DISSERTATION

PROBING MOLECULAR KINETICS USING HIGHER-ORDER FLUORESCENCE CORRELATION SPECTROSCOPY

Submitted by

Farshad Abdollah-Nia

Department of Physics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2019

Doctoral Committee:

Advisor: Martin P. Gelfand
Co-Advisor: Alan Van Orden

Diego Krapf
Ashok Prasad
Jacob Roberts

ABSTRACT

PROBING MOLECULAR KINETICS USING HIGHER-ORDER FLUORESCENCE CORRELATION SPECTROSCOPY

Fluorescence correlation spectroscopy (FCS) is a powerful tool in the time-resolved analysis of non-reacting or reacting molecules in solution, based on fluorescence intensity fluctuations. However, conventional (second-order) FCS alone is insufficient to measure all parameters needed to describe a reaction or mixture, including concentrations, fluorescence brightnesses, and forward and reverse rate constants. For this purpose, correlations of higher powers of fluorescence intensity fluctuations can be calculated to yield additional information from the single-photon data stream collected in an FCS experiment. To describe systems of diffusing and reacting molecules, considering cumulants of fluorescence intensity results in simple expressions in which the reaction and the diffusion parts factorize. The computation of higher-order correlations in experiments is hindered by shot noise and common detector artifacts, the effects of which become worse with increasing order. We introduce a technique to calculate artifact-free higher-order correlation functions with improved time resolution, and without any need for modeling and calibration of detector artifacts. The technique is formulated for general multi-detector experiments and verified in both two-detector and single-detector configurations. Good signal-to-noise ratio is achieved down to $1\,\mu$s in correlation curves up to order $(2, 2)$.

Next, we demonstrate applications of the technique to analyze systems of fast and slow reactions. As an example of slow- or non-reacting systems, the technique is applied to resolve two-component mixtures of labeled oligonucleotides. Then, the protonation reaction of fluorescein isothiocyanate (FITC) in phosphate buffer is analyzed as an example of fast reactions (relaxation time $< 10\,\mu$s). By reference to an (apparent) non-reacting system, the simple factorized form of cumulant-based higher-order correlations is exploited to remove the dependence on the molecular

detection function (MDF). Therefore, there is no need to model and characterize the experimental MDF, and the precision and the accuracy of the technique are enhanced. It is verified that higher-order correlation analysis enables complete and simultaneous determination of number and brightness parameters of mixing or reacting molecules, the reaction relaxation time, and forward and reverse reaction rates.

Finally, we apply the technique to analyze the conformational dynamics of DNA hairpins. Previous FCS measurements of DNA hairpin folding dynamics revealed at least three conformational states of the DNA are present, distinguished by the brightness of fluorescent dye-quencher labels. Rapid fluctuations between two of the states occurred on time scales observable by FCS. A third state that was static on the FCS time scale was also observed. We show that conventional FCS alone cannot uniquely distinguish the conformational states or assign their roles in the observed mechanism. The additional information offered by higher-order FCS makes it possible (i) to uniquely identify the static and rapidly-fluctuating states; and (ii) to directly measure the brightnesses and populations of all three observed states. The rapid fluctuations occurring on the FCS time-scale are due to a reversible reaction between the two lowest brightness levels, attributed to the folded and random-coil conformations of the DNA. The third state, which is the brightest, is attributed to spatially extended unfolded conformations that are isolated from the more compact conformations by a substantial energy barrier. These conformations attain a maximum equilibrium population of nearly 10% near physiological temperatures and salt concentrations.

ACKNOWLEDGEMENTS

When I arrived in the US in 2011 to begin my graduate studies in the Physics Department at CSU, I found myself surrounded by the most welcoming and intelligent people I could have imagined. I have been extremely lucky to have had Marty as my advisor and Alan as my co-advisor. Marty taught me how to think when analyzing various research topics, starting with my Master's project and throughout my entire PhD research. He was my go-to person any time I needed help with a theoretical problem, or for any academic support. He was admittedly way more diligent in keeping track of my progress and steps than myself. He is and will always be a role model for me in life.

Working with Alan and in his lab has been an absolute blessing for me. Alan taught me the experimental concepts and techniques, and gave me the direction for my PhD research. He was always available to teach and advise, and at the same time gave me the support and freedom that I needed to explore on my own and try. For this, and for everything he has done for me, I cannot be more grateful.

Throughout my graduate studies I benefited enormously from the discussions with and support from my Master's and PhD committee members, Diego, Ashok, Kristen, and Jake. I am thankful to all of them, especially to Diego and Ashok who helped through closer collaborations and support.

Alan's lab provided me with the environment to grow and thrive, and to meet many people who touched my life and to whom I owe very much. I would like to thank the former senior graduate students Raj, Kevin, and Duncan, and our former post-doc Rebeca, who trained and helped me in the lab and made excellent examples of successful colleagues and students. I am grateful to have met and worked with Megan, David, John, Gracie, Eric, Amanda, and other past and present members of the Van Orden group. The growth and knowledge that I have acquired by experiencing each relationship has made me a better person today and indebted to every single one of these people.

I consider my experience in the physics graduate program at CSU especially rewarding and successful. I learned far more than I hoped for through the well-organized coursework and from the excellent professors I had, including Richard Eykholt, Mark Bradley, Martin Gelfand, William Fairbank, Jacob Roberts, Kristen Buchanan, and Siu Au Lee. I was particularly blessed to have met and worked with Brian Jones both in teaching undergraduate courses and in the Little Shop of Physics. He gave us, his TA's, his passion for presenting science in fun and accessible ways to the younger students and the general public, and taught us how to be better teachers and TA's. My extensive teaching experience in the labs and recitations, during which I had the pleasure of working with our amazing lab coordinator Kenn Lonnquist, has been very rewarding and educational for me. Kenn was particularly good at keeping us organized and efficient.

I would like to extend special thanks to Elliott Forney and Leif Anderson for creating and maintaining the CSU thesis package for LaTeX. Their contribution, combined with the power of LaTeX, made life significantly easier for me when merging separate documents and handling all the references, math, tables, and figures. The output looks pleasant and meets the graduate school formatting guidelines automatically.

I am going to remember my years in Fort Collins not only for the scenic Colorado mountains and rivers, but also for the wonderful friendships that I made, most of which will hopefully last for life. I feel lucky to have experienced this journey with my beloved wife, Sanaz, who has been relentless in giving me her love and support the entire time. Above all, I am eternally grateful to my parents who raised, loved, and educated me to lead a happy and successful life. I owe every achievement to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

A broad range of molecular kinetics are commonly studied using single-molecule spectroscopy techniques [1–5]. These kinetics include spatial diffusion of molecules and interactions with other atoms or molecules, such as binding-unbinding reactions in biological systems. They also include rearrangements of atoms or groups of atoms within molecules, such as the conformational dynamics of proteins and nucleic acids. Very often, such kinetics can be effectively probed by monitoring the fluorescence properties of the molecules. For example, the fluorescence brightness, lifetime, or wavelength properties of a molecule can change with the molecule's state in a reaction. If the molecules of interest are not fluorescent, fluorophore labels may be attached to them to report the desired kinetics.

Single-molecule techniques can resolve the information collected from individual or small groups of molecules. This makes the techniques successful in identifying the molecular states involved in a reaction and characterizing the population of the molecules in each state. Time-resolved single-molecule techniques, such as fluorescence correlation spectroscopy (FCS) [6–10], can further characterize the rates of transitions between the various states, ultimately revealing the energetics of the dynamics being investigated. Almost all of this information remains hidden in bulk measurements which probe only the average properties of a large ensemble of molecules.

Single-molecule experiments include a range of measurements, from stretching macromolecules using cantilevers [11] or optical tweezers [12] to tracking individual fluorescent proteins diffusing on live cell membrane [13] or monitoring enzymes undergoing conformational change as they are tethered to a coverslip [14]. The techniques based on imaging usually employ highly sensitive cameras which confer the ability to spatially track individual molecules with sub-micron resolution. However, the time resolution in these methods is typically limited to milliseconds. On the other hand, fast single-photon detectors, such as photomultiplier tubes (PMT) or single-photon avalanche diodes (SPAD), are used in FCS or fluorescence lifetime measurements which provide

time resolutions down to picoseconds. The fluorescence signal is usually collected from a small confocal region to reduce the out-of-focus light, and individual molecules are probed as they diffuse through the probe region. If imaging functionality is desired, an array of detectors or a scan of the sample area may be used at the expense of collection efficiency and time resolution.

FCS is based on computing and analyzing temporal correlations of the fluorescence signal with itself or with the signal from other detection channels. These correlation functions describe the fluorescence fluctuations that arise from molecular kinetics or the heterogeneity of molecular species, as well as the diffusion of molecules through the confocal probe region. Conventionally, only second-order correlation functions are used in FCS studies. However, second-order correlation functions are not sufficient to characterize a mixture of two or more molecular species, nor can they characterize a reaction between two or more molecular states. Therefore, complementary ensemble measurements, usually with auxiliary theoretical modeling and assumptions, are used together with second-order FCS to assist the characterization of the system in conventional studies. For reactions that involve more than two states, even such complementary methods cannot reveal the underlying multi-state mechanism and help obtain a valid description of the system. This will become clear in the case of DNA hairpin dynamics in this dissertation. Second-order correlation functions cannot distinguish the more complex hairpin dynamics from a two-state mechanism, unless a careful analysis of dependence on temperature or salt concentration is performed. Even then, the correct multi-state description of the system remains obscure, and the molecular states remain uncharacterized.

The stream of fluorescence photons collected from the confocal probe volume in principle contains all of the information needed to fully describe the molecular brightness and number fluctuations. This full description becomes available when the second- and a sufficient number of higher-order correlation functions are used together, not just the second-order correlation function alone. This idea of higher-order FCS was first investigated three decades ago by Palmer and Thompson [15, 16] who demonstrated successful application in resolving multi-component mixtures. However, the technique remained almost entirely unused for two decades due to practical

difficulties: first, a theoretical formulation of the higher-order correlation functions for simultaneously reacting and diffusing molecules seemed too complicated to be viable; second, increased shot noise in higher-order correlations prevented access to short enough time scales necessary to study fast reactions; and third, common detector artifacts which are manageable in second-order FCS became increasingly problematic in higher-order analysis.

In 2009 Melnykov and Hall [17] solved the first problem by formulating higher-order correlations in terms of the cumulants, rather than the moments, of the fluorescence intensity signal. To address the issue of artifacts, they used methods to approximately model and calibrate the detector artifacts in their setup. They achieved $\sim 10\,\mu$s time resolution and demonstrated the unique capability of higher-order FCS to characterize irreversible reactions. The artifact removal methods were still complex and setup-dependent, and the time resolution was not sufficient for some applications.

In this dissertation and the relevant publications, we introduce a "sub-binning" method for the computation of higher-order correlation functions which yields artifact-free results, with no dependence on the specifics of each setup or detector. We also use closer-to-optimal bin sizes to achieve time resolutions of $0.1\,\mu$s for third-order and $1\,\mu$s for fourth-order correlation functions. Using these advances, we show applications of higher-order FCS to resolve two-component non-reacting mixtures, fast two-state reactions, and the three-state dynamics of DNA hairpin conformations. In all of these cases, we achieve a complete characterization of the brightnesses and populations of all molecular states (or species). Additionally, we characterize the forward and reverse rate constants of those fluctuations that occur within the transition time of the molecules in the probe region. For hairpin dynamics, we uniquely identify which molecular states are associated with the fast reaction step, and which state is isolated by slow rate constants. This has important consequences in identifying the hairpin conformations associated with each molecular state.

In the remainder of this section, we briefly review the significance of DNA/RNA hairpin structures in biological systems and in understanding nucleic acid interactions. Then we introduce the basics of FCS and review the performance of second-order analysis in the three mentioned cases:

two-component mixtures, and two- and three-state reactions. Lastly, an overview of the dissertation is provided.

1.1 Nucleic acid hairpins

A hairpin structure is formed when a single strand of DNA or RNA molecule folds back and makes base pairs with itself. The region containing the base pairs in a folded hairpin is called the "stem" and the unpaired segment is known as the "loop" (see Figure 1.1, left). Hairpin structures appear very commonly in nucleic acids in their natural form; for example, almost 70% of the 16S ribosomal RNA folds into hairpins [18]. When double-stranded DNA is unzipped for replication, hairpins may form in the resulting single strands, especially where discontinuous replication leaves single-stranded gaps behind (see Figure 1.1, right). Additionally, hairpins might extrude from double-stranded DNA in slipped strand or cruciform structures (Figure 1.2, left).

Important biological functions have been associated with DNA and RNA hairpins in cellular processes. For example, hairpins direct the binding of a variety of proteins which initiate DNA synthesis for replication [19]. An example of such proteins is RNA polymerase which directly recognizes certain DNA hairpins and binds to them to make either a primer for replication or a transcription for gene expression. Viruses with single-stranded DNA genome extensively use hairpin structures in their life cycle: for complementary-strand synthesis, replication, integration into the host chromosome (recombination), and finally packaging. A variety of other functions, for example in transferring genetic material between bacteria (DNA conjugation), have also been described for DNA hairpins [19].

More than 30 genetic diseases in humans are caused by expanded repeats of a simple sequence of nucleotides in DNA [20, 21]. For example, in Huntington's disease, the triplet repeat $(\mathrm{CAG})_{\mathrm{n}}$ expands to $\mathrm{n} > 40$ in a particular gene [22]. These repeat sequences have the propensity to form hairpin structures during DNA replication. Part of the nascent DNA copy may fold into a hairpin and cause the replication of the repeat sequence to erroneously continue to a longer length (Figure 1.2, right). Consequently, the repeat expands to longer lengths in successive generations [23].

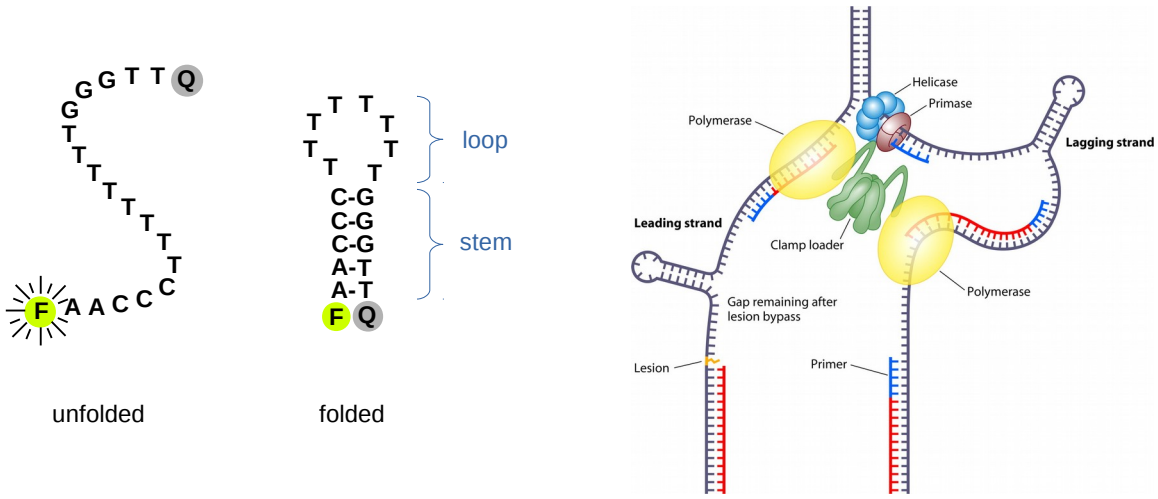

Figure 1.1: Left: Schematic examples of individual hairpin molecules used in our study. Hairpins are shown in example folded and unfolded conformations. F and Q denote the fluorophore and quencher labels, respectively. The fluorophore is brighter in the unfolded conformation. Right: Hairpins form naturally during DNA replication in the single-stranded regions. The right panel is adopted from Bikard et al. [19]

This results in earlier onset and increased severity of the disease in individuals of subsequent generations. The thermodynamic stability of the hairpins has been shown to determine their potential for large scale expansion in diseases. [24] Mechanisms to suppress hairpin formation in such repeats have been proposed to stop the disease-causing expansion. [25]

To understand the hairpin formation mechanism in various biological systems, including the thermodynamics of the mechanism and its response to possible perturbations, a solid theoretical understanding of the fundamental molecular interactions is needed first. Then, sufficient theoretical and computational analysis is required to model and predict the complex interactions in biologically relevant scales. Limitations exist in both of these aspects today. Various types of interactions have been identified to play key roles in the structures and dynamics of nucleic acids, such as standard and non-canonical base pairing, base stacking, and the short- and long-range electrostatic interactions. Debate is still ongoing as to, for example, what constitutes the atomic origin of base stacking and how to properly model this interaction in large scale calculations: different studies have inconsistently suggested either electrostatic [26], dispersion [27], solvent hydrophobic [28], or solvent entropic [29] forces, or some combination of them [30,31], as the main factor that drives

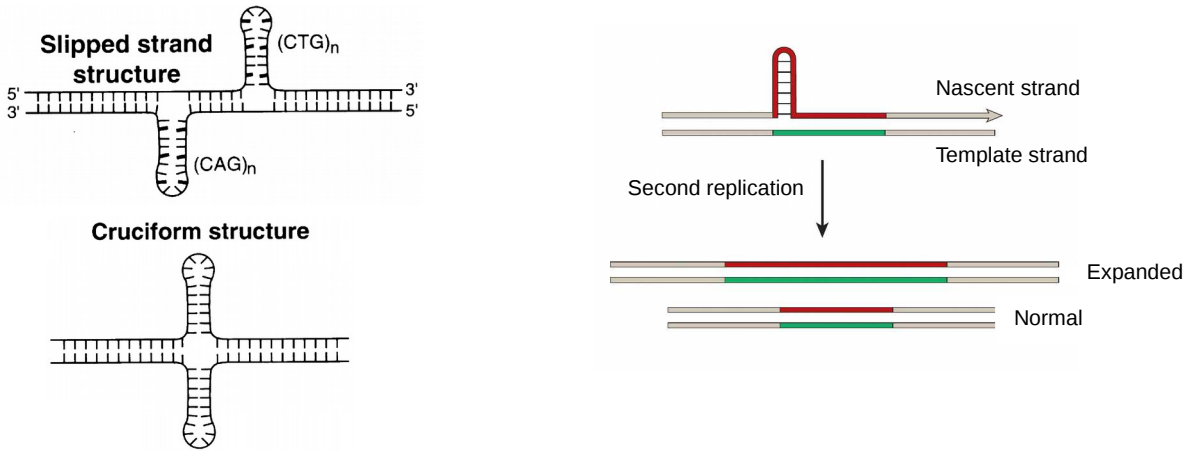

Figure 1.2: Left: Hairpin extrusions on double-stranded DNA in the forms of slipped strand and cruciform. The slipped strand structure is shown for triplet repeats that are relevant in neurodegenerative diseases. Although base paring in the stem is not perfect, hairpins still form. Right: In a disease, repetitive DNA sequence expands during replication because the nascent strand folds into hairpin structure. As the repeat segment becomes longer, the expansion becomes progressively more likely in successive generations. The left panel is adopted from Sinden et al. [22]. The right panel is adopted from Mirkin [20].

the aromatic stacking between only two nucleobases. The problem becomes increasingly challenging for molecules of the size of a hairpin. The hairpin dynamics involve large rearrangements and motions of the hairpin and solvent molecules, on number and time scales far beyond current ab initio calculations. On the other hand, the scales are not large enough for field-theoretical approximations to apply conveniently. Investigations are in progress to find and refine appropriate force fields for molecular dynamics simulations. [32–34]

There is hence a clear need for experimental measurements on simple hairpin molecules as model systems to help improve our theoretical understanding of nucleic acid interactions and refine our computational tools to predict nucleic acid structures and dynamics. For this purpose, we use short synthetic DNA strands which can fold into hairpins, as shown in Figure 1.1 (left), in our experimental research in this dissertation. These small hairpin molecules dynamically fluctuate between various folded and unfolded conformations, as do hairpins in larger DNA molecules. We seek to develop and apply experimental methods to reveal these conformations; in particular, identify the stable conformational states and the group of conformations represented by each state. We aim to measure the thermodynamical properties of each state, and understand the molecular

interactions that stabilize the hairpin in each state. Further, we intend to probe the transitions between different conformational states; characterize the possible pathways, the intermediate steps, the rates, and the energy barriers associated with these transitions. In addition, we plan to investigate the dependence of the kinetics and thermodynamics of hairpin formation on the length and sequence of the stem and loop segments. Ideally, the same techniques might be used to target larger nucleic acids in more complex biological environments to test our models and verify our predictions.

In our experimental approach we use fluorescence as a measure of end-to-end distance in various conformations of the hairpin. To this end, we attach a fluorophore and a quencher label to the opposite termini of the hairpin as shown in Figure 1.1, left. As the hairpin folds, the fluorophore gets closer to the quencher and fluorescence intensity decreases. In an unfolded conformation, the fluorescence is higher because the fluorophore and quencher are farther apart. The rapid folding-unfolding transitions will then appear as fluorescence fluctuations. In our FCS technique, we try to detect these fluorescence fluctuations and translate them into the underlying molecular kinetics.

It is noteworthy that similar hairpin molecules with a fluorophore and quencher pair are also commonly used in biochemical studies to detect specific sequences of nucleic acids. The fluorescence of these probes increases upon binding to the specific target molecules and reports their presence in the sample. For this application, the hairpins are commonly known as molecular beacons. However, in this dissertation we use short hairpin molecules not as molecular beacons, but for the analysis of hairpin conformational dynamics as described above.

We can study hairpin conformational dynamics within a more general category of two- or multi-state reactions that can be probed based on fluorescence fluctuations. In the following section, we describe the basics of conventional FCS and review its performance in the analysis of certain categories of two- and multi-state reactions. This will clarify the limitations encountered in second-order FCS and will motivate the development of a higher-order analysis.

1.2 FCS principles

In this section we explain the principles of FCS and how it can be applied to probe molecular kinetics of various time scales and complexities. We will consider limiting cases of fast and slow two-state kinetics and compare the experimental results in these cases with the data from DNA hairpins. The analysis will reveal limitations that necessitate complementary measurements, in particular higher-order FCS.

In over four decades, FSC and its derivative techniques have evolved into successful methods to monitor molecular kinetics and electronic transitions over a broad range of time scales based on fluorescence fluctuations. [6–9] In a basic FCS experiment, fluorescent molecules are probed as they diffuse into and out of a small, laser-illuminated volume. This probe volume (also known as the probe region) is usually less than a micron in diameter and is obtained by focusing a laser beam into the sample volume. A schematic diagram of a simple FCS setup is shown in Figure 1.3. As seen in the picture, a microscope objective lens is used for focusing the laser light. When the molecules enter the probe region they emit fluorescent light. The emitted photons are collected using the same microscope objective and directed toward the detector(s). Along the way, a pinhole is placed to ensure that the fluorescent light from only a small volume around the laser focal point reaches the detectors ("confocal" setup): any out-of-focus light is blocked to reduce the background and increase the signal-to-noise ratio. (In two-photon excitation the pinhole is not necessary since excitation is much more efficient near the focal point.) The dichroic mirror reflects and transmits the laser and the fluorescent lights, respectively. The emission filter also ensures that only the desired fluorescent signal is detected.

A theoretical fluorescence intensity function, $I(t)$, can be considered to denote the probability of detecting a photon by the detector at any moment t. The schematic fluorescence intensity function in Figure 1.3 shows the fluorescence intensity increasing and decreasing as a molecule enters and exits the probe region. The average time a molecules spends in the probe region is denoted as the "diffusion time". The diffusion time is typically ~ 0.1–$1\,\mathrm{ms}$ depending on the size of the molecule. For simplicity in this introduction, we assume that the concentration of the fluo-

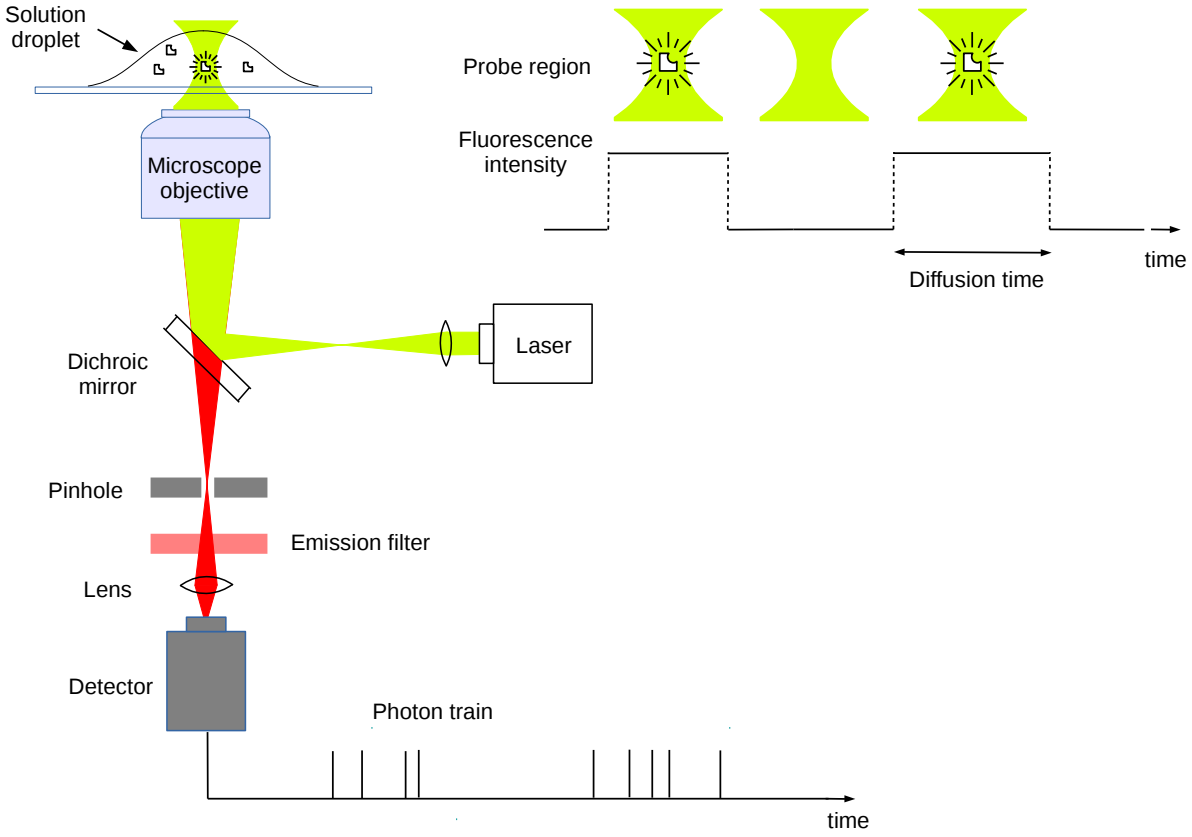

Figure 1.3: Schematic setup of a simple FCS experiment. A laser beam (green) is directed into a microscope objective lens which focuses the beam into a small volume (probe region) inside the solution. As the fluorescent molecules in the solution diffuse through the probe region they emit fluorescent light (red) which is collected by the same objective lens and directed toward the detector. The pinhole ensures that only the light from a small volume around the focal region reaches the detector. By lowering the sample concentration, only a single molecule at a time will occupy the probe region. A theoretical fluorescence intensity signal is shown corresponding to when the probe region is empty or occupied by a fluorescent molecule. A uniform laser beam profile with sharp edges is assumed in this diagram for simplicity. In actual experiments, the single-photon detector generates a voltage spike for each detected photon; therefore the detector output is represented by a photon train.

rescent molecules is low enough that the probe region is either empty or occupied by one molecule at a time. This usually corresponds to concentrations significantly below 1 nanomolar in experiments, at which FCS can be successfully applied. Despite this assumption, the FCS formulation accounts for fluctuations in the number of molecules and the technique can be applied to higher concentrations as well, maintaining single-molecule sensitivity in several orders of magnitude in concentration.

As each molecule enters the probe region, it may undergo a reaction that results in changes in its fluorescence brightness. The reaction may involve two or multiple states, with a distinct level of fluorescence brightness for each state. For example, a simple two-state reaction is depicted as a binding-unbinding mechanism in Figure 1.4, top. The molecule in the unbound state is assumed to be brighter than in the bound state. We note that this binding reaction is only a representation, the analysis applies equally to conformational dynamics and other kinetics that show distinct fluorescence brightness levels. In Figure 1.4, we consider two-state reactions in two limiting cases based on the reaction time scale; the performance of FCS can then be understood on any time scale in between. A "fast" reaction occurs on a time scale significantly shorter than the average diffusion time of the molecules: as a molecule enters the probe region, the molecule is likely to undergo several transitions between the bright and dim states before it leaves the probe region (Figure 1.4, second from above). In a "slow" reaction, the reaction transitions occur on a time scale much longer than the diffusion time. When a molecule enters the probe region, it is unlikely to change its brightness state before leaving the probe region; the entire transition takes place at either a bright or dim state (Figure 1.4, third from above). It should be clear that in the slowest possible limit, the system is equivalent to and can be modeled by a static or non-reacting mixture. As far as FCS is concerned, a reaction on time scales much slower than the diffusion time is indistinguishable from a non-reacting mixture.

The first task expected from FCS to achieve is to distinguish the fast or slow nature of the reaction; if the reaction is fast enough, measure its characteristic time scale. On top of that, reveal the probability of finding the molecule in each brightness state, regardless of the reaction time scale.

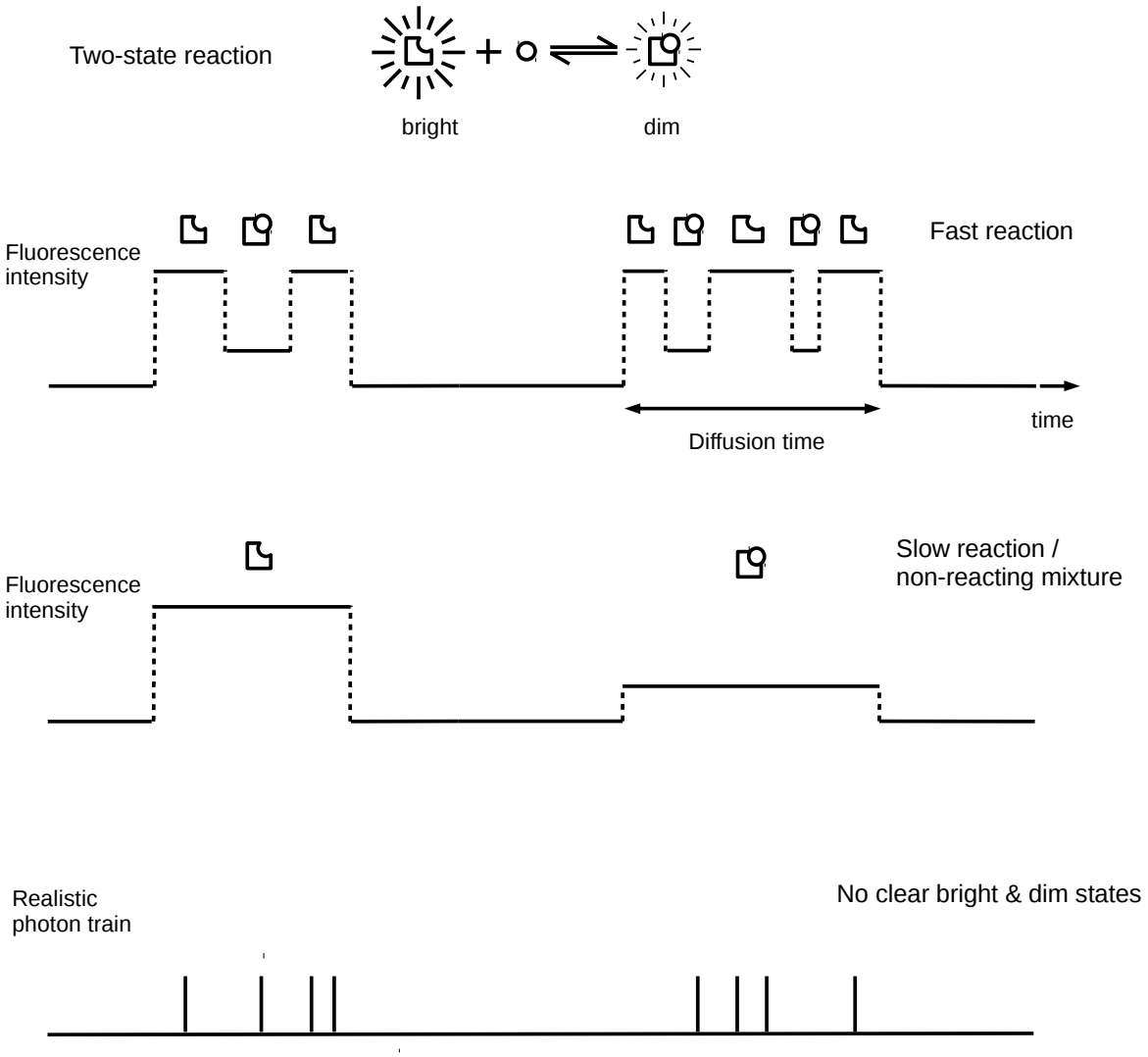

Figure 1.4: Schematic representation of fast and slow reactions using fluorescence intensity traces and photon train. A two-state reaction between bright and dim states is represented by a binding mechanism. In a fast reaction, the transitions between the bright and dim states occur on a time scale much shorter than the average diffusion time of the molecules through the probe region. In a slow reaction, the transitions (if any) occur on a time scale much slower than the diffusion time, so they are not likely to be observed during the transition of a molecule through the probe region. A very slow reaction can be considered and modeled as a non-reacting mixture in this approach. A photon train more realistically represents what is obtained in an experiment. The individual bright and dim transitions cannot be identified in the photon train, therefore a statistical analysis approach is needed to find the type of the mechanism and the characteristic time scales and populations from the photon train.

In other words, characterize the relative population of each state in a reaction, or the concentration of each species in a mixture. When the reaction time scale (relaxation time) and the relative population numbers are found, the individual forward and reverse rate constants also become available through straightforward detailed balance relations. All of this information can be obtained if an intensity trace such as those shown in Figure 1.4 is available. One can simply see the distinct pattern in each trace showing the type of the reaction. For the fast reaction, one can measure the average time spent in each state, and from there, the relative populations, rate constants, and relaxation time can be calculated. For the slow reaction, the intensity trace still yields the relative population of the two species: one needs to simply count the number of times a molecule of each species enters the probe reaction.

In a real FCS experiment, the fluorescence intensity trace is not available. The intensity traces shown in Figure 1.4 represent the probability of detecting a photon by the detector. What is in fact collected in an experiment is a photon train that may look like the diagram shown at the bottom of Figure 1.4 in comparison to the intensity traces above it. Looking at the photon train, there is no way to distinguish the individual bright and dim states of a reaction, because there are not enough photons in each state. In single-molecule techniques where the molecules are immobilized or slowly moving and the kinetic transitions are on millisecond time scales or slower, enough photons may be collected to effectively reconstruct the intensity traces. However, in a typical FCS experiment collecting photons at a rate of say $50\,\mathrm{kHz}$, the average diffusion time of the molecules can be $\sim 100\,\mu\mathrm{s}$ and the average number of molecules in the probe region may be ~ 1. This means that there are on average only ~ 5 photons collected per each transition of a molecule through the probe region. The number of photons available in each transient state of a fast reaction may be even smaller. The observed kinetics are fast and we usually need to keep the excitation power low for practical reasons. Increasing the laser power usually results in undesired photo-induced effects, such as fluorophore blinking which may complicate the analysis, photobleaching, or other photodegradation, saturation, or sample damage effects.

For the above reason, and for other practical details such as non-uniform laser illumination and particle number fluctuations, a direct analysis of the individual bright and dim states in a photon train is difficult or impossible. However, as one would intuitively expect, the photon train (considered as a whole with many photons) should still contain all the information about the statistical properties of the underlying intensity trace or kinetics. Statistical analysis methods are needed to extract such information. Among the most successful of these statistical methods is correlation analysis.

The correlation analysis can be explained more conveniently using the intensity traces, then a similar approach will apply to photon trains. For this introduction, we only describe the autocorrelation function, which is obtained by correlating a function with itself in time. The autocorrelation function can be considered as a measure of the similarity of the intensity trace with a copy of itself that is shifted in time by a small "lag time", τ. Figure 1.5 shows a representation of this idea. The correlation can be quantified as the average of the product of the copies: $\langle I(t)I(t+\tau)\rangle$, with brackets denoting averaging over time, t. The correlation is plotted as a function of lag time in Figure 1.5, bottom (ignoring any normalization or scale for now). At zero or short lag times, the similarity (overlap) between the two copies of the signal is higher than when the lag time is large. The transition from high to low similarity occurs on the time scale of the intensity fluctuations. In particular, for the intensity trace of molecules that undergo a fast reaction as shown in Figure 1.5, two time scales can be identified in the correlation function. The slower time scale corresponds to the diffusion time, τ_{D}, and the faster time scale is the reaction relaxation time, τ_{R}. The relaxation time is a combination of the average bright and dim times, τ_{bri} and τ_{dim}, but does not yield these parameters individually. We will later consider if information from the amplitude of the correlation function can help solve for these parameters individually.

A similar approach can be taken to compute the autocorrelation using the photon train. The main difference is that photons need to be divided into small sampling times known as bins, as shown in Figure 1.6 . The number of photons in a bin centered around the time t is counted and denoted by $n(t)$. Then the correlation function is computed using a relation similar to what we had

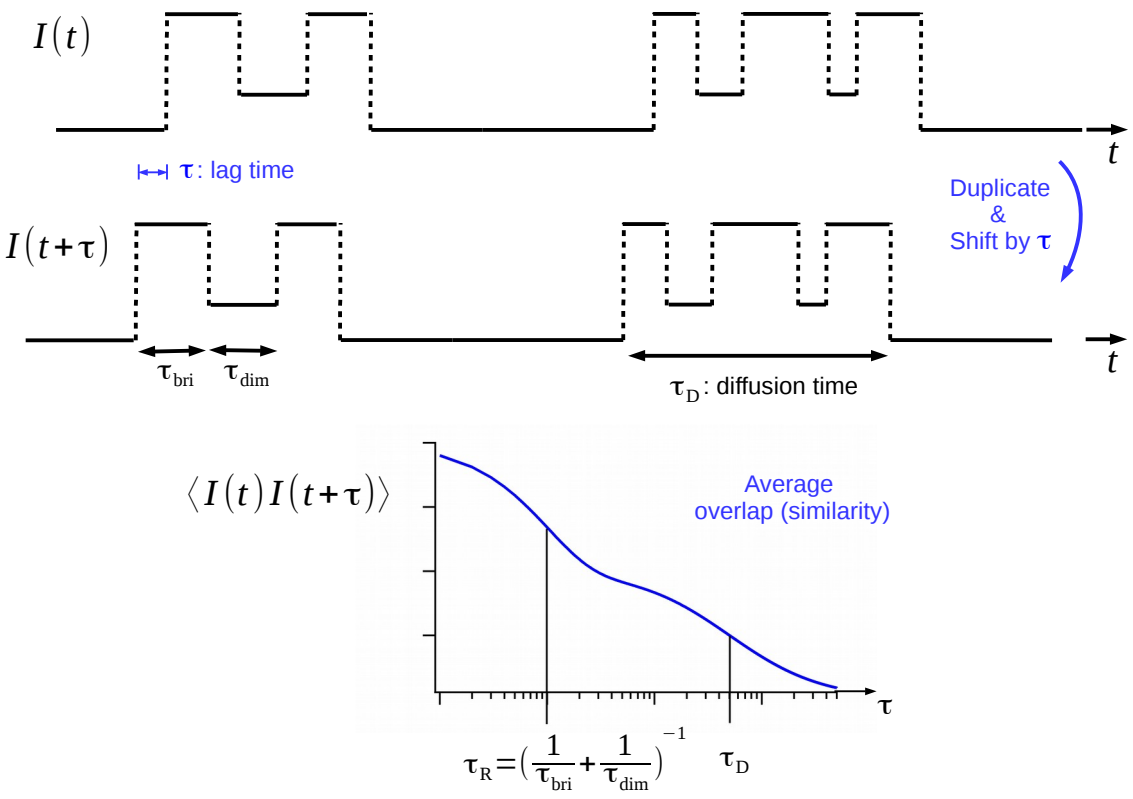

Figure 1.5: Schematic representation of the autocorrelation function calculation. The intensity trace is duplicate and shifted by a lag time, τ. Autocorrelation represents the similarity or overlap between the two copies of the signal. The (unnormalized) autocorrelation is plotted as a function of lag time. Characteristic reaction and diffusion times can be identified in the two decaying components of the autocorrelation function.

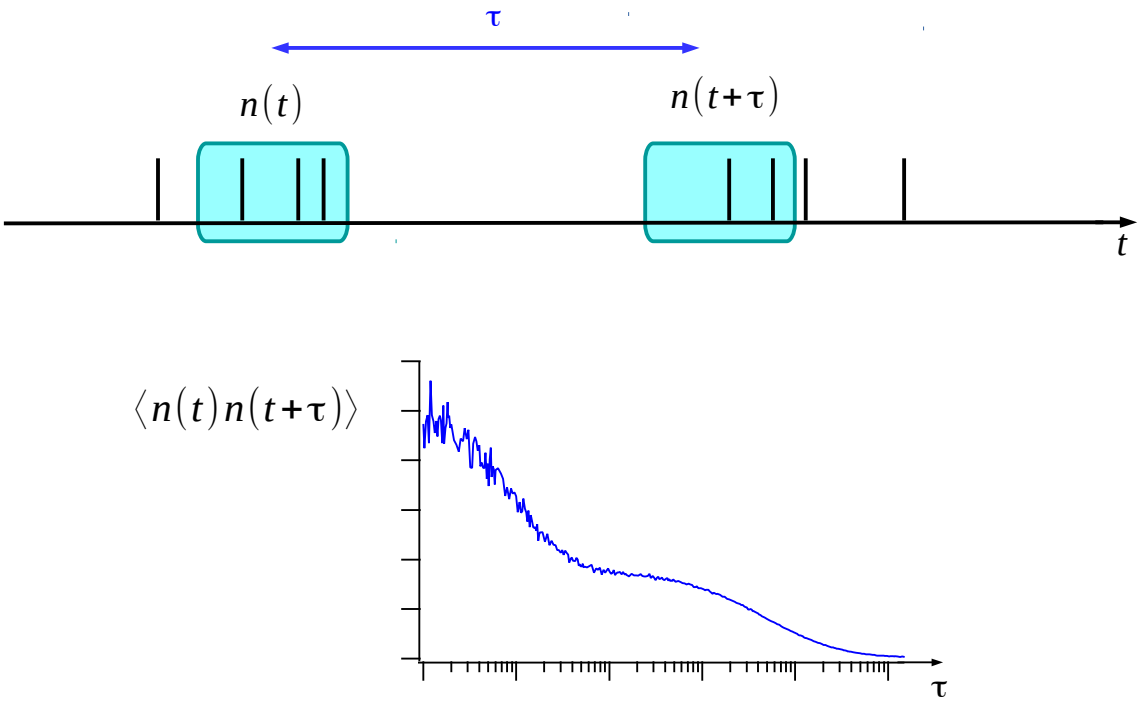

Figure 1.6: Schematic representation of binning the photon train to compute the autocorrelation function. The relation to compute the autocorrelation function is similar to that involving the intensity traces (in Figure 1.5). The resulting correlation function may contain some shot noise due to the discrete nature of the photons.

for intensity traces: $\langle n(t)n(t+\tau)\rangle$. An autocorrelation function obtained in this way will contain some noise, especially at shorter lag times where bins need to be taken smaller (see Figure 1.6 , bottom). This is called shot noise and is due to the discrete nature of the photons. The shot noise can be reduced by collecting more photons in a longer experiment and/or improving collection efficiency.

The autocorrelation function needs to be normalized properly for easier analysis. The normalized autocorrelation function is defined as:

$$g(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I\rangle^2} - 1 \tag{1.1}$$

This way of normalization ensures that the amplitude of the correlation function at short (zero) lag times, which contains useful information about the number of molecules, is preserved (as opposed

to the common Pearson normalization, for example). It also ensures that the baseline decays to zero at very long lag times.

The autocorrelation function can be modeled and calculated for various systems of reacting and diffusion molecules. An important category of systems are those with independent reaction and diffusion processes. In these systems, the diffusional properties of the fluorescent molecule do not change significantly during the reaction. This is a good approximation for many systems, especially since FCS is not as sensitive to changes in diffusion constant as it is to changes in brightness. For example binding of a small ion to a larger fluorescent molecule may not change the diffusion constant of the larger molecule significantly. The conformational dynamics of hairpin molecules have also been shown to satisfy this condition within experimental uncertainty. Larger changes in the diffusion constant, if measurable, inherently provide additional information that helps with resolving the reaction components, although the analysis may be complicated in the case of fast reactions.

With the assumption that the reaction and diffusion are independent processes, it can be shown that the autocorrelation function factorizes into two components [6]:

$$g(\tau) = R(\tau)Y(\tau) \tag{1.2}$$

The reaction function, $R(\tau)$, depends only on the reaction process, and the function $Y(t)$ depends only on the diffusion constant and the spatial shape of the probe region (the laser beam profile). For example, for an approximately two-dimensional Gaussian beam profile, we have

$$Y(\tau) = \frac{1}{1 + \tau/\tau_{\mathrm{D}}}$$

Regardless of the exact form of $Y(\tau)$, this function can be characterized experimentally using a reference or control measurement. Then by normalizing $g(\tau)$ by the experimental $Y(\tau)$, the reaction function $R(\tau)$ can be extracted in an experiment. $R(\tau)$ contains all the information about the reaction that can be found in the autocorrelation function; however, this may or may not be all

the information that is available in the photon train (or the corresponding intensity trace). In the next section, we investigate this question using three example cases.

1.3 Conventional FCS: performance overview

To examine the performance of conventional FCS, as was described in the previous section, in characterizing different types of molecular reactions, we consider three example cases; first we look at two-state reactions in the slow and fast limits. The slow limit is modeled by a non-reacting mixture of a bright fluorophore (R6G) and a dimmer one (TAMRA). The fluorophores are attached to short single-stranded DNA molecules only to ensure more similar diffusion properties, no interaction occurs between or within the molecules. The reaction function for this system is shown in Figure 1.7. As theory predicts, the reaction function is effectively a constant line (the small variations at short short lag times are due to residual triplet blinking of the fluorophores and can be ignored). The constant value of the reaction function is a combination of three parameters: the individual populations of the bright and dim states, N_{bright} and N_{dim}, as well as the relative brightness of the two states, q. Unfortunately, the values of these three parameters cannot be individually resolved using a single reaction function value.

The fast two-state reaction is modeled by the protonation reaction of FITC, a fluorophore which becomes dimmer upon binding to hydrogen ions in solution. This binding-unbinding reaction occurs on time scales shorter than $10\,\mu$s. The reaction function for a two-state mechanism has the following exponential form:

$$R(\tau) = \frac{1}{N}\left[1 + B\exp(-\tau/\tau_{\mathrm{R}})\right] \tag{1.3}$$

The data for the FITC protonation reaction is shown in Figure 1.7, middle, and is fitted by Equation (1.3). Other than the relaxation time, τ_{R}, there are two parameters that can be obtained from this function: N is the sum of N_{bright} and N_{dim}, and B is a function of N_{bright}, N_{dim}, and q. Therefore there are not enough parameters to solve for N_{bright}, N_{dim} and q individually (and con-

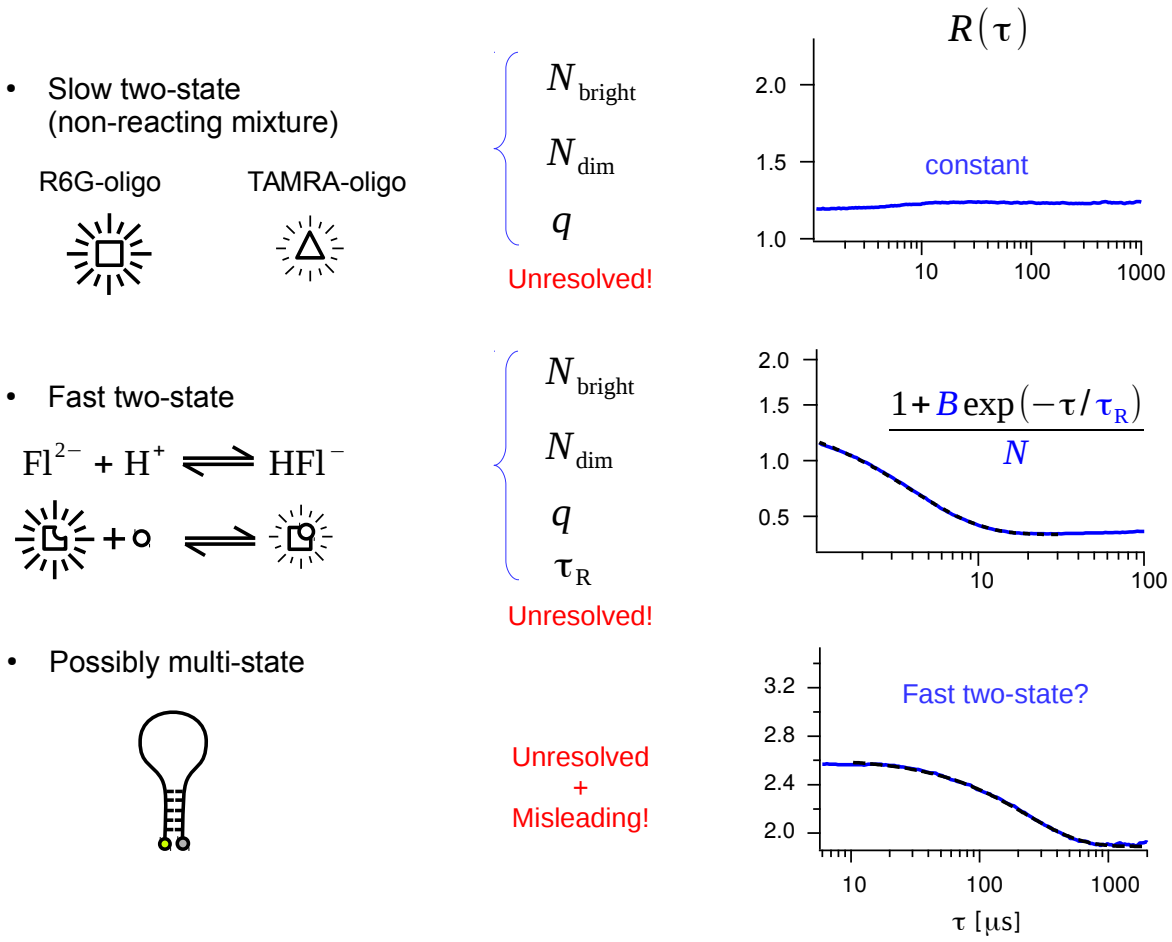

Figure 1.7: Conventional (second-order) reaction functions and their performance in resolving two-state reactions in the slow and fast limits; also in the analysis of hairpin folding dynamics. Experimental data are shown in blue and the fitted functions are in dashed black. The number of parameters obtained from the reaction function in each case is fewer than the parameters that describe the system, leaving the critical population parameters (and hence the rate constants) unresolved. Additionally, the reaction function suggests a two-state mechanism for the hairpin dynamics which was later found to be incorrect.

sequently for the individual rate constants). The system remains unresolved using conventional FCS alone.

To compensate for these limitations, conventional FCS experiments are usually accompanied by some sort of complementary, independent measurement(s). The complementary measurements are sometimes difficult to find, perform, or relate to the FCS measurements in a reliable way; the analysis commonly requires some additional modeling, assumptions, and approximations about the system. These efforts would be unnecessary if one could extract all of the information available in the photon train: as was pointed out, the photon train contains all of the statistical information that the corresponding intensity trace contains. The problem is that conventional FCS does not extract all of this information.

Lastly, we look at the reaction function for the system we set out to study: the DNA hairpin. The reaction function is shown in Figure 1.7, bottom, and is fitted almost perfectly by the two-state function, Equation (1.3). This suggests that the hairpin dynamics follow a simple two-state mechanism between folded and unfolded conformations. In fact, this was the assumption for many years since the earliest FCS measurements on hairpins [35, 36]. However, Jung et al. later reported that the two-state model does not hold when the dependence of the autocorrelation function on salt concentration is considered [37] (see Section 6.2 for additional temperature dependence studies and an overview of the problem using reaction functions). They assumed that a three-state mechanism was involved comprising a fast step (in the sense described in this section) and two slow steps; only the fast step of the reaction was observable within the FCS time frame. They further assumed that the fast step occurred between the brightest and the intermediate states; the third state was thought to be an invisible, dark state. These assumptions were made since conventional FCS did not provide enough information to nail down the correct model. The reaction parameters could not be measured either, using conventional FCS alone. Higher-order FCS has later revealed that in the correct model the fast reaction is between the intermediate and the dimmest states; the analysis has also assisted with measuring many parameters and characterizing the underlying hairpin conformations.

Although conventional FCS was able to reveal the slow or fast nature of a two-state mechanism and measure the characteristic time scale of a fast reaction, it was unable to independently resolve any of the individual population, brightness, or rate constant components. In the case of a more complex mechanism such as hairpin conformational dynamics, second order FCS alone appeared to suggest incorrect results. It has become clear that the conventionally used correlation function, Equation (1.1) (also known as the second-order correlation function) does not extract all of the statistical information that is available in the photon train. In an analogous comparison, conventional FCS is like measuring the standard deviation (and possibly the mean) of a statistical distribution. For any distribution more complex than a Gaussian, this will be insufficient information to fully describe the distribution. Higher statistical moments will be needed to complete the description. The same idea is behind higher-order FCS. See the Conclusion chapter (Figure 8.1) for a review of the performance of higher-order FCS in the three cases discussed in this section.

1.4 Dissertation overview

Although the idea of higher-order FCS has been around for nearly three decades, the technique has found little applications until recent developments. The main reasons for this were three major difficulties. The first difficulty was that the formulation of higher-order correlation functions using higher-order statistical moments turned out to yield extremely complicated results for systems of diffusing and reacting species. Palmer and Thompson [15, 16] developed the complex formulation for non-reacting mixtures in 1987, however, the formulation could not be easily extended to describe reactions; any result would certainly not factorize into separate reaction and diffusion parts as it did in second-order FCS (Equation (1.2)). It wasn't until 2009 that Melnykov and Hall [17] proposed a formulation based on higher-order cumulants which yielded a simple factorized formulation for reactions.

The second problem was complications arising from known detector artifacts, most notably dead-time and after-pulsing in single-photon detectors. These artifacts affect second-order correlation functions only at short lag times; these lag times are either not needed in conventional FCS

or the artifacts can be removed by cross-correlating two detectors. In higher-order FCS, these artifacts affect the correlation functions on *all* time scales. Attempts to remove the artifacts using two detectors may face new challenges: signal-to-noise ratio may be reduced and a new cross-talk artifact may be encountered.

The third problem was the dramatical increase of shot noise in higher-order correlations. Higher order correlations rely on the detection of two or photons within a bin; these incidents become increasingly unlikely with higher numbers of photons required in a bin. This limited the time resolution in the work of Melnykov and Hall [17] to $\sim 10\,\mu$s or longer (using standard fluorophores and laser power) which was not enough for many fast reactions.

In **Chapters 2 and 3** of this dissertation, we introduce methods to solve the latter two problems facing higher-order FCS analysis. A method called "sub-binning" is introduced which, as the name suggests, is based on splitting each sampling bin into two or more smaller bins. This will allow the artifact-free calculation of higher-order correlation functions using one or more detectors. We also show that the bin size can be more optimally selected to reduce the shot noise without creating significant systematic errors. Using this approach, we show that time resolutions down to $0.1\,\mu$s for third-order correlations, and $1\,\mu$s for fourth-order correlation functions can be achieved using standard laser power and fluorophores.

In **Chapters 4 and 5**, we show proof-of-concept applications of our technique developments in resolving fast and slow two-state reactions. The technique is used to resolve a two-component mixture using the reaction formalism in the slow time-scale limit. This will greatly simplify the analysis compared to the early formulation by Palmer and Thompson. Then we characterize conditions in which higher-order FCS can yield better or worse statistical precisions. Lastly, we apply the technique to analyze the fast protonation reaction of FITC, and we show that the technique can fully characterize the reaction parameters independently, as expected.

In **Chapters 6 and 7**, we apply higher-order FCS to study DNA hairpin folding dynamics. Higher-order FCS immediately reveals the multi-state nature of the mechanism and allows the characterization of many important parameters. A three-state model, with fast reaction step between

the intermediate and dimmest states, is found to fit the data reasonably. A study of the dependence of the measured population numbers and the rate constants on temperature and salt concentration reveals that the intermediate and dimmest states correspond to the well-known random-coil unfolded and folded hairpin conformation. The brightest state seems to correspond to unusually extended conformations which are separated from the more compact conformations by a large energy barrier.

There will be a small change in our mathematical notation for correlation functions. In the relations containing averages over time, t, for example $\langle I(t)I(t+\tau)\rangle$, the parameter t has no notation significance and can be taken to be 0 (and the averaging can be thought to be on an ensemble). Instead, in the rest of this dissertation we use the Roman letter t to denote the lag time (τ in this Chapter); for example we write $\langle I(0)I(t)\rangle$ as a more convenient notation.

Chapter 2

Improving higher-order fluorescence correlation spectroscopy

2.1 Introduction

Fluorescence correlation spectroscopy (FCS) is a technique widely used to probe concentration, diffusion, and chemical reaction kinetics of single fluorescent molecules in solution on the sub-microseconds to milliseconds time scale. [6–8] When the underlying dynamics are not too complex, many of these molecular parameters can be derived from autocorrelation analysis of the fluorescence intensity fluctuations observed from a microscopic optical probe region:

$$g_{11}(t) = \frac{\langle \delta I(0) \delta I(t) \rangle}{\langle I \rangle^2} \tag{2.1}$$

Here, $g_{11}(t)$ is the conventional, second-order autocorrelation function, $\delta I(t) = I(t) - \langle I \rangle$ denotes the centered fluorescence intensity at lag time t, and angle brackets indicate averaging over the entire experimental time. The non-invasive nature of FCS, its sub-micron spatial resolution, and its sensitivity to sub-nanomolar concentrations make this technique suitable for a wide range of biological and chemical applications.

An important application of FCS is chemical kinetics, including biomolecule conformational fluctuations and binding reactions. Examples include folding and unfolding of DNA or RNA hairpin structures [37–40], and ligand-receptor binding and unbinding [41–45]. In many such cases, the experiment is designed such that the chemical reactions induce fluorescence intensity fluctuations between various brightness levels of the fluorescent labels. FCS characterizes these fluctuations through fitting a theoretical model to the experimental autocorrelation function. As

a simple example, consider a reversible process between two brightness levels Q_1 and Q_2 with average populations N_1 and N_2 in the probe region, respectively:

$$\text{state 1} \underset{k_\text{b}}{\overset{k_\text{f}}{\rightleftharpoons}} \text{state 2} \tag{2.2}$$

Here, k_f and k_b refer to forward and reverse (backward) rates, respectively. Using a two-dimensional Gaussian beam profile, the autocorrelation function takes the form [6]:

$$g_{11}(t) = \frac{1 + Be^{-t/t_\text{R}}}{N}\left(\frac{1}{1 + t/t_\text{D}}\right) \tag{2.3}$$

The parameters that can be obtained through fitting, namely N, B, t_R, and t_D, are related to the underlying parameters as follows:

$$\begin{aligned} N &= N_1 + N_2 \\ B &= \frac{k(1-q)^2}{(1+kq)^2} \\ t_\text{R} &= (k_\text{f} + k_\text{b})^{-1} \end{aligned} \tag{2.4}$$

where $k = N_2/N_1 = k_\text{f}/k_\text{b}$ is the equilibrium constant, and $q = Q_2/Q_1$ is the brightness ratio. The parameter t_D is the characteristic diffusion time of the molecules through the probe region and is assumed to be independent of the reaction.

The system of equations 2.4 cannot be solved to determine *any* of the underlying parameters on the right hand side. In general, the number of fitted parameters that can be obtained from a single measurement of the autocorrelation function is limited. For a typical reaction, there are usually more underlying parameters of interest than can be fitted to obtain a unique solution. To obtain all the desired parameters, independent or complementary methods must be applied in addition to the conventional autocorrelation analysis to fix or constrain some parameters. In the case of Equations (2.4), if only one of the parameters on the right hand side is somehow additionally known, or if another independent relation is given, then all the parameters are uniquely determined.

Despite the difficulty in extracting the reaction parameters using conventional (second-order) FCS, the stream of photons collected in an FCS experiment in principle contains all the dynamical information within the dwell time of the molecules in the probe region. As an analogy, consider a simple probability distribution of arbitrary shape. Second-order FCS is analogous to measuring the variance of that distribution. However, if the distribution shape is more complex than a simple Gaussian, the mean and the variance are insufficient to describe the distribution completely. Higher statistical moments, such as skewness, are then required to capture more details. In an analogous way, higher-order correlations are a natural extension of conventional second-order correlation, Equation (2.1), to provide more information on the underlying dynamics. For example, the third-order correlation can be defined as

$$g_{21}(t) = \frac{\langle \delta I^2(0) \delta I(t) \rangle}{\langle \delta I^2 \rangle \langle I \rangle} \tag{2.5}$$

Such higher-order correlation functions provide independent relationships between parameters, which can be used in combination with second-order FCS, Equation (2.1), to fully determine the system dynamics.

