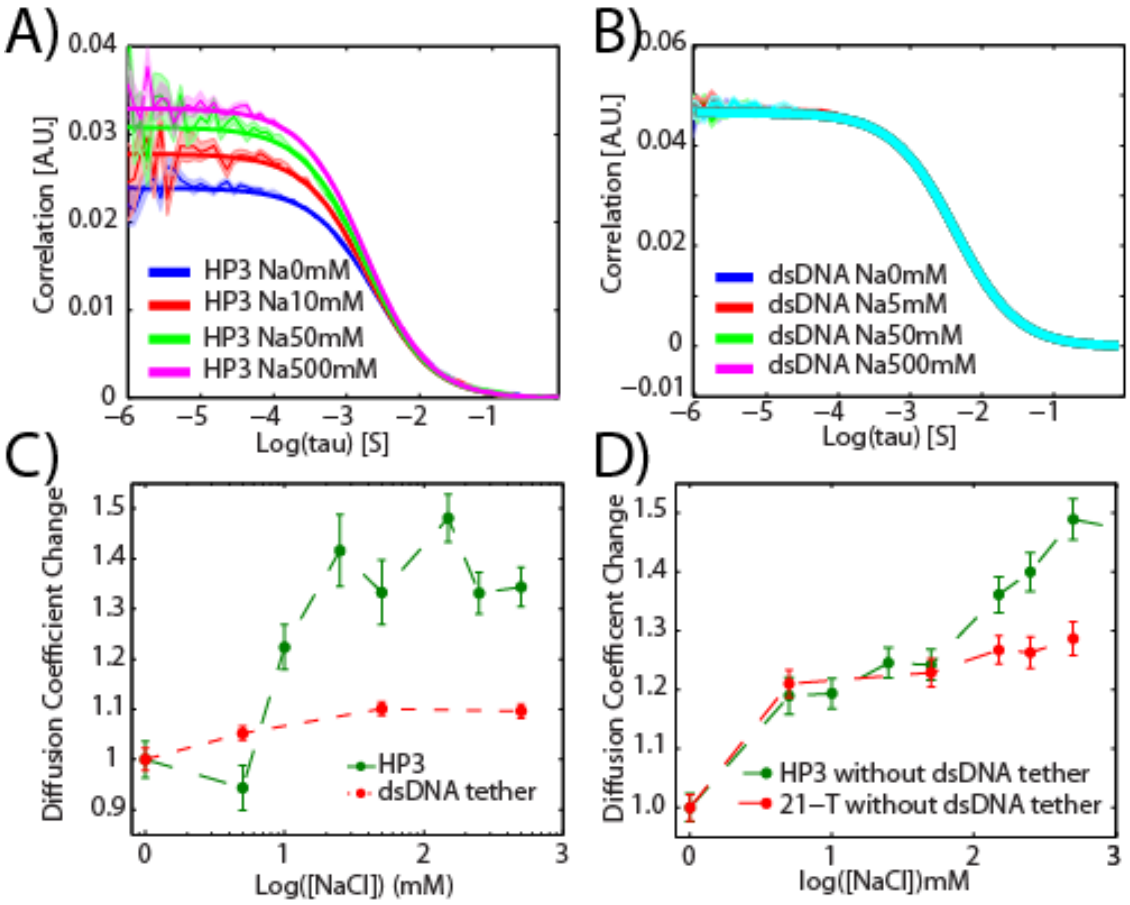

Figure 3.13: Salt dependence of diffusion coefficients and dye brightness of HP3 and dsDNA tether. A). FCS from a Cy3 only HP3 at four different salt concentrations. The shaded regions are standard deviation of the mean of a series of identical measurements (dashed), and solid curves are from fitting results which employed the standard diffusion model. The FCS curves are normalized by multiplying constants which minimized the least square differences for $\tau > 10^{-2}S$. The elevated amplitudes indicated that the diffusion coefficients increased as NaCl concentrations were increased. B). FCS from an Atto700 labeled 26-bp dsDNA tether at four different salt concentrations. Solid cyan curve is from fitting result from measurements in 500 mM NaCl. C). Normalized diffusion coefficients extracted from A) and B) as a function of salt. The error bars are from fitting (for more details, please see main text). D). Salt dependence of diffusion coefficients of ssDNA obtained using FCS. Red curve is the diffusion coefficients of a polythymine(ALLT-noTether); green curve is the diffusion coefficients of the hairpin under investigation (HP3-noTether). Notice both sequences do not have the 26 bp dsDNA.

50 mM (Fig.3.13D). The first step coincides with what we saw for the poly-thymine chain. It is likely that this increment of the diffusion coefficient is due to salt electrostatic screening of the backbone charges, which makes the chain more flexible. The second step increased the diffusion coefficient for another 30%. As our later folding data suggests, this second step is caused by the dominance of the folded hairpin in these experimental conditions. This is one interesting result and we will come back to compare it with the results we obtain from our folding data. All our experimental results suggest that the hairpin stem formation is the origin of the increased diffusion coefficient of the molecule. This result indicates that the folding and diffusion is not independent processes, which is usually assumed in the hairpin literature(As stated in Eq.3.7). This unexpected coupling could resolve controversies regarding different folding models proposed in previous studies.

3.6 FCS Experiments on Dye-quencher Labeled Three Base-pair Hairpin Suggests that the Folding Is a Two-State Process

After characterizing the fluorescent probes and our apparatus, we investigated the folding of the 3 bp stem hairpin by dye-quencher system and FCS. This hairpin can have all the possible states in the GFM. The manifestation of these states lies on the functional form of the folding FCS signal. If it can be fitted satisfactorily by an exponential decay function, then a two-state folding model could be a possible candidate to describe the folding. This can be one evidence supporting that misfolds are not significantly presented in the conformational state population. It also should be noted that the stem formation might start around 50 mM salt as the molecule's diffusion coefficient starts to increase at this salt concentration. Thus it would be interesting to see at what salt concentration the folding dynamics starts to appear in our kinetic measurement and whether the values we obtained are consistent with our diffusion data.

We performed FCS measurements on dye-quencher labeled 3bp hairpin at various salt concentrations (Fig.3.14). In salt concentrations less than 50mM, we only observed one single FCS relaxation contributed by diffusion. When the FCS signals of HP3 compared to ALLT, and HP3-CTL, we could not distinguish them at the same salt concentration(Fig.3.14A), suggesting the conformational similarity among these molecules at these low salt concentrations. At the same time, we found the dye brightness stayed constant within measurement error bounds(Fig.3.14D). These results all suggest that the quencher and dye were most of the time ($> 99\%$) not in contact at low salt concentrations ($< 50mM$). We also performed FCS measurements on FRET labeled 3bp hairpin, and compared that to a FRET labeled poly-thymine chain ALLT. In $< 50mM$ salt concentrations, both average FRET efficiencies and FCS signals for these two samples are not statistically different. This further supports that at low cation concentrations, the hairpin behaves like a flexible coil. In solutions with salt

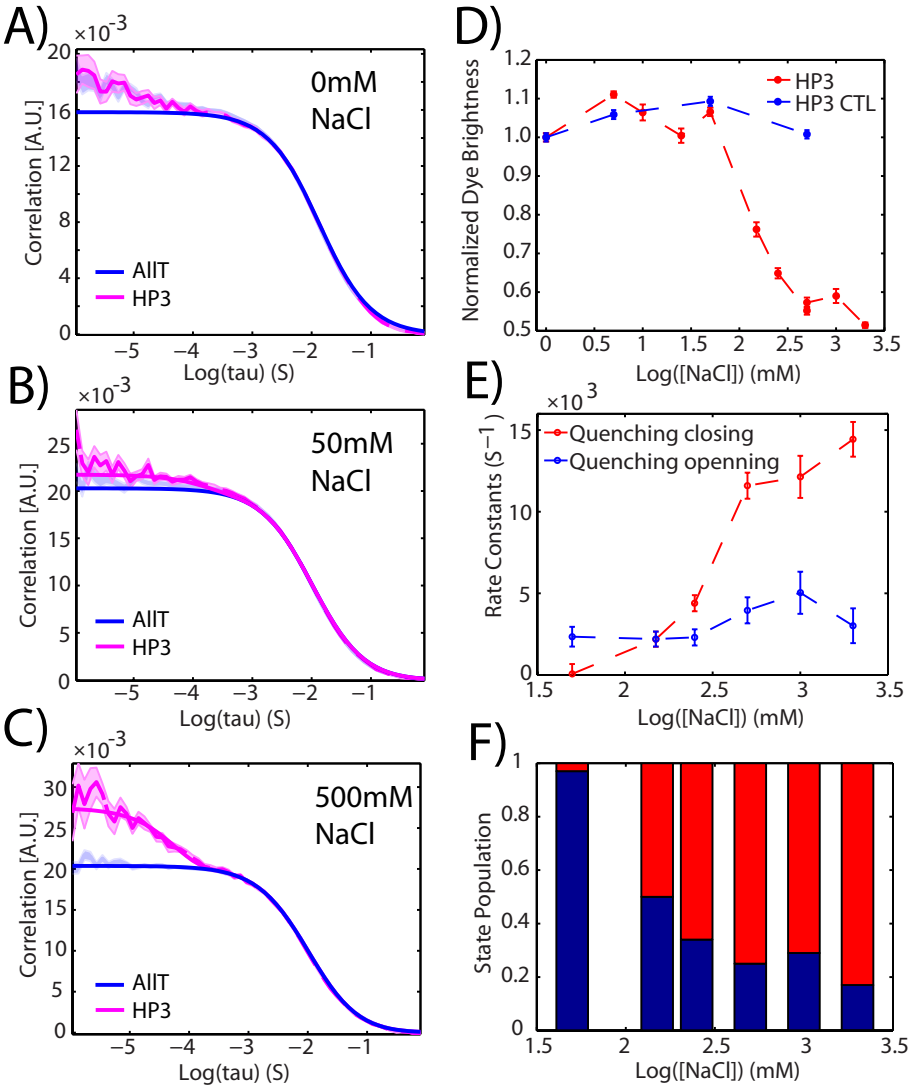

Figure 3.14: Hairpin dynamics at various salt concentrations probed by Atto700-dG dye-quencher FCS. In A, B, and C, we plotted the FCS curves from ALLT, T20G, GTA, GTC and HP3 at 0, 50 and 500 mM of NaCl concentrations. D) The normalized brightness of Atto700 in controls and hairpins are plotted. E), transition rate constants from NaCl 50mM to 2M are plotted. F) Population fractions of the two states are plotted based on the calculated rate constants.

concentrations more than 50mM, we see very different behavior from the 3bp stem hairpin(Fig.3.14B, C). All our results suggest that at salt concentrations more than 50 mM, hairpin stem starts to form. First we see a distinguished FCS relaxation at around 100 uS in addition to the diffusion contributed FCS (Fig.3.14C). As salt concentration is increased, this relaxation is more and more distinguished (amplitude increased) and finally stabilized when the salt concentration reached 500 mM. Using HP3-CTL or ALLT as control, we extracted this folding dynamics and found that a single exponential decaying function could fit it satisfactorily. This suggests that at these salt concentrations, the conformation changes of the hairpin make the dye blinking in a two-state fashion. Further, the dark state became the dominant state as salt concentration was elevated, increasing its occupation of the states from 10% at 50mM salt to 80% at 500mM salt (Fig.3.14F). This is consistent with the well-known fact that the folded hairpin becomes more stable at higher salt concentrations and thus the dye and quencher were able to stay at proximity. The decreased dye brightness data is another independent measure of the folding dynamics at salt concentrations more than 50mM NaCl(Fig.3.14D). The dye brightness of HP3-CTL stayed constant in the entire salt concentration range (from 0mM to 2,000 mM). However for HP3, the dye brightness at 500mM NaCl decreased to 50% of the value at 5mM NaCl. It is very exciting for us to see that our folding data corresponds well with the diffusion data. The folding data is about local information: it is the result of end-end contact of the chain (which is one indicator for the stem formation, a local structure); the diffusion data is global information: how fast a molecule diffuses depends on the overall conformation of the chain. The consistency of these two suggests that it is the whole chain is opening and closing that cause the end of the chain colliding and forming the hairpin stem. Although we now have one evidence to suggest that the hairpin folding is a two-state process with the whole chain transiting between a random coil and a fully folded chain, we would like to further investigate this process and obtain more evidences. This is partially because we learned from the literature that hairpin folding is complex, and partially because that our evidence so far is not conclusive yet (for example, one can certainly construct misfolded confirmations that can give us all the results we have observed and yet are significantly presented in the state

population.). The conservative conclusion that we could have here is that the stem is necessary for the observed folding FCS signal. Now we want to further investigate, by engineering sequences that could isolate and favor certain states, 1) whether or not misfolds are significantly presented that they could affect the folding pathway, 2) whether or not it is the whole chain opening and closing that causes the end of the chain colliding and forming the hairpin stem. The following several sections answer these two requests in turn.

Extracting Folding Rate Constants

There are four free parameters determining a two-state FCS curve (Eq.3.3):k_{lh} the transition rate from the dark state to the bright state, k_{hl} the transition rate from the bright state to the dark state, Q_h the dark state dye brightness, and Q_l the bright state dye brightness. Fitting FCS data to a two-state model, however, can only give us two fitting parameters (amplitude and decaying rate k of the FCS curve) that are functions of the four free parameters. Instead assuming a total dark state as commonly done in the literature, we looked into our fluorescence data and directly obtained brightness data that can give us the two additional parameters: we measured the time-averaged brightness of the dye, which is $\pi_h \cdot Q_h + \pi_l \cdot Q_l$; we also measured the unfolded-state brightness by measuring a sample without the quencher, which we assumed, based on the quenching mechanism, that the unfolded-state brightness is the same as a dye on the control sample without any quencher. With these two extra parameters, we solved the dye-quencher rate constants, which are plotted in Fig. 3.14E. One interesting aspect about this procedure is that it gave consistent dark state dye brightness for all the HP3 sample at different NaCl concentrations, a hint that our procedure is correctly predicting the dark state brightness since as we can experimentally prove later, the physical molecular conformations corresponding to this dark state are the same for all the molecules at different salt concentrations.

Dye Brightness Extraction

By fitting measured $g_s(\tau)$ at $tau \gg \tau_D$ with the pure diffusion model, we extracted the average number of molecules in the laser focus. In the meantime, since we recorded all the photon-arriving events by our instruments, we could also calculate the average fluorescence rate from those molecules. Thus the average brightness of the dyes under the Gaussian laser excitation is easily extracted: $\langle B(t) \rangle = \langle I(t) \rangle / N$, which gave us another important parameter to peek into folding.

3.7 Engineered DNA Sequences that Favor/Isolate the Presence of Specific Misfolded States Did Not Find Significant Presence of Such States

We now check the open hairpin to misfolded hairpin branch in the GFM. As proposed by previous studies ([AKS01, KDNS06]) that the interaction of the stem and loop can form misfolded conformations and trap the hairpin from forming the stem, we designed sequences that promote the interaction between the loop and the stem (GTA, T4C, and HP3-CTL. For their sequence information, please see Table 3.3.1). For example, GTA has a adenine at the stem location, which can form basepair with thymine in the loop region. We used the dye-quencher system to study their folding dynamics. We first expected that GTA, GTC, and T4C would display relatively fast dynamics (in microseconds range) other than just the diffusion contributed FCS relaxation. This is because the trapped states are only hold together by one or two base pairs and thus should not be very stable. To our surprise, when we examined these samples by FCS, all these samples displayed FCS signals within error bounds to a FCS signal given by ALLT - a ssDNA chain without any quencher. This is true for all the salt range we investigated, from zero mM to 2,000 mM NaCl (Fig.3.15). We also extended our observation time in order to observe dynamics as fast as 10 nS. We did not find any evidence of folding dynamics in those samples across our measurement sensitivity and solution conditions. Besides FCS signals, the average dye brightness is another independent parameter that allows us to check for signs of folding. We found that the dye brightness of the these samples did not differ from that of the ALLT. This means that the dye on average is not quenched. Lastly, the extracted diffusion coefficients of these samples also stayed within error bounds when compared to that of ALLT. At this point, we do not know whether the conformation of these misfolded samples are the same as ALLT, but we can safely conclude that it is more similar to the ALLT than to the three base pair stem hairpin, as the three base pair stem hairpin would change dramatically its diffusion coefficient at these high salt concentrations.

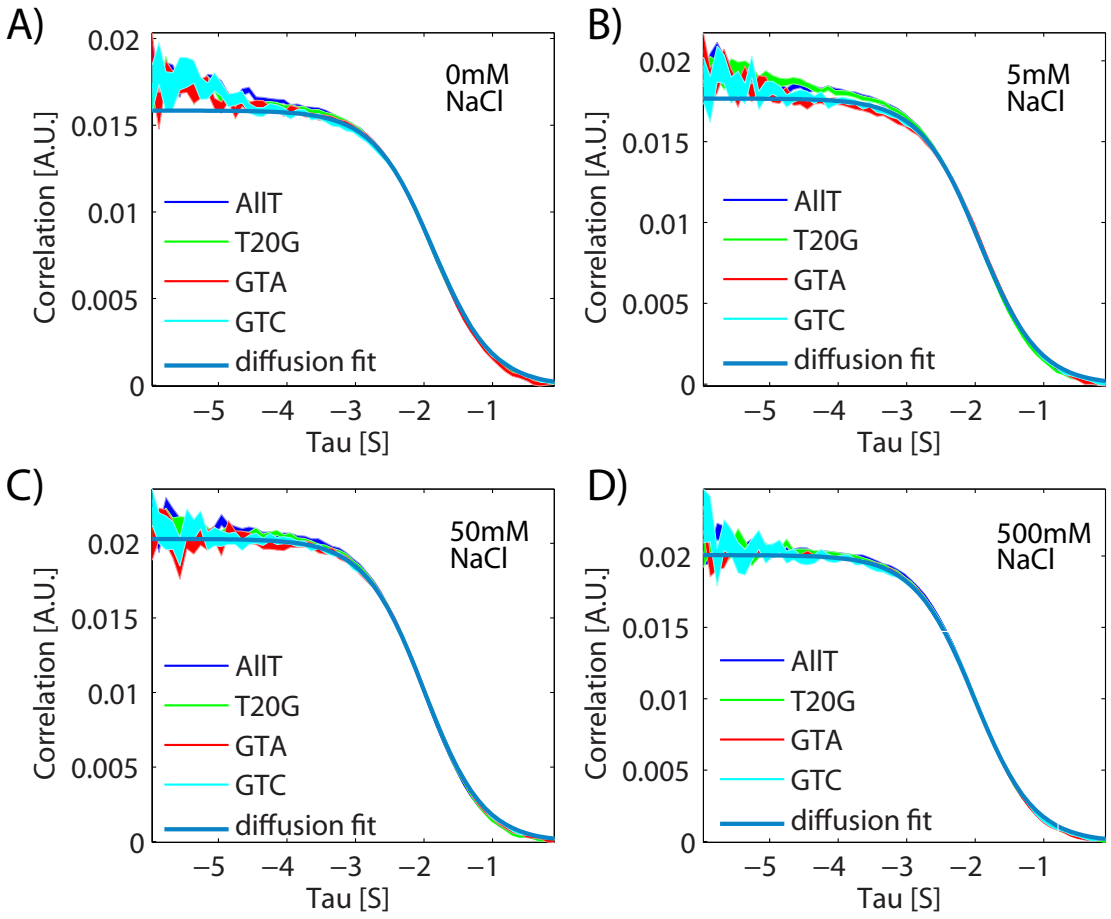

Figure 3.15: FCS signals from sequences that isolate and promote misfolded states at four different salt concentrations. All these samples displayed FCS signals within error bounds to a FCS signal given by ALLT - a ssDNA chain without any quencher. This is observed for all the salt range we investigated, from zero mM to 2,000 mM NaCl. This result suggest that one bp misfolds by either canonical or non-canonical base-pairing is not strong enough to hold the 21 nt chain together to affect the folding pathway.

All our experimental evidence suggests that the ends of the ssDNA chain most of the time ($> 99\%$ of the time) are not in contact. Their conformational state is more similar to ALLT than to the three base pair stem hairpin. The one basepair misfolds by either canonical or non-canonical base-pairing is not strong enough to hold the 21 nt chain together to affect the folding pathway.

3.8 Kinetics of Abasic Hairpins Suggests the Stem-loop Interaction Is Not Important in the Folding Reaction

The combined results of the above two experiments suggest that the three basepair stem is necessary for the observed dynamics, but it does not prove that the stem is sufficient for such dynamics. In this experiment, we provide such evidence.

The chief attribute to our experimental design is the use of a set of hairpins that could not possibly form any stem-loop basepairs - a set of stable mimic of abasic hairpins. These samples' loop is a base-free sugar-phosphate backbone (consisting of 15 dSpacer, or abasic furan, instead of 15 dT), thereby preventing stem-loop interactions. We investigated three abasic samples, ALLX, GX, and HP3X. Their sequence information and purpose is on Table 3.3.1. ALLX has a loop that entirely consists dSpacer. We use ALLX to measure diffusion of the molecule with abasic loops. GX has an abasic loop but with a quencher at the end. We use GX to measure end-to-end collision of the abasic chain. HP3X has an abasic loop but a normal 3 base-pair stem. We use HP3X to measure the stem formation/detachment in the case of abasic loop.

We used the dye-quencher system to probe the dynamics of the abasic hairpins. We predict that if the hairpin folds according to the complex folding model with stem-loop interactions, a two-state folding model is still possible; on the other hand, owing to the extra folding pathways, the rate constants we measure for the abasic hairpins would be different from those of the ordinary hairpins in Experiment 3.

Abasic hairpins lack bases on the loop but the our results with these samples are strikingly similar to what we have found out for the normal hairpins. Diffusion coefficient of abasic samples again increased at higher salt concentration similar to what we observed with normal hairpins (Fig.3.17). Dye brightness of HP3X showed the same

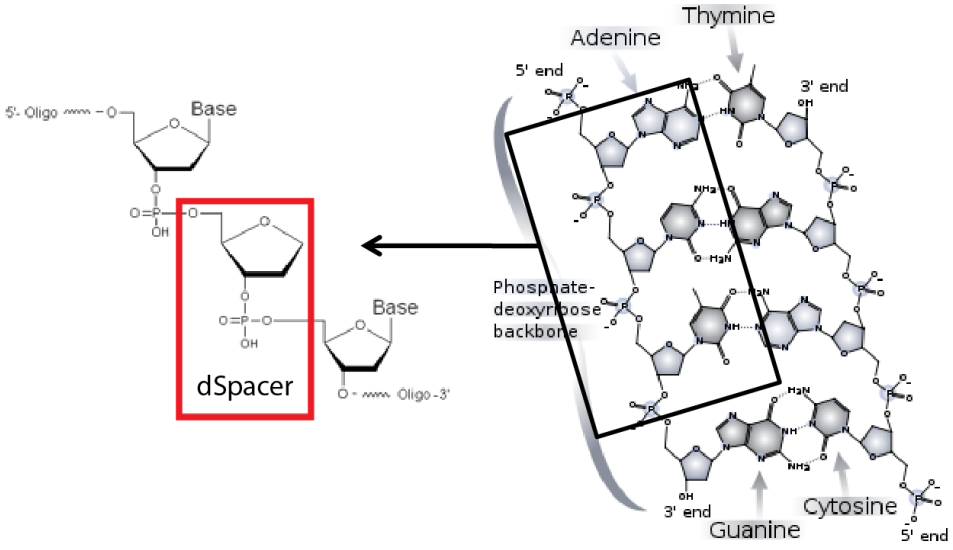

Figure 3.16: Illustration of the chemical structure of dSpacer, or abasic furan (red box). A normal double helix DNA is shown on the right (photo from www.wikipedia.com). dSpacer lacks the bases and is only the sugar-phosphate DNA backbone. Our abasic sequences have dSpacer as the loop and normal bases as the stem. This minimizes the stem-loop interactions but still preserves the stem-stem interactions.

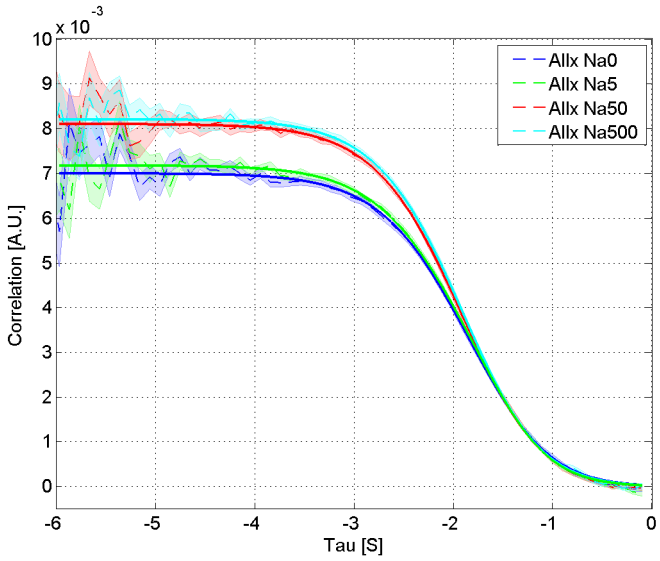

Figure 3.17: AllX FCS at different salt concentrations. The elevated amplitude at higher salt concentration is the result of increased diffusion coefficients.

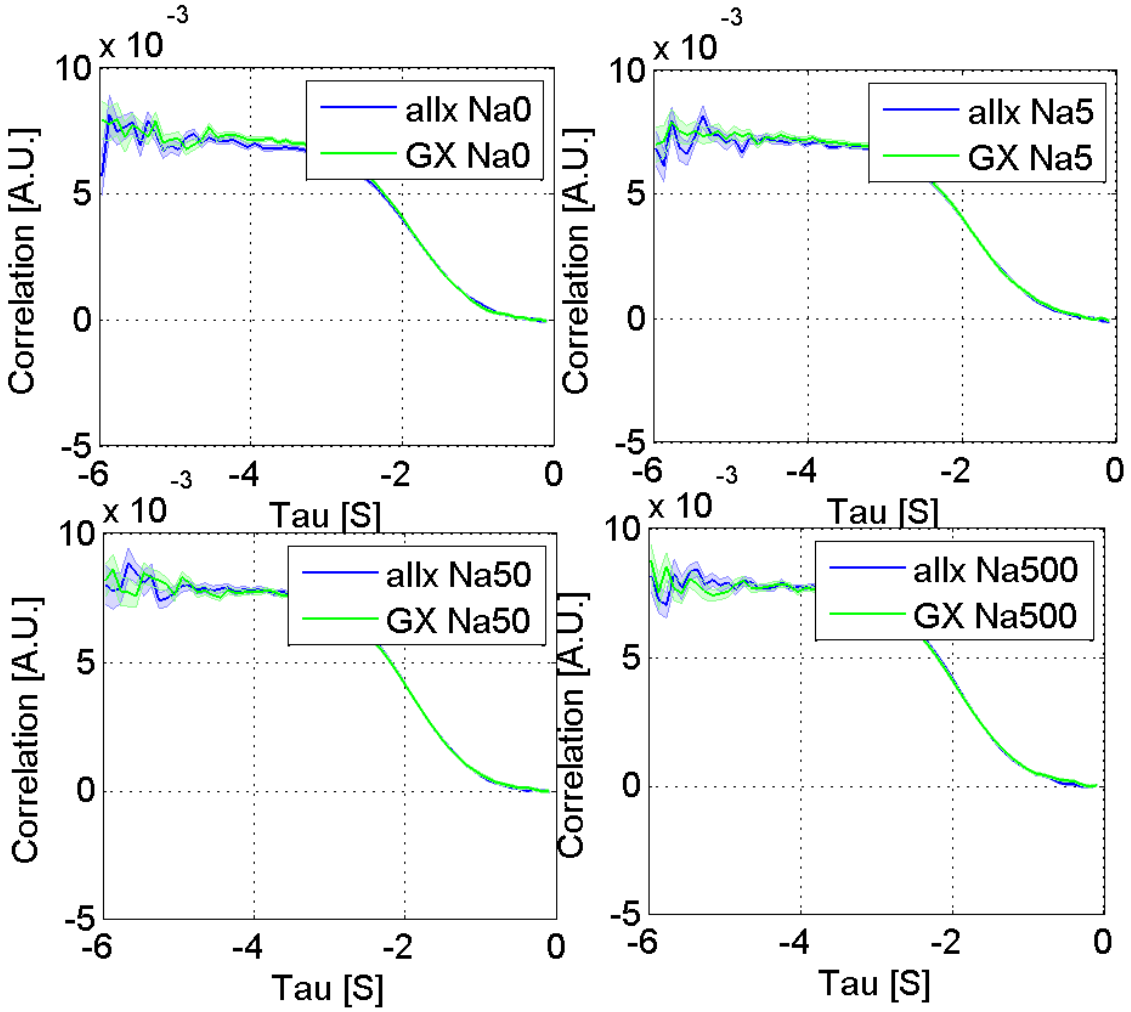

Figure 3.18: FCS signals of ALLX vs GX at four different salt concentrations. Notice that both signals are within error bounds of one another. This is very similar to what we have seen in Fig.3.15.

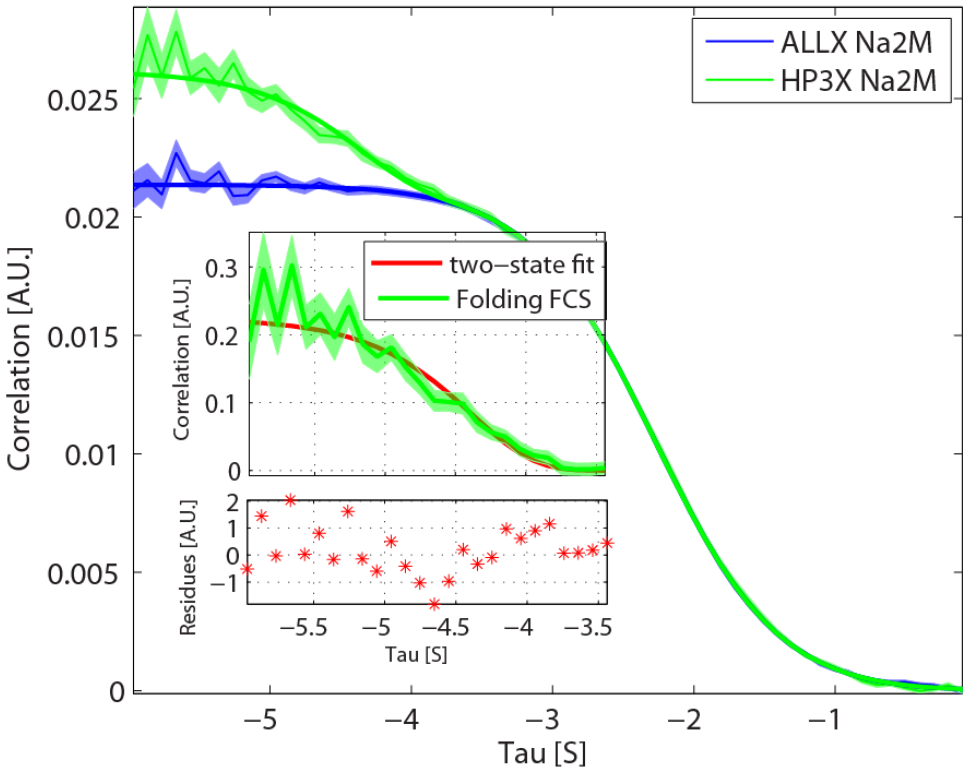

Figure 3.19: FCS signals of HP3X vs ALLX at 2M NaCl. The signals are aligned by minimizing sum of the square differences of these signals from $10^{-3.5}S$ to 1 S. Insert: the extracted folding FCS signals of HP3X using ALLX as control. Green curves are data and the red curve is a fit to a two-state folding model. The extracted parameters are listed in Table.3.2.

	HP3 Na500	HP3 Na1M	HP3 Na2M		HP3x Na500	HP3x Na1M	HP3x Na2M
Amp	0.29	0.25	0.26		0.23	0.19	0.22
δAmp	0.04	0.04	0.08		0.03	0.02	0.02
$k(\times 10^3 S^{-1})$	13	14	16		16	21	24
$\delta k(\times 10^3 S^{-1})$	3	3	7		3	4	5
$\bar{Q}(kHz)$	0.97	1.02	0.89		1.04	1.05	0.96
$k_{lh}(\times 10^3 S^{-1})$	3.0	3.4	2.6		6.2	7.7	6.2
$k_{hl}(\times 10^3 S^{-1})$	10.0	10.6	13.4		9.8	13.3	17.8
$Q_h(kHz)$	1.92	1.92	1.92		1.67	1.67	1.67
$Q_l(kHz)$	0.68	0.73	0.69		0.65	0.71	0.67

Table 3.2: Folding parameters of normal and abasic hairpin at high salt concentrations.

decrease as HP3 as a function of salt concentration. ALLX and GX did not show FCS relaxation other than diffusion, and their FCS signals are statistically identical(See Fig.3.18). The FCS signal of HP3X had very clear and distinguished 100uS relaxation that can be described by a two-state process(Fig.3.19). The extracted rate constants are within one standard deviation of the rate constants extracted from HP3 FCS measurements(See Table.3.2). Both qualitatively and quantitatively, abasic hairpins are similar to normal hairpins.

Thus our results with the normal and abasic three-basepair stem hairpin suggest that the three-basepair stem is both sufficient and necessary in observing the same 100uS FCS relaxation. The striking similarity between the results in these experiments strongly suggests that misfolded states do not present significantly in the state population to affect the hairpin folding pathway.

3.9 Results on FRET labeled DNA Hairpin Did not Find Significant Presence of Semifolded States

To see whether it is the whole chain or part of the stem opening and closing caused the observed two-state kinetics, we performed FCS measurements on Cy3-Cy5 FRET labeled HP3. The FRET pair has a Forster Radius of 5.4 nm. Comparing this value with the contour length of the chain (7nm), this FRET pair is sensitive to the long distance changes in the hairpin folding process. If it is the global change of the conformation caused the dynamics in three-basepair stem hairpins, we should be able to observe similar dynamics from a FRET pair. On the other hand, if what we observed is really due to local conformational changes, such as semifolded state, then we will see FCS signals that have very different functional forms and parameters. Because the efficiency of the FRET process is a continuous function of the distance between the donor and acceptor, the extraction of stem formation/detachment procedures is very different from the dye quencher case. There are three stochastic processes contributed to the FRET FCS signal: center-of-mass diffusion, chain conformation fluctuation, and stem formation/detachment. Thus the folding FCS from the FRET hairpin is:

$$g_2(\tau) = \frac{1}{N} \cdot [g_{chain}(\tau) + 1] \cdot [g_{folding}(\tau) + 1] \cdot Diff(\tau)$$

We do not really know the exact functional form of $g_{chain}(\tau)$, but we can directly measure it by a FRET labeled ALLT strand. Once we have the chain FCS and the overall FCS from the hairpin, we can peek into the folding FCS.

Similar to our previous measurements, HP3-FRET and ALLT-Cy3 showed indistinguishable FCS curves (Fig.3.20 A), diffusion coefficients and FRET efficiencies (Fig.3.22) in solutions with less than 50mM NaCl. Together with dye quencher data, this indicates that the hairpin in less than 50 mM NaCl solutions does not undergo

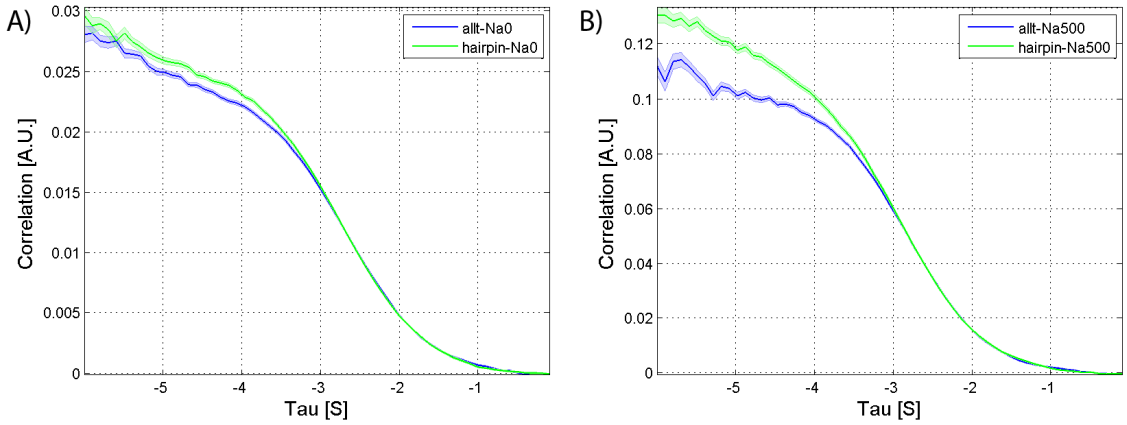

Figure 3.20: FCS of FRET labeled samples at zero and 500 mM of NaCl, donor channel. Samples are 21nt ssDNA attached to a 26 bp dsDNA. The blue curve is ALLT, and the green curve is hairpin. A) Solution is zero mM NaCl. B) Solution is 500 mM NaCl.

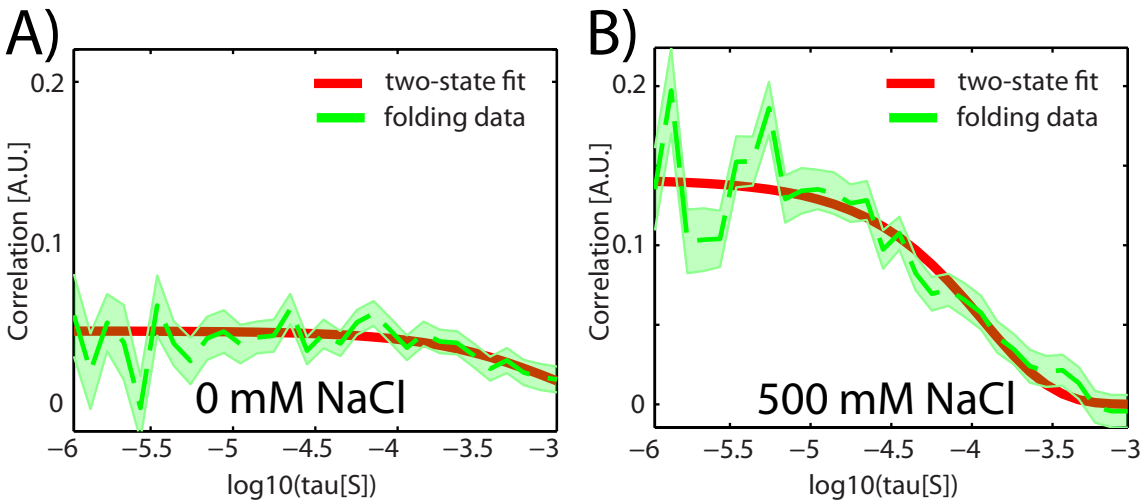

Figure 3.21: Hairpin dynamics probed by FRET-HP3 and FRET-ALLT as control. Green curves are the extracted folding dynamics (normalized at 10^{-2} S.) The fitting model is a two-state model (Red). A) Extracted folding dynamics at 0 mM NaCl. B) Extracted folding dynamics at 500 mM NaCl. Residue analysis cannot reject the two-state model. Fitting parameters: $g_2(0) = 0.14 \pm 0.01, k = (8.4 \pm 1.8) \cdot 10^3 S^{-1}$. Folding rate constants: $k_f = (5.6 \pm 1.2) \times 10^3 S^{-1}$, $k_{uf} = (2.8 \pm 0.6) \times 10^3 S^{-1}$.

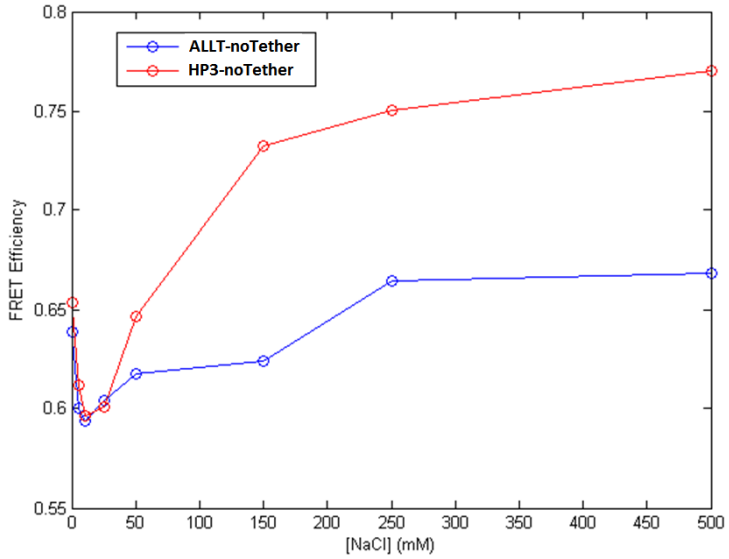

Figure 3.22: FRET efficiencies of ALLT-noTether and HP3-noTether vs salt concentration. The initial overlapping of the FRET efficiency of these two samples is one indication that at low salt concentrations, the hairpin resembles a random coil.

stem formation and detachment. In NaCl concentrations higher than 50mM, both donor channel and acceptor channel showed additional FCS relaxation than the diffusion induced relaxation (Fig.3.20 B). The extracted stem formation/detachment contributed FCS again can be satisfactorily fit by a two-state model, and the rate constants are within a factor of two of the dye-quencher data (Fig.3.21). Since the physics of FRET system does not allow our FRET pair to distinguish the semifolded state from the folded state, the two states in stem formation/detachment must involve distance changes comparable to the Forster Radius of the FRET pair (5.4nm). This observation, together with the fact that the measured rate constants from FRET and dye-quencher system yield similar values, suggests that the hairpin folding is mainly transitions between a fully folded hairpin and a fully open hairpin.

3.10 Surface Measurement and Calculations Verified that the Whole Folding Reaction Has Been Probed

Previous literature studies established the fact that hairpin folding is complex: it can involve many conformational states that are stable at different time scales, from microsecond end-to-end collision event to milliseconds lived folded hairpin. Because complexity and the fact we can only observe limited temporal range by our FCS experiment, it is necessary to check thoroughly on whether we have measured the whole reaction or not. Thus we performed calculations and additional experiment. All results in this section indicate that our FCS experiments captured the whole folding reaction.

First, we tethered our molecules on surfaces and measured the folding dynamics to check if there is any slow folding dynamics. This compliments the FCS results by extending the upper bound of the temporal range to more than 30 seconds. Using a diffraction limited laser beam, we individually probed dye-quencher labeled ALLT and HP3 at 0 and 500 mM of NaCl. (Fig.3.23). Measurements of surface-confined ALLT revealed that Atto700 blinked in our solution conditions (Fig.3.23A, blue trace). The time scale of Atto700's blinking is at least one order of magnitude slower than diffusion time scale, and is therefore not captured by our FCS measurements. Measurements of surface-confined HP3 also showed the blinking(Fig.3.23A, red trace). Since the transition rates of HP3 and ALLT form one cluster (Fig.3.23B), we attribute this blinking to the photophysics of Atto700 and not to hairpin folding. The brightness levels of the on-states of ALLT and HP3, however, are very different. ALLT is about two times brighter than HP3(Fig.3.23A), which agrees with our brightness measurements on freely-diffusing hairpins (vide supra, Fig.3.14D). Using the fluorescence signal of the on-state, we calculated the surface FCS of ALLT and HP3. ALLT gave a flat FCS curve(Fig.3.23C), consistent with what we saw in its solution data. HP3 revealed dynamics that is in the same time scales as our FCS measurements(Fig.3.23D). There is no slower dynamics as compared to what we have

observed in the solution. Unless the folded hairpin is completely dark and stable for more than 30 seconds (which is very unlikely given its energy levels), the hairpin dynamics we observed in the solution FCS constitutes the full folding reaction of the hairpin.

Second, we performed calculations directly on solution data to see if we have probe the whole folding reaction. Surface measurements could be different from the solution FCS measurements as the local micro-environment of the hairpin has changed. Thus it is the best if we could have direct evidence from the solution data to see if we actually have observed the full folding reaction. For this purpose, we calculated the total number of molecules in the solution using the observed FCS: if what we observed is the full folding reaction, even though the fractions of the population of the observed states might change, the total number of molecules from all the observed states should be constant. If there is a long lived dark state of the hairpin, then the total number of observed molecules in the solution should decrease as the NaCl concentration is increased. What we observed as shown in Fig.3.24, however, is that the total number of observed molecules is the same, within error bounds, for the full range of salt concentration (0 to 2M of NaCl). This is one direct confirmation from solution data that we have observed the full folding reaction of the hairpins.

The significance of the surface data not only proves that we have observed the whole folding reaction of our hairpins, at the same time, the consistency of the surface data and solution data suggests that the functional surface did not interfere with the molecular folding.

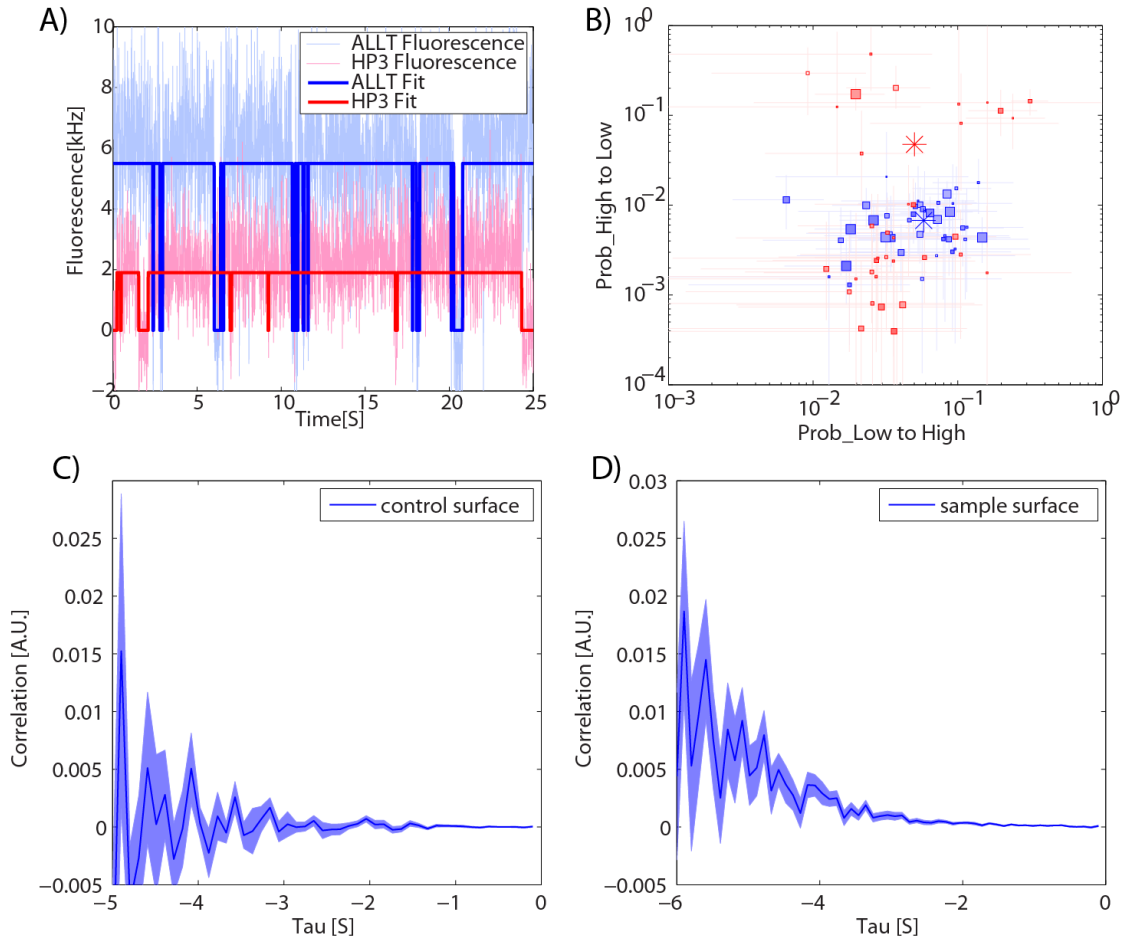

Figure 3.23: Probing hairpin folding dynamics by attaching the hairpin on the surface. A) gives examples of the fluorescence traces of ALLT(blue) and HP3 (red) hairpin fluorescence traces in 500mM NaCl. The solid curves are filtering results, assuming Possion emitters and a two-state model. Both traces showed blinking. Note that the ALLT traces is about two time brighter than HP3, which is consistent with our solution FCS measurement. B) the inferred transition probabilities of ALLT and HP3. The overlapping of the transition probabilities (length of the cross is the 95% confidence interval of the filtering) suggests that the blinking is independent of the molecular folding - one evidence supporting that this long time blinking is from the photophysics of the dye. C) FCS signal of ALLT obtained from the high state of the blinking. D) FCS signal of HP3 obtained from the high state of the blinking.

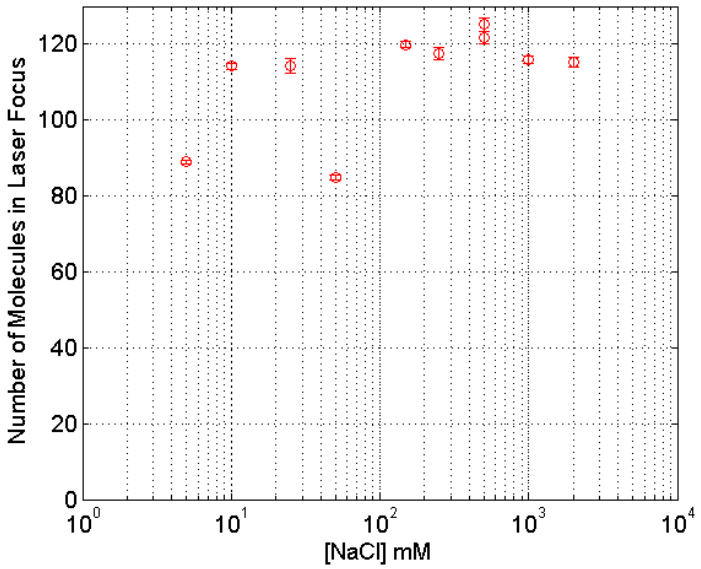

Figure 3.24: The number of molecules in the laser focus in FCS measurement vs salt concentration. While the molecules stay more fraction of time in the folded and dimmer state as salt concentration is increased, the total number of molecules stayed constant. This is one direct confirmation that the folding of the hairpin does not involve with a total dark, population sink state, which could change the total number of molecules in the reaction.