Other analysis techniques which are not based on correlation have also been devised that can extract complementary information from the single-photon data stream. Photon counting histogram (PCH) or fluorescence intensity distribution analysis (FIDA) [46–48] uses the distribution of the number of photons within a small bin size to deduce the molecular brightness values and concentrations. The technique is mathematically equivalent to non-correlated fluctuation moment analysis [49, 50] and cumulant analysis (FCA) [51, 52]. When fast reaction dynamics are involved, the bin size has to be taken much smaller than the reaction time scale because such techniques commonly ignore reaction dynamics for simplicity. At the same time, the excitation intensity usually has to be reduced to minimize fast photochemical blinking of the fluorophores. Few photons, if any, occur in a small bin size at reduced fluorescence intensity in a limited data set. This leaves us with few histogram points for fitting in PCH, and large uncertainties in higher moments and cumulants. To overcome these difficulties, PCH (FIDA) and FCA have been extended to arbitrary bin sizes for

non-interacting diffusing molecules. [53–56] However, no treatment of the case of reacting species has been proposed. In contrast to such techniques, second- and higher-order correlation analysis enable studying dynamics at arbitrary lag times using simple analytical models that incorporate both diffusion and reaction dynamics. Some more elaborate correlation-related techniques have also been suggested, for example three time correlation analysis [57], or a two-dimensional extension of FCS by adding a photon counting dimension to the correlation function [58]. Higher-order FCS provides the desired complementary information with simpler formulation and analysis.

The generalization of FCS to higher orders was first developed by Palmer and Thompson [15, 16, 59] in a series of papers in late 1980s. These authors defined higher-order correlations based on a natural extension to higher moments. The results included rather complicated analytical functions describing the higher-order correlation functions for non-reacting molecules in solution. The authors applied the technique to characterize mixtures of fluorescent proteins using high-order correlations at lag times above $0.3\,\mathrm{ms}$ [16]. No extension of this moment-based formulation to include reaction dynamics has been proposed.

More recently, Melnykov and Hall introduced a definition of higher-order correlations based on statistical cumulants instead of moments of the signal intensity [17]. The cumulant-based formulation significantly reduced the analytical complexity of the models in systems of diffusing molecules, and was an important step towards making the technique applicable to the study of reactions. The resulting relations contain separate reaction-related and diffusion-related factors, as in ordinary FCS. Another significant improvement over earlier work was the reduction of the experimentally accessible time scale to $\sim 10\,\mu\mathrm{s}$ to enable the observation of fast dynamics. Those authors demonstrated an application of high-order FCS to characterize irreversible processes, which is one of the important advantages of this technique. For their experiments, the authors used a single detector to preserve signal-to-noise ratio (SNR). However, because only a single detector was used, the authors had to account for common detector artifacts, such as dead-time and after-pulsing, using empirical modeling. These artifacts become increasingly problematic in higher-order correlations. Several approximations were used in the analytical modeling to account for artifacts,

and calibration of artifact parameters using separate measurements was required, adding to the complexity and cost of the technique.

This chapter presents a new approach to higher-order FCS data collection and analysis that addresses the issue of detector artifacts, enables improved time-resolution down to 1 microsecond, and simplifies the theoretical modeling. We introduce simple methods to improve data collection through a combination of sub-binning and bin size selection that can be applied in one-detector or two-detector experiments to remove artifacts without the need to analytically model and calibrate for such effects. By "sub-binning" we refer to the process of dividing a whole sampling interval (bin) from a single detector into non-overlapping partitions (sub-bins) to create separate channels which are treated as if they came from independent detectors. We take the photon count in a sub-bin (or a group of them) to represent the count in the whole bin, and we use all sub-bins for maximum SNR. We show, in single-detector and two-detector experiments, that the SNR is acceptable down to $\sim 1\,\mu\mathrm{s}$ lag times in measuring higher-order correlations up to fourth order, using common fluorophores and low excitation powers. We also present theory modifications specific to multi-detector experiments and sub-binning. The improvements presented here result in a very portable implementation of higher-order FCS that does not require any setup-dependent calibration and can be applied to any FCS experiment just as easily as ordinary autocorrelation analysis.

2.2 Methods

2.2.1 Experiment

Rhodamine 6G-labeled and HPLC purified oligonucleotide of sequence $5'$-$\mathrm{AACCC(T)_8}$ GGGTT-$3'$ (Modification: $5'$ 5(6)CR 6G) was purchased from AnaSpec (Fremont, CA) and diluted to $\sim 1\,\mathrm{nM}$ concentration in $^1\!/_4\,\mathrm{X}$ TE buffer PH 8.0. The sample is referred to as "R6G-oligo" in this report.

A commercial inverted microscope (TE2000-U, Nikon) was modified to perform fluorescence fluctuations spectroscopy. The beam from a $532\,\mathrm{nm}$ continuous wave solid state laser (B&W Te, Newark, DE) was collimated and attenuated to $19.0 \pm 0.2\,\mu\mathrm{W}$ (measured at objective back aper-

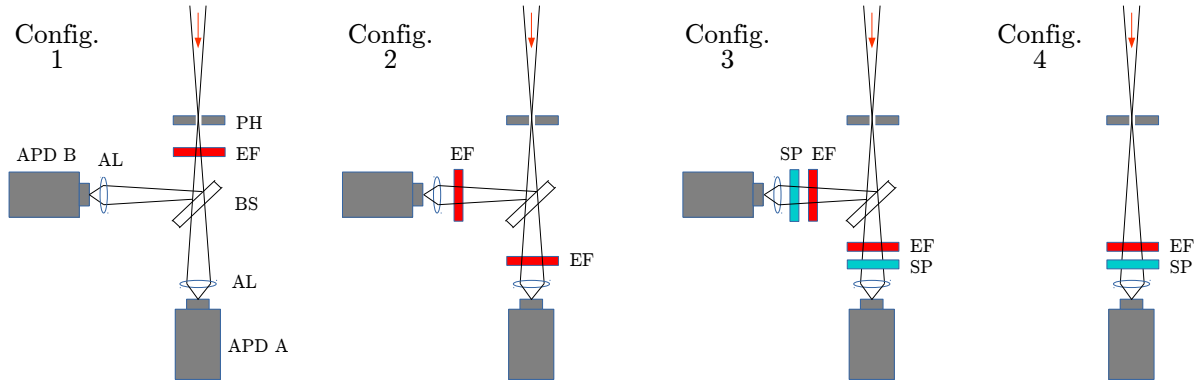

Figure 2.1: In a simple two-detector setup configuration (1), the emission light is directed through a pinhole (PH) and an emission filter (EF), then is split among the detectors using a beamsplitter (BS) and focused on the detection area using aspherical lenses (AL). Setup configurations 2 and 3 are considered to reduce or remove detector cross-talk, with the possible addition of a short-pass filter (SP). Configuration 4 shows a single-detector setup with optical filters similar to those in configuration 3 for similar spectral detection efficiency.

ture) to minimize triplet blinking, then focused by the objective (Plan Fluor $100\times/1.30$ Oil, Nikon) at roughly $5\,\mu\mathrm{m}$ deep into the solution. Peak laser intensity at the center of the probe region was estimated to be $14 \pm 1\,\mathrm{kW/cm^2}$ and dropped to $1/e^2$ of maximum at radius $0.30 \pm 0.02\,\mu\mathrm{m}$ assuming a Gaussian profile. The Fluorescence was collected by the same objective, passed through a dichroic beamsplitter ($550\,\mathrm{nm}$ cut-off midpoint), a $50\,\mu\mathrm{m}$ pinhole, and directed toward one or two avalanche photodiode (APD) detectors (SPCM-AQR-14, EH&G) equipped with aspherical lenses, using one of the four configurations shown in Figure 2.1. In these setup configurations, one or two emission filters (550LP) were combined with one or two short-pass filters (RPE640SP, Omega Optical, Inc., Rattlebrain, VT) to study detector cross-talk as described in Results and discussion. Data was collected for $15\,\mathrm{min}$ to provide sufficient SNR at low excitation power for high order correlations. Molecular brightness was observed to be 125 ± 4 and $264 \pm 8\,\mathrm{kHz}$ per molecule per detector in configurations 3 and 4, respectively. Lidded glass-bottom dishes were used to suppress sample evaporation during measurements. Raw photon arrival times were recorded by a DPC-230 card and SPCM software (Becker & Hicks, Berlin, Germany). Higher order correlations were later calculated using an in-house program written in C.

2.2.2 Simulation

Computer simulations of single-photon detection experiments with known true correlation functions were used to verify the artifact-free correlation computation programs and investigate the effects of finite bin size. By including various types and degrees of detector artifacts in simulation, the artifact removal methods were also investigated. The simulation consisted of a single immobile fluorescent molecule alternating between two brightness levels Q_1 and Q_2 with forward and backward transition rates $k_\mathrm{f} = 2.0 \times 10^4\,\mathrm{s}^{-1}$ and $k_\mathrm{b} = 1.3 \times 10^4\,\mathrm{s}^{-1}$, respectively, as in a two-detector experiment run for a total of 300 seconds. Diffusion was not included in the simulation for simplicity. Theoretical derivation of cumulant-based higher order correlation functions for this stationary single-molecule system appears in Section 3.4. The individual brightness levels were set according to

$$Q_1 = \frac{\bar{I}(1+k)}{1+kq}\,,\;\; Q_2 = qQ_1 \tag{2.6}$$

where $\bar{I} = 50\,\mathrm{kHz}$ was set as the count rate per detector, and $q = Q_2/Q_1 = 0.1$ was set as the brightness ratio. Consistent with the rest of this document, $k = k_\mathrm{f}/k_\mathrm{b}$. For single-detector simulations, $\bar{I} = 100\,\mathrm{kHz}$ and only one of the simulated detection channels was used. $\bar{I}$ was doubled for smoother view of finite-bin-size residuals.

Inverse transform sampling [60] was used to generate state waiting times with exponential distribution. Then the state transition times (from alternately adding the waiting times) were recorded to later calculate the "Ideal" shot-noise-free higher order correlations through segment-by-segment multiplication and integration over time, using the brightness levels Q_1 and Q_2 as known values. These curves are shown in all simulation graphs and labeled as "Ideal" for comparison (see Figures 2.3, 2.6 and 2.8).

Similarly, exponentially distributed waiting times were generated between photon events in each state. Then the photons were randomly distributed among the two detection channels with equal probability, and the photon timestamps were recorded in a file format identical to that produced by our correlator card software. To simulate dead-time artifact in each detector, we discarded any pulse that occurred within $50\,\mathrm{ns}$ after a recorded pulse in that detection channel. To simulate

after-pulsing, a false pulse was generated with a certain probability (0.7, 1, 1.4, or 2% in different simulations) for each recorded pulse, and was placed after the corresponding dead-time period ($50\,\mathrm{ns}$) with an exponential distribution with $25\,\mathrm{ns}$ decay time constant.

2.3 Theory

We outline the relations describing higher order fluorescence correlations for a system of single-species multi-state diffusing molecules, assuming the molecules have the same diffusion constant in all states. This single-species multi-state model can also more generally describe a mixture of reacting and (or purely) non-reacting species when the reaction rates between the non-reacting species are set to zero, with distinct brightness levels and identical diffusion properties for all species assumed. Palmer and Thompson [15] extended the definition of correlation to higher orders using higher-order moments of intensity and they obtained results that depend on lower-order correlations for non-interacting diffusing species. Later, Melnykov and Hall [17], following earlier work on non-correlated cumulant analysis [51, 55, 56], presented a simpler formulation based on cumulants. Here, we follow the latter approach, because, while the moments of the fluorescence data are straightforward to define and calculate directly from experimental data, it is in fact the cumulants that yield compact and useful analytical expressions for systems of diffusing and reacting molecules. Next we describe the relation between models and measurements in single-detector and multi-detector experiments. Finally, normalized correlation functions are presented for each method. A list of useful relations between moments and (factorial) cumulants appears in Section 3.1.

2.3.1 Modeling correlations for molecules in solution

The modeling presented in this section is independent of the number of detectors in an experiment. The relation between the model and measurements in single- and multi-detector experiments is explained in the next section. Here, we assume that if more than one detector is used, the beamsplitters and the detection efficiencies are adjusted such that the effective fluorescence light

intensity (that is, the actual intensity scaled by any efficiency factors) is equal for all detectors. Then take $I(0)$ to denote the effective intensity of the fluorescence light arriving at any detector at lag time zero, and $I(t)$ to denote that intensity at lag time t, and define the random vector $\vec{I} = [I(0), I(t)]$.

Consider a system of diffusing and reacting fluorescent molecules alternating between two or more states with distinct brightness levels, and assume the diffusion constant remains unchanged by reaction. It has been shown [17] for such systems that the bivariate cumulants of intensity can be expressed as a product of reaction and spatial factors (for an alternative derivation, see Section 3.2):

$$\kappa_{m,n}(\vec{I}) = \gamma_{m+n} X_{m,n}(t) Y_{m,n}(t) \tag{2.7}$$

in which the reaction factor $X_{m,n}(t)$ is

$$X_{m,n}(t) = \sum_{s=1}^{J} \sum_{s'=1}^{J} N_s Q_s^m Q_{s'}^n Z_{s',s}(t) \tag{2.8}$$

and the spatial factors γ_k and $Y_{m,n}(t)$ are

$$\gamma_k = \frac{\int_V L^k(\vec{r}) \mathrm{d}^3 r}{\int_V L(\vec{r}) \mathrm{d}^3 r} \tag{2.9}$$

$$Y_{m,n}(t) = \frac{1}{\gamma_{m+n} V_{\mathrm{MDF}}} \int_V \int_V L^m(\vec{r}) L^n(\vec{r}') \frac{\exp\left[-|\vec{r} - \vec{r}'|^2 / 4Dt\right]}{(4\pi D t)^{3/2}} \mathrm{d}^3 r \mathrm{d}^3 r' \tag{2.10}$$

In the above relations, $L(\vec{r})$ is the laser illumination profile normalized to its peak value, or more accurately, the molecular detection function (MDF) defined as the combination of laser intensity distribution, collection point-spread function, and pinhole aperture [61]. J is the number of molecular states, Q_s is the brightness of state s at peak illumination in counts per unit time per molecule, and V is an integration volume that includes the illuminated region. In the limit $V \to \infty$, $V_{\mathrm{MDF}} = \int_V L(\vec{r}) \mathrm{d}^3 r$ approaches the volume of the MDF (observation volume, or probe region), and N_s approaches the average number of molecules in state s in the observation volume. $Z_{s_2,s_1}(t)$ denotes the probability that a particle is found in state s_2 at time t given it was in state s_1

at time 0, and is obtained by solving linear rate equations. The diffusion constant D is assumed to be the same for all molecules in all states.

In experiments, the instantaneous intensities $I(0)$ and $I(t)$ are not directly measured. Instead, their corresponding *integrated intensities* $W(0)$ and $W(t)$ are measured over sampling intervals (bins) of finite size T:

$$W(t) = \int_t^{t+T} I(t')\mathrm{d}t' \tag{2.11}$$

The random vector $\vec{W} = [W(0), W(t)]$ is then defined accordingly. For a short bin size T over which the variations of intensity can be neglected, we get

$$\kappa_{m,n}(\vec{W}) \approx T^{m+n}\gamma_{m+n}X_{m,n}(t)Y_{m,n}(t) \tag{2.12}$$

The is a good approximation for most practical applications (see Results and discussion). A modeling approach with arbitrary bin size and approximate corrections for finite bin size is presented in Section 3.5.

2.3.2 Relating models and measurements in multi-detector and single-detector experiments

For computing higher-order correlations, the discrete nature of photon counting has to be taken into account. We can use n_d independent detectors to obtain cross-correlations of order (p, q) with $p, q \leq n_\mathrm{d}$. The photon count $n_i(t)$ in the ith detector during a bin is a distinct random variable for each detector, i. Therefore we define the vector

$$\vec{n} = [n_1(0), n_2(0), \ldots, n_p(0), n_1(t), n_2(t), \ldots, n_q(t)] \tag{2.13}$$

Using a generalized form of Mandel's formula [62], one can show [63] for a general k-channel experiment that the $\vec{r}$th order cumulant of integrated intensities, $\kappa_{\vec{r}}(W_1, W_2, \ldots, W_k)$, is equal to

the $\vec{r}$th order factorial cumulant of photon counts, $\kappa_{[\vec{r}]}(n_1, n_2, \ldots, n_k)$. This, in the case of two-time correlations with $\vec{n}$ defined in Equation (2.13), yields

$$\begin{aligned}
\kappa_{\vec{1}_{p+q}}(\vec{n}) &= \kappa_{[\vec{1}_{p+q}]}(\vec{n}) \\
&= \kappa_{\vec{1}_{p+q}}\left[W(0), \ldots, W(0), W(t), \ldots, W(t)\right] \\
&= \kappa_{p,q}\left[W(0), W(t)\right] \\
&= \kappa_{p,q}(\vec{W})
\end{aligned} \tag{2.14}$$

where $\vec{1}_{p+q} = (1, 1, \ldots, 1)$ has $p + q$ elements, and the first line holds by definition.

In particular, for the case of two detectors named A and B we have:

$$\begin{aligned}
\kappa_{21}\left[W(0), W(t)\right] &= \kappa_{111}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t)\right] \\
\kappa_{12}\left[W(0), W(t)\right] &= \kappa_{111}\left[n_{\mathrm{A}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right] \\
\kappa_{22}\left[W(0), W(t)\right] &= \kappa_{1111}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]
\end{aligned} \tag{2.15}$$

When only one detector is used in the experiment, let $n(0)$ and $n(t)$ denote the number of photons detected in the bins at lag times 0 and t, respectively. Then define $\vec{n}_{1\mathrm{d}} = [n(0), n(t)]$ accordingly. The relation between fluorescence intensity and photon counting distributions then yields

$$\kappa_{p,q}(\vec{W}) = \kappa_{[p,q]}(\vec{n}_{1\mathrm{d}}) \tag{2.16}$$

As a result, to measure $\kappa_{p,q}(\vec{W})$ experimentally, it is straightforward to first compute the moments of $\vec{n}$ (or $\vec{n}_{1\mathrm{d}}$) based on their definition (see Equation (3.1)), then use the relations between moments and (factorial) cumulants (see Equations (3.3) and (3.4)) to compute $\kappa_{\vec{1}_{p+q}}(\vec{n})$ (or $\kappa_{[p,q]}(\vec{n}_{1\mathrm{d}})$).

2.3.3 Normalized correlations

Given the simple analytical form of cumulants for diffusing molecules in solution, Equation (2.12), we define normalized higher order correlations as

$$g_{m,n}(t) = \frac{\kappa_{m,n}(\vec{W})}{\kappa_{m,0}(\vec{W})\kappa_{0,n}(\vec{W})} \tag{2.17}$$

which, in a multi-detector experiment (or single-detector with sub-binning) becomes, following Equation (2.14),

$$g_{m,n}(t) = \frac{\kappa_{\vec{1}_{m+n}}(\vec{n})}{\kappa_{\vec{1}_m,\vec{0}_n}(\vec{n})\kappa_{\vec{0}_m,\vec{1}_n}(\vec{n})} \tag{2.18}$$

In a single-detector experiment without sub-binning, Equation (2.17) becomes, following Equation (2.16),

$$g_{m,n}(t) = \frac{\kappa_{[m,n]}(\vec{n}_{1\mathrm{d}})}{\kappa_{[m,0]}(\vec{n}_{1\mathrm{d}})\kappa_{[0,n]}(\vec{n}_{1\mathrm{d}})} \tag{2.19}$$

which was obtained by Melnykov and Hall [17]. In this report, we consider single-detector and two-detector experiments with and without sub-binning to examine their success in avoiding detector artifacts.

In the limit $T \to 0$, the normalized correlation function becomes, using Equation (2.12):

$$g_{m,n}(t) = \gamma_{m,n} R_{m,n}(t) Y_{m,n}(t) \tag{2.20}$$

where

$$\gamma_{m,n} = \frac{\gamma_{m+n}}{\gamma_m \gamma_n} \tag{2.21}$$

and

$$\begin{aligned} R_{m,n}(t) &= \frac{X_{m,n}(t)}{X_{m,0}X_{0,n}} \\ &= \frac{\sum_{s=1}^{J}\sum_{s'=1}^{J} N_s Q_s^m Q_{s'}^n Z_{s',s}(t)}{\left(\sum_{s=1}^{J} N_s Q_s^m\right)\left(\sum_{s'=1}^{J} N_{s'} Q_{s'}^n\right)} \end{aligned} \tag{2.22}$$

is the normalized reaction factor.

The importance of the factorized form of Equation (2.20) in the study of reactions or mixtures cannot be overemphasized. Even though the spatial factors $Y_{m,n}(t)$ and $\gamma_{m,n}$ depend on

the specifics of the illumination profile, and $Y_{m,n}(t)$ additionally depends on the diffusion constant, it is usually possible in experiments to normalize $g_{m,n}(t)$ by that of a single-state or pure sample, obtained using the same experimental setup. This removes the dependence on excitation and collection profiles, i.e. no actual determination of $Y_{m,n}(t)$ and $\gamma_{m,n}$ is required, enhancing the fitting accuracy, convenience, and chance of success. In Chapter 4, we show how the model, Equation (2.20), is applied to fully determine all of the parameters in Equations (2.4) for static mixtures and reversible chemical reactions. In this chapter, we discuss measurement techniques for acquiring artifact-free higher-order correlation functions.

2.3.4 Variance of correlations

Understanding the uncertainty in higher-order correlations can help improve signal quality in experiment design, and estimate fitting weights and residuals in data analysis. Taking $\tilde{g}_{m,n}(t)$ to denote a sampling estimator of $g_{m,n}(t)$, the signal-to-noise ratio is defined as

$$\mathrm{SNR}_{m,n} = \sqrt{\frac{\tilde{g}_{m,n}^2}{\mathrm{var}(\tilde{g}_{m,n})}} \tag{2.23}$$

For simplicity, we assume a single species/state with N molecules in the probe region ($N_{\mathrm{FCS}} = 1/g_{11}(0) = N/\gamma_2$ is more commonly used in the literature on conventional FCS) and peak brightness Q. Here, we focus on the limit $t \to 0$ where shot noise is dominant and assume $QT \ll 1$. For a sample of size n we can show [17, 63, 64]:

$$\mathrm{SNR}_{11} \approx \sqrt{n}\gamma_2 QT \sqrt{\frac{N}{N+\gamma_2}} \tag{2.24}$$

Extending the approach to multi-detector experiments of higher order we obtain [63]:

$$\mathrm{SNR}_{21} \approx \sqrt{n}\gamma_3 (QT)^{3/2} \sqrt{\frac{N}{\gamma_3 + 3\gamma_2 N + N^2}} \tag{2.25}$$

Comparing this with $\mathrm{SNR}_{21}^{(1\mathrm{d})}$ reported for single-detector experiment by Melnykov and Hall [17] shows that SNR_{21} is larger by a factor of $\sqrt{2}$ due to the presence of independent channels in multi-detector and/or sub-binning approach. However, for full comparison, the effects of reduced brightness per detector, reduced effective bin size in sub-binning, and increase in SNR by averaging independent combinations of sub-bins (see Results and discussion) should also be considered.

Finally,

$$\mathrm{SNR}_{22} \approx \sqrt{n}\gamma_4 (QT)^2 \sqrt{\frac{N}{\gamma_4 + (4\gamma_3 + 3\gamma_2^2)N + 6\gamma_2 N^2 + N^3}} \tag{2.26}$$

has been derived [63] for multi-detector and/or sub-binning approach. No single-detector case, $\mathrm{SNR}_{22}^{(1\mathrm{d})}$, has been reported yet. As the lowest SNR among the correlation orders considered here, SNR_{22} is maximized at $N \approx 0.22$ with half maxima at 0.016 and 1.9, using Gaussian profile approximation $\gamma_m = m^{-3/2}$.

2.4 Results and discussion

In this section, we describe the detector artifacts that become relevant in the calculation of higher-order correlations, and the scope of these effects. We explain how to eliminate these artifacts using experimental or data analysis methods without any need to model and calibrate the artifact-related parameters which are specific to each experimental setup. Finally, we examine the effects of bin size on the calculation of higher order correlations.

2.4.1 Detector artifacts

Dead-time, after-pulsing, and cross-talk are known artifacts of APD detectors at very short time scales. In this section, we show these artifacts and their native time scale in conventional order-$(1,1)$ correlation functions, which is necessary before we can explore the effects on higher-order correlation functions. We discuss the origin of the cross-talk artifact, and consider experimental removal options for it. At the end of this section, we explain how these artifacts can have dramatic effects on higher-order correlation functions over much longer time scales. In the following sec-

tions, we show these effects on higher-order correlation functions and discuss methods to remove them.

Using conventional order-$(1,1)$ auto- and cross-correlation curves, we can clearly see the three mentioned artifacts in their native time scale (Figure 2.2). Figures 2.2a to 2.2c are obtained using experimental setup configurations 1 to 3, respectively (see Figure 2.1), and Figure 2.2d is obtained using computer simulation. In the auto-correlation curves, detector dead-time is seen as the roughly $50\,\mathrm{ns}$ lag-time interval in which no photon is recorded and the value of $g_{1,1}$ drops to -1 based on its definition. After-pulsing is seen as the large spike immediately after the dead-time, which extends to tens of nanoseconds afterwards, depending on the detector quality. In the absence of these artifacts and rapid processes such as anti-bunching, the $g_{1,1}$ function would have an approximately constant value at lag times below $0.1\,\mu\mathrm{s}$, as shown by the black "Ideal" curve in the simulation panel, (d) (see Methods for description of the "Ideal" curves and the simulation).

It is common practice in conventional second-order FCS experiments to cross-correlate the signal from two detectors to avoid dead-time and after-pulsing artifacts. Excellent agreement between the blue and the black curves in the simulated data (Figure 2.2d) shows that dead-time and after-pulsing are completely removed in order-$(1,1)$ correlation curves by cross-correlating two detectors. In real experiments, however, the cross-correlation curves may show strong cross-talk effects between the two detectors at lag times shorter than $\sim 20\,\mathrm{ns}$ (Figures 2.2a and 2.2b, blue and red curves). Single-photon APD detectors are known to emit a small amount of light in the spectral range 700 to $1000\,\mathrm{nm}$ each time an avalanche of charge carriers is triggered, signifying a photon detection event [65, 66]. This "breakdown flash" of light is reflected from optical surfaces in the experimental setup and picked up by the other detector, resulting in false cross-correlation peaks. The peaks in the forward and reverse cross-correlation curves may show different locations due to differences in the optical path and cable lengths. If the cable length difference is large enough, two cross-talk peaks may be seen in one cross-correlation curve.

Due to its very short native time scale, detector cross-talk is usually of little concern in conventional second-order FCS conducted in molecular reaction and diffusion time scales. However,

detector cross-talk can cause major problems in multi-detector higher-order FCS at any time scale, if not experimentally or computationally removed. Even in second-order FCS, the presence of an uncorrelated noise may cause systematic error in the mean detection intensity and hence the order-$(1,1)$ correlation amplitudes. Therefore it is advisable to minimize optical cross-talk in second- and higher-order FCS when possible. On the other hand, the known methods to completely remove cross-talk experimentally usually involve some loss in detection efficiency or other disadvantages, or may not be compatible with the experiment design. For such cases, we report a data analysis method to avoid cross-talk artifacts (see Two detectors with sub-binning).

Referring to Figure 2.1, the cross-talk artifact is present in setup configurations 1 and 2, but is removed in setup configuration 3. In setup configuration 1, the reflection of the cross-talk photons, evidenced in Figure 2.2a, occurs largely from the emission filter even though a $550\,\mathrm{nm}$ long-pass filter is used. In setup configuration 2, we removed the main emission filter before the beam-splitter, and placed two identical emission filters (550 long pass) before the detectors. As seen in Figure 2.2b, the cross-talk is significantly reduced but not eliminated. The remaining cross-talk must be due to reflections from the tube and objective lenses inside the microscope.

In FCS experiments either long-pass or band-pass emission filters can be used. In setup configuration 1, the cross-talk was worse than shown in Figure 2.2a when a band-pass filter was used as the emission filter, reflecting more cross-talk photons (data not shown). The results of configuration 2 were better than shown in Figure 2.2b with band-pass filters, however, cross-talk was not fully eliminated (data not shown) because band-pass filters are usually not fully efficient over a broad spectral range.

The best results were achieved with setup configuration 3 in which highly efficient $650\,\mathrm{nm}$ short-pass filters were coupled with the long-pass emission filters. Figures 2.2c shows these cross-talk-free results, with a clear view of an anti-bunching decay below $10\,\mathrm{ns}$. The slight asymmetry between the forward and reverse anti-bunching curves is due to optical and electronic path length differences. Correcting this asymmetry was not attempted because of lack of significance in our

current analysis. Besides efficient spectral separation of detectors [67], a polarizing beamsplitter may also be used to help prevent detector cross-talk [68].

This chapter is focused on the calculation of higher-order correlation functions at lag times $1\,\mu$s and beyond. Dead-time, after-pulsing, and cross-talk can have serious effects on higher-order correlations in such longer time scales, far beyond the native time scale of these artifacts seen in order-$(1,1)$ correlation curves. There are two major ways in which the detector artifacts can affect correlation functions. The first way is by creating false correlation (or lack of correlation) *between* bins separated by a lag time. Effects of this type are limited to the native time scales of the artifacts, typically well below $1\,\mu$s, and can appear in order-$(1,1)$ as well as higher-order correlation curves. The second way is by affecting the probability of counting two or more photons *within* a bin. Effects of this type can appear at any lag time, and become significant only in higher-order correlations (as well as non-correlated counting histogram or distribution analysis methods). Correlations of order-$(1,1)$ are not significantly affected in this way because the probability of counting two or more photons within a bin are usually much smaller than the probability of counting one photon, hence the total photon count within a bin is not significantly affected. However, higher-order correlations rely primarily on such higher counts of photons. Here, we are concerned primarily with the latter type of effects in time scales beyond the native time scale of the artifacts.

In the following sections, we explore the effects of detector artifacts on higher-order correlation functions and discuss several approaches for mitigating them. First, we demonstrate the effects of dead-time and after-pulsing artifacts in single-detector experiments. We propose two approaches to avoid these artifacts: either transitioning to a multi-detector experiment, or using a "sub-binning" data analysis approach with one detector. In multi-detector experiments, cross-talk artifact becomes relevant. We propose either removing cross-talk experimentally, and/or using the sub-binning approach when experimental removal is inefficient or impractical.

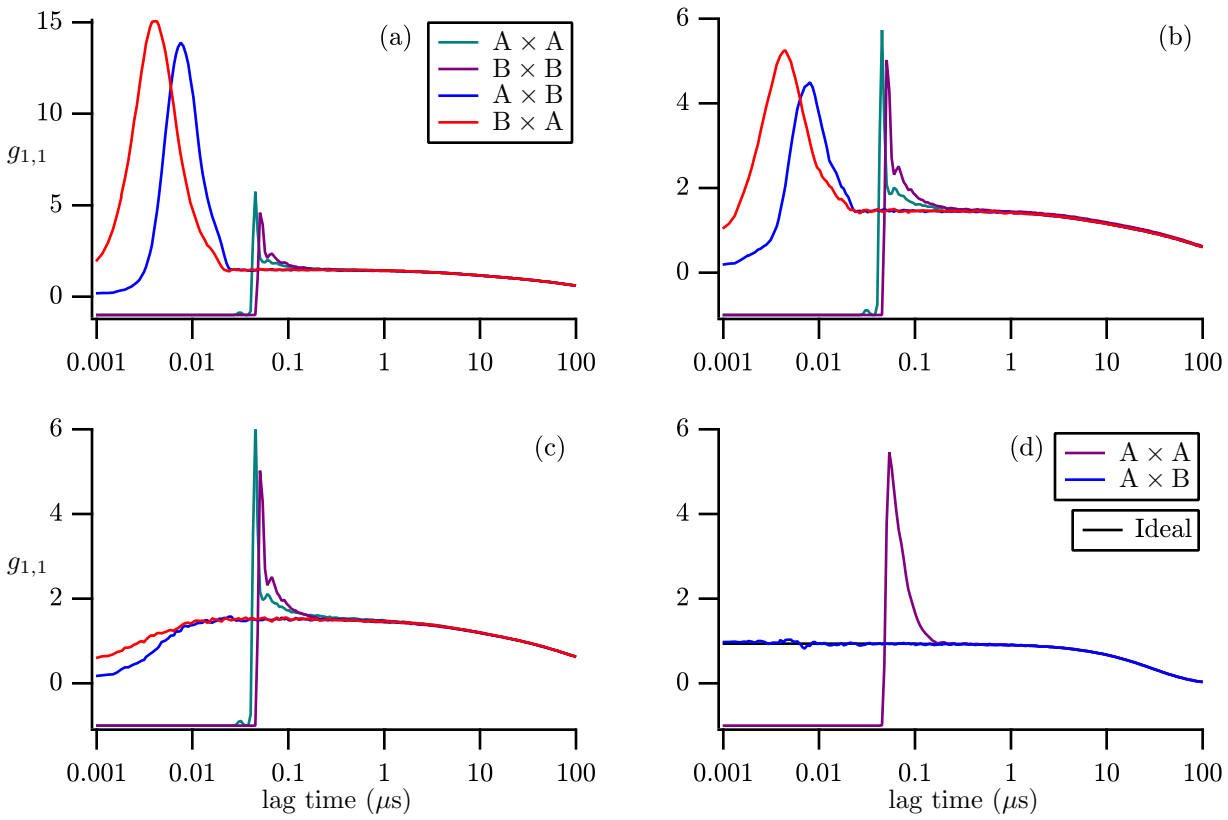

Figure 2.2: Auto-correlation and cross-correlation of two detection channels A and B at very short lag times show detector artifacts. Absence of correlation up to time $\sim 50\,\mathrm{ns}$, and the unusual peaks immediately after, show dead-time and after-pulsing effects in auto-correlation curves A $\times$ A and B $\times$ B. Cross-talk artifacts are seen as peaks extending to $\sim 20\,\mathrm{ns}$ in cross-correlation curves A $\times$ B and B $\times$ A (panels (a) and (b)). Panels (a) through (c) show experimental data using setup configurations 1 through 3, respectively, and share the same color scheme. Panel (d) shows simulated data with dead-time ($50\,\mathrm{ns}$) and after-pulsing (0.7%, $25\,\mathrm{ns}$ exponential decay time). Cross-talk and anti-bunching were not included in simulations.

2.4.2 Single detector without sub-binning

Figure 2.3 shows the higher order correlation curves resulting from the simulation of a single stationary molecule alternating between two brightness states (no diffusion included for simplicity). The cumulant-based definition of correlations may result in negative higher-order correlation amplitudes for this system (contrary to the diffusing multi-particle system). The "Ideal" curves (black, see Methods for details) are in full agreement with the theoretical relations for such system derived in Section 3.4, and fitting to the ideal curves fully and exactly yields the reaction parameters of the simulation (We omit such analysis here. Instead, Chapter 4 presents applications in experimental reacting systems.) We will use these curves to compare with various methods of calculating higher-order correlations from the simulated photon data stream.

In an experiment with only one detector, the higher order correlations are given by Equation (2.19). Each element of $\vec{n}_{1\mathrm{d}} = [n(0), n(t)]$ refers to a (whole) bin at a particular lag time, therefore only two bins are required to calculate any high-order correlation. This is the basis of the single-detector method *without* sub-binning.

In Figure 2.3, the red curves show the results of the single-detector approach without sub-binning. Figure 2.3a shows that the single-detector method without sub-bins works well when no artifact is introduced in simulation. The very small deviation from the black curve is due to finite bin size effects which will be discussed in Bin size selection section. As noted earlier, dead-time and after-pulsing artifacts, if not corrected for, can significantly throw off the results of the single-detector method without sub-binning. The red curves in Figure 2.3b significantly deviate from the "Ideal" curves when a $0.05\ \mu\mathrm{s}$ dead time is introduced into the simulation without any after-pulsing. In Figure 2.3c we have introduced a 1.4% chance of recording a false after-pulse after any detector pulse event. We can see that the deviation of the single-detector method without sub-binning (red curves) from the artifact-free curves (black) is reduced because after-pulsing partially compensates (or slightly over-compensates) for the photons lost in dead time. This "moderate" amount of after-pulsing closely resembles the behavior of real detectors used in our experiments. The after-pulsing probability and its distribution over lag time varies from detector to detector, and strongly depends

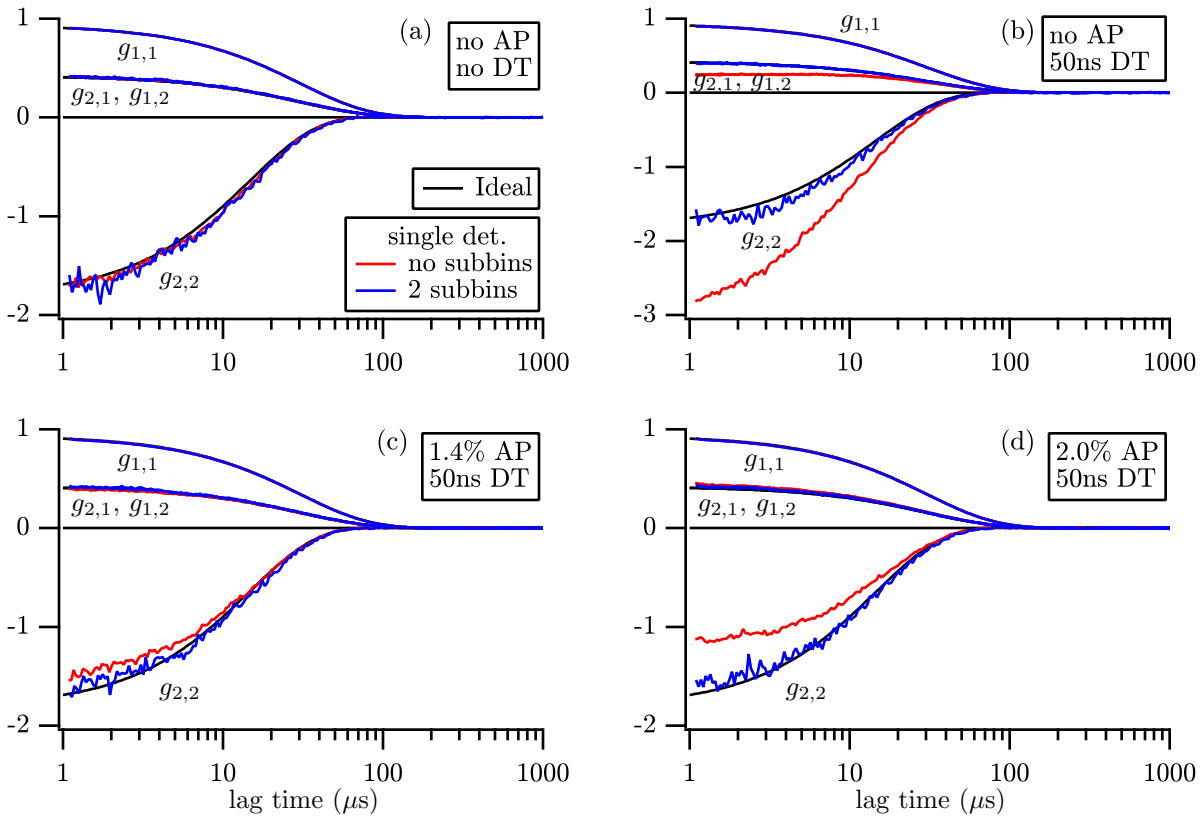

Figure 2.3: Single-detector methods with and without sub-binning are compared in simulated data. (a): Simulation with no artifacts, (b)–(d): with dead-time (ST) and various amounts of after-pulsing (AP). In all panels, the single-detector curves without sub-binning are shown in red, and the single-detector curves with two sub-bins are shown in blue. The "Ideal" noise-free curves are shown in black.

on the detector count rate (data not shown). In Figure 2.3d we have simulated a case of more severe after-pulsing (2%). It is seen that the single-detector method without sub-binning is very sensitive to the after-pulsing probability.

In Figure 2.4a the red curve shows the result of the single-detector method without sub-binning applied to experimental data from R6G-labeled DNA molecules diffusing in solution. On the same graph, the results of a cross-talk-free two-detector method without sub-binning — which are considered to be artifact-free for reasons to be explained — are presented in black as benchmark curves for comparison. The deviation of the single-detector curves without sub-binning from the artifact-free curves seems rather moderate, consistent with a moderate amount of after-pulsing in our detectors, as explained above.

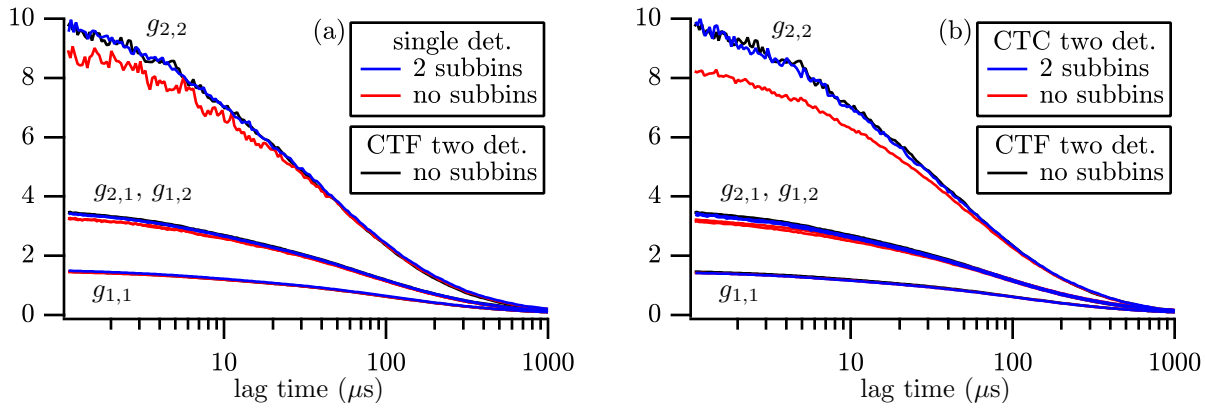

Figure 2.4: Different methods of obtaining higher order correlations are compared using experimental data from R6G-oligo sample. The single-detector method without sub-binning (a, red) suffers from dead-time and after-pulsing artifacts. The single-detector method with sub-binning (a, blue) is free of artifacts. In cross-talk-containing (CTC) two-detector experiment (setup configuration 1), the curves without sub-binning (b, red) are strongly affected by cross-talk artifact, while the those with sub-binning (b, blue) are free of artifacts. The cross-talk-free (CTF) two-detector experiment (setup configuration 3) also yields artifact-free curves without sub-binning (a and b, black).

2.4.3 Two detectors without sub-binning

The photon detection process and detector artifacts are mutually independent in different detection channels. Therefore a multi-detector approach is expected to remove such artifacts. In an experiment with two detectors A and B, the higher order correlations are given by Equation (2.18). For example,

$$g_{2,1}(t) = \frac{\kappa_{1,1,1}(\vec{n})}{\kappa_{1,1,0}(\vec{n})\kappa_{0,0,1}(\vec{n})} \tag{2.27}$$

where $\vec{n} = [n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t)]$ or $[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{B}}(t)]$ (both variants are possible and partially independent, thus averaging them can improve the SNR). Each element of $\vec{n}$ refers to a bin in a particular channel and lag time, therefore three bins are required in this calculation.

As another example,

$$g_{2,2}(t) = \frac{\kappa_{1,1,1,1}(\vec{n})}{\kappa_{1,1,0,0}(\vec{n})\kappa_{0,0,1,1}(\vec{n})} \tag{2.28}$$

where $\vec{n} = [n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)]$, requires four bins. The situation is shown in Figure 2.5, top.

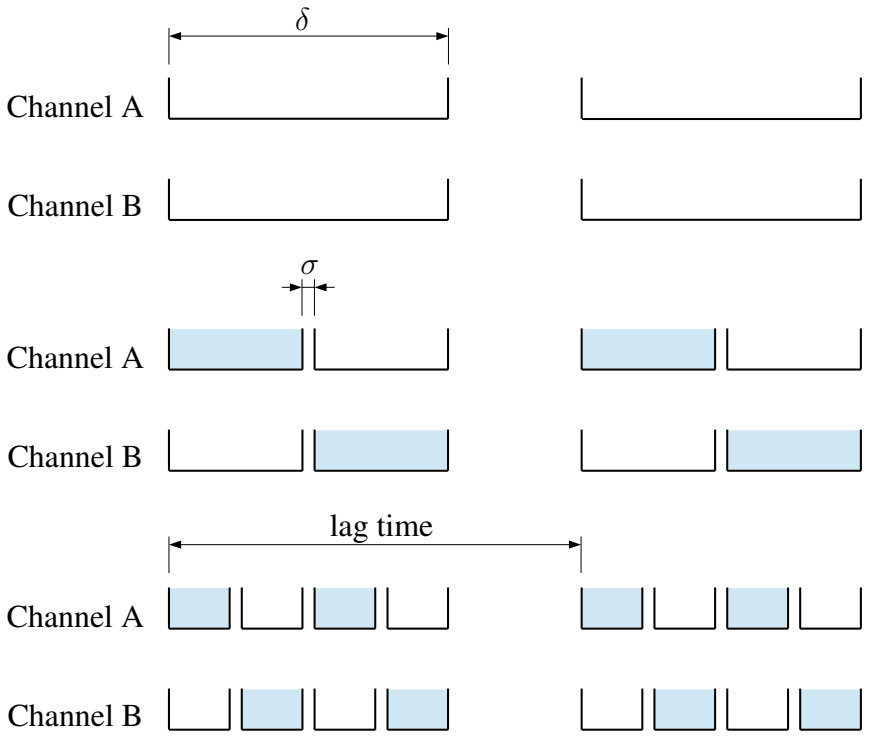

Figure 2.5: The choice of ordinary bins (top), bins with two sub-bins (middle), and bins with four sub-bins (bottom) in a two-detector experiment (also a single-detector experiment, considering only one of the channels).

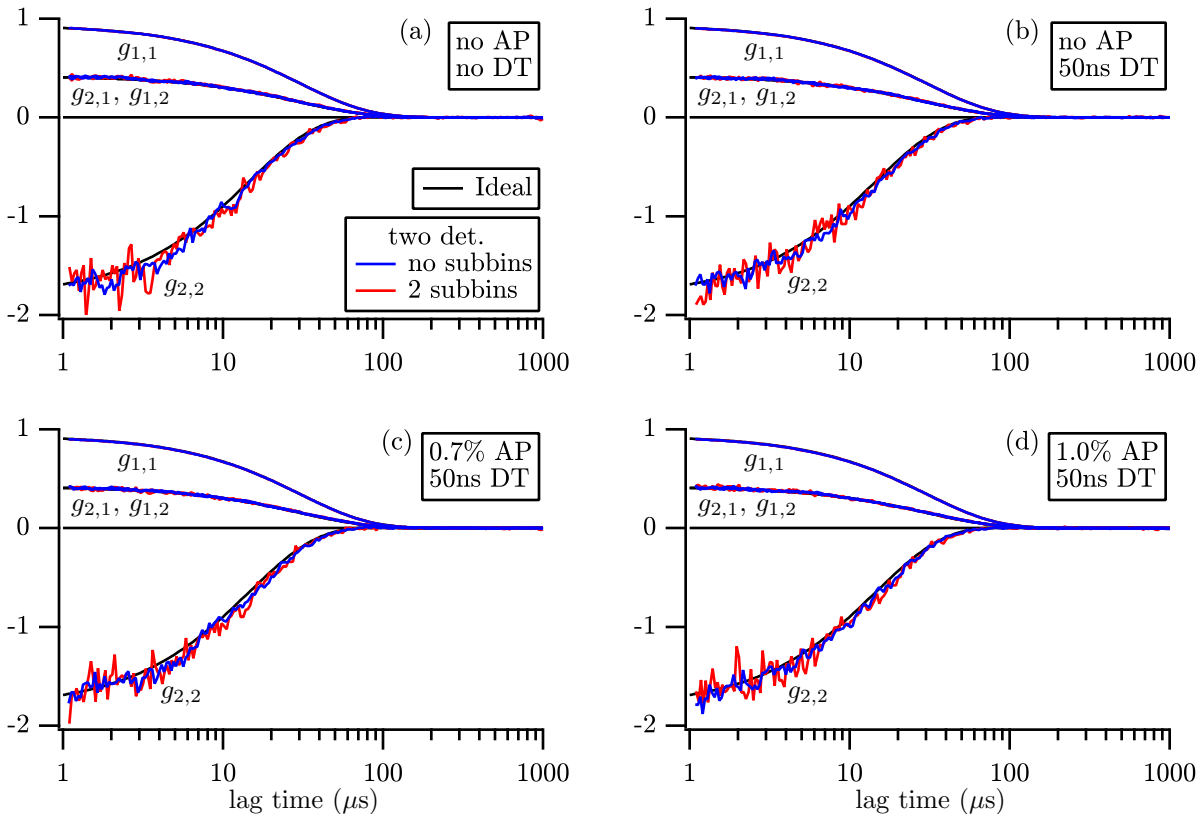

Figure 2.6: Two-detector method with and without sub-binning are robustly resistant to various degrees of dead-time and after-pulsing artifacts in cross-talk-free simulated data. (a): Simulation with no artifacts, (b)–(d): with dead-time and various amounts of after-pulsing. In all panels, the two-detector curves without sub-binning are shown in blue, and the two-detector curves with two sub-bins are shown in red. The "Ideal" noise-free curves are shown in black. When cross-talk is present, only the two-detector method without sub-binning remains artifact-free (see Figure 2.4b).

Figure 2.6 shows the results of the two-detector method without sub-binning in simulation. No detector cross-talk has been introduced in any simulation presented in this chapter. The cross-talk-free two-detector method without sub-binning shows robust resistance to various degrees of dead-time and after-pulsing artifacts (Figure 2.6). To successfully apply the two-detector method without sub-binning to experiments, the cross-talk artifact has to be fully removed experimentally, as in setup configuration 3. The results are shown with black curves in Figure 2.4 (both panels). If cross-talk exists, the two-detector method without sub-binning will be significantly affected (Figure 2.4b, red).

In the experimental results presetented here, we believe that higher-order correlation curves reported as artifact-free are indeed free of the known artifacts based on several pieces of evidence.

First, the artifact-free methods have been tested under various artifact conditions in simulated data and shown consistently in agreement with the ideal curves (Figures 2.3 and 2.6). Second, all three different methods that are found to be artifact-free in simulation, whether single-detector or two-detector, produce identical results (within noise) when applied to experimental data. This is very unlikely to be a matter of coincidence. Finally, Chapter 4 is dedicated to verifying the artifact-free methods in experiments, providing extensive experimental evidence supporting the validity of the methods presented here.

One might suggest using the fit parameters of the amplitudes of higher-order correlation functions to show that they agree with predictions for a single species. However, it should be noted that even the amplitudes of the correlation functions in Equation (2.20) depend on the properties of illumination and detection profiles (MDF as a whole) through $g_{m,n}(0) = \gamma_{m,n}$. A three-dimensional Gaussian beam model predicts $\gamma_m = m^{-3/2}$. This equation is known to be incorrect from fluorescence intensity distribution analysis [47]. Instead, in Chapter 4 we remove the dependence on MDF using a reference measurement and show good agreement of the fit parameters with predictions in two-species mixture experiments.

2.4.4 Two detectors with sub-binning

Experimental removal of cross-talk between detectors usually requires particular setup design and equipment. Even though reduced cross-talk has been achieved in setup configuration 2 only at the cost of an extra emission filter, complete removal of detector cross-talk as in setup configuration 3 usually comes with reduced detection efficiency, to which higher-order correlations are excessively sensitive. Also, it is possible to envision an experiment in which spectral decoupling of the detectors is not possible because the emission spectrum largely overlaps with the cross-talk spectrum. In another situation, one might be interested in analyzing higher-order correlations using pre-existing FCS raw data which may contain cross-talk. For these reasons, a method to remove the effects of detector cross-talk in data analysis may be useful.

Figure 2.4b shows that two-detector curves without sub-binning (red) deviate significantly from the correct curves (black) when cross-talk exists (setup configuration 1). To avoid these artifacts, each whole bin can be divided into two *sub-bins*, with a small gap of size σ between them, as shown in Figure 2.5, middle. Then only the non-overlapping sub-bins, for example shaded or non-shaded in Figure 2.5, are used to obtain the count numbers $n_i(t)$ in multi-detector relations 2.18 (for example $n_{\mathrm{A}}(0)$, $n_{\mathrm{B}}(0)$, $n_{\mathrm{A}}(t)$, etc. in Equations (2.27) and (2.28)). In this case, $n_{\mathrm{i}}(t)$ represents the number of photons in an *effective* bin size, T, equal to the size of one of the two sub-bins. The whole bin size, $\delta = 2T + \sigma$ for two sub-bins, will be relevant in the analysis of bin size effects.

Two-detector method with sub-binning was applied to cross-talk-containing data and the resulting curves are shown in blue in Figure 2.4b, in good agreement with the results of the cross-talk-free experiments (black). Also, Figure 2.6 (red curves) shows using simulated data that this method is resistant to dead-time and after-pulsing artifacts, because of independent channels being used.

Detector cross-talk occurs at a short time-scale, therefore a sufficient gap of size σ between the sub-bins can prevent any cross-talk spillover between non-overlapping bins. However, when the sub-bins are sufficiently large themselves, the likelihood of a false cross-talk count at their boundary is already small compared to true counts elsewhere. (A minimum bin size of $\delta_0 = 1\,\mu\mathrm{s}$ is used, see Bin size selection section.) We have observed that the cross-talk artifact is successfully removed with practically any choice of σ. Figure 2.7 shows the robustness of the results when σ is changed by orders of magnitude. Setting σ strictly equal to zero shows no problem in the presented data. However, in some severe cases of detector cross-talk, we have observed very small deviations with $\sigma = 0$ only (data not shown). For this reason and for definiteness, we have used a value of $\sigma = 0.01\,\mu\mathrm{s}$ in all of our sub-binning results in this report (except Figure 2.7). We recommend a non-zero value for σ on the order of cross-talk time scale, which is usually much smaller than the whole bin size, hence has no significant effect on any analysis results. No particular calibration of this parameter should be usually needed, further enhancing the portability and ease of use of correlator software based on this method.

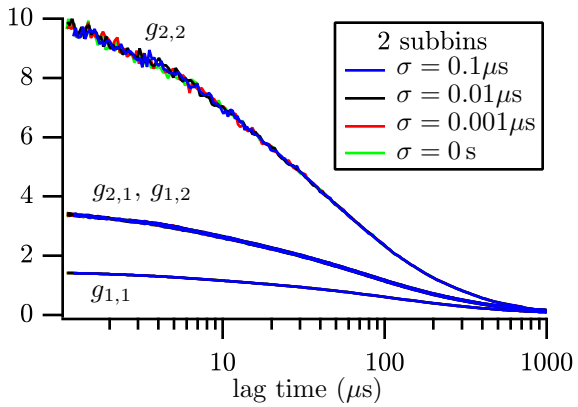

Figure 2.7: The sub-binning method is not sensitive to the gap between the sub-bins.

One can also consider the possibility of dividing a whole bin into more than two sub-bins, as shown in Figure 2.5, bottom. In that Figure, the shaded sub-bins form a whole sub-bin together (their photon counts added together) and the non-shaded sub-bins are similarly grouped together. In this case, $\delta = 2T + (N_s - 1)\sigma$ with N_s being the number of sub-bins in a whole bin. In principle, any even number of sub-bins may be used this way. The SNR is not significantly affected with more sub-bins as long as the the total size of gaps between sub-bins, $(N_s - 1)\sigma$, remains small compared to the whole bin size. There is a systematic variation in the results due to non-zero bin size, which is also affected by the number of sub-bins. We examine these effects in Bin size selection section. Because such variations are small when the bin size effects themselves are small, we use only two sub-bins for simplicity in our analyses and experiments in this report, unless otherwise stated.

2.4.5 Single detector with sub-binning

The photon counts in the two adjacent sub-bins within a whole bin in *one* detection channel are also statistically independent, ignoring the spillover of dead-time and after-pulsing at the boundary between the sub-bins. In fact, the probability of this spillover is usually very small compared to the size of sub-bins. This suggests that the sub-binning method can also be applied to a single detector. Doing so is only a matter of routing the output of a single detector to both inputs of the the two-detector program with sub-binning described in the two-detector case. By creating independent

sub-bins, we virtually convert a single-detector experiment to a two-detector variant and use the multi-detector formulation, Equation (2.18).

The blue curves in Figure 2.4a show the results of the single-detector method with sub-binning in experiment. No deviation is observed from the black curves within noise. The blue curves in Figure 2.3 show this method tested under various conditions of dead-time and after-pulsing in simulation. As expected, only little variations may result from artifact spillover at the boundary between sub-bins. The method is robustly resistant to artifacts considering all the examined cases. In Section 3.6 we present a detailed calculation which demonstrates how sub-binning removes artifacts in one particular case. The costs of sub-binning are remarkably small. The minimum lag-time for which correlations could be calculated at a desired SNR is increased, and some systematic errors (negligible, in practice) are introduced since the channels introduced by sub-binning are offset in time.

2.4.6 Which method to use?

We have described three artifact-free methods for obtaining higher-order correlation functions, namely single- and two-detector methods with sub-binning, and a two-detector method without sub-binning, which can be applied when there is no detector cross-talk. Which method should be preferred in a given experimental application? The answer generally boils down to the specific experimental requirements and conditions of that application. Another factor to consider is the signal-to-noise ratio, which becomes particularly significant in higher-order correlations. In this section, we briefly discuss advantages and disadvantages of each method. But first, we argue why any of the presented methods may be preferred to the previously reported single-detector method involving modeling and calibration of detector artifacts.

To preserve SNR in experiments, earlier works on higher-order correlations [17] and intensity distribution analysis [56] have used a single detector in combination with mathematical modeling and experimental calibration of detector artifacts. However, besides adding to the complexity of the technique, little has been done to demonstrate the success of such corrections in sufficiently

higher orders. In their time-integrated fluorescence cumulant analysis (TIFCA), Wu et al. [56] demonstrate the success of these corrections in cumulants of up to order 2 only, while cumulants of order 4 appear in correlations of order $(2,2)$ needed here. Approximations are used in such models, while it should be noted that any residual errors can become amplified in higher orders. Finally, a major drawback of such artifact modeling is the need for calibration measurement of the artifact parameters in each experimental setup. This whole process requires a separate analysis by itself [69], adding expense and difficulty to the application of this technique.

All of the artifact-removal methods described presented in this chapter benefit from one simple concept: splitting the signal between independent (sub-)bins, whether in a single detection channel or more. This, in principle, results in reduced effective bin size and therefore some loss in SNR. However, the cost in SNR can be usually compensated by wise selection of bin size, described in the following section. Additionally, the independence of the counts in different channels partially makes up for the mentioned loss (though not completely by itself, see Equation (2.25) and its discussion). When multiple detectors are used with sub-binning, (semi-)independent variants of sub-bin selection (for example shaded or non-shaded in Figure 2.5) can be averaged to partially make up for the SNR loss. Benefiting from all of these possibilities, our results show good SNR up to order $(2,2)$ for practical applications using either of the artifact-free methods. In fact, we have improved the time resolution down to $\sim 1\,\mu\mathrm{s}$, unprecedented in higher order correlations, using standard fluorophores (e.g. Rhodamine 6G here, additionally TAMRA and FITC in Chapter 4) at low excitation powers ($\sim 19\,\mu\mathrm{W}$).

The single-detector method with sub-binning can be the first artifact-free method of choice in simple applications. In a single-detector experiment the entire signal is directed to one detector without splitting, resulting in the highest brightness (count rate per molecule per channel). However, splitting a bin into two sub-bins makes the SNR theoretically identical to that of a two-detector method without sub-binning. This would only be true assuming no cross-talk and no detection efficiency lost in the beamsplitter and the cross-talk removal apparatus in the two-detector method. Since this assumption is usually not true in practice, we believe that the single-detector method

with sub-binning can provide the highest SNR among the three artifact-free methods. The technique is also the easiest and least expensive to set up.

Many applications such as dual-color or anisotropy experiments may require more than one detection channel to be used. The two-detector methods with and without sub-binning provide these potentials. They also provide more robust artifact removal than single-detector with sub-binning. In practice, however, the experimenter should try to minimize detection efficiency loss and remove cross-talk between detectors. If experimental cross-talk removal is not possible or preferred, sub-binning in two detectors can be used to remove the cross-talk artifact, at a small cost of SNR (compare noise between two-detector methods with sub-binning and without sub-binning in Figures 2.4b and 2.6. Much SNR is preserved by including all possible combinations of sub-bins). Finally, in two-detector experiments, it is crucial to achieve an equal splitting ratio between the two channels. Usually the signal is not exactly equally split by beamsplitters, and the detectors do not have exactly identical efficiencies. The alignment of the detection pinhole has also been observed to affect the symmetry of the MDF between the detectors. This source of error does not exist in single-detector experiments.

When applied to the analysis of mixtures and reactions, no notable difference is observed between the results obtained using either of the artifact-free methods within statistical uncertainty, as we will show in Chapter 4.

2.4.7 Bin size selection

As shown in the theory section, the derived analytical results become exact only in the limit of infinitesimally small bin size, $T \to 0$. This means that any finite bin size introduces deviations from the theoretical curves, causing systematic errors. On the other hand, SNR is proportional to T in second-order correlation, and scales with higher powers of T in higher-order correlations. Therefore, it is critical to find a good balance between noise and the effects of finite bin size.