3.11 DNA Hairpin Folding Conclusion

Leveraging on the previous literature findings, we constructed a very general model of the hairpin folding (as shown in Fig 3.1.). It constitutes basic states such as the flexible coil and the fully folded hairpin, and complex states such as the misfolded and semifolded hairpin. We isolated these branches by engineering DNA sequences that favor certain states and measured their folding dynamics. The data we obtained by these experiments, while is complex, collectively suggests that the hairpin folding generally only involves two states: the open, random coil state and the closed, stem fully formed hairpin state. Other previously proposed states, such as the misfolded state and semifolded states, do not present significantly in the folding process in our sequence. This conclusion is supported by independent measurements using different sequences, by probing systems with complementary sensitivities, and in a wide range of salt concentrations.

In addition, our findings suggest that hairpin folding kinetics as well as its center-of-mass diffusion highly depends on the salt concentration. In salt concentrations less than 50 mM, the hairpin most of the time ($> 99\%$) is a flexible chain. Once the salt concentration is more than 50 mM, the stem starts to form, and the molecule diffuses faster. As salt concentration is further increased, the folded hairpin becomes the dominant state, and the diffusion coefficient can be increased by as much as 50%. This trend stabilizes until the salt concentration reaches 500 mM. Cation concentration controls how much hairpin folds and how fast it diffuses (Fig.3.25).

The end-to-end collision event is the precursor of stem formation, and now we understand how salt affects it. Although we did not directly observe the end-to-end collision events by our data, we inferred it by diffusion coefficient measurements and chain conformational fluctuation measurements. Both measurements suggest that end-to-end collision rate of the chain is relatively indifferent to the cation concentration once it is more than 5 mM. The reason we see increased folded hairpin at high salt concentration is in fact due to increased reaction probability per collision. This fascinating finding further illustrates that the cation affects the folding of the hairpin in many ways.

An important extension of this work is to experimentally check what we learned from our three-bp hairpin applicable to other sequences with different loop and stem compositions. We think while some findings in the current study can be generally applied to other sequences, we do not anticipate everything we learned here can be directly transferred to all sequences. As far as what states are involved, we expect that for sequences with long stems ($> 5bp$) and similar loop/stem length ratio, we can see semi-folded states at low salt concentrations. This is because that we already see end-fraying directly from the dsDNA tether at low salt concentrations in our study. At high concentrations (¿ 50mM), however, since our data suggests that hairpin folding is highly corporative, we would still see two-state folding for these long stem sequences. This is supported by findings from experiments by [WBPL+06]l, and [GGL+01] using optical trapping and TIR FRET studies respectively. We think what is important is that now we have established a systematic way to examine hairpin folding. Many important questions, such as how a poly-admine chain would fold compared to a poly-thymine chain, can all be learned in this way.

We need to further investigate the coupling of diffusion and folding. We hypothesize that the phenomenon of folding and diffusion coupling could give rise to artifacts, which had led to over-interpretation of the data in the literature. A scan of current studies reveals that studies that used freely-diffusing hairpins reported non-twostate folding of the chain, while studies using surfaces or optical trapping reported two-state folding (See Table 2.1). This trend can be the result of the above-mentioned diffusion-folding coupling.

Finally, the tools and analysis used in this study can be generally applied to other folding investigations. For example, by simply expanding the laser waist, we would be capable of expanding the upper temporal range of FCS by nearly 100 folds. The simplicity of FCS with increased temporal range can make a real difference in our abilities to probe complex macromolecular folding.

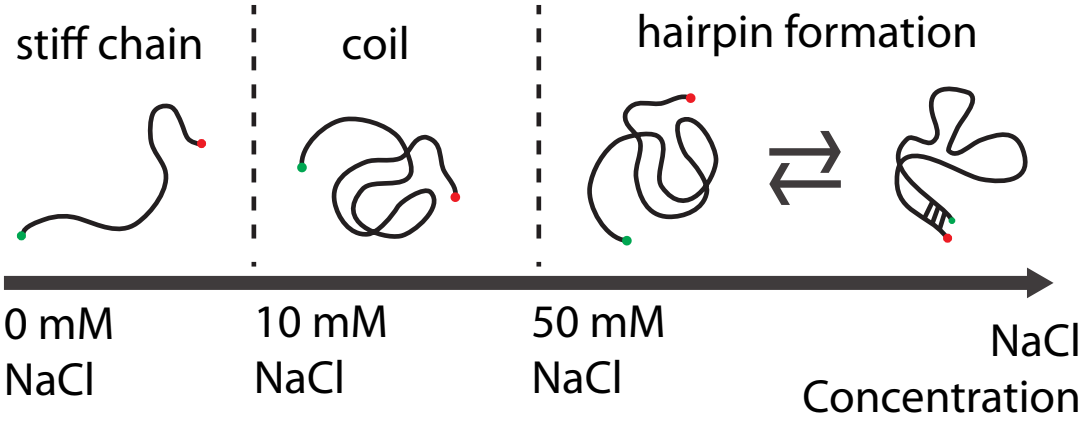

Figure 3.25: Global and local conformation fluctuations of the hairpin at different salt concentrations. Globally, the hairpin adopts different conformational forms at different salt concentrations. In NaCl concentrations < 10 mM, the hairpin is more extended due to electrical repulsion from backbone charges. In NaCl concentrations > 10 mM, the hairpin adopts a more condensed form and diffuses faster. Locally, it requires at least 50 mM NaCl for the hairpin to form its stem. The folding rate constant strongly depends on the cation concentration due to the fact that the ends of the hairpins need to brought into contact to form hydrogen bonds in the stem. Our data indicates that the formation of the stem and deformation of the stem is a two-state process.

4

Tracking and Monitoring Molecular Folding Reaction

In this chapter of the thesis, we present our results of tracking a molecular complex while observing its folding. The molecule under investigation is still the DNA hairpin, the folding of which has been thoroughly investigated in the previous chapter. With this information at hand, it is convenient for us to investigate the potential of tracking-FRET in its ability of studying folding. After introducing the basic tracking experiment, we will present basic theories and data analysis procedures associated with tracking-FRET. We will then present our main results. It will be clear at the end of this chapter, that tracking FRET, while holds much potential as a method to investigate macromolecular folding in general, still needs much work in both theory and experiment.

4.1 Introduction

In this section, we will introduce the tracking microscope. Since the development of the tracking microscope was not my focus, I will not cover all the details. Rather, I will only cover some basic concepts so that the reader could be equipped with information needed to understand tracking FRET. For a complete and detailed description of the tracking apparatus, please refer to Andy Berglund and Kevin McHale's Ph.D

theses[Ber06, McH08]. Andy first developed basic theory and built the the tracking apparatus in 2D[BMM07]. Kevin expanded both theory and experiment of the tracking into three dimensional and performed tracking experiments to study the anti-bunching properties of quantum dots[MBM07] and the intramolecular folding dynamics of λ-phage DNA while they freely diffused in the solution[MM09, MM09]. Both theses had extensive description and calculations on the tracking microscope, such as the position sensing techniques and tracking system dynamics. These calculations pave the foundations to understand the statistics of tracking error, which is manifested in the statistics of the fluorescence signal (in the simplest case). Since fluorescence signal is also being used to monitor folding dynamics, it is important to factor this tracking contributed part out. This is one of the many reasons why major portions of Andy and Kevin's theses were devoted to understand and characterize the statistic of tracking errors.

4.1.1 Tracking Basics and Setup

It can be very simple to summarize how fluorescence tracking microscope works: sense where the particle is and follow as close as possible to that particle. This description, although not very scientific, covers the essence of the tracking system.

The first step is to sense where the particle is. A simple Gaussian beam in the solution could tell how far the particle is away from the center of the beam by the magnitude of the fluorescence rate but could not tell exactly the x and y positions because of the symmetric shape of the beam. To get the x, and y positions of the particle (relative to the center of the microscope objective), we need to break this symmetry. Any non-symmetric scanning of the laser could break this symmetry and could allow the encoding and later extraction of the position information of this particle, but we chose to rotate the laser beam in the xy plane, as shown in Fig.4.1 A). In this specific configuration, the fluorescence rate detected is:

$$f(\bar{e},t) = f_o \cdot Exp(-\frac{2\left|\bar{e}\right|^2}{\omega^2})$$

where $\bar{e}$ is the difference between the position of the laser beam and the position of

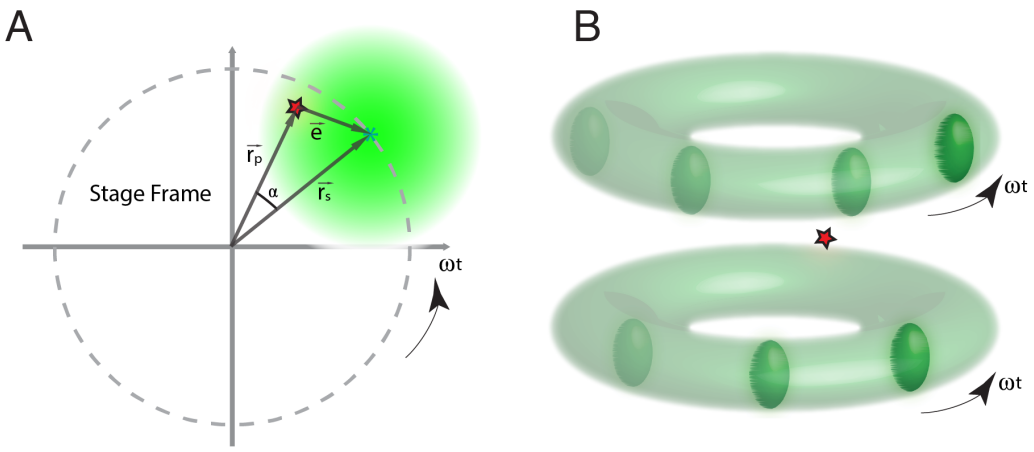

Figure 4.1: The tracking laser scan patterns in 2D(A) and 3D(B). Note the patterns are not drawn in scale to illustrate the general symmetry breaking idea to sense the position of the particle, which is represented by the red star in the figure.

the particle, f_o is defined as the fluorescence rate at the center of the beam, and ω is the radius of the beam. Since the angle $\alpha = \phi_p - \omega_{xy}t$, we have:

$$f(\bar{e},t) = f_o \cdot Exp[-\frac{2(\rho^2 + r^2 - 2\rho r\cos(\phi_p - \omega_{xy}t))}{\omega^2}]$$

where ρ is the length of the particle position vector, and r is the radius of the rotation laser beam, and ω_{xy} is the rotational frequency of the laser. Thus the fluorescence signal encodes the position information of the particle. The last step is to decode this information by lock-in detection: we modulated the fluorescence signal at the rotational frequency ω_{xy}, and we only look at the frequency component at that specific frequency. In reality, this calculation is done by multiplying the fluorescence signal with a reference signal and integrate for a certain periods of time. Any signal component that is not in the same frequency of the reference signal will be attenuated to close to zero. Mathematically, we decode by:

$$v(B) \propto B\int_0^{1/B} dte^{-i\omega_{xy}t} f(\bar{e},t) \propto \rho e^{-i\phi_p}$$

where B is the bandwidth of the lockin detection, and $v(B)$ is the output of the lockin amplifier. Thus we can see that the output of the lockin amplifier is directly proportional to the difference between the center of the microscope objective and the position of the particle. In turns of tracking, the objective essentially is to have the reminding of the feedback loop, namely the feedback controller and sample stages, keep the both x and y component of $V(B)$ close to zero.

The three-dimensional signal modulation and demodulation scheme is very similar to the 2D case(See Fig.4.1 B)). We used another laser beam whose focus is separated by a few μm, and switched it on and off at a different frequency. The exact frequency value of the rotation or switching does not matter as long as they are well separated and not integral multiples of each other.

A couple of notes before we going into more details of the tracking:

- The photon emission process by the dyes are discrete stochastic events, with each arrival of the photon generating a TTL pulse in our detector. The calculations above assumed a continuous process. But it can be shown, under a few very reasonable conditions, that there is no signal loss if one directly feed the TTL pulses into the lockin amplifier. For details of this, please read Andy or Kevin's thesis.

- Generally, the signal to noise ratio of the localization signal is proportional to $\sqrt{f_o/B}$ (for all three axes). This means that to have good signal-to-noise ratio, we need to have set the integration time of the lockin amplifier to be long. However, if we average it for long time, the particle allows the particle to diffuse around, adding additional uncertainty to the position estimation. We can also see that the tracking error is a battle (ratio) between how much fluorescence we can collect and how fast we can set the tracking bandwidth. Ideally, we would want to have infinitely bright sample with infinite infinitesimal integration time. We do not have that, and it means we will always have tracking error associated with the fluorescence tracking microscope. This in turn means that the particle, although being tracked in the stage frame, but is undergoing random motion inside the rotational beams. The beam modulations, together with the tracking

error, all contributes to additional the correlational functions of the fluorescence signal, which should be take into account when folding dynamics is probed.

- There is a mis-conception in the tracking/trapping community, that one needs to feedback upon every photon received. This simply is not true, because the stochastic nature of the emission process and the noisy background - doing so will only feed noise in the system and will not have any benefit.

- Align the two beams together is NOT fun. It can be constantly mis-aligned and requires a lot of trial-and-error. Charles and I perhaps have spent months just performing the alignment. We have improved the techniques over time, but it is not fun when all biological samples were ready and then we discovered the setup needed to be re-aligned. This prevents us working on the biological systems we really want to study. In my opinion, we need to get rid of this two-beam setup to really unleash the potential of tracking.

Fig.4.2 shows the overall design of the tracking apparatus. It can be divided into three groups. First group is the beam modulation optics and electronics. This group is responsible for creating the rotating beams and intensity stabilization and feedback. Specifically, we used 532nm as our tracking laser. Acoustic Optic Modulators (AOMs) are used to deflect the beam by modulating the RF frequency driving the AOM. AOM Y deflect the beam along the y axis, where as AOM X1 and AOM X2 deflected beam at a 90 degree phase delay to complete the beam rotation at 95 kHz. The AOM X1 and X2 were also alternatively turned on and off at 100 kHz, which is the z modulation of the beams. We also modulated the RF signal amplitude in AOM Y to control the power going through the first order deflection beam. This controls the total power going to the molecule and is used in our fluorescence feedback system. We also aligned a 634nm probe beam at the fixed point of the tracking feedback system. This is done by scanning the stage positions over a fluorescent bead on the glass surface while monitoring the collected fluorescence from the probe beam and compare it to the error signals to ensure good alignment. For example, in Fig.4.3,

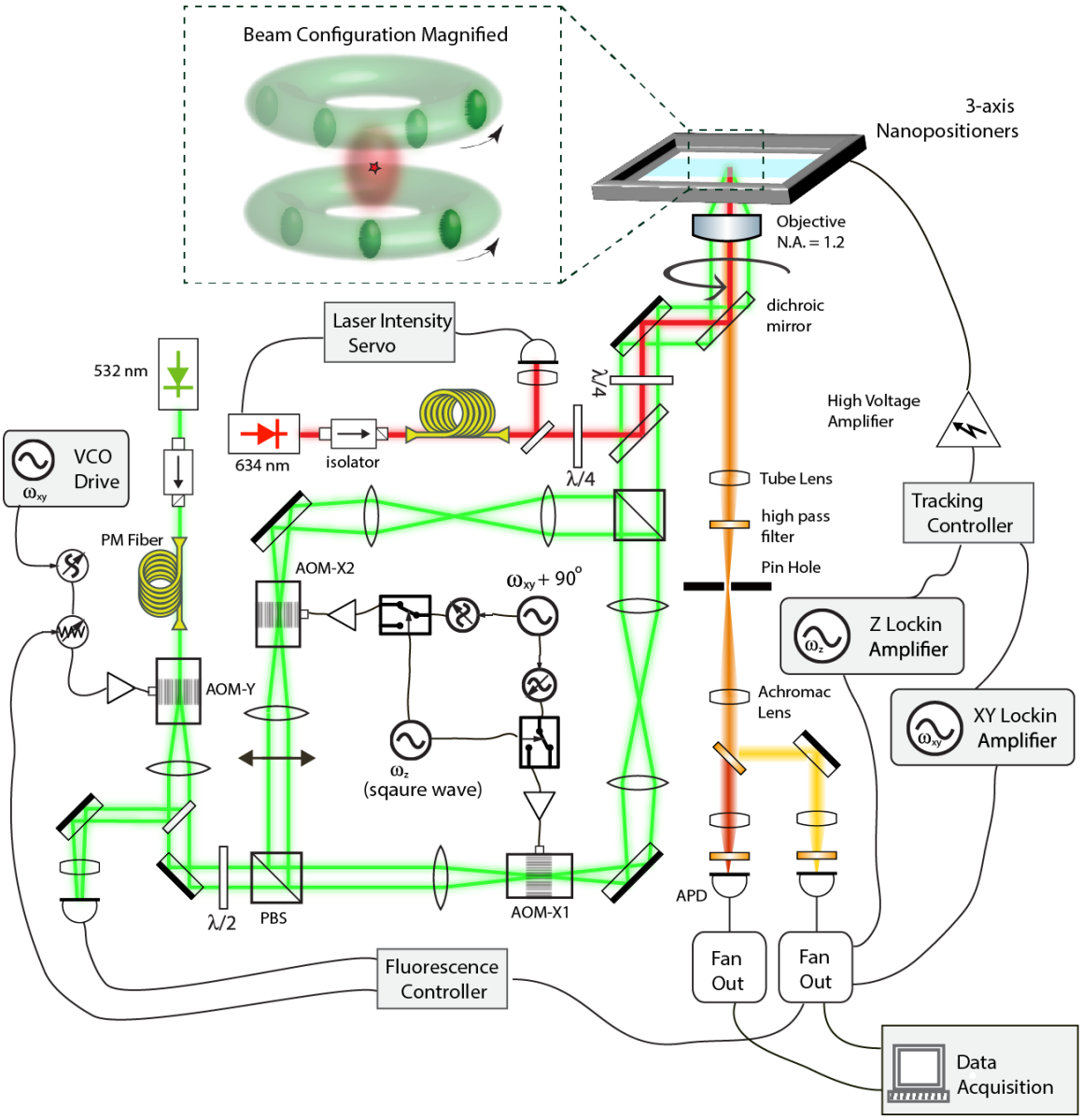

Figure 4.2: The schematic of the tracking setup (this figure is a remake from [McH08]). The apparatus can be divided into three groups. First group is the beam modulation optics and electronics. AOM-Y deflect the beam in the vertical direction, while as the AOM-X1 and AOM-X2 deflect the beam in the horizontal direction with 90 degree phase delay to complete the beam rotation. The second group is the fluorescence collection optics, which is essentially a con-focal microscope. The final block is the position feedback electronics. For more details, please see text.

when the error signal reaches zero, we can see that the fluorescence from the probe beam reaches its maximum. In this case, a Gaussian fit to the probe beam fluorescence generates a beam waist size of 0.65 μm.

The second group is the fluorescence collection optics, which is essentially a confocal setup with dual spectrum channels. The Zeiss $f = 160mm$ tube lens and the Zeiss C-Apochromat water immersion (with NA = 1.2) microscope objective work as a pair to create high fidelity optics that is crucial to collect a large spectrum of the light. At the imaging plane of the tube lens, we aligned a 300 μm pinhole to reject out of focus light. Then the fluorescence signal is separated by a Chroma dichroic filter (Chroma 625 DCXR), filtered by Chroma ET series bandpass filters, and finally focused to the active area of Avalanche Photon Detector (APDs. Model: Perkin Elmer SPCM-AQR 13/15) by a Edmund Optics achromat.

The final block is the position feedback electronics. The TTL signals from the APDs are fed into the lockin amplifiers, whose output is the input signals for the stage feedback controllers. Finally, the controllers drive the stage high voltage amplifier, which drives the piezo stage to complete the feedback loop.

Fig.4.3 shows one alignment result we have for the red probe laser. This alignment is done by scanning the stage positions over a fluorescent bead on the glass surface while monitoring the collected fluorescence from the probe beam and compare it to the error signals to ensure good alignment. When the error signal reaches zero(black trace), we can see that the fluorescence from the probe beam(red trace) reaches its maximum. In this case, a Gaussian fit to the probe beam fluorescence generates a beam waist size of 0.65 μm.

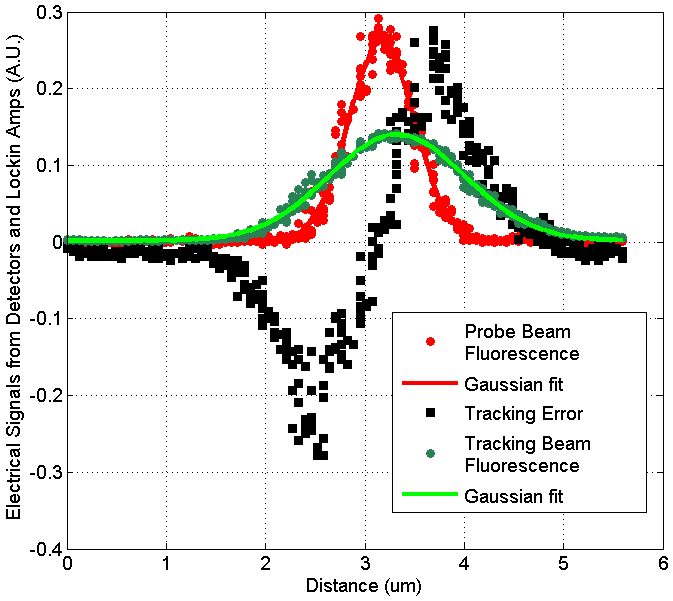

Figure 4.3: We aligned a 634nm probe beam at the fixed point of the tracking feedback system. This alignment is done by scanning the stage positions over a fluorescent bead on the glass surface while monitoring the collected fluorescence from the probe beam and compare it to the error signals to ensure good alignment. When the error signal reaches zero, we can see that the fluorescence from the probe beam reaches its maximum. In this case, a Gaussian fit to the probe beam fluorescence generates a beam waist size of 0.65 μm.

4.2 Theory Related to Tracking FRET FCS

In this section of the thesis, we will present some basic theory that would later guide us in the experiment, help us in understanding the statistics of the fluorescence signal, and allow us to extract the folding dynamics out of the complex data we obtained. We will first calculate how tracking errors and our beam modulation scheme contribute to the statistics of the collected fluorescence signal. Next, we will focus on the effect of multiple FRET pairs on one tracked molecule as well as the effects of imperfect labeling and crosstalk among the channels. The majorities of the calculations will be focused on calculating the theoretical form of the auto- and cross-correlations of the fluorescence signals.

4.2.1 Tracking FCS

The tracking errors as well as the beam modulation scheme render non-uniform excitation intensity over both time and space for the tracked molecule, which is manifested, together with the folding reactions, in the fluorescence signal. To correctly extract the folding dynamics out of the statistics of the overall fluorescence signal, it is first necessary to understand how these two sources of dynamics come into play. This calculation here follows closely of Kevin's awesome work[MM10, McH08] with extensions.

Atto700 Quenching System Case - with a stationary Gaussian beam as probe beam.

To probe the folding dynamics of the hairpin by the Atto700 quenching system, we coupled a 634 nm beam at the fixed point of the tracking system. This red laser excites the Atto700 labeled on the hairpin, yielding fluorescence in the 705/50 channel. The green 532 nm tracking laser excites Atto532 coupled to the bead by biotin-NeutrAvidin binding, yielding fluorescence in the 585/40 channel. There was no observed crosstalk in the Atto dyes system - the green tracking laser could only excite Atto532, the red laser could only excite Atto700, and no crosstalk of these two channels due to their far spectral separation. This simplifies our calculation quite a

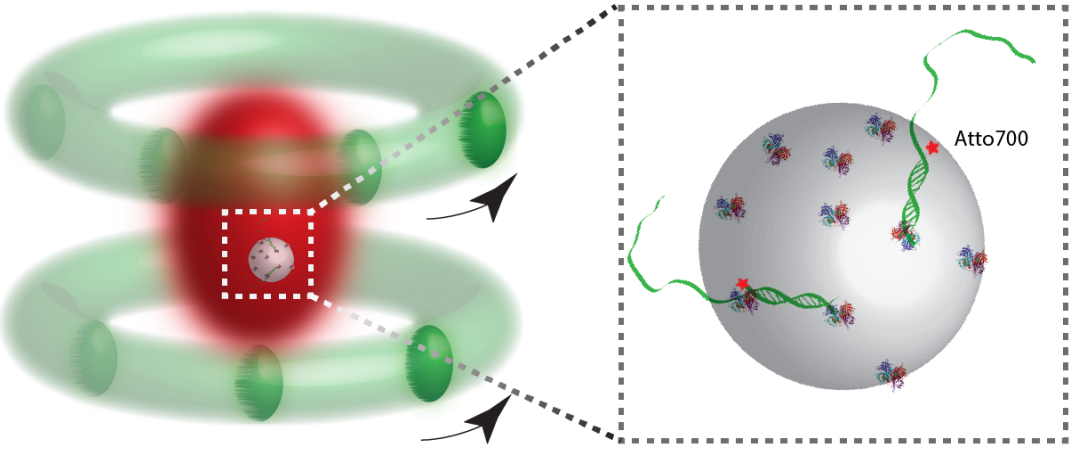

Figure 4.4: Tracking and probe beam illustration with bead hairpin complex. The two green rings represent the tracking beam, the red oval represents the probe 634 nm beam. The bead hairpin complex is tracked by the green laser while the folding dynamics is probed by exciting the Atto700 by the probe beam. Note that there can be multiple hairpins per complex. We deal with this situation in a later section.

bit, because this means that tracking is decoupled from the molecular folding. We only need to consider how the tracking error contributes to the auto-correlation functions of the Atto700 fluorescence. We will also treate the particles as a point particle, because the small relative size of the particle to the tracking beam geometries.

We start by the standard definition of FCS.

$$g_2(\tau) = \frac{< I(t)I(t+\tau) >}{< I(t) >^2} - 1$$

Here $I(t)$ is the signal from an Atto700 dye, which is given by:

$$I(t) = \int d^3\overrightarrow{x}\, b(\vec{x}, t)\phi(\vec{x} - \vec{y}, t) \tag{4.1}$$

Here $\vec{x} - \vec{y}$ is the distance between the point of integration to the center of the gaussian laser at $\vec{y}$. Note that the position of the laser beam $\vec{y}$ is a function of time as the laser beam is tracking the position of the particle. $b(\vec{x}, t)$ is the brightness of the

dye. $b(\vec{x},t)$ is a function of time for two reasons: the first reason is that the location of the molecule is a function of time as the partible is diffusing and thus experience variations in excitation intensity; the second reason is that the intrinsic brightness $B(t)$ of the dye can change because of folding:

$$b(\vec{x},t) = B(t)\delta(\vec{x} - \vec{x}_P^t) \tag{4.2}$$

It is a delta function: only at the position of the particle, the dye is bright. Combining Eq.4.1 and Eq.4.2, and applying Fourier transform to facilitate our calculation, we have:

$$I(t) = B(t)\int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}^T(\vec{x}_P^t - \vec{y}^t)}\tilde{\phi}(-\vec{k},t) \tag{4.3}$$

To calculate autocorrelation functions of the fluorescence signal, we have:

$$\begin{aligned} < I(t)I(t+\tau) > &= < B(t)\int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}^T(\vec{x}_P^t - \vec{y}^t)}\tilde{\phi}(-\vec{k},t) \\ &\quad \times B(t+\tau)\int \frac{d^3\vec{k'}}{(2\pi)^3} e^{i\vec{k'}^T(\vec{x}_P^{t+\tau} - \vec{y}^{t+\tau})}\tilde{\phi}(-\vec{k'},t+\tau) > \\ &= < B(t)B(t+\tau) >< Trk(t)Trk(t+\tau) > \end{aligned}$$

since the folding of the hairpin cannot change the tracking fluorescence or modulate the diffusion process, we can factor them as in the last step above. The folding dynamics will be encoded in $< B(t)B(t+\tau) >$, where as dynamics of the tracking will also present itself as a factor in the overall autocorrelation function of the fluorescence signal.

To calculate the tracking contributed autocorrelation, we first notice that we can factor out the laser profile out of the average. This is because that the laser profile is

deterministic:

$$< Trk(t)Trk(t+\tau) >= \int \frac{d^3\vec{k}}{(2\pi)^3} \int \frac{d^3\vec{k}'}{(2\pi)^3} < Exp[i\vec{k}(\vec{x}_P^t - \vec{y}^t)]Exp[i\vec{k}'(\vec{x}_P^{t+\tau} - \vec{y}^{t+\tau})] > \times \tilde{\phi}(-\vec{k}, t)\tilde{\phi}(-\vec{k}', t+\tau) \quad (4.4)$$

To continue, it is worth while to notice the following property. For a random variable that is Gaussian distributed with zero mean, we have

$$< Exp[ikx] >= Exp[-\frac{1}{2}k^2 < x^2 >]$$

Fortunately this is true for the random variable $\vec{x}_P - \vec{y}$, since over time we are closely tracking the particle. Thus Eq.4.4 reduces to:

$$\begin{aligned} &\langle Trk(t)Trk(t+\tau)\rangle \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3} \int \frac{d^3\vec{k}'}{(2\pi)^3} \\ &\times Exp[-\frac{1}{2}\begin{pmatrix} \vec{k} & \vec{k}' \end{pmatrix} \left\langle \begin{pmatrix} \vec{x}_P^t - \vec{y}_P^t \\ \vec{x}_P^{t+\tau} - \vec{y}_P^{t+\tau} \end{pmatrix} \begin{pmatrix} \vec{x}_P^t - \vec{y}_P^t & \vec{x}_P^{t+\tau} - \vec{y}_P^{t+\tau} \end{pmatrix} \right\rangle \begin{pmatrix} \vec{k} \\ \vec{k}' \end{pmatrix}] \\ &\times \tilde{\phi}(-\vec{k}, t)\tilde{\phi}(-\vec{k}', t+\tau) \end{aligned} \quad (4.5)$$

Thus the overall calculation simplifies to calculating the autocorrelation functions for the tracking error:

$$\begin{aligned} \sigma^\tau &= \langle e(t)e(t+\tau)\rangle \\ &= \left\langle (\vec{x}_P^t - \vec{y}_P^t)(\vec{x}_P^{t+\tau} - \vec{y}_P^{t+\tau}) \right\rangle \end{aligned}$$

We can already expect how the autocorrelation of the tracking error looks: at τ zero, it reduces to the average square distance of the particle and the laser beam, which certainly should increase for bigger diffusion coefficient of the particle, slower tracking system response, and greater noise level. At large delay time, we would expect that the correlation goes to zero, since the tracking error is Gaussian distributed, and more

importantly it no long term memory of the tracking errors.

The procedure to calculate the autocorrelation of the tracking error is conceptually simple, but computationally intense. It starts by first writing out the state space representation of our tracking system with the particle's random position $\vec{x_P}^t$ and noise as input[Ber06]:

$$\frac{d}{dt}\vec{q(t)} = \hat{A}\vec{q(t)} + \hat{B}\vec{u(t)}$$
$$\vec{y}(t) = \hat{C}\vec{q(t)}$$

where $\vec{q(t)}$ is the internal state vector, $\hat{A}$, $\hat{B}$, and $\hat{C}$ are state space representation of the tracking system. The exact expressions of these matrices depends on the transfer function of the plant and controller. Now $\vec{u(t)}$ is the input signal, which is the particle position estimation with a noise input:

$$\vec{u(t)} = \vec{x_P(t)} + \vec{F}(t) \tag{4.6}$$

where $\vec{F(t)}$ is a white noise term to take into account noises in the tracking feedback system. Then we solve the tracking system's position $\vec{y}^t$ as a function of time:

$$\vec{y}(t) = \hat{C} \cdot Exp[\hat{A}t]\vec{q_o} + \hat{C} \cdot Exp[\hat{A}t] \int_0^t d\tau Exp[-\hat{A}\tau]\hat{B}\vec{u(\tau)}$$

Now having system's response, we can calculate the autocorrelation of tracking error. This is one incredible complex calculation, and I thank Dr.McHale for calculating this general expression[MM10]:

$$\sigma^\tau = \hat{C}e^{\hat{A}\tau}[2D\hat{A}^{-1}\Gamma^\infty(\hat{A}^T)^{-1} + f^2\Gamma^\infty]C^T \tag{4.7}$$

where f is the power spectral density of the white noise term in Eq. 4.6. The exact expression certainly depends on how we model the tracking system response. Before we plug the exact expression of σ^τ to Eq.4.5, it is instructive to first write a simplified

expression for $< Trk(t)Trk(t+\tau) >$:

$$< Trk(t)Trk(t+\tau) >= \frac{a}{(2\pi)^3}\{det[\Sigma + \frac{1}{4}\begin{pmatrix} W & 0 \\ 0 & W \end{pmatrix}]\}^{-\frac{1}{2}}$$

where $a = (I_o\omega_x\omega_y\omega_z/8)^2$, ω_x is the probing laser waist in x axis, and I_o is the peak laser intensity in the center of the beam. And,

$$\Sigma = \begin{pmatrix} \sigma^0 Id_3 & \sigma^\tau Id_3 \\ \sigma^\tau Id_3 & \sigma^0 Id_3 \end{pmatrix}$$

$$W = \begin{pmatrix} \omega_x & 0 & 0 \\ 0 & \omega_y & 0 \\ 0 & 0 & \omega_z \end{pmatrix}$$

where σ^τ is a row vector defined as the autocorrelation function of tracking error given by Eq.4.7.

Finally, let's write out $< Trk(t)Trk(t+\tau) >$ explicitly:

$$\begin{aligned} < Trk(t)Trk(t+\tau) > &= \frac{a}{(2\pi)^3}\{\prod_{\rho\in\{x,y,z\}}[(\sigma_\rho^0 + \frac{1}{4}\omega_\rho^2)^2 - (\sigma_\rho^\tau)^2]\}^{-\frac{1}{2}} \\ &= \frac{a}{(2\pi)^3}\frac{1}{(\sigma_{xy}^0 + \frac{1}{4}\omega_{xy}^2)^2 - (\sigma_{xy}^\tau)^2}\frac{1}{\sqrt{(\sigma_z^0 + \frac{1}{4}\omega_z^2)^2 - (\sigma_z^\tau)^2}} \end{aligned}$$

where we assumed the identical tracking performance in x and y axis.

To work out the normalized FCS curve, we also need to calculate $< Trk(t) >^2$:

$$\begin{aligned} < Trk(t) >^2 &= \frac{a}{(2\pi)^3} \frac{1}{\prod\limits_{\rho \in \{x,y,z\}} [\sigma_\rho^0 + \frac{1}{4}\omega_\rho^2]} \\ &= \frac{a}{(2\pi)^3} \frac{1}{(\sigma_{xy}^0 + \frac{1}{4}\omega_{xy}^2)^2 (\sigma_z^0 + \frac{1}{4}\omega_z^2)} \end{aligned}$$

Therefore, we have:

$$\begin{aligned} g_{tFRET}(\tau) = & \quad \frac{< B(t)B(t+\tau) >}{< B(t) >^2} \frac{< Trk(t)Trk(t+\tau) >}{< Trk(t) >^2} - 1 \\ = & \quad [DNA(\tau) + 1][Tracking(\tau) + 1] - 1 \end{aligned} \tag{4.8}$$

with

$$Tracking(\tau) = \frac{(\sigma_{xy}^0 + \frac{1}{4}\omega_{xy}^2)^2}{(\sigma_{xy}^0 + \frac{1}{4}\omega_{xy}^2)^2 - (\sigma_{xy}^\tau)^2} \frac{\sigma_z^0 + \frac{1}{4}\omega_z^2}{\sqrt{(\sigma_z^0 + \frac{1}{4}\omega_z^2)^2 - (\sigma_z^\tau)^2}} \tag{4.9}$$

Depending on how we model the feedback loop, we will arrive at different expressions for $Tracking(\tau)$. Andy and Kevin had specified two different models for the tracking system. In both cases, we will use an simple integrator as the controller. For the first case, we model the plant as we can drive it instantly without any amplitude or phase roll off penalty. This will translate to:

$$\sigma^\tau = e^{-\gamma\tau}[\frac{D}{\gamma} + \frac{1}{2}f^2\gamma]$$

where γ is the entire gain of the feedback loop. The exact value of γ depends on the fluorescence rate, the gain of the lockin amplifier, the controller, and voltage to position conversation factor of the piezo stage. Note in this case, there exists an optimum gain of the whole system to minimize the average tracking error:$\gamma_s = \sqrt{2D/f^2}$.

We can also model the tracking system by adding a more realistic model to the plant (stage). Both Andy and Kevin have investigated when we use a low-pass filter

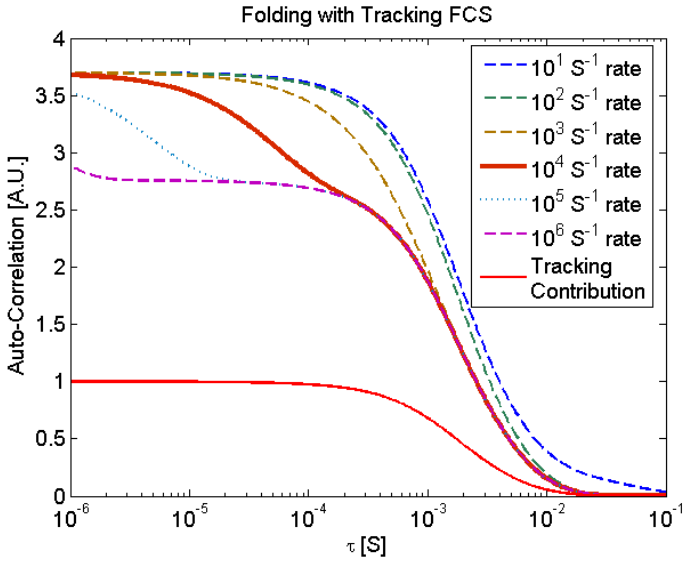

Figure 4.5: An example of tracking FCS with folding at various time scales. The tracking contributed FCS has the following parameters. $\gamma_c = 10\ Hz$, $\gamma_p = 200\ Hz$, $D = 5\ \mu m^2/S$, $\omega_{xy} = 0.65\ \mu m$, $\omega_z = 3\ \mu m$. We plotted a range of dynamics on top of the tracking FCS, showing that tracking FCS is capable of probing dynamics slower than the tracking bandwidth. This directly contrasts to the case of free solution FCS in Fig.3.2, in which dynamics slower than the typical diffusion time is not measured.

transfer function as a description of the stage response, and have shown that experimental data obtained can be very closely predicted by this theoretical understanding.

As one example, in Fig.4.5 we plot tracking FRET FCS with a range of dynamics. Modeling the tracking system by the second order system, we first plot the tracking contributed FCS (the detailed values of the parameters are in the caption). We can see that the tracking contributed FCS mostly decays around $10^{-3}\ S$, as it is limited by the limited bandwidth of the plant. Additional folding dynamics factors into the total FCS in the fashion of Eq.4.8. We plotted a range of dynamics on top of the tracking FCS, showing that tracking FCS is capable of probing dynamics slower than the tracking bandwidth. This directly contrasts to the case of free solution FCS in Fig.3.2, in which dynamics slower than the typical diffusion time is not measured.

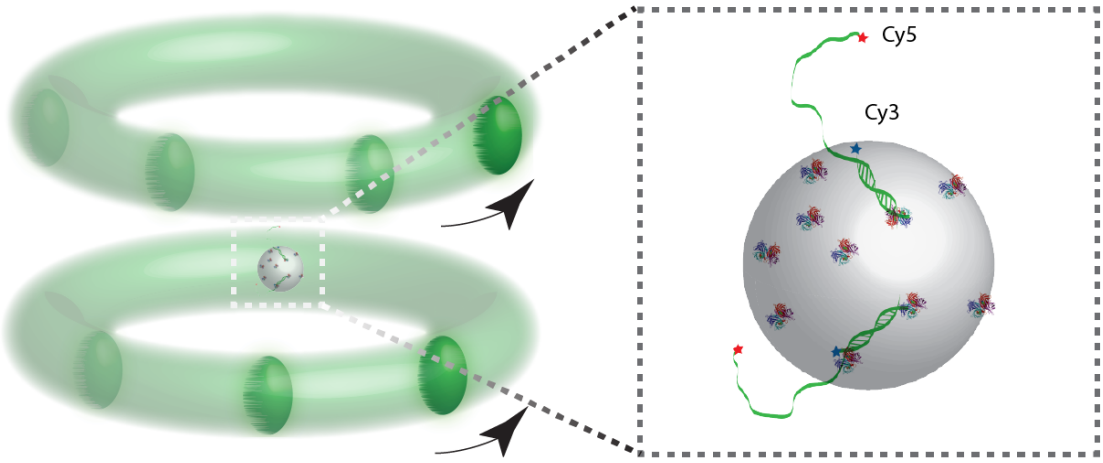

Figure 4.6: Tracking beam illustration with bead hairpin complex coated by Cy3Cy5 labeled DNA hairpin. The two green rings represent the tracking beam. Fluorescence of the FRET pair is recorded for folding dynamics. The sum of the donor and acceptor signals are also fed to the tracking system to provide spatial information for tracking. Note in this case, the tracking beam spatial temporal modulation needs to factor into the calculations of the autocorrelation functions of the fluorescence signal.

FRET case - use the tracking beam to probe the folding dynamics

In probing the Cy3-Cy5 FRET pair, we directly used the 532 nm tracking beam to excite the dye pair. The fluorescence of the donor and acceptor channels are recorded separately, and are summed together to drive the tracking system. In this special situation, not only the tracking error contributes to the variation in the signals, but also the beam modulation to probe the positions of the molecule will contributes to the statistics of the fluorescence we collected. The following calculation is to calculate the theoretical FCS curves when using tracking beams as probe beams. It follows the work done by Berglund[BMM07] and McHale.

The excitation profile of our tracking system is approximated by a pair of identical Gaussian beams, which are rotating at a radius r parallel on the xy plane with intercepts of z axis at $\pm z_0$. In this configuration, the fixed point of the tracking system is

at the origin. Now the spatially and temporally dependent fluorescence rate is:

$$\begin{aligned}\Gamma(\vec{e^t},t) &= \frac{\Gamma_0}{2}Exp[-\frac{2}{\omega^2}(e_x^t - rCos\omega_{xy}t)^2 - \frac{2}{\omega^2}(e_y^t - rSin\omega_{xy}t)^2] \\ &\times\{T(t)Exp[-\frac{2\zeta^2}{\omega^2}(e_z^t + z_0)^2] + T^{'}(t)Exp[-\frac{2\zeta^2}{\omega^2}(e_z^t - z_0)^2]\}\end{aligned}$$

where $\vec{e}$ is the tracking error vector, ω_{xy} is the angular rotational frequency, ω is the rotational laser's xy waist, ω/ζ is the rotational laser's z waist, $T(t)$ and $T^{'}(t)$ are two square waves that are 180 degree out of phase:

$$\begin{aligned}T(t) &= \frac{2}{\pi}\sum_{k=1}^{\infty}\frac{Sin[(2k-1)2\pi ft]}{2k-1} + \frac{1}{2} \\ T^{'}(t) &= \frac{2}{\pi}\sum_{k=1}^{\infty}\frac{Sin[(2k-1)2\pi f(t-1/2f)]}{2k-1} + \frac{1}{2}\end{aligned} \tag{4.10}$$

which turning on and off the laser beams to add modulation in the axial direction. Procedures in calculating for the autocorrelation function is again computationally intense, and not much can be gained from showing those messy calculations. Only results are given here:

$$\begin{aligned}SqTracking(\tau) &= \frac{\bar{\sigma}_0^4\bar{s}_0^2}{2[\bar{\sigma}_0^4 - (\sigma_x^\tau)^4]\sqrt{\bar{s}_0^4 - (\sigma_z^\tau)^4}}Exp[-r^2(\frac{\bar{\sigma}_0^2 - (\sigma_x^\tau)^2Cos\omega_{xy}\tau}{\bar{\sigma}_0^4 - (\sigma_x^\tau)^4}) + \frac{r^2}{\bar{\sigma}_0^2} + \frac{z_0^2}{\bar{s}_0^2}] \\ &\times\{(1 + \sum_{k=1}^{\infty}\frac{8Cos[2\pi f(2k-1)\tau]}{\pi^2(2k-1)^2})Exp[-\frac{z_0^2}{\bar{s}_0^2 + (\sigma_z^\tau)^2}] \\ &+(1 + \sum_{k=1}^{\infty}\frac{8Cos[2\pi f(2k-1)(\tau - \frac{1}{2f})]}{\pi^2(2k-1)^2})Exp[-\frac{z_0^2}{\bar{s}_0^2 - (\sigma_z^\tau)^2}]\} - 1\end{aligned} \tag{4.11}$$

with

$$\begin{aligned}\bar{\sigma}_\tau^2 &= (\sigma_x^\tau)^2 + \omega^2/4 \\ \bar{s}_\tau^2 &= (\sigma_z^\tau)^2 + \omega^2/4\zeta^2\end{aligned}$$

This result can directly be plugged into 4.8 to yield the total tFCS curves.

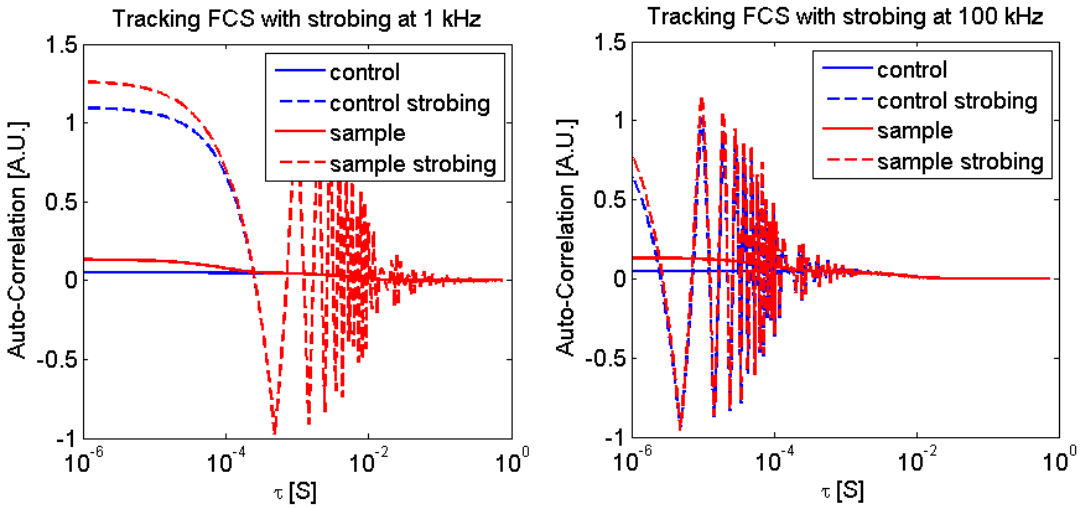

Figure 4.7: FCS oscillations due to tracking beam oscillation renders difficulties in visual comparison and data extraction of folding FCS difficult. In both panel, control is tracking only FCS with the following parameters. $\gamma_c = 10\ Hz$, $\gamma_p = 200\ Hz$, $D = 5\ \mu m^2/S$, $\omega_{xy} = 0.65\ \mu m$, $\omega_z = 3\ \mu m$. Sample is tracking FCS with two state folding dynamics with folding rate constant at $10^4\ S^{-1}$, and unfolding rate constant at $10^2\ S^{-1}$. The dashed lines are control and sample FCS with triangular waves. The diminish of the triangular wave at large τ is because of averaging within a bin.

In the experiment, we will see that the oscillations in Eq.4.11 can make the visual examination of tracking FCS curves and extraction of folding FCS difficult. This can be illustrated in Fig.4.7. We first plot two groups of FCS curves: tracking contributed FCS only (blue solid curve) and tracking plus folding dynamics FCS (Red solid curve.). Without the added oscillation, it is relatively easy to distinguish them. However when a 100 kHz square wave is added in the tracking laser (called strobing in the plot, since we substituted the complex functions in Eq.4.11 by one triangular wave to illustrate this point).), we will see that the contrast of the FCS curves is reduced (Fig.4.7, right panel). Keep in mind that, noises in the FCS measurement also will further reduce the contrast. One certainly can use frequencies that are very different from the folding dynamics(like Fig.4.7, left panel), however in many cases, we either do not know what the folding dynamics is at or it spans a large temporal range, ideal situations like these might not be available in experimental settings.

It should be noted that the experimental FCS curves depends on the exact laser profiles, such as its rotational radius, separation, and tracking errors. It is nearly impossible to measure all the parameters to determine the tracking contributed FCS purely from our theoretical calculations. However, results like Eq.4.9and Eq.4.11 helps us in designing the apparatus to measure holding dynamics by tracking.

4.2.2 Effects of Multiple Constructs, Non-perfect Labeling, and Crosstalk on tFCS.

Experimental conditions are more complicated than the situation in the above calculations for tFCS. The particles we tracked can have multiple hairpins with acceptor dyes bleached, and crosstalk from the donor channel to the acceptor channel is present. To obtain the kinetics of holding, it is necessary to take these effects into consideration in analyzing our tFCS curves. These results should find its applications in protein folding when subunits of the proteins are labeled and later on assembled together to the native state of the protein.
Signals from the acceptor channel is:

$$aa(t) = A_1(t) + A_2(t) + ... + A_x(t) + c(D_1(t) + D_2(t) + ... + D_x(t)) \tag{4.12}$$

where $A(t)$ is acceptor signal and $D(t)$ is the corresponding donor signal, and c is the cross-talk coefficient. We estimate that $c \backsimeq 0.1$ from measurement on bulk FCS as well as total internal reflection experiments.
For the simplest situation, let's first ignore the corsstalks from the donor channel,

and the bead have x FRETing hairpins. In this case, we have:

$$\begin{aligned} g_{2\_NoCrosstalk}(\tau) &= \frac{< aa(t)aa(t+\tau) >}{< aa(t) >^2} - 1 \\ &= \frac{x < BA(t)BA(t+\tau) >< Trk(t)Trk(t+\tau) >}{x^2 < BA(t) >^2 < T(t) >^2} \\ &\quad + \frac{x(x-1) < BA(t) >^2 < Trk(t)Trk(t+\tau) >}{x^2 < BA(t) >^2 < T(t) >^2} - 1 \quad (4.13) \\ &= [\frac{DNA(\tau)}{x} + 1][Tracking(\tau) + 1] - 1 \quad (4.14) \end{aligned}$$

Here $BA(t)$ is the bightness of the dye, $T(t)$ is the effect of tracking to the brightness, $DNA(\tau)$ is the DNA hairpin dynamics that we want to measure, and $Tracking(\tau)$ is the tracking caused correlation. We assumed that different hairpin's dynamics is not coupled.