In the theory section, we used T to represent the whole bin size without any sub-binning. However, when sub-binning is used, the theoretical results are valid only if T represents an effective

bin size. With any even number of sub-bins, the effective bin size, T, is (almost) equal to half of an undivided bin size, δ (ignoring the gaps between the sub-bins). Without sub-binning, T and δ can be used interchangeably. Inspired by the well-known multiple-tau method [70], we choose the whole bin size, δ, proportional to the lag time, t. The bin size can become too small and the correlations too noisy at microsecond lag times if bin size is strictly linear in lag time. Therefore, we introduce an initial bin size, δ_0, and take

$$\delta = \delta_0 + \rho t \tag{2.29}$$

where ρ is the ratio of bin size to lag time at large lag times.

Figure 2.8 shows the dependence of the correlation curves on ρ and δ_0 using simulated data. Figure 2.8a shows that sufficiently large δ_0 is required to reduce shot-noise at small lag times. In Figure 2.8b, as ρ increases the curves get smoother, though the systematic deviations from the "Ideal" curve increase. The residual deviations from the ideal curve of order $(2, 2)$ are depicted in Figure 2.9a for different values of ρ. Since δ_0 is fixed at 1μs, the deviations near 1μs are similar for all curves. In Figures 2.9b through 2.9d we show the average deviation over 10–100μs for all three correlation orders and for different calculation methods. To ensure that the scaling dependence of residuals on ρ reflects the same functional dependence on bin size, δ_0 has been reduced to 0.1μs in panels b–d. Figure 2.10 shows a similar analysis using experimental data. In this case, an "Ideal" curve is not available, therefore the deviations are measured from the curve with $\rho = 0.44$ as the smoothest curve among the set. In Figures 2.10b–d the average deviation over 10–500μs is shown as a function of ρ.

Comparing different methods of calculating higher-order correlations in Figures 2.9b–d shows that when only 2 sub-bins are used, the systematic deviation is increased compared to the method without sub-binning. However, the additional deviation is almost completely removed with the use of more sub-bins, here as few as six. The number of detectors is not relevant in the analysis of bin size effects (compare single-detector and two-detector methods with 2 sub-bins). As seen on the graphs, the distinction between two and more sub-bins becomes significant only when the

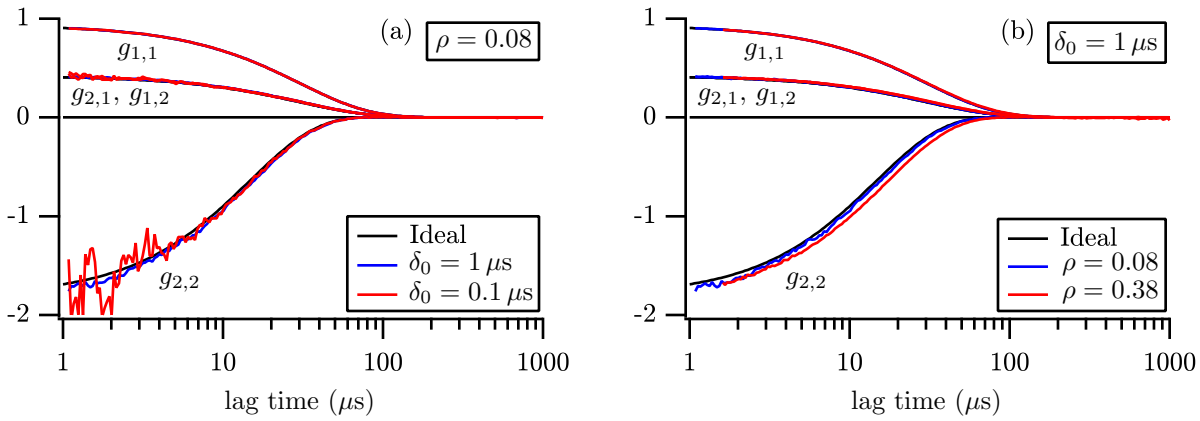

Figure 2.8: Dependence of correlation curves on the bin-size parameters, δ_0 and ρ, is shown using the two-detector method without sub-binning applied to artifact-free simulated data. (a): The starting bin size, δ_0, must be sufficiently large to limit shot noise at short lag times. (b): Increased scaling ratio, ρ, causes smoother curves but more deviation from the ideal curves.

bin size effects are significant themselves. With an appropriate selection of bin size parameters, generally δ_0 much smaller than the characteristic time scale of the fastest fluctuations to be studied and $\rho \ll 1$, we can safely ignore bin size effects and use only two sub-bins for simplicity.

Modeling the exact dependence of correlation functions on bin size requires detailed characterization of the MDF and the resulting multi-point correlation functions integrated over two sampling times, hence can be complicated. However, an approximate analysis using slow and rapid intensity fluctuations presented in Section 3.5 suggests that the leading approximation terms are linear or quadratic in T/t_{F}, where t_{F} is the characteristic time scale of the fluctuations of interest (e.g. t_{R} in the case of a reaction). Therefore, a second-degree polynomial is sufficient to approximate the dependence of the residuals on the bin size. The fit curves in Figures 2.9b–d and Figures 2.10b–d verify this relationship. For order $(1, 1)$, the leading term is purely quadratic in ρ, while the higher-order deviations may contain a linear term as well.

This analysis implies that at small lag times, the bin size has to be much shorter than the time scale of the (fastest) fluctuations of interest. However, this condition can be relaxed at lag times significantly larger than the fluctuations time scale, because the resulting correlations and hence the deviations caused by integration over sampling time are small. Our bin size selection scheme, Equation (2.29), ensures these conditions are met, with a suitable choice of δ_0 and ρ values.

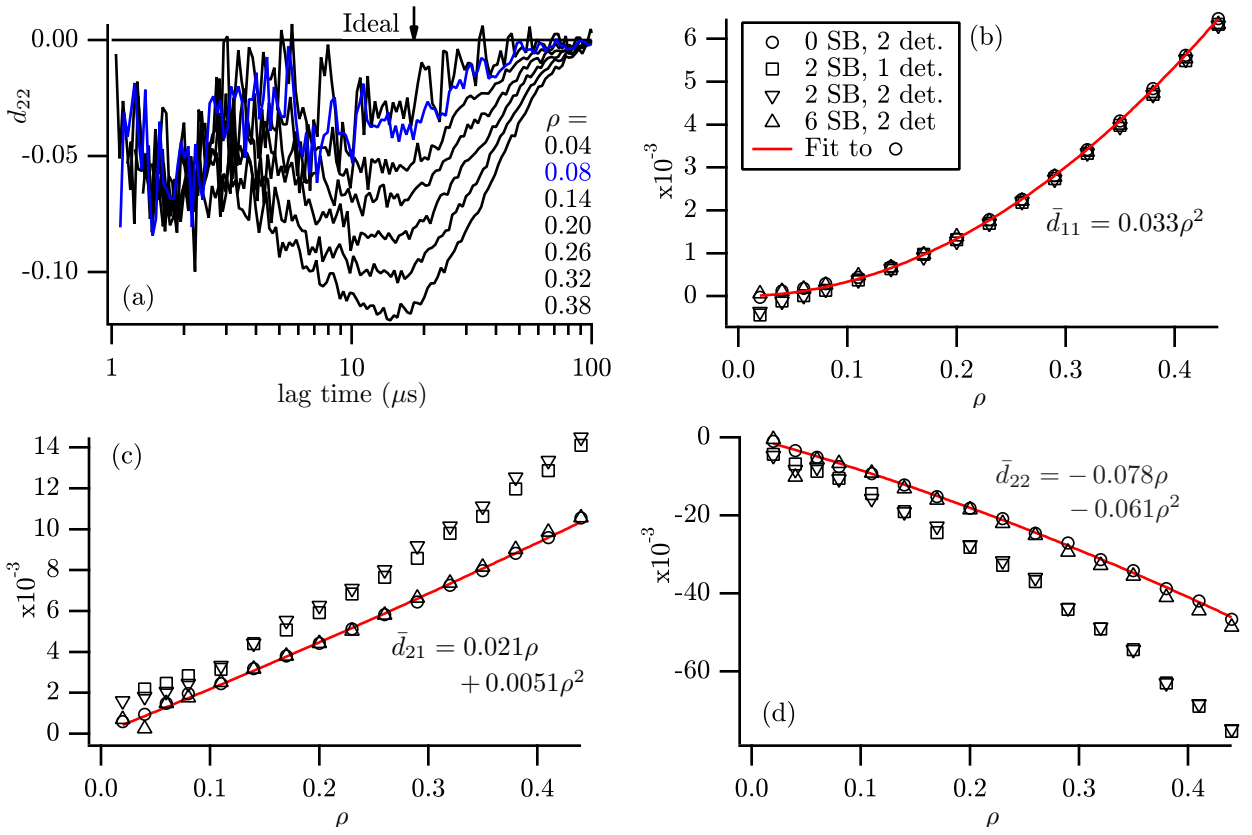

Figure 2.9: The residual deviation of calculated correlations from the "Ideal" curve, $d_{m,n}(t) = g_{m,n}(t) - g^{\text{Ideal}}_{m,n}(t)$, in artifact-free simulated data: Increased scaling ratio, ρ, results in smoother curves but more deviation from the true curves. (a): $d_{22}(t)$ shown for various ρ values and $\delta_0 = 1\,\mu$s using two-detector method without sub-binning. Curves appear in the listed ρ order. (b)–(d): Average $d_{m,n}$ over 10–100μs as a function of ρ for two-detector method (2 det.) without sub-binning (0 SB), with 2 sub-bins (2 SB), and 6 sub-bins (6 SB), as well as single-detector (1 det.) with two sub-bins. The same color scheme and $\delta_0 = 0.1$ apply to panels (b)–(d).

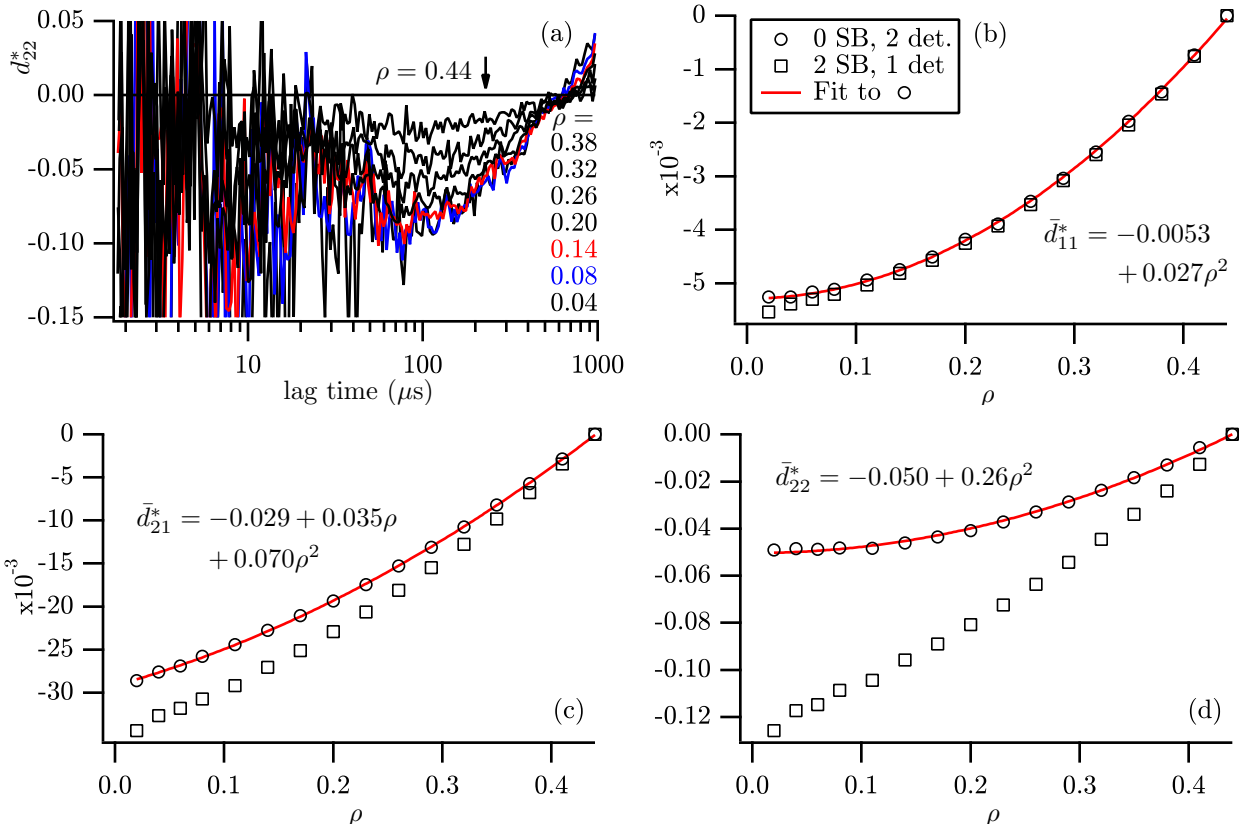

Figure 2.10: The residual deviation of calculated correlations from the curve with $\rho = 0.44$, $d^*_{m,n}(t) = g_{m,n}(t) - g^{\rho=0.44}_{m,n}(t)$, in experimental data from R6G-oligo sample: Increased scaling ratio, ρ, results in smoother curves but more deviation from the true curves in the limit $\rho \to 0$. (a): $d^*_{22}(t)$ shown for various ρ values and $\delta_0 = 1$ using cross-talk-free two-detector experiment without sub-binning. Curves appear in the listed ρ order. (b)–(d): Average $d_{m,n}$ over 10–500μs as a function of ρ for cross-talk-free two-detector method without sub-binning (0 SB, 2 det.) and single-detector method with two sub-bins (2 SB, 1 det.). The same color scheme and $\delta_0 = 0.1$ apply to panels (b)–(d).

Figure 2.9a shows that with $\rho = 0.08$ and $\delta_0 = 1\,\mu$s the maximum deviation in the simulated data is less than 5% of the amplitude of $g_{2,2}$. The average deviation ($\bar{d}_{22} = 0.011$) is less than 8% of the average order-$(2,2)$ correlation function ($\bar{g}_{2,2} = 0.14$) over 10–100μs. In the experimental data (Figure 2.10), the average deviation at $\rho = 0.08$ from the extrapolated value at $\rho = 0$ (i.e. $\bar{d}^*_{22}(0.08) - \bar{d}^*_{22}(0) = 0.0021$) is less than 0.2% of the average correlation function ($\bar{g}_{2,2} = 1.47$) over the same interval, 10–500μs. The deviations arising from pure diffusion (experimental data) are understandably much smaller than those of the reaction simulation, given the significantly different time scales of these processes. This suggest that in modeling the finite-bin-size effects, the diffusion-related correction terms can be safely ignored compared to the reaction-related terms, simplifying the analysis by lifting the dependence on the MDF (see Section 3.5)

Another implication of this analysis is that a second-degree polynomial extrapolation of the correlation values at each lag time, or of the overall fit parameters, toward zero bin size can be done in principle to reduce systematic deviations caused by finite bin size. This approach is particularly useful in situations where shot noise in higher-order correlations does not allow a choice of bin size parameters with negligible effects, depending on the measurement accuracy required. For most of our applications down to 6μs reaction time scale [71], we have found the values $\rho = 0.08$ and $\delta_0 = 1\,\mu$s to have negligible finite-bin-size effects within the stochastic uncertainty. Further adjustment of these parameters can be helpful depending on the application at hand. One has to consider the fluctuation time scale of interest, and the available experimental SNR which depends on the molecular brightness and the experimental time.

2.4.8 Can we reach even shorter time scales?

The question whether we can extend higher-order FCS to time scales shorter than $1\,\mu$s requires two considerations: detector artifacts and shot noise. With a single detector, the artifacts that we considered so far were the result of dead-time and after-pulsing within each sampling time (bin), while the lag time between the bins was larger than the native time scale of the detector artifacts (i.e. ~ 50ns for dead-time and ~ 100ns for after-pulsing in the detectors used in this

study, see Figure 2.2). When lag time is shorter than the native time scale of the artifacts, the resulting effects on higher-order correlations cannot be removed by using two detectors or sub-binning, because there is no way to choose three or more independent (sub-)bins within one or two detection channels. Instead, more than two detectors will be required to remove the artifacts, in principle; that is, three detectors for correlations of order $(2,1)$ and $(1,2)$ and four detectors for order-$(2,2)$ correlation. Although we have not studied cases of more than two detectors using simulation or experiment in this report, the general multi-detector formulation developed in the theory section incorporates such cases as well. In practice, the SNR can become extremely limiting in higher-order correlations at very short lag times, due to very small bin size. Splitting the signal among multiple detectors will result in even lower SNR, given the higher sensitivity of higher-order correlations to molecular brightness value. Added to the challenge might be the necessity to remove detector cross-talk, which usually affects detection efficiency negatively.

Figure 2.11 shows the described effects in sub-microsecond time scale using cross-talk-free two-detector simulated and experimental data, without sub-binning. The after-pulsing effects are more clearly seen in Figure 2.11a as a deviation from ideal curves of order $(2,1)$ and $(1,2)$ near 0.1μs lag time. The after-pulsing effects in the experimental curves of order $(2,1)$ and $(1,2)$ (Figure 2.11b) are less pronounced, because the bulk of the experimental after-pulsing peak (as opposed to the simulated one, Figure 2.2) is very narrow compared to the bin size. To capture such rapid change in the correlation function the bin size had to be much smaller, which would result in extremely noisy curves. When lag time is further reduced to below the dead time, the higher-order correlation curves using two detectors become even more erratic (not shown). If more than two detectors are used, shot noise may still prevent useful computation of fourth-order correlation functions at nanosecond lag times, assuming otherwise similar experimental conditions to those reported here. However, third-order correlation functions may be feasibly calculated and used in such time scales. For example, in Figure 2.11 we have calculated third-order correlations down to $0.1\,\mu$s.

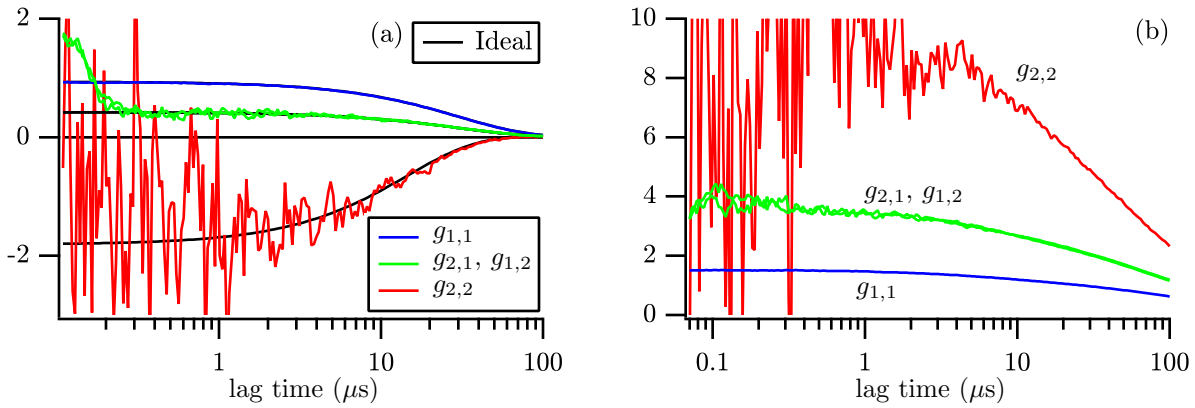

Figure 2.11: Higher-order correlation functions calculated in sub-microsecond time scale using cross-talk-free two-detector method without sub-binning (a): Simulation with 50ns dead-time and 0.7% after-pulsing. Correlations calculated with $\delta_0 = 0.1\mu\text{s}$ and $\rho = 0.08$ (b): Experiment. Correlations calculated with $\delta_0 = 0.065\mu\text{s}$ and $\rho = 0.08$.

Higher-order correlations can be calculated simultaneously with the conventional second-order correlation in a single scan of the photon data, for each lag time. As a result of sub-binning, no particular modification of an existing FCS setup or calibration of detector artifacts is required. Bin size control parameters (δ_0 and ρ), as well as the number of sub-bins, can be adjusted to desired values for best results. With our choice of these values, we demonstrate a time resolution of $\sim 1\ \mu\text{s}$ in correlations up to order $(2, 2)$ using a standard two-detector FCS setup and a modified cross-talk-free setup. Using a single detector is also possible, with no modification to the sub-binning correlator program required and the artifact-removal benefits retained. Our data collection time has been set to 15 minutes, which should be achievable in typical stable systems. However, the experimental time is not critical since the SNR depends only on the square root of the experimental time for all correlation orders. The simplified modeling resulting from cumulant-based formulation, Equation (2.20), enables easy analysis of reaction dynamics without any need to model and calibrate the MDF and the diffusion constant. This feature is more thoroughly discussed and used in Chapter 4.

2.5 Conclusion

With the addition of higher order correlation analysis, FCS is complete and particularly suitable for the time-resolved study of molecular dynamics, be it diffusion, a reversible reaction, or an irreversible one. The time scale of the target dynamics can vary over a wide range, from $\sim 1\,\mu$s achieved in this report for correlations up to fourth order, to the average dwell time of the molecules in the probe region, typically a millisecond or beyond. As demonstrated in Chapter 4, even in reactions occurring at slower rates, or in non-reacting mixtures, higher-order correlation analysis still provides important information about the equilibrium populations, which is beyond the capacity of second-order FCS. To collect sufficient photons at a low excitation intensity, an experimental time of at least a few minutes is recommended. However, this can vary based on the specifics of each system and the time resolution needed. The total analysis time is not otherwise significantly impacted, as the computation of all correlation orders can be performed simultaneously with little additional time for each higher order. With the multi-detector or sub-binning approach suggested in this report, detector artifacts are no longer of any concern, and the technique is easy and ready to use with any existing FCS setup with one or more detectors, or even pre-recorded single-photon data.

Chapter 3

Improving higher-order fluorescence correlation spectroscopy—supporting information

3.1 Useful relations between moments and cumulants

In this section we summarize some relations between multivariate moments, central moments, cumulants, and their factorial counterparts. These relations are useful in the computation of cumulant-based correlation functions and their theoretical derivations and discussions. An introduction to these quantities and full derivations of the relations are presented in [63]. Here, we only list the results. The reader may also refer to [72, 73] for more information.

Let $\vec{X} = (X_1, X_2, \ldots, X_k)$ be a multivariate random vector. For $\vec{r} = (r_1, r_2, \ldots, r_k)$ the $\vec{r}$th *moment* of $\vec{X}$ is defined as

$$\mu'_{\vec{r}} = \mathrm{E}[\prod_{i=1}^{k} X_i^{r_i}] \tag{3.1}$$

where E denotes the expectation operator. The $\vec{r}$th *central moment* of $\vec{X}$ is defined as

$$\mu_{\vec{r}} = \mathrm{E}[\prod_{i=1}^{k} (X_i - \mathrm{E}[X_i])^{r_i}]$$

The multivariate *cumulants* of $\vec{X}$ are defined through their generating function [63, 73]. Take $\kappa_{\vec{r}}$ to denote the $\vec{r}$th cumulant of $\vec{X}$.

For a single random variable, X,

$$\mu_0 = 1$$

$$\mu_1 = 0$$

$$\kappa_0 = 0$$

$$\kappa_1 = \mathrm{E}[X]$$

$$\mu_2 = \kappa_2$$

$$\mu_3 = \kappa_3$$

And for $m \geq 4$:

$$\mu_m = \kappa_m + \sum_{i=2}^{m-2} \binom{m-1}{i} \kappa_{m-1} \mu_i$$

For example,

$$\mu_4 = \kappa_4 + 3\kappa_2^2$$

conversely,

$$\kappa_4 = \mu_4 - 3\mu_2^2$$

and so forth.

For a bivariate distribution, $\vec{X} = (X_1, X_2)$, we have:

$$\mu_{0,0} = 1$$

$$\mu_{1,0} = \mu_{0,1} = 0$$

$$\kappa_{0,0} = 0$$

$$\kappa_{0,1} = \mathrm{E}[X_2]$$

$$\kappa_{1,0} = \mathrm{E}[X_1]$$

As evident from basic definitions, $\kappa_{m,0}$, $\mu_{m,0}$, $\kappa_{0,m}$ and $\mu_{0,m}$ are identical to univariate cases and follow their relations:

$$\kappa_{2,0} = \mu_{2,0} \qquad \kappa_{0,2} = \mu_{0,2}$$

$$\kappa_{3,0} = \mu_{3,0} \qquad \kappa_{0,3} = \mu_{0,3}$$

$$\kappa_{4,0} = \mu_{4,0} - 3\mu_{2,0}^2 \qquad \kappa_{0,4} = \mu_{0,4} - 3\mu_{0,4}^2$$

and so forth.

A useful symmetry property also follows from the definitions: exchanging the subscripts in any valid relation between κs and μs will produce a valid relation. Thus we can find the expression for $\kappa_{n,m}$ (or $\mu_{n,m}$) from that of $\kappa_{m,n}$ (or $\mu_{m,n}$).

We continue with:

$$\mu_{1,1} = \kappa_{1,1}$$

$$\mu_{2,1} = \kappa_{2,1}$$

By symmetry,

$$\mu_{1,2} = \kappa_{1,2}$$

In general, for $m + n \geq 4$,

$$\mu_{m,n} = \kappa_{m,n} + \sum_{\substack{i=0 \\ 2 \leqslant i+j \leqslant m+n-2}}^{m} \sum_{j=0}^{n} \binom{m}{i} \binom{n-1}{j} \kappa_{m-i,n-j} \mu_{i,j}$$

For example,

$$\mu_{2,2} = \kappa_{2,2} + 2\kappa_{1,1}^2 + \kappa_{0,2}\kappa_{2,0}$$

Conversely,

$$\kappa_{2,2} = \mu_{2,2} - 2\mu_{1,1}^2 - \mu_{0,2}\mu_{2,0}$$

As other examples,

$$\mu_{3,1} = \kappa_{3,1} + 3\kappa_{1,1}\kappa_{2,0}$$

$$\mu_{1,3} = \kappa_{1,3} + 3\kappa_{1,1}\kappa_{0,2}$$

Conversely,

$$\kappa_{3,1} = \mu_{3,1} - 3\mu_{1,1}\mu_{2,0}$$

$$\kappa_{1,3} = \mu_{1,3} - 3\mu_{1,1}\mu_{0,2}$$

and so forth.

The conversion relations between the (non-centered) moments of $\vec{X}$, denoted by $\mu'_{m,n}$, and its cumulants will also be useful:

$$\begin{aligned}
\mu'_{1,0} &= \kappa_{1,0} \\
\mu'_{2,0} &= \kappa_{2,0} + \kappa_{1,0}^2 \\
\mu'_{1,1} &= \kappa_{1,1} + \kappa_{1,0}\kappa_{0,1} \\
\mu'_{2,1} &= \kappa_{2,1} + \kappa_{2,0}\kappa_{0,1} + 2\kappa_{1,0}\kappa_{1,1} + \kappa_{1,0}^2\kappa_{0,1} \\
\mu'_{2,2} &= \kappa_{2,2} + 2\kappa_{0,1}\kappa_{2,1} + 2\kappa_{1,0}\kappa_{1,2} + \kappa_{0,1}^2\kappa_{2,0} + \kappa_{1,0}^2\kappa_{0,2} \\
&\quad + 4\kappa_{1,0}\kappa_{0,1}\kappa_{1,1} + \kappa_{2,0}\kappa_{0,2} + 2\kappa_{1,1}^2 + \kappa_{1,0}^2\kappa_{0,1}^2
\end{aligned} \tag{3.2}$$

The inverse relations are:

$$\begin{aligned}
\kappa_{1,0} &= \mu'_{1,0} \\
\kappa_{2,0} &= \mu'_{2,0} - \mu'^2_{1,0} \\
\kappa_{1,1} &= \mu'_{1,1} - \mu'_{1,0}\mu'_{0,1} \\
\kappa_{2,1} &= \mu'_{2,1} - \mu'_{2,0}\mu'_{0,1} - 2\mu'_{1,0}\mu'_{1,1} + 2\mu'^2_{1,0}\mu'_{0,1} \\
\kappa_{2,2} &= \mu'_{2,2} - 2\mu'_{0,1}\mu'_{2,1} - 2\mu'_{1,0}\mu'_{1,2} + 2\mu'^2_{0,1}\mu'_{2,0} + 2\mu'^2_{1,0}\mu'_{0,2} \\
&\quad + 8\mu'_{1,0}\mu'_{0,1}\mu'_{1,1} - \mu'_{2,0}\mu'_{0,2} - 2\mu'^2_{1,1} - 6\mu'^2_{1,0}\mu'^2_{0,1}
\end{aligned} \tag{3.3}$$

The factorial forms of moments or cumulants are particularly useful in the case of discreet random variables. In what follows, suppose X_i can take only non-negative integer values $\{0, 1, \ldots\}$. We tabulate the relations between cumulants and factorial cumulants up to 4th order:

$$\left.\begin{aligned}
\kappa_{[1]} &= \kappa_1 \\
\kappa_{[2]} &= \kappa_2 - \kappa_1 \\
\kappa_{[3]} &= \kappa_3 - 3\kappa_2 + 2\kappa_1 \\
\kappa_{[4]} &= \kappa_4 - 6\kappa_3 + 11\kappa_2 - 6\kappa_1
\end{aligned}\right\}$$

$$\left.\begin{aligned}
\kappa_{[1,1]} &= \kappa_{1,1} \\
\kappa_{[1,2]} &= \kappa_{1,2} - \kappa_{1,1} \\
\kappa_{[1,3]} &= \kappa_{1,3} - 3\kappa_{1,2} + 2\kappa_{1,1} \\
\kappa_{[2,2]} &= \kappa_{2,2} - \kappa_{2,1} - \kappa_{1,2} + \kappa_{1,1}
\end{aligned}\right\} \tag{3.4}$$

Conversely

$$\left.\begin{aligned}
\kappa_2 &= \kappa_{[2]} + \kappa_{[1]} \\
\kappa_3 &= \kappa_{[3]} + 3\kappa_{[2]} + \kappa_{[1]} \\
\kappa_4 &= \kappa_{[4]} + 6\kappa_{[3]} + 7\kappa_{[2]} + \kappa_{[1]}
\end{aligned}\right\}$$

$$\left.\begin{aligned}
\kappa_{1,2} &= \kappa_{[1,2]} + \kappa_{[1,1]} \\
\kappa_{1,3} &= \kappa_{[1,3]} + 3\kappa_{[1,2]} + \kappa_{[1,1]} \\
\kappa_{2,2} &= \kappa_{[2,2]} + \kappa_{[2,1]} + \kappa_{[1,2]} + \kappa_{[1,1]}
\end{aligned}\right\}$$

3.2 Modeling correlations for molecules in solution

In this section we derive the relations describing higher order fluorescence correlations for a system of single-species multi-state simultaneously reacting and diffusing molecules, assuming the molecules have the same diffusion constant in all states. This multi-state system reduces to a multi-species non-interacting system when the reaction rates are set to zero, with identical diffusion

constant for all species assumed. A mixture of reacting and non-reacting species can also be described by setting only a subset of the reaction rates equal to zero.

Palmer and Thompson [15] defined higher-order correlations using higher-order moments of intensity. For mixtures of diffusing molecules, the moment-based definition of correlation functions leads to complex expressions that depend on lower-order correlation functions. No such expression has been proposed to include reactions of diffusing molecules due to the increased complexity of expressions. Later, Melnykov and Hall [17], following the approach developed by Müller in the study of Fluorescence Cumulant Analysis [51], presented a definition of higher-order correlations based on higher-order cumulants. In their derivation, those authors used the additive property of cumulants to arrive at simple factorized expressions for the cumulant-based higher-order correlation functions describing systems of diffusing molecules with reactions. In comparison to moments, the computation of cumulants is only slightly more complicated, with tabulated conversion relations between the two sets of quantities available. However, with the cumulant-based formulation, the resulting theoretical relations for systems of diffusing molecules are greatly simplified compared to the moment-based formulation. Most importantly, the expressions factorize into pure reaction and diffusion parts for systems with independent reaction and diffusion processes. This allows for the experimental removal of any dependence on the molecular detection function (MDF, defined as the combination of laser intensity distribution, collection point-spread function, and pinhole aperture [61]) and on the diffusion constant, using a reference measurement [71].

In this section, rather than presenting the derivation reported by Melnykov and Hall, we outline an alternative derivation starting from simpler premises and use a reverse reasoning process: we start from the explicit integrals following the definition of higher-order moments, (3.11), and find conversion relations by only demanding simple final expressions which are factorized into reaction and diffusion parts, (3.12), without assuming any knowledge about cumulants, their properties, and their relation to moments. Only then, we show that such conversion relations are in general equivalent to the well-known conversion relations between moments and cumulants. We label

this approach the Palmer-Thompson approach because the explicit expression of integrals using Dirac and Kronecker delta functions was inspired by the work of those authors. On the other hand, Melnykov and Hall used the well-known additive property of cumulants to directly derive the simple factorized expressions for a multi-particle system based on those for a single particle. While the approach by Melnykov and Hall is more concise and elegant, the Palmer-Thompson approach is more elaborate and instructive.

Take $I(0)$ to denote the effective intensity of the fluorescence light arriving at the detector(s) at lag time zero, and $I(t)$ to denote that intensity at lag time t. The corresponding *integrated intensities* $W(0)$ and $W(t)$ are defined by integration over a sampling interval (bin) of size T,

$$W(t) = \int_t^{t+T} I(t')\mathrm{d}t' \tag{3.5}$$

The random vectors $\vec{I} = [I(0), I(t)]$ and $\vec{W} = [W(0), W(t)]$ are then defined accordingly. We limit our attention to the case of small bin sizes and obtain results which become exact in the limit $T \to 0$. For a short bin size T over which the variations of intensity can be neglected, we have

$$W(t) \approx TI(t) \tag{3.6}$$

The (m, n)th moment of $\vec{W}$ is defined as

$$\mu'_{m,n}\left[W(0), W(t)\right] = \langle W^m(0)W^n(t)\rangle \tag{3.7}$$

and the (m, n)th moment of $\vec{I}$ is

$$\mu'_{m,n}\left[I(0), I(t)\right] = \langle I^m(0)I^n(t)\rangle \tag{3.8}$$

Thus, using (3.6), we get

$$\mu'_{m,n}(\vec{W}) \approx T^{m+n}\mu'_{m,n}(\vec{I}) \tag{3.9}$$

Having absorbed any detection efficiency factors into I, we can write

$$I(t) = \sum_{s}^{J} Q_s \int_V L(\vec{r}) C_s(\vec{r}, t) \mathrm{d}^3 r \tag{3.10}$$

where $L(\vec{r})$ is the laser illumination profile normalized to its peak value, J is the number of molecular states, Q_s is the brightness of state s at peak illumination in counts per unit time per molecule, $C_s(\vec{r}, t)$ is the concentration of the particles in state s at position $\vec{r}$ and time t, and V is an integration volume that includes the illuminated region. V can be taken to be the entire sample volume containing a fixed number of molecules, M.

Plugging (3.10) into (3.8) we get

$$\mu'_{m,n}\left[I(0), I(t)\right] = \sum_{s_1=1}^{J} \dots \sum_{s_{m+n}=1}^{J} Q_{s_1} \dots Q_{s_{m+n}} \int \dots \int \mathrm{d}^3 r_1 \dots \mathrm{d}^3 r_{m+n} \\ \times L(\vec{r}_1) \dots L(\vec{r}_{m+n}) \mathcal{G}'_{m,n}(s_1, \dots, s_{m+n}, \vec{r}_1, \dots, \vec{r}_{m+n}; t) \tag{3.11}$$

where

$$\mathcal{G}'_{m,n}(s_1, \dots, s_{m+n}, \vec{r}_1, \dots, \vec{r}_{m+n}; t) \\ = \left\langle C_{s_1}(\vec{r}_1, 0) \dots C_{s_m}(\vec{r}_m, 0) C_{s_{m+1}}(\vec{r}_{m+1}, t) \dots C_{s_{m+n}}(\vec{r}_{m+n}, t) \right\rangle$$

The concentration of particles in state s at position $\vec{r}$ at time t is given by

$$C_s(\vec{r}, t) = \sum_{j=1}^{M} \delta[s, s_j(t)] \delta[\vec{r} - \vec{r}_j(t)]$$

where $s_j(t)$ and $\vec{r}_j(t)$ are the state and position of the jth particle at time t, respectively. $\delta(s, s')$ and $\delta(\vec{r} - \vec{r}')$ denote the Kronecker and the Dirac delta functions, respectively.

For use in the upcoming expressions, we define

$$U(s_1, s_2, \vec{r}_1, \vec{r}_2; t) = \sum_{j=1}^{M} \langle \delta[s_1, s_j(0)]\delta[\vec{r}_1 - \vec{r}_j(0)]\delta[s_2, s_j(t)]\delta[\vec{r}_2 - \vec{r}_j(t)]\rangle$$

which can be shown [63] to be

$$= \langle C_{s_1} \rangle \, Z_{s_2,s_1}(t) \frac{\exp\left[-|\vec{r}_1 - \vec{r}_2|^2/4Dt\right]}{(4\pi Dt)^{3/2}}$$

where $Z_{s_2,s_1}(t)$ denotes the probability that a particle is found in state s_2 at time t given it was in state s_1 at time 0, and is obtained by solving linear rate equations (see Section 5.5 for the two-state case). The diffusion constant D is assumed to be the same for all molecules in all states.

Now we define the functions

$$F_{m,n}(t) = \sum_{s_1=1}^{J} \cdots \sum_{s_{m+n}=1}^{J} Q_{s_1} \ldots Q_{s_{m+n}} \int \cdots \int \mathrm{d}^3 r_1 \ldots \mathrm{d}^3 r_{m+n}$$
$$\times L(\vec{r}_1) \ldots L(\vec{r}_{m+n}) \mathcal{F}_{m,n}(s_1, \ldots, s_{m+n}, \vec{r}_1, \ldots, \vec{r}_{m+n}; t) \quad (3.12)$$

Where

$$\begin{aligned}
&\mathcal{F}_{m,n}(s_1, \ldots, s_{m+n}, \vec{r}_1, \ldots, \vec{r}_{m+n}; t) \\
&= \delta(s_1, s_2)\delta(\vec{r}_1 - \vec{r}_2) \ldots \delta(s_1, s_m)\delta(\vec{r}_1 - \vec{r}_m) \\
&\quad \times \delta(s_{m+1}, s_{m+2})\delta(\vec{r}_{m+1} - \vec{r}_{m+2}) \ldots \delta(s_{m+1}, s_{m+n})\delta(\vec{r}_{m+1} - \vec{r}_{m+n}) \\
&\quad \times U(s_1, s_{m+1}, \vec{r}_1, \vec{r}_{m+1}; t)
\end{aligned}$$

The functions $F_{m,n}(t)$ are defined such that when the sums and integrals of the delta functions are directly evaluated, they collapse to the simple factorized form

$$F_{m,n}(\vec{I}) = \gamma_{m+n} X_{m,n}(t) Y_{m,n}(t) \quad (3.13)$$

where we have defined:

$$\gamma_k = \frac{\int_V L^k(\vec{r})\mathrm{d}^3 r}{\int_V L(\vec{r})\mathrm{d}^3 r}$$

The reaction factor $X_{m,n}(t)$ is

$$X_{m,n}(t) = \sum_{s=1}^{J}\sum_{s'=1}^{J} N_s Q_s^m Q_{s'}^n Z_{s',s}(t) \tag{3.14}$$

and the spatial factor $Y_{m,n}(t)$ is:

$$Y_{m,n}(t) = \frac{1}{\gamma_{m+n} V_{\mathrm{MDF}}} \int_V \int_V L^m(\vec{r}) L^n(\vec{r}') \frac{\exp\left[-|\vec{r}-\vec{r}'|^2/4Dt\right]}{(4\pi Dt)^{3/2}} \mathrm{d}^3 r \mathrm{d}^3 r'$$

In the limit $V,\ M \to \infty$, V_{MDF} approaches the volume of the molecular detection function (observation volume, or probe region), and N_s approaches the average number of molecules in state s in the observation volume.

It is possible to obtain conversion relations between $\mu'_{m,n}(\vec{I})$ and $F_{m,n}(t)$, and show in general that those relations are identical to the relations between moments and cumulants of a bivariate distribution. This derivation and proof turns out to be a lengthy yet instructive task which is presented in a separate document [63]. The concluding result is that $F_{m,n}$ are indeed the cumulants of $\vec{I}$:

$$F_{m,n}(t) = \kappa_{m,n}(\vec{I})$$

Finally, we have, parallel to (3.9),

$$\kappa_{m,n}(\vec{W}) \approx T^{m+n}\kappa_{m,n}(\vec{I})$$

for a small T. Thus, using (3.13),

$$\kappa_{m,n}(\vec{W}) \approx T^{m+n}\gamma_{m+n} X_{m,n}(t) Y_{m,n}(t) \tag{3.15}$$

3.3 Some computational notes

Here we discuss some computational details implemented in our correlator program.

3.3.1 Minimum lag time

The starting lag time, t_1, must be greater than the smallest bin size plus a safety margin of the size of the gap between sub-bins, σ, to properly avoid bin overlap and cross-talking artifacts. Therefore, using $\delta = \delta_0 + \rho t$ for the bin size, we have $t_1 \geq \delta_0 + \rho t_1 + \sigma$, or

$$t_1 \geq \frac{\delta_0 + \sigma}{1 - \rho}$$

This gives the minimum starting lag time used in our graphs. Earlier, we observed no significant spillover of artifacts between two sub-bins within a whole bin, even when applied to a single detector. For the same reasons, we observe no significant artifacts here between two bins at a lag time as small as t_1, using the selected bin size parameter values $\rho = 0.08$ and $\delta_0 = 1\,\mu$s.

3.3.2 Computation of cumulants

In the two-detector calculation of higher order correlations with or without sub-binning, or in single-detector with sub-binning, the (sub-)bins are all statistically independent (ignoring artifacts). The cumulants that appear in multi-detector correlations (Equation (2.18)), such as $\kappa_{\vec{1}_{m+n}}(\vec{n})$, can be calculated using ordinary cumulants, because they are equal to factorial cumulants of order $\vec{1}_{m+n}$, that is, $\kappa_{\left[\vec{1}_{m+n}\right]}(\vec{n}) = \kappa_{\vec{1}_{m+n}}(\vec{n})$. To compute ordinary cumulants, we use their relation to ordinary moments (Equations (3.3)). For example, assuming $\vec{n} = [n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t)]$,

$$\begin{aligned}
\kappa_{1,1,1}(\vec{n}) &= \kappa_{2,1}(\vec{W}) \\
&= \mu'_{2,1}(\vec{W}) - \mu'_{2,0}(\vec{W})\mu'_{0,1}(\vec{W}) - 2\mu'_{1,0}(\vec{W})\mu'_{1,1}(\vec{W}) + 2\mu'^2_{1,0}(\vec{W})\mu'_{0,1}(\vec{W}) \\
&= \mu'_{1,1,1}(\vec{n}) - \mu'_{1,1,0}(\vec{n})\mu'_{0,0,1}(\vec{n}) - 2\mu'_{1,0,0}(\vec{n})\mu'_{0,1,1}(\vec{n}) + 2\mu'_{1,0,0}(\vec{n})\mu'_{0,1,0}(\vec{n})\mu'_{0,0,1}(\vec{n})
\end{aligned}$$

where we have used Equation (2.14), and its parallel for moments:

$$\mu'_{p,q}(\vec{W}) = \mu'_{\vec{1}_{p+q}}(\vec{n})$$

We compute each ordinary moment, $\mu'_{\vec{1}_{p+q}}(\vec{n})$, using by averaging over all possible combinations of (sub-)bins in a two-detector method, to maximize signal-to-noise ratio. For example, to calculate $\mu'_{2,1}(\vec{W})$ in a two-detector method with two sub-bins, we average over the following eight variants:

$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0), n^*_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n^*_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0), n_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0), n_{\mathrm{B}}(t)\right]$$

where the asterisk indicates the shaded sub-bin(s) as shown in Figure 2.5. For example, $n^*_{\mathrm{A}}(t)$ stands for the number of photons counted in the shaded sub-bin (or sub-bins, if more than two sub-bins are used,) of channel A at lag time t, and $n_{\mathrm{A}}(t)$ stands for the number of photons in the non-shaded sub-bin(s). Notice that the sub-bins at a certain lag time are either both asterisked or both not asterisked.

Because of the fact that the average number of photons per bin is usually much smaller than one, the computation algorithm becomes much more time efficient if the loop scans over photons, rather than bins. To further keep the program simple and time-efficient, we choose a "primary channel" and a "primary lag time" for which all photons are scanned, and the photons in all other channels and/or lag times are counted only if the primary (whole) bin contains at least a photon. Our primary channel and lag time are chosen to be channel A and lag time 0. For example, in calculating

$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t)\right] = \frac{\sum n_{\mathrm{A}}(0) n_{\mathrm{B}}(0) n_{\mathrm{A}}(t)}{N_{\mathrm{bins}}}$$

only the non-zero terms are calculated and summed in the numerator. The denominator, N_{bins}, is the total number of bins, independently found by dividing the whole experimental time by a bin size.

In some cases, the scanning of photons based on a primary channel and lag time causes a limitation to the possible variants that can be calculated and averaged in a single execution of the program. For example, consider the following variants in the calculation of $\mu'_{1,2}(\vec{W})$:

$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{A}}(t), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1}\left[n_{\mathrm{A}}(0), n^*_{\mathrm{A}}(t), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1}\left[n^*_{\mathrm{A}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]$$
$$\cancel{\mu'_{1,1,1}\left[n_{\mathrm{B}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]}$$
$$\cancel{\mu'_{1,1,1}\left[n^*_{\mathrm{B}}(0), n^*_{\mathrm{A}}(t), n^*_{\mathrm{B}}(t)\right]}$$
$$\cancel{\mu'_{1,1,1}\left[n_{\mathrm{B}}(0), n^*_{\mathrm{A}}(t), n^*_{\mathrm{B}}(t)\right]}$$
$$\cancel{\mu'_{1,1,1}\left[n^*_{\mathrm{B}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]}$$

The first four variants are normally calculated. However, the last four do not involve any bin/sub-bin of channel A at lag time 0 (the primary channel and lag time), thus are not calculated. The average is then calculated over four variants only.

To calculate $\mu'_{2,0}(\vec{W})$ we average the following two variants

$$\mu'_{1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0)\right]$$
$$\mu'_{1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0)\right]$$

However, to calculate $\mu'_{0,2}(\vec{W})$ the variants at lag time t do not involve the primary channel and lag time, thus are not available. We therefore simply use $\mu'_{0,2}(\vec{W}) = \mu'_{2,0}(\vec{W})$.

To finish the discussion of these variants, we list the eight variants used to calculate $\mu'_{1,1}(\vec{W})$:

$$\mu'_{1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1}\left[n_{\mathrm{A}}(0), n^*_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1}\left[n^*_{\mathrm{A}}(0), n_{\mathrm{A}}(t)\right]$$
$$\mu'_{1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1}\left[n_{\mathrm{A}}(0), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1}\left[n^*_{\mathrm{A}}(0), n_{\mathrm{B}}(t)\right]$$

and the four variants we have used to calculate $\mu'_{2,2}(\vec{W})$:

$$\mu'_{1,1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1,1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0), n^*_{\mathrm{A}}(t), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1,1,1}\left[n_{\mathrm{A}}(0), n_{\mathrm{B}}(0), n^*_{\mathrm{A}}(t), n^*_{\mathrm{B}}(t)\right]$$
$$\mu'_{1,1.1.1}\left[n^*_{\mathrm{A}}(0), n^*_{\mathrm{B}}(0), n_{\mathrm{A}}(t), n_{\mathrm{B}}(t)\right]$$

These are not affected by the choice of a primary channel and lag time.

3.4 Modeling correlations for a stationary reacting molecule

In our simulation we consider a stationary single particle alternating between J states of fluorescence brightness Q_i with no spatial motion. Following the approach and notation of Section 3.2, the fluorescence intensity at time t is given by

$$I(t) = \sum_{i=1}^{J} Q_i \delta[i, s(t)]$$

where $s(t)$ is the state of the particle at time t, and $\delta[\,,]$ denotes the Kronecker delta function.

The mean intensity is given by

$$\begin{aligned}\langle I(t)\rangle &= \sum_{i=1}^{J} Q_i \langle \delta[i, s(t)]\rangle \\ &= \sum_{i=1}^{J} Q_i P(i)\end{aligned}$$

where $P(i)$ is the probability of finding the particle in state i.

The mth power of intensity is given by

$$I^m(t) = \sum_{i=1}^{J} Q_i^m \delta[i, s(t)]$$

because the particle is only at one state $s(t)$ at any moment t.

The (m, n)th moment of the two-time intensity vector $\vec{I} = [I(0), I(t)]$ is, therefore,

$$\begin{aligned}\mu'_{m,n}(\vec{I}) &= \langle I^m(0) I^n(t)\rangle \\ &= \left\langle \sum_{i=1}^{J}\sum_{j=1}^{J} Q_i^m Q_j^n \delta[i, s(0)]\delta[j, s(t)] \right\rangle \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J} Q_i^m Q_j^n \langle \delta[i, s(0)]\delta[j, s(t)]\rangle\end{aligned}$$

Thus we need to calculate the following expected value:

$$\begin{aligned}\langle \delta[i, s(0)]\delta[j, s(t)]\rangle &= P(i, j; t) \\ &= P(j|i; t)P(i)\end{aligned}$$

which is the joint probability that the particle is found in state i at time 0 and in state j at time t. It depends on the conditional probability that the particle is found in state j at time t given it was in state i at time 0:

$$P(j|i; t) = Z_{j,i}(t)$$

which is calculated for two states in Section 5.5.

The two-time moments of intensity then become

$$\begin{aligned} G'_{m,n}(t) &:= \mu'_{m,n}[I(0), I(t)] \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J} P(i) Q_i^m Q_j^n Z_{j,i}(t) \end{aligned} \tag{3.16}$$

Evidently, a *moment-based* correlation function for this system of a single non-diffusing molecule results in a similar expression to Equation (3.14), the reaction function that appears in the *cumulant-based* correlation function for a system of diffusing and reacting molecules in solution. Since we are going to test our cumulant-based correlation calculation program on a simulated stationary single-particle system, we should proceed to calculate the two-time cumulants of intensity for this system. This will make the resulting expressions more complicated, and even negative correlation amplitudes may appear, which may be hard to interpret intuitively. The conversion is done via the standard conversion relations (3.3) between bivariate moments, $G'_{m,n}$, and bivariate cumulants, $F_{m,n}$,of $\vec{I}$. Then, assuming small binning time T, we can calculate the cumulant-based higher-order correlation functions

$$g_{m,n}(t) = \frac{F_{m,n}(t)}{F_{m,0}F_{0,m}}$$

The results are summarized below for a single two-state immobile molecule:

$$g_{1,1}(t) = \frac{k(1-q)^2}{(1+kq)^2} e^{-t/t_{\mathrm{R}}}$$

$$\begin{aligned} g_{2,1}(t) &= g_{1,2}(t) \\ &= \frac{(k-1)(1-q)}{1+kq} e^{-t/t_{\mathrm{R}}} \end{aligned}$$

$$g_{2,2}(t) = \frac{(k-1)^2}{k} e^{-t/t_R} - 2e^{-2t/t_{\mathrm{R}}} \tag{3.17}$$

where

$$k = \frac{k_\mathrm{f}}{k_\mathrm{b}} = \frac{P(2)}{P(1)}$$

and

$$1/t_\mathrm{R} = k_\mathrm{f} + k_\mathrm{b}$$

k_f and k_b indicate forward and backward (reverse) rates, respectively.

In summary, the moment-based correlations for a single reacting particle have the simple known forms with positive amplitudes. In the thermodynamic limit, the cumulant-based correlations for a sample of many reacting particles converge proportionally to the moment-based correlations for a single particle (see the Melnykov-Hall derivation [17, 63]), hence they are also simple and positive. On the contrary, the cumulant-based correlations for a single particle, and the moment-based correlations for many particles, will both be non-trivially complex expressions with positive or negative amplitudes. Cumulants and centered moments are identical up to the third order, hence the complexities start to show up in the forth order and above, as in Equation (3.17).

3.5 Finite-bin-size effects

Recall the bivariate moments

$$\mu'_{m,n}\left[W(0), W(t)\right] = \langle W^m(0) W^n(t) \rangle \tag{3.18}$$

of the integrated intensity over sampling time (bin size), T,

$$W(t) = \int_t^{t+T} I(t')\mathrm{d}t' \tag{3.19}$$

Assuming non-overlapping bins ($t > T$), we get, by substituting (3.19) into (3.18),

$$\langle W^m(0)W^n(t)\rangle = \int_0^T \mathrm{d}t_1 \ldots \int_0^T \mathrm{d}t_m \int_t^{t+T} \mathrm{d}t_{m+1} \ldots \int_t^{t+T} \mathrm{d}t_{m+n} \langle I(t_1)I(t_2)I(t_3)\ldots I(t_r)\rangle \tag{3.20}$$

Notice that the integrand in (3.20) is symmetric with respect to the exchange of t_i variables. Therefore we can assume

$$0 \leq t_1 \leq \ldots \leq t_m \leq T < t \leq t_{m+1} \leq \ldots \leq t_{m+n} \leq t+T$$

and write

$$\langle W^m(0)W^n(t)\rangle = m!n! \int_{0\leq t_1\leq\ldots\leq t_m\leq T} \mathrm{d}t_1 \ldots \int \mathrm{d}t_m \int_{t\leq t_{m+1}\leq\ldots\leq t_{m+n}\leq t+T} \mathrm{d}t_{m+1} \ldots \int \mathrm{d}t_{m+n}\, G'_{\vec{1}_{m+n}}(t_2-t_1, t_3-t_2, \ldots, t_{m+n}-t_{m+n-1}) \tag{3.21}$$

where we have defined the r-point correlation function

$$G'_{\vec{1}_r}(t_2-t_1, t_3-t_2, \ldots, t_r-t_{r-1}) = \langle I(t_1)I(t_2)I(t_3)\ldots I(t_r)\rangle \tag{3.22}$$

which, assuming a stationary process, depends only on time differences of the ordered t_i. Therefore, we consider a change of variables from $\{t_1, t_2, \ldots, t_{m+n}\}$ to $\{t_1, \tau_2, \ldots, \tau_{m+n}\}$ where $\tau_i = t_i - t_{i-1}$ for $i \geq 2$. The integral (3.21) becomes

$$\begin{aligned}
&\langle W^m(0)W^n(t)\rangle = \\
&m!n! \int \mathrm{d}t_1 \int \mathrm{d}\tau_2 \ldots \int \mathrm{d}\tau_m \qquad \int \mathrm{d}\tau_{m+1} \ldots \int \mathrm{d}\tau_{m+n} \qquad G'_{\vec{1}_{m+n}}(\tau_2, \tau_3, \ldots, \tau_{m+n}) \\
&\quad t_1, \tau_2, \tau_3, \ldots, \tau_m \geq 0 \qquad t_1+\tau_2+\ldots+\tau_m+\tau_{m+1} \geq t \\
&\quad t_1+\tau_2+\ldots+\tau_m \leq T \qquad \tau_{m+2}, \tau_{m+3}, \ldots, \tau_{m+n} \geq 0 \\
&\qquad\qquad\qquad\qquad t_1+\tau_2+\ldots+\tau_{m+n} \leq t+T
\end{aligned} \tag{3.23}$$

Since $G'_{\vec{1}_{m+n}}(\tau_2, \tau_3, \ldots, \tau_{m+n})$ does not depend on t_1, the integral over t_1 can already be taken from $\max(0, t-\tau_2-\tau_3-\ldots-\tau_m-\tau_{m+1})$ to $\min(T-\tau_2-\tau_3-\ldots-\tau_m, t+T-\tau_2-\tau_3-\ldots-\tau_{m+n})$.

In general, to determine the exact correlation functions with arbitrary finite bin size, one has to determine $G'_{\vec{1}_r}(\tau_2, \tau_3, \ldots, \tau_r)$ for the system of interest, and calculate the integral (3.21) or (3.23) analytically or numerically. Then $\mu'_{m,n}[W(0), W(t)]$ can be converted to cumulants, and normalized higher-order correlations will be obtained using their definition. Alternatively, one can use a Taylor expansion of $G'_{\vec{1}_{m+n}}(\tau_2, \tau_3, \ldots, \tau_{m+n})$ assuming this function changes slowly compared to the bin size, then find the first few correction terms to be added to the zero-bin-size relations.

In typical FCS experiments with diffusing molecules, $G'_{\vec{1}_{m+n}}(\tau_2, \tau_3, \ldots, \tau_{m+n})$ depends on the molecular detection function which has been observed to deviate significantly from the Gaussian approximation, particularly its higher-order integrals γ_m which appear in higher-order correlation functions. Therefore, $G'_{\vec{1}_{m+n}}(\tau_2, \tau_3, \ldots, \tau_{m+n})$ can be difficult to characterize analytically. In the following sections, we show that the Taylor expansion approach is valid for two important limiting cases which cover typical molecular reaction applications. In the first case, we assume that the variations in $I(t)$ are much slower than the selected bin size, T. In the second case, we show using the model of a stationary reacting molecule that even though $I(t)$ is discontinuous, the multi-point correlation functions, $G'_{\vec{1}_r}(\tau_2, \tau_3, \ldots, \tau_r)$, are smooth and differentiable. It can be a valid assumption for most practical applications that a Taylor expansion of $G'_{\vec{1}_r}(\tau_2, \tau_3, \ldots, \tau_r)$ exists and converges at any point. Then we find the leading correction orders in the corresponding expansions which can help extrapolate finite-bin-size results towards zero bin size in experiments, or help characterize negligible bin size within an error threshold.

Luckily, in an open system with many diffusing molecules, the *cumulants* of integrated intensities converge to the *moments* of integrated intensities for a single molecule [17, 63]. Therefore $\mu'_{m,n}[W(0), W(t)]$ calculated for a single molecule is all we need for such systems. Furthermore, if the characteristic reaction time is small compared to the characteristic diffusion time, then the correction terms due to diffusion (which depend on the MDF) may also be neglected. Experimental and simulation results in Chapter 2 support this claim. This yields further motivation to start

with $\mu'_{m,n}[W(0), W(t)]$ of the stationary single-molecule system. In this section we show that a polynomial expansion is valid and the leading terms are linear or quadratic.

3.5.1 Slow-changing intensity

When fluctuations of $I(t)$ are slow compared to the bin size and derivatives of $I(t)$ exist such that its Taylor expansion around any point is convergent, we have

$$I(t) = I(0) + I'(0)t + \frac{1}{2!}I''(0)t^2 + \mathcal{O}(t^3)$$

in which I' and I'' indicate the first and second derivatives of I. Using the stationary property of $I(t)$ we can replace the integration from 0 to T with $-T/2$ to $T/2$ and write

$$\begin{aligned} W(0) &= \int_{-T/2}^{T/2} I(t)\mathrm{d}t \\ &= I(0)T + \frac{1}{24}I''(0)T^3 + \mathcal{O}(T^5) \end{aligned} \tag{3.24}$$

In what follows, we focus on the first few $G'_{\vec{1}_{m+n}}(\tau_2, \tau_3, \ldots, \tau_{m+n})$ written explicitly as

$$G'_{1,1}(\tau_2) = \langle I(0)I(\tau_2)\rangle$$

$$G'_{1,1,1}(\tau_2, \tau_3) = \langle I(0)I(\tau_2)I(\tau_2+\tau_3)\rangle$$

$$G'_{1,1,1,1}(\tau_2, \tau_3, \tau_4) = \langle I(0)I(\tau_2)I(\tau_2+\tau_3)I(\tau_2+\tau_3+\tau_4)\rangle$$

We will also need the derivatives of the above multi-point correlation functions. We define the notation

$$\mathrm{D}_i = \frac{\partial}{\partial\tau_i}$$
$$\mathrm{D}_i^r = \frac{\partial^r}{\partial\tau_i^r}$$

In particular,

$$\mathrm{D}_2^2 G'_{1,1}(\tau_2) = \langle I(0) I''(\tau_2) \rangle$$

$$\mathrm{D}_3^2 G'_{1,1,1}(\tau_2, \tau_3) = \langle I(0) I(\tau_2) I''(\tau_2 + \tau_3) \rangle$$

$$\mathrm{D}_4^2 G'_{1,1,1,1}(\tau_2, \tau_3, \tau_4) = \langle I(0) I(\tau_2) I(\tau_2 + \tau_3) I''(\tau_2 + \tau_3 + \tau_4) \rangle$$

Using the above definitions and the expansion of $W(t)$, Equation (3.24), around 0 and t we get

$$\begin{aligned} \langle W(0) W(t) \rangle &= \langle I(0) I(t) \rangle T^2 + \frac{1}{12} \langle I(0) I''(t) \rangle T^4 + \mathcal{O}(T^6) \\ &= G'_{1,1}(t) T^2 + \frac{1}{12} \mathrm{D}_2^2 G'_{1,1}(t) T^4 + \mathcal{O}(T^6) \end{aligned} \tag{3.25}$$

where we have assumed reversibility property for the observed process:

$$\langle I(0) I''(t) \rangle = \langle I(0) I''(-t) \rangle = \langle I(t) I''(0) \rangle$$

A similar analysis yields

$$\begin{aligned} \left\langle W(0) W^2(t) \right\rangle &= \left\langle I(0) I^2(t) \right\rangle T^3 + \frac{1}{12} \langle I(0) I(t) I''(t) \rangle T^5 \\ &\quad + \frac{1}{24} \left\langle I^2(0) I''(t) \right\rangle T^5 + \mathcal{O}(T^7) \\ &= G'_{1,2}(t) T^3 + \frac{1}{12} \mathrm{D}_3^2 G'_{1,1,1}(t, 0) T^5 \\ &\quad + \frac{1}{24} \mathrm{D}_3^2 G'_{1,1,1}(0, t) T^5 + \mathcal{O}(T^7) \end{aligned}$$

in which we have used the reversibility property again: $\langle I''(0) I^2(t) \rangle = \langle I^2(0) I''(t) \rangle$.