Next let's include crosstalk from the donor channel. The simplest case is that we only have one construct with both dyes active. To calculate the FCS, we need to first calculate:

$$\begin{aligned} < I(t)I(t+\tau) > &= < [A(t) + cD(t)][A(t+\tau) + cD(t+\tau)] > \\ &= < A(t)A(t+\tau) > + 2c \cdot < A(t)D(t+\tau) > + < c^2 \cdot D(t)D(t+\tau) > \\ &\backsimeq < A(t)A(t+\tau) > + 2c \cdot < A(t)D(t+\tau) > \end{aligned}$$

In the last step we ignored terms in the second order of c, since $c \ll 1$.

Now for $< A(t)A(t+\tau) >$ we have:

$$\begin{aligned} < A(t)A(t+\tau) > &= < BA(t)Trk(t) \cdot BA(t+\tau)Trk(t+\tau) > \quad (4.15) \\ &= < BA(t)BA(t+\tau) > \cdot < Trk(t)Trk(t+\tau) > \\ &= [\sum_{i=1}^{K} \sum_{j=1}^{K} \pi_i \cdot Trans_{ij}(\tau) \cdot QA_i \cdot QA_j] \cdot < T(t)T(t+\tau) > \end{aligned}$$

in which $BA(t)$ is the brightness of the acceptor dye at time t, the π_i is the equilibrium state distributions, $Trans_{ij}(\tau)$ is the transition rate of starting at state i and ending at state j at τ seconds later, QA_i is the acceptor fluorescence rate at state i .
Similarly,

$$\begin{aligned}
\langle A(t)D(t+\tau)\rangle &= < BA(t)Trk(t)\cdot BD(t+\tau)Trk(t+\tau) > \\
&= < BA(t)BD(t+\tau) > \cdot < T(t)T(t+\tau) > \qquad (4.16)\\
&= [\sum_{i=1}^{K}\sum_{j=1}^{K}\pi_i \cdot Trans_{ij}(\tau)\cdot QA_i \cdot QD_j]\cdot < Trk(t)Trk(t+\tau) > \\
&= [\kappa < BA(t) > - < BA(t)BA(t+\tau) >]\cdot < Trk(t)Trk(t+\tau) > \qquad (4.17)
\end{aligned}$$

Apparently, $< A(t)D(t+\tau) >=< D(t)A(t+\tau) >$.
Therefore,

$$\begin{aligned}
\langle I(t)I(t+\tau)\rangle &= [(1-2c) < BA(t)BA(t+\tau) > +2c\kappa < BA(t) >] \qquad (4.18)\\
&\quad \times < Trk(t)Trk(t+\tau) > \qquad (4.19)
\end{aligned}$$

For the single construct on bead with cross talk situation, average fluorescence intensity is:

$$\begin{aligned}
< I(t) > &= < A(t) + c\cdot D(t) > \\
&= < BA(t)\cdot Trk(t) + c\cdot BD(t)\cdot Trk(t) > \qquad (4.20)\\
&= [(1-c) < BA(t) > +c\kappa] < Trk(t) >
\end{aligned}$$

Therefore the autocorrelation function in this simple case is:

$$\begin{aligned} g_{2aa\_HP1}(\tau) &= \frac{< I(t)I(t+\tau) >}{< I(t) >^2} - 1 \\ &\backsimeq \frac{[(1-2c)\langle BA(t)BA(t+\tau)\rangle + 2c\kappa\langle BA(t)\rangle]\langle Trk(t)Trk(t+\tau)\rangle}{[(1-c)\langle BA(t)\rangle + c\kappa]^2\langle T(t)\rangle^2} \quad (4.21) \\ &= \frac{(1-2c)[DNA(\tau)+1] + 2ce^{-1}}{[1-c+ce^{-1}]}[Track(\tau)+1] - 1 \quad (4.22) \end{aligned}$$

where e is the average FRET efficiency for the dynamics.

One can easily repeat the above calculation for the case if there are x FRET pairs with y donor dyes (designated by D^*'s) that are not FRETing. That is:

$$aa(t) = \sum_{i=1}^{x} A_i(t) + c(\sum_{i=1}^{x} D_i(t) + \sum_{j=1}^{y} D_j^*(t))$$

In this general case, we will have:

$$\begin{aligned} g_{2\_Acceptor}(\tau) &= \{\frac{(x-2cx^2)[DNA(\tau)+1] + 2cx(1+\frac{y}{x})e^{-1} + x(x-1)}{[x(1-c)+c(x+y)e^{-1}]^2}\} \\ &\quad \times\{Track(\tau)+1\} - 1 \quad (4.23) \end{aligned}$$

We can see Eq.4.23 reduces to Eq. 4.14 by setting c to zero. Note again here we omitted terms in the second order in c.
In the donor channel, if we have x donor FRET pair with y donor not FRETing, then the autocorrelation function is:

$$g_{2\_Donor}(\tau) = \{\frac{x[DNA(\tau)+1] + 2xy(1-e)^{-1}}{[x+y(1-e)^{-1}]^2}\}\{Track(\tau)+1\} - 1 \quad (4.24)$$

4.3 Track and Probe Folding Dynamics of Dye Quencher Samples

In this part of the thesis, we present the results of tracking and probing the folding dynamics of dye-quencher labeled hairpin. The goal of the experiments is two folds. One is to to practice basic theory we have developed in the previous section. For example, how to factor out the tracking contributed signal statistics from the overall signal. Two is to use tracking as a tool to see if our hairpin samples have long lived states that have not observed in our previous experiments. Certainly, this can also go the other way around: the comprehensive results that we have already learned on the DNA hairpin can serve as references for us to investigate tracking as a potential tool to monitor molecular folding. If the tracking can recover what we have observed in the stationary FCS results in the fast time scales ($< 10\ mS$), this means that we could also use the tracking results to check whether we have long lived states by taking advantage of the long observation time in the tracking experiments. However, if the tracking results cannot recover the stationary FCS results, then we need to investigate why. It can be the interaction between the bead surface and the hairpin or the technique itself.

4.3.1 The Tracking Complex and Experimental Conditions

The sequence of the sample sequence is HP3, a three base pairs stem hairpin with a Atto700 dye at its 3' and a quencher, dG, at it's 5' end. We will use HP3-CTL, a sequence without the 5' end quencher as control. For the exact sequence of the DNA hairpin, please refer to Table.3.3.1. In theory, the fluorescence of the control sequence codes the statistics of tracking errors as discussed in the theory section. The sample sequence, however, will display additional folding dynamics. The usage of control sequences is necessary for us to extract the folding dynamics of the hairpin.

Our hairpin constructs have diffusion coefficient around $150\mu m^2/S$, which is currently too fast to be tracked by the tracking microscope. To slow down the molecule,

we attached the Atto700 labeled hairpins on a functionalized, non-fluorescent microsphere (Invitrogen FluoSpheres NeutrAvidin labeled microspheres, 0.04 um, nonfluorescent *1% solids*.) by the strong biotin-streptavidin bonding. The sample complex is assembled by mixing mM concentrations of bead solution and DNA hairpin solutions together at 150 mM of NaCl. By mixing the two reactants at known concentrations, we can control the average number of DNA per bead. To provide tracking fluorescence, concentrated biotinylated Atto532 were added in the DNA bead solution. Based on the mixing concentrations, on average each bead can have about 200 Atto532 attached. The final tracking solution is made by diluting the assembled complex solution to 10 picomolar in various concentrations of NaCl.

It should be noted that this attachment can potentially affect the hairpin folding dynamics, and is not the ideal construct to learn the folding dynamics of hairpin. It however should not stop us from investigating the potential of tracking in probing molecular folding dynamics, because 1) what we can learn from the present study, such as the dyes, theories, and good practices can still be used later for true single molecule tracking FRET experiments; 2) our abilities of tracking fast diffusing molecules are constantly improved and will allow us tracking fast molecules like the hairpin (Charles Limouse, unpublished results).

The tracking beam configuration and constructs are illustrated in Fig.4.4. In probing the Atto700-Quencher hairpin folding dynamics, we focused a 634 nm probe laser (Coherent) in the tracking system's fixed point in addition to the two 532nm tracking beams. Tracking fluorescence was induced by the green laser and fed to the tracking system, while the folding fluorescence was induced by the 634 nm red laser (7 μW in power and waist size is 0.65 μm), and was filtered by 705/50 (Chroma). There was no observed crosstalk in the Atto dyes system - the green tracking laser could only excite Atto532, the red laser could only excite Atto700, and no crosstalk of these two channels due to their far spectral separation.

The solution conditions are exactly the same as the Atto FCS experiments. For the tracking sample assembly, we used one quartz coverslip (Esco Products), one glass coverslip (VWR micro cover glass) and two pieces of double sided tape (3M) to form a 5 mm wide, 20 mm long, and about 100 μm deep channel as our solution space.

Quartz coverslips were used in order to further reduce background fluorescence. One percent casein solution (Sigma) was applied to the channel for 5 minutes before the channel was washed three times using 25μl of 18 $M\Omega$ Milipore water and another three times using 25 μl of tracking sample solutions in order to reduce nonspecific absorption of the tracking complex to the glass/quartz surfaces.

4.3.2 Data on HP3-CTL Control Sequence

Fig.4.8 and Fig.4.9 show two examples of tracking events on HP3-CTL sequences. In the first example, the top panel is the 634nm probe beam induced fluorescence of Atto700 in 705/50 nm channel; the middle panel shows the trajectories of the stage position; the third panel shows the 532nm tracking beam induced fluorescence of Atto532 in 585/40 nm channel.

In the first example, the complex is tracked for more than 50 seconds. The complex was was lost because of limited traveling range of the z stage (40 μm traveling range) at 50 second. The fluorescence of the tracking dyes (Atto532) was locked at 20 KHz by modulating the tracking laser power. This helps by providing a consistent gain in the tracking feedback loop. The probe beam's power, however, is locked around 100 kHz at 7 μW. The variations that we see in the top panel (such as the one at 43 second) is certainly not due to tracking error or the laser power blinking but due to the intrinsic blinking of the Atto700, which is consistent with what we have already seen in the surface confocal measurement in our hairpin studies.

In the second example, we used a different format to best display similar data to show a new feature of the fluorescence signal: step-wise bleaching of the Atto700 dyes. Fluorescence signal, filtering (based on forward-backward algorithm) results such as posterior probability of the states and emission rates, and 3D showcase of the stage trajectories are presented. This illustrates that on a single bead, there can be

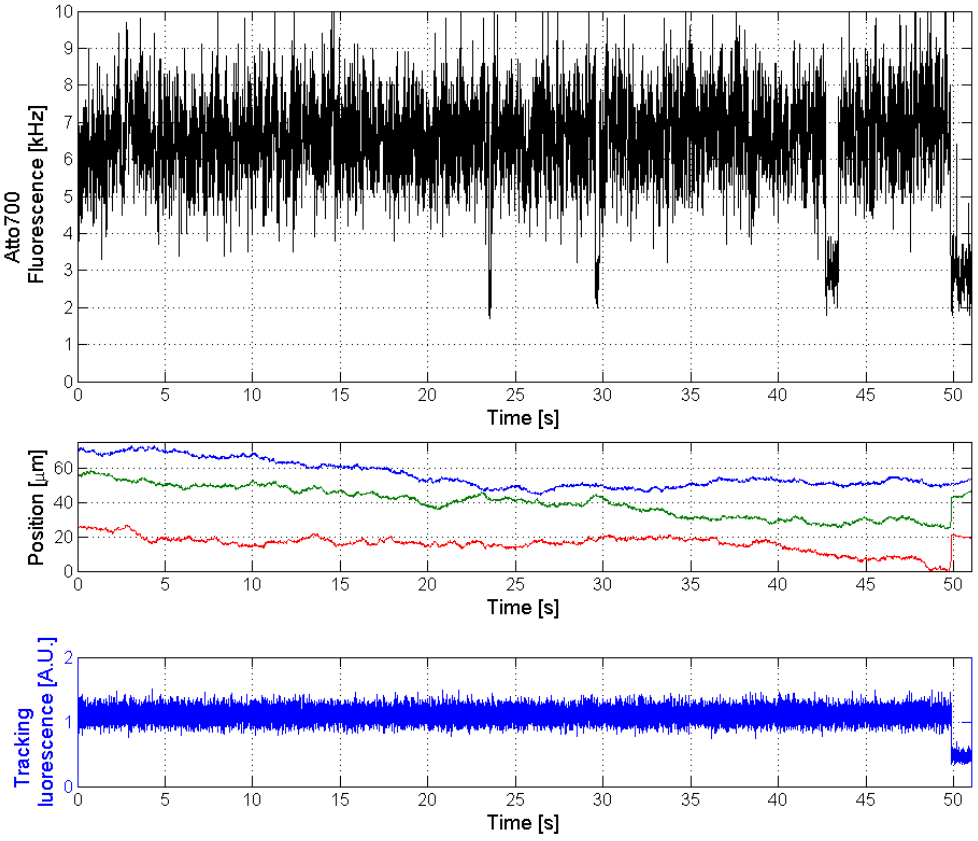

Figure 4.8: Examples of tracking events of HP3-CTL and bead complex. Blinking events are apparent. The top panel is the 634nm probe beam induced fluorescence of Atto700 in 705/50 nm channel; the middle panel shows the trajectories of the stage position; the third panel shows the 532nm tracking beam induced fluorescence of Atto532 in 585/40 nm channel. The variations that we see in Atto700 fluorescence is not due to tracking error or the laser power blinking but due to the intrinsic blinking of the Atto700, which is consistent with what we have already seen in the surface confocal measurement in our hairpin studies.

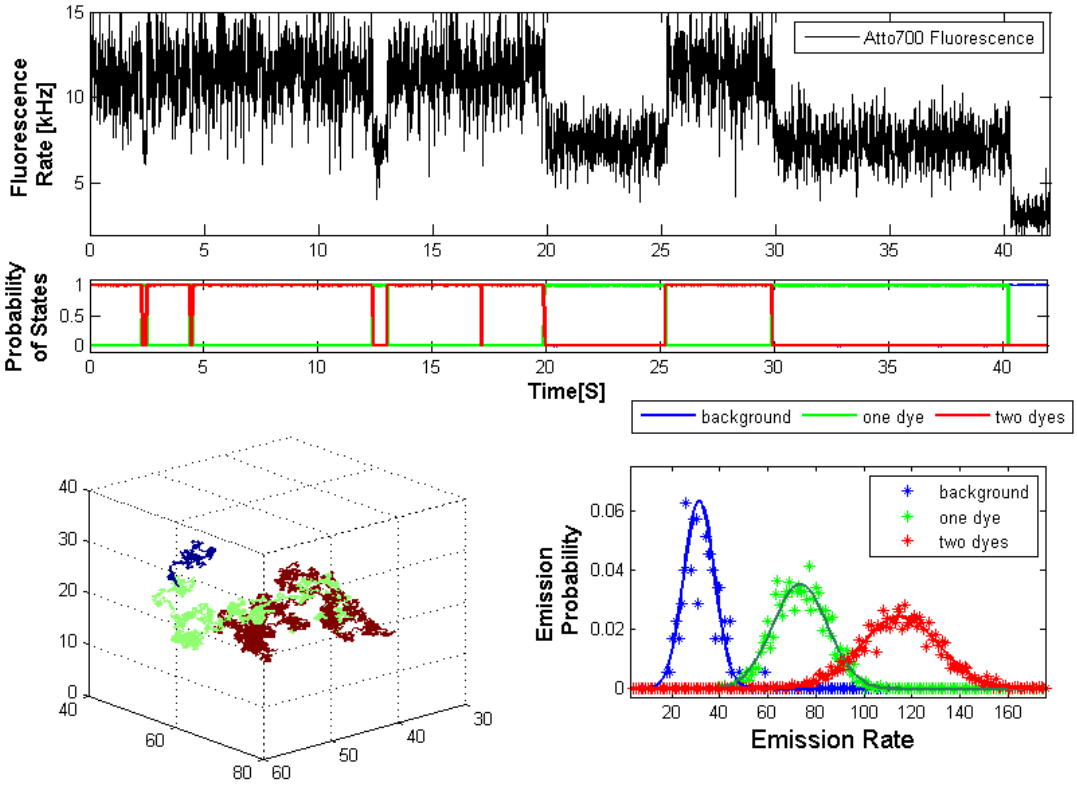

Figure 4.9: Examples of tracking events of HP3-CTL and bead complex. Bleaching events are apparent. Fluorescence signal, filtering (based on forward-backward algorithm) results such as posterior probability of the states and emission rates, and 3D showcase of the stage trajectories are presented. This illustrates that on a single bead, there can be multiple hairpin constructs.

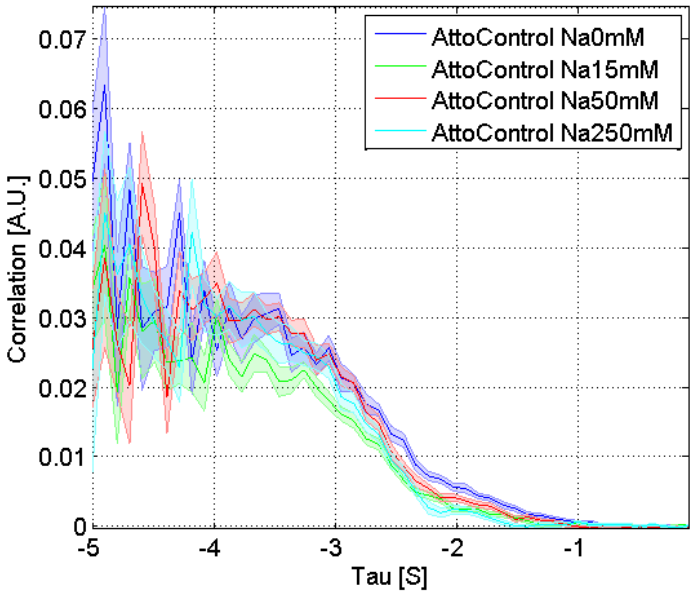

Figure 4.10: Tracking FCS signals calculated from the on-state of Atto700 labeled HP3-CTL at a range of salt concentrations (0 to 500 mM). The FCS curves of HP3-CTL bead complex are within one error bound of each other. Given the same distribution of diffusion coefficient and tracking parameters, this is is expected from a sample that is supposed to lack folding dynamics.

multiple hairpin constructs. And we will see later, that this will not prevent us from observing folding dynamics.

Next we calculated the FCS signals on the on-state of the fluorescence signal, and the results are displayed in Fig.4.10(each curve represents the averaging results of around 26 tracked molecules). The identical experiments were performed at a range of salt concentrations to see if there is any salt dependence of the FCS signals. We have seen that in the FCS studies of the DNA hairpin, the dye's photophysics is independent of the salt concentration. Thus as expected, given the same distribution of diffusion coefficient and tracking parameters, the FCS curves of HP3-CTL bead complex are within one error bound of each other. What we have measured by the HP3-CTL bead complex is a direct measurement of the tracking contributed FCS.

4.3.3 Data on HP3 Sample sequence

Potential Folding Dynamics from HP3

The FCS signals of the HP3 bead complex, under identical experimental conditions, however, displayed a very different dependence on the salt concentration, as shown in Fig.4.11. For each salt concentration, we tracked about 60 hairpin bead complex (a total of 477 tracking events), calculated FCS curves for each tracking event, and the tracking time weighted averages and standard deviations of the mean are displayed. At zero mM of NaCl concentration, the FCS signals of sample and control (grey) are within error bounds of one another; but as salt concentration in increased, the amplitude of the FCS signal is elevated, a feature that was not observed in the HP3-CTL control experiment.

The elevated amplitude of the FCS curves at higher concentrations of salt is same qualitative behavior in our previous FCS measurement on the freely diffusing hairpin. However the folding rate constants that we can learn from this measurement is very different from the stationary FCS results. To obtain the folding rate constants, we first note Eq.4.14:

$$g_2(\tau) = [\frac{DNA(\tau)}{x} + 1][Tracking(\tau) + 1] - 1$$

where $Tracking(\tau)$ is the tracking contributed FCS, and $DNA(\tau)$ is the folding dynamics of DNA. To extract the folding dynamics, we first performed a fit of the FCS signal of the control sample to a theoretical function of the second order of the tracking system with variable gain[BMM07]. The extracted theoretical curve was used as $Tracking(\tau)$. Then we fitted the FCS curves at higher concentrations with this $Tracking(\tau)$ in addition of a two-state model in Eq.4.14. χ^2 goodness-of-fit suggests that a two-state model can be used to describe the observed dynamics at all the salt concentrations we probed. The resulting fitting parameters are shown in Fig.4.12.

It should be noted, however, that the dynamics we observed here is orders of magnitudes slower than what we have seen in the case of freely diffusing hairpins probed by FCS. We will come back and discuss about this important point after we

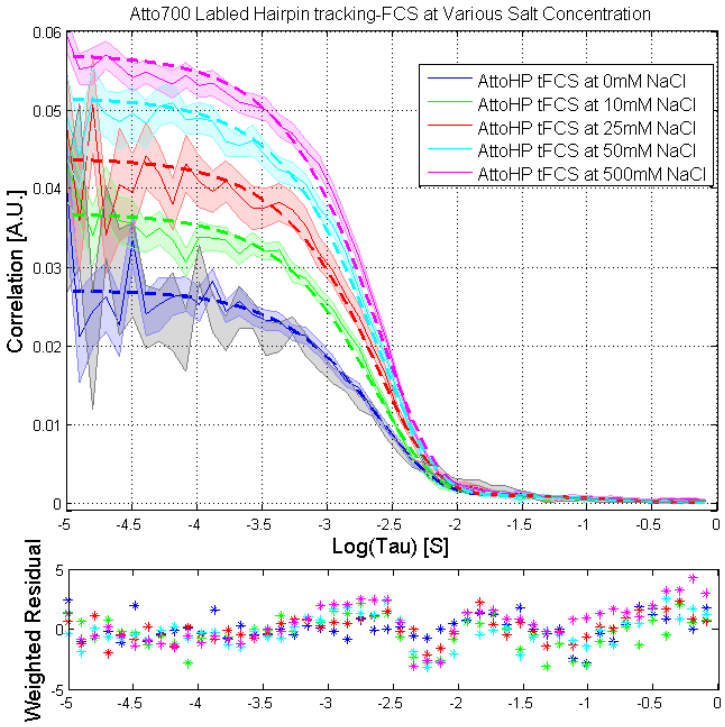

Figure 4.11: The raw FCS signals and the fitting results HP3 tracking FCS curves at various salt concentrations. The tracking contributed FCS is first calculated from the FCS fits of control tracking results shown in Fig.4.10. Then a two-state model together with the tracking contributed FCS curves are combined to perform the fit. χ^2 goodness-of-fit suggests that a two-state model can be used to describe the observed dynamics at all the salt concentrations we probed.

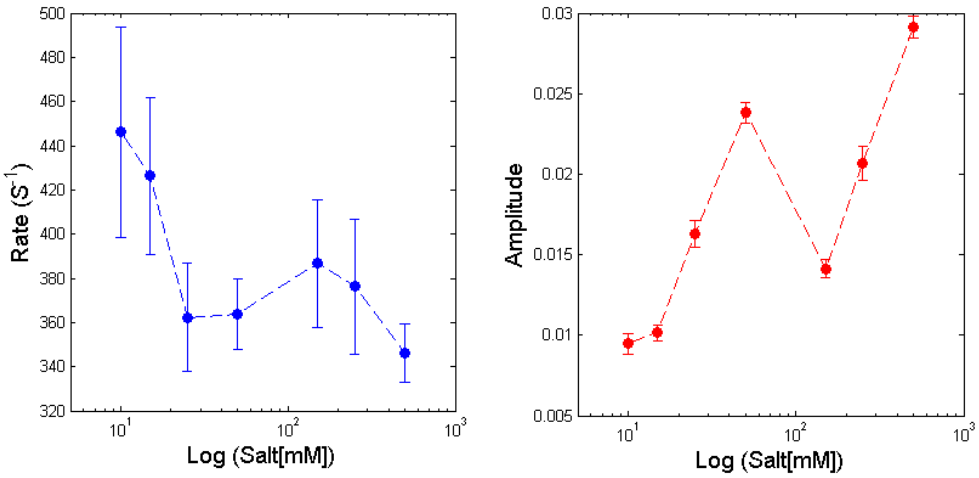

Figure 4.12: Two-state fitting results of the HP3 folding FCS. The amplitude is shown on the left panel, and the extracted rate is shown on the right panel.

present the complimentary folding dynamics data provided by tracking FRET labeled hairpins.