Similarly,

$$\begin{aligned}\left\langle W^2(0)W^2(t)\right\rangle &= \left\langle I^2(0)I^2(t)\right\rangle T^4 + \frac{1}{6}\left\langle I^2(0)I(t)I''(t)\right\rangle T^6 + \mathcal{O}(T^8) \\ &= G'_{2,2}(t)T^4 + \frac{1}{6}\mathrm{D}_4^2 G'_{1,1,1,1}(0,t,0)T^6 + \mathcal{O}(T^8)\end{aligned}$$

where we have used the reversibility property again: $\langle I(0)I''(0)I^2(t)\rangle = \langle I^2(0)I(t)I''(t)\rangle$.

It is straightforward to see that the results for $\mu'_{m,n}\left[W(0),W(t)\right]$ are all of the form

$$\langle W^m(0)W^n(t)\rangle = T^{m+n}\left[G'_{m,n}(t) + \mathcal{O}(T^2)\right]$$

The resulting cumulants and normalized correlation functions will also be void of any linear term in T and the leading correction term will be quadratic. As a result, if the nth derivative of $I(t)$ is of order $1/t_{\mathrm{F}}^n$ where t_{F} is the characteristic time scale of intensity fluctuations, then it is sufficient to have $(T/t_{\mathrm{F}})^2 \ll 1$ for bin size effects to be very small.

3.5.2 Fast-changing intensity

When the intensity changes rapidly, for example in a step-wise fashion, the derivatives of $I(t)$ do not exist and the Taylor expansion of $I(t)$ is not valid. As a simplified example of such cases, consider the non-diffusing multi-state single particle studied in Section 3.4. Using the notation and results of that section we have (Equation (3.16)):

$$\begin{aligned}G'_{m,n}(t) &= \mu'_{m,n}[I(0),I(t)] \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J} P(i)Q_i^m Q_j^n Z_{j,i}(t) \\ &= \vec{Q}\mathbf{Q}^{n-1}\mathbf{Z}(t)\mathbf{Q}^m\vec{P}\end{aligned} \tag{3.26}$$

where $\vec{Q}$ is a $1 \times J$ row vector with elements Q_i , $\mathbf{Q}$ is a $J \times J$ diagonal matrix with elements Q_i on the main diagonal, $\mathbf{Z}(t)$ is a $J \times J$ matrix with elements $Z_{i,j}(t)$, and $\vec{P}$ is a $J \times 1$ column vector with elements $P(i)$.

We seek to find the multi-point correlation functions $G'_{\vec{1}_{m+n}}$. The first of such functions is immediately given by Equation (3.26):

$$G'_{1,1}(\tau_2) = \vec{Q}\mathbf{Z}(\tau_2)\mathbf{Q}\vec{P} \tag{3.27}$$

For the next higher order we have

$$\begin{aligned} G'_{1,1,1}(\tau_2, \tau_3) &= \langle I(0)I(\tau_2)I(\tau_2+\tau_3)\rangle \\ &= \left\langle \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{k=1}^{J} Q_iQ_jQ_k\delta[i, s(0)]\delta[j, s(\tau_2)]\delta[k, s(\tau_2+\tau_3)] \right\rangle \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{k=1}^{J} Q_iQ_jQ_k \left\langle \delta[i, s(0)]\delta[j, s(\tau_2)]\delta[k, s(\tau_2+\tau_3)]\right\rangle \end{aligned}$$

Thus we need to calculate the following expected value:

$$\begin{aligned} \langle \delta[i, s(0)]\delta[j, s(\tau_2)]\delta[k, s(\tau_2+\tau_3)]\rangle &= P(i, j, k; \tau_2, \tau_3) \\ &= P(k|i, j; \tau_2, \tau_3)P(i, j; \tau_2) \\ &= P(k|j; \tau_3)P(j|i; \tau_2)P(i) \end{aligned}$$

where we have used the independence property of each transition from its previous step. We have also defined

$$P(j|i; t) = Z_{j,i}(t)$$

Thus we have

$$G'_{1,1,1}(\tau_2,\tau_3) = \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{k=1}^{J} P(i)Q_iQ_jQ_kZ_{j,i}(\tau_2)Z_{k,j}(\tau_3)$$
$$= \vec{Q}\mathbf{Z}(\tau_3)\mathbf{Q}\mathbf{Z}(\tau_2)\mathbf{Q}\vec{P} \tag{3.28}$$

Similarly,

$$G'_{1,1,1,1}(\tau_2,\tau_3,\tau_4) = \langle I(0)I(\tau_2)I(\tau_2+\tau_3)I(\tau_2+\tau_3+\tau_4)\rangle$$

yields

$$G'_{1,1,1,1}(\tau_2,\tau_3,\tau_4) = \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{k=1}^{J}\sum_{l=1}^{J} P(i)Q_iQ_jQ_kQ_lZ_{j,i}(\tau_2)Z_{k,j}(\tau_3)Z_{l,k}(\tau_4)$$
$$= \vec{Q}\mathbf{Z}(\tau_4)\mathbf{Q}\mathbf{Z}(\tau_3)\mathbf{Q}\mathbf{Z}(\tau_2)\mathbf{Q}\vec{P} \tag{3.29}$$

and so forth.

Two-state transitions

For two-state reversible transitions, the $Z_{i,j}$ functions are given by

$$\mathbf{Z}(t) = \mathbf{A} + \mathbf{C}e^{-t/t_{\mathrm{R}}} \tag{3.30}$$

where

$$\mathbf{A} = \frac{1}{1+k}\begin{bmatrix} 1 & 1 \\ k & k \end{bmatrix}$$
$$\mathbf{C} = \frac{1}{1+k}\begin{bmatrix} k & -1 \\ -k & 1 \end{bmatrix}$$

Then using (3.27) through (3.29) we have

$$
\begin{aligned}
G'_{1,1}(\tau_2) &= \vec{Q}\mathbf{Z}(\tau_2)\mathbf{Q}\vec{P} \\
&= \vec{Q}\mathbf{A}\mathbf{Q}\vec{P} + \vec{Q}\mathbf{C}\mathbf{Q}\vec{P}e^{-\tau_2/t_{\mathrm{R}}}
\end{aligned}
$$

$$
\begin{aligned}
G'_{1,1,1}(\tau_2,\tau_3) &= \vec{Q}\mathbf{Z}(\tau_3)\mathbf{Q}\mathbf{Z}(\tau_2)\mathbf{Q}\vec{P} \\
&= \vec{Q}(\mathbf{A}\mathbf{Q})^2\vec{P} + \vec{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\vec{P}e^{-\tau_2/t_{\mathrm{R}}} \\
&\quad + \vec{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\vec{P}e^{-\tau_3/t_{\mathrm{R}}} + \vec{Q}(\mathbf{C}\mathbf{Q})^2\vec{P}e^{-(\tau_2+\tau_3)/t_{\mathrm{R}}}
\end{aligned}
$$

$$
\begin{aligned}
G'_{1,1,1,1}(\tau_2,\tau_3,\tau_4) &= \vec{Q}\mathbf{Z}(\tau_4)\mathbf{Q}\mathbf{Z}(\tau_3)\mathbf{Q}\mathbf{Z}(\tau_2)\mathbf{Q}\vec{P} \\
&= \vec{Q}(\mathbf{A}\mathbf{Q})^3\vec{P} + \vec{Q}(\mathbf{A}\mathbf{Q})^2\mathbf{C}\mathbf{Q}\vec{P}e^{-\tau_2/t_{\mathrm{R}}} \\
&\quad + \vec{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\vec{P}e^{-\tau_3/t_{\mathrm{R}}} + \vec{Q}\mathbf{C}\mathbf{Q}(\mathbf{A}\mathbf{Q})^2\vec{P}e^{-\tau_4/t_{\mathrm{R}}} \\
&\quad + \vec{Q}\mathbf{A}\mathbf{Q}(\mathbf{C}\mathbf{Q})^2\vec{P}e^{-(\tau_2+\tau_3)/t_{\mathrm{R}}} + \vec{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\vec{P}e^{-(\tau_2+\tau_4)/t_{\mathrm{R}}} \\
&\quad + \vec{Q}(\mathbf{C}\mathbf{Q})^2\mathbf{A}\mathbf{Q}\vec{P}e^{-(\tau_3+\tau_4)/t_{\mathrm{R}}} + \vec{Q}(\mathbf{C}\mathbf{Q})^3\vec{P}e^{-(\tau_2+\tau_3+\tau_4)/t_{\mathrm{R}}}
\end{aligned}
$$

Now we calculate the leading correction terms for integrals of the firs few $G'_{\vec{1}_r}(\tau_2,\tau_3,\ldots,\tau_r)$. Assuming non-overlapping bins ($t > T$) we have

$$
\begin{aligned}
\langle W(0)W(t)\rangle &= \int_0^T \mathrm{d}t_1 \int_t^{t+T} \mathrm{d}t_2 G'_{1,1}(t_2-t_1) \\
&= T^2\vec{Q}\mathbf{A}\mathbf{Q}\vec{P} + \vec{Q}\mathbf{C}\mathbf{Q}\vec{P}\int_0^T \mathrm{d}t_1 \int_t^{t+T} \mathrm{d}t_2 e^{-(t_2-t_1)/t_{\mathrm{R}}}
\end{aligned} \tag{3.31}
$$

The integral can be exactly evaluated to be

$$
\int_0^T \mathrm{d}t_1 \int_t^{t+T} \mathrm{d}t_2 e^{-(t_2-t_1)/t_{\mathrm{R}}} = 4t_{\mathrm{R}}^2 e^{-t/t_{\mathrm{R}}} \sinh^2(\frac{T}{2t_{\mathrm{R}}})
$$

and expanded as

$$= T^2 e^{-t/t_{\mathrm{R}}} \left[1 + \frac{1}{12} \frac{T^2}{t_{\mathrm{R}}^2} + \mathcal{O}(\frac{T^4}{t_{\mathrm{R}}^4}) \right]$$

Substituting into Equation (3.31) we obtain

$$\langle W(0)W(t) \rangle = T^2 \left[\vec{Q}\mathbf{Z}(t)\mathbf{Q}\vec{P} + \frac{1}{12} \frac{T^2}{t_{\mathrm{R}}^2} \vec{Q}\mathbf{C}\mathbf{Q}\vec{P} e^{-t/t_{\mathrm{R}}} + \mathcal{O}(\frac{T^4}{t_{\mathrm{R}}^4}) \right]$$

where we have used Equation (3.30). We see that the leading correction term for $\langle W(0)W(t) \rangle$ is quadratic.

In a similar way, we see that

$$\begin{aligned} \left\langle W^2(0) \right\rangle &= 2 \int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 G'_{1,1}(t_2 - t_1) \\ &= T^2 \vec{Q}\mathbf{A}\mathbf{Q}\vec{P} + 2\vec{Q}\mathbf{C}\mathbf{Q}\vec{P} \int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 e^{-(t_2 - t_1)/t_{\mathrm{R}}} \\ &= T^2 \left[\vec{Q}\mathbf{Q}\vec{P} - \frac{1}{3} \frac{T^2}{t_{\mathrm{R}}^2} \vec{Q}\mathbf{C}\mathbf{Q}\vec{P} + \mathcal{O}(\frac{T^4}{t_{\mathrm{R}}^4}) \right] \end{aligned}$$

where we have used $\mathbf{Z}(0) = \mathbf{I}$ (identity matrix). We see that the leading correction term for $\langle W^2(0) \rangle$ (which is used for normalization of higher-order correlations) is quadratic.

Using a similar analysis we have

$$\begin{aligned} \left\langle W^2(0)W(t) \right\rangle &= 2 \int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \int_t^{t+T} \mathrm{d}t_3 G'_{1,1,1}(t_2 - t_1, t_3 - t_2) \\ &= T^3 \vec{Q}(\mathbf{A}\mathbf{Q})^2 \vec{P} + 2\vec{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\vec{P} \int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \int_t^{t+T} \mathrm{d}t_3 e^{-(t_2 - t_1)/t_{\mathrm{R}}} \\ &\quad + 2\vec{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\vec{P} \int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \int_t^{t+T} \mathrm{d}t_3 e^{-(t_3 - t_2)/t_{\mathrm{R}}} \\ &\quad + 2\vec{Q}(\mathbf{C}\mathbf{Q})^2 \vec{P} \int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \int_t^{t+T} \mathrm{d}t_3 e^{-(t_2 - t_1)/t_{\mathrm{R}}} e^{-(t_3 - t_2)/t_{\mathrm{R}}} \end{aligned} \tag{3.32}$$

which, after evaluating and expanding the integrals, yields:

$$\left\langle W^2(0)W(t)\right\rangle = T^3\left[\vec{Q}\mathbf{Z}\mathbf{Q}^2\vec{P} - \frac{1}{3}\frac{T}{t_{\rm R}}\vec{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\vec{P}\right.$$
$$\left. + \frac{1}{6}\frac{T}{t_{\rm R}}\vec{Q}\mathbf{C}\mathbf{Q}(\mathbf{A}-\mathbf{C})\mathbf{Q}\vec{P}e^{-t/t_{\rm R}} + \mathcal{O}(\frac{T^2}{t_{\rm R}^2})\right]$$

Therefore the leading correction term is linear for $\langle W^2(0)W(t)\rangle$.

We use a similar analysis for order $(2,2)$:

$$\begin{aligned}
\left\langle W^2(0)W^2(t)\right\rangle &= 4\int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \int_t^{t+T} \mathrm{d}t_3 \int_{t_3}^{t+T} \mathrm{d}t_4 G'_{1,1,1,1}(t_2-t_1, t_3-t_2, t_4-t_3) \\
&= 4\int_0^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \int_t^{t+T} \mathrm{d}t_3 \int_{t_3}^{t+T} \mathrm{d}t_4 \left[\vec{Q}(\mathbf{A}\mathbf{Q})^3\vec{P} + \vec{Q}(\mathbf{A}\mathbf{Q})^2\mathbf{C}\mathbf{Q}\vec{P}e^{-(t_2-t_1)/t_{\rm R}}\right. \\
&\quad + \vec{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\vec{P}e^{-(t_3-t_2)/t_{\rm R}} + \vec{Q}\mathbf{C}\mathbf{Q}(\mathbf{A}\mathbf{Q})^2\vec{P}e^{-(t_4-t_3)/t_{\rm R}} \\
&\quad + \vec{Q}\mathbf{A}\mathbf{Q}(\mathbf{C}\mathbf{Q})^2\vec{P}e^{-(t_3-t_1)/t_{\rm R}} + \vec{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\vec{P}e^{-(t_2-t_1)/t_{\rm R}}e^{-(t_4-t_3)/t_{\rm R}} \\
&\quad \left. + \vec{Q}(\mathbf{C}\mathbf{Q})^2\mathbf{A}\mathbf{Q}\vec{P}e^{-(t_4-t_2)/t_{\rm R}} + \vec{Q}(\mathbf{C}\mathbf{Q})^3\vec{P}e^{-(t_4-t_1)/t_{\rm R}}\right] \qquad (3.33)
\end{aligned}$$

Evaluating and expanding the integrals in Equation (3.33) we see they have linear or quadratic deviations from zero-bin-size values. Keeping the linear terms we get:

$$\begin{aligned}
\left\langle W^2(0)W^2(t)\right\rangle = T^4\left[\vec{Q}Q\mathbf{Z}\mathbf{Q}^2\vec{P} - \frac{1}{3}\frac{T}{t_{\rm R}}\vec{Q}(\mathbf{A}\mathbf{Q})^2\mathbf{C}\mathbf{Q}\vec{P} - \frac{1}{3}\frac{T}{t_{\rm R}}\vec{Q}\mathbf{C}\mathbf{Q}(\mathbf{A}\mathbf{Q})^2\vec{P}\right. \\
- \frac{2}{3}\frac{T}{t_{\rm R}}\vec{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\vec{P} + \frac{1}{3}\frac{T}{t_{\rm R}}\vec{Q}\mathbf{A}\mathbf{Q}\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{Q}\vec{P}e^{-t/t_{\rm R}} \\
\left. -\frac{1}{3}\frac{T}{t_{\rm R}}\vec{Q}(\mathbf{C}\mathbf{Q})^3\vec{P}e^{-t/t_{\rm R}} + \mathcal{O}(\frac{T^2}{t_{\rm R}^2})\right]
\end{aligned}$$

which shows the leading correction term is linear for $\langle W^2(0)W^2(t)\rangle$.

Finally, we point out that the average of $W(t)$, which is used for normalization of correlation functions, is always

$$\begin{aligned}\langle W(t)\rangle &= \frac{1}{N_{\text{bins}}}\sum_{T}\int_{t}^{t+T} I(t')\mathrm{d}t' \\ &= \frac{T}{t_{\text{expt}}}\int_{0}^{t_{\text{expt}}} I(t')\mathrm{d}t' \\ &= T\,\langle I(t)\rangle\end{aligned}$$

hence does not require finite-bin-size correction. Here, $\sum_T$ indicates summation over all N_{bins} number of bins in the entire experimental time t_{expt}.

3.6 How sub-binning works

Statistical independence of artifact photons in different bins is the principle that allows us to remove artifacts by cross-correlating two detectors in conventional, second-order FCS. Before removal, each type of artifact in a conventional order-(1,1) correlation function is limited to a short time scale referred to as the "native" time scale of that artifact. For example, the native time scales of dead-time and after-pulsing are $\sim 50\,\text{ns}$ and $\sim 0.1\,\mu\text{s}$ in our detectors, respectively, as seen in the autocorrelation functions shown in Figure 2.2. Using two detectors, artifacts are removed in second-order correlation functions at lag times shorter than the native time scale of the artifacts, with no sub-binning needed. However, for higher-order correlations, more than two detectors would be required to correct artifacts at lag times below their native time scales. Shot noise would significantly increase with more detectors in higher-order FCS, limiting the accessible time scale in practice. Also for higher-order correlations, detector cross-talk effects need to be handled properly. If cross-talk photons are not removed by spectral filtering or some other method during data collection, sub-binning can be used to avoid cross-talk artifacts when computing the higher-order correlation functions using multiple detectors, down to the native cross-talk time scale.

If we try to compute higher-order correlation functions using a single detector, detector artifacts will extend to time scales much longer than their native time scales (explained below). Sub-binning can remove these extended artifacts in higher-order correlation functions obtained from a single detector. This approach again works based on the principle that the artifact photons in one sub-bin

are statistically independent from another sub-bin. For this condition to hold in a single-detector experiment, the sub-bins and the lag time must be significantly larger than the native time scale of the artifacts. The smallest bin size we usually use in this approach is $1\,\mu$s, which contains two sub-bins of size $\sim 0.5\,\mu$s. The non-zero size of the bins and sub-bins causes systematic deviations from ideal correlation functions. However, these deviations are small when the lag time is significantly larger than the bin size and/or the variations of the correlation function are much slower than the bin size. Under such conditions, a whole bin and the sub-bins within it represent a single time point. This can be represented mathematically as

$$\left\langle I^2(0)I^2(t)\right\rangle = \lim_{t'\to 0}\lim_{t''\to t}\left\langle I(0)I(t')I(t'')I(t)\right\rangle$$

for a fourth-order correlation function as an example. In this case, $I(0)$ and $I(t')$ denote the signal intensity at the two sub-bins at lag time 0, and $I(t'')$ and $I(t)$ denote the signal intensity at the sub-bins at lag time t.

In higher-order correlation functions, the detector artifacts extend to time scales far beyond their native time scale, and sub-binning is intended only to remove such effects at longer time scales. To understand how the artifacts affect longer time scales and how sub-binning (in one or multiple detectors), or independent bins (in multiple detectors), remove these affects, we consider a simple mathematical model. For simplicity, we assume after-pulsing is the only artifact that exists.

First, we consider a single-detector experiment with no sub-binning. Take n to denote the number of (true) photons in a given bin, and k to denote the number of (false) after-pulses. For each detected photon, an after-pulse may be recorded with the small probability p, inside the same bin (remember that the bin size is significantly larger than the native after-pulsing time scale). Therefore, for a given n, the probability of recording k after-pulses is given by the binomial distribution:

$$P(k|n) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \tag{3.34}$$

As a simple case, suppose we are interested in computing the quantity $\langle I^2 \rangle$. This quantity is relevant in computing higher-order correlations since higher powers of intensity appear in higher-order correlation functions. Assuming a small bin size, T, the expected number of photons in a bin is

$$W \approx TI$$

More precisely, W, also known as the integrated intensity, is the integral of $I(t)$ over the bin interval (Equation (3.5)). To approximate $\langle I^2 \rangle$, we compute the quantity $\langle W^2 \rangle / T^2$ in experiments. In a single-detector experiment, the moments of W are equal to the factorial moments of n, i.e.

$$\langle W \rangle = \langle n \rangle \tag{3.35}$$

$$\langle W^2 \rangle = \langle n(n-1) \rangle \tag{3.36}$$

and so forth. These follow from Mandel's formula [62, 63]:

$$\begin{aligned} P(n) &= \int \mathrm{Poi}(n; W) P(W) \mathrm{d}W \\ &= \int \frac{W^n}{n!} e^{-W} P(W) \mathrm{d}W \end{aligned}$$

Th total count of (true and false) pulses in a bin is $n' = n + k$, thus what we compute for $\langle W^2 \rangle$ using Equation (3.36) is

$$\langle n'(n'-1) \rangle = \langle n(n-1) \rangle + 2 \langle kn \rangle + \langle k(k-1) \rangle \tag{3.37}$$

The last two terms in Equation (3.37) are caused by the artifact, therefore we estimate their sizes. For the middle term we have:

$$\begin{aligned}\langle kn\rangle &= \sum_{n=0}^{\infty} nP(n) \sum_{k=0}^{n} kP(k|n) \\ &= p\sum_{n=0}^{\infty} n^2 P(n) \\ &= p\left\langle n^2\right\rangle \\ &= p\left(\left\langle W^2\right\rangle + \langle W\rangle\right) \end{aligned} \tag{3.38}$$

where we have used Equations (3.34), (3.35), and (3.36) . For the last term in Equation (3.37) we have:

$$\begin{aligned}\langle k(k-1)\rangle &= \sum_{n=0}^{\infty} P(n) \sum_{k=0}^{n} k(k-1)P(k|n) \\ &= p^2\sum_{n=0}^{\infty} n(n-1)P(n) \\ &= p^2\left\langle n(n-1)\right\rangle \\ &= p^2\left\langle W^2\right\rangle \end{aligned} \tag{3.39}$$

Substituting Equations (3.36), (3.38), and (3.39) into (3.37) we obtain:

$$\left\langle n'(n'-1)\right\rangle = \left\langle W^2\right\rangle (1+p)^2 + 2p\left\langle W\right\rangle \tag{3.40}$$

Next, we consider two independent "channels", named A and B, that share the same $I(t)$ at any moment. These "channels" can be two overlapping bins at two different detectors without sub-binning, or two adjacent, non-overlapping sub-bins in a single-detector experiment with sub-binning. Either way, we assume that the same W is shared by the two bins, or sub-bins, at any moment, but that they have independent photon counts, n_{A} and n_{B}, as well as independent artifact counts, k_{A} and k_{B}, respectively. With two independent channels, one can show that [63]

$$\begin{aligned}\left\langle W^2 \right\rangle &= \langle n_{\mathrm{A}} n_{\mathrm{B}} \rangle \\ &= \int P(W) \mathrm{d}W \sum_{n_{\mathrm{A}}} n_{\mathrm{A}} \mathrm{Poi}(n_{\mathrm{A}}; W) \sum_{n_{\mathrm{B}}} n_{\mathrm{B}} \mathrm{Poi}(n_{\mathrm{B}}; W)\end{aligned} \tag{3.41}$$

which follows from the multivariate form of Mandel's formula [63]:

$$P(n_{\mathrm{A}}, n_{\mathrm{B}}) = \int \mathrm{Poi}(n_{\mathrm{A}}; W) \mathrm{Poi}(n_{\mathrm{B}}; W) P(W) \mathrm{d}W$$

With the artifact present, what we compute is

$$\begin{aligned}\langle n'_{\mathrm{A}} n'_{\mathrm{B}} \rangle &= \langle (n_{\mathrm{A}} + k_{\mathrm{A}})(n_{\mathrm{B}} + k_{\mathrm{B}}) \rangle \\ &= \langle n_{\mathrm{A}} n_{\mathrm{B}} \rangle + \langle k_{\mathrm{A}} n_{\mathrm{B}} \rangle + \langle k_{\mathrm{B}} n_{\mathrm{A}} \rangle + \langle k_{\mathrm{A}} k_{\mathrm{B}} \rangle\end{aligned} \tag{3.42}$$

The last three terms are related to the artifact. For the second term we have

$$\begin{aligned}\langle k_{\mathrm{A}} n_{\mathrm{B}} \rangle &= \sum_{k_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} k_{\mathrm{A}} n_{\mathrm{B}} P(k_{\mathrm{A}}, n_{\mathrm{B}}) \\ &= \sum_{n_{\mathrm{A}}} \sum_{k_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} k_{\mathrm{A}} n_{\mathrm{B}} P(k_{\mathrm{A}}, n_{\mathrm{B}} | n_{\mathrm{A}}) P(n_{\mathrm{A}})\end{aligned}$$

Given a particular n_{A}, the probabilities of k_{A} and n_{B} are independent, i.e. $P(k_{\mathrm{A}}, n_{\mathrm{B}} | n_{\mathrm{A}}) = P(k_{\mathrm{A}} | n_{\mathrm{A}}) P(n_{\mathrm{B}} | n_{\mathrm{A}})$, thus,

$$
\begin{aligned}
\langle k_{\mathrm{A}} n_{\mathrm{B}} \rangle &= \sum_{n_{\mathrm{A}}} \sum_{k_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} k_{\mathrm{A}} n_{\mathrm{B}} P(k_{\mathrm{A}}|n_{\mathrm{A}}) P(n_{\mathrm{B}}|n_{\mathrm{A}}) P(n_{\mathrm{A}}) \\
&= \sum_{n_{\mathrm{A}}} \sum_{k_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} k_{\mathrm{A}} n_{\mathrm{B}} P(k_{\mathrm{A}}|n_{\mathrm{A}}) P(n_{\mathrm{A}}, n_{\mathrm{B}}) \\
&= \int \mathrm{d}W P(W) \sum_{n_{\mathrm{A}}} \mathrm{Poi}(n_{\mathrm{A}}; W) \sum_{k_{\mathrm{A}}} k_{\mathrm{A}} P(k_{\mathrm{A}}|n_{\mathrm{A}}) \sum_{n_{\mathrm{B}}} n_{\mathrm{B}} \mathrm{Poi}(n_{\mathrm{B}}; W) \\
&= p \int \mathrm{d}W P(W) W \sum_{n_{\mathrm{A}}} n_{\mathrm{A}} \mathrm{Poi}(n_{\mathrm{A}}; W) \\
&= p \int \mathrm{d}W P(W) W^2 \\
&= p \left\langle W^2 \right\rangle
\end{aligned} \tag{3.43}
$$

A similar expression can be found for $\langle k_{\mathrm{B}} n_{\mathrm{A}} \rangle$. For the last term in Equation (3.42) we have

$$
\begin{aligned}
\langle k_{\mathrm{A}} k_{\mathrm{B}} \rangle &= \sum_{n_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} \sum_{k_{\mathrm{A}}} \sum_{k_{\mathrm{B}}} k_{\mathrm{A}} k_{\mathrm{B}} P(k_{\mathrm{A}}, k_{\mathrm{B}}|n_{\mathrm{A}}, n_{\mathrm{B}}) P(n_{\mathrm{A}}, n_{\mathrm{B}}) \\
&= \sum_{n_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} \sum_{k_{\mathrm{A}}} \sum_{k_{\mathrm{B}}} k_{\mathrm{A}} k_{\mathrm{B}} P(k_{\mathrm{A}}|n_{\mathrm{A}}) P(k_{\mathrm{B}}|n_{\mathrm{B}}) P(n_{\mathrm{A}}, n_{\mathrm{B}}) \\
&= \sum_{n_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} P(n_{\mathrm{A}}, n_{\mathrm{B}}) \sum_{k_{\mathrm{A}}} k_{\mathrm{A}} P(k_{\mathrm{A}}|n_{\mathrm{A}}) \sum_{k_{\mathrm{B}}} k_{\mathrm{B}} P(k_{\mathrm{B}}|n_{\mathrm{B}}) \\
&= \sum_{n_{\mathrm{A}}} \sum_{n_{\mathrm{B}}} P(n_{\mathrm{A}}, n_{\mathrm{B}}) p n_{\mathrm{A}} p n_{\mathrm{B}} \\
&= p^2 \langle n_{\mathrm{A}} n_{\mathrm{B}} \rangle \\
&= p^2 \left\langle W^2 \right\rangle
\end{aligned} \tag{3.44}
$$

Substituting Equations (3.41), (3.43), and (3.44) into (3.42) we obtain:

$$
\langle n'_{\mathrm{A}} n'_{\mathrm{B}} \rangle = \left\langle W^2 \right\rangle (1 + p)^2 \tag{3.45}
$$

Now we can compare the result for single-channel (single-detector) experiment without sub-binning, Equation (3.40), with the result for two-channel (sub-binning or two-detector) experiment, Equation (3.45). The probability of after-pulsing, p, is usually very small, $\sim 1\%$. So in both cases

a small relative error of $\sim 2p$ is introduced by after-pulsing. However, in the single-channel experiment the extra term $2p\langle W\rangle$ exists. The relative error caused by this term is

$$\begin{aligned} r &= \frac{2p\langle W\rangle}{\langle W^2\rangle} \\ &= \frac{2p}{\langle W\rangle\left(1+g_{1,1}(0)\right)} \end{aligned}$$

where $g_{1,1}(0)$ is the amplitude of the second-order correlation function. This relative error r can be significant or small, depending on the ratio $2p/\langle W\rangle$. In typical FCS experiments, the average signal intensity can be $\langle I\rangle \approx 50\,\mathrm{kHz}$ and the bin size can be $T \approx 1\,\mu\mathrm{s}$. This means that the average photon count per bin is $\langle n\rangle = \langle W\rangle \approx 0.05$. This, using $p \approx 0.01$ and $g_{1,1}(0) \approx 1$, yields $r \approx 20\%$, showing that a significant error can be caused by artifacts of small probability in single-detector experiments without sub-binning. This large relative error is consistent with our simulations of either after-pulsing or dead-time (Figure 2.3). The error becomes smaller as the lag time, and consequently the bin size, increases. In real detectors, or in simulations with both dead-time and after-pulsing, the errors are somewhat smaller because dead-time and after-pulsing have canceling effects. For modeling dead-time, we could follow a similar approach to the above analysis, with the number of photons lost in the dead-times subtracted from the true counts. A major difference would be that the probability of losing j photons in the dead-times is $P(j|n) = \mathrm{Poi}(j; ndW/T)$, where d is the dead-time after each photon detection.

Chapter 4

Testing higher-order fluorescence correlation spectroscopy

4.1 Introduction

In this chapter, we first exploit the factorized reaction and diffusion parts of higher order correlations to develop a procedure that greatly simplifies the analysis of reaction kinetics by eliminating the dependence on the molecular detection function (MDF, defined as the combination of laser intensity distribution, collection point-spread function, and pinhole aperture [61]) and the diffusional properties of the system. Next, we carry out two experimental demonstrations of this procedure. In the first set of experiments, we resolve multi-component mixtures of non-reacting species, by treating such cases as very slow reactions. This constitutes a foundationally important special case because at short enough lag times any reaction behaves as a static mixture, with correlation amplitudes only depending on the average populations of the reacting species (states) in the probe region. The systems we study are mixtures of TAMRA-labeled and rhodamine 6G-labeled oligonucleotides at various concentration ratios. The results are then compared with the known concentration and brightness ratios of the species in each sample. In the second set of experiments, we study the fast protonation reaction of FITC, which occurs at time scales below $10\,\mu$s. The individual reaction rates are obtained and compared with independent measurements and previously reported values.

In each set of measurements, we employ our new data acquisition approach to calculate higher-order fluorescence correlations, from which higher-order "relative reaction" functions are extracted. In the case of mixtures, the time-resolved relative reaction functions clearly reveal the multi-component, non- or slow-reacting nature of the system. In the protonation reaction of FITC,

the relative reaction functions immediately show the two-state mechanism of the reaction. Obtaining the brightness and concentration ratios of the mixing or reacting species, as well as the time scale of the reaction, is then simply a matter of fitting to the relative reaction curves. The successful results of our experiments verify the applicability of the developed procedure to the analysis of mixtures as well as slow and fast reactions. The success of our improved data acquisition method is additionally verified in these applications. Along the way, we also verify the theoretical relations developed in the preceding chapters for signal-to-noise ratio by evaluating their performance in estimating weighted fit residuals.

Considering the dependence of the uncertainty of the results on the concentration and brightness ratios of the components of a mixture or reaction, it is important to evaluate the applicability of the technique to systems of different compositions. Using the analysis of mixtures as the simple yet fundamental case, we carry out a detailed numerical analysis of the uncertainty of the measured parameters over a wide range of brightness and concentration ratios. We clarify regions of the parameter space over which the technique can be reliably applied. The results guide us in the design of experiments in the study of fast reactions.

FCS has been a primary tool in the time-resolved analysis of non-reacting or reacting molecules in solution, based on fluorescence intensity fluctuations. However, conventional FCS alone is insufficient to fully determine the reaction or mixture parameters. Other relevant techniques that can complement FCS by further analysis of the same photon stream have primarily been applied to the case of non- or slow-reacting mixtures, or at time scales shorter than any reaction time scale to avoid such dynamical effects. These include non-correlated analysis of moments [49, 50] or cumulants [51, 52, 55, 56], and intensity distribution analysis or photon counting histogram [46–48, 53, 54]. Limited bin size and neglect of reaction dynamics results in increased uncertainty and inaccuracy of such techniques if applied to reactions. Efforts have been made to increase the bin size in the non-correlated techniques to include translational diffusion of the molecules [53–55], though no formulation of molecular reaction dynamics has been proposed. The moment-based formulation of higher-order correlation analysis [15, 16, 59] has not been extended

and applied to reaction dynamics either, due to the increased complexity that would ensue. All of these non-correlated techniques and the moment-based formulation of higher-order correlations rely on modeling, usually with approximations, and characterization of the MDF. This makes the analysis more difficult and dependent on the agreement of the experimental MDF with the models, and on the accuracy of the relevant calibration measurements. In comparison, we demonstrate in this chapter how the simple factorized form of cumulant-based higher-order correlation analysis can be exploited to remove the dependence on, and hence the need to characterize, the MDF. In addition to simpler application and setup independence, the reduced number of modeled and calibrated parameters greatly enhances the accuracy and the chance of success in a broader variety of systems compared to the alternatives in the analysis of static or slow-reacting mixtures. It also provides a unique capability for the analysis of fast reactions. With our new sub-binning approach, the requirement to model and calibrate detector artifacts is also lifted. We are therefore reporting a technique that can be largely automated as a portable (experimental-setup-independent) program, and broadly applied to the study of non-, slow-, or fast-reacting systems.

4.2 Methods

4.2.1 Mixture experiments

Labeled and HPLC purified oligonucleotides of sequence $5'$-$\mathrm{AACCC(T)_8GGGTT}$-$3'$ were purchased and diluted to $\sim 1\,\mathrm{nM}$ concentration in $0.25\,\mathrm{X}$ TE buffer PH 8.0. The sample referred to as "TAMRA-oligo" contained oligonucleotides labeled with $5'$ TAMRA (NHS Ester) and was purchased from Integrated DNA Technologies (Coralville, IA). The sample referred to as "R6G-oligo" contained oligonucleotides labeled with $5'$ 5(6)-CR 6G (Carboxyrhodamine 6G) and was purchased from AnaSpec (Fremont, CA). Mixtures of TAMRA-oligo and R6G-oligo samples were prepared at 20:80, 50:50, and 80:20 volume ratios (TAMRA to R6G) for mixture experiments.

The experimental setup and the calculation of higher order correlations are described in Chapter 2. As a reminder, the main setup components include a laser excitation wavelength of $532\,\mathrm{nm}$ at $19\,\mu\mathrm{W}$ power, a $100\times$/$1.30\,\mathrm{NA}$ oil objective in a commercial microscope, a dichroic beamsplit-

ter and emission filter with cut-offs at $540\,\mathrm{nm}$ and $550\,\mathrm{nm}$ respectively, a $50\,\mu\mathrm{m}$ pinhole, and two avalanche photodiode detectors. Data was collected for $15\,\mathrm{min}$ for each sample. For this study, correlation curves were calculated using the two-detector two-sub-bin method, with binning parameters $\delta_0 = 1\,\mu\mathrm{s}$ (starting bin size), $\rho = 0.08$ (bin/lag time ratio), and $\sigma = 0.01\,\mu\mathrm{s}$ (sub-bin gap). The resulting curves were analyzed in IGOR Pro 6.3.7.2 (WaveMetrics, Lake Oswego, OR) using weighted least-squares Global Fit package. Direct numerical solutions to mixture equations were calculated to absolute precision of 10^{-9} using a version of Powell's Hybrid method implemented in the GNU Scientific Library.

4.2.2 Reaction experiments

Phosphate buffers at three different pH values were prepared by mixing monobasic and dibasic sodium phosphate solutions at various ratios and diluted to total phosphate concentration of $1.0\,\mathrm{mM}$. The pH was measured using a freshly calibrated pH-meter (Hanna Instruments, Woonsocket, RI) to be 5.73, 5.94, and 6.05 (all ± 0.03) for the three buffers used in this report. Solutions of fluorescein-5-isothiocyanate (FITC) (Invitrogen, Carlsbad, CA) were prepared in these buffers at final concentrations of $7.6\,\mathrm{nM}$ FITC and $0.88\,\mathrm{mM}$ total phosphate concentration. A solution of $3.8\,\mathrm{nM}$ FITC and $8.8\,\mathrm{mM}$ dibasic sodium phosphate (pH 9.03 ± 0.03) was used as the reference sample.

For FITC protonation reaction experiments, the following changes were applied to the experimental setup used in mixture experiments: The beam from a $488\,\mathrm{nm}$ laser (Novalux, Inc., Sunnyvale, CA) was cleaned up using a $488/10$ bandpass filter, then attenuated, expanded and cleaned up with a $50\,\mu\mathrm{m}$ pinhole. The dichroic beamsplitter ($503\,\mathrm{nm}$ cut-off), the emission filter ($525/36$BP), and the emission pinhole ($100\,\mu\mathrm{m}$) were selected accordingly for reaction experiments. A drop of solution was placed on a normal coverslip and covered with an inverted glass vial containing a moist sponge to suppress instabilities due to air exposure during measurement. For two-detector reaction experiments with $62\,\mu\mathrm{W}$ excitation power (measured at objective back aperture), the rest of the setup elements and configuration were preserved as in mixture experiments. The data collec-

tion time and the analysis tools and procedure were also identical to those in mixture experiments. For single-detector reaction experiments with $25\,\mu\mathrm{W}$ excitation power, one of the detectors and their beamsplitter were removed from the setup. Correlation curves were calculated by using the recorded photon stream from the one detector as both inputs of the two-detector two-sub-bin correlator program. All other parameters and analysis steps were identical to two-detector experiments. Due to observed instability of FITC in low-pH buffers, experiments were repeated six times at each pH value to obtain the associated uncertainty.

4.3 Theory

Previously we demonstrated that the normalized higher-order correlations, defined using cumulants of integrated signal intensity or detector photon counts over a small sampling time (bin size), reduce to

$$g_{m,n}(t) = \gamma_{m,n} R_{m,n}(t) Y_{m,n}(t) \tag{4.1}$$

for a system of diffusing and reacting molecules, assuming the molecules have same diffusivity in all states and the reaction rates are independent of excitation intensity. The factors $\gamma_{m,n}$ and $Y_{m,n}(t)$ depend on the spatial illumination and detection profiles and their specific form will not matter here, for reasons that will soon become clear. The reaction factor is

$$R_{m,n}(t) = \frac{\sum_s \sum_{s'} N_s Q_s^m Q_{s'}^n Z_{s',s}(t)}{\left(\sum_s N_s Q_s^m\right)\left(\sum_s N_s Q_s^n\right)} \tag{4.2}$$

in which the summations over s and s' run from 1 to J, the total number of states (species). As a reminder, Q_s and N_s denote the peak brightness and the average number of molecules in state s in the probe region, respectively, and $Z_{s',s}(t)$ denotes the probability that a particle is found in state s' at time t given it was in state s at time 0.

To fully characterize a system of J reacting states/species, the sets of independent parameters $\{N_1, \ldots, N_J\}$ and $\{Q_1, \ldots, Q_J\}$ are to be determined, in addition to the reaction rate constants that appear in $Z_{s',s}(t)$. In practice, one of these parameters can always be expressed in terms of the

mean channel count rate,

$$\langle I \rangle = \sum_{s=1}^{J} Q_s N_s \tag{4.3}$$

and the remaining parameters. Without loss of generality, we take that parameter to be the first brightness level Q_1. Furthermore, we can define $N = \sum_{s=1}^{J} N_s$ as the total number of molecules in the probe region regardless of their state, $k_s = N_s/N_1$ to denote the relative concentration of state s, and $q_s = Q_s/Q_1$ to be the relative brightness of state s. We then have

$$R_{m,n}(t) = \left(\frac{\sum_s k_s}{N}\right) \frac{\sum_s \sum_{s'} k_s q_s^m q_{s'}^n Z_{s',s}(t)}{\left(\sum_s k_s q_s^m\right)\left(\sum_s k_s q_s^n\right)} \tag{4.4}$$

Without loss of generality, state 1 can be taken to be the brightest state, thus $0 \leq q_s \leq 1$ for all s. Obviously, $q_1 = 1$, $k_1 = 1$, and $N_1 = N/\sum_s k_s$. Therefore the new independent parameters are $\{N, k_2, \ldots, k_J\}$, $\{q_2, \ldots, q_J\}$, and the rate constants, which can be determined using the second- and higher-order correlations.

Here we take advantage of the factorized form of Equation (4.1). The factors $\gamma_{m,n}$ and $Y_{m,n}(t)$ depend only on the MDF and the diffusion constant. Therefore, if only the reaction parameters are of interest, the higher order correlations $g_{m,n}^{(\mathrm{ref})}(t)$ from a "reference" sample with identical diffusional properties can be used to extract the *relative reaction functions*:

$$\begin{aligned} R_{m,n}^{(\mathrm{rel})}(t) &= \frac{g_{m,n}(t)}{g_{m,n}^{(\mathrm{ref})}(t)} \\ &= \frac{R_{m,n}(t)}{R_{m,n}^{(\mathrm{ref})}(t)} \end{aligned} \tag{4.5}$$

This removes the dependence on the excitation beam profile, and there is no need to characterize and calibrate the MDF and the diffusion parameters. In addition to theoretical and experimental simplification of the technique compared to the alternatives, the dependence on fewer parameters results in a more accurate and successful fitting procedure. The reference sample, in the simplest case, consists of the same molecule of interest labeled or conditioned such as to remain in a single brightness state. In this case,

$$R_{m,n}^{(\mathrm{ref})}(t) = 1/N^{(\mathrm{ref})} \tag{4.6}$$

and we have

$$R_{m,n}^{(\mathrm{rel})}(t) = \left(\frac{\sum_s k_s}{N^{(\mathrm{rel})}}\right) \frac{\sum_s \sum_{s'} k_s q_s^m q_{s'}^n Z_{s',s}(t)}{\left(\sum_s k_s q_s^m\right)\left(\sum_s k_s q_s^n\right)} \tag{4.7}$$

where

$$\begin{aligned} N^{(\mathrm{rel})} &= \frac{N}{N^{(\mathrm{ref})}} \\ &= \frac{\langle C \rangle}{\langle C \rangle^{(\mathrm{ref})}} \end{aligned} \tag{4.8}$$

is the ratio of the concentration of the sample of interest (test sample) to that of the reference sample. $N^{(\mathrm{rel})}$ can be determined as an independent parameter in higher-order fluorescence correlation analysis, replacing N. Alternatively, it might be possible to measure $N^{(\mathrm{rel})}$ through independent techniques such as UV-Vis to reduce the number of higher order correlations required and the measurement uncertainty. The values of the absolute parameters N, $N^{(\mathrm{ref})}$, N_s and Q_s are dependent on the specifics of each experimental setup, and hence of no general interest. Therefore, the target parameters are usually only $\{N^{(\mathrm{rel})}, k_2, \ldots, k_J\}$, $\{q_2, \ldots, q_J\}$, and the rate constants. However, if need be, $N^{(\mathrm{ref})}$ can be determined for one reference sample, for example using ordinary FCS, then the absolute parameters N and N_s for any test sample can also be determined using that reference sample. Also using $\langle I \rangle$ all Q_s values can be determined, following Equation (4.3). The rate constants can in general be determined using the decay time of the correlation curves together with the equilibrium population ratios, k_s.

For a mixture of slow- or non-reacting fluorescent species, we have $Z_{s',s}(t) = \delta_{s',s}$ and Equation (4.7) becomes

$$R_{m,n}^{(\mathrm{rel})}(t) = \left(\frac{\sum_s k_s}{N^{(\mathrm{rel})}}\right) \frac{\sum_s k_s q_s^{m+n}}{\left(\sum_s k_s q_s^m\right)\left(\sum_{s'} k_s q_s^n\right)} \tag{4.9}$$

which is independent of lag time. In the absence of rate constants, there are $2J - 1$ parameters to determine for a mixture of J components: $\{N^{(\mathrm{rel})}, k_2, \ldots, k_J\}$ and $\{q_2, \ldots, q_J\}$. A sample of a

single component (preferably the brightest) can be used as the reference, to remove the diffusion and beam shape effects.

The assumption that diffusion is independent from reaction, which leads to the factorized form of Equation (4.1), is a good approximation in a broad range of FCS applications. For example, in Section 5.3 we directly verify this model for FITC protonation reaction. We significantly distort the laser beam profile and observe that the isolated relative reaction functions are robustly independent of the MDF, while the correlation functions themselves are affected by the beam geometry. Several other reaction kinetics probed by FCS, such as conformational dynamics of DNA hairpin molecules [37,38], have also shown little dependence of the diffusion properties on the reaction state, and that a single average diffusion constant is sufficient to describe the FCS results. This approximation becomes even better when the reaction occurs in a much faster time scale than the diffusion, such that the difference in the displacement of different species in the observation volume is negligible over the reaction time scale.

4.3.1 Two-component mixtures

For a two-component mixture, we have $N^{(\mathrm{rel})}$, $k := k_2 = N_2/N_1$, and $q := q_2 = Q_2/Q_1$ as the three independent parameters to determine. The relative reaction functions become:

$$R^{(\mathrm{rel})}_{m,n} = \frac{(1+k)(1+kq^{m+n})}{N^{(\mathrm{rel})}(1+kq^m)(1+kq^n)} \tag{4.10}$$

We will show that orders $(1,1)$, $(2,1)$, and $(2,2)$ are sufficient to determine the three parameters of interest. (Orders $(1,2)$ and $(2,1)$ are identical in the absence of fast irreversible processes.) $N^{(\mathrm{rel})}$ is of less practical interest here because we are free to select the reference concentration, therefore we simplify our analysis by focusing on q and k. We define:

$$\begin{aligned} a &= \frac{R^{(\mathrm{rel})}_{2,1}}{R^{(\mathrm{rel})}_{1,1}} = \frac{(1+kq^3)(1+kq)}{(1+kq^2)^2} \\ b &= \frac{R^{(\mathrm{rel})}_{2,2}}{R^{(\mathrm{rel})}_{1,1}} = \frac{(1+kq^4)(1+kq)^2}{(1+kq^2)^3} \end{aligned} \tag{4.11}$$

The relative uncertainties in these parameters are nearly identical to the relative uncertainties of the $R^{(\mathrm{rel})}_{m,n}$ in the numerators, since the uncertainty in order $(1,1)$ is much smaller than the uncertainty in any higher order. Because the two-dye mixture and the single-dye reference sample have different triplet blinking factors in their corresponding correlation curves, there will be some residual triplet blinking effects in the relative reaction functions. The definition of a and b parameters removes these residual effects almost completely, and facilitates our general analysis of uncertainties in the parameter space using fewer parameters.

In the case that $N^{(\mathrm{rel})}$ is reliably known through independent measurements, q and k can be determined using only $R^{(\mathrm{rel})}_{1,1}$ and $R^{(\mathrm{rel})}_{2,1}$ which have much smaller uncertainties than $R^{(\mathrm{rel})}_{2,2}$. We will see that this method has the additional advantage of being applicable when q is very small (< 0.1); however, the results are sensitive to errors in $N^{(\mathrm{rel})}$.

4.3.2 Two-state reactions

For a fast two-state reaction, the transition factors $Z_{s',s}(t)$ are derived in Section 5.5. Using those factors, the relative reaction functions become, following Equation (4.7),

$$R^{(\mathrm{rel})}_{m,n} = \frac{1}{N^{(\mathrm{rel})}}(1 + B_{m,n}e^{-t/t_{\mathrm{R}}}) \tag{4.12}$$

where

$$B_{m,n} = \frac{k(1-q^m)(1-q^n)}{(1+kq^m)(1+kq^n)} \tag{4.13}$$

The three independent parameters $N^{(\mathrm{rel})}$, k, and q are defined as in the case of two-component mixtures, and the fourth parameter, t_{R}, is the overall reaction relaxation time. A reference sample in this case may be a system either with suppressed reaction (for example by removing a binding ligand), or labeled such that the reaction does not affect the fluoresce signal (for example by removing the quencher tag).

At lag times significantly shorter than the reaction time, $t/t_{\mathrm{R}} \ll 1$, Equation (4.12) reduces to an alternative, but equivalent, form of Equation (4.10) describing mixtures. This illustrates the

importance of a thorough study of the case of mixtures. At lag times significantly longer than the reaction time, $t/t_{\mathrm{R}} \gg 1$, all the relative reaction curves decay to the same baseline value, $1/N^{(\mathrm{rel})}$. This, together with the single-exponential decay of the curves, provides a way to clearly differentiate two-state reactions from multi-state reactions.

We saw in the previous section that three correlation orders were required to fully resolve the parameters for two-component mixtures. In contrast, only two correlation orders, $(1,1)$ and $(2,1)$, are required for two-state reversible reactions to fully determine all the parameters of interest: $N^{(\mathrm{rel})}$, k, and q, and t_{R}. This is true because the two correlation amplitudes at short lag times, $t/t_{\mathrm{R}} \ll 1$, together with the single baseline at long lag times, $t/t_{\mathrm{R}} \gg 1$, provide three independent measurements to uniquely determine q, k, and $N^{(\mathrm{rel})}$. The relaxation time, t_{R}, can be measured by fitting to either of the two correlation curves. In the two-state reaction experiments presented in this chapter, we include the correlation curve of order $(2,2)$ in our global fit analysis for additional statistical precision and to verify that the models work for all higher-order correlation functions up to the fourth order. The inclusion of the fourth-order correlation function will be crucial in the study of reactions with more than two states.

4.4 Results and discussion

Experiments were conducted to test higher order correlations in resolving binary mixtures of TAMRA-labeled and R6G-labeled DNA at various concentration ratios. We also present estimated uncertainties as functions of various underlying and measured parameters to understand the parameter regions in which the technique can be successfully applied. The results also provide insight into the reliability of the technique in the case of fast reactions, and guide us in the design and analysis of reaction experiments. As an example of a two-state fast reaction, the protonation reaction of FITC occurring at time scales shorter than $10\,\mu$s was studied at different pH values.

4.4.1 Mixture experiments

Solutions of pure TAMRA-oligo and R6G-oligo were prepared as described in Methods section, then mixed at 20:80, 50:50, and 80:20 TAMRA-oligo to R6G-oligo volume ratios, corresponding to $k \approx 0.25$, 1, 4, respectively. We take $s = 1$ to denote the brighter species, R6G-oligo, thus $0 < q < 1$. Higher order correlation functions $g_{1,1}(t)$, $g_{2,1}(t)$ and $g_{2,2}(t)$ were calculated for each sample (shown in Figure 4.1). Then pure R6G-oligo was chosen as the reference sample and all mixture curves were normalized by their corresponding curve from the reference sample to obtain the relative reaction functions $R^{(\mathrm{rel})}_{1,1}(t)$, $R^{(\mathrm{rel})}_{2,1}(t)$ and $R^{(\mathrm{rel})}_{2,2}(t)$ for each mixture. The resulting curves are shown in Figure 4.2, left. The corresponding a and b functions were also calculated, according to definitions 4.11, and shown in Figure 4.2, right. (a' was also defined and calculated as $R^{(\mathrm{rel})}_{1,2}(t)/R^{(\mathrm{rel})}_{1,1}(t)$).

As predicted by theory, the curves are roughly constant. The curves $R^{(\mathrm{rel})}_{2,1}(t)$ and $R^{(\mathrm{rel})}_{1,2}(t)$ are nearly identical, which shows no significant photobleaching or other irreversible processes are present. We consider three-parameter and two-parameter fitting approaches. In the three-parameter approach, $N^{(\mathrm{rel})}$, q, and k are determined simultaneously by global fitting to the three relative reaction functions, $R^{(\mathrm{rel})}_{1,1}$, $R^{(\mathrm{rel})}_{2,1}$, and $R^{(\mathrm{rel})}_{2,2}$. Even though the excitation power was reduced to $\sim 19\,\mu\mathrm{W}$, small but noticeable triplet blinking effects are still present in the relative correlation curves. Since the two dyes have different triplet blinking characteristics, their blinking reaction factors partially cancel out upon normalization by the reference sample. (The cancellation would be complete in studies using a single dye, see the reaction experiments below.) To avoid these effects, only a limited number of data points, N_{p}, on each relative reaction curve ($N_{\mathrm{p}} = 42$, $1.1\mu\mathrm{s} \leq t \leq 4.5\mu\mathrm{s}$) were used for parameter estimations. In the two-parameter approach, q and k are simultaneously determined by global fitting to a and b. The blinking effects are more effectively canceled out when a and b are calculated, thus more data points were used on each of those curves (and per degree of freedom) ($N_{\mathrm{p}} = 64$, $1.1\mu\mathrm{s} \leq t \leq 9.5\mu\mathrm{s}$). Due to these advantages, the two-parameter approach is preferred and presented in Table 4.1. For comparison, the three-parameter fitting results are presented in Tables 5.1 and 5.4.

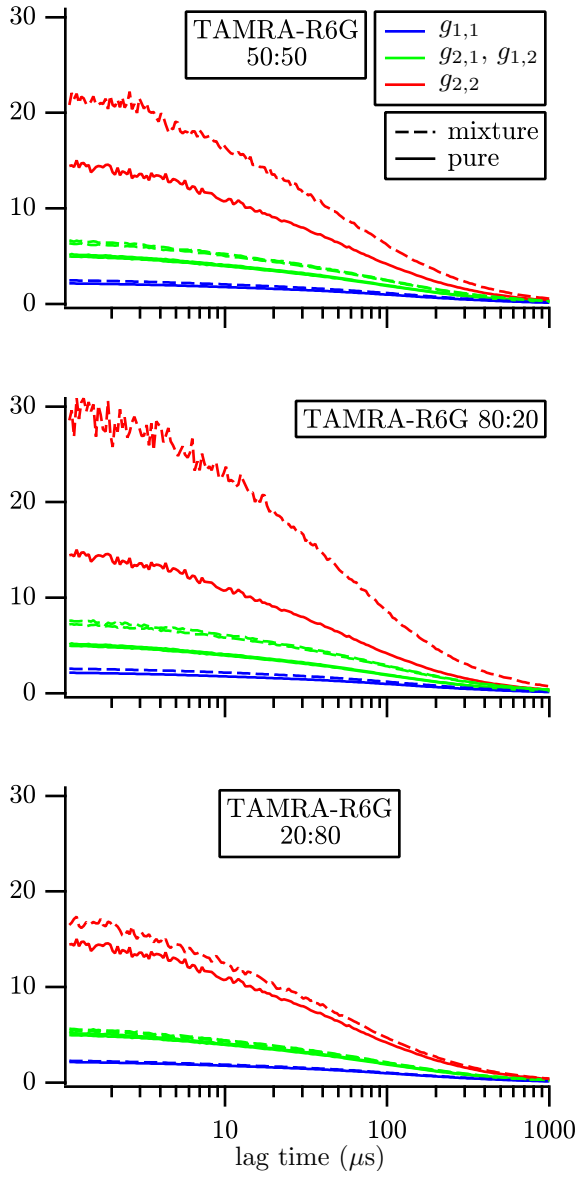

Figure 4.1: Second- and higher-order correlation functions obtained for mixtures of TAMRA-oligo and R6G-oligo at various concentration ratios (dashed lines). Pure R6G-oligo was used as reference in all cases (solid lines). The same color scheme and axis scale applies to all graphs.

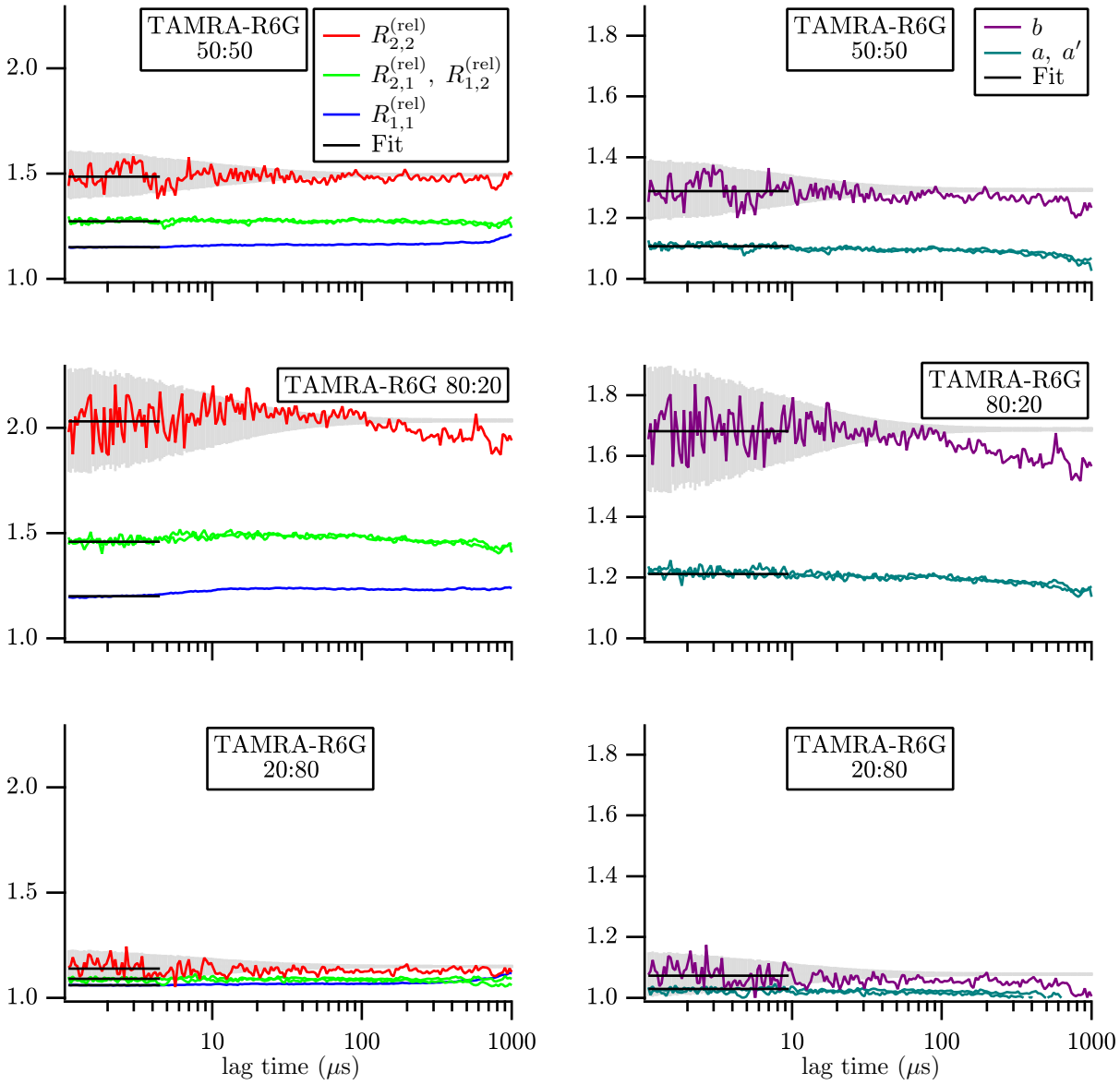

Figure 4.2: Relative reaction functions (left) and their corresponding a and b functions obtained for mixtures of TAMRA-oligo and R6G-oligo at various concentration ratios. Pure R6G-oligo was used as reference in all cases. The global fit lines to the first N_p data points on each curve are also shown. The theoretical uncertainties are visualized as light gray shade around the mean of the first N_p points, for curves of the highest order only.

Table 4.1: TAMRA- and R6G-oligo mixture components resolved by global fitting to a and b. Numbers in parentheses indicate fitting uncertainty in the corresponding last digits. Data standard deviations have been used as fitting weights.