Just like our previous FCS measurement, another independent sign of folding events is the average brightness of the dye. At higher salt concentration, the dye and the quencher stay at close proximity, resulting in dimmer average fluorescence per molecule. To measure this brightness, we can directly look into the step size of the bleaching event. The result of this measurement at different salt concentrations is shown in Fig.4.13. We plot the relative dye brightness change, using the brightness at zero mM of salt as reference. Overall, it shows that the average dye brightness decreased as salt concentration is increased - up until 250 mM of salt. The return and large variation at 500mM salt is due to the fact that at that high salt, the signal to noise ratio of the steps is poor and we missed many such steps. Thus the overall population at 500 mM salt has more fraction of the molecules that do not have quenchers. This results into average value that is similar to zero mM of salt with relative large variance at this data point.

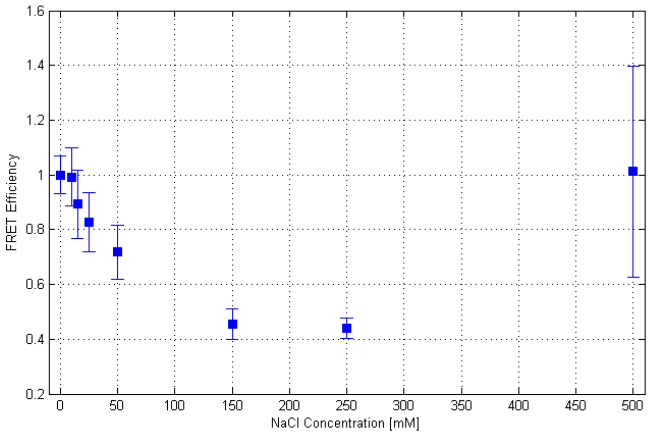

Figure 4.13: The brightness of the Atto dyes obtained from the bleaching steps durign tracking as a function of salt. This result is very similar to what we have observed in the solution data (Fig.3.14D)

Tracking Statistics

The results we have presented so far are potential results for the folding of the hairpin from the analysis of the Atto700 dye. However, another dynamical source of Atto700 fluorescence signal is the tracking error, which is again the distance of the molecule from the fixed point of the tracking system (and thus to the center of the probe beam.). The statistics of the tracking error directly factors into the statistics of Atto700 in a fashion governed by Eq.4.3.3. We have already demonstrated by control samples that the statistics of the tracking should be the same, however it would be the best if we could provide another evidence directly from the data of HP3 hairpin to show that. Fortunately, from the statistics of the stage positions, we can also learn about the statistics of tracking, and the statistics of the tracking is in fact independent of the folding. This is because that there is no observable crosstalk between the tracking spectral channels and folding spectral channels. If the tracking behaves differently at different salt concentration, it can render different statistics of the tracking error, and result into the FCS dynamics we seen in Fig.4.11.

Here we briefly outline the procedures to obtain the tracking model parameters.

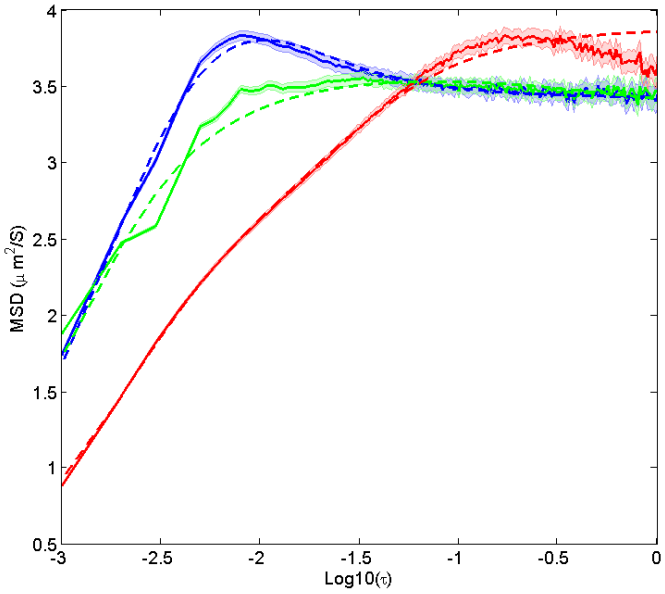

Figure 4.14: $MSD(\tau)$ from 477 tracking events in Fig.4.11.

For each tracking session, we calculate the scaled variance of the stage position increment, $MSD(\tau)$, as a function of delay time τ for each axis (scaled by $(2\tau)^{-1}$). At τ values smaller than the inverse of the tracking bandwidth, the stage could not follow the particle and thus $MSD(\tau)$ starts from zero and approach the diffusion coefficient of the molecule. The exact shape of $MSD(\tau)$ depends on the bandwidth of the tracking system, the diffusion coefficient, and noise density. Fitting the measured $MSD(\tau)$ with its theoretical expressions, we can extract these parameters. As an example, when we aggregate all the 450+ tracking, we obtained the $MSD(\tau)$ as shown in Fig.4.14. The solid line is the average value of $MSD(\tau)$, the shaded region is the standard deviation of the mean value, and the dashed line is a fit to the measured value. In this specific case, the extracted diffusion coefficient is $D = 3.55 \pm 0.25$ $\mu m^2/S$. We performed such analysis on each tracking event at different salt concentrations. The averages and standard deviations of the extracted tracking parameters are shown in Fig.4.16, and Fig.4.15 shows the histogram of the extracted diffusion

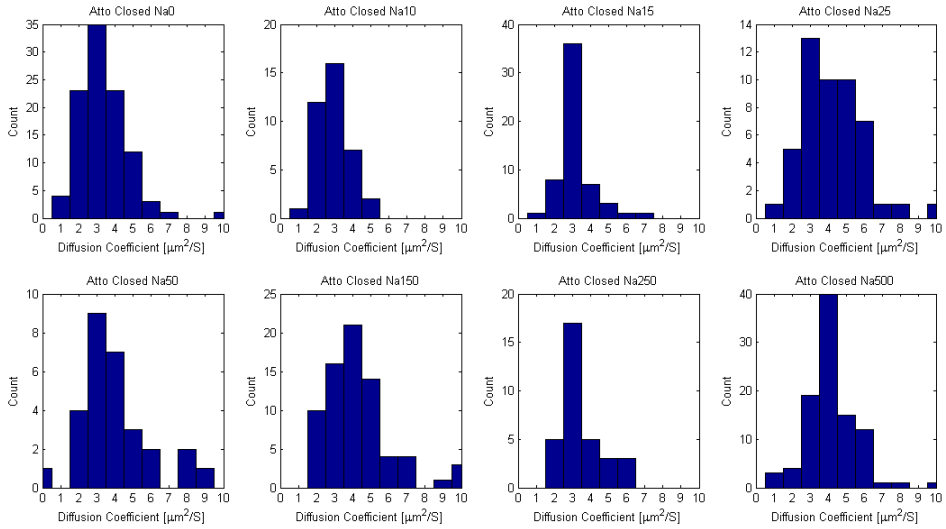

Figure 4.15: The histogram of diffusion coefficient of the hairpin bead complex at different salt concentrations.

coefficient. Examining the plots, we do not see a specific dependence of the tracking parameters on the solution condition. Together with the control HP3-CTL FCS results, we conclude that tracking contributed FCS is independent of the solution condition. What we see in Fig.4.11 is indeed due to some interaction of the hairpin to the Atto700 dye.

A puzzling result is that we do not see the $100nS$ FCS relaxation in the hairpin folding dynamics that we see in the solution experiment. Instead, we see a dynamics that is independent of tracking, can also be described by a two-state model, but is slower by about one order of magnitude. We will come back and address this when we present the tracking FRET data and analysis.

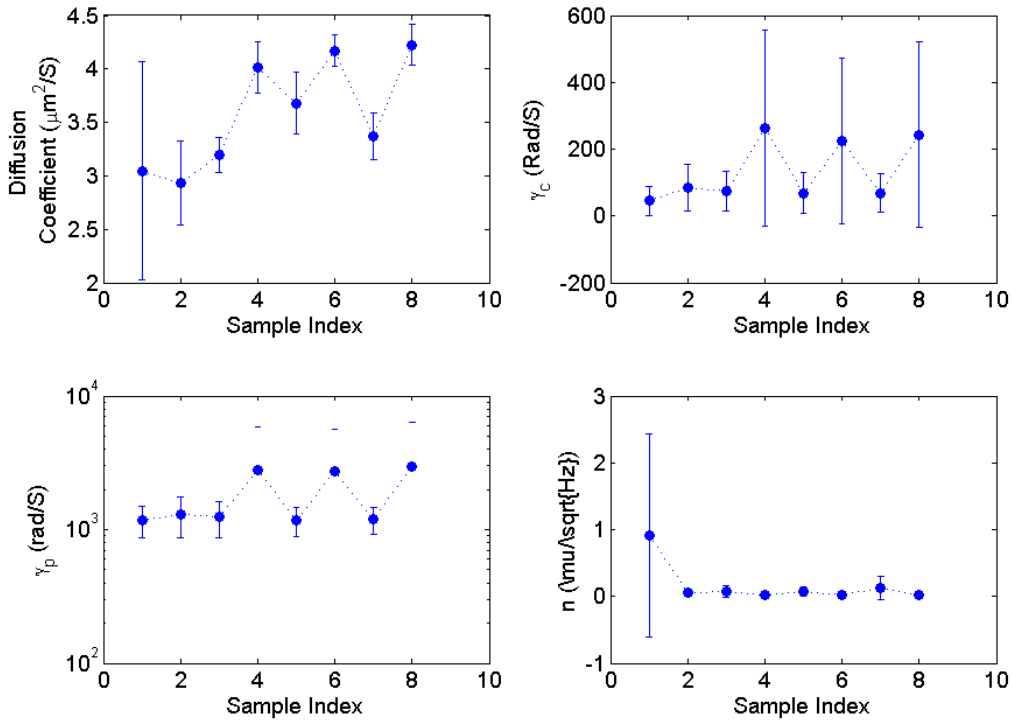

Figure 4.16: Extracted tracking parameters as a function of salt. The lack of dependence of the tracking parameters on the solution condition indicates that tracking dynamics is independent of solution conditions.

4.4 Track and Probe Folding Dynamics of FRET labeled Samples

In this part of the thesis, we present our results of tracking Cy3-Cy5 FRET labeled hairpin complex on a transparent 40nm bead. One benefit of using a FRET labeled sample is to extend the spatial sensitivity of our probes; another benefit is that, the unti-correlation feature of the donor and acceptor dye in theory should be a very clean signal. This signal cannot be affected by the tracking dynamics, as the statistics of the tracking errors will contribute to both acceptor and donor channels but would not affect the anti-correlation dynamics between the acceptor and donor dye itself. This will help us in answering a puzzling result that we obtained in the previous section.

After presenting the details of the tracking complex and experimental conditions, I will first give several tracking examples, from which we will clearly see examples of donor and acceptor bleaching event. Next I will present and focus on the tracking FCS signal. We will see that the spatial temporal modulation of the tracking laser renders non-trivial features to the overall FCS signal, which makes the visual inspection of folding dynamics quite difficult. To get rid of these tracking contributed FCS signal and extract the folding dynamics of the hairpin, I will discuss several strategies. Finally, I will present the extracted folding dynamics from the FRET labeled sample, and offer an possible explanation of the inconsistency between the tracking and solution data.

4.4.1 The Tracking Complex and Experimental Conditions

The sequence of the sample is a FRET labeled HP3, with a Cy5 at the 5' of the molecule and a CY3 at the 3' end. The control we used in this case is just a donor only, Cy3 labeled HP3. To slow down the diffusion of the hairpins, the same 40 nm, NeutrAvidin labeled nonfluorescent microsphere is used.

The beam configuration is shown in Fig. 4.6. We only used the 532 nm tracking beam. The bandpass filters are 585/40 (Chroma) and 655/40 (Chroma) for donor and acceptor, respectively. The signals from these two spectral channels were used

to drive one input, four outputs TTL fan-outs, whose outputs were recorded on a computer and were also summed together by a digital OR gate to drive the tracking system.

In the Cy labeled hairpin tracking experiments, we also applied 2.5mM protocatechuic acid (PCA), 10nM protocatechuate-3,4-dioxygenase (PCD), and 1mM trolox to the solution. PCD/PCA/Trolox oxygen scavenging system has been shown to improve Cy3 and Cy5 stability dramatically[AMP08]. Everything else is the same as the Atto tracking case.

4.4.2 Tracking Data Examples.

We first present the tracking data of the donor only hairpin sample. As shown in Fig.4.17, the top panel is the fluorescence of donor, acceptor, and their average; the middle panel shows the tracking stage positions; the bottom panel shows the overall tracking laser power. It should be noted that the sum of the donor and acceptor fluorescence is used to track the complex and is locked at 20 kHz. One subtle feature we can see in the laser power is that there is a sudden jump around 32 second, which corresponds to a bleaching event of a donor dye. This event made the whole complex dimmer. To achieve the same fluorescence rate, the fluorescence controller increased the laser power, as shown by the elevated average power from 32 second to the end of the tracking session. This also induced more background signal in the acceptor channel. These bleaching events are constantly observed at different salt concentrations.

The FRET labeled HP3 data is present by the same format in Fig.4.18. This particular example is in solution of 500 mM NaCl, and shows an acceptor bleaching event around 8 second. Because not every energy transferring event is converted back to a photon by one acceptor, a bleaching event of the acceptor made the whole complex brighter. Thus the required laser power to lock the total fluorescence went down from 8 seconds.

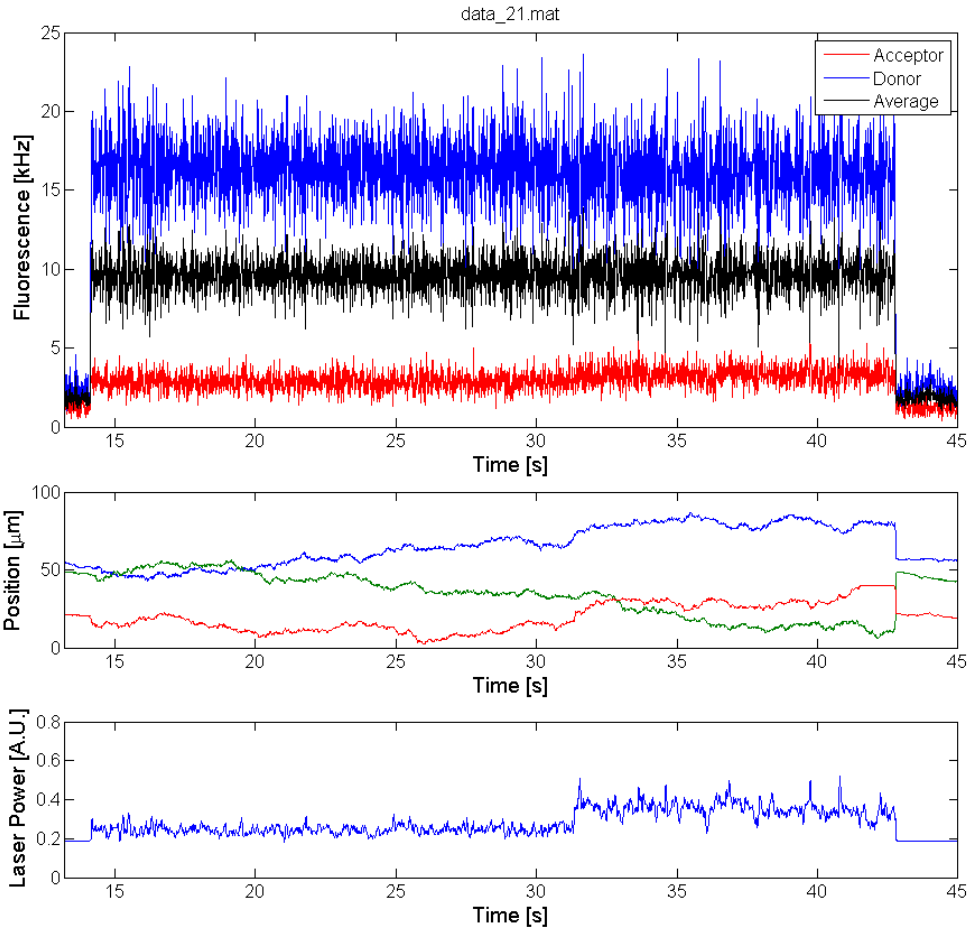

Figure 4.17: Examples of tracking events of donor only HP3 and bead complex. Fluorescence signal, stage positions, and tracking laser power are displayed. It should be noted that the sum of the donor and acceptor fluorescence is used to track the complex and is locked at 20 kHz. One subtle feature we can see in the laser power is that there is a sudden jump around 32 second. This is the result of one donor dye bleached, which made the whole complex dimmer. To achieve the same fluorescence rate, the fluorescence controller increased the laser power, which also induced more background signal in the acceptor channel.

By looking at signatures like the two above examples, it is easy to spot these events along a tracking event, evidently shown in Fig.4.19. In this particular example, we cam clearly see two donor bleaching events and two acceptor bleaching events.

To calculate the tracking FCS, we hand picked traces that have stable fluorescence levels. Before we present that results, we first can perform a direct calculation of the FRET efficiency over the traces we select at different solution conditions. The result is shown in Fig.4.20. This FRET efficiency dependence on the solution condition is very similar to our solution case. The relatively bigger error bars are due to bleaching events on the tracking traces.

4.4.3 Tracking FRET FCS

From each stable region of the fluorescence traces, we calculate donor-donor autocorrelation, acceptor-acceptor autocorrelation, and donor-acceptor cross-correlation functions. Then we perform a tracking time weighted average for these signals at one salt concentration. The average and sample standard deviation for each salt concentration in the donor channel for both sample (FRET labeled HP3) and control (Donor only labeled HP3) sequence is displayed in Fig.4.21. There are several distinctive features of these FCS curves, and I will discuss them below. While some features prevent us from identifying folding dynamics, some features provide evidences to support that we have observed folding induced FCS, not some artifacts.

First is the zigzag shown in the plot. This zigzag feature makes the visual inspection of the FCS curves very difficult. This is in fact due to the beam modulation in time. A simple square wave modulation (turning the beam on-and-off) will result into a triangular FCS curve at the same frequency. Our tracking beams are rotating and turning on and off around 100 kHz, resulting a theoretical expression of FCS curve as described by Eq.4.11. This modulation shows up in τ ranges comparable to the modulation frequency (in this specific case, around 10 μS). But in τ ranges that

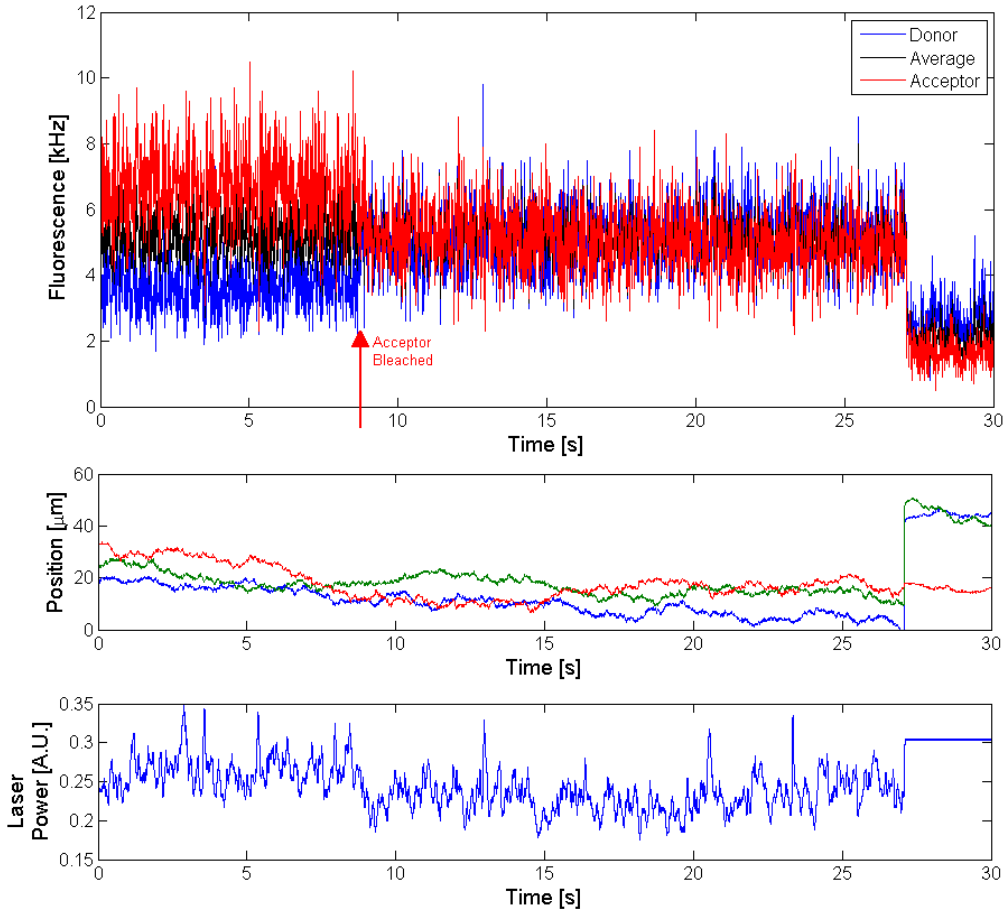

Figure 4.18: An example of tracking event of FRET labeled HP3 and bead complex, showing an acceptor bleaching event. Fluorescence signal, stage positions, and tracking laser power are displayed. It should be noted that the sum of the donor and acceptor fluorescence is used to track the complex and is locked at 20 kHz. This particular example is in solution of 500 mM NaCl, and shows an acceptor bleaching event around 8 second. Because not every energy transferring event is converted back to a photon by one acceptor, a bleaching event of the acceptor made the whole complex brighter. Thus the required laser power to lock the total fluorescence went down from 8 seconds.

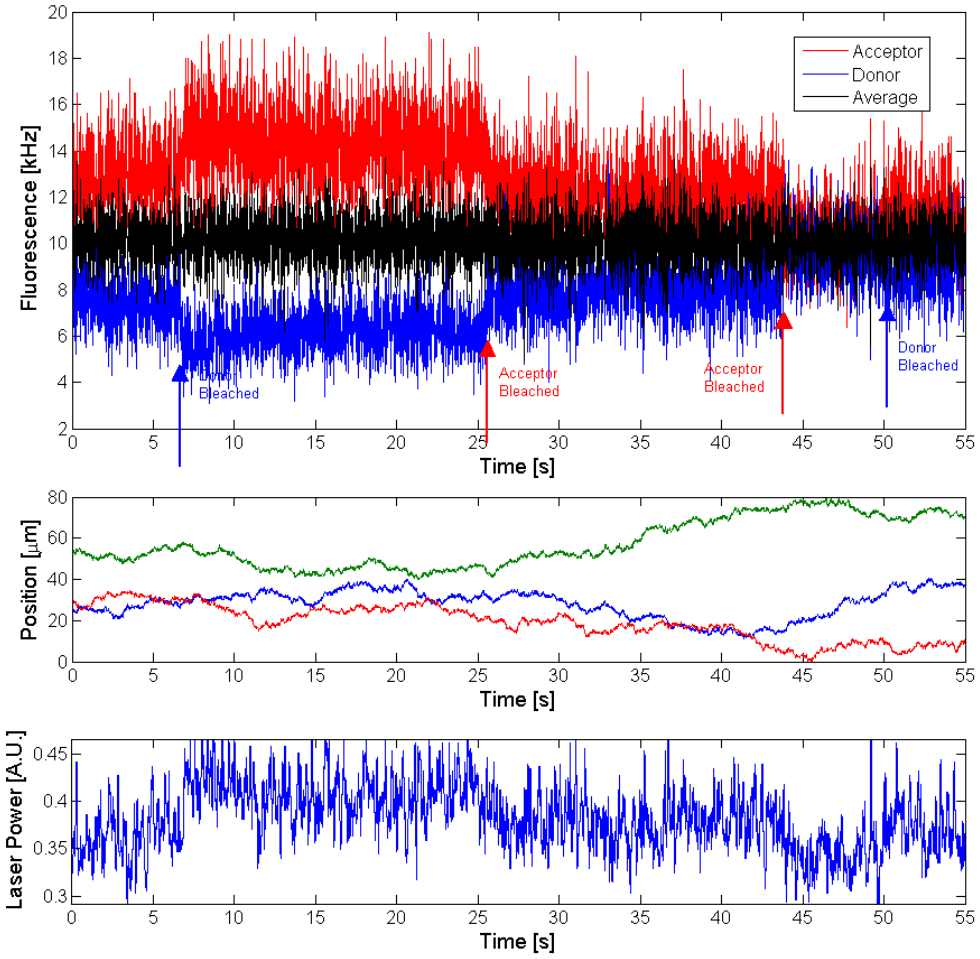

Figure 4.19: A example of tracking events of FRET labeled HP3 and bead complex, showing multiple donor and acceptor bleaching events. Fluorescence signal, stage positions, and tracking laser power are displayed. It should be noted that the sum of the donor and acceptor fluorescence is used to track the complex and is locked at 20 kHz. In this particular example, we cam clearly see two donor bleaching events and two acceptor bleaching events.

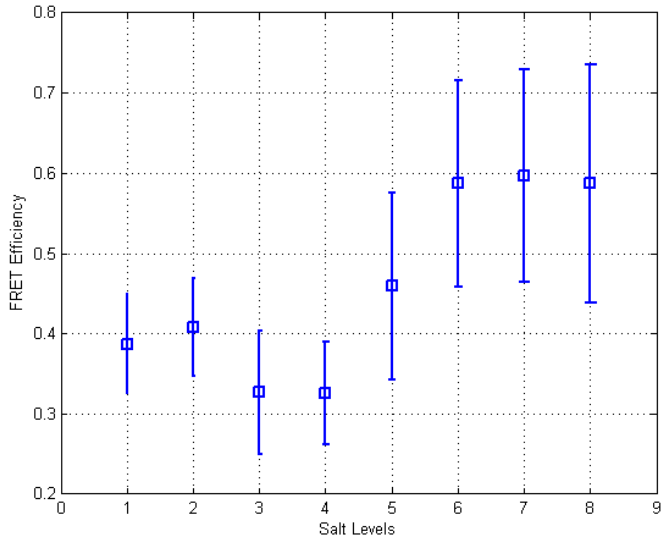

Figure 4.20: FRET efficiencies of the FRET labeled HP3 obtained from tracking events. This FRET efficiency dependence on the solution condition is very similar to our solution case. The relatively bigger error bars are due to bleaching events on the tracking traces.

are orders of magnitude larger than the inverse of the modulation frequency, we will again see a smoothed curve because of the averaging calculation we perform in each bin (bin size is about 0.025 τ).

The difference between the donor only labeled construct and FRET labeled construct is their relative amplitude of the FCS signals (Note that what we displayed in Fig.4.21 is the raw FCS signal, without any normalization procedure). While the control grouped together from low to high salt concentration, the sample construct displayed elevated FCS signals in sequential order of the salt concentration. This is consistent with what we have seen already in the dye-quencher case, and is the first evidence that we have observed folding induced fluorescence signal variation.

A second qualitative evidence is apparent if we can also examine the all three correlation functions of the sample construct together, as shown in Fig.4.22. Each

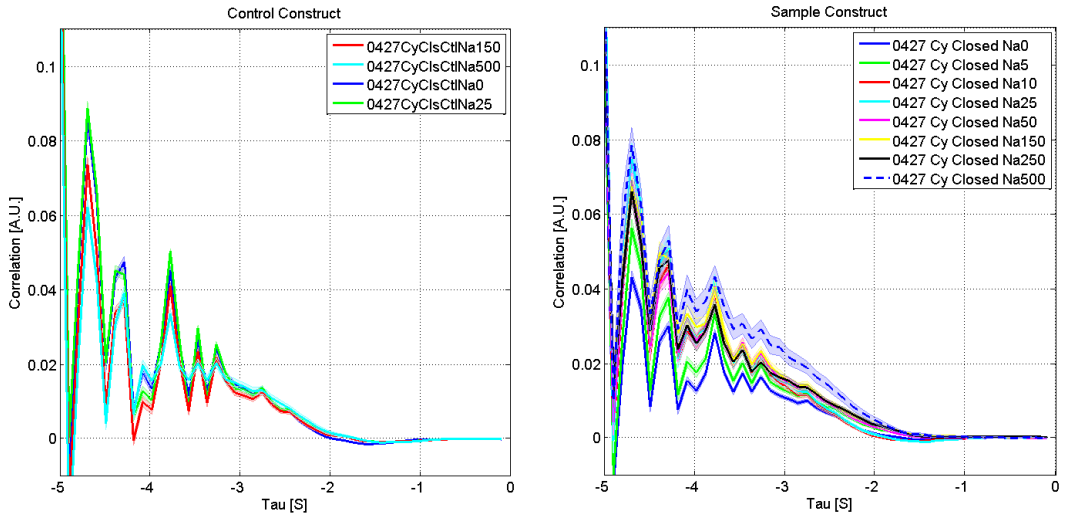

Figure 4.21: Donor tracking FCS of control (left panel) and sample constructs. The control is a hairpin that only has donor; the sample construct is a FRET labeled hairpin.

plot is at one salt concentration with donor-donor autocorrelation (blue), acceptor-acceptor autocorrelation (red), and donor-acceptor cross-correlation (black). In this format, it is easy to spot that in low salt concentration (¡ 50mM NaCl), the acceptor autocorrelation is above the donor; but as salt concentration is more than 50 mM of NaCl, we see that the donor autocorrelation is above the acceptor. The amplitude of the autocorrelation at zero time delay is the variance divided by the mean of the fluorescence rate. At low salt concentration, folded structures are not stable and thus for donor fluorescence signal, the variance is relatively small, and the mean is relatively large (as compared to higher salt concentration.). Therefore the amplitude of donor autocorrelation function is smaller than the acceptor case (another way to think about this is that the donor $g_2(0)$ is in general inverse proportional to equilibrium constant of the folding reaction.) .

To proceed, we need to extract the folding dynamics from the overall FCS signals.

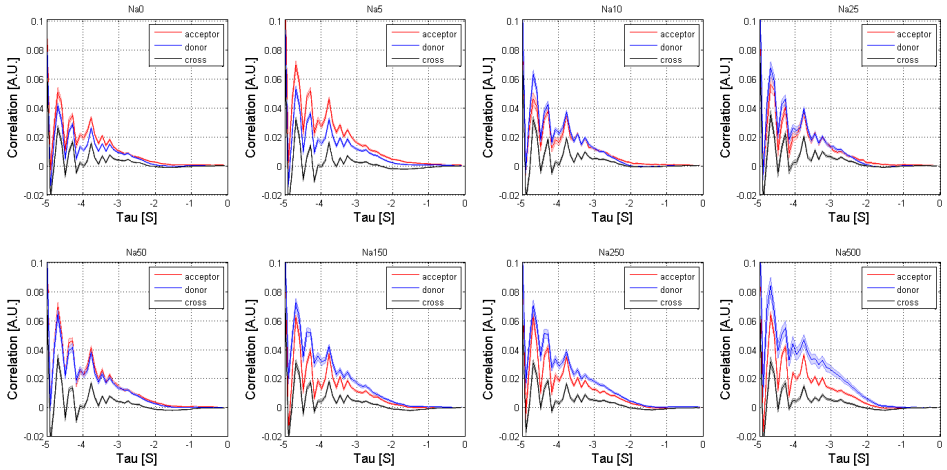

Figure 4.22: Tracking FCS from donor-donor, acceptor-acceptor, and donor-acceptor correlation functions.

4.4.4 Extract the Folding Dynamics

To extract the folding dynamics, we need to have an accurate estimation of the tracking contributed FCS, which is a function of tracking errors, and spatial temporal modulation of the tracking beam. This is not a trivial task, given how complicated our experimental system is. There are at least three ways that we can estimate $Tracking(\tau)$. I will give descriptive advantages and dis-advantages for each method below.

1. Utilizing the theory we developed, we can first analyze the statistics of the stage positions and extract the tracking parameters. Based on these tracking parameters, we can calculate the theoretical form of $Tracking(\tau)$. Then we can use this to factor out the tracking contributed FCS for the folding dynamics. However to accurately extract the tracking parameters, long tracking sessions are needed, which is not always guaranteed when we track fluorescence from a few dyes. Even we can have long tracking sessions, the accuracy of the tracking contributed FCS calculated by theory depends on accurate modeling and measurement of the controller, plant, beam modulation frequencies and geometries - all of which are subject to change over time and are not possible to measure

prior to each experiment. Thus it is not a surprise that this procedure could not extract the folding dynamics, which can be shown by performing experiments on control samples and compare the measured and calculated FCS curves.

2. The second method is to use control sample and measure the tracking contributed FCS directly. This is the definition of tracking FCS. We don't know how exactly we can evaluate the fidelity of tracking FCS extraction, but using this method, we still left with zigzags (from beam modulation) in $DNA(\tau)$, which should not show such behavior as folding dynamics.

3. The third method depends on one validated assumption. For Cy3 and Cy5 FRET pair, it has been shown that $A(t) + D(t) = \alpha D_0$, where D_0 is a constant and $\alpha < 1$. Therefore we can directly sum the donor and acceptor fluorescence to drive the tracking system. Since the time arriving events of the photons in both channels are recorded, it is trivial to combine them and calculate the FCS curves on it. Thus:

$$DNA(\tau) = \frac{\langle X_1(t)X_2(t+\tau)\rangle / \langle X_1(t)\rangle\langle X_2(t+\tau)\rangle}{\langle D_0(t)D_0(t+\tau)\rangle / \langle D_0(t)\rangle^2} - 1$$

Here X can be donor or acceptor fluorescence. Using this method, we have been successfully extracted the folding dynamics.

By the last method, in Fig.4.23, the extracted $DNA(\tau)$ from the donor channel at different salt levels are displayed. The triangular waves from the beam modulation is not a dominant feature on this point, and we can see that in addition to the amplitude increase at higher salt concentration, the FCS curves also illustrate that at higher salt concentration, the dynamics of the folding becomes slower. For example, the 10^{-2} S dynamics was not evident in the Na0, but is very apparent in any other solution condition. This behavior is what we expect at higher salt concentration solution, in which the hairpin structure is more stable.

A simple two-state model cannot fit any of the extracted folding dynamics; the dashed lines are weight fits to a stretched exponential function. This is consistent

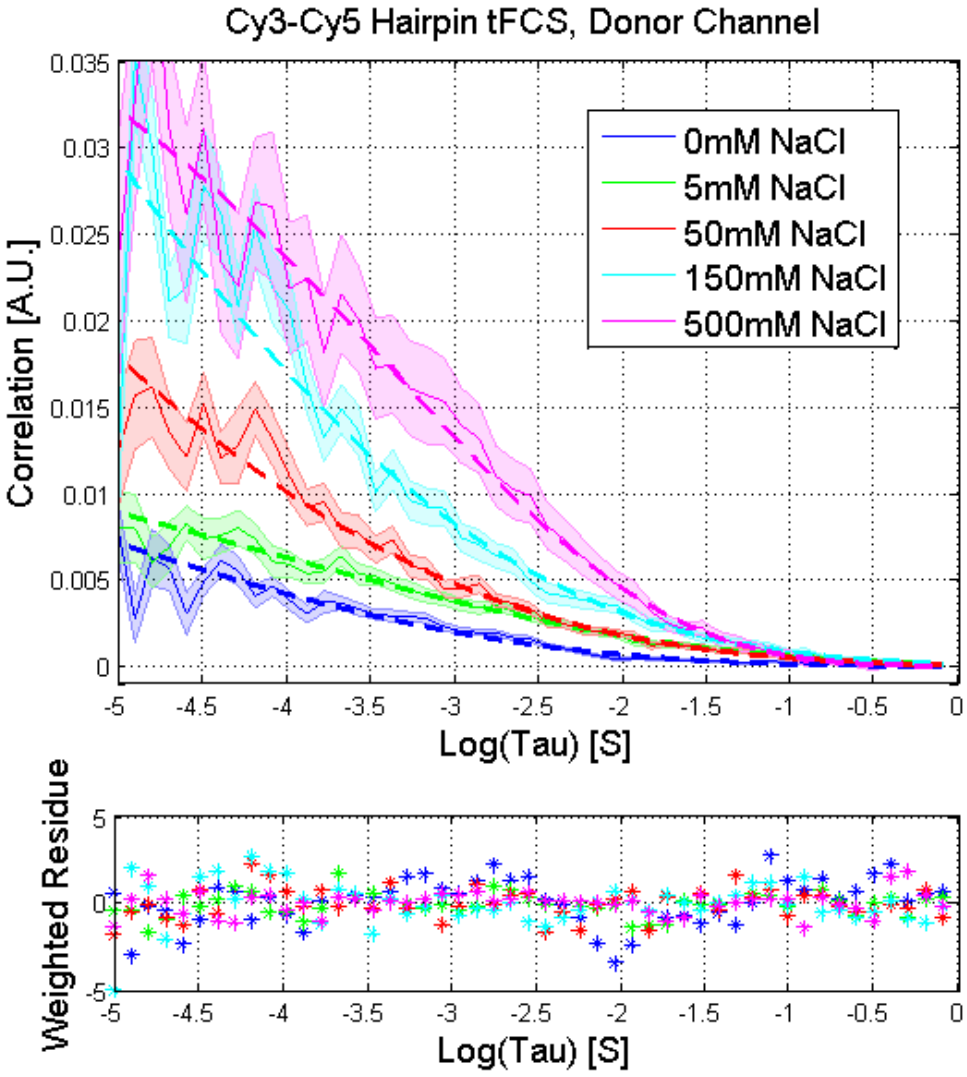

Figure 4.23: The extracted folding dynamics from donor channels at different salt concentration. The triangular waves from the beam modulation is not a dominant feature on this point, and we can see that in addition to the amplitude increase at higher salt concentration, the FCS curves also illustrate that at higher salt concentration, the dynamics of the folding becomes slower. For example, the 10^{-2} S dynamics was not evident in the Na0, but is very apparent in any other solution condition. This behavior is what we expect at higher salt concentration solution, in which the hairpin structure is more stable.

with what we had seen before in the solution case, although the dynamics we see here is again orders of magnitude slower than what we had observed by solution FCS.

Finally, we can extract the folding dynamics from acceptor autocorrelation (red traces) and donor-acceptor cross-correlation (black traces). The results are shown in Fig.4.24. One important feature we can see here is the negative correlation between the donor and acceptor signals. What we have seen in the donor case can still be artifacts: perhaps the tracking system behaves very differently at different salt concentrations (although we showed that tracking behaves the same by control samples in different concentrations, but the molecular samples are not exactly the same) and contributes to what we see in Fig.4.23. However, effects like such will not result into anti-correlation. Therefore the negative-correlation in the donor-acceptor anti-correlation function is a strong evidence that we have observed folding of some sort.

In the next section, we will discussion the discrepancies between our solution data and tracking data. We will combine what we have seen in the dye-quencher and FRET case and propose a possible cause of this.

4.5 Track and Observe Folding Reaction

In this chapter, we have demonstrated the first study of tracking a molecular complex and observing the folding of which by dye-quencher and FRET probe systems. After introducing the basics of tracking apparatus, we dived into theories of how to accurately extract the folding dynamics out of a group of systematic noises, such as tracking contributed dynamics, multiple FRETs on one tracked complex, and imperfect labeling. With two labeling strategies and vary the folding energy landscape by performing the experiment in eight different salt concentration solutions, we obtained comprehensive data set on over 1,000 tracking sessions. We have shown examples of tracking sessions, which clearly illustrate dye blinking, step-wise bleaching, and energy transferring between FRET donor and acceptor dyes. Then we applied the theories

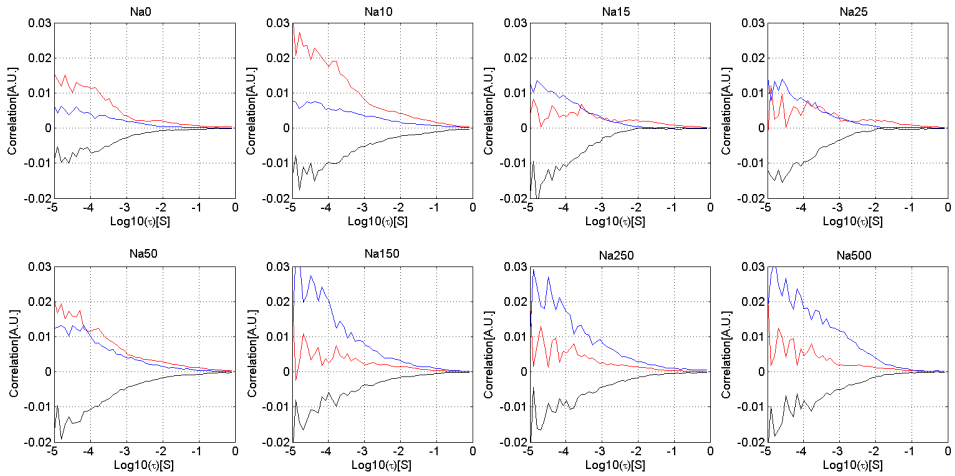

Figure 4.24: Folding dynamics from donor-donor (blue), acceptor-acceptor (red), acceptor-donor (black) correlation. The negative correlation between the donor and acceptor is a strong evidence that we have observed energy transferring and thus folding of the molecule.

that we have developed and show that we can successfully extract the folding FCS. Several features of the FCS curves demonstrates that what we have observed is not artifacts, but folding of the molecule on the transparent bead. These features includes the dramatic difference between the control and sample sequence, and the negative correlation between the donor and acceptor fluorescence. Modeling the extracted folding dynamics, however, the time scales of these folding dynamics are at least one order of magnitude slower than our solution result in the previous chapter. Now we will discuss this discrepancy.

There is overwhelming evidence that our solution measurement has observed the folding of DNA hairpins. This is confirmed with more than a dozen constructs with different labeling and measurement conditions. Tracking experiments also demonstrated that we have observed folding of some sort, however we did not recover the same dynamics that we have observed in the solution case (for example, the two-state exponential relaxation around 100 nS for the dye-quencher). One possible source of discrepancies here is that the bead surface is interacting with the hairpin. The NeutrAvidin coating of the bead can alter the local environment of the hairpin and

make the hairpin folds much slower than in the solution case. This result, although disappointing, demonstrates that tracking still needs improvement for it to fully unleash its power in observing single molecule folding. The current development in the lab, such as using AOMs to defect the tracking beam and simplifying the beam modulation scheme, can really turn tracking into a powerful tool in monitoring folding of biological molecules in their native conditions. The future is indeed exciting for working in the Mabuchi Lab!

Appendix A

Error of FCS

A.1 Introduction

An adequate analysis of any FCS data requires the knowledge about the statistical accuracy of the measured data. Throughout the thesis, we estimated the error by preparing identical experiments and compute the sample standard deviation as our estimation of error of the mean value. However, in a tracking experiments, sometimes one can observe one molecule for extended periods of time and should have enough data to resolve the dynamics of one molecule at a time. Given that we want to observe molecule to molecule differences by FCS, how can we estimate the error in this case? This note is to document how to calculate error in such case.

This error estimation calculations in FCS in general is quite difficult, and sometimes theoretically impossible - for example the exact theoretical expression of FCS error in the case of three-dimensional Gaussian excitation profile does not exist due to non-converging integrals. Many papers on this subject used approximation methods to estimate error in FCS measurement, such as [Kop74, Qia90, KGA97, WRV01].

[WRV01] outlined three methods of estimating error of FCS directly from data. Although in the experiments presented in this thesis we all used One of their conclusions is that one can directly use fluorescence data to estimate noise in FCS, and this method gave reasonable fitting results. This note first outline how FCS computation

is done in physical correlators; how this multi-tau method is theoretically equivalent to the Laurence method[LFH06](this is the method we use in the experiment); how we estimate error in FCS in general; finally give the computation method to calculate the error.

A.2 Noise of in a FCS Measurement

We follow the definition of FCS in [WRV01]: $G_2(\tau) =< I(t)I(t+\tau) > / < I(t) >^2$ (this is the one without minus 1.)

$$G_2(m\Delta\tau) = \frac{Y}{M_{dir}M_{del}} \tag{A.1}$$

with

$$\begin{aligned} Y =& < n(k\Delta\tau)n(k\Delta\tau + m\Delta\tau) > \\ =& \frac{1}{M-m}\sum_{k=1}^{M-m}[n(k\Delta\tau)n(k\Delta\tau + m\Delta\tau)] \end{aligned}$$

also with definition of

$$\begin{aligned} M_{dir} =& < n(k\Delta\tau) > \\ =& \frac{1}{M-m}\sum_{k=m}^{M} n(k\Delta\tau) \end{aligned}$$

$$\begin{aligned} M_{del} =& < n(k\Delta\tau) > \\ =& \frac{1}{M-m}\sum_{k=1}^{M-m} n(k\Delta\tau) \end{aligned}$$

This expression is symmetrically normalized to give good statistics[SDS88]. For later analysis, we will approximate the measurement noise of n's by its square root (as if it is Poission distributed). Note this is the exact form of autocorrelation functions for discrete signals.

Notice that this expression for the FCS is done by calculating the mean of a set of random numbers: $y_k = n(k\Delta\tau)n(k\Delta\tau + m\Delta\tau)$. If one asks how much the measured G (which is also averaged) is deviated from the actual value (what is your estimated error given this set of measurement), we can answer this by looking at the standard deviation of the mean of y_k. Therefore, what we need to do is calculate the standard deviation of sample set y_k and then scale it to get the standard deviation of its mean value.
A minor point is that $n(k\Delta\tau)n(k\Delta\tau + m\Delta\tau)$ is correlated over time, which makes using their std as estimation of error unjustified. But we still use their std as an approximation of error.
So how do we calculate error associated with y_k's? One natural way to do this is first binning the photon arriving events by $\Delta\tau$, record all the n 's, perfrom the calculation of $n(k\Delta\tau)n(k\Delta\tau + m\Delta\tau)$, and then calculate the standard deviation of this set of random variables. The first three steps of this calculation is performed in the so-called multi-tau algorithm developed by Schatzel in his correlator papers, and is the algorithm in those physical correlators. To perfrom the last step of this calculations in the multi-tau algorithm, it is super simple:

$$Var[n(k\Delta\tau)n(k\Delta\tau+m\Delta\tau)] = E[\{n(k\Delta\tau)n(k\Delta\tau+m\Delta\tau)\}^2]-\{E[n(k\Delta\tau)n(k\Delta\tau+m\Delta\tau)]\}^2$$

Since in the multi-tau algorithm you do record all the $n(k\Delta\tau)n(k\Delta\tau + m\Delta\tau)$ one by one and then calculate $E[n(k\Delta\tau)n(k\Delta\tau+m\Delta\tau)]$. Thus it is really trivial to calculate $E[\{nn\}^2]$ as well as $E[nn]^2$.

However this is quite different in our current Laurence Method.

Simply put: all the n's are just not accessible by Laurence Method. The Laurence method calcualte the average FCS curve by counting photons and photon paris. Its normalization is done by dividing the whole time. See Eq.A.3,A.4,A.5 for details.

The Laurence Method does have the advantage of using any tau vectors you specify, whereas in multi-tau algorithm, you have the tau vector in some strange way that can complicated your calculation later on. Specifically, the multi-tau algorithm will break the tau vectors into several groups, with each group the same tau bin. And tau bin window increase a factor of two after each group. It should be noted that the Laurence's method of calculating the averaged FCS curves is equivalent to the multi-tau method. However to fundamentally understand what FCS really is, we should use the definition in Eq.A.1 - we do calculate $< I(t)I(t+\tau) >$, but perform the actual calculation using discrete definition autocorrelation functions.

So given our current Laurence method of calculating FCS, how do we calculate $Var[n1, n2]$? ($n1 = n(k\Delta\tau)$ and $n2 = n(k\Delta\tau + m\Delta\tau)$)

We will use propagation of error to estimate the error in our measured mean of y_k:

$$\begin{aligned}\sigma^2_{y_k} =& \bar{n2}^2\sigma^2_{n1} + \bar{n1}^2\sigma^2_{n2} + 2\bar{n1}\bar{n2}Cov[n1, n2] \\ =& \bar{n2}^2\bar{n1} + \bar{n1}^2\bar{n2} + 2\bar{n1}\bar{n2}Cov[n1, n2]\end{aligned}$$

in which variances are replaced by averages. If $n1$ and $n2$ is not correlated, the above expression reduces to the familiar form of error propagation formular. To calculate

$Cov[n1, n2]$, we use its definition:

$$\begin{aligned} Cov[n1, n2] =& E[(n1 - \bar{n1})(n2 - \bar{n2})] \\ =& E[n1n2 - n1\bar{n2} - \bar{n1}n2 + \bar{n1}\bar{n2}] \\ =& E[n1n2] - \bar{n1}\bar{n2} \\ =& Y - \bar{n1}\bar{n2} \end{aligned}$$

Therefore, we have:

$$\begin{aligned} \sigma^2_{y_k} =& \bar{n2}^2\bar{n1} + \bar{n1}^2\bar{n2} + 2\bar{n1}\bar{n2}(Y - \bar{n1}\bar{n2}) \\ =& \bar{n1}\bar{n2}\{\bar{n1} + \bar{n2} + 2(Y - \bar{n1}\bar{n2})\} \\ =& M_{dir}M_{del}\{M_{dir} + M_{del} + 2(Y - M_{dir}M_{del})\} \end{aligned} \tag{A.2}$$

All the above quantities are calculated in the Laurence method.
For example, to calculate M_{dir}:

$$M_{dir} = \frac{N[u_i \geq \tau]}{(T - \tau)/\Delta\tau} \tag{A.3}$$

in which N is the operater that counts the photon tag stream t_i for photons that have tags equal or bigger than τ. Similarly, M_{del} is:

$$M_{del} = \frac{N[t_i \leq (T - \tau)]}{(T - \tau)/\Delta\tau} \tag{A.4}$$

And Y is calcualted by:

$$Y = \frac{N[(\tau - \Delta\tau) \leq (u_i - t_i) \leq (\tau + \Delta\tau)]}{(T - \tau)/\Delta\tau} \tag{A.5}$$

Therefore,

$$\begin{aligned} G_2(\tau) =& \frac{Y}{M_{dir}M_{del}} \\ =& \frac{N[(\tau-\Delta\tau) \le (u_i - t_i) \le (\tau+\Delta\tau)](T-\tau)}{N[u_i \ge \tau]N[t_i \le (T-\tau)]\Delta\tau} \end{aligned}$$

Laurence Method is very efficient in calculating $N[(\tau-\Delta\tau) \le (u_i - t_i) \le (\tau+\Delta\tau)]$. This algorithm is the current FCS calculator.