	50:50		**80:20**		**20:80**	
	known	**measured**	**known**	**measured**	**known**	**measured**
q	$0.399(10)$	$0.413(26)$	$0.399(10)$	$0.402(4)$	$0.399(10)$	$0.57(11)$
k	$1.05(3)$	$1.05(9)$	$4.21(11)$	$4.07(35)$	$0.263(7)$	$0.34(17)$

Uncertainties were also estimated for each data point using the theoretical relations in Chapter 2, referred to here as theoretical data errors, and shown on graphs as gray shade around the hypothetical constant mean line of the first N_{p} data points, for highest orders only. Initially, fits were weighted with theoretical data errors. In the case of static mixtures, the curves are theoretically constant, therefore the standard deviation of data points around the mean value of each curve can also be used as fitting weights for that curve. In this way, curves of different orders have different uncertainties, but the lag-time dependence is ignored. The fits weighted with data standard deviations were found to yield more reasonable uncertainties, hence they are presented in the main text. The fits weighted with theoretical errors are presented and discussed in Section 5.1.1. In fits weighted with data standard deviations, the reduced chi-squared values, χ^2_{red}, were obtained equal to 1.00, as expected and confirming the fit lines actually converged to the mean values.

The "known" values were calculated and presented in the tables as follows: The ratio of the concentration of the pure TAMRA-oligo to that of pure R6G-oligo, $C_{\mathrm{T/R}}$, was obtained from the ratio of their order $(1,1)$ correlation curves (ordinary FCS) averaged over $N_{\mathrm{p}} = 42$ data points ($1.1\mu\mathrm{s} \leq t \leq 4.5\mu\mathrm{s}$). Then the $N^{(\mathrm{rel})}$ and k values for each mixture were calculated using $C_{\mathrm{T/R}}$ and the known mixing volumes. The brightness ratio, q, was also calculated using the the pure-solution count rates (average of the two detectors) and $C_{\mathrm{T/R}}$. The dominant uncertainties in these estimations were assumed to stem from pipetting error ($\sim 1.0\%$), the unequal splitting of the signal between detectors ($\sim 2.0\%$), and systematic variations in $C_{\mathrm{T/R}}$ caused by fast triplet blinking

Table 4.2: TAMRA- and R6G-oligo mixture components resolved by global fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$, and fixed $N^{(\mathrm{rel})}$. In all Tables, numbers in parentheses indicate uncertainty in the corresponding last digits. In the measured parameters, the first number in parentheses shows the variations in fitting results when $N^{(\mathrm{rel})}$ is varied by one standard deviation, and the second number shows the individual fit uncertainty with $N^{(\mathrm{rel})}$ fixed at the mean point. The "held" parameters were held fixed at the "known" mean value and at one standard deviation above and below to find the associated uncertainty.

	50:50		**80:20**		**20:80**	
	known	**measured**	**known**	**measured**	**known**	**measured**
q	$0.399(10)$	$0.404(47)(2)$	$0.399(10)$	$0.381(18)(1)$	$0.399(10)$	$0.360(142)(9)$
k	$1.05(3)$	$1.02(18)(2)$	$4.21(11)$	$3.07(59)(6)$	$0.263(7)$	$0.202(50)(8)$
$N^{(\mathrm{rel})}$	$1.03(3)$	held	$1.04(3)$	held	$1.01(3)$	held

(1.0% standard deviation over the 42 data points). A total uncertainty of 2.5% was then assumed in all known values.

As an alternative, we also consider a two-parameter approach in which $N^{(\mathrm{rel})}$ is independently measured (and not treated as a fitting parameter), and only the two lowest-order relative reaction curves are used for fitting. The results are shown in Table 4.2. Fits were weighted with data standard deviations, and the parameters where evaluated with $N^{(\mathrm{rel})}$ fixed at the mean known value, and one standard deviation above and below. It is observed that the reported fit values are strongly sensitive to variations in $N^{(\mathrm{rel})}$, making the success of this approach contingent upon the accuracy of $N^{(\mathrm{rel})}$. For better comparison, the uncertainty of individual fits with $N^{(\mathrm{rel})}$ fixed at the mean value are also presented in the Table.

In the simple case of a set of constant fit functions, the least-squares fit lines converge to the data mean values, given the existence of a unique solution at that point. Therefore it is possible to manually extract the mean values of a and b and find the exact numerical solutions to Equations (4.11). This was done, and the solutions were also found at a distance of 1 (or $1/2$) data standard deviation on each side of the mean a and b points to estimate the solution uncertainties. The results fully agree with the fitting approach, Table 4.1, with details appearing in Section 5.1.3. Comparison between different samples suggests that the accuracy of the technique can be strongly

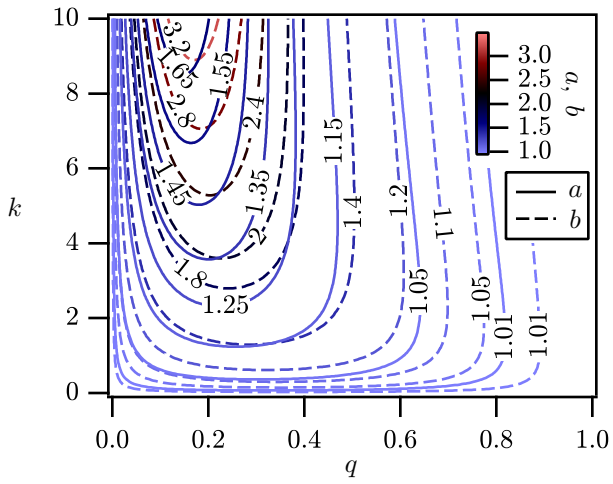

Figure 4.3: Contour maps of a and b as a function of q and k.

affected by the mixture concentration ratio, k. In the next section, we use this direct numerical solution method to explore the reliability of the technique over a wide range of k and q parameters.

4.4.2 General study of uncertainties in parameter space

It is beneficial to understand the dependence of the fitted parameter uncertainties on the underlying q and k parameters. In other words, we would like to know the region of (q, k) parameter space in which the technique of higher order correlations can be reliably used to resolve mixtures (and thus reactions). We start by using a and b parameters, then consider the approach of fixed $N^{(\mathrm{rel})}$ value.

Using a and b parameters

Consider the system of equations 4.11 that relate a and b to q and k. As a preliminary visual inspection of existence and uniqueness of solutions at each point (a, b), we present a graph of a and b contour lines versus (q, k) in Figure 4.3. The graph shows examples of existing solutions where lines of constant a and b cross. However, there are clearly regions where contour lines become parallel and the solution values can be either extreme or non-existent, and the uncertainties are large. Our numerical solver may fail to converge at these points.

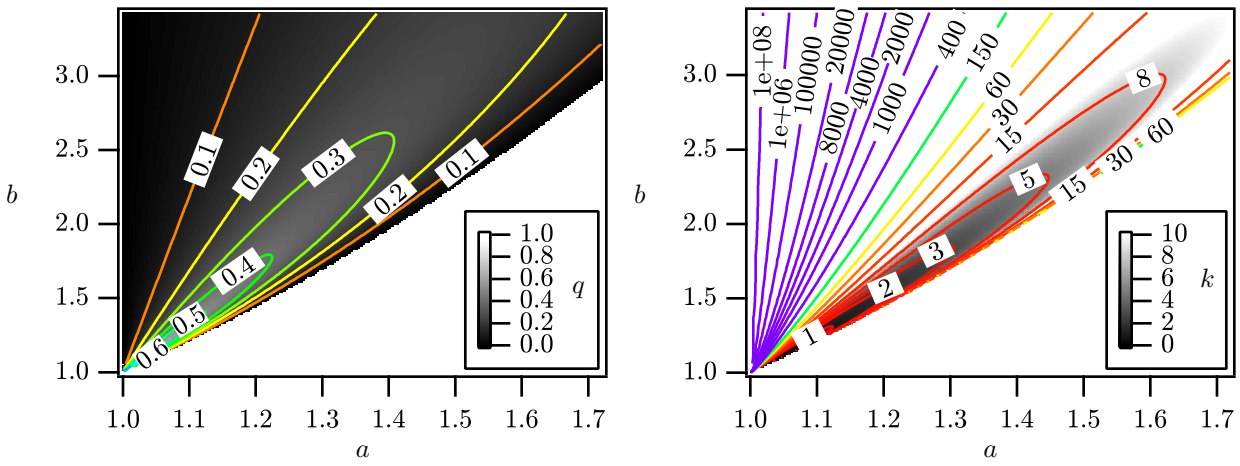

Figure 4.4: Solutions to Equations (4.11). No physical solutions exist in the right lower region.

Figure 4.4 shows the values of q and k calculated over a rectangular region of a and b large enough to include $0 < k < 10$. No physical solutions were found in the lower right part of the graph, and extremely large values of k are also present. However, this does not necessarily translate to large *relative* errors.

We proceed by calculating relative uncertainties in q and k given fixed relative uncertainties of 0.1% and 0.5% in a and b respectively. These relative uncertainties in a and b are reminiscent of those obtained from data standard deviations (Table 4.1). Full standard deviation in q was calculated using half-standard-deviation steps around the mean of a and b,

$$\delta q = \sqrt{\left[q(\bar{a} + \delta a/2, \bar{b}) - q(\bar{a} - \delta a/2, \bar{b})\right]^2 + \left[q(\bar{a}, \bar{b} + \delta b/2) - q(\bar{a}, \bar{b} - \delta b/2)\right]^2} \tag{4.14}$$

then normalized by the value at the mean point, $q(\bar{a}, \bar{b})$, to obtain the relative uncertainty, and similarly for k. The results are shown as functions of q and k in Figure 4.5. The relative uncertainties in q and k at points $q \approx 0.4$ and $k \approx 0.25$, 1, 4 agree with those in experiments (Table 4.1). In fact, these graphs are extensions of such estimations assuming similar relative uncertainties in a and b. Figure 5.1 shows the results for 0.2% and 1.0% errors assumed in a and b, respectively. Approximate linear scaling is observed in most regions.

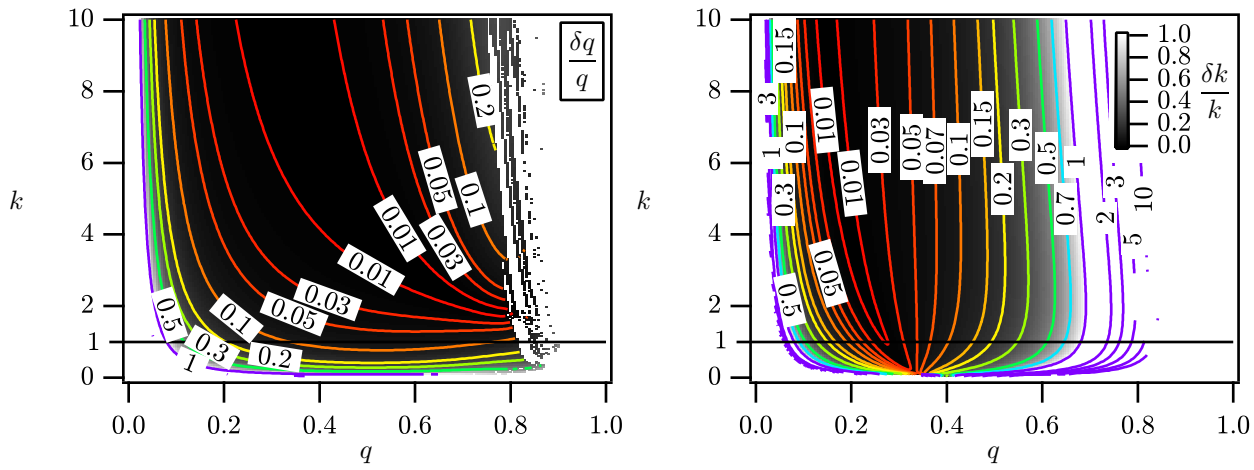

Figure 4.5: Relative uncertainties in q and k assuming $\delta a/a = 0.001$ and $\delta b/b = 0.005$ shown as functions of q and k. The image gray scale is between 0 and 1 in all relative uncertainty graphs. Near the edges, where relative uncertainties in k are very large, solutions are hard to reach and some points are left out from both graphs.

The graphs of relative uncertainty show that the brightness ratio, q, can be generally determined with good precision. The relative uncertainty in k is generally much greater. The graphs show that in the regime $k \gtrsim 1$ the technique can be successfully applied when $0.1 \lesssim q \lesssim 0.5$. This range of q, however, slowly varies as k increases. Both graphs suggest that useful results generally cannot be obtained when k is significantly smaller than 1 (many more bright molecules than dim molecules).

In our experiments we set k by mixture design and set q by choice of fluorophores. However, in most applications, these values are not known a priori and are in fact to be determined. Therefore, an experimentalist might be more interested in knowing the applicability of the technique solely based on the observed a and b values. Figure 5.2 and 5.3 show the relative uncertainties in q and k as functions of a and b, which can be more useful in mixture resolving experiments.

Using fixed $N^{(\mathrm{rel})}$

As explained earlier, in situations that $N^{(\mathrm{rel})}$ can be independently and reliably determined, the fits can be applied to only the two lowest order relative correlations, $R^{(\mathrm{rel})}_{1,1}$ and $R^{(\mathrm{rel})}_{2,1}$, which in principle have much smaller uncertainties than order $(2, 2)$. To make this analysis simpler, we assume that the uncertainties in $R^{(\mathrm{rel})}_{1,1}$ and $R^{(\mathrm{rel})}_{2,1}$ are also much less than the uncertainty in $N^{(\mathrm{rel})}$, thus we define

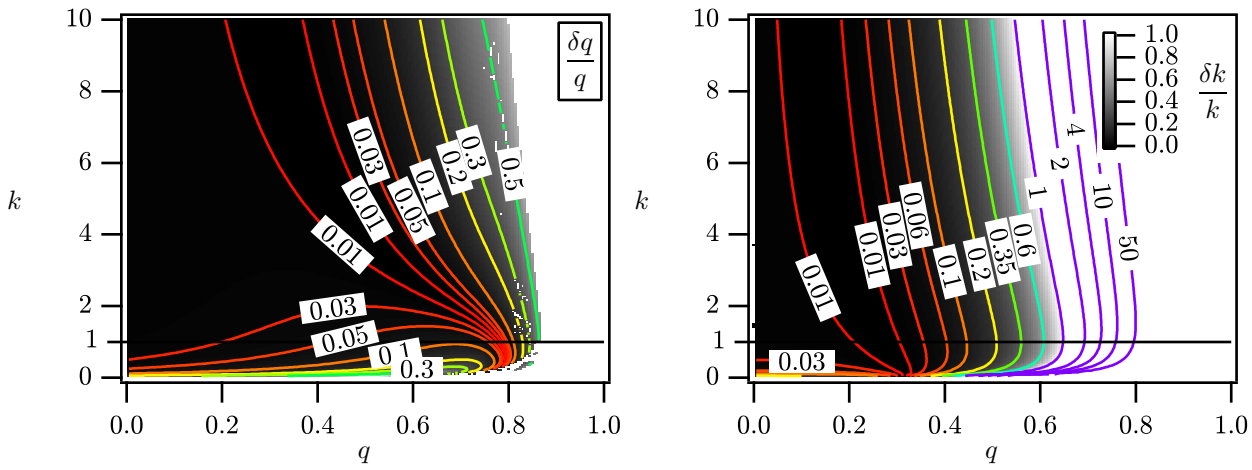

Figure 4.6: Relative uncertainties in q and k assuming $\delta N^{(\mathrm{rel})}/N^{(\mathrm{rel})} = 0.01$ shown as functions of q and k. Similar to the analysis using a and b, relative uncertainties grow rapidly when q approaches 1. However, the restriction is removed on q approaching 0.

$$\begin{aligned} A_{1,1} &= N^{(\mathrm{rel})} R^{(\mathrm{rel})}_{1,1} = \frac{(1+k)(1+kq^2)}{(1+kq)^2} \\ A_{2,1} &= N^{(\mathrm{rel})} R^{(\mathrm{rel})}_{2,1} = \frac{(1+k)(1+kq^3)}{(1+kq^2)(1+kq)} \end{aligned} \tag{4.15}$$

This leads to correlated uncertainties in $A_{1,1}$ and $A_{2,1}$, thus we combine them directly rather than in quadrature:

$$\frac{\delta q}{q} = \frac{|q(A_{1,1} + \delta A_{1,1}/2, A_{2,1} + \delta A_{2,1}/2) - q(A_{1,1} - \delta A_{1,1}/2, A_{2,1} - \delta A_{2,1}/2)|}{q(A_{1,1}, A_{2,1})} \tag{4.16}$$

and similarly for k.

Figure 4.6 shows the relative uncertainties in q and k as functions of q and k, when $N^{(\mathrm{rel})}$ is known with a constant relative uncertainty of 1%. Figure 5.4 shows the results assuming 3% uncertainty in $N^{(\mathrm{rel})}$. Comparison of these graphs with their counterparts in the (a, b) approach (Figure 4.5) shows that the reliability region is similarly restricted to q values sufficiently smaller than 1. However, the region $k < 1$ is more available in the known $N^{(\mathrm{rel})}$ approach. Also in this approach, there is no restriction on q being arbitrarily small, and the uncertainties in q and k remain on the order of the uncertainty in $N^{(\mathrm{rel})}$ when q is very small.

To make sense of the fact that having an arbitrarily dim species causes no problem when $N^{(\mathrm{rel})}$ is known, consider the most extreme case, $q = 0$. In this case, the experiment yields identical $R_{1,1}^{(\mathrm{rel})}$ and $R_{2,1}^{(\mathrm{rel})}$ (within their experimental uncertainties, of course, which were assumed to be small). The equation $R_{1,1}^{(\mathrm{rel})} = R_{2,1}^{(\mathrm{rel})}$ has three possible solutions $q = 0$, $q = 1$, and $k = 0$. We then compare $1/R_{1,1}^{(\mathrm{rel})}$ to the known $N^{(\mathrm{rel})}$ value. If they are equal, the conclusion is that no second species exists and the sample is pure ($q = 1$ and/or $k = 0$). If they are not equal, we conclude that $q = 0$ and solve $R_{1,1}^{(\mathrm{rel})} = (1+k)/N^{(\mathrm{rel})}$ to determine k. The uncertainties then rely on that of $N^{(\mathrm{rel})}$ (again, ignoring those in $R_{1,1}^{(\mathrm{rel})}$ and $R_{2,1}^{(\mathrm{rel})}$).

As before, we present the relative uncertainties in q and k, given a constant uncertainty in $N^{(\mathrm{rel})}$, as functions of $A_{1,1}$ and $A_{2,1}$ in Figures 5.6 and 5.5.The graphs can be useful in actual mixture resolving experiments where q and k are not known beforehand.

As a brief discussion, we consider two possible approaches to resolve a mixture of two species: using three correlation orders (equivalent to a, b approach), or, measuring $N^{(\mathrm{rel})}$ independently and using two correlation orders. Both approaches suggest that systems with $k > 1$, i.e. more dim molecules than bright ones, are more suitable for this type of analysis (yielding more distinct relative reaction curves, see Figure 4.2). In both approaches, the brightness levels have to be sufficiently distinct ($q < 0.5$) to provide successful estimation of the k value. However, in the approach of three correlations, a species too dark compared to the other can also cause uncertainty issues (by yielding indistinguishable relative reaction curves). On the contrary, the approach of fixed $N^{(\mathrm{rel})}$ remains useful even when q is very small. Choosing the fixed $N^{(\mathrm{rel})}$ approach is advised only if a reliable independent measurement of $N^{(\mathrm{rel})}$ is available. Both approaches, or the technique in general, are more precise in determining the relative brightness value, q, compared to k, with q up to near 0.8 possible for such applications. However, the most favorable region for measurement of q does not necessarily overlap with that for k, in either approach.

4.4.3 Reaction experiments

FITC, a derivative of fluorescein, is commonly used as a fluorescent label and pH indicator in cellular and molecular biology experiments. The fluorescence of FITC is strongly decreased upon binding of a hydrogen ion, in a reaction of the form

$$\mathrm{Fl}^{2-} + \mathrm{H}^{+} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{HFl}^{-} \tag{4.17}$$

where Fl^{2-} and Fl^{-} denote the dianion (bright) and anion (dark) states of FITC, and k_+ and k_- denote the association and dissociation rate constants, respectively. Following our notation in the Theory section, $k = [\mathrm{HFl}^{-}]/[\mathrm{Fl}^{2-}]$ is the average concentration ratio of dark/bright molecules, and q denotes the brightness ratio of the two states in each molecule. Higher-order correlation analysis of a single measurement is sufficient to obtain q and k simultaneously. The analysis also directly yields the overall relaxation time of the reaction, t_{R}. In the absence of any other reaction, one would expect the overall kinetic rate t_{R}^{-1} to be equal to the sum of the forward and reverse rates, $[\mathrm{H}^{+}]k_+$ and k_- , respectively. However, it has been shown previously [74] using conventional FCS that the overall relaxation time depends not only on the concentration of H^{+} , but also on the concentration of the phosphate buffer. This indicates direct donation and acceptance of protons between FITC and phosphate ions, which adds to the overall relaxation rate in the form

$$t_{\mathrm{R}}^{-1} = [\mathrm{H}^{+}]k_+ + k_- + c[\mathrm{P}_{\mathrm{tot}}] \tag{4.18}$$

$[\mathrm{P}_{\mathrm{tot}}]$ denotes the total concentration of the phosphate ions of all aqueous forms in the solution, and c depends on pH and buffer properties. By successive dilutions of the buffer and extrapolation to zero buffer concentration we obtain

$$t_{\mathrm{R0}}^{-1} = [\mathrm{H}^{+}]k_+ + k_- \tag{4.19}$$

independent of the buffer. This, combined with $k = [\mathrm{H}^+]k_+/k_-$ yields the forward and reverse rates, $[\mathrm{H}^+]k_+$ and k_- , with no further measurement required. Finally, if the pH of the buffer is also known, the k_+ value can be isolated from $[\mathrm{H}^+]k_+$, and the dissociation constant of the reaction can be found using the Henderson-Hasselbalch equation:

$$\mathrm{p}K_\mathrm{a} = \mathrm{pH} + \log_{10}(\mathrm{k}) \tag{4.20}$$

Here, the solution of interest (test solution) contains FITC in thdesired phosphate buffer for reaction analysis. Guided by the results of the uncertainty analysis in mixtures, we chose test solutions with pH values less than $\mathrm{p}K_\mathrm{a}$, which correspond to $k > 1$. As a reference solution, FITC at a high pH value can be used, since no significant protonation reaction was observed in solutions ranging from pH 8.4 to 10.3 (data not shown). Therefore, a single measurement of a solution at pH 9.03 has been used as the reference for all measurements in this section. Figure 4.7 shows the higher-correlation curves (left) and their corresponding relative reaction curves (right) at pH values 5.94, 5.73, and 6.05. At each pH value, the reaction parameters were obtained by simultaneously fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$ and $R_{2,2}^{(\mathrm{rel})}$ using Equation (4.12). The two-state mechanism is verified by the decay of all relative reaction curves to the same baseline level, and good fit to the single-exponential model. Due to some instability of FITC observed in low-pH environment, the experiment at each pH was repeated six times to find the associated uncertainty across measurements. Table 4.3 shows the average reaction parameters over the six trials, the standard error of the mean, and the average reported fit uncertainty in each trial. The variations across measurements were found to be small enough to provide reliable measurements of each parameter. They are, however, generally greater than the uncertainty of each individual fit. We will use the former values for further uncertainty propagation.

To find the zero-buffer relaxation time, $t_{\mathrm{R}0}$, separate measurements were performed with successive dilutions of the buffer, then t_R was extrapolated to zero buffer concentration as shown in Figure 4.8 and reported in Table 4.3. No dependence of the other measured parameters on buffer concentration was observed in the examined concentration range (data not shown). A good linear

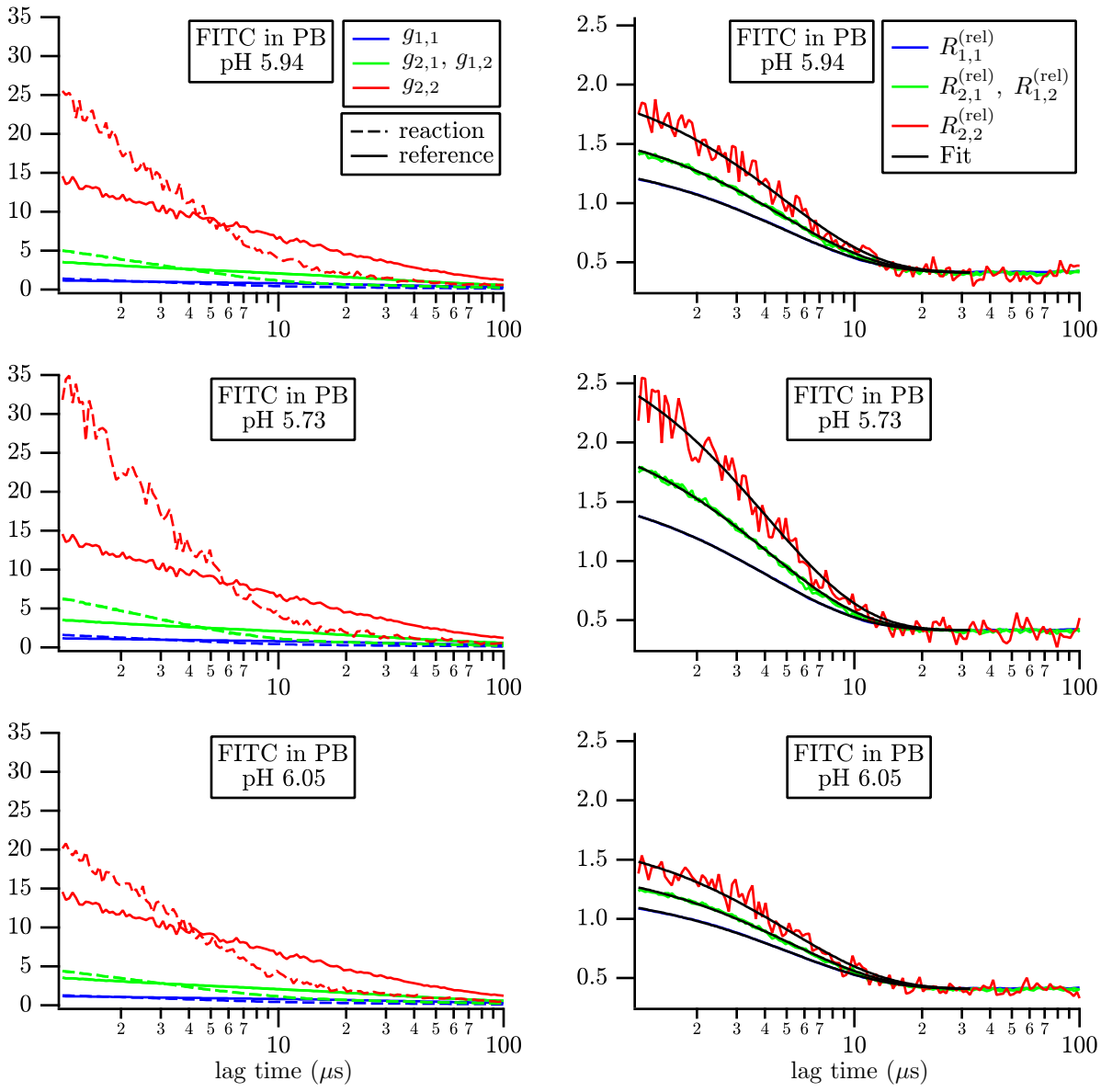

Figure 4.7: Left column: Second- and higher-order correlation functions obtained for FITC in phosphate buffer (PB) at different pH values (dashed lines). FITC at pH 9.03 was used as reference in all cases (solid lines). The same color scheme and axis scale applies to all graphs in each column. Right column: The corresponding relative reaction functions obtained from the higher-order correlation functions on the left. The global fit lines to the first $N_\mathrm{p} = 100$ data points (each curve) are shown in black. Data was collected with 25 μW excitation power and a single detector.

Table 4.3: Protonation reaction of FITC in phosphate buffer resolved by fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$ and $R_{2,2}^{(\mathrm{rel})}$ (Data for $25\,\mu$W excitation power, single detector). Where two numbers appear in parentheses, the first number shows the standard error of mean over six trials, and the second number shows the average reported fit uncertainty in one trial.

pH	**5.94(3)**	**5.73(3)**	**6.05(3)**
q	0.0586(2)(4)	0.0584(5)(2)	0.0586(6)(5)
k	4.23(2)(2)	6.60(9)(4)	3.36(3)(2)
$N^{(\mathrm{rel})}$	2.484(49)(3)	2.497(47)(3)	2.436(47)(3)
$t_{\mathrm{R}}\,(\mu\mathrm{s})$	4.845(8)(12)	4.118(6)(8)	5.118(15)(13)
$t_{\mathrm{R0}}\,(\mu\mathrm{s})$	11.27(18)	7.82(14)	13.68(41)
$k_-\,(10^4\,\mathrm{s}^{-1})$	1.70(3)	1.68(3)	1.68(5)
$[\mathrm{H}^+]k_+\,(10^4\,\mathrm{s}^{-1})$	7.17(12)	11.1(2)	5.63(17)
$k_+\,(10^{10}\,\mathrm{s}^{-1}\mathrm{M}^{-1})$	6.2(4)	6.0(4)	6.3(5)
$\mathrm{p}K_{\mathrm{a}}$	6.57(3)	6.55(3)	6.58(3)

relationship is observed between t_{R}^{-1} and $[\mathrm{P_{tot}}]$, in agreement with Equation (4.18). It should be noted that knowing only the dilution factors is sufficient for correct extrapolation to zero, since the starting $[\mathrm{P_{tot}}]$ value does not affect the $y-$intercept. The variations in the slope of the lines with pH reflect the dependence of c on pH, and/or possible error in the value of the starting buffer concentration, both of which are inconsequential in our current analysis.

In Table 4.3, the values of $k_- = t_{\mathrm{R0}}^{-1}/(1+k)$ and $[\mathrm{H}^+]k_+ = t_{\mathrm{R0}}^{-1}k/(1+k)$ were obtained using k and t_{R0} alone. Then the value of k_+ was calculated using $[\mathrm{H}^+]k_+$ and the pH value, which increases the resultant uncertainty of k_+ compared to $[\mathrm{H}^+]k_+$. Finally, the $\mathrm{p}K_{\mathrm{a}}$ value in each case was obtained by substituting k and pH values into Equation (4.20), with the uncertainty in pH measurement dominating that of $\log_{10}(k)$.

The values of reaction constants k_+ , k_- , and $\mathrm{p}K_{\mathrm{a}}$ agree within uncertainty between measurements at different pH values. To independently verify the obtained $\mathrm{p}K_{\mathrm{a}}$ value, the overall fluorescence intensity of FITC, I, was measured as a function of pH and shown in Figure 4.9. Error bars were calculated by repeating each measurement 5 times and combining the resulting uncertainty with that of pH. Data was then fitted to the Henderson-Hasselbalch equation in the form $\mathrm{p}K_{\mathrm{a}} = \mathrm{pH} + \log_{10}[(\mathrm{I_{max}} - \mathrm{I})/(\mathrm{I} - \mathrm{I_{min}})]$. The result, $\mathrm{p}K_{\mathrm{a}} = 6.57 \pm 0.03$, is in good agree-

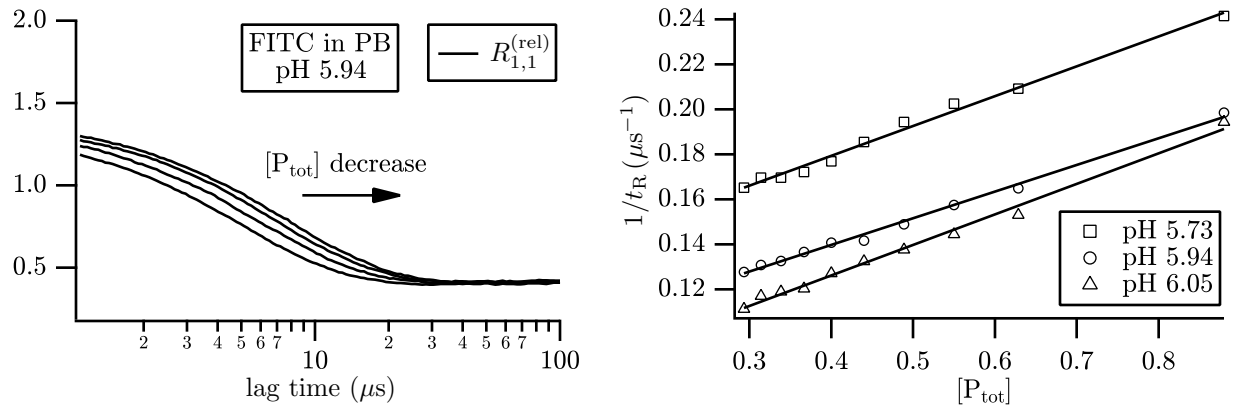

Figure 4.8: Left: The reaction relaxation time, t_{R}, increases with decreased total phosphate buffer concentration, $[\mathrm{P_{tot}}]$. Examples of $R_{1,1}^{(\mathrm{rel})}$ are shown at $[\mathrm{P_{tot}}]$ ranging from 0.88 to $0.29\,\mathrm{mM}$. Right: The relaxation time at zero buffer concentration has been found by extrapolation.

ment with the results from higher-order correlation analysis, Table 4.3. Using the dependence of the fluorescence intensity on pH, Highsmith [75] has reported a $\mathrm{p}K_{\mathrm{a}}$ value of 6.4 for free FITC, measured at ionic strengths above $100\,\mathrm{mM}$. Since our measurements are at sub-millimolar buffer concentrations, a slight decrease in the apparent $\mathrm{p}K_{\mathrm{a}}$ at higher ionic strengths can be expected. Leonhardt et al. [76] have reported 6.7 for $\mathrm{p}K_{\mathrm{a}}$ of fluorescein, while Sjöback et al. [77] have obtained 6.43. Geisow [78] has reported $\mathrm{p}K_{\mathrm{a}}$ values ranging from 6.30 to 6.80 for FITC conjugated to dextran or ovalbumin.

Widengren et al. [74] have found the full reaction parameters in a pioneering application of conventional (order-$(1,1)$) FCS. Since conventional FCS is insufficient to determine the reaction parameters independently, those authors modeled the dependence of the order-$(1,1)$ reaction amplitude and the relaxation time on total buffer concentration and pH. The model containing seven parameters was then globally fitted to a series of measurements at various buffer concentrations and pH values. Their reported values are $k_+ = 4.0 \times 10^{10}\,\mathrm{M^{-1}s^{-1}}$ and $k_- = 2.5 \times 10^{4}\,\mathrm{s^{-1}}$ which yield $\mathrm{p}K_{\mathrm{a}} = -\log_{10}(k_-/k_+) = 6.2$ (no uncertainty reported). They also reported $q = 0.05$, not far from our results. In comparison, we get some reaction parameters (q, k, $N^{(\mathrm{rel})}$, t_{R}) with as few as two measurements (including the reference). Adding a single measurement of pH allows us to determine $\mathrm{p}K_{\mathrm{a}}$ immediately. Since the inevitable buffer interactions affect t_{R} (not the case in

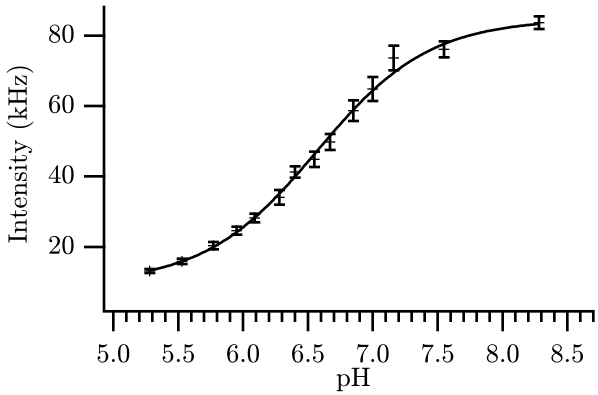

Figure 4.9: The fluorescence intensity of FITC in phosphate buffer is measured as a function of pH. Fitting to Henderson-Hasselbalch equation yields $\mathrm{p}K_\mathrm{a} = 6.57 \pm 0.03$. Error bars show variations across five measurements at each pH value.

all systems), we remove this effect by extrapolation to zero and then the forward and reverse rate constants are also determined. Higher-order correlation analysis thus provides a direct approach in which additional solution preparations and measurements at different known pH values and known buffer concentrations are not required (we only diluted the test sample by known factors), avoiding their associated systematic and stochastic uncertainties. Moreover, the modeling and approximations involving the reaction of FITC with buffer, which may introduce unknown systematic errors, have been avoided.

Looking at Figure 4.7 again, it can be seen that solutions with larger k (lower pH) have more distinct higher-order relative reaction curves, similar to the case of mixtures, resulting in less fitting errors. However, this comes at the cost of reduced overall fluorescence intensity in more quenched solutions, which can be partially compensated by using higher fluorophore concentrations. The concentration of FITC in test solutions was therefore prepared at twice that of the reference solution. The applicability of this approach is limited because, as seen in Chapter 2, the signal-to-noise ratio (SNR) of higher order correlations decreases at very high concentrations. Therefore, the suitable range of pH for this analysis is broad but still limited by these conditions.

One last observation to note is that the measured $N^{(\mathrm{rel})}$ value is slightly larger than 2, the ratio of the test to reference concentrations. This has been verified to be caused by a slow photoinduced blinking of FITC with relaxation time scale $> 300\,\mu$s. We present further study of this slow blink-

ing in Section 5.4. At short time scales, the slow blinking effect simply acts as a constant scaling factor which only affects the measured value of $N^{(\mathrm{rel})}$, but has no effect on the other measured parameters of the fast protonation reaction. We show that this slow blinking is photoinduced, and does not affect the protonation reaction results, by conducting measurements at two different excitation powers. The results that appear in this section were obtained with $25\,\mu\mathrm{W}$ excitation power, and we present replicated results with $62\,\mu\mathrm{W}$ excitation power in Table 5.7 and Figure 5.10. Apart from the value of $N^{(\mathrm{rel})}$ which is not intrinsic to the reaction, no significant effect on the results of the reaction analysis is observed. Contrary to the fast triplet blinking, the slow blinking is not independent from the protonation reaction, therefore not canceled out by normalization by the reference data. We also verify in Section 5.4 that the fast triplet blinking is independent of the protonation reaction and is effectively removed upon normalization by the reference data, consistent with the theory.

Finally, our proposed sub-binning approach to calculate higher-order correlations is verified when applied to both single-detector and two-detector experiments. The measurements at the higher excitation power, $62\,\mu\mathrm{W}$, were conducted using two detectors, as in mixture studies reported earlier. The experiments at the lower excitation power, $25\,\mu\mathrm{W}$, were conducted using only one of the detectors, to avoid significant loss in SNR. In the latter case, the output of the detector was fed to both inputs of the two-detector two-sub-bin program with no modification of the algorithm. The methods provide similar reaction analysis results within uncertainty (Tables 4.3 and 5.7). This shows that the two-detector two-sub-bin program is also capable of analyzing the signal from a single detector and successfully removing detector artifacts.

4.5 Conclusion

The analysis of higher-order correlations formulated based on bivariate cumulants of signal intensity provides a simple and powerful framework for full characterization of the parameters in molecular reaction experiments. Direct measurement of parameters using higher-order correlation analysis eliminates the need for many additional measurements and assumptions that would be

required using conventional FCS alone. The technique has been successfully applied to resolve two-component mixtures at various concentration ratios, by analyzing these systems as very slow reactions within the general framework. The results confirm the applicability of the technique to analyze slow- or non-reacting systems, and pave the way for its application to faster reacting systems. As an example of such fast reactions, we have applied the technique to fully characterize the protonation reaction of FITC. The results also confirm our proposed modifications to improve time resolution and remove detector artifacts without any setup-specific modeling. In addition, they verify that the systematic effects of finite bin size remain insignificant within the stochastic uncertainty of the measurements. The analysis of the uncertainties using mixtures shows the technique can be reliably applied over a wide range of brightness and concentration ratios. Any system of fast or slow reacting components at distinct brightness levels can in principle be the target of higher-order fluorescence correlations analysis.

Chapter 5

Testing higher-order fluorescence correlation spectroscopy—supporting information

5.1 Mixture Experiments

5.1.1 Fits weighted with theoretical errors

Non-linear least-squares fits were weighted with theoretical data errors and the Global Fit package in Igor Pro was used to simultaneously fit all curves for each sample. The fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$, and $R_{2,2}^{(\mathrm{rel})}$ were conducted using Equations (4.10) and the results are shown in Table 5.1. The initial fitted parameter values were set to the "known" values (explained in Chapter 4). The reduced chi-squared values were calculated as $\chi^2_{\mathrm{red}} = \chi^2/(N_{\mathrm{p}} - n_{\mathrm{p}})$, where n_{p} is the number of (non-held) fitted parameters, and χ^2 is the weighted sum of squared errors reported by the fit program following its standard definition.

Comparison with the actual standard deviations of the curves showed that theoretical data uncertainties were slightly larger in most cases. This explains the χ^2_{red} values less than 1 for samples 50:50 and 20:80 in Tables 5.1 and 5.2. In the case of sample 80:20 in Table 5.1, the data variations in order $(1, 1)$ due to triplet blinking exceeds the theoretical errors by almost a factor of two. This causes a jump in χ^2_{red} to greater than 1. The numbers in parentheses in the "measured" column show the reported uncertainties by the fit procedure. Since the accuracy of the reported fit uncertainties depends on the accuracy of the weighting values (theoretical errors), and given the deviations of χ^2_{red} from unity, we conclude that the reported uncertainties are not accurate estimates in Tables 5.1 and 5.2. We have improved these values in Chapter 4 by using data standard deviations as errors.

Using theoretical data uncertainties and the same number of data points, the values of $N^{(\mathrm{rel})}$ were held fixed at the known values and the global fit was applied to only $R_{1,1}^{(\mathrm{rel})}(t)$ and $R_{2,1}^{(\mathrm{rel})}(t)$.

The results are shown in Table 5.3. The increase in the reduced chi-squared values compared to the previous Tables can be understood by noting that due to triplet blinking and smaller theoretical uncertainties, lower-order curves contribute larger weighted residuals to χ^2_{red} . As in Chapter 4, the sensitivity of the reported fit values, as well as the reported uncertainties, to the accuracy of the fixed $N^{(\mathrm{rel})}$ value is observed.

Table 5.1: TAMRA- and R6G-oligo mixture components resolved by global fitting to $R^{(\mathrm{rel})}_{1,1}$, $R^{(\mathrm{rel})}_{2,1}$, and $R^{(\mathrm{rel})}_{2,2}$, using theoretical uncertainties. In all tables, numbers in parentheses indicate the uncertainty in the corresponding last digits.

	50:50		**80:20**		**20:80**	
	$\chi^2_{\mathrm{red}} = 0.53$		$\chi^2_{\mathrm{red}} = 1.83$		$\chi^2_{\mathrm{red}} = 0.40$	
	known	**measured**	**known**	**measured**	**known**	**measured**
q	0.399(10)	0.406(61)	0.399(10)	0.398(13)	0.399(10)	0.53(26)
k	1.05(3)	1.03(20)	4.21(11)	3.95(96)	0.263(7)	0.28(24)
$N^{(\mathrm{rel})}$	1.03(2)	1.03(3)	1.04(3)	1.01(2)	1.01(3)	0.99(3)

Table 5.2: TAMRA- and R6G-oligo mixture components resolved by global fitting to a and b**,** using theoretical uncertainties.

	50:50		**80:20**		**20:80**	
	$\chi^2_{\mathrm{red}} = 0.50$		$\chi^2_{\mathrm{red}} = 0.41$		$\chi^2_{\mathrm{red}} = 0.57$	
	known	**measured**	**known**	**measured**	**known**	**measured**
q	0.399(10)	0.396(43)	0.399(10)	0.401(9)	0.399(10)	0.42(19)
k	1.05(3)	0.99(12)	4.21(11)	3.94(67)	0.263(7)	0.22(7)

Table 5.3: TAMRA- and R6G-oligo mixture components resolved by global fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$, and fixed $N^{(\mathrm{rel})}$, using theoretical uncertainties. In the measured parameters, the first number in parentheses shows the variations in fitting results when $N^{(\mathrm{rel})}$ is varied by one standard deviation, and the second number shows the individual fit uncertainty with $N^{(\mathrm{rel})}$ fixed at the mean point. The "held" parameters were held fixed at the "known" mean value and at one standard deviation above and below to find the associated uncertainty.

	50:50		**80:20**		**20:80**	
	$\chi^2_{\mathrm{red}} = 0.64$		$\chi^2_{\mathrm{red}} = 2.68$		$\chi^2_{\mathrm{red}} = 0.43$	
	known	**measured**	**known**	**measured**	**known**	**measured**
q	$0.399(10)$	$0.404(47)(2)$	$0.399(10)$	$0.380(18)(1)$	$0.399(10)$	$0.359(142)(13)$
k	$1.05(3)$	$1.02(18)(3)$	$4.21(11)$	$3.08(59)(7)$	$0.263(7)$	$0.201(48)(10)$
$N^{(\mathrm{rel})}$	$1.03(2)$	held	$1.04(3)$	held	$1.01(3)$	held

5.1.2 Fits weighted with data standard deviations

The two-parameter fitting methods (the a, b approach, and the fixed $N^{(\mathrm{rel})}$ approach) weighted with data standard deviations were discussed in Chapter 5. The only remaining case is three-parameter fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$, and $R_{2,2}^{(\mathrm{rel})}$ using data standard deviations, which is presented in Table 5.4. As expected, all chi-squared values are practically equal to 1, showing the fit lines have actually converged to the mean values. The reported fitting uncertainties are more reasonable in this case. Only in the case of three-parameter fitting for sample 20:80, the fit procedure did not converge due to generally higher uncertainties at this mixture concentration ratio and fewer data points (N_{p}) available per fitting parameters, compared to two-parameter methods.

Table 5.4: TAMRA- and R6G-oligo mixture components resolved by global fitting to $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$, and $R_{2,2}^{(\mathrm{rel})}$, using data standard deviations as fitting weights. In the last column the fit failed to properly converge.

	50:50		**80:20**		**20:80**	
	$\chi^2_{\mathrm{red}} = 1.00$		$\chi^2_{\mathrm{red}} = 1.00$		-	
	known	**measured**	**known**	**measured**	**known**	**measured**
q	0.399(10)	0.435(31)	0.399(10)	0.401(6)	0.399(10)	-
k	1.05(3)	1.15(15)	4.21(11)	4.13(48)	0.263(7)	-
$N^{(\mathrm{rel})}$	1.03(2)	1.01(2)	1.04(3)	1.01(1)	1.01(3)	-

5.1.3 Numerical solutions at individual points

In the simple case of a set of constant fit functions, the least-squares fit lines converge to the data mean values, given the existence of a unique solution at that point. Therefore it is reasonable to manually extract the mean values and find the exact numerical solutions to Equations (4.11).What we obtained earlier with fitting remains valuable in giving us a sense of the performance of the fit procedure, its reported uncertainties, and the reliability of the theoretical fitting weights.

To simplify matters, we focus on the two-parameter model only. Table 5.5 shows the mean, standard deviation, and error of the mean for $N_{\mathrm{p}} = 64$ data points of a and b, each. Precise values of q and k were found using a numerical solver at the mean point. Assuming a linear approximation, the values were also evaluated at a distance of $1/2$ standard deviation on each side of the mean point along the a and b coordinates. These values are also reported in the table to demonstrate the suitability of linear approximation, particularly in the regime $k \geq 1$. The uncertainty in q was calculated in the following way:

$$\delta q = \sqrt{\left[q(\bar{a} + \delta a/2, \bar{b}) - q(\bar{a} - \delta a/2, \bar{b})\right]^2 + \left[q(\bar{a}, \bar{b} + \delta b/2) - q(\bar{a}, \bar{b} - \delta b/2)\right]^2} \tag{5.1}$$

and similarly for k. The relative uncertainty $\delta q/q$ denotes normalization by the value at the mean point, $q(\bar{a}, \bar{b})$, and similarly for k.

Table 5.5: TAMRA- and R6G-oligo mixture components resolved by numerical solution at mean a and b. Uncertainties were evaluated by variation by $1/2$ standard deviation around the mean (Equation (5.1)). $\bar{a}$, σ_a, and δa represent the mean, standard deviation, and standard deviation of the mean of a (similarly for b).

	50:50	**80:20**	**20:80**
$\bar{a}$	1.1076	1.2129	1.0289
σ_a	0.0070	0.0101	0.0056
δa	0.00088	0.0013	0.00070
$\delta a/a$	0.079%	0.10%	0.068%
$\bar{b}$	1.2924	1.6876	1.0778
σ_b	0.0414	0.0695	0.0278
δb	0.0052	0.0087	0.0043
$\delta b/b$	0.40%	0.51%	0.40%
$q(\bar{a}, \bar{b})$	0.4128	0.4020	0.5704
$q(\bar{a}, \bar{b} + \delta b/2)$	0.4239	0.4035	0.6189
$q(\bar{a}, \bar{b} - \delta b/2)$	0.4012	0.4002	0.5160
$q(\bar{a} + \delta a/2, \bar{b})$	0.4069	0.4006	0.5466
$q(\bar{a} - \delta a/2, \bar{b})$	0.4188	0.4033	0.5939
δq	0.0256	0.0043	0.1133
$\delta q/q$	6.2%	1.1%	20%
q (known)	0.399(10)	0.399(10)	0.399(10)
$k(\bar{a}, \bar{b})$	1.0510	4.0685	0.3388
$k(\bar{a}, \bar{b} + \delta b/2)$	1.0954	4.2299	0.4393
$k(\bar{a}, \bar{b} - \delta b/2)$	1.0113	3.9145	0.2758
$k(\bar{a} + \delta a/2, \bar{b})$	1.0357	3.9986	0.3113
$k(\bar{a} - \delta a/2, \bar{b})$	1.0675	4.1408	0.3737
δk	0.0899	0.3460	0.1750
$\delta k/k$	8.6%	8.5%	52%
k (known)	1.05(3)	4.21(11)	0.263(7)

We also calculated the solutions at points located one full standard deviation on each side of the mean point, and the uncertainties in q were calculated according to

$$\delta q = \frac{1}{2}\sqrt{\left[q(\bar{a}+\delta a,\bar{b}) - q(\bar{a}-\delta a,\bar{b})\right]^2 + \left[q(\bar{a},\bar{b}+\delta b) - q(\bar{a},\bar{b}-\delta b)\right]^2} \tag{5.2}$$

and similarly for k. The results are presented in supporting Table (5.6). Larger deviation steps should, in principle, encompass more non-linear effects. However, the agreement of final uncertainties with those of half-standard-deviation steps shows the linear approximation is good in both scales, particularly for 50:50 and 80:20 mixture ratios.

Table 5.6: TAMRA- and R6G-oligo mixture components resolved by numerical solution at mean a and b. Uncertainties were evaluated by variation by one standard deviation around the mean (Equation (5.2)). $\bar{a}$, σ_a, and δa represent the mean, standard deviation, and standard deviation of the mean of a (similarly for b).

	50:50	**80:20**	**20:80**
$\bar{a}$	1.1076	1.2129	1.0289
σ_a	0.0070	0.0101	0.0056
δa	0.00088	0.0013	0.00070
$\delta a/a$	0.079%	0.10%	0.068%
$\bar{b}$	1.2924	1.6876	1.0778
σ_b	0.0414	0.0695	0.0278
δb	0.0052	0.0087	0.0043
$\delta b/b$	0.40%	0.51%	0.40%
$q(\bar{a}, \bar{b})$	0.4128	0.4020	0.5704
$q(\bar{a}, \bar{b} + \delta b)$	0.4344	0.4049	0.6593
$q(\bar{a}, \bar{b} - \delta b)$	0.3891	0.3982	0.4572
$q(\bar{a} + \delta a, \bar{b})$	0.4008	0.3991	0.5224
$q(\bar{a} - \delta a, \bar{b})$	0.4246	0.4046	0.6166
δq	0.0256	0.0043	0.1115
$\delta q/q$	6.2%	1.1%	20%
q (known)	0.399(10)	0.399(10)	0.399(10)
$k(\bar{a}, \bar{b})$	1.0510	4.0685	0.3388
$k(\bar{a}, \bar{b} + \delta b)$	1.1449	4.3988	0.6017
$k(\bar{a}, \bar{b} - \delta b)$	0.9762	3.7677	0.2365
$k(\bar{a} + \delta a, \bar{b})$	1.0215	3.9308	0.2895
$k(\bar{a} - \delta a, \bar{b})$	1.0852	4.2154	0.4184
δk	0.0902	0.3462	0.1936
$\delta k/k$	8.6%	8.5%	57%
k (known)	1.05(3)	4.21(11)	0.263(7)

5.1.4 Discussion of fits and direct solution

The results using direct numerical solution and their calculated uncertainties (Table 5.5) verify those obtained from fits weighted with data standard deviations (Table 4.1). Direct numerical solution is more reliable than non-linear least squares fitting in avoiding fitting complications such as false local minima, but is only applicable in simpler cases such as mixture analysis. Reducing the number of simultaneously fitted parameters by defining a and b reduces the uncertainties and likelihood of fitting failure. It also reduces the effects of systematic variations due to triplet blinking. Care must be taken if parameter reduction is done by fixing $N^{(\mathrm{rel})}$ because of high sensitivity of results to uncertainty in $N^{(\mathrm{rel})}$.

Comparison with the results with theoretical data uncertainties shows the theoretical relations can provide reasonable estimates when shot noise is dominant, but they fail to account for residual triplet blinking and other systematic variations. Looking at longer lag times ($> 100\,\mu\mathrm{s}$) on the graphs, our current uncertainty theory does not account for variations in those time scales. Use of theoretical relations must therefore be with these caveats in mind.

As noted in Chapter 2, systematic errors exist due to finite bin size effects. However, the good overlap of the known and measured uncertainty intervals in our experiments suggest that they are small compared to stochastic errors and other systematic variations such as triplet blinking which increase data standard deviations.

5.1.5 General study of uncertainties in parameter space

Here we present additional Figures referenced in Chapter 4.

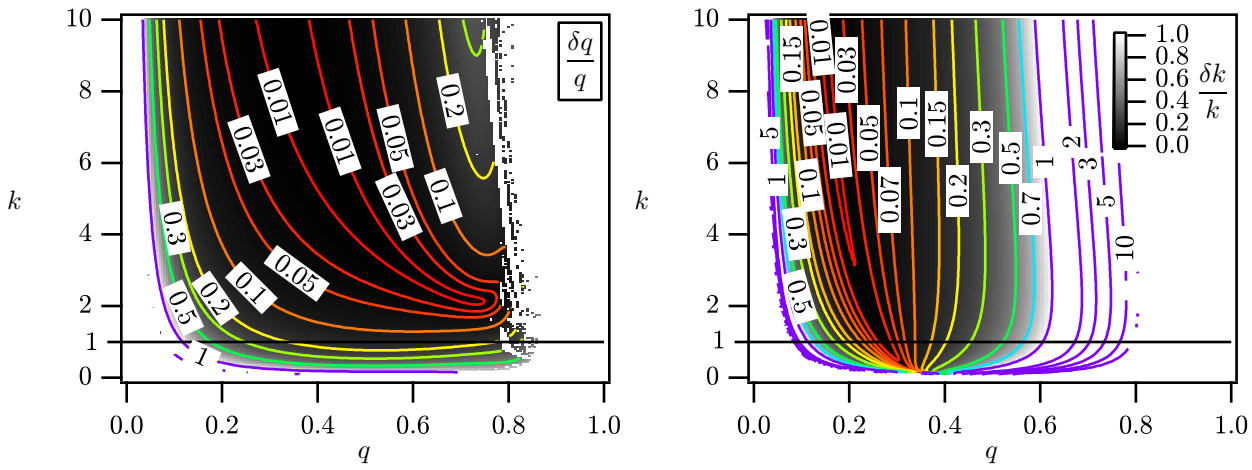

Figure 5.1: Relative uncertainties in q and k assuming $\delta a/a = 0.002$ and $\delta b/b = 0.010$ shown as functions of q and k. The image gray scale is between 0 and 1 in all relative uncertainty graphs. Near the edges, where relative uncertainties in k are very large, solutions are hard to reach and some points are left out from both graphs.

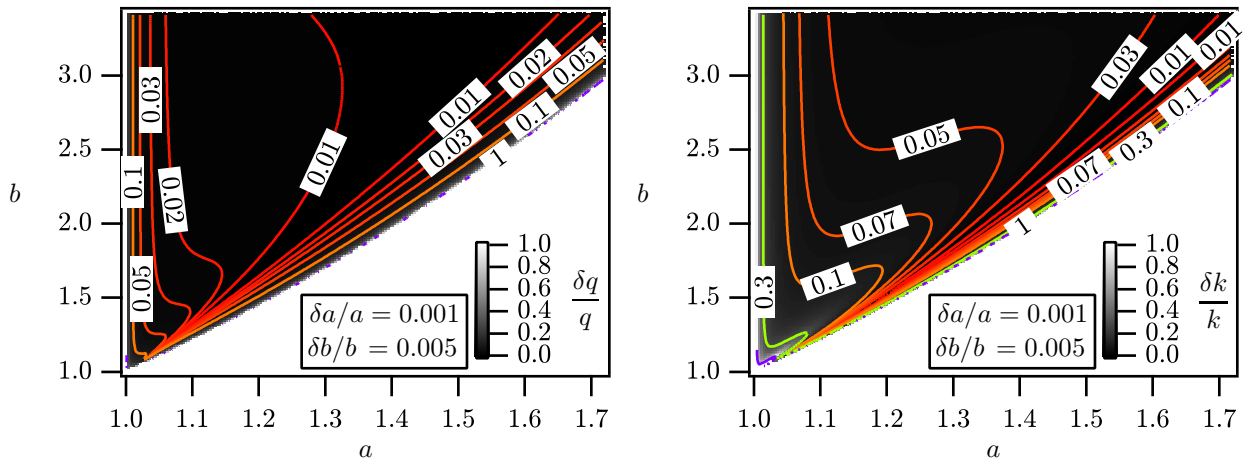

Figure 5.2: Relative uncertainties in q and k assuming $\delta a/a = 0.001$ and $\delta b/b = 0.005$ shown as functions of a and b. No physical solutions exist in the lower right region.

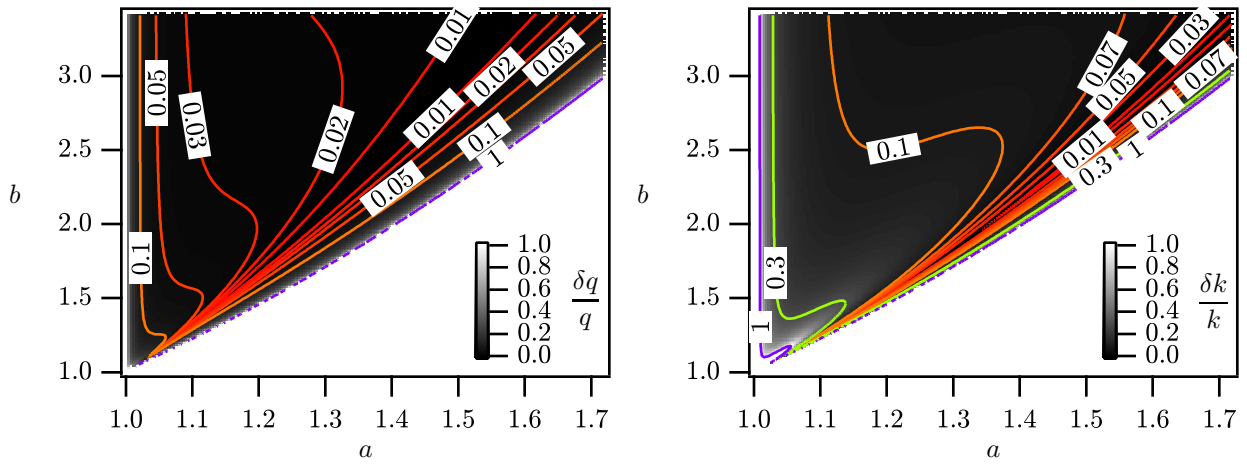

Figure 5.3: Relative uncertainties in q and k assuming $\delta a/a = 0.002$ and $\delta b/b = 0.010$ shown as functions of a and b. No physical solutions exist in the lower right region.

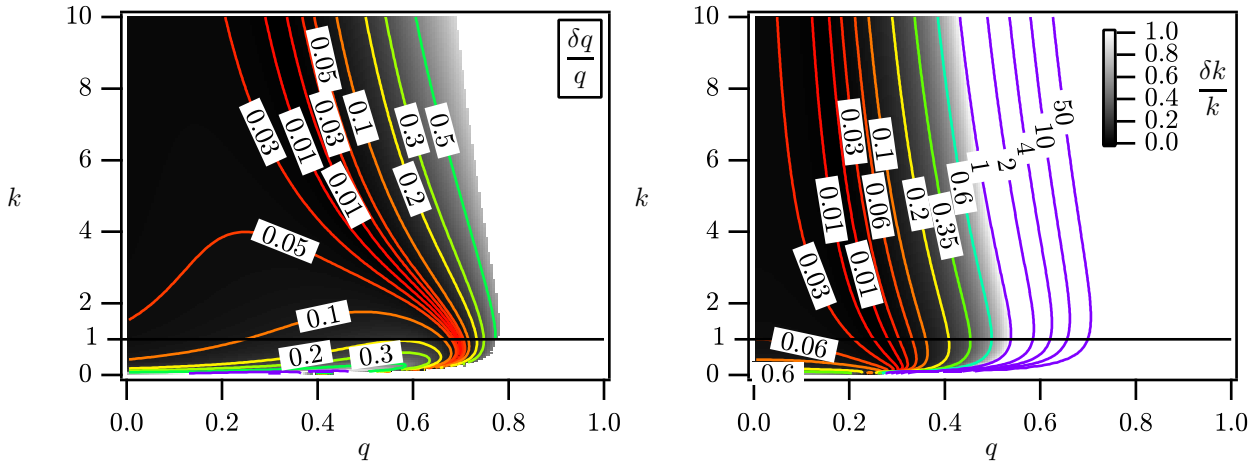

Figure 5.4: Relative uncertainties in q and k assuming $\delta N^{(\mathrm{rel})}/N^{(\mathrm{rel})} = 0.03$ shown as functions of q and k. Similar to the analysis using a and b, relative uncertainties grow rapidly when q approaches 1. However, the restriction is removed on q approaching 0.

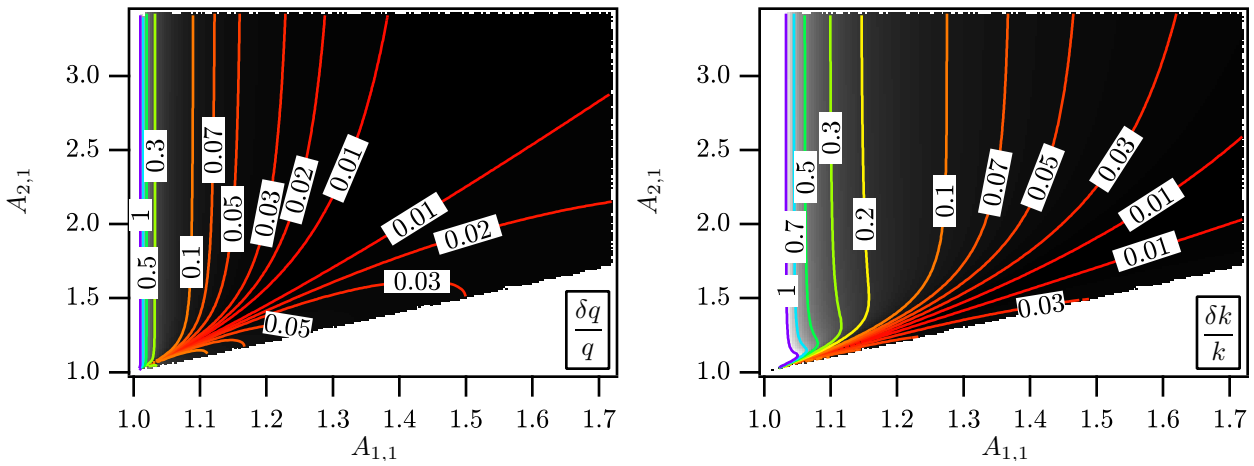

Figure 5.5: Relative uncertainties in q and k assuming $\delta N^{(\mathrm{rel})}/N^{(\mathrm{rel})} = 0.01$ shown as functions of $A_{1,1}$ and $A_{2,1}$. No physical solutions were found in the lower right region.

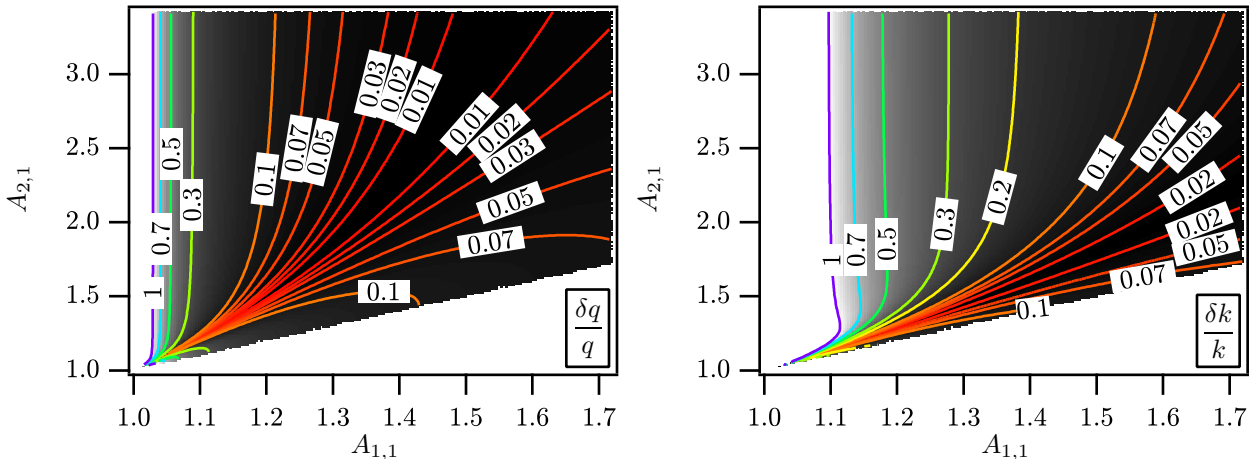

Figure 5.6: Relative uncertainties in q and k assuming $\delta N^{(\mathrm{rel})}/N^{(\mathrm{rel})} = 0.03$ shown as functions of $A_{1,1}$ and $A_{2,1}$. No physical solutions were found in the lower right region.

5.2 Other verifications

5.2.1 Methods

For comparison of fluorophores in mixture experiments, the following additional labeled oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and diluted to $\sim 1\,\mathrm{nM}$ concentration in $0.25\,\mathrm{X}$ TE buffer PH 8.0:

- ALEXA-oligo: Sequence $5'$-$(\mathrm{T})_{18}$-$3'$, Modification: $5'$ Alexa Fluor 532 (NHS Ester)
- ATTO-oligo: Sequence $5'$-$(\mathrm{T})_{18}$-$3'$, Modification: $5'$ ATTO 532 (NHS Ester)

5.2.2 Comparison of fluorophores

The brightness of the fluorescent molecules has significant impact on improving the signal to noise ratio. Oligonucleotides labeled with other bright fluorophores ATTO532 ($Q/Q_{\mathrm{R6G}} = 1.13$) and Alexa532 ($Q/Q_{\mathrm{R6G}} = 0.78$) were also tested at $19\,\mu\mathrm{W}$ excitation power. Higher order correlation functions for those pure samples and pure R6G-oligo were calculated and normalized by those of pure TAMRA-oligo. TAMRA was assumed to have negligible triplet blinking at this excitation power. The results are shown in Figure 5.7, left. Blinking effects that extended to long lag times were observed in ATTO- and Alexa-labeled samples, most clearly seen in $R^{(\mathrm{rel})}_{1,1}$. Only in the case of R6G the triplet blinking decayed out quickly. The hypothesis that such extended variations in the case of ATTO and Alexa normalized by TAMRA were due to slightly different DNA sequences (paired stem in TAMRA-oligo vs non-hairpin ATTO- and Alexa-oligo), and therefore different diffusion times, was rejected by normalizing ATTO curves by Alexa curves, which have identical sequences. The resulting variations were even more intense in this case (Figure 5.7, top right).

These effects are not limited to higher orders and can be seen in order $(1, 1)$ curves as well. We conclude that the choice of fluorophore must be with caution in all fluorescence correlation studies, including higher order correlation analysis.

5.2.3 Setup stability

The idea of normalization by a reference sample to remove diffusion and beam profile effects relies on the assumption that the profile remains unchanged among different measurements. To test this assumption, samples of pure ATTO-oligo and Alexa-oligo were prepared and tested on two consecutive days, and the newer correlation curves were normalized by the older ones in each case. The results are shown in Figure 5.7, right middle and right bottom. The resultant relative curves

are very close to constant lines, most clearly seen in curves of order $(1, 1)$. The concentrations were not intended to be identical at the times of preparation.

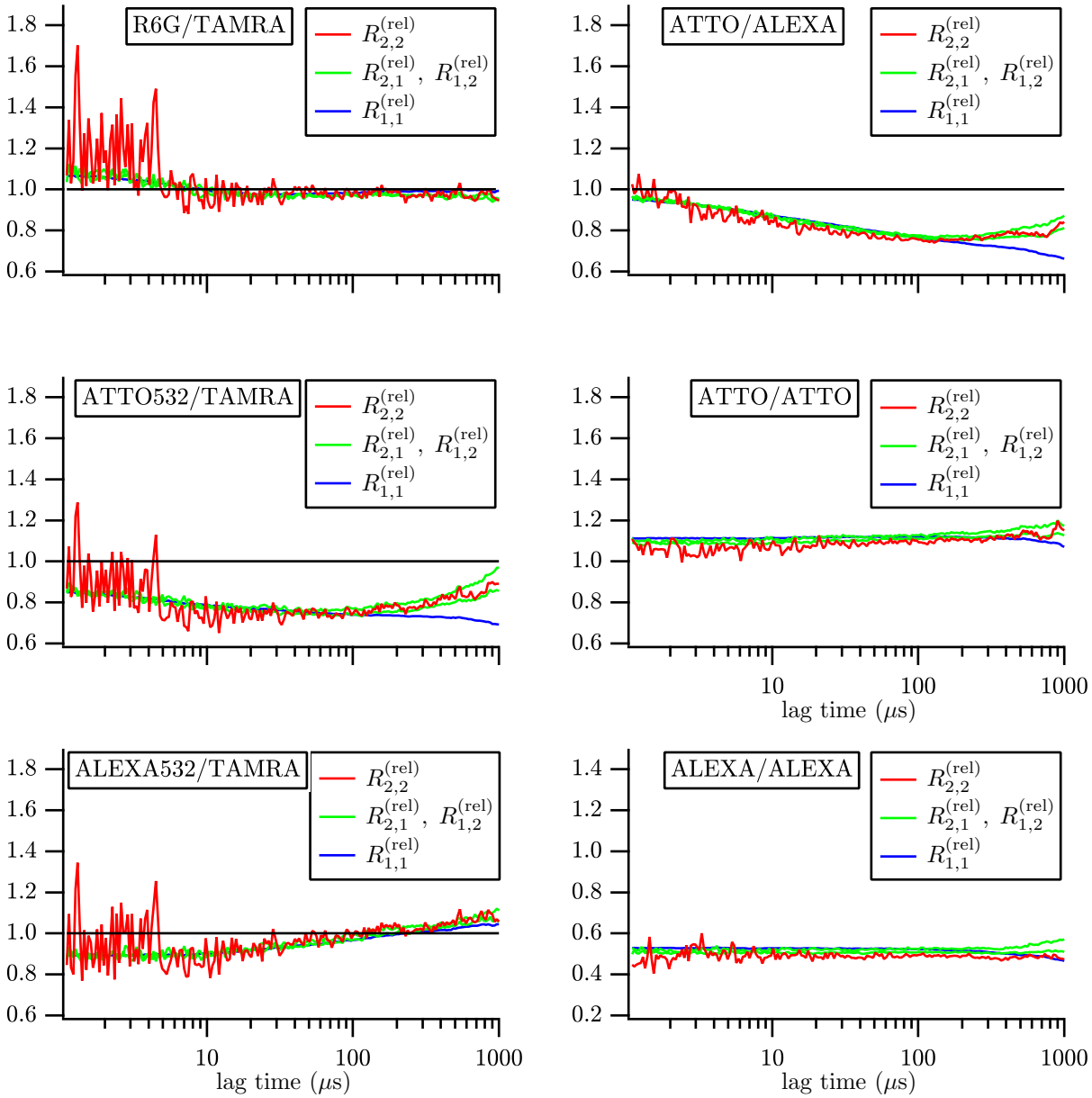

Figure 5.7: Left: Relative correlation curves for samples of pure R6G-, ATTO-, and Alexa-oligo, with TAMRA-oligo used as reference. Top right: Those of ATTO-oligo with Alexa-oligo as reference. Right middle and right bottom: Setup stability verified by looking at relative correlation curves for pure dye-oligo with reference being samples of similar dye-oligo prepared and tested a day earlier.

5.3 MDF independence test

For FITC protonation reaction, we verify the assumption that the diffusion is fully independent from the reaction, hence normalization by a reference sample completely removes the dependence on the molecular detection function (MDF). We perform the test by severely distorting the excitation beam profile and observing how the correlation functions and the relative reaction functions are affected. Figure 5.8 shows the full, undistorted laser beam profile and the same laser beam severely distorted (clipped) by an obstacle (razor edge). Figure 5.9 shows the resulting correlation functions and the relative reaction functions obtained using the full and clipped beam profiles. The analysis is shown up to third order only, because clipping a large portion of the excitation beam significantly reduces the SNR. As expected, the correlation functions are strongly affected by beam geometry modification, while the corresponding relative reaction functions are robustly intact (except in noise).

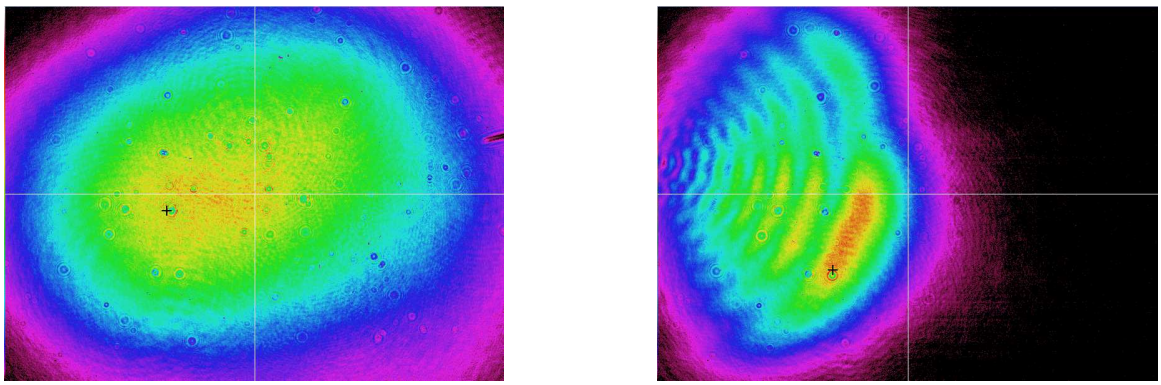

Figure 5.8: Left: Full laser beam profile. Right: Clipped (distorted) laser beam profile

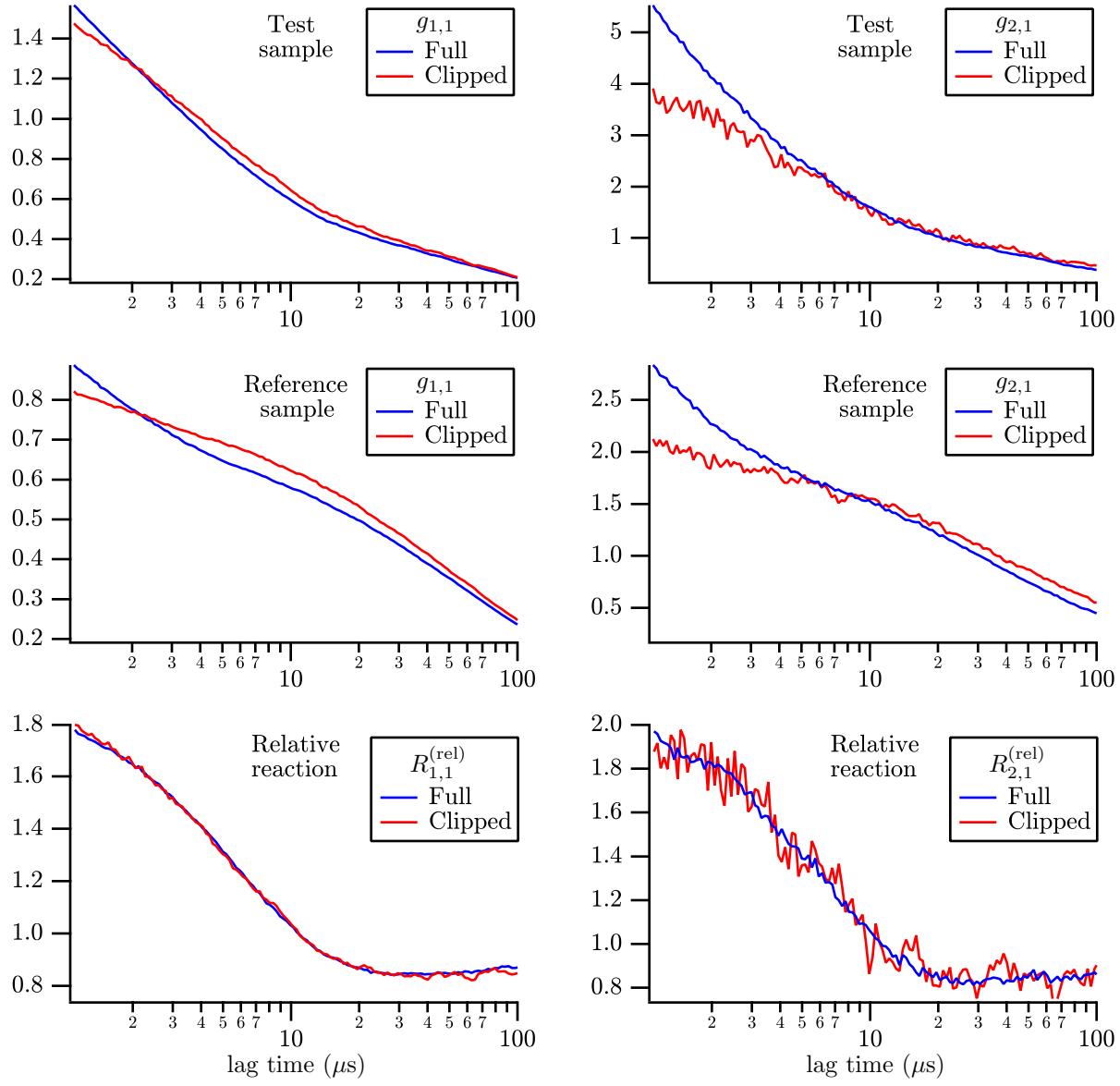

Figure 5.9: Relative reaction functions are fully independent of the MDF (bottom panels), while correlation functions are strongly affected by distortion of the beam profile (top and middle panels). The effect is larger on higher-order correlation functions. Data is shown for FITC in phosphate buffer at pH 6.4 as the test sample and FITC at pH 9.0 as the reference sample. The blue and red curves correspond to the full and clipped beam profiles shown in Figure 5.8, respectively.

5.4 Fast and slow photoinduced blinking of FITC

In this section, we first present replicated measurements of FITC protonation reaction at a higher excitation power (62 μW) than in measurements reported in Chapter 4 (25 μW). Figure 5.10

shows the second- and higher-order correlation curves and their corresponding relative reaction curves. The results of fitting to the relative reaction curves and the calculated reaction parameters are shown in Table 5.7. The measured reaction parameters agree very well with those at a lower excitation power (Table 4.3 in Chapter 4). The difference in the $N^{(\mathrm{rel})}$ values is caused by a slow photoinduced blinking reaction which will be discussed shortly.

Like other common fluorophores, FITC exhibits a fast photoinduced triplet blinking at time scales $< 10\,\mu$s. When no chemically-induced blinking is present, such as in a reference sample, the photoinduced blinking is still seen as a small rise in the second- and higher-order correlation curves, $g_{m,n}$, at short lag times (Figure 5.10, left column, solid lines) similar to a typical reaction. The photoinduced nature of this reaction can be verified by its dependence on the excitation intensity (compare Figure 5.10 with Figure 4.7). The photoinduced reactions can be effectively diminished by lowering the excitation power. However, this results in less signal-to-noise ratio (SNR) in particular in higher order correlations. Luckily, the fast triplet blinking reaction is effectively independent of the chemical protonation reaction and is efficiently removed by calculation of the relative reaction curves, $R^{(\mathrm{rel})}_{m,n}$, given the reference and the test sample are measured at the same excitation power. The first immediate verification of the successful removal of this fast triplet blinking is the perfect fitting of a single-exponential model to the relative reaction curves (Figures 5.10 and 4.7, right column) and the successful measurement of the reaction parameters. The agreement between the measured protonation reaction parameters (in particular k and t_{R}) at low and high excitation powers is another verification of the effective removal of the fast blinking effects. (c.f. Tables 5.7 and 4.3. The small difference in q could be because the bright and dim state brightness values do not both scale proportionally with the excitation intensity.) As a last verification, we consider mismatched normalization of the correlation curves measured at the higher excitation power by those of the reference sample measured at the lower excitation power. The resulting reaction curves are shown in Figure 5.11 for pH 6.05. The single-exponential model fails to fit the resulting relative reaction curves because the fast triplet blinking factor is not effec-

tively removed. We conclude that fast triplet blinking effects are present and efficiently removed by proper normalization by the reference sample.

Further examination of the relative reaction curves at longer lag times reveals a slowly increasing baseline after the exponential decay, which is not compatible with the reaction-free assumption about the reference solution. Figure 5.12 shows order-$(1,1)$ relative reaction curves,$R_{11}^{(\mathrm{rel})}$, at two different excitation powers at lag times up to $1000\,\mu$s (for FITC at pH 5.73). Visualization of the higher-order curves has been omitted because of excessive noise in those curves at longer lag times (The noise stems from the decay of the correlation curves to zero as fewer molecules stay in the probe region for longer lag times). At lag times close to $1000\,\mu$s both curves relax to 0.5 which corresponds to the true $N^{(\mathrm{rel})}$ value of 2. At lag times below $10\,\mu$s the protonation reaction is seen as before. However, at intermediate lag times, a negative slow decay is observed. We propose that a slow photoinduced blinking exists in the reference solution but not in the test solution. To verify this proposal, we fit each curve with a function of the form

$$R_{1,1}^{(\mathrm{rel})} = \frac{1}{N^{(\mathrm{rel})}} \left(\frac{1 + B_{\mathrm{f}} \exp(-t/t_{\mathrm{R_f}})}{1 + B_{\mathrm{s}} \exp(-t/t_{\mathrm{R_s}})} \right) \tag{5.3}$$

in which B_{f} and $t_{\mathrm{R_f}}$ are the fast reaction amplitude and relaxation time, respectively, and B_{s} and $t_{\mathrm{R_s}}$ are those of the slow reaction. The resulting fit parameters are shown in Table 5.8. The amplitude of the slow reaction, B_{s}, is strongly dependent on the excitation power, showing its photoinduced nature. We have verified that the slow blinking exists in reference solutions ranging from pH 8.4 to 10.3 almost invariably (data not shown). However, this blinking is absent in test solutions at pH values close to or below 6. Since the slow blinking is not independent of the protonation reaction, it cannot be removed by normalization by the reference sample. Luckily, since the timescales of the two reactions are far apart, the analysis of the protonation reaction is not significantly affected by the slow blinking. At lag times much smaller than the slow reaction, the denominator in Equation (5.3) is almost constant, explaining the inaccurate, yet inconsequential measurements of $N^{(\mathrm{rel})}$ when the slow reaction is ignored. At intermediate lag times, the deviation of higher-order relative reaction curves from order $(1,1)$ due to the slow reaction is small, because the reaction

amplitude is small and the dark-state has very low brightness ($q \approx 0$). Therefore no modification of higher-order relations is required when ignoring the slow reaction (a single $N^{(\mathrm{rel})}$ can be used).

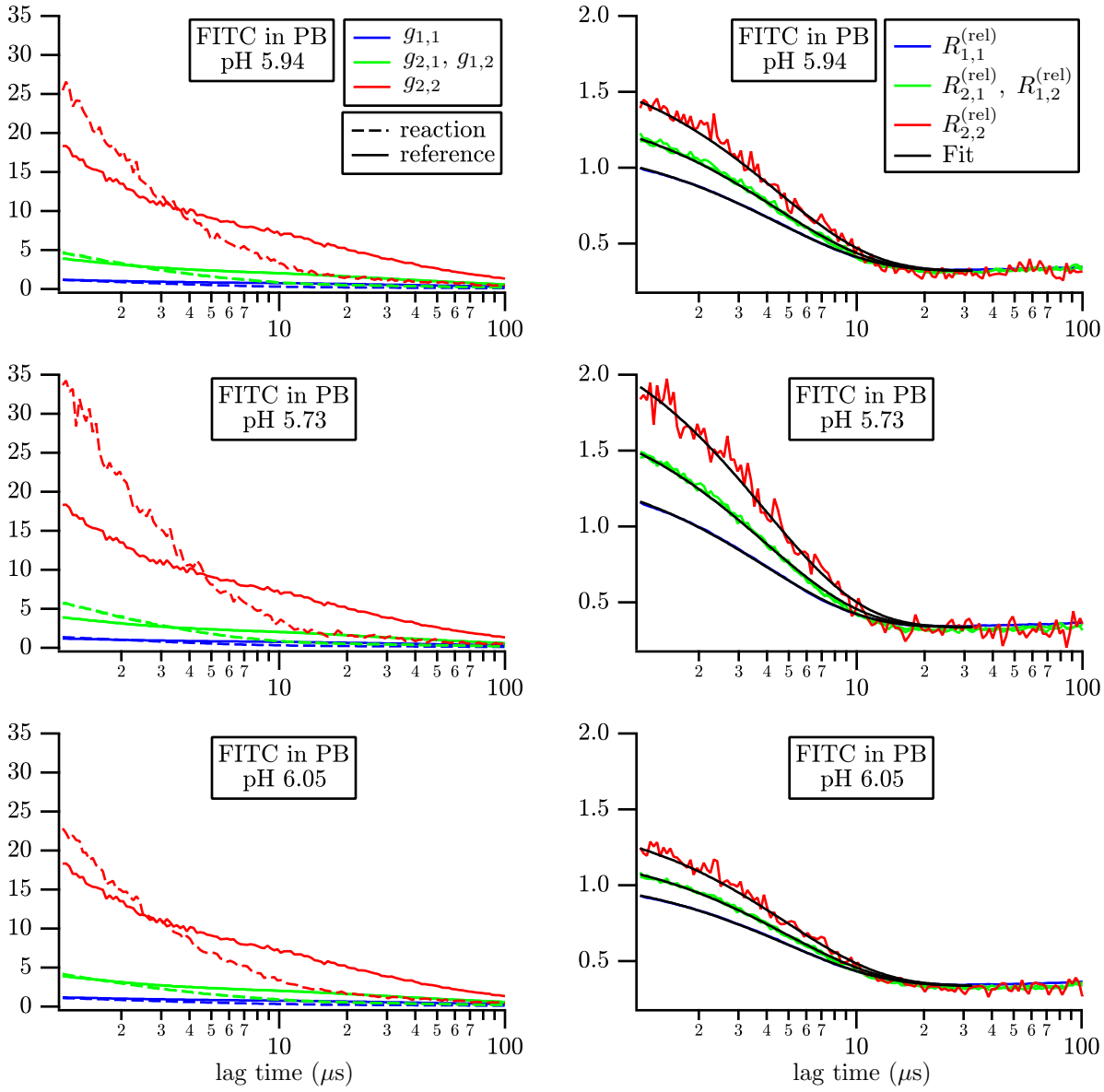

Figure 5.10: Left column: Second- and higher-order correlation functions obtained for FITC in phosphate buffer (PB) at different pH values (dashed lines). FITC at pH 9.03 was used as reference in all cases (solid lines). The same color scheme and axis scale applies to all graphs in each column. Right column: The corresponding relative reaction functions obtained from the higher-order correlation functions on the left. The global fit lines to the first $N_\mathrm{p} = 100$ data points (each curve) are shown in black. Data was collected with 62 μW excitation power and two detectors.

Table 5.7: Protonation reaction of FITC in phosphate buffer is resolved by fitting to the relative reaction curves, $R_{1,1}^{(\mathrm{rel})}$, $R_{2,1}^{(\mathrm{rel})}$ and $R_{2,2}^{(\mathrm{rel})}$. Where two numbers appear inside parentheses, the first number shows the standard error of mean over six trials, and the second number shows the average reported fit uncertainty in one trial. Data is shown for 62 μW excitation power and two detector.

pH	**5.94(3)**	**5.73(3)**	**6.05(3)**
q	0.0514(3)(4)	0.0523(4)(3)	0.0500(12)(6)
k	4.23(12)(2)	6.70(22)(4)	3.42(4)(2)
$N^{(\mathrm{rel})}$	3.211(42)(6)	3.035(25)(6)	2.993(7)(7)
t_{R} (μs)	4.606(46)(15)	3.862(69)(11)	4.931(31)(21)
t_{R0} (μs)	11.27(18)	7.82(14)	13.68(41)
k_- ($10^4\,\mathrm{s}^{-1}$)	1.70(5)	1.66(6)	1.66(5)
$[\mathrm{H}^+]k_+$ ($10^4\,\mathrm{s}^{-1}$)	7.17(12)	11.1(2)	5.66(17)
k_+ ($10^{10}\,\mathrm{s}^{-1}\mathrm{M}^{-1}$)	6.2(4)	6.0(4)	6.3(5)
$\mathrm{p}K_{\mathrm{a}}$	6.57(3)	6.56(3)	6.58(3)

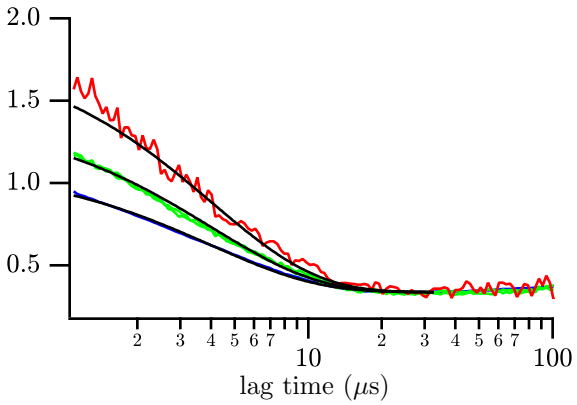

Figure 5.11: Mismatched normalization of the correlation curves obtained at 62 μW excitation power by reference curves obtained at 25 μW power does not fully remove the fast triplet blinking effect. The color scheme follows that of Figure 5.10, right column. The curves clearly deviate from the single-exponential fits. Data is shown for FITC in phosphate buffer pH 6.05.

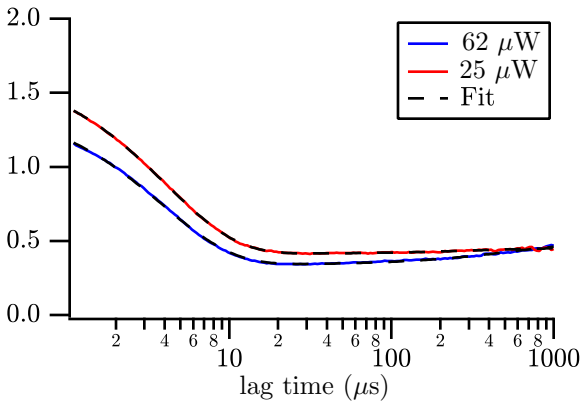

Figure 5.12: Relative reaction curves of order (1,1), $R_{11}^{(\mathrm{rel})}$, are shown at two different excitation powers over extended lag times. The curve at higher excitation power shows stronger slow blinking in the reference sample. Data is shown for FITC in phosphate buffer pH .5.73.

Table 5.8: The results of fitting to the two curves in Figure 5.12 using Equation (5.3).

	62 μW	**25 μW**
$N^{(\mathrm{rel})}$	2.08(2)	2.21(1)
B_{f}	3.24(1)	3.06(1)
$t_{\mathrm{R_f}}$ (μs)	3.90(1)	4.13(1)
B_{s}	0.42(1)	0.098(4)
$t_{\mathrm{R_s}}$ (μs)	432(20)	332(33)

5.5 Two-state transition factors

Consider a fluorescent particle alternating between two states:

$$\text{state 1} \underset{k_{\mathrm{b}}}{\overset{k_{\mathrm{f}}}{\rightleftharpoons}} \text{state 2}$$

where "state 1" is usually the brighter (unfolded) state, "state 2" is usually the darker (folded) state, and k_{f} and k_{b} are the forward and backward (reverse) rates respectively. Denote the probability

that the particle is found in state 1 at time time t with $P_1(t)$, and similarly for state 2. The following equations describe the reaction:

$$\frac{\mathrm{d}P_1(t)}{\mathrm{d}t} = -k_\mathrm{f} P_1(t) + k_\mathrm{b} P_2(t)$$
$$P_2(t) = 1 - P_1(t)$$

The solutions are

$$\begin{aligned} P_1(t) &= \frac{1 + [(1+k)P_1(0) - 1]e^{-t/t_\mathrm{R}}}{1+k} \\ P_2(t) &= \frac{k - [k - (1+k)P_2(0)]e^{-t/t_\mathrm{R}}}{1+k} \end{aligned} \tag{5.4}$$

where $P_1(0)$ and $P_2(0)$ are the initial probabilities at time 0. We have defined

$$k = \frac{k_\mathrm{f}}{k_\mathrm{b}} = \frac{N_2}{N_1}$$

with N_i being the number of molecules in state i in the ensemble, and

$$t_\mathrm{R} = (k_\mathrm{f} + k_\mathrm{b})^{-1}$$

as the overall reaction time constant. Also, defining

$$P(i) = P_i(\infty) = \frac{N_i}{N_1 + N_2}$$

as the probability of finding the particle in state i independent of initial conditions, we have

$$k = \frac{P(2)}{P(1)}$$

The transition factor $Z_{s_2,s_1}(t)$, denoting the probability that the particle is found in state s_2 at time t given it was in state s_1 at time 0, can be found by setting the initial probabilities in (5.4)

equal to 1 or 0:

$$
\begin{aligned}
Z_{1,1}(t) &= \frac{1 + ke^{-t/t_{\mathrm{R}}}}{1+k} \\
Z_{1,2}(t) &= \frac{1 - e^{-t/t_{\mathrm{R}}}}{1+k} \\
Z_{2,1}(t) &= \frac{k(1 - e^{-t/t_{\mathrm{R}}})}{1+k} \\
Z_{2,2}(t) &= \frac{k + e^{-t/t_{\mathrm{R}}}}{1+k}
\end{aligned}
$$

Chapter 6

Conformational dynamics of DNA hairpins

6.1 Introduction

Nucleic acid hairpins are stem-loop structures that occur naturally in DNA and RNA molecules, and play key roles in biological functions [19] such as DNA replication [79], transcription [80], and genetic recombination [81]. Individual hairpin-forming segments of single-stranded DNA and RNA are also commonly synthesized in the lab to serve as reporter molecules for the presence of specific nucleic acids [36], or used as model systems. A large body of work on DNA hairpin dynamics [12,35,37,39,40,82–114] has contributed to the understanding of nucleotide interactions and their structural and energetic implications in both single-stranded (the loop part of a hairpin) and double-stranded (the stem part) DNA. However, many discrepancies and questions remain.

Experimental studies employing various techniques have suggested a wide variety of characteristic time-scales for the folding-unfolding dynamics of comparable hairpin structures. Techniques most sensitive to sub-millisecond time scales, such as fluorescence correlation spectroscopy (FCS) [35, 82–87] and temperature-jump spectroscopy (T-jump) [88–95], have found the time scales to be a few tens to a few hundreds of microseconds. In contrast, experiments based on stopped flow [96, 97] have measured a few millisecond time scale, while surface immobilization methods [98–100], show dynamical phenomena over tens to hundreds of milliseconds. Finally, optical trapping measurements on single DNA hairpins of 6-bp stem and longer reveal unloaded unfolding rates on the order of 0.1 seconds and longer [12, 101–103], which compare favorably to those observed by FCS for shorter stems by extrapolation. [39]

Few techniques can give insights into DNA hairpin folding dynamics over scales from microseconds to seconds. One such method is diffusion-decelerated FCS, which has directly shown

two or more relaxation times ranging from $\sim 10\,\mu$s to $\sim 100\,$ms [40] and provided direct evidence that DNA hairpin folding and unfolding is a multi-step process. However, given the limitations of conventional FCS, which uses the second-order autocorrelation function alone (herein referred to as second-order FCS), little was revealed about the nature of the underlying conformation states. This reflects a general weakness of second-order FCS, which is unable to independently determine the underlying parameters of a folding-unfolding reaction, including species concentrations, brightnesses, and kinetic rate constants, even when the entire reaction time scale is observed.

In many studies on DNA hairpins [35, 82, 93, 94, 98–100] the systems seemed to be described well by activated dynamics between two states, described as folded and unfolded conformations, whereas other studies have suggested more complicated free-energy landscapes. [37, 40, 87–91, 104–112] A second-order FCS study in our lab by Jung et al [37] demonstrated, by a somewhat complicated argument, that no two-state mechanism could account for the observed data. Rather, a three-state model consisting of bright, intermediate, and dark states was proposed that successfully described the observations. The fluctuations observable on the FCS time scale were attributed to transitions between the bright and intermediate states, while transitions to the dark state were thought to occur on longer time scales. However, other models with three or more states could not be ruled out, because of the limitations of the second-order FCS.

The inadequacy of two-state models to describe DNA hairpin dynamics is evident by a simpler line of argument when the reaction components of correlation functions, the so-called *reaction functions*, are considered side-by-side in a series of second-order temperature or salt dependence FCS measurements. This has apparently been overlooked in almost all the earlier FCS experiments, notably the pioneering temperature dependence studies by Libchaber and co-workers [35, 82]. Even in the salt dependence study by Jung et al. [37] the reaction functions were not extracted and directly compared. In the section that immediately follows, we briefly present this line of argument within second-order FCS, which will motivate and lay some groundwork for the higher-order analysis that follows.

Higher-order FCS refers to the computation and analysis of third- and higher-order correlation functions, along with the second-order function, from the photon stream collected in an FCS experiment. [15, 17, 115, 116] The technique allows the characterization of the reaction parameters [71, 116], including the species populations and brightnesses and rate constants, which is not possible through second-order analysis alone. It can report and characterize the irreversible nature of a reaction if present [17]. Most importantly, it can reveal and characterize the multi-component nature of a mixture [16, 71], or as demonstrated here, the multi-state nature of a reaction — all inaccessible by any single second-order measurement.

We carried out FCS measurements for DNA hairpins of 8 and 21 poly-thymidine loops, with either TAMRA or Rhodamine 6G used as the fluorophore, and either dabcyl or guanine used as the quencher. In all cases, higher-order analysis definitively ruled out the two-state model, and showed that the *only* three-state model consistent with second- and higher-order FCS data has the fast reaction between the intermediate and dimmest states, with the brightest state isolated by slower reaction rates. The analysis ruled out the three-state model previously proposed by Jung et al. [37], in which the fast step of the reaction was assumed to be between the brightest and intermediate states. Using higher-order FCS we directly measured the relative populations and the brightnesses of the three states, as well as the forward and reverse rate constants of the fast reaction. The dependencies of these parameters on temperature and NaCl concentration gave insight into the energy landscape of the conformational dynamics and the hairpin conformations associated with the three observed states.

6.2 Background

A brief discussion of second-order reaction functions will help clarify how the original conclusions by Jung et al. were reached, and serve as useful preparation for the complementary higher-order analysis that follows. When the reaction (in this case, between conformational states) and the spatial diffusion are effectively independent, as will be verified for the hairpin molecules, the correlation function in second-order (also known as order-$(1,1)$) FCS is given by [6, 9, 17]:

$$g_{1,1}(t) = \gamma_{1,1} R_{1,1}(t) Y_{1,1}(t) \tag{6.1}$$

where $R_{1,1}(t)$ is the "reaction function" which depends only on the reaction and population properties of the involved states. The factors $\gamma_{1,1}$ and $Y_{1,1}(t)$ depend only on the spatial illumination and detection profiles, collectively called the molecular detection function (MDF), and the diffusional properties of the molecules. The correlation function from a "reference" sample which has similar diffusional properties but is designed to lack or not exhibit the reaction is also measured using the same experimental setup under identical conditions of temperature and salt concentration. The reference sample is most conveniently the same hairpin structure labeled with a fluorophore but without a quencher. We will later show that even pure fluorophore, with time rescaled to match the hairpin diffusion time, can be used as the reference. The reference correlation function is consequently of the form

$$g_{1,1}^{(\mathrm{ref})}(t) = \frac{1}{N^{(\mathrm{ref})}} \gamma_{1,1} Y_{1,1}(t) \tag{6.2}$$

Here $N^{(\mathrm{ref})}$ is the average number of molecules in the observation volume for the reference sample, and $1/N^{(\mathrm{ref})}$ is the reaction function of the reference sample. By dividing the correlation function from the test sample, Equation (6.1), by that from the reference sample, (6.2), we can extract the "relative reaction function"

$$R_{1,1}^{(\mathrm{rel})}(t) = N^{(\mathrm{ref})} R_{1,1}(t) \tag{6.3}$$

which differs from the reaction function, $R_{1,1}(t)$, only by a scaling factor. This procedure also removes at least partially the effects of photochemical blinking of the fluorophore on the reaction function [71].

For a two-state reaction with N_1 and N_2 denoting the populations of the two states, and first-order kinetics, the relative reaction function has the form

$$R_{1,1}^{(\mathrm{rel})}(t) = \frac{1}{N^{(\mathrm{rel})}} \left(1 + B e^{-t/t_{\mathrm{R}}}\right) \tag{6.4}$$

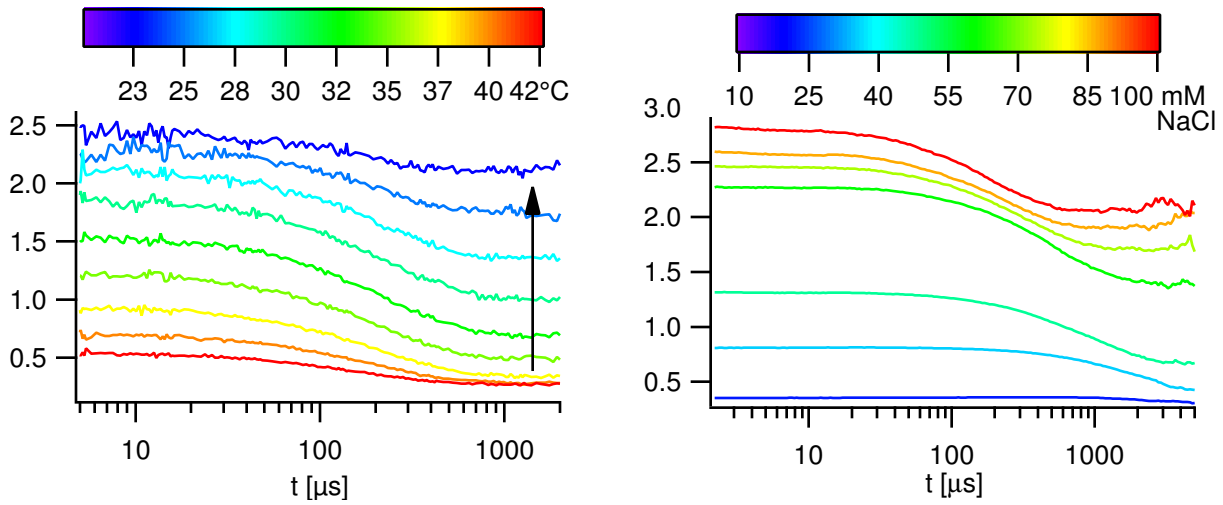

Figure 6.1: Temperature and salt dependence of the relative reaction functions, $R_{1,1}^{(\mathrm{rel})}(t)$, obtained using second-order FCS. Left: Baselines increase as temperature is decreased (arrow) for hairpin TT8D with reference TT8. Right: Baselines increase as [NaCl] is increased for hairpin RT21D with reference RT21. See the Experimental Section for sample descriptions.

in which $N^{(\mathrm{rel})} = (N_1 + N_2)/N^{(\mathrm{ref})}$ is the ratio of the concentration of the test sample to that of the reference sample, B depends on the equilibrium constant and the relative brightness of the states, and the relaxation time, t_{R}, is the inverse of the sum of the forward and reverse rate constants. Note that $N^{(\mathrm{rel})}$ does not depend on temperature or salt conditions. Therefore, when t_{R} is within the timescale accessible by FCS, $R_{1,1}^{(\mathrm{rel})}(t)$ decays to a baseline value of $1/N^{(\mathrm{rel})}$ at larger lag times *regardless* of the temperature or salt concentration. Figure 6.1 shows $R_{1,1}^{(\mathrm{rel})}(t)$ at various temperatures and [NaCl] for two types of DNA hairpins. The baselines vary by nearly an order of magnitude. This is manifestly inconsistent with the two-state model. Many studies of DNA hairpins have considered the temperature or salt dependence of hairpin conformational dynamics by FCS; however, they have mostly analyzed the correlation or reaction curves individually, extracting t_{R} and B values, but not comparing the baselines. The analysis by Jung et al., though conducted on the overall correlation function, $g_{1,1}(t)$, rather than the reaction functions, was precise enough to reveal this discrepancy in a more indirect way.

Looking more closely at Figure 6.1, as the temperature is decreased or as the salt concentration is increased, the baselines corresponding to $1/N^{(\mathrm{rel})}$ increase (upward arrow), suggesting that the

$N^{(\mathrm{rel})}$ value decreases. It appears as though the hairpin molecules are driven into a third, invisible state not previously accounted for by the two-state model. In other words, $N^{(\mathrm{rel})} = (N_1 + N_2 + N_3)/N^{(\mathrm{ref})}$ seems plausible, in which N_3 is the population of a non-fluorescent, dark state. Since the fluctuations to and from this state are not directly observed within the time scale of FCS, this third state must be long-lived compared to the other two states. Jung et al. proposed this model and described the third state as "a sink into which DNA molecules disappear and never return [within the FCS time frame]". We refer to this as the "dark sink" model. We will show below that it is not the only model that can fit the second-order FCS data, and that it is unable to account for higher-order FCS data.

In a two state model, it is reasonable to describe the two states as the fully folded and unfolded conformations. But with three states, it may not be clear what the third state is, or which state is static (long-lived) on the FCS time scale. Second-order FCS alone is insufficient to measure the reaction rates and population parameters even for a two-state model; and if three states are present it does not even allow one to identify which states are implicated in the reaction occurring within the FCS time frame (the fast reaction). Therefore, Jung et al. had to make assumptions on which states are involved in the fast reaction, and assumed a non-detectable brightness for the third state. Here, we also examine three-state models, but make no further assumptions. The states are initially characterized only in terms of their relative brightness: the brightest state, the intermediate state, and the dimmest state. By means of simultaneously fitting the second- and higher-order reaction functions, we determine the relative brightnesses and populations of the states and reveal which pair of states are implicated in the fast reaction. We will then characterize the relative entropy and enthalpy of the states and the role of ionic strength. Only after these kinetic and thermodynamic analyses will we make any claims regarding the conformational character of these states.

6.3 Theory

Consider a system of diffusing fluorescent molecules which undergo reactions (e.g. binding or conformational dynamics) between J states and in which the fluorescence brightness of the ith

state is Q_i. Suppose that there are N such molecules in the probe region, of which N_i are in the ith state, on average. For brevity, in the remainder of this dissertation we always use the term "reaction function" and its symbol $R_{m,n}(t)$ to denote the "relative reaction function" ($R^{(\mathrm{rel})}_{m,n}(t)$). Also N and N_i will always refer to those values divided by $N^{(\mathrm{ref})}$ of the reference sample.

Recall that reaction functions for the considered system have the form [115]

$$R_{m,n}(t) = \frac{(\overleftarrow{1}\vec{K})}{N} \frac{\overleftarrow{1}\mathbf{Q}^n\mathbf{Z}(t)\mathbf{Q}^m\vec{K}}{(\overleftarrow{1}\mathbf{Q}^m\vec{K})(\overleftarrow{1}\mathbf{Q}^n\vec{K})} \tag{6.5}$$

where $\overleftarrow{1}$ is a $1 \times J$ row vector with all elements 1, $\mathbf{Q}$ is a $J \times J$ diagonal matrix with elements $q_i = Q_i/Q_1$ on the main diagonal, and $\vec{K}$ is a $J \times 1$ column vector with elements $K_i = N_i/N_1$. $\mathbf{Z}(t)$ is a $J \times J$ matrix the elements of which, $Z_{i,j}(t)$, denote the probability that a molecule is in state i at time t given that it was in state j at time 0. Thus if $\vec{P}(t)$ is a $J \times 1$ column vector with elements specifying the probability of finding a molecule in state i at time t, we have

$$\vec{P}(t) = \mathbf{Z}(t)\vec{P}(0)$$

The time-dependence of the probabilities satisfies the equation

$$\frac{\mathrm{d}\vec{P}(t)}{\mathrm{d}t} = \mathbf{A}\vec{P}(t) \tag{6.6}$$

where $\mathbf{A}$ is the matrix of reaction rate constants, and the matrix $\mathbf{Z}(t)$ can be expressed as

$$\mathbf{Z}(t) = e^{\mathbf{A}t}$$

The vector of the relative equilibrium populations (equilibrium constants), $\vec{K}$, can be expressed in terms of the rate constants, because $\vec{K}$ is a steady-state solution to Equation (6.6), or in other words, the eigenvector of $\mathbf{A}$ corresponding to the zero eigenvalue. The set of the rate constants together with the brightness ratios $\{q_2, q_3, \ldots, q_J\}$ and N form a complete set of independent parameters to be measured from higher-order FCS analysis.

For comparison, we first consider a two-state model for hairpin folding, as shown in Figure 6.2(a). Panel (b) shows an example of higher-order reaction curves for the two-state system calculated up to fourth order using the parameters shown on the graph. Higher-order reaction functions for two-state systems have a single relaxation time after which they all decay to the same baseline value. On the graph, we also show the timescale typically accessible by (higher-order) FCS measurements, limited by the diffusion time of the molecules in the probe region.

Next we examine a three-state model, as shown in Figure 6.2(c), which in general can include six rate constants. Panel (d) shows an example of the reaction functions for two pairs of rate constants, a fast pair and a slow pair, as indicated on the graph. The resulting curves exhibit two distinct relaxation times, fast and slow. In general, the reaction functions may exhibit more complex behavior. However, based on the symmetry of higher-order reaction functions observed in the experimental data which will be presented, $R_{2,1}(t) = R_{1,2}(t)$, we conclude that the reaction is reversible and detailed balance holds. As a result, $\mathbf{A}$ must have two real eigenvalues (other than 0) which correspond to the two relaxation times. Under these conditions, if one relaxation time is much longer than the other, there can be only one pair of fast rate constants, together with two pairs of slow rate constants. In other words, one state is isolated by slow rate constants, and the two other states are connected by a fast pair. See Section 7.1 for details.

We therefore limit our attention to models with one fast reaction pair only. Denoting the fast rate constants a and b**,** defined always between states labeled as 1 and 2, and ignoring the other rate constants in comparison, $\mathbf{Z}(t)$ can be approximated as

$$\mathbf{Z}(t) \approx \frac{1}{1+K_2} \begin{bmatrix} 1+K_2 e^{-t/t_\mathrm{F}} & 1-e^{-t/t_\mathrm{F}} & 0 \\ K_2(1-e^{-t/t_\mathrm{F}}) & K_2+e^{-t/t_\mathrm{F}} & 0 \\ 0 & 0 & 1+K_2 \end{bmatrix} \tag{6.7}$$

where $t_\mathrm{F} = (a+b)^{-1}$ is the fast relaxation time, and $K_2 = a/b$ (detailed balance) has been used. Equation (6.7) describes a reaction between states 1 and 2 in the presence of a stationary population in state 3. The higher-order reaction functions decay to separate baseline values, as

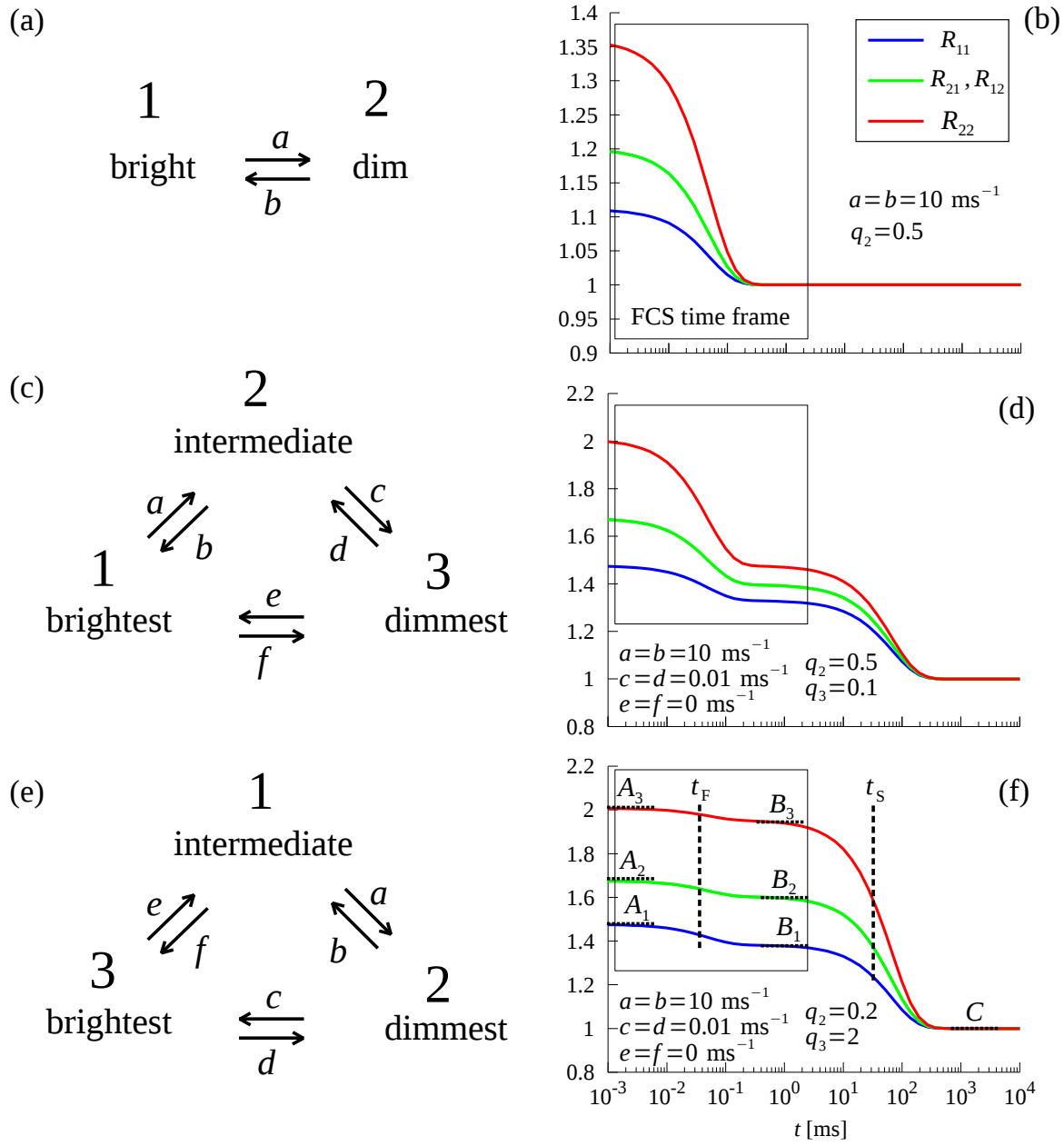

Figure 6.2: Two-state and three-state models (left) and higher-order reaction functions calculated for $N = 1$ and example rate constants and brightness values (right). (a, b): A two-state model yields a single relaxation time and all the higher order curves decay to the same baseline value $(1/N)$. (c): A three state model with the fast rate constants, a and b, between the brightest and intermediate states. (e): A three state model the fast rate constants, a and b, between the intermediate and dimmest states. The states are numbered accordingly. (d, f): Example plots corresponding to the models on the left, calculated with the parameters shown on each graph. The dashed lines in (f) indicate significant measurable parameters from the graph.

seen for example within the FCS time frame in Figure 6.2(d), making the case of three (or more) states clearly distinguishable from a two-state scenario (unless $q_3 = 0$, that is, the stationary state is fully dark), despite the fact that the entire reaction functions are not accessible. Notice that although we assume the slow rate constants are negligible compared to the fast rate constants, their ratios still determine the population of state 3 which is measurable: the approximate $\mathbf{Z}(t)$ in Equation (6.7) does not depend on K_3, but the parameter still exists in Equation (6.5).

There are three qualitatively distinct classes of three-state models differentiated by which species is effectively static. For example, Figure 6.2(c) presents the class in which state 1 is the brightest state and undergoes fast transformations to and from the intermediate state 2; the dimmest state 3 is static on FCS timescales. The model considered by Jung et al. is of this class. Figure 6.2(e) presents the class in which state 1 is the intermediate state and undergoes fast transformations to and from the dimmest state 2; the brightest state 3 is static on FCS timescales. Figure 6.2(d) and (f) present reaction curves for specific examples of those classes. Notice how the associated reaction functions are readily distinguished. In Figure 6.2(d), curves get closer over the course of the fast-reaction timescale, while in Figure 6.2(f) the curves move apart. For discussion relating to the third class of models see Section 7.1.2 and Figure 7.1. In our analysis of the data presented below, we will make no assumptions about the appropriate class of models, rather the relative brightnesses will reveal which one is appropriate. If through fitting we obtained $0 \leq q_3 \leq q_2 \leq 1$ we would infer that the fast reaction is between the brightest and intermediate states. If we obtained $0 \leq q_2 \leq 1$ and $q_3 \geq 1$ we would conclude that the fast transformations are between the intermediate and the dimmest states.

To conclude this section, we examine the questions of how many higher-order correlation curves are needed to determine a full set of reaction parameters, and how many (or which) reaction parameters we can measure with a limited FCS time scale. In general, the set of 6 rate constants, together with q_2, q_3, and N, leave us with 9 parameters to measure. For now, assume that detailed balance holds and the two relaxation times are largely distinct, as shown in Figure 6.2(f). On that graph, we can identify some significant independent parameters directly measurable from the re-

action functions. The first two are the fast and the slow relaxation times, t_{F} and t_{S}, respectively. In addition to those are the amplitude parameters (A_1, A_2, and A_3) corresponding to the limit $t \to 0$. The baselines (B_1, B_2, and B_3) correspond to $t_{\mathrm{F}} \ll t \ll t_{\mathrm{S}}$, and the overall baseline $C = 1/N$ corresponds to $t \to \infty$. If both timescales happen to be measurable by higher-order FCS (for example using diffusion deceleration) then correlations up to third order will provide 7 independent measurables (A_1, A_2, B_1, B_2, C, t_{F}, t_{S}), and detailed balance yields 2 constraining equations (Section 7.1.2), thereby constraining a unique solution in the 9 dimensional parameter space. If the FCS timescale restricts access to t_{S} and C, then two parameters, A_3 and B_3, can be obtained from the fourth order reaction function to still fully determine the system (using the exact form of $\mathbf{Z}(t)$). In practice, however, the correction to the zeroth-order approximation in Equation (6.7) might be small compared to the stochastic and systematic uncertainties in the measured reaction curves, preventing a reliable measurement of all the parameters when FCS timescale is limited. Therefore, in this work, we only use the zeroth-order approximation, Equation (6.7), to fit the experimental data. This prevents direct measurement of the slow rate constants, but we still measure six independent parameters (K_2, K_3, t_{F}, q_2, q_3, N) (alternatively, a, b, K_3, q_2, q_3, N) instead of nine. The main systematic sources of error considered for this decision stem from the simplistic assumption of a three-state model for the more complex hairpin folding dynamics. The main stochastic source of error is the shot noise in higher-order correlations.

The number of required correlation curves may also be pre-evaluated for a more general case where detailed balance does not hold. In that case, the distinction of $R_{2,1}(t)$ and $R_{1,2}(t)$ provides two new measurables (an amplitude and a baseline, assuming distinct relaxation times) which can replace the detailed balance constraints. However, instead of two real relaxation times, we generally have the two eigenvalues of $\mathbf{A}$ to measure, which might be complex numbers. These two are in principle measurable if a full reaction function is available. In a limited timescale, the two may or may not be measurable.

6.4 Experimental Section

6.4.1 Temperature dependence measurements

Labeled and dual HPLC purified oligonucleotides listed in Table 6.1 were purchased from AnaSpec (Fremont, CA) and Integrated DNA Technologies (Coralville, IA). Except for TT21G which has the sequence $5'$-d$(\mathrm{AACCC(T)}_{21}\mathrm{GGGTTG})$-$3'$, all other samples are of sequence $5'$-d$(\mathrm{AACCC(T)}_{x}\mathrm{GGGTT})$-$3'$ in which x is the length of the thymidine loop, either 8 or 21, as listed in Table 6.1. The samples were labeled at the $5'$ end with either rhodamine 6G (R6G) or carboxytetramethylrhodamine (TAMRA). At the $3'$ end, they carried either dabcyl, deoxyguanosine, or, in the case of reference samples, no extra attachment. Reference samples were diluted to $\sim 1\,\mathrm{nM}$ concentration in $0.25\,\mathrm{X}$ TE buffer pH 8.0. The dabcyl-quenched test samples were prepared at 5–10 times higher oligo concentrations in the same buffer solution. The guanine-quenched test sample (TT21G) was prepared at the same concentration as that of its reference (TT21) due to weaker quenching. Sample TT8D/T refers to a fresh measurement of hairpins of type TT8D but normalized by reference obtained from a solution of pure 5(6)-TAMRA (EMD Millipore, Burlington, MA). Lag time was scaled such that the diffusion time of TAMRA matched that of the hairpin. The buffer environment was found to induce sufficient folding in T_8-loop hairpins, so no salt was added to those samples. For T_{21}-loop hairpins, NaCl was added at $100\,\mathrm{mM}$ to facilitate folding. In the presence of NaCl, addition of a nonionic surfactant was necessary to stabilize hairpin concentration by suppressing adsorption to the vials during measurements; proteomic grade Triton X-100 (G-Biosciences, St. Louis, MO) was therefore added at $10^{-4}\%\,\mathrm{w/v}$ ($\sim 1.5\,\mu\mathrm{M}$) to the T_{21}-loop samples. Figure 7.12 shows no dependence of melting curves on surfactant concentrations up to $0.01\%\,\mathrm{w/v}$.