σ_{y_k} is the standard deviation of random variables y_k. Now we need the standard deviation of Y, which is the average of y_k. And,

$$\sigma_Y = \frac{\sigma_{y_k}}{\sqrt{(T-\tau)/\Delta\tau}}$$

Finally, we have:

$$\begin{aligned} \sigma^2_{G_2(\tau)} \backsimeq& \frac{\sigma^2_Y}{(M_{dir}M_{del})^2} \\ \backsimeq& \frac{\{M_{dir}+M_{del}+2(Y-M_{dir}M_{del})\}}{M_{dir}M_{del}(T-\tau)/\Delta\tau} \\ =& \frac{\frac{nb}{m}+\frac{na}{m}+2(\frac{g\_tmp}{m}-\frac{na\cdot nb}{m^2})}{\frac{nb}{m}\cdot\frac{na}{m}\cdot m} \\ =& \frac{na+nb+2(g\_tmp-\frac{na\cdot nb}{m})}{na\cdot nb} \end{aligned}$$

Bibliography

[AK05] Anjum Ansari and Serguei V. Kuznetsov. Is hairpin formation in single-stranded polynucleotide diffusion-controlled? *The Journal of Physical Chemistry B*, 109(26):12982–12989, July 2005. URL: http://dx.doi.org/10.1021/jp044838a. 17, 21

[AKS01] Anjum Ansari, Serguei V. Kuznetsov, and Yiqing Shen. Configurational diffusion down a folding funnel describes the dynamics of dna hairpins. *Proceedings of the National Academy of Sciences of the United States of America*, 98(14):7771–7776, 2001. URL: http://www.pnas.org/content/98/14/7771.abstract, doi:10.1073/pnas.131477798. 12, 17, 19, 21, 66

[AMP08] Colin Echeverra Aitken, R. Andrew Marshall, and Joseph D. Puglisi. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal*, 94(5):1826–1835, March 2008. URL: http://www.sciencedirect.com/science/article/B94RW-4V0KNJF-17/2/016974d248b0a017aae3e31e585ae062. 41, 121

[BB08] Philip C. Bevilacqua and Joshua M. Blose. Structures, kinetics, thermodynamics, and biological functions of rna hairpins. *Annual Review of Physical Chemistry*, 59(1):79–103, 2008. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.59.032607.093743, doi:10.1146/annurev.physchem.59.032607.093743. iv, 11, 12, 13

[Ber06] Andrew J. Berglund. *Feedback Control of Brownian Motion for Single-Particle Fluorescence Spectroscopy*. PhD thesis, California Institute of Technology, 2006. URL: http://resolver.caltech.edu/CaltechETD:etd-10092006-165831. 85, 96

[BFY+11] Samuel Bockenhauer, Alexandre Furstenberg, Xiao Jie Yao, Brian K. Kobilka, and W. E. Moerner. Conformational dynamics of single g protein-coupled receptors in solution. *J. Phys. Chem. B*, 115(45):13328–13338, September 2011. URL: http://dx.doi.org/10.1021/jp204843r, doi:10.1021/jp204843r. 3

[BKL98] Gregoire Bonnet, Oleg Krichevsky, and Albert Libchaber. Kinetics of conformational fluctuations in dna hairpin-loops. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8602–8606, July 1998. URL: http://www.pnas.org/content/95/15/8602.abstract. 10, 16, 17, 18, 21, 25

[BLBM10] David Bikard, C??line Loot, Zeynep Baharoglu, and Didier Mazel. Folded dna in action: Hairpin formation and biological functions in prokaryotes. *Microbiology and Molecular Biology Reviews*, 74(4):570–588, December 2010. URL: http://mmbr.asm.org/content/74/4/570.abstract, doi:10.1128/MMBR.00026-10. iv, 14, 15

[BMA+10] Jillian G Baker, Richard Middleton, Luke Adams, Lauren T May, Stephen J Briddon, Barrie Kellam, and Stephen J Hill. Influence of fluorophore and linker composition on the pharmacology of fluorescent adenosine a1 receptor ligands. *British Journal of Pharmacology*, 159(4):772–786, 2010. URL: http://dx.doi.org/10.1111/j.1476-5381.2009.00488.x. 5

[BMM07] Andrew J. Berglund, Kevin McHale, and Hideo Mabuchi. Fluctuations in closed-loop fluorescent particle tracking. *Opt. Express*, 15(12):7752–7773, June 2007. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-15-12-7752. 85, 100, 113

[Cri70] F. Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970. 14

[DML08] Ashok A Deniz, Samrat Mukhopadhyay, and Edward A Lemke. Single-molecule biophysics: at the interface of biology, physics and chemistry. *Journal of The Royal Society Interface*, 5(18):15–45, 2008. URL: http://rsif.royalsocietypublishing.org/content/5/18/15.abstract, doi:10.1098/rsif.2007.1021. 3

[Dob03] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, December 2003. URL: http://dx.doi.org/10.1038/nature02261. 2

[EFG+09] John Eid, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, David Rank, Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, Alex deWinter, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangxu Kong, Ronald Kuse, Yves Lacroix, Steven Lin, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Maxham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomaney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach, and Stephen Turner. Real-time dna sequencing from single polymerase molecules. *Science*, 323(5910):133–138, January 2009. URL: http://www.sciencemag.org/content/323/5910/133.abstract. 1

[GBKL00] Noel L. Goddard, Gregoire;goire Bonnet, Oleg Krichevsky, and Albert Libchaber. Sequence dependent rigidity of single stranded dna. *Phys. Rev. Lett.*, 85(11):2400–, September 2000. URL: http://link.aps.org/doi/10.1103/PhysRevLett.85.2400. 12, 21

[GGL+01] Jocelyn R. Grunwell, Jennifer L. Glass, Thilo D. Lacoste, Ashok A. Deniz, Daniel S. Chemla, and Peter G. Schultz. Monitoring the conformational fluctuations of dna hairpins using single-pair fluorescence resonance energy transfer. *Journal of the American Chemical Society*, 123(18):4295–4303, May 2001. URL: http://dx.doi.org/10.1021/ja0027620. 18, 21, 82

[GWAB05] William J. Greenleaf, Michael T. Woodside, Elio A. Abbondanzieri, and Steven M. Block. Passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.*, 95(20):208102–, November 2005. URL: http://link.aps.org/doi/10.1103/PhysRevLett.95.208102. 18

[HKPS03] Thomas Heinlein, Jens-Peter Knemeyer, Oliver Piestert, and Markus Sauer. Photoinduced electron transfer between fluorescent dyes and guanosine residues in dna-hairpins. *The Journal of Physical Chemistry B*, 107(31):7957–7964, August 2003. URL: http://dx.doi.org/10.1021/jp0348068. 36, 48

[HL88] MS Horwitz and LA Loeb. An e. coli promoter that regulates transcription by dna superhelix-induced cruciform extrusion. *Science*, 241(4866):703–705, 1988. URL: http://www.sciencemag.org/content/241/4866/703.abstract, doi:10.1126/science.2456617. 13

[HTL+99] Taekjip Ha, Alice Y. Ting, Joy Liang, Daniel S. Chemla, Peter G. Schultz, Shimon Weiss, and Ashok A. Deniz. Temporal fluctuations of fluorescence resonance energy transfer between two dyes conjugated to a single protein. *Chemical Physics*, 247(1):107–118, August 1999. URL: http://www.sciencedirect.com/science/article/B6TFM-3X29GBX-C/2/1d06a5617d95f9dbd8cad76acac6bb85. 52

[JBI+08] Chirlmin Joo, Hamza Balci, Yuji Ishitsuka, Chittanon Buranachai, and Taekjip Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.*, 77(1):51–76, June 2008. URL: http://dx.doi.org/10.1146/annurev.biochem.77.070606.101543. 1

[JIS+08] Jaemyeong Jung, Rachelle Ihly, Eric Scott, Ming Yu, and Alan Van Orden. Probing the complete folding trajectory of a dna hairpin using dual beam fluorescence fluctuation spectroscopy. *The Journal of Physical Chemistry B*, 112(1):127–133, January 2008. URL: http://dx.doi.org/10.1021/jp076248t. 18, 21

[JVO06] Jaemyeong Jung and Alan Van Orden. A three-state mechanism for dna hairpin folding characterized by multiparameter fluorescence fluctuation spectroscopy. *Journal of the American Chemical Society*, 128(4):1240–1249, 2006. URL: http://pubs.acs.org/doi/abs/10.1021/ja0560736, arXiv:http://pubs.acs.org/doi/pdf/10.1021/ja0560736, doi:10.1021/ja0560736. 17, 18, 19, 21

[KB02] Oleg Krichevsky and Gregoire Bonnet. Fluorescence correlation spectroscopy: the technique and its applications. *Reports on Progress in Physics*, 65(2):251, 2002. URL: http://stacks.iop.org/0034-4885/65/i=2/a=203. 25

[KDNS06] Jiho Kim, Soren Doose, Hannes Neuweiler, and Markus Sauer. The initial step of dna hairpin folding: a kinetic analysis using fluorescence correlation spectroscopy. *Nucleic Acids Research*, 34(9):2516–2527, 2006. URL: http://nar.oxfordjournals.org/content/34/9/2516.abstract, doi:10.1093/nar/gkl221. 16, 21, 48, 52, 66

[KGA97] Peet Kask, R. G??nther, and Peter Axhausen. Statistical accuracy in fluorescence fluctuation experiments. 25(3):163–169–, 1997. URL: http://dx.doi.org/10.1007/s002490050028. 134

[KMS00] Jens-Peter Knemeyer, Nicole Marme, and Markus Sauer. Probes for detection of specific dna sequences at the single-molecule level. *Analytical Chemistry*, 72(16):3717–3724, August 2000. URL: http://dx.doi.org/10.1021/ac000024o. 4, 36, 48, 52

[Kop74] Dennis E. Koppel. Statistical accuracy in fluorescence correlation spectroscopy. *Phys. Rev. A*, 10(6):1938–1945, December 1974. URL: http://link.aps.org/doi/10.1103/PhysRevA.10.1938. 134

[KRWA08] Serguei V. Kuznetsov, Cha-Chi Ren, Sarah A. Woodson, and Anjum Ansari. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Research*, 36(4):1098–1112, 2008. URL: http://nar.oxfordjournals.org/content/36/4/1098.abstract, doi:10.1093/nar/gkm1083. 16, 17

[KSBA01] Serguei V. Kuznetsov, Yiqing Shen, Albert S. Benight, and Anjum Ansari. A semiflexible polymer model applied to loop formation in dna hairpins. *Biophysical Journal*, 81(5):2864–2875, November 2001. URL: http://www.sciencedirect.com/science/article/B94RW-4TXGN6H-1B/2/e5baf606191b193a1538545079b37501. 16

[Lev68] Cyrus Levinthal. Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65:44–45, 1968. 2

[Lev10] Marcia Levitus. Relaxation kinetics by fluorescence correlation spectroscopy: Determination of kinetic parameters in the presence of fluorescent impurities. *The Journal of Physical Chemistry Letters*, 1(9):1346–1350, May 2010. URL: http://dx.doi.org/10.1021/jz100231v. 25

[LFH06] TA Laurence, S Fore, and T Huser. Fast, flexible algorithm for calculating photon correlations. *Opt Lett*, 31(6):829–831–, March 2006. URL: http://europepmc.org/abstract/MED/16544638. 51, 135

[LMSZ08] Milo M. Lin, Lars Meinhold, Dmitry Shorokhov, and Ahmed H. Zewail. Unfolding and melting of dna (rna) hairpins: the concept of structure-specific 2d dynamic landscapes. *Phys. Chem. Chem. Phys.*, 10(29):4227–4239, 2008. URL: http://dx.doi.org/10.1039/B804675C. 12

[MBM07] Kevin McHale, Andrew J. Berglund, and Hideo Mabuchi. Quantum dot photon statistics measured by three-dimensional particle tracking. *Nano Lett.*, 7(11):3535–3539, October 2007. URL: http://dx.doi.org/10.1021/nl0723376. 85

[McH08] Kevin L. McHale. *Feedback tracking and correlation spectroscopy of fluorescent nanoparticles and biomolecules*. PhD thesis, California Institute of Technology, 2008. URL: http://resolver.caltech.edu/CaltechETD:etd-05072008-204627. 85, 89, 92

[MM09] Kevin McHale and Hideo Mabuchi. Precise characterization of the conformation fluctuations of freely diffusing dna: Beyond rouse and zimm. *Journal of the American Chemical Society*, 131(49):17901–17907, December 2009. URL: http://dx.doi.org/10.1021/ja906979j. 49, 50, 85

[MM10] Kevin McHale and Hideo Mabuchi. Intramolecular fluorescence correlation spectroscopy in a feedback tracking microscope. *Biophysical Journal*, 99(1):313–322, July 2010. URL: http://www.sciencedirect.com/science/article/B94RW-50G667C-1C/2/cc2c2c0478e512495acf8a51301f02ac. 92, 96

[MN08] Mathieu G. McPhie and Gerhard Nagele. Nonmonotonic density dependence of the diffusion of dna fragments in low-salt suspensions. *Phys. Rev. E*, 78(6):060401–, December 2008. URL: http://link.aps.org/doi/10.1103/PhysRevE.78.060401. 58

[MP09] Dmitrii E. Makarov and Kevin W. Plaxco. Measuring distances within unfolded biopolymers using fluorescence resonance energy transfer: The

effect of polymer chain dynamics on the observed fluorescence resonance energy transfer efficiency. *The Journal of Chemical Physics*, 131(8):085105, 2009. URL: http://dx.doi.org/10.1063/1.3212602. 20

[MPK+06] Hairong Ma, David J. Proctor, Elzbieta Kierzek, Ryszard Kierzek, Philip C. Bevilacqua, and Martin Gruebele. Exploring the energy landscape of a small rna hairpin. *Journal of the American Chemical Society*, 128(5):1523–1530, 2006. URL: http://pubs.acs.org/doi/abs/10.1021/ja0553856, arXiv:http://pubs.acs.org/doi/pdf/10.1021/ja0553856, doi:10.1021/ja0553856. 16

[MRC+04] M.C. Murphy, Ivan Rasnik, Wei Cheng, Timothy M. Lohman, and Taekjip Ha. Probing single-stranded dna configurational flexibility using fluorescence spectroscopy. *Biophysics Journal*, 86(4):2530–2537, 2004. 12, 20

[MWWZ07] Hairong Ma, Chaozhi Wan, Aiguo Wu, and Ahmed H. Zewail. Dna folding and melting observed in real time redefine the energy landscape. *Proceedings of the National Academy of Sciences*, 104(3):712–716, 2007. URL: http://www.pnas.org/content/104/3/712.abstract, doi:10.1073/pnas.0610028104. 16

[OJ08] Alan Van Orden and Jaemyeong Jung. Review fluorescence correlation spectroscopy for probing the kinetics and mechanisms of dna hairpin formation. *Biopolymers*, 89(1):1–16, 2008. URL: http://dx.doi.org/10.1002/bip.20826. iv, 18

[OSG12] Dylan M. Owen, Markus Sauer, and Katharina Gaus. Fluorescence localization microscopy: The transition from concept to biological research tool. *cib*, 5(1942-0889):345–349, July 2012. URL: http://www.landesbioscience.com/journals/cib/article/20348/. 48, 52

[Qia90] Hong Qian. On the statistics of fluorescence correlation spectroscopy. *Biophysical Chemistry*, 38(1??2):49–57, October 1990. URL: http://www.sciencedirect.com/science/article/pii/030146229080039A. 134

[QYL+10] Peng Qu, Xinxing Yang, Xun Li, Xiaoxue Zhou, and Xin Sheng Zhao. Direct measurement of the rates and barriers on forward and reverse diffusions of intramolecular collision in overhang oligonucleotides. *J. Phys. Chem. B*, 114(24):8235–8243, May 2010. URL: http://dx.doi.org/10.1021/jp101173y. 20

[RHH08] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule fret. *Nat Meth*, 5(6):507–516, June 2008. URL: http://dx.doi.org/10.1038/nmeth.1208. 4, 40

[RMH06] Ivan Rasnik, Sean A McKinney, and Taekjip Ha. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Meth*, 3(11):891–893, November 2006. URL: http://dx.doi.org/10.1038/nmeth934. 40, 48

[SBRN06] X. Schlagberger, J. Bayer, J. O. R?dler, and R. R. Netz. Diffusion of a single semiflexible charged polymer. *EPL (Europhysics Letters)*, 76(2):346, 2006. URL: http://stacks.iop.org/0295-5075/76/i=2/a=346. 58

[SC06] P. Svoboda and A.Di. Cara. Hairpin rna: a secondary structure of primary importance. 63(7-8):901–908–, 2006. URL: http://dx.doi.org/10.1007/s00018-005-5558-5. iv, 5, 13

[SDS88] Klaus Schatzel, Martin Drewel, and Sven Stimac. Photon correlation measurements at large lag times: Improving statistical accuracy. *Journal of Modern Optics*, 35(4):711–718, April 1988. URL: http://www.tandfonline.com/doi/abs/10.1080/09500348814550731. 136

[SEM05] Chandran R. Sabanayagam, John S. Eid, and Amit Meller. Using fluorescence resonance energy transfer to measure distances along individual dna molecules: Corrections due to nonideal transfer. *The Journal of Chemical Physics*, 122(6):061103, 2005. URL: http://dx.doi.org/10.1063/1.1854120. 40

[SGCH10] Sergey V. Solomatin, Max Greenfeld, Steven Chu, and Daniel Herschlag. Multiple native states reveal persistent ruggedness of an rna folding landscape. *Nature*, 463(7281):681–684, February 2010. URL: http://dx.doi.org/10.1038/nature08717. 4

[SH04] John SantaLucia and Donald Hicks. The thermodynamics of dna structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33(1):415–440, 2004. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.32.110601.141800, doi:10.1146/annurev.biophys.32.110601.141800. 12, 16

[SH09] S Solomatin and D. Herschlag. Methods of site-specific labeling of rna with fluorescent dyes. *Methods in Enzymology*, 469:47–67, 2009. 42

[SR98] M.C.Ramachandra Shastry and Heinrich Roder. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nat Struct Mol Biol*, 5(5):385–392, May 1998. URL: http://dx.doi.org/10.1038/nsb0598-385. 2

[SRNP03] Eric J. Sorin, Young Min Rhee, Bradley J. Nakatani, and Vijay S. Pande. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophysical Journal*, 85(2):790–803, August 2003. URL: http://www.sciencedirect.com/science/article/B94RW-4V3HBYY-B/2/b75504c6cc4c95caf5970349cc6644b2. 16

[SRP05] Eric J. Sorin, Young Min Rhee, and Vijay S. Pande. Does water play a structural role in the folding of small nucleic acids?,

April 2005. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349505733063. 16

[TBRB03] Andrew Tsourkas, Mark A Behlke, Scott D. Rose, and Gang Bao. Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Research*, 31(4):1319–1330, 2003. 15

[TC08a] Zhi-Jie Tan and Shi-Jie Chen. Electrostatic free energy landscapes for dna helix bending, April 2008. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349508704703. 20

[TC08b] Zhi-Jie Tan and Shi-Jie Chen. Salt dependence of nucleic acid hairpin stability, July 2008. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349508702492. 20

[TL07] Tedman Torres and Marcia Levitus. Measuring conformational dynamics:a new fcs-fret approach. *The Journal of Physical Chemistry B*, 111(25):7392–7400, June 2007. URL: http://dx.doi.org/10.1021/jp070659s. 25

[TM08] Christina L. Ting and Dmitrii E. Makarov. Two-dimensional fluorescence resonance energy transfer as a probe for protein folding: A theoretical study. *The Journal of Chemical Physics*, 128(11):115102, 2008. URL: http://link.aip.org/link/?JCP/128/115102/1, doi:10.1063/1.2835611. 5

[Var95] G Varani. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.*, 24(1):379–404, June 1995. URL: http://dx.doi.org/10.1146/annurev.bb.24.060195.002115. 12, 14

[VCF+09] Jan Vogelsang, Thorben Cordes, Carsten Forthmann, Christian Steinhauer, and Philip Tinnefeld. Controlling the fluorescence of ordinary oxazine dyes for single-molecule switching and superresolution microscopy. *Proceedings of the National Academy of Sciences*, 106(20):8107–8112,

2009. URL: http://www.pnas.org/content/106/20/8107.abstract, doi:10.1073/pnas.0811875106. 40, 48, 52

[Wad00] Randy M. Wadkins. Targeting dna secondary structures. *Current Medical Chemistry*, 7:1–15, 2000. URL: http://www2.uah.es/farmamol/Public/Curr_Med_Chem/DNA_2ndstr_CMC2000.pdf. 5, 15

[WBPL+06] Michael T. Woodside, William M. Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences*, 103(16):6190–6195, 2006. URL: http://www.pnas.org/content/103/16/6190.abstract, doi:10.1073/pnas.0511048103. 10, 16, 18, 21, 82

[WF74] Gerald Wilemski and Marshall Fixman. Diffusion-controlled intrachain reactions of polymers. ii results for a pair of terminal reactive groups. *The Journal of Chemical Physics*, 60(3):878–890, 1974. URL: http://dx.doi.org/10.1063/1.1681163. 20

[WGGB08] Michael T Woodside, Cuauhtmoc Garca-Garca, and Steven M Block. Folding and unfolding single rna molecules under tension. *Current Opinion in Chemical Biology*, 12(6):640–646, December 2008. URL: http://www.sciencedirect.com/science/article/B6VRX-4TDBD62-3/2/a14a2a911961c9a966cb0d82a121b65a. 16, 18, 25

[WMR95] Jerker Widengren, Uelo Mets, and Rudolf Rigler. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. *J. Phys. Chem.*, 99(36):13368–13379, September 1995. URL: http://dx.doi.org/10.1021/j100036a009. 41

[WN03] Xiaojuan Wang and Werner M. Nau. Kinetics of end-to-end collision in short single-stranded nucleic acids. *J. Am. Chem. Soc.*, 126(3):808–813, December 2003. URL: http://dx.doi.org/10.1021/ja038263r. 20

[Woe83] Gutell R. Gupta R. & Noller H. Woese, C. R. Detailed analysis of the higher-order structure of 16s-like ribosomal ribonucleic acids. *Microbiol. Rev.*, 47:621–669, 1983. 13, 14

[WRV01] Thorsten Wohland, Rudolf Rigler, and Horst Vogel. The standard deviation in fluorescence correlation spectroscopy. *Biophysical Journal*, 80(6):2987–2999, June 2001. URL: http://www.sciencedirect.com/science/article/pii/S0006349501762649. 134, 135

[WSG+08] Anna K. Wozniak, Gunnar F. Schroder, Helmut Grubmuller, Claus A. M. Seidel, and Filipp Oesterhelt. Single-molecule fret measures bends and kinks in dna. *Proceedings of the National Academy of Sciences*, 105(47):18337–18342, 2008. URL: http://www.pnas.org/content/105/47/18337.abstract, doi:10.1073/pnas.0800977105. 39

[WvH95] K S Wilson and P H von Hippel. Transcription termination at intrinsic terminators: the role of the rna hairpin. *Proceedings of the National Academy of Sciences*, 92(19):8793–8797, 1995. URL: http://www.pnas.org/content/92/19/8793.abstract. 14

[WYBK00] Mark Ian Wallace, Liming Ying, Shankar Balasubramanian, and David Klenerman. Fret fluctuation spectroscopy: Exploring the conformational dynamics of a dna hairpin loop. *The Journal of Physical Chemistry B*, 104(48):11551–11555, December 2000. URL: http://dx.doi.org/10.1021/jp001560n. 16, 21, 25

[WYBK01] Mark I. Wallace, Liming Ying, Shankar Balasubramanian, and David Klenerman. Non-arrhenius kinetics for the loop closure of a dna hairpin. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10):5584–5589, 2001. URL: http://www.pnas.org/content/98/10/5584.abstract, doi:10.1073/pnas.101523498. 16

[ZC02] Wenbing Zhang and Shi-Jie Chen. Rna hairpin-folding kinetics. *Proceedings of the National Academy of Sciences of the United States of America*, 99(4):1931–1936, 2002. URL: http://www.pnas.org/content/99/4/1931.abstract, doi:10.1073/pnas.032443099. 16

[ZC06] Wenbing Zhang and Shi-Jie Chen. Exploring the complex folding kinetics of rna hairpins: I. general folding kinetics analysis, February 2006. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349506722652. 16