For temperature dependence measurements, a variant of FCS setup was built in-house as shown in Figure 6.3. A glass capillary tube (Polymicro Technologies, Phoenix, AZ) of square cross section (50 and $300\,\mu\mathrm{m}$ inner and outer dimensions) was uncoated and mounted on a thin coverslip, which itself was mounted on the surface of a Peltier device. Both ends of the capillary were submerged in solution and pneumatic pressure was used to fill the capillary, with no flow during the

Table 6.1: Oligo samples used in this study. The terminal dG in TT21G is part of the DNA sequence.

type	name	loop length	5′ label	3′ label	source
test	RT8D	8	R6G	Dabcyl	AnaSpec
	TT8D TT8D/T	8	TAMRA	Dabcyl	IDT
	RT21D	21	R6G	Dabcyl	AnaSpec
	TT21G	21	TAMRA	(dG)	IDT
reference	RT8	8	R6G	–	AnaSpec
	TT8	8	TAMRA	–	IDT
	RT21	21	R6G	–	AnaSpec
	TT21	21	TAMRA	–	IDT

measurements. The Peltier device was run by a temperature controller (TC-48-20, TE Technology, Traverse City, MI) using a small thermistor placed adjacent to the capillary. The temperature was stable within $0.1\,^{\circ}\mathrm{C}$ from 5 to $90\,^{\circ}\mathrm{C}$. Temperature readings were calibrated using the dependence of the fluorescence of single- and dual-labeled hairpin samples on temperature as described in Section 7.2.3. The main optical components of the setup consisted of a $532\,\mathrm{nm}$ cw solid state laser (B&W Te, Newark, DE), a finite conjugate microscope objective (Achromatic $100\times/1.25$ Oil, Edmund optics, Barrington, NJ), a dichroic beamsplitter and a long-pass emission filter (both Omega Optical, Brattleboro, VT) with cut-offs at $540\,\mathrm{nm}$ and $550\,\mathrm{nm}$, respectively, a $75\,\mu\mathrm{m}$ pinhole, and an avalanche photodiode (APD) detector (SPCM-AQR-14, EH&G). Beam radius in the observation region was estimated to be $0.33 \pm 0.03\,\mu\mathrm{m}$ using Gaussian approximation and a diffusion coefficient of $(4.0 \pm 0.4) \times 10^{-6}\mathrm{cm}^2/\mathrm{s}$ for TAMRA [117, 118]. Total laser power was adjusted to $\sim 100\,\mu\mathrm{W}$ for TAMRA-labeled samples and $\sim 32\,\mu\mathrm{W}$ for R6G-labeled samples. Data was collected for $15\,\mathrm{min}$ for each sample. Non-drying high viscosity immersion oil (Cargille Laboratories, Cedar Grove, NJ) was used to support the horizontal geometry of the objective and operation at higher temperatures. Optical variations in the objective and immersion oil caused by temperature changes were inconsequential to our measurements due to normalization by reference data collected under identical conditions.

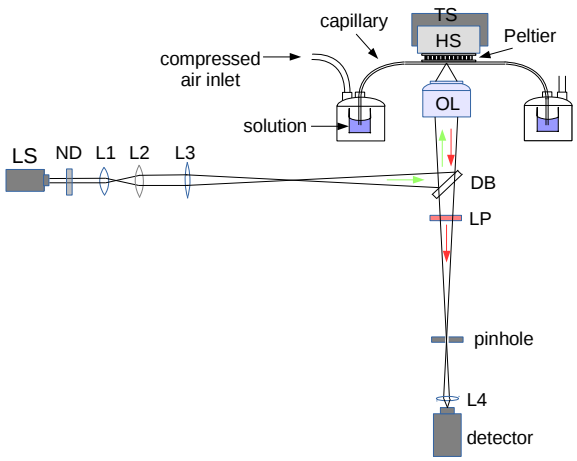

Figure 6.3: Experimental setup for temperature dependence studies. Excitation light (green arrow) from the laser source (LS) is passed through a neutral density filter (ND) and a Keplerian collimation and expansion pair of lenses (L1 and L2), then focused using a lens (L3) at the image plane of a DIN microscope objective lens (OL) which in turn focuses the light into the capillary tube. The fluorescence light (red arrow) is collected using the same objective and passed through a dichroic beamsplitter (DB), a long-pass filter (LP), a pinhole, and focused using an aspherical lens (L4) on the detection area. The capillary is mounted on a Peltier device attached to a heat sink (HS) placed on a translation stage (TS).

Raw photon data were recorded using a commercial card (DPC-230, Becker & Hicks, Berlin, Germany). Correlation functions were computed using the single-detector two-sub-bin method described in Chapter 2 with two small modifications described in Section 7.2.1, and implemented with an in-house C program. The resulting correlation curves were analyzed in IGOR Pro 6.3.7.2 (WaveMetrics, Lake Oswego, OR) using the weighted least-squares Global Fit package.

6.4.2 Salt dependence measurements

Sample RT21D, with reference RT21, was used for measurements at NaCl concentrations ranging from 10 to 100 mM. The buffer was the same as in the temperature dependence measurements. Addition of the surfactant Triton X-100 was also necessary in the presence of NaCl.

Three sets of data were collected: in trial 1, the laser power was $\sim 100\mu$W, the surfactant concentration was $10^{-4}\%\,\mathrm{w/v}$, and data was collected for 15 minutes for each sample. Trial 2 was similar to trial 1 except for shorter data collection times of 2-3 minutes. In trial 3, the laser power

was reduced to $\sim 50\mu W$, the surfactant concentration was increased to $5 \times 10^{-4}\%\,\mathrm{w/v}$, and data was collected for 10 minutes.

Salt dependence measurements were carried out using a commercial inverted microscope (TE2000-U, Nikon, Japan) modified for FCS, similar to the setup described elsewhere [115]. The notable differences were that the beam from the $532\,\mathrm{nm}$ laser source was collimated to underfill the objective back aperture, yielding a slightly larger focal radius ($0.37 \pm 0.04\,\mu\mathrm{m}$). A $100\,\mu\mathrm{m}$ pinhole was used on the detection side, followed by a 550 nm long-pass filter coupled with a 640 nm short-pass filter (to reduce the Raman-scattered background noise), and a single APD detector. Optical variations caused by changes in the refractive index of the solution with salt concentration were inconsequential to our measurements due to normalization by the reference data collected at the same salt concentration.

6.5 Results

6.5.1 Temperature dependence measurements

The temperature dependence measurements presented in this section used test sample RT8D and reference sample RT8. Additional samples with different loop length, fluorophore, and quencher were also tested; the results are presented in Section 7.2 and will be briefly described at the end of this section.

Figures 6.4(a–e) show the second- and higher-order reaction functions at different temperatures. Recall that these curves are obtained by normalizing the correlation curves $g_{m,n}(t)$ obtained from RT8D by those obtained from RT8. Panel (f) shows the second-order reaction functions side by side to aid comparison. The reaction curves show an exponential decay followed by a horizontal baseline. The flatness of the baselines confirms that the test and the reference samples have nearly identical diffusion properties. It also shows that the observed step of the reaction is complete within the FCS time window. Slower reactions, if present, would have time constants much greater than $2\,\mathrm{ms}$.

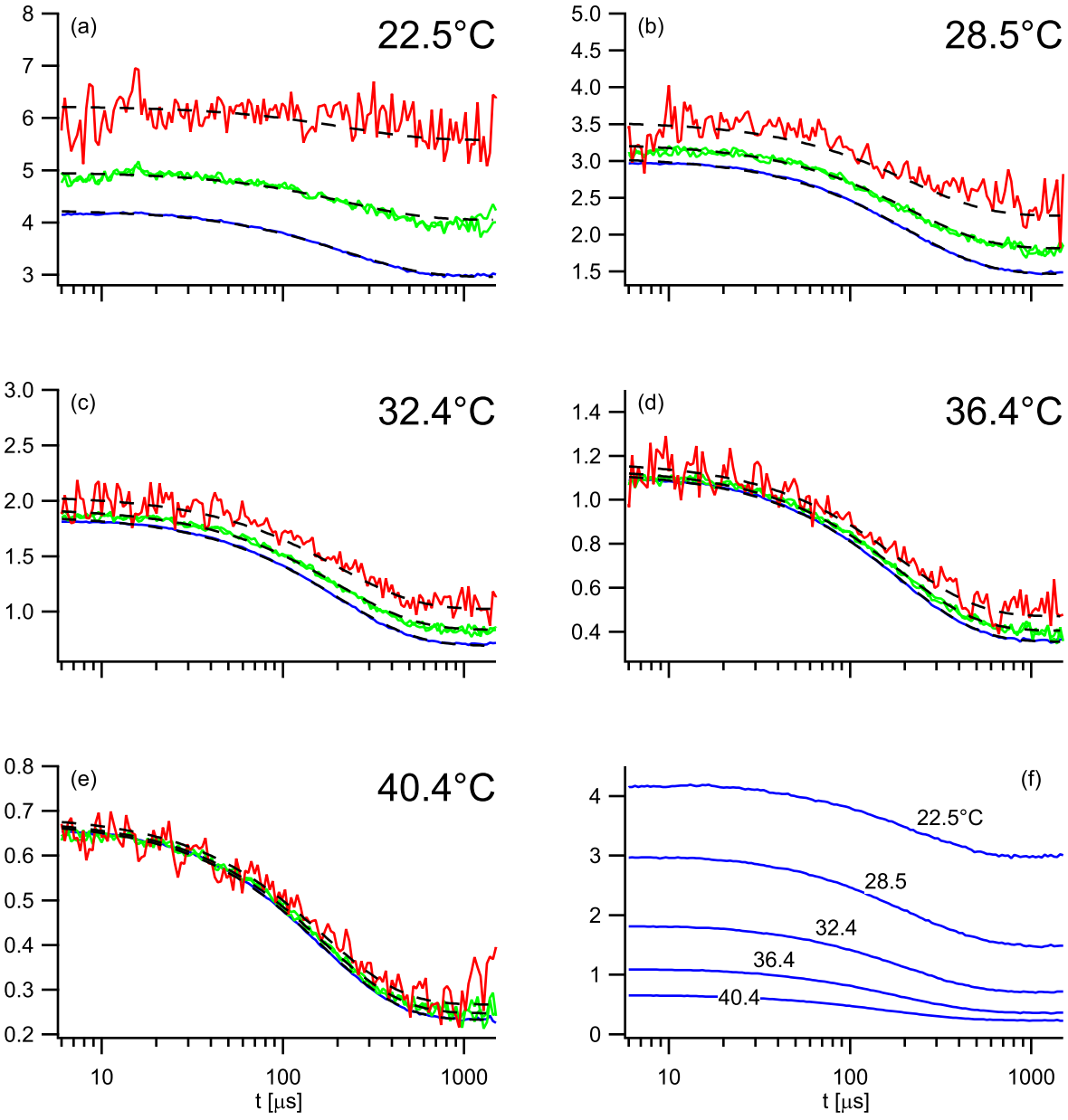

Figure 6.4: (a–e): $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different temperatures. (f) The second-order curves, $R_{11}(t)$, shown together on the same y-axis for better comparison. The shown temperature readings are post-calibration.

It has been shown [71] that higher-order FCS generally yields better precision in parameter determination when the dimmer molecules outnumber the brighter molecules, rather than the reverse. This agrees with the qualitative observation that as the temperature is increased and the hairpin molecules unfold, the reaction amplitude (the difference between its value at short lag times and its baseline value) decreases, and the higher-order reaction functions get closer to each other. Therefore, the temperatures for measurements were selected below the overall melting temperature of the hairpin at $43.2 \pm 0.1\,^{\circ}\mathrm{C}$ (measured for sample TT8D, see Section 7.2.3).

At each temperature, the higher-order reaction curves decay to separate baselines, rather than a single overall baseline, which shows that more than two states are involved in the reaction. This line of argument follows from a *single* measurement of the test and reference samples; it is independent from the variations of the baselines of the second-order reaction functions with temperature (Figure 6.4(f)) or ionic strength. A third-order FCS study of hairpin dynamics by Wu et al. [116] did not report this deviation from two-state mechanism since the reaction functions were not extracted and analyzed; only the reaction amplitudes were measured from the correlation functions.

Another qualitative observation in Figure 6.4 is that at any temperature the higher-order reaction curves decay to baselines slightly farther apart than the amplitude separations. Among three-state models, this is characteristic of the regime in which the fast step of the reaction is between the intermediate and the dimmest states, matching the situation shown in Figures 6.2(e) and (f).

To fit the data in Figure 6.4 using a three-state model, equation (6.5) was used with $\vec{K} = (1, K_2, K_3)$, $\vec{Q} = (1, q_2, q_3)$, and the approximate form of $\mathbf{Z}(t)$ as given in Equation (6.7). Since the N value should be temperature-independent, it was linked across the temperatures for improved fitting precision; we refer to this as "global fit". This also ensures that a single N value accounts for all the baseline variations. N values could alternatively be fitted at each temperature individually, and such approach was taken for salt dependence measurements. The resulting fit parameters are listed in Table 7.1 and the fit curves are shown as dashed lines in Figure 6.4.

To obtain reasonable fitting uncertainties, fitting weights were set as follows: data was fit using uniform fitting weights for a preliminary result. The mean residual errors from the lines of fit were then calculated and used as the new fitting weights for each correlation order. The process was repeated once or twice until convergence of χ^2_{red} to ~ 1 was reached. In this approach, we assumed only stochastic sources of error with normal and lag-time-independent distribution around the data expected values. To obtain a sense of systematic errors, including possible effects of the fluorophore and reference selection, measurements were repeated using TAMRA-labeled hairpin, and with pure TAMRA as reference; a comparison will be summarized at the end of this section. Other sources of systematic errors, including deviation from the three-state model and the approximate form of $\mathbf{Z}(t)$ in Equation (6.7) were not quantified in this study.

The fitted q parameters were free to take values less than or greater than 1, to ensure an unbiased fitting of models with respect to the fast reacting pair of states. The results definitively converged to $q_2 < 1$ and $q_3 > 1$, which corresponds to the fast step between the intermediate and the dimmest states. For example at $32.4^{\circ}\mathrm{C}$, the ratio of the brightness of state 2 to state 1 was $q_2 = 0.0018 \pm 0.0003$, and the ratio of the brightness of state 3 to state 1 was $q_3 = 1.36 \pm 0.02$. The obtained q values varied slightly with temperature, which can be attributed to the temperature dependence of the quenching and the resonance energy transfer (FRET) [119] mechanisms, as well as variations in the distribution of microstates within each measured state. However, the result that the fast step occurs between the intermediate and the dimmest states was true in all measurements. Attempts to forcibly fit the "dark sink" model in which the fast step is assumed between the brightest and the intermediate states ruled out that class explicitly, as described in Section 7.2.2 and Figure 7.3.

The equilibrium constants K_i were obtained directly from the fits and plotted against temperature in a van 't Hoff graph, Figure 6.5, left. A linear dependence of $\ln K_i$ as a function of $1/T$ was observed, in particular in the range 28.5–$40.4^{\circ}\mathrm{C}$. This signifies that the standard enthalpy changes, ΔH_i, and entropy changes, ΔS_i, are nearly constant over the measured temperature range. These values were obtained by fitting and reported in Table 6.2 for both the entire temperature range and the more linear subrange. From the latter, the enthalpy of the dimmest state

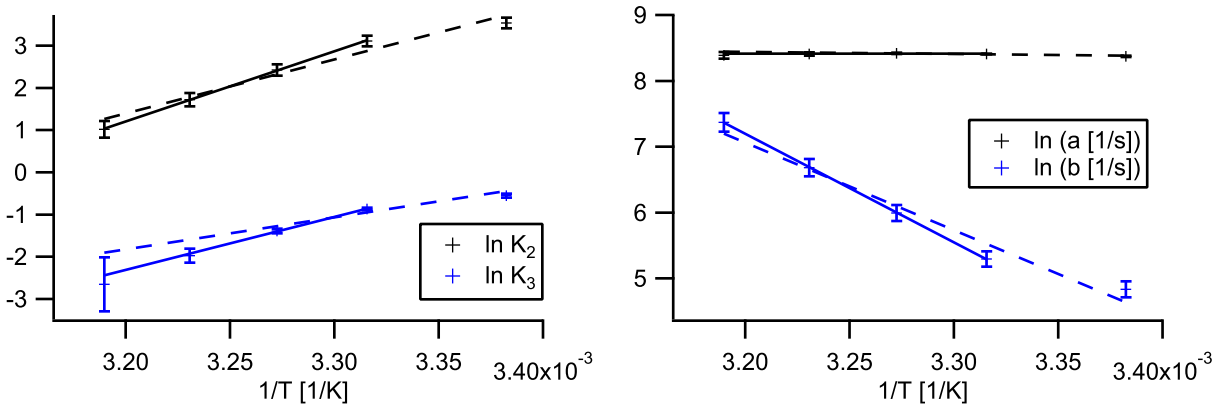

Figure 6.5: Left: Van 't Hoff plots of population ratios K_2 (dimmest/intermediate) and K_3 (brightest/intermediate). Right: Arrhenius plots of the rate constants, a (intermediate to dimmest) and b (dimmest to intermediate), of the fast step of the reaction. Linear fits on both graphs are shown using the four highest temperatures (solid) and all temperatures (dashed).

Table 6.2: Enthalpies and entropies relative to state 1, and the activation energies of the reaction intermediate $\rightleftharpoons$ dimmest. Numbers in parentheses indicate fitting uncertainties in the corresponding last digits.

	22.5–40.4°C	28.5–40.4°C
$\Delta H_2[\mathrm{kcal\,mol^{-1}}]$	$-25.4(2.0)$	$-32.9(3.3)$
$\Delta S_2[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.0784(66)$	$-0.103(11)$
$\Delta H_3[\mathrm{kcal\,mol^{-1}}]$	$-15.1(1.2)$	$-24.9(2.3)$
$\Delta S_3[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.0519(39)$	$-0.0844(75)$
$E_{\mathrm{a,fwd}}[\mathrm{kcal\,mol^{-1}}]$	$0.64(25)$	$0.07(42)$
$\ln(A_{\mathrm{fwd}}[s^{-1}])$	$9.48(42)$	$8.52(70)$
$E_{\mathrm{a,rev}}[\mathrm{kcal\,mol^{-1}}]$	$26.5(1.7)$	$32.7(2.8)$
$\ln(A_{\mathrm{rev}}[s^{-1}])$	$49.7(2.8)$	$59.9(4.5)$

is $\sim 33\,\mathrm{kcal/mol}$ below that of the intermediate state, and the enthalpy of the brightest state is $\sim 25\,\mathrm{kcal/mol}$ lower than that of the intermediate state. The intermediate state also has the largest entropy, about $0.1\,\mathrm{kcal\,mol^{-1}\,K^{-1}}$ above the two other states.

The fits to the reaction functions also directly yielded the relaxation time of the fast step of the reaction, t_{F}. The relaxation time, $t_{\mathrm{F}} = (a+b)^{-1}$, combined with the equilibrium constant, $K_2 = a/b$, yielded the forward and reverse rate constants, a and b, for the fast reaction step, intermediate $\rightleftharpoons$ dimmest. These rate constants are shown in an Arrhenius plot in Figure 6.5, right. The forward rate constant, a, shows very small temperature dependence, while the reverse rate constant,

b, increases significantly with increasing temperature. A generally Arrhenius behavior is observed for the rate constants over the measured temperatures, in particular in the range 28.5–$40.4^{\circ}\mathrm{C}$. The Arrhenius function $A\exp[-E_{\mathrm{a}}/(RT)]$ was fitted to the data and the results are included in Table 6.2. The activation energy in the forward direction is zero or negligibly small. The enthalpic rise from the dimmest to the intermediate state almost entirely accounts for the reverse reaction activation energy.

The temperature-independent N value and the equilibrium constants K_2 and K_3, all obtained from the fits, were used to calculate the absolute population of each state. These values are shown as a function of temperature in Figure 6.6, left (in semi-log scale). As the temperature is increased, the population of the intermediate state, N_1, increases significantly, with almost the entire contribution coming from the dimmest state population, N_2. The population of the brightest state does not change significantly over the measured temperature range. This is consistent with the overall energetics of the three states. Using the relative entropy and enthalpy values reported in Table 6.2 for the range 28.5–$40.4^{\circ}\mathrm{C}$ for a broader temperature range, we can approximately predict the populations over a broader temperature range. These are plotted in Figure 6.6, right, as fractions of the total population. Although small, the measured populations of the brightest state are near their global maximum at about $40\,^{\circ}\mathrm{C}$, explaining their slow variations with temperature in the measured range.

Figure 6.7 shows a schematic of the free energy levels of the three states with respect to state 1 at two selected temperatures of 30 and $40^{\circ}\mathrm{C}$, calculated using the enthalpy and entropy differences reported in Table 6.2. No energy barrier is considered between the intermediate and the dimmest states. The activation energy between the intermediate and the brightest state is known to be significant, though not directly measured; therefore it is represented by an arbitrary shape. At $30^{\circ}\mathrm{C}$, the majority of the hairpin molecules reside in the dimmest state with the lowest free energy. As the temperature is increased by $10\,^{\circ}\mathrm{C}$, the free energy levels of the dimmest and the brightest states both rise roughly by the same amount which is governed by the comparable entropies of the two states. This brings the dimmest level closer to the intermediate level, driving a significant

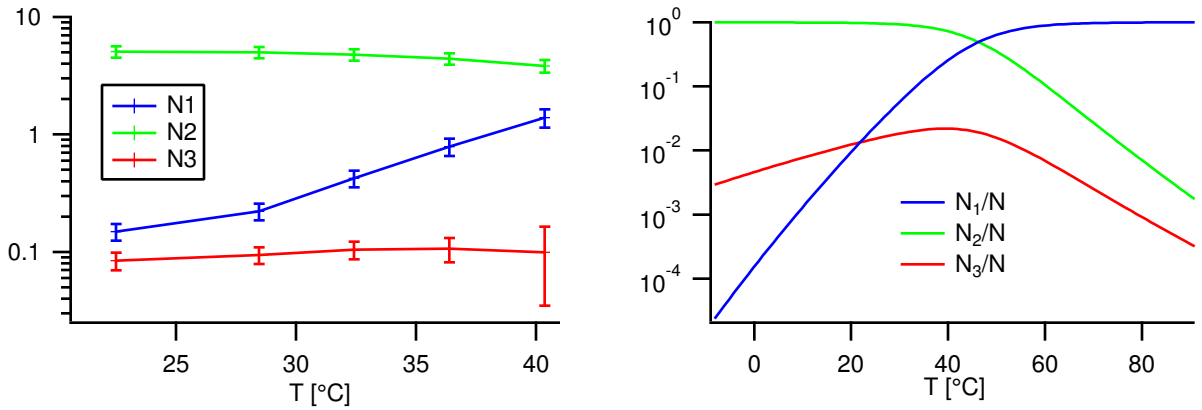

Figure 6.6: Left: Population numbers for the brightest (N_3), intermediate (N_1) and dimmest (N_2) states as a function of temperature. Right: Population numbers (normalized by the total population) predicted over a broader temperature range using the thermodynamic parameters reported in Table 6.2 for the range 28.5–40.4°C.

portion of the hairpin molecules into the intermediate state. However, the brightest state remains almost equally out of reach as before. This explains the small change in the brightest population over the examined temperature interval. The population of the brightest state is expected to slowly fade away at much higher or much lower temperatures, as was observed in Figure 6.6, right.

The three-state model with the measured brightness and population values should reproduce the overall fluorescence of the bulk sample as a function of temperature, commonly referred to as the melting curve. To verify this, the predicted overall brightness calculated using $\sum_{i=1}^{3} q_i N_i$ was compared with the average background-corrected intensity at each temperature. The results are shown in Figure 6.8. For ease of comparison, the curves were normalized by their average values (which avoids systematic bias by any single-point error). Good agreement is observed. The complete melting curve fits a two-state model acceptably (see Section 7.2.3), meaning that the presence of the small N_3 population does not significantly deviate the overall melting curve from a two-state appearance. This explains why the third state has been previously undetected by such bulk measurements.

Temperature dependence measurements were conducted for different types of hairpins for comparison. In Section 7.2.3 we present results for T_8-loop hairpins labeled with TAMRA and dabcyl as the dye and quencher, respectively (test sample TT8D, reference TT8). General agreement with

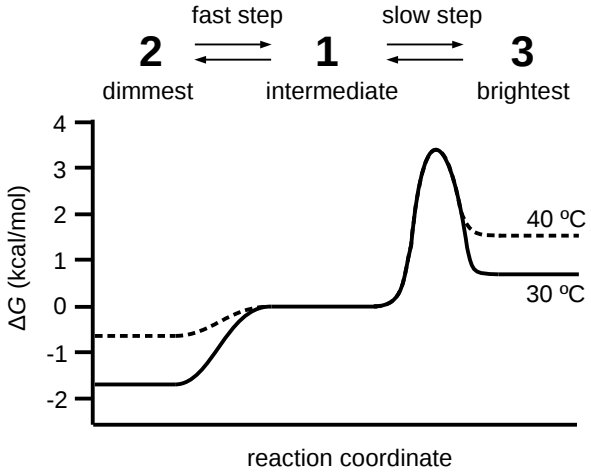

Figure 6.7: Schematic free energy landscape along an appropriate reaction coordinate. The horizontal levels represent experimental values obtained from Table 6.2 at $30°\mathrm{C}$ (solid) and $40°\mathrm{C}$ (dashed). The possibly small barrier between states 1 and 2 is ignored in this diagram, with arbitrary lines connecting the levels. The height and temperature dependence of the barrier between states 1 and 3 is also sketched arbitrarily.

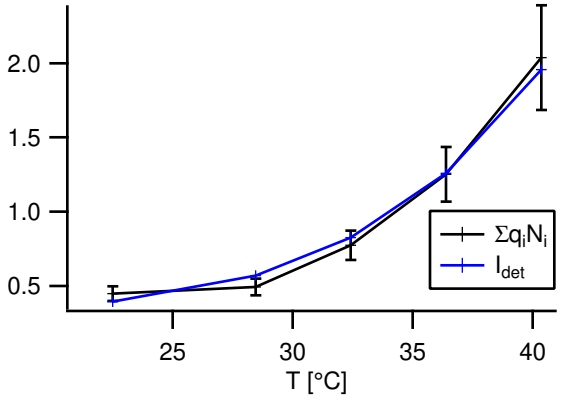

Figure 6.8: The overall brightness of the bulk sample as predicted by the measured N and q values (black) compared against the average detector count (blue). Uncertainties in the latter are relatively small hence not shown. The curves are normalized by their average values.

results from the R6G-labeled sample is observed. In particular, the qualitative behavior of the reaction curves shows the inadequacy of the two-state model in a similar way: the higher-order reaction curves decay to separate baselines, and the second-order baselines depend strongly on the temperature. The three state model with the fast step of the reaction between the intermediate and dimmest states still provides the best fit to the data. However, minor fitting imperfections are observed with TAMRA-labeled samples, which can point to more significant deviation from the three-state model and/or the approximate form of $\mathbf{Z}(t)$ with this fluorophore. The magnitudes of the energy and entropy differences obtained from the TAMRA-labeled samples are larger by about 20% than the R6G-labeled counterparts. Also, the small folding activation energy is slightly negative, which is not unprecedented by itself [88, 90, 105]. These differences, though on the order of the thermodynamical analysis uncertainties, might be due to the possible effects of the fluorophore on hairpin stability, dynamics, and the brightness distribution of the microstates that constitute the three observed states.

The possible effects of the reference sample on our measurement were investigated by using pure TAMRA as reference in a repeated set of measurements with TAMRA-labeled T_8-hairpin (sample TT8D/T). As explained in Section 7.2.3, the lag time in the correlation function of TAMRA was rescaled such that the diffusion time matched the diffusion time of the reference hairpin TT8. The scaled TAMRA diffusion function served as the new reference data by which the TT8D/T correlation functions were normalized. The resulting reaction curves were analyzed as before and the results were very similar to those from the measurement with a hairpin reference. This strongly supports the assumption that diffusion and conformational dynamics are independent processes in DNA hairpins. It also shows that experiments of this sort are not strongly constrained by the availability of reference samples.

More temperature dependence measurements were carried out for T_{21}-loop hairpins with the same stem. $100\,\mathrm{mM}$ of sodium chloride was added to the buffer to induce sufficient folding in these longer hairpins. In one case, R6G-labeled hairpin with dabcyl quencher was used (sample RT21D, Section 7.2.3). In another, TAMRA-labeled hairpin was used with guanine as the photoinduced

electron transfer (PET) pair (sample TT21G, Section 7.2.3). The results confirmed the three state model in each case. The quenching with guanine was much weaker, resulting in greater measurement uncertainties. At high [NaCl], the concentration of hairpins were found to be unstable due to surface adsorption to the vials (confirmed by single-labeled hairpins and by switching vials). A nonionic surfactant was added in small amounts to suppress these variations, as described in the Experimental Section. Residual concentration fluctuations remained in some measurements, and these needed to be taken into account in the analysis described below.

6.5.2 Salt dependence measurements

A series of measurements were carried out on the longer T_{21}-loop hairpins (test sample RT21D, reference RT21) at different NaCl concentrations. Three data sets (trials) were collected at different laser powers, as described in the Experimental Section. The reaction curves are shown in Figure 6.9 for one of the trials, along with the fitted curves. (The other trials are presented in Section 7.2.4.) In the global fitting procedure, the brightness values were linked across all salt concentrations since they were not expected to vary with [NaCl]. Instead, the N values were allowed to depend on NaCl concentrations since surface adsorption was observed to cause concentration fluctuations at higher ionic strengths.

The three state model with the fast step between the intermediate and dimmest states again provided the best fit to the data. Consistent with the temperature dependence measurements, brightness ratios of $q_2 = 0.020 \pm 0.005$ and $q_3 = 1.19 \pm 0.08$ (averaged over the trials) were obtained for the dimmest and the brightest states relative to the intermediate state.

It is seen in Figure 6.9 that the relaxation time of the fast step of the reaction increases significantly as [NaCl] decreases. At $10\,\mathrm{mM}$ NaCl, the relaxation time is far outside of the FCS time scale and cannot be directly obtained by fitting, therefore a value of $7\,\mathrm{ms}$ was assumed. Also, the uncertainty of the technique is higher in the limit of very low salt concentrations, since the reaction amplitude and baselines are not fully available, and the higher-order curves get very close to each

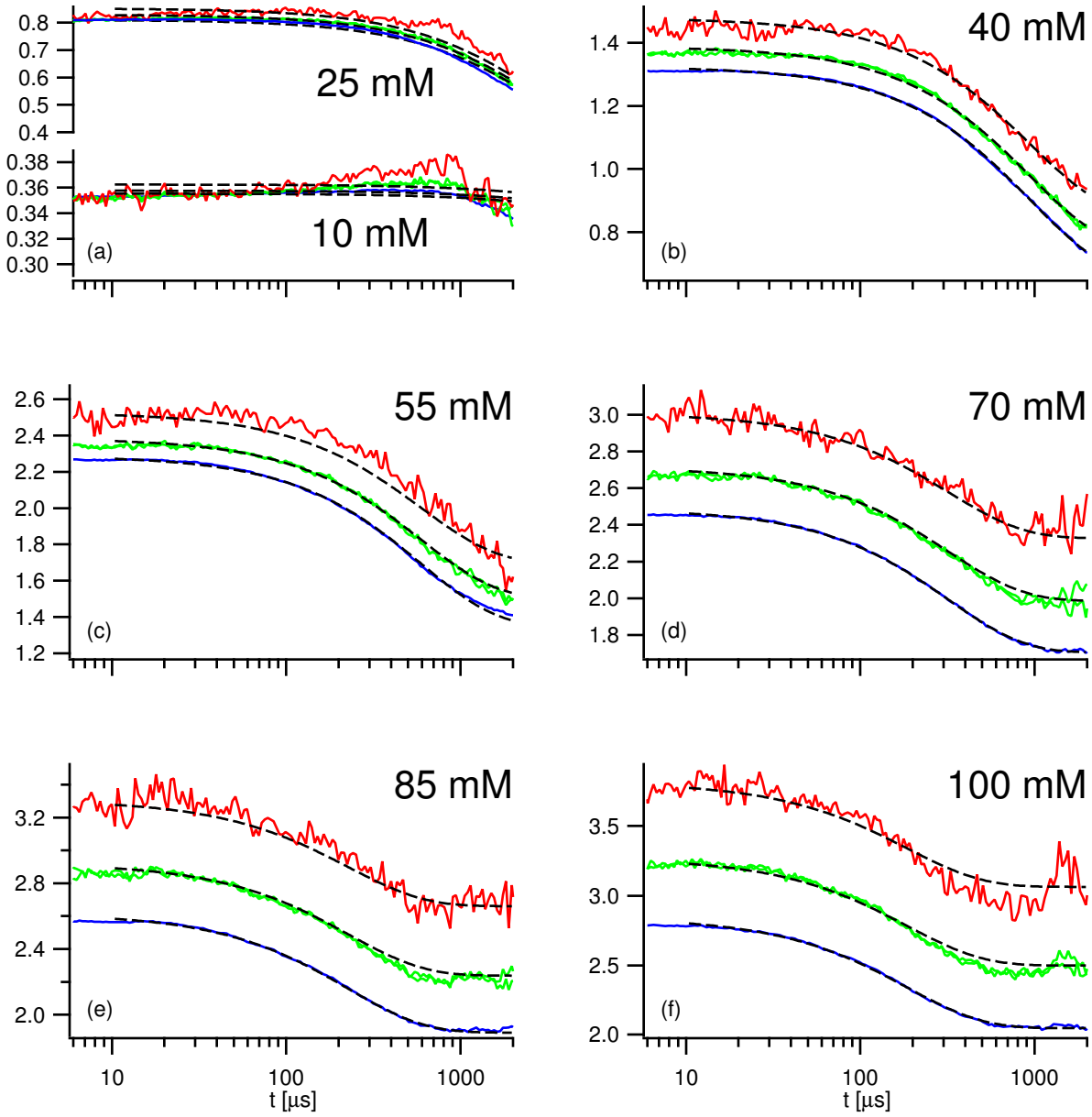

Figure 6.9: (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different NaCl concentrations. (sample: RT21D, trial 3)

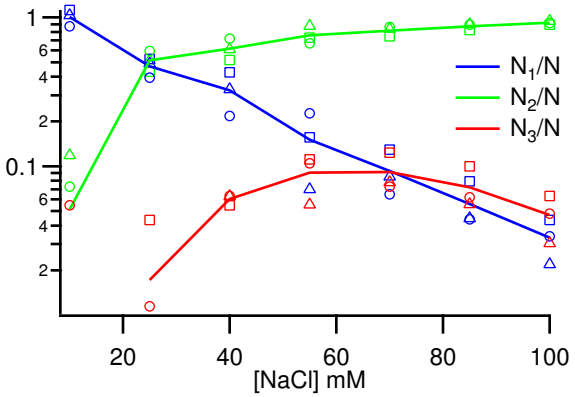

Figure 6.10: Population numbers for the brightest (N_3), intermediate (N_1) and dimmest (N_2) states as a function of NaCl concentration. Solid lines connect the average points of trials 1 (triangle), 2 (square), and 3 (circle). The peaking trend in N_3 seems to be robust since it is repeated in each trial within a systematic drift error. At very low [NaCl] some N_3 values were slightly negative due to higher uncertainty and the small value of N_3; see Figure 7.22 for linear axes.

other. However, at the medium or higher salt concentrations, the uncertainty of the technique is acceptable.

The values of N and the equilibrium constants K_i were directly obtained by fitting (listed along with all other fitting parameters in Table 7.8). From there, the state populations N_i were calculated and normalized by the total N value at each [NaCl]. The results are plotted in Figure 6.10. As before, the majority of the hairpins are found in the intermediate and dimmest states, shifting from the intermediate to the dimmest state as the ionic strength is increased. However, the small population in the brightest state also shows significant fractional variations with salt concentration, going from almost zero at the lowest salt concentrations to a maximum of $\sim 10\%$ of the total population near $70\,\mathrm{mM}$ NaCl, then slowly declining again at higher salt concentrations. This trend is reproduced within each dataset, suggesting it is robust. At very low salt concentrations, the uncertainty in N_3 is relatively large, however, based on extrapolations from higher salt concentration, the population appears to be negligibly small (see the linear scale graph in Figure 7.22).

Using the relaxation time of the fast step of the reaction, intermediate $\rightleftharpoons$ dimmest, and the equilibrium constants obtained from fitting, we calculated the forward and reverse rate constants,

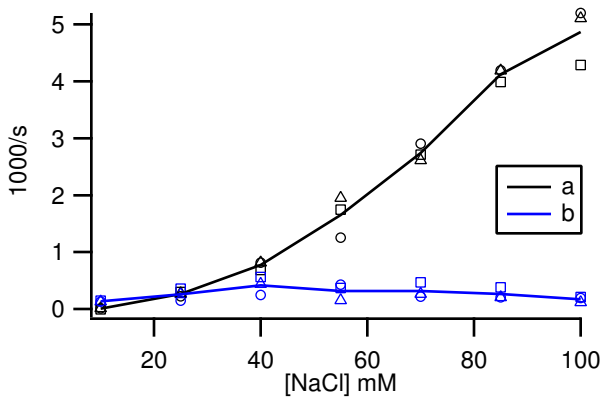

Figure 6.11: The rate constants, a (intermediate to dimmest) and b (dimmest to intermediate), of the fast step of the reaction as a function of [NaCl]. Solid lines connect the average points of trials 1 (triangle), 2 (square), and 3 (circle).

a and b, between the intermediate and dimmest states. The results are shown in Figure 6.11 as functions of NaCl concentration. Increased [NaCl] increases a, while b is almost unaffected.

Finally, we verified again that the measured number and brightness parameters for the three states reproduce the overall fluorescence of the sample. In Figure 6.12, the predicted sample brightness, $\Sigma_{i=1}^{3} q_i N_i$, is plotted together with the measured detector counts, I_{det}, as a function of the salt concentration. To assist comparison, each data set is normalized by the average of the values at $40\,\mathrm{mM}$ NaCl and higher due to the higher uncertainties in the measurements at 10 and $25\,\mathrm{mM}$ NaCl. Except at the two lowest NaCl concentrations, reasonable agreement is observed between the two types of measurements.

6.6 Discussion

Second-order FCS shows that no two-state dynamical model can account for the temperature or salt dependence data in the thymidine-loop DNA hairpins we have studied; however, second-order analysis alone cannot restrict a more complete model nor measure any rate constants. Higher-order FCS reveals that the only three-state model consistent with the data has a fast reaction, observable on the FCS time scale, between the intermediate brightness state (defined as relative brightness 1) and the dimmest state (relative brightness $q_2 < 1$). The third state, which is static on FCS

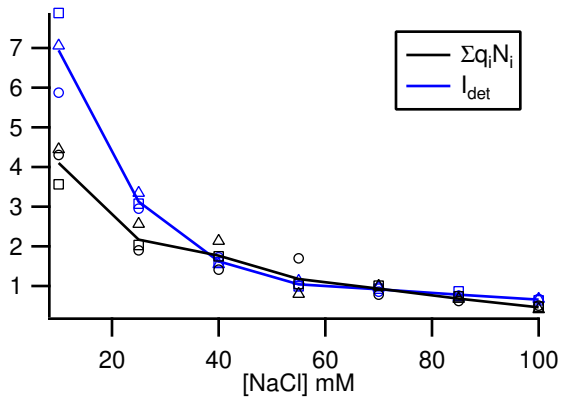

Figure 6.12: The overall brightness of the bulk sample as predicted by the measured N and q values (black) compared against the average detector count (blue). Solid lines connect the average points of trials 1 (triangle), 2 (square), and 3 (circle). Each trial is normalized by its average value excluding the two first date points.

timescales, is the brightest (relative brightness $q_3 > 1$). The state populations and brightnesses and the fast reaction rate constants are measurable at each temperature and ionic strength. Global fitting over multiple temperature and ionic strengths improves the precision of those parameters.

In the samples that had dabcyl as quencher, q_2 was tiny: typically around 1%. When guanine replaced dabcyl, q_2 was about 50%. The corresponding quenching efficiencies are comparable with, and generally higher than, those reported for dabcyl and guanine in contact mode [119]. These results suggest that the dimmest state is associated with the folded (native) hairpin conformations.

Thermodynamical analysis reveals that the intermediate-brightness state has significantly higher entropy than the folded state. With increasing ionic strength, the rate of transitions from the intermediate to the folded state increases, which suggests that the rate is controlled by the electrostatic repulsion of the hairpin phosphate backbone. The reverse rate does not depend on ionic strength, consistent with that rate being controlled by the breakage of the base pairs in the hairpin stem. Transitions from the intermediate to the folded state required negligibly small positive or negative activation energies. The activation energy from the folded to the intermediate state was almost equal to the enthalpy difference between the two states, which can be attributed to the enthalpy of unzipping the stem ($\sim 7\,\mathrm{kcal/mol/bp}$ [120,121]). All of these results suggest that the intermediate

state corresponds to random coil conformations, and support the idea that the stem duplex is fully formed in the dimmest (folded) state.

Small energy barriers and relaxation times of a few hundred microseconds have been previously observed and attributed to transitions between folded and random-coil conformations [35, 82, 86, 90, 94, 113]. Some studies, in particular using T-jump, have reported time scales in a few tens of microseconds [38, 88, 89]. We observed no faster reaction components in our data; a single exponential model fitted the curves for intermediate-dimmest transitions acceptably. Stretched exponentials have been reported to better fit the results of FRET-FCS measurements [83–85, 87], but only minimally improved the fits in FCS with contact-mode quenchers [38, 85, 86, 112]. It has been argued that the appearance of stretched exponentials in FRET-FCS might be an artifact of the technique [122, 123]. Fast triplet blinking of the fluorophore may also cause measurement complications if the correlation functions, $g_{m,n}$, are fitted in FCS. These effects are reduced in reaction function analysis (fitting to $R_{m,n}$) employing the same type of fluorophore in the test and reference samples.

We were perplexed that the static state was the brightest, at first. If the long-lived state is not the fully-folded state (its role in the dark sink model [37]) nor some mis-folded state (its role in models with intermediate "trap" states [88–90, 108, 109, 124, 125]), what could it be? Surely it could not be the random-coil state. But if not that, what state could be brighter?

One hypothesis is that the brightest state is not associated with hairpin conformations, but is rather an experimental artifact coming from "impurities": labels which became detached from the DNA, or DNA from which quencher had detached, or some fluorescent contaminant. We are convinced this is not the case. All samples were of dual HPLC purity. We confirmed that the static state was the brightest in samples of different fluorophore and quencher types, and obtained from different sources. New and aged samples yielded consistent results, ruling out hairpin degradation. Free fluorophores have much shorter diffusion times than the reference DNA hairpin; this would prevent the baselines of the relative reaction functions, which are sensitive to the static population, from appearing flat (see Section 7.1.2 for details). Contaminants in reference hairpin samples can-

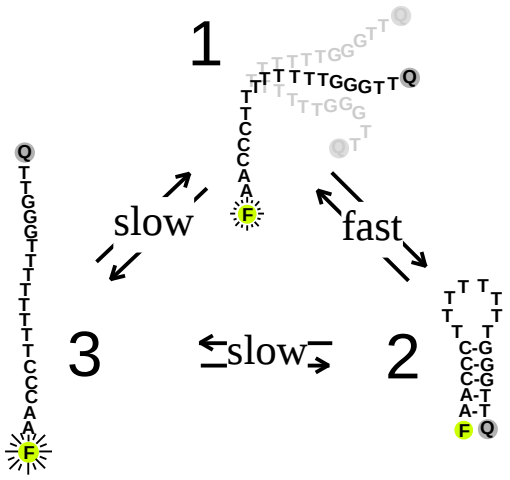

Figure 6.13: Schematic examples of DNA hairpin conformations associated with the three states as revealed by brightness and thermodynamical analyses. F and Q denote the fluorophore and quencher labels.

not be the cause either, since the results were reproduced using pure dye as reference. Incomplete quencher labeling cannot explain the results in the guanine-quenched samples, since guanine is part of the DNA chain. Finally, while changes in the population of the brightest state were small over the limited temperature range, as one would expect for a contaminant, in the salt dependence studies its population ranged from zero to 10%, which is far greater variation than the measurement uncertainty. We are thus confident that the brightest state is associated with some class of DNA hairpin conformations.

What are the conformations associated with the brightest state? Since it is (slightly) brighter than the random-coil state, the probability that the ends of a hairpin are near each other is lower than in random-coil conformations. This implies that the brightest state corresponds to a set of relatively extended conformations (and we will refer to it as the extended state, hereafter) which are somehow separated from more compact conformations by a substantial energy barrier. The set of these conformations is small as revealed by the entropy of the extended state, which is comparable to the folded state. In terms of enthalpy, it is between the folded and random-coil states. Figure 6.13 shows a schematic of the DNA hairpin conformations and mechanism we are proposing.

Recent small-angle x-ray scattering (SAXS) studies of T_{30} and A_{30} single-stranded DNA by Plumridge et al. [126, 127] show distinct extended conformations in T_{30} with salt-dependent properties similar to the extended state reported here. Looking at their data at $100\,\mathrm{mM}$ NaCl for T_{30}, (see Figure 6.14, C) nearly 10% of the population resides in a distinct set of conformations, which they refer to as a secondary population, with anomalously large end-to-end distance and radius of gyration. The primary population corresponds to more compact random-coil conformations (no hairpin stem thus no folded conformation exists). The secondary population is absent at $20\,\mathrm{mM}$ NaCl; and at $200\,\mathrm{mM}$ the secondary population is less than at $100\,\mathrm{mM}$ (see Figure 6.14, left column). This is consistent with the salt dependence of the extended population in the studies reported here (see Figure 6.10). The isolation of the extended conformations seen in Figure 6.14 also suggests a potential barrier responsible for slow transitions between the extended and more compact conformations. Plumridge et al. were concerned that this secondary population could be an experimental artifact. It would be a remarkable coincidence if that were the case: it seems more likely that higher-order FCS and SAXS independently show evidence of an unexpected extended state in poly-T single-stranded DNA which has a maximum population at near-physiological salt concentration at room temperature.

A computational study by Chakraborty et al. [107] on RNA tetraloop hairpins shows that when folding starts from an extended conformation, the folding rate is several orders of magnitude slower than when it starts from a "compact unfolded" state. This is consistent with our observations for DNA hairpins, with the intermediate-brightness, random-coil state corresponding to the compact unfolded state. The study also finds that the extended conformations have very low equilibrium occupation probability and are disconnected from the rest of the free energy landscape by the highest barriers, again consistent with our observations. Those authors explain that the interplay of competing base-pairing, base-stacking, and electrostatic interactions results in the partitioning of the unfolded population into slow and fast folders.

A disparity in time scales when folding starts from distinct unfolded states is consistent with T-jump and rapid mixing experiments by Narayanan et al. [90] Those authors found that the time

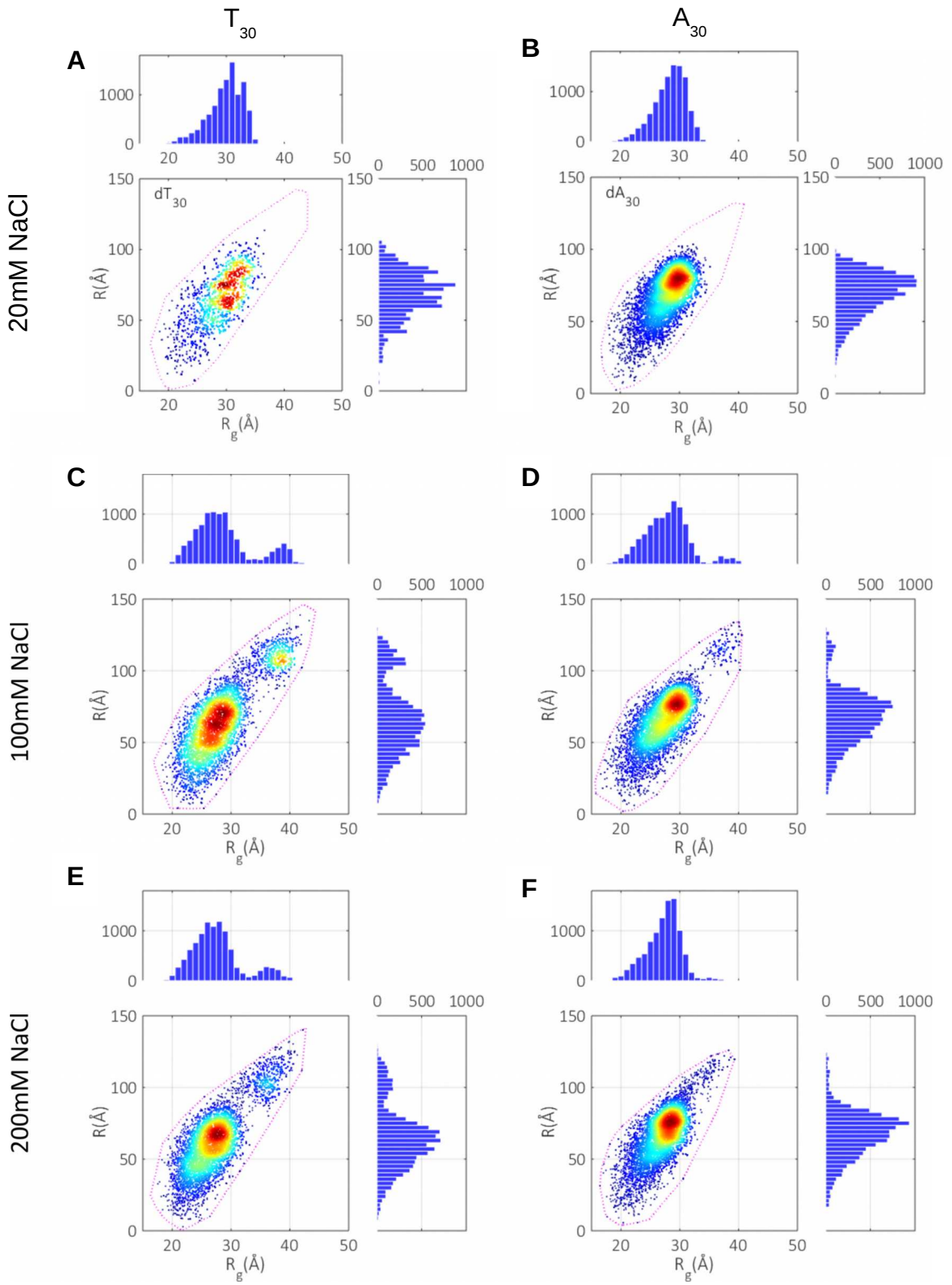

Figure 6.14: Distributions of T_{30} and A_{30} at 20, 100, and 200 mM NaCl visualized in a two-dimensional space of end-to-end distance (R) and radius of gyration (R_g) as obtained in SAXS measurements by Plumridge et al. [126, 127]. Data reproduced under CC BY-NC 4.0; panels rearranged from original.

scale is nearly an order of magnitude slower in rapid mixing ($\sim 1\,\mathrm{ms}$), where the folding starts from a less compact conformation, than in T-jump. The rapid mixing experiments in our lab by Nayak et al. [96, 97] also showed slower time scales than those observed by FCS. However, the initial conformations in rapid mixing are probably not the same as the extended conformations reported here, for two reasons: First, the time scale observed in rapid mixing measurements is still much faster than the four orders of magnitude increase suggested by Chakraborty et al. [107] The present results suggest folding from the extended conformation takes longer than a few milliseconds, more in line with the results of Chakraborty et al.. Second, the present results and the SAXS data [127] show relatively small extended populations at all salt concentrations. This population is likely too small to be detected in bulk measurements such as rapid mixing and T-jump; moreover, the population of extended conformations is practically zero at the low-ionic premixing conditions of rapid mixing experiments.

The SAXS studies found that A_{30} behaved differently from T_{30} in a few ways; most relevant to us, the secondary population was far smaller than for T_{30} at the same salt concentration. There is thus an urgent need for FCS (and particularly higher-order FCS) studies of similar DNA hairpins with adenine loops, to check for full consistency of FCS with SAXS with regards to the extended state populations. At least one hypothetical explanation consistent with both poly(A) and poly(T) measurements (and other details in the SAXS data) can be provided based on competing stacking and electrostatic interactions. See Section 7.3 for details.

It would be especially interesting to extend the higher-order FCS to longer time scales using diffusion deceleration [40], gel [128], or time gating [129] methods. The second-order diffusion-decelerated FCS by Yin et al. [40] on a T_{21}-loop DNA hairpin has already shown that new relaxation times in tens and hundreds of milliseconds appear at NaCl concentrations over $125\,\mathrm{mM}$. Although their examined [NaCl] range does not overlap with our measurements, their work does support the notion of a slowly reacting state which becomes more stable with increased ionic strength.

Major questions remain open for future investigations. What causes the extended conformation basin and its associated barrier to appear in the free energy landscape? How is the dependence of the extended population on salt concentration explained? Why is the extended population larger in poly-T than in poly-A? Important clues were presented by Chakraborty et al. who employed realistic force fields involving base-stacking and electrostatic interactions in their computational work. More theoretical and experimental studies are needed on poly-A and poly-T DNA hairpins at various ionic and temperature conditions for more clear answers.

6.7 Conclusion

We have used higher-order FCS to investigate the conformational fluctuations of DNA hairpins in aqueous solution as a function of temperature and salt concentration. These analyses revealed the non-two state nature of the DNA conformational dynamics under all conditions studied, based on single measurements of test and reference samples. Further, the only class of three-state models consistent with the analysis of autocorrelation functions up to fourth order shows rapid fluctuations between the lowest and intermediate brightness states, while the brightest state is static on the FCS time scale. This rules out the previously assumed dark sink model. Higher-order analysis yields the kinetic parameters previously inaccessible by second-order FCS, including the relative brightnesses and populations of the three states. In addition, both the forward and reverse rate constants of the transitions within the FCS time scale are obtained. The dependence of these parameters on temperature and ionic strength suggests that the dimmest and intermediate states are associated with the familiar folded and random-coil conformations, respectively. It is concluded that the brightest state is associated with unfolded conformations that are anomalously spatially extended. These extended conformations are separated from the more compact conformations by a substantial energy barrier. Supporting evidence for such conformations exists in independent SAXS data [126, 127] and computational studies [107]. Consistent with the SAXS results, we find that the extended conformations are practically absent at low salt concentrations, while they reach

a maximum equilibrium population of $\sim 10\%$ in the vicinity of physiological temperatures and salt concentrations.

Chapter 7

Conformational dynamics of DNA hairpins—supporting information

7.1 Theory

7.1.1 General considerations

Considerations regarding reversibility and detailed balance become relevant mainly in systems with more than two states, where, for example, cyclic dynamics can appear in a broad category of system configurations. One main advantage of higher-order correlation analysis is that it can reveal the reversible or non-reversible nature of the dynamics, simplifying the analysis and providing a more comprehensive and definitive picture of the system. We formulate the results in this section for a general multi-state system.

Reversibility condition

Recall that $Z_{i,j}(t)$ is the conditional probability that a particle is found in state i at time t given it was in state j at time 0:

$$Z_{i,j}(t) = P(i,t|j,0) \tag{7.1}$$

Reversibility is the condition that the transition probability between any two states is independent of the direction of time flow, i.e. 0 and t can be interchanged in $P(i,t|j,0)$:

$$P(i,t|j,0) = P(i,0|j,t) \tag{7.2}$$

In other words, if time flows backward, the reaction process remains statistically indistinguishable from the forward flow reaction. An example of non-reversible dynamics in this sense would be cyclic behavior in multi-state systems (see Section 7.1.1 for more details).

We can re-write Equation (7.2) in a more useful form by applying Bayes' theorem:

$$P(i,0|j,t) = \frac{P(j,t|i,0)P_i}{P_j} \tag{7.3}$$

where a stationary process is assumed: $P_i = P(i,t)$ (for any t) is the (marginal or stationary) probability of finding the particle in state i regardless of the state of the particle at any other time. Substituting (7.3) into (7.2), the reversibility condition becomes:

$$P(i,t|j,0) = \frac{P(j,t|i,0)P_i}{P_j}$$

or, using the notation of Equation (7.1),

$$Z_{i,j}(t)P_j = Z_{j,i}(t)P_i \tag{7.4}$$

It is evidently convenient to define a $J \times J$ diagonal matrix $\mathbf{P}$ with elements P_i on the main diagonal. Then (7.4) can be written as

$$(\mathbf{ZP})_{i,j} = (\mathbf{ZP})_{j,i}$$

which is valid for all i, j and any lag time t. Thus, reversibility is the condition that the matrix $\mathbf{Z}(t)\mathbf{P}$ is symmetric: $\mathbf{ZP} = (\mathbf{ZP})^{\mathrm{T}}$, or,

$$\mathbf{Z}(t)\mathbf{P} = \mathbf{P}\mathbf{Z}^{\mathrm{T}}(t) \tag{7.5}$$

Detailed balance condition

Suppose the reaction can be described using a system of linear first-order rate equations

$$\frac{\mathrm{d}\vec{P}(t)}{\mathrm{d}t} = \mathbf{A}\vec{P}(t)$$

where $\mathbf{A}$ is a matrix of linear rate constants as described in Section 6.3. More specifically, the non-diagonal element $A_{i,j}$ is the rate constant from j to i. Detailed balance is the condition that the forward and reverse rate constants between any (non-identical) pair of states i, j satisfy

$$A_{i,j}P_j = A_{j,i}P_i \tag{7.6}$$

which holds for time-independent (stationary) marginal probabilities (P_i). Equation (7.6) is trivially valid for identical indices as well. Thus, detailed balance is the condition that the matrix $\mathbf{AP}$ is symmetric: $\mathbf{AP} = (\mathbf{AP})^{\mathrm{T}}$, or,

$$\mathbf{AP} = \mathbf{PA}^{\mathrm{T}} \tag{7.7}$$

Reversibility and detailed balance are equivalent

Assuming a linear first-order reaction for which $\mathbf{A}$ exists, the reversibility and detailed balance conditions imply one another. To show, recall that:

$$\mathbf{Z}(t) = e^{\mathbf{A}t} \tag{7.8}$$

If the reaction is reversible, we simply take the first derivative of the reversibility equation, (7.5), with respect to the lag time t:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{Z}(t)\mathbf{P} = \mathbf{P}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{Z}^{\mathrm{T}}(t)$$

which, using (7.8), yields the detailed balance condition, Equation (7.7), at $t = 0$.

To show the converse, we rewrite the detailed balance equation (7.7) as

$$\mathbf{A}^{\mathrm{T}} = \mathbf{P}^{-1}\mathbf{AP} \tag{7.9}$$

$\mathbf{P}$ is invertible with no zero element on the diagonal. (Any state with zero occupation ($P_i = 0$) would be effectively non-existent, hence not included in the model.) Equation (7.9) then yields, for any non-negative integer power n,

$$(\mathbf{A}^{\mathrm{T}})^n = \mathbf{P}^{-1}\mathbf{A}^n\mathbf{P}$$

As a result

$$\sum_{n=0}^{\infty} \frac{1}{n!}(\mathbf{A}^{\mathrm{T}})^n t^n = \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{P}^{-1}\mathbf{A}^n\mathbf{P}t^n$$

by term-by-term agreement. This becomes

$$\left(\sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n t^n\right)^{\mathrm{T}} = \mathbf{P}^{-1}\left(\sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n t^n\right)\mathbf{P}$$

The expressions inside parentheses are the Taylor expansions of $\mathbf{Z}(t) = e^{\mathbf{A}t}$, thus we have

$$\mathbf{Z}^{\mathrm{T}}(t) = \mathbf{P}^{-1}\mathbf{Z}(t)\mathbf{P} \tag{7.10}$$

which is the reversibility condition, Equation (7.5).

Relation to eigenvalues of $\mathbf{A}$ and $\mathbf{Z}$

If a reaction is in detailed balance, or equivalently reversible, the eigenvalues of $\mathbf{A}$ and $\mathbf{Z}$ are real. This result is useful because it significantly reduces the categories of models that need to be considered (Section 7.1.2) for analyzing the data when reversibility is readily established through the symmetry of higher-order correlation functions (see Section 7.1.1). To show this theorem for $\mathbf{A}$, we obtain from Equation (7.9) that

$$\mathbf{P}^{1/2}\mathbf{A}^{\mathrm{T}}\mathbf{P}^{-1/2} = \mathbf{P}^{-1/2}\mathbf{A}\mathbf{P}^{1/2}$$

which is manifestly a symmetric real matrix, and hence has real eigenvalues. This implies that $\mathbf{A}$ also has real eigenvalues, following the properties of similarity transforms. A similar result about the eigenvalues of $\mathbf{Z}(t)$ follows from Equation (7.10).

In a rather intuitive argument, if the eigenvalues of $\mathbf{A}$ include imaginary parts, then the dynamical system will no longer relax with simple exponential rates, but will also include cyclic dynamics (in microscopic level) in preferred cyclic direction(s). Such cyclic direction(s) would be reversed, in a statistically distinguishable sense, by the reversal of time direction, which would be in direct contradiction with the reversibility condition.

However, the converse of the theorem does not hold, i.e. the "realness" of the eigenvalues of $\mathbf{A}$ is not a sufficient condition to guarantee reversibility or non-cyclic behavior. A notable counter-example is the type of three-state models proposed for triplet blinking of fluorophores [130], in which a fast bidirectional process is assumed between the ground and the excited states, and a slower, cyclic, and unidirectional mechanism is assumed through the triplet state. This results in two distinct and real relaxation rates, very consistent with the observations about DNA hairpin dynamics, while the underlying dynamics are in fact irreversible. The symmetry of higher-order correlation functions disproves such irreversible models for DNA hairpin dynamics, as explained in Section 7.1.1. However, to directly observe asymmetric higher-order FCS curves for the irreversible process of triplet blinking, at least two fluorescent states would be required, which may not be the case for common fluorophores.

Relation to higher-order correlations

The reaction factor, $R_{m,n}(t)$, in higher-order correlation functions can provide information about the reversibility (detailed balance) of the reaction. When the reaction is reversible, the higher-order reaction factors (and hence the corresponding correlation functions) are symmetric: $R_{m,n}(t) = R_{n,m}(t)$. To show this result, let us re-write the reaction function, Equation (6.5), in the following form:

$$R_{m,n}(t) = \frac{1}{N} \frac{\overleftarrow{Q^n}\mathbf{Z}(t)\mathbf{P}\overrightarrow{Q^m}}{\overleftarrow{Q^m}\mathbf{P1P}\overrightarrow{Q^n}}$$

Here, we have defined $\overrightarrow{Q^m}$ and $\overleftarrow{Q^n}$ to be a column and row vectors with elements Q_i^m and Q_i^n, respectively. $\mathbf{1}$ is a $J \times J$ matrix of 1's (for all elements). Now note that $R_{m,n}(t)$ is a scalar, therefore,

$$\begin{aligned} R_{m,n}(t) &= [R_{m,n}(t)]^{\mathrm{T}} \\ &= \frac{1}{N} \frac{\overleftarrow{Q^m}\mathbf{P}\mathbf{Z}^{\mathrm{T}}(t)\overrightarrow{Q^n}}{\overleftarrow{Q^n}\mathbf{P1P}\overrightarrow{Q^m}} \end{aligned}$$

Using the reversibility condition, Equation (7.5), we get

$$\begin{aligned} R_{m,n}(t) &= \frac{1}{N} \frac{\overleftarrow{Q^m}\mathbf{Z}(t)\mathbf{P}\overrightarrow{Q^n}}{\overleftarrow{Q^n}\mathbf{P1P}\overrightarrow{Q^m}} \\ &= R_{n,m}(t) \end{aligned}$$

showing the symmetry of higher-order correlation functions, $g_{m,n}$, as well.

The converse theorem involves a more challenging question: Can higher-order correlations definitively identify an irreversible process? The answer can be shown to be positive if $R_{m,n}(t) = R_{n,m}(t)$ holds for all higher orders up to order (J, J) with a set of non-identical and non-zero Q_i values.

To show this, assume the symmetry of the reaction functions:

$$\begin{aligned} R_{m,n}(t) &= R_{n,m}(t) \\ &= [R_{n,m}(t)]^{\mathrm{T}} \end{aligned}$$

This can be written as

$$\frac{1}{N} \frac{\overleftarrow{Q^n}\mathbf{Z}(t)\mathbf{P}\overrightarrow{Q^m}}{\overleftarrow{Q^m}\mathbf{P1P}\overrightarrow{Q^n}} = \frac{1}{N} \frac{\overleftarrow{Q^n}\mathbf{P}\mathbf{Z}^{\mathrm{T}}(t)\overrightarrow{Q^m}}{\overleftarrow{Q^m}\mathbf{P1P}\overrightarrow{Q^n}}$$

The denominators cancel out and we get

$$\overleftarrow{Q^n}\mathbf{Z}(t)\mathbf{P}\overrightarrow{Q^m} = \overleftarrow{Q^n}\mathbf{P}\mathbf{Z}^{\mathrm{T}}(t)\overrightarrow{Q^m} \tag{7.11}$$

It should be clear that if

$$\overleftarrow{u}\,\mathbf{Z}(t)\mathbf{P}\,\overrightarrow{v} = \overleftarrow{u}\,\mathbf{P}\mathbf{Z}^{\mathrm{T}}(t)\,\overrightarrow{v}$$

for all vectors $\overleftarrow{u}$ and $\overrightarrow{v}$, then the operators $\mathbf{Z}(t)\mathbf{P}$ and $\mathbf{P}\mathbf{Z}^{\mathrm{T}}(t)$ are equal by the definition of operator equality. However, it is not necessary to test for every $\overleftarrow{u}$ and $\overrightarrow{v}$ due to linearity. As long as the vectors $\{\overrightarrow{Q^1}, \overrightarrow{Q^2}, \ldots, \overrightarrow{Q^J}\}$ are linearly independent (J, the number of states, is also the dimensionality of the space) then the equality $\mathbf{Z}(t)\mathbf{P} = \mathbf{P}\mathbf{Z}^{\mathrm{T}}(t)$ is guaranteed provided Equation 7.11, or equivalently $R_{m,n}(t) = R_{n,m}(t)$, holds for $m, n = 1, 2, \ldots, J$. For linear independence of $\{\overrightarrow{Q^1}, \overrightarrow{Q^2}, \ldots, \overrightarrow{Q^J}\}$ it is necessary and sufficient to have distinct, non-zero Q_i values.

In our current higher-order FCS study of DNA hairpin dynamics we observe the symmetry of reaction functions for $m, n = 1, 2$, while a three-state model, $J = 3$, is assumed. Despite this limitation and that one of the brightness values is very small, we still consider it highly unlikely to have an irreversible process that is somehow hidden by these limitations.

7.1.2 The three-state model

Here, we apply some of the above results to the general three state model.

Eigenvalues and eigenvectors of A

For a three-state model with 6 rate constants as shown in Figure 6.2(c) or (e), we have

$$\mathbf{A} = \begin{bmatrix} -(a+f) & b & e \\ a & -(c+b) & d \\ f & c & -(d+e) \end{bmatrix}$$

The three eigenvalues are

$$\lambda_0 = 0$$

$$\lambda_\pm = -\frac{1}{2}(\Sigma \pm \sqrt{\Delta})$$

where

$$\Sigma = a + b + c + d + e + f$$

$$\Delta = \Sigma^2 - 4(r_1 + r_2 + r_3)$$

in which

$$r_1 = be + bd + ce$$

$$r_2 = ad + ae + df$$

$$r_3 = cf + ac + bf$$

The eigenvectors are

$$\vec{v}_0 = \{r_1, r_2, r_3\}$$

$$\vec{v}_\pm = \{-\gamma_2^\pm - \gamma_3^\pm, \gamma_2^\pm, \gamma_3^\pm\}$$

where

$$\gamma_2^\pm = a\lambda_\pm + r_2$$

$$\gamma_3^\pm = f\lambda_\pm + r_3$$

It should be noted that since $\vec{v}_0$ is the steady-state solution, r_is are proportional to the the equilibrium population numbers, N_is, or the marginal probabilities, P_is, and by definition $K_2 = r_2/r_1$ and $K_3 = r_3/r_1$.

Large distinction of relaxation times implies an isolated state

By "isolated" we mean surrounded by slow rate constants. An explanation follows. We assume that the two eigenvalues $\lambda_{\pm} = -\frac{1}{2}(\Sigma \pm \sqrt{\Delta})$ are both real and negative, because, as explained in Section 7.1.1, an imaginary part would result in cyclic and hence irreversible reaction behavior which is not supported by our data. If the two relaxation times are different by order(s) of magnitude, i.e. $|\lambda_-| \ll |\lambda_+|$, we have $\Sigma - \sqrt{\Delta} \ll \Sigma + \sqrt{\Delta}$, or $\Sigma^2 - \Delta \ll (\Sigma + \sqrt{\Delta})^2$. This, using $\Delta = \Sigma^2 - 4(r_1 + r_2 + r_3)$, yields

$$r_1 + r_2 + r_3 \ll (\Sigma + \sqrt{\Delta})^2/4$$

The right hand side is smaller than Σ^2 (because $\sqrt{\Delta} < \Sigma$), thus

$$r_1 + r_2 + r_3 \ll \Sigma^2 \tag{7.12}$$

Since r_1, r_2 and r_3 are non-negative, Equation (7.12) implies that each is much smaller than Σ^2 .

Without loss of generality, we examine the case for $r_1 = be + bd + ce$. For r_1 to be small ($\ll \Sigma^2$) at least two of the rate constants involved in r_1, i.e. e, b, c, and d, must be small ($\ll \Sigma$). If only one of the rate constants is small (say e), a term (e.g bd) will remain non-small. Therefore at least two rate constants must be small. There are three possible ways to choose two rate constants that make r_1 small. Each of these ways isolates an state by leaving only the outward or inward rates to that state. These three ways are:

$$e, b \ll \Sigma \text{ (state 1 is isolated)}$$

$$e, d \ll \Sigma \text{ (state 3 is isolated)}$$

$$b, c \ll \Sigma \text{ (state 2 is isolated)}$$

An isolated state cannot be almost entirely full or entirely empty (data shows more than two detectable states), therefore for each small rate constant, its reverse rate constant must also be small

(by detailed balance). For example, if state 1 is isolated, a, b, e, and f must all be small. As a result, there is only one pair of fast reaction rate constants that might be within the timescale of FCS measurements, while the slow rate constants are still relevant in determining the equilibrium populations of all the states. For example, if state 1 is isolated,

$$\begin{aligned}\frac{r_2}{r_1} &= \frac{ad + ae + df}{be + bd + ce} \\ &\approx \frac{ad + df}{bd + ce}\end{aligned}$$

Detailed balance

For a three-state system, detailed balance provides two independent relations:

$$\frac{r_2}{r_1} = \frac{a}{b}$$

$$\frac{r_3}{r_1} = \frac{f}{e}$$

These two can yield the third relation as a dependent: Combining the two, we find

$$\frac{r_3}{r_2} = \frac{fb}{ea}$$

which together with $ace = bdf$ (obtainable from either of the two) gives

$$\frac{r_3}{r_2} = \frac{c}{d}$$

Supplementary example figure

To supplement Figure 6.2, in this section we show example calculated higher-order reaction curves for the regime in which the fast step of the reaction occurs between the brightest and the dimmest states, and the isolated state is the intermediate state. The calculated curves are shown in panel (d) of Figure (7.1). Panels (a) and (b) are replicated from Figure 6.2 for comparison. As described in Section 6.3, it is more convenient for the fitting process to define states 1 and 2 as always

being the fast-reacting states. When the fast step is between the dimmest and brightest states, this leads to the configuration shown in panel (c); we choose to keep a clockwise permutation, so state 1 becomes the dimmest state and state 2 is the brightest state, i.e. $q_2 \geq q_3 \geq 1$. The rate constants and the brightnesses values shown in panel (d) correspond to this definition.

The regime in which the fast step of the reaction occurs between the brightest and the dimmest states corresponds to the proposed "trap state" models, in which the pathway through the intermediate (trap) state either does not lead to the native folded state, or leads to it very slowly. As indicated by Figure 6.2(d), in this regime, the qualitative behavior of the higher-order reaction curves is that their separation is significantly and *asymmetrically* reduced after the fast relaxation time. This is not what we have observed in our experiments, which rules out this regime of rate constants between the observed states. However, the theories of trap states are not fully ruled out since the implicated time constants or brightness fluctuations in such theories may not have been measurable by our technique.

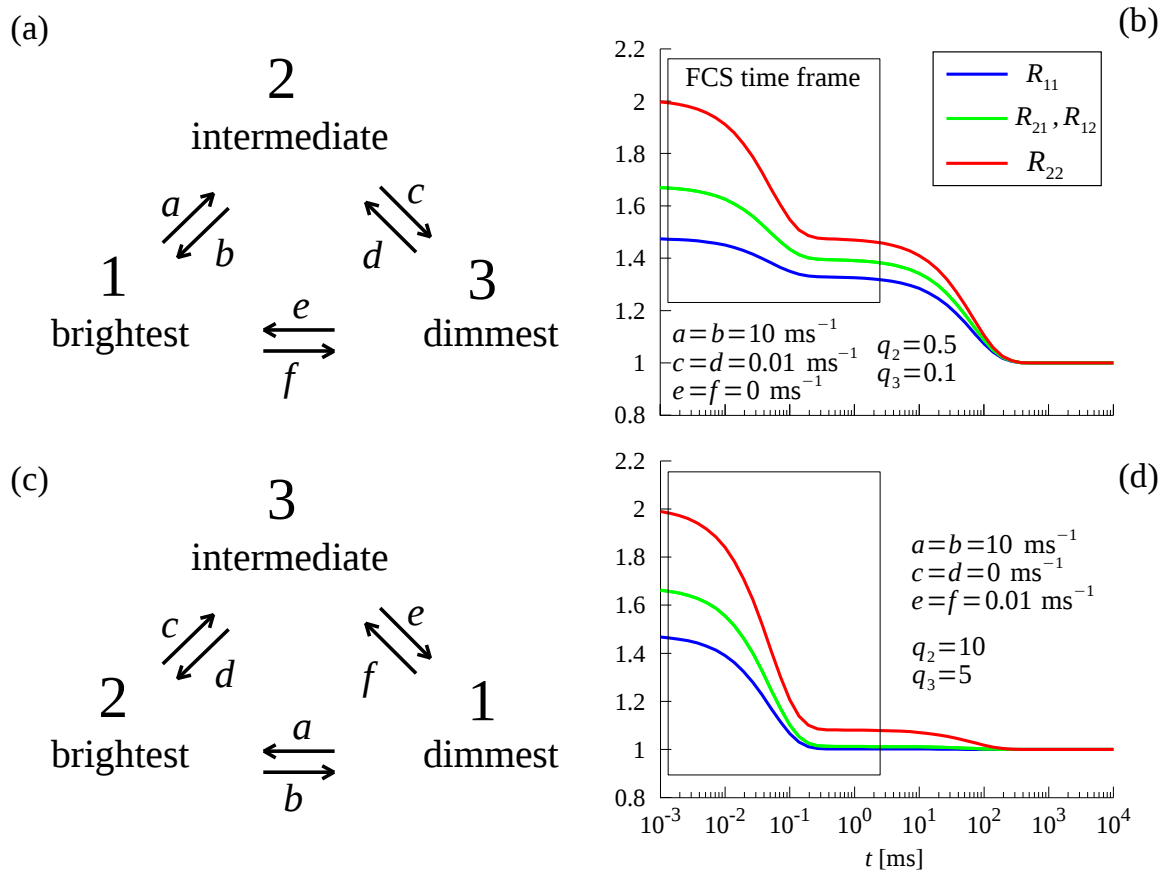

Figure 7.1: Three-state models (left) and calculated higher-order reaction functions for $N = 1$ and select rate constants and brightness values (right). (a): A three state model with the fast rate constants, a and b, between the brightest and intermediate states. (c): A three state model the fast rate constants, a and b, between the dimmest and brightest states. The states are numbered accordingly. (d, f): Example plots corresponding to the models on the left, calculated with the parameters shown on each graph.

Modeling free fluorophores as the static state

In this section we further examine the hypothesis that free fluorophores or contaminants of some sort are responsible for the third, static state. This would be very unlikely because the "contaminant" would have to have identical or very similar diffusion constant to that of the reference DNA in order to produce the observed flat (horizontal) baselines of the relative reaction functions. The short- and long-looped hairpins both yielded consistent results; a single type of contaminant cannot match the diffusion constants of both hairpins. Free fluorophores in general diffuse much

faster than the hairpin molecules. For more clarity, here we model free fluorophores mixed with a two-state hairpin mechanism and compare with data.

The derivation below follows steps similar to what was done for a single reacting species elsewhere [17, 63]. See those references for more details on each step.

We start more generally by considering a mixture of two non-interacting fluorescent species, A and B, that have different diffusion constants, D_{A} and D_{B}, respectively. Each species may undergo reaction between multiple fluorescent states (for example through conformational change or binding-unbinding with non-fluorescent molecules), but the fluorescent molecules do not interact with each other. Let us denote the fluorescence intensity collected at time t from the detection volume by $I(t)$ and define the two-time vector $\vec{I} := [I(0), I(t)]$. In a sample containing M_{A} and M_{B} molecules of each type, $\vec{I}$ is given by

$$\vec{I} = \sum_{j=1}^{M_{\mathrm{A}}} \vec{I}^{(1)}_{j,\mathrm{A}} + \sum_{j=1}^{M_{\mathrm{B}}} \vec{I}^{(1)}_{j,\mathrm{B}}$$

where $\vec{I}^{(1)}_{j,\mathrm{A}}$ is the fluorescence intensity vector due to the jth (single) molecule of type A, and similarly for B. Since the molecules are independent, the additive property of cumulants yields:

$$\kappa_{m,n}(\vec{I}) = M_{\mathrm{A}}\kappa_{m,n}[\vec{I}^{(1)}_{\mathrm{A}}] + M_{\mathrm{B}}\kappa_{m,n}[\vec{I}^{(1)}_{\mathrm{B}}]$$

where $\kappa_{m,n}(\vec{I})$ denotes the (m,n)th cumulant of $\vec{I}$. In the thermodynamical limit where the sample volume is large (compared to the detection volume), $M_{\mathrm{A}}, M_{\mathrm{B}} \to \infty$, we obtain

$$\kappa_{m,n}(\vec{I}) = M_{\mathrm{A}}\mu'_{m,n}[\vec{I}^{(1)}_{\mathrm{A}}] + M_{\mathrm{B}}\mu'_{m,n}[\vec{I}^{(1)}_{\mathrm{B}}]$$

in which $\mu'_{m,n}[\vec{I}^{(1)}_{\mathrm{A}}]$ denotes the (m,n)th moment of $\vec{I}^{(1)}_{\mathrm{A}}$. Assuming the diffusion constant of each species does not change significantly by reaction (i.e. independent reaction and diffusion processes), the above equation becomes

$$\kappa_{m,n}(\vec{I}) = \gamma_{m+n}[X_{m,n,\mathrm{A}}(t)Y_{m,n,\mathrm{A}}(t) + X_{m,n,\mathrm{B}}(t)Y_{m,n,\mathrm{B}}(t)] \tag{7.13}$$

where

$$X_{m,n,\mathrm{A}}(t) = \sum_{s=1}^{J_\mathrm{A}} \sum_{s'=1}^{J_\mathrm{A}} N_{s,\mathrm{A}} Q_{s,\mathrm{A}}^m Q_{s',\mathrm{A}}^n Z_{s',s,\mathrm{A}}(t)$$

is the reaction factor for species A. $Y_{m,n,\mathrm{A}}(t)$ depends on the molecular detection function (MDF) and the diffusion constant of species A, D_A. Functions with subscript B similarly apply to species B. γ_{m+n} only depends on the MDF. The explicit integral forms of $Y_{m,n}(t)$ and γ_{m+n} can be found elsewhere [17, 63].

The univariate cumulants

$$\kappa_{m,0}(\vec{I}) = \kappa_{0,m}(\vec{I}) = \gamma_m(\sum_{s=1}^{J_\mathrm{A}} N_{s,\mathrm{A}} Q_{s,\mathrm{A}}^m + \sum_{s=1}^{J_\mathrm{B}} N_{s,\mathrm{B}} Q_{s,\mathrm{B}}^m) \tag{7.14}$$

can be obtained for this mixture in a similar fashion.

The normalized correlation functions, $g_{m,n}(t)$, generally depend on the the bin size (sampling interval), T. The bin size effects are usually negligible for small bin sizes. In the limit $T \to 0$, we have

$$g_{m,n}(t) = \frac{\kappa_{m,n}(\vec{I})}{\kappa_{m,0}(\vec{I})\kappa_{0,n}(\vec{I})} \tag{7.15}$$

These functions can be expressed for the mixture of A and B by substituting Equations (7.13) and (7.14) into (7.15).

As a special case, the resulting $g_{m,n}(t)$ functions can be written for a mixture containing DNA hairpins which undergo conformational change between two states (species A) and independent fluorophores (impurity) which remain in a single state. For extracting the relative reaction functions we normalize $g_{m,n}(t)$ from the test sample by that from a reference sample, $g_{m,n}^{(\mathrm{ref})}(t)$. In this case the reference sample has the same diffusion constant as that of the DNA hairpin (the reference sample may be DNA hairpin without quencher, or pure fluorophore with time rescaled to match

that of DNA as in Section 7.2.3.) The result of such normalization in this case is

$$R_{m,n}^{(\mathrm{rel})*} := \frac{g_{m,n}(t)}{g_{m,n}^{(\mathrm{ref})}(t)} = \frac{(\overleftarrow{1}\vec{K})}{N} \frac{\overleftarrow{1}\mathbf{Q}^n\mathbf{Z}^*(t)\mathbf{Q}^m\vec{K}}{(\overleftarrow{1}\mathbf{Q}^m\vec{K})(\overleftarrow{1}\mathbf{Q}^n\vec{K})}$$

where we have used notation similar to Equation (6.5) (the fluorophore is element 3 in the matrices, with parameters q_3 and K_3). We have

$$\mathbf{Z}^*(t) = \frac{1}{1+K_2} \begin{bmatrix} 1+K_2e^{-t/t_\mathrm{F}} & 1-e^{-t/t_\mathrm{F}} & 0 \\ K_2(1-e^{-t/t_\mathrm{F}}) & K_2+e^{-t/t_\mathrm{F}} & 0 \\ 0 & 0 & (1+K_2)\frac{Y_{m,n,\mathrm{dye}}(t)}{Y_{m,n,\mathrm{DNA}}(t)} \end{bmatrix} \tag{7.16}$$

which differs from $\mathbf{Z}(t)$ (Equation (6.7)) only in the third diagonal (the "static") element. This element now includes the ratio of $Y_{m,n}(t)$ of the fluorophore (dye) to that of the DNA hairpin, which is not canceled out by normalization by the reference since the two species have different diffusion times.

To compare with data, we estimate $Y_{m,n}(t)$ (for either species) by assuming a Gaussian beam profile. This yields [17]

$$Y_{m,n}(t) = \frac{1}{\left(1+\frac{2mn}{m+n}\frac{t}{t_\mathrm{D}}\right)} \frac{1}{\left(1+\frac{2mn}{m+n}\frac{t}{\alpha^2 t_\mathrm{D}}\right)^{1/2}}$$

where t_D is the characteristic diffusion time of either species through the detection volume, and α is the beam elongation in the z direction.

In Figure 7.2(left) we have used second-order correlation functions to obtain the diffusion parameters for a free fluorophore (TAMRA; rhodamine derivatives have similar diffusion coefficients [117, 118]) and R6G-labeled T_{21}-loop hairpin (sample RT21 which has no quencher). The R6G label in the DNA sample also shows a fast photoinduced triplet-blinking component which is modeled as a two-state mechanism [131] to improve the fit quality. (The triplet blinking com-

ponent is almost entirely canceled out when the test and reference samples have the same type of fluorophore). The fit parameters relevant to $Y_{m,n}(t)$ are

$$t_{\mathrm{D,TAMRA}} = 85.3 \pm 0.2\,\mu\mathrm{s}$$
$$t_{\mathrm{D,DNA}} = 346.6 \pm 0.6\,\mu\mathrm{s}$$
$$\alpha = 4.3 \ \ (\mathrm{held})$$

α was held at 4.3 for the fits presented in Figure 7.2, since two distinct values of 3.2 and 5.5 were obtained from fitting TAMRA and DNA curves individually.

In the right panel of Figure 7.2 we present example data from salt dependence measurements (see figure caption for details) which is fitted using the three-state model as developed in Section 6.3 (dashed black lines). The resulting parameters where then used to plot $R_{m,n}^{(\mathrm{rel})*}$ developed in this section, with $K_3 = N_{\mathrm{dye}}/N_{1,\mathrm{DNA}}$ in this case representing the ratio of the fluorophore (contaminant) concentration to the concentration of DNA in state 1. The resulting functions are dramatically and qualitatively different from the data; most notably, the baselines continue decaying and get closer to each other at longer lag times.

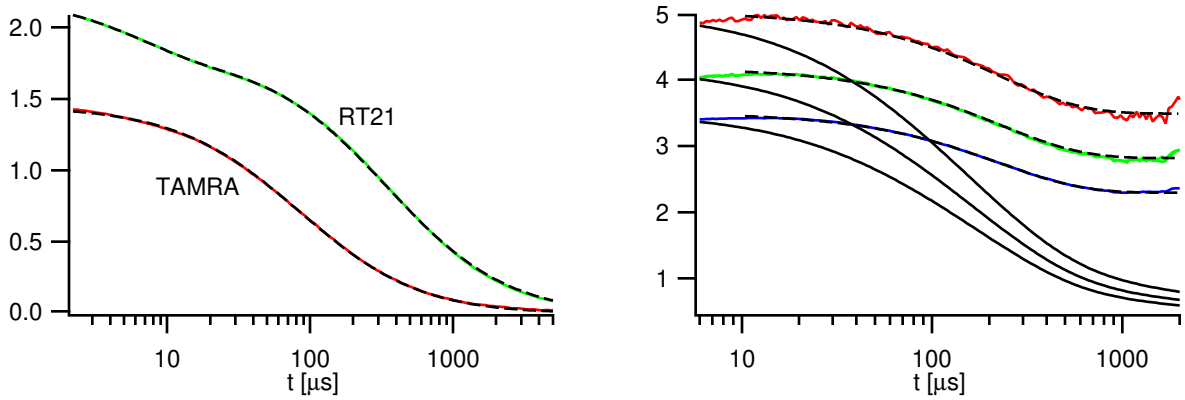

Figure 7.2: Modeling free fluorophores as the "static" state. Left: Second-order correlation functions, $g_{1,1}(t)$, for TAMRA (red) and reference hairpin sample RT21 (green) are fitted (dashed black) to find their diffusion times. Right: Relative reaction functions, $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves using the three-state model with $\mathbf{Z}(t)$ as in Equation (6.7) (dashed black). The solid black curves are calculated using $\mathbf{Z}^*(t)$ as developed in this section (Equation (7.16)) with the same brightness and population parameters obtained from fitting with $\mathbf{Z}(t)$. Data is shown for sample RT21D (reference RT21) at 85 mM NaCl (trial 3 in salt-dependence measurements, see Figure 7.20(e)).

7.2 Experiment

7.2.1 Correlation computation modifications

Two modifications were made in the previously reported correlation computation method [115] for improved results. First, in a new bin size selection scheme, the bin size, δ, grows with lag time, t, from a starting size of $\delta_0 = 2\,\mu$s to a maximum size of $\delta_\mathrm{m} = 10\,\mu$s through the relation

$$\delta = \delta_\mathrm{m} + (\delta_0 - \delta_\mathrm{m})e^{-\rho t/(\delta_\mathrm{m}-\delta_0)}$$

where $\rho = 0.08$ consistent with the previous research. The bin size needs to be capped at a maximum value much smaller than the reaction time scale for reactions that extend to large lag times (the case for hairpin folding dynamics), because even though keeping the bin size much smaller than the lag time keeps the absolute uncertainties small as shown in [115], the *relative* systematic errors caused by finite bin size (on the order of the reaction time scale) become amplified upon normalization by the small correlation values from the reference sample at larger lag times.

Luckily if the reaction of interest occurs at larger lag times (say $100\,\mu$s) there is sufficient room at smaller lag times (up to $10\,\mu$s) to select bin sizes yielding good signal-to-noise ratio.

The second modification is correction for background noise (measured from blank buffer) in computing the correlation functions of order $(1,1)$, $(2,1)$, and $(1,2)$ which involve the mean integrated intensity, $\langle W \rangle$, (or mean photon count $\langle n \rangle$, in a bin) in their computation (Equations (2.17)–(2.19)). This was done by simply subtracting the mean background count from the mean signal count in computing correlations up to third order. Correlations of higher orders are not systematically (but only stochastically) affected by the uncorrelated background noise, therefore no correction was applied to order-$(2,2)$ curves.

7.2.2 Trying the "dark sink" class of models

Forcing a fit using the assumption that the fast-reacting states (states 1 and 2) are the brightest and the intermediate states clearly fails. This is done by ensuring that $0 \leq q_3 \leq q_2 \leq 1$. Figure 7.3 shows an attempt using sample RT8D. To assist the fitting procedure, global fit was attempted using only the first 4 temperatures, and the values of the relaxation times were found from $R_{1,1}(t)$ and constrained. The value of q_3 (in this model: the lowest brightness, of state 3, relative to the highest brightness, of state 1) was held at a very low value of 0.01. From the (failed) fit, $q_2 = 0.30 \pm 0.01$ was obtained, confirming the class of models being investigated. If we set $q_3 = 0$ (fully dark stationary population), the curves in each graph converge to the same baseline, because the model effectively describes a two-state system at each temperature. (In comparison, the successful fitting model looks like a binary mixture, with one species blinking. This supports largely separate baselines). Some population parameters (N and K_is) obtained from these attempts are unphysical or extreme. For example K_is are extremely large at 22.5°C in Figure 7.3 (the reaction is sacrificed to keep the baselines apart). If we make q_3 slightly larger than 0.01, some population parameters become even negative.

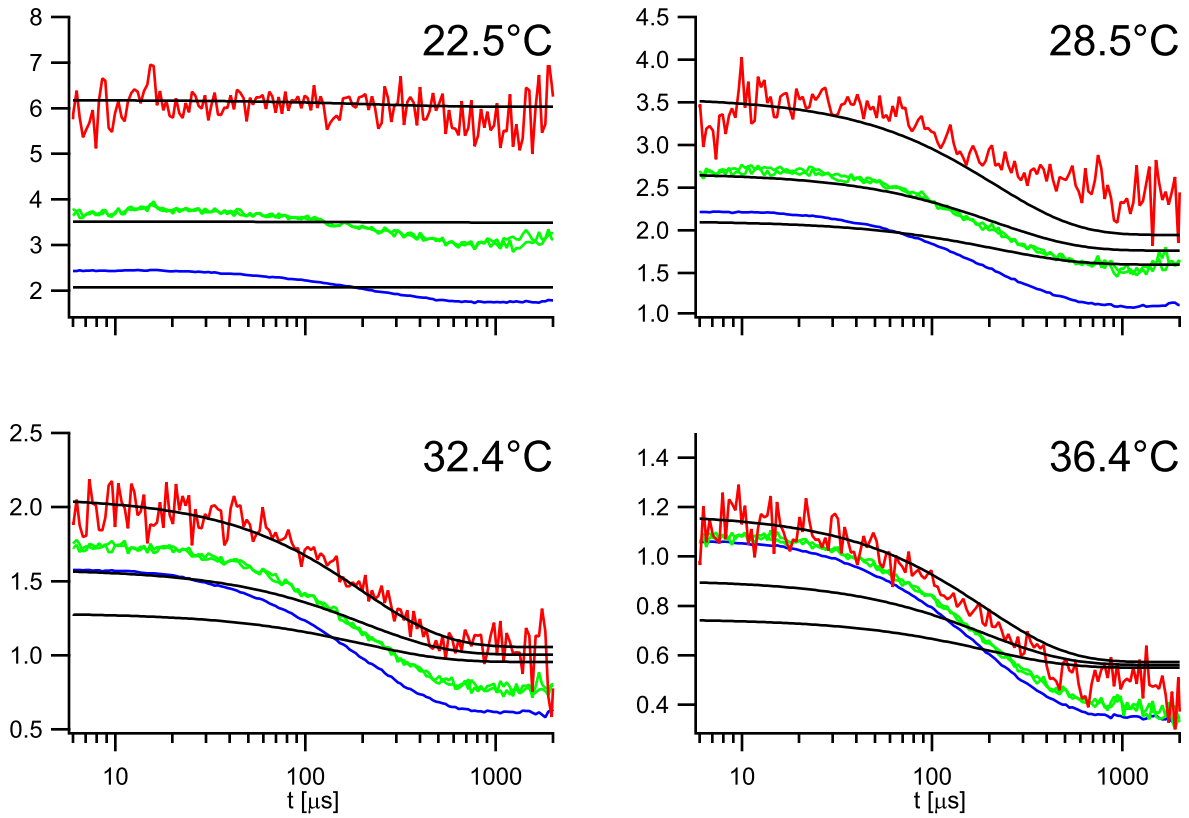

Figure 7.3: Plots showing a forced global fit using a three-state model assuming the fast reaction step between the brightest and the intermediate states. Data are shown for sample RT8D (reference RT8).

7.2.3 Temperature dependence measurements

TT8D melting and temperature calibration

We used the decay in the fluorescence of TAMRA as a function of temperature in the single-labeled (reference) hairpin sample TT8 to calibrate the setup temperature readings. The cross marks in Figure 7.4(a) show the data from our setup as a function of the uncalibrated temperature, T_{UC}. Figure 7.4(b) shows the data from a commercial calibration-verified spectrofluorometer (model ATF 105, Aviv Biomedical, inc., Lakewood, NJ). To obtain the true temperature, T, from the uncalibrated temperature, T_{UC}, the following relation was assumed:

$$T[^{\circ}\mathrm{C}] = T_{\mathrm{UC}}[^{\circ}\mathrm{C}] - \alpha(T_{\mathrm{UC}}[^{\circ}\mathrm{C}] - 22.5) \tag{7.17}$$

To find α, both data sets were normalized to their value at room temperature then Equation (7.17) was used to fit the data from the fluorometer to that from the setup, as shown in Figure 7.4(c) (linear interpolation was used). The value was found as $\alpha = -0.206 \pm 0.007$.

Additionally, the melting curves from the dual-lebeled (test) sample TT8D were measured using both the setup and the machine (square marks in Figures 7.4(a) and (b)). To remove the effects caused by the dependence of the fluorescence of TAMRA on temperature, data were normalized by those from single-labeld measurements and scaled back (for visualization) by a value near room temperature (circle marks in the Figures). Then a two-state model was used to fit the corrected melting data:

$$I(T) = \frac{I_{\min} + rI_{\max}}{1 + r} \tag{7.18}$$

where

$$r = \exp\left[-\frac{\Delta H}{RT}(1 - \frac{T}{T_{\mathrm{m}}})\right]$$

is the ratio of the unfolded to the folded populations. From the commercial fluorometer we found $T_{\mathrm{m}} = 43.29 \pm 0.04\,^{\circ}\mathrm{C}$ and $\Delta H = 44.2 \pm 0.3\,\mathrm{kcal/mol}$. From the uncalibrated setup data we found $T_{\mathrm{m,UC}} = 48.5 \pm 0.1\,^{\circ}\mathrm{C}$ and after calibration using Equation (7.17) with $\alpha = -0.206$ we found $T_{\mathrm{m}} = 43.2 \pm 0.1\,^{\circ}\mathrm{C}$ and $\Delta H = 42.8 \pm 1.0\,\mathrm{kcal/mol}$, verifying the calibration.

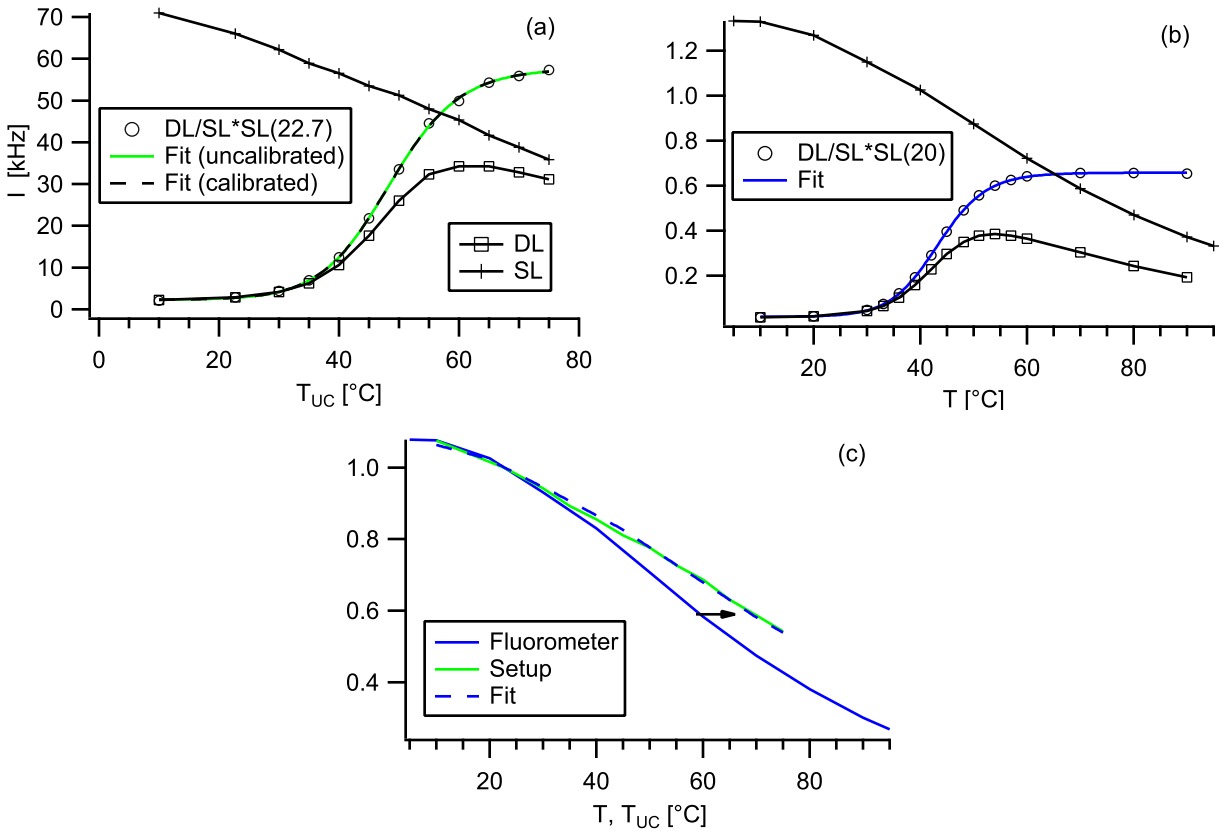

Figure 7.4: Fluorescence from single-labeled (SL, sample TT8) and dual-labeled (DL, sample TT8D) measured as a function of temperature using (a) our setup and (b) a commercial fluorometer. The circles show the values of the DL curve divided by those of SL curve to remove the dependence of the fluorescence of TAMRA alone on temperature, then multiplied by a value of the SL curve near room temperature for better visualization. The values where then fitted to Equation (7.18) to obtain the melting temperatures using uncalibrated and calibrated temperatures (for calibration verification) (c) To calibrate the setup, the fluorescence of the single-labeled sample from the commercial fluorometer was used to fit the corresponding data from the setup, using Equation (7.17) and interpolation.

Sample TT8D

Table 7.1: Global fit parameters directly measured by fitting to the higher-order reaction functions shown in Figures 6.4, 7.5, and 7.9. The "linked" parameter is shared by the fit functions at all temperatures (global fit). In all tables, numbers in parentheses indicate fitting uncertainty in the corresponding last digits. A q_2 parameter was held at 0, otherwise a small negative value within the fitting uncertainty would be obtained.

		RT8D	TT8D	TT8D/T
linked	N	5.32(58)	11.9(1.1)	16.0(6)
22.5°C	K_2	34.2(4.3)	65.5(7.5)	75.1(3.7)
	q_2	0.00388(62)	0.0204(21)	0.0156(7)
	K_3	0.568(28)	4.01(18)	4.66(12)
	q_3	1.96(5)	1.13(2)	1.00(1)
	t_F	225.2(2.6)	158.4(2.8)	139.4(1.3)
28.5°C	K_2	22.5(2.8)	30.5(3.3)	28.1(1.4)
	q_2	0.00222(32)	0.0124(13)	0.0163(7)
	K_3	0.423(12)	1.97(5)	1.74(4)
	q_3	1.40(2)	1.04(1)	1.09(1)
	t_F	213.3(1.1)	176.7(1.2)	162.2(8)
32.4°C	K_2	11.3(1.5)	13.2(1.5)	13.1(7)
	q_2	0.00177(34)	0.0119(13)	0.0187(9)
	K_3	0.247(13)	0.811(28)	0.805(19)
	q_3	1.36(2)	1.12(1)	1.13(1)
	t_F	202.5(1.0)	185.9(8)	183.1(9)
36.4°C	K_2	5.61(87)	6.03(81)	5.41(33)
	q_2	0.00010(49)	0.0092(11)	0.0235(11)
	K_3	0.135(22)	0.399(33)	0.292(17)
	q_3	1.32(3)	1.07(1)	1.21(1)
	t_F	189.5(8)	191.9(5)	178.7(5)
40.4°C	K_2	2.76(54)	2.31(4)	2.52(16)
	q_2	0(held)	0.0136(11)	0.0209(23)
	K_3	0.071(45)	0.047(52)	0.086(16)
	q_3	1.29(10)	1.31(21)	1.25(held)
	t_F	167.2(6)	183.2(6)	166.4(6)

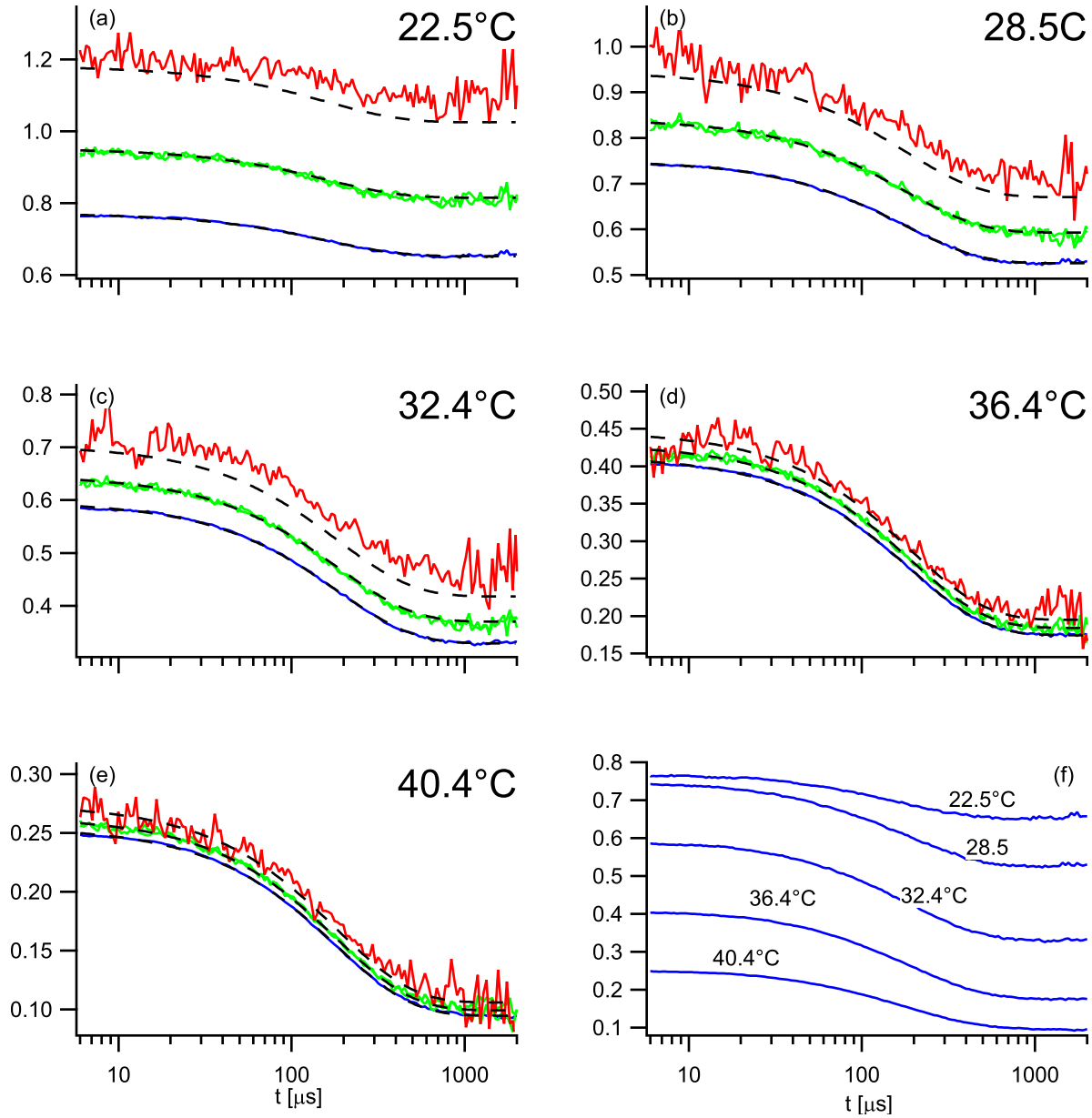

Figure 7.5: Sample TT8D: (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different temperatures. (f) The second-order curves, $R_{11}(t)$, shown together on the same y-axis for better comparison.

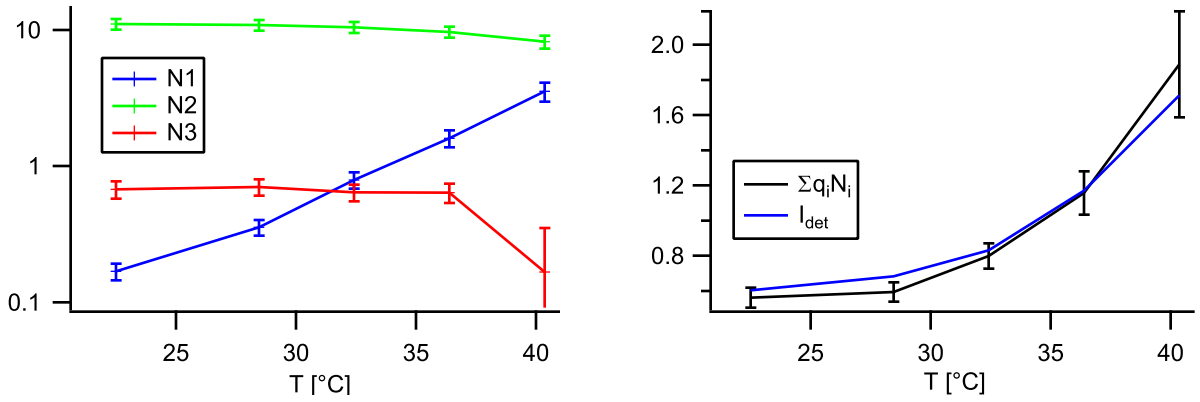

Figure 7.6: Sample TT8D: Left: Population numbers for the unfolded (N_3), intermediate (N_1) and folded (N_2) states as a function of temperature. Right: The overall brightness of the molecules as predicted by the fit parameters (black) compared against the measured detector counts (blue). Uncertainties in the latter are considered relatively small hence not shown. The curves are normalized by their average values.

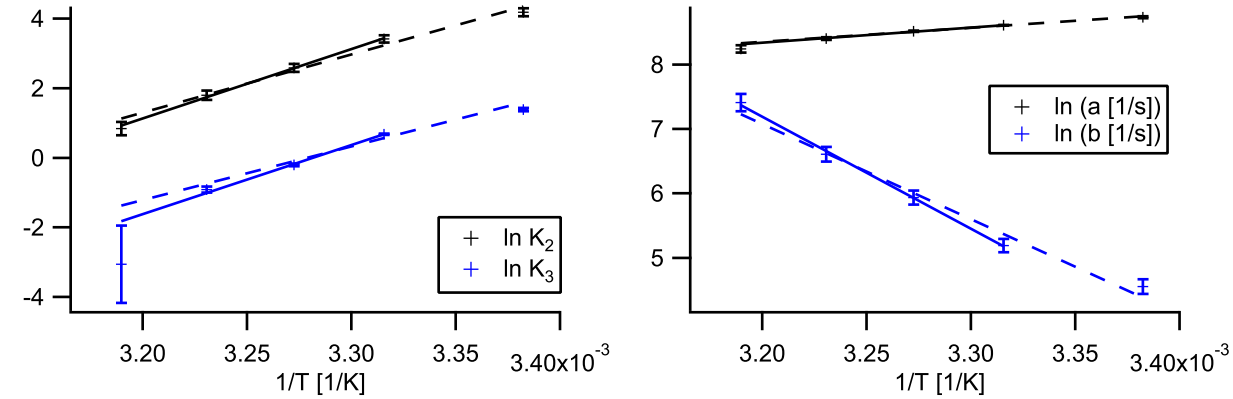

Figure 7.7: Sample TT8D: Left: Van 't Hoff plots of population ratios K_2 (folded/intermediate) and K_3 (unfolded/intermediate). Right: Arrhenius plots of the rate constants, a and b, of the fast step of the reaction (between folded and intermediate states). Linear fits on both graphs are using the four highest temperatures (solid) and all points (dashed).

Table 7.2: Sample TT8D: Enthalpy change, entropy change, activation energy, and the pre-exponential factors obtained from the van 't Hoff and Arrhenius plots in Figure 7.7.

		22.5–40.4°C	28.5–40.4°C
K_2	$\Delta H_2[\mathrm{kcal\,mol^{-1}}]$	$-33.2(1.8)$	$-39.7(3.0)$
	$\Delta S_2[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.104(6)$	$-0.125(10)$
K_3	$\Delta H_3[\mathrm{kcal\,mol^{-1}}]$	$-30.6(9)$	$-39.4(1.5)$
	$\Delta S_3[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.100(3)$	$-0.129(5)$
a	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	$-4.35(28)$	$-4.70(38)$
	$\ln(A[s^{-1}])$	$1.35(46)$	$0.77(63)$
b	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	$29.4(1.6)$	$34.5(2.5)$
	$\ln(A[s^{-1}])$	$54.4(2.6)$	$62.8(4.1)$

Sample TT8D/T

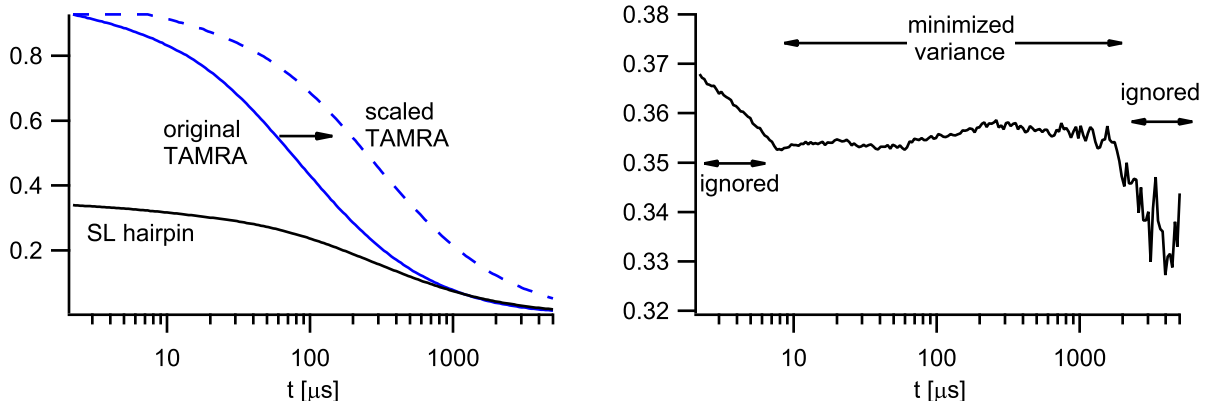

Figure 7.8: Preparing reference for sample TT8D/T. Left: Lag time in $g_{11}(t)$ from TAMRA (solid blue) is scaled (shifted in logarithmic scale) to obtain the dashed blue curve which matches the diffusion time of the single-labeled (SL) hairpin (solid black). To do this, $g_{11}(t)$ of scaled TAMRA is normalized by $g_{11}(t)$ of SL hairpin (right panel) then its variance is minimized over the fitting region.

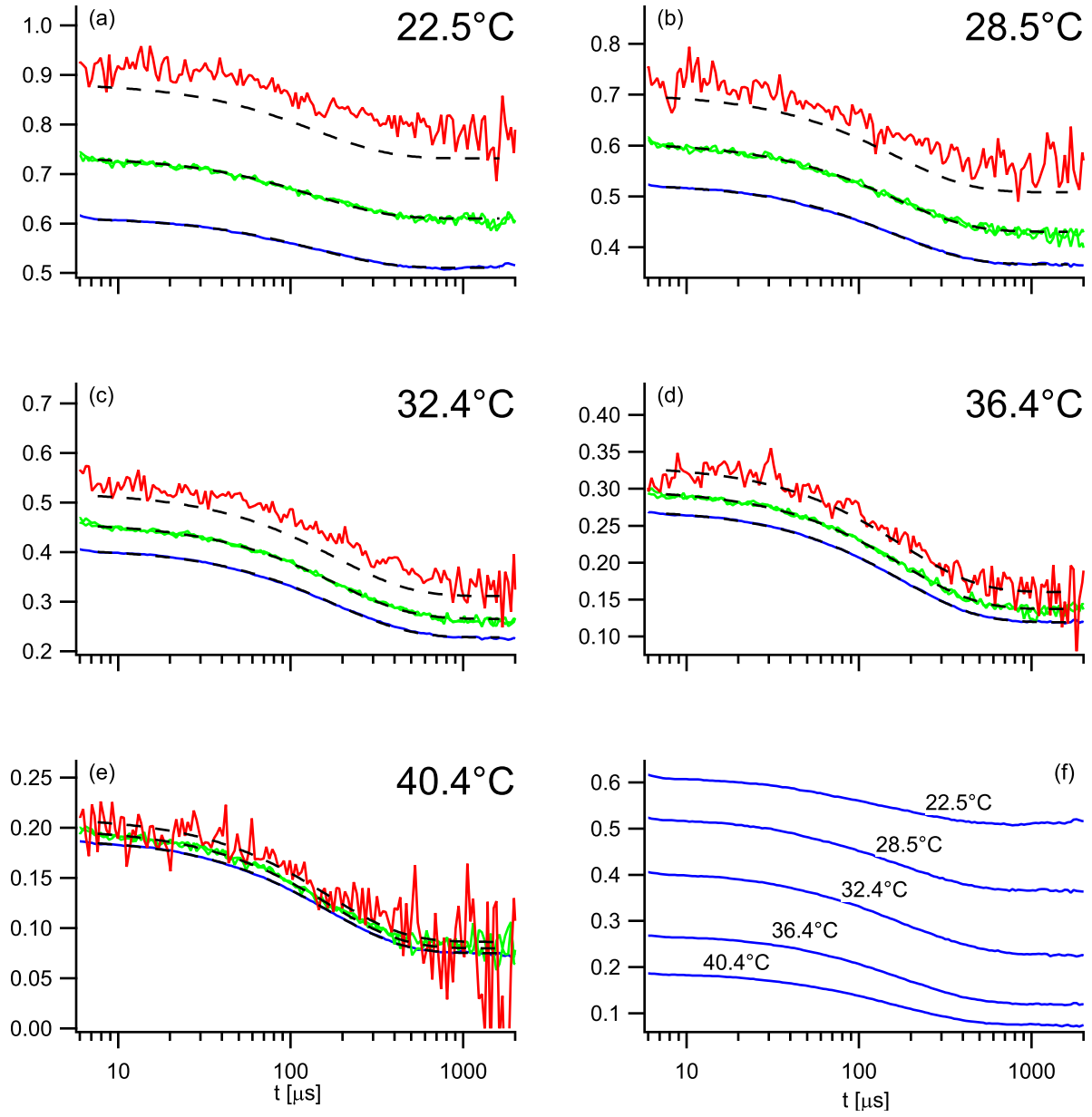

Figure 7.9: Sample TT8D/T: (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different temperatures. (f) The second-order curves, $R_{11}(t)$, shown together on the same y-axis for better comparison.

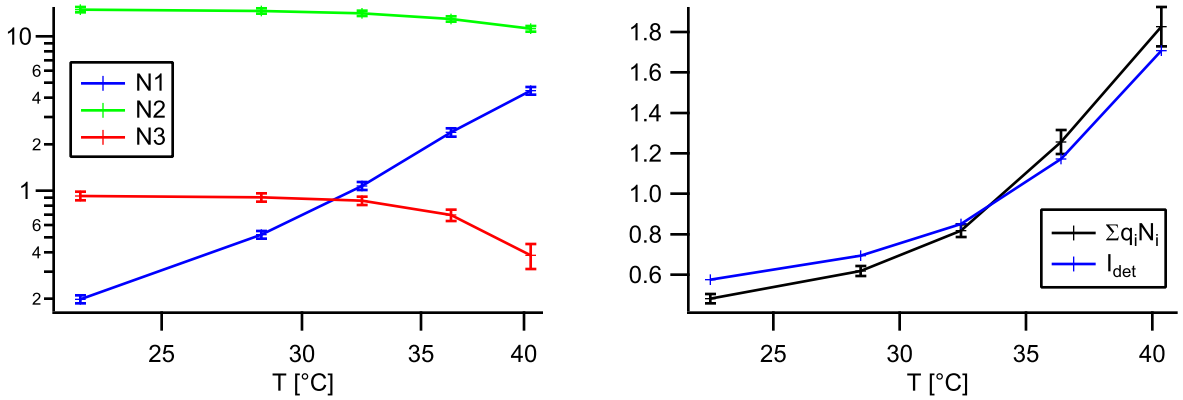

Figure 7.10: Sample TT8D/T: Left: Population numbers for the unfolded (N_3), intermediate (N_1) and folded (N_2) states as a function of temperature. Right: The overall brightness of the molecules as predicted by the fit parameters (black) compared against the measured detector counts (blue). Uncertainties in the latter are relatively small hence not shown. The curves are normalized by their average values.

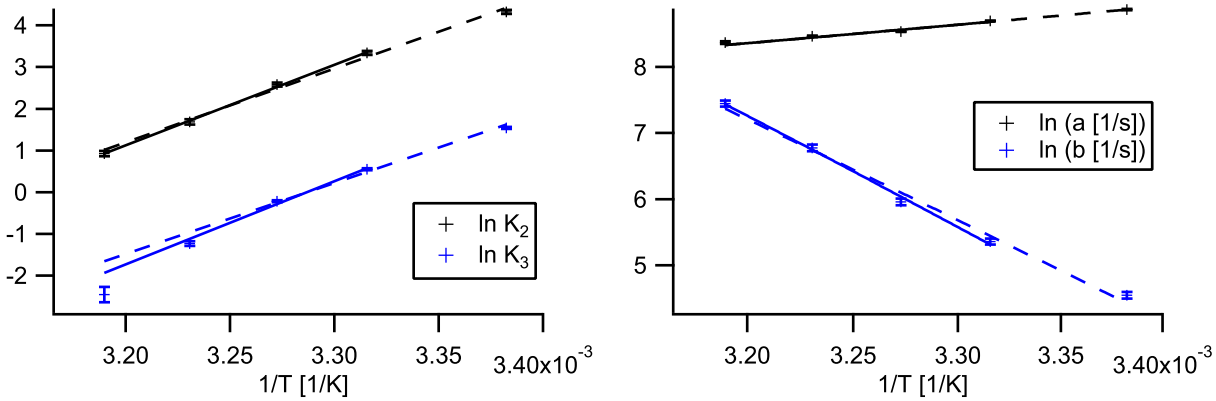

Figure 7.11: Sample TT8D/T: Left: Van 't Hoff plots of population ratios K_2 (folded/intermediate) and K_3 (unfolded/intermediate). Right: Arrhenius plots of the rate constants, a and b, of the fast step of the reaction (between folded and intermediate states). Linear fits on both graphs are using the four highest temperatures (solid) and all points (dashed).

Table 7.3: Sample TT8D/T: Enthalpy change, entropy change, activation energy, and the pre-exponential factors obtained from the van 't Hoff and Arrhenius plots in Figure 7.11.

		22.5–40.4°C	28.5–40.4°C
K_2	$\Delta H_2[\mathrm{kcal\,mol^{-1}}]$	$-35.1(7)$	$-38.3(1.2)$
	$\Delta S_2[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.110(2)$	$-0.120(4)$
K_3	$\Delta H_3[\mathrm{kcal\,mol^{-1}}]$	$-33.8(6)$	$-39.7(1.1)$
	$\Delta S_3[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.111(2)$	$-0.130(4)$
a	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	$-5.54(14)$	$-5.55(20)$
	$\ln(A[s^{-1}])$	$-0.56(24)$	$-0.57(34)$
b	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	$30.4(6)$	$33.5(1.0)$
	$\ln(A[s^{-1}])$	$56.1(1.0)$	$61.1(1.6)$

Sample RT21D

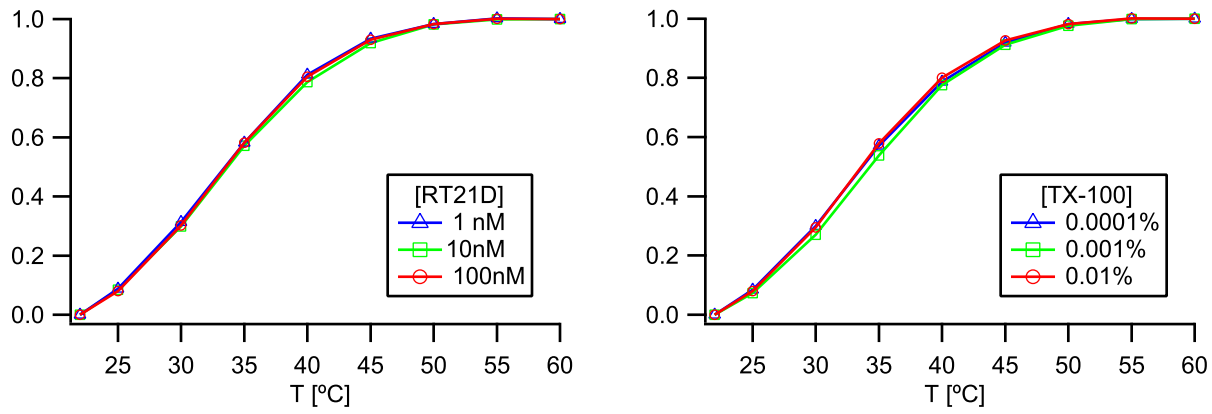

Figure 7.12: Melting curves for sample RT21D (in 54mM NaCl) scaled to fill the interval $[0, 1]$. Left: Different hairpin concentrations at $10^{-4}\%$ Triton X-100 concentration, showing no hairpin dimerization. Right: Various Triton X-100 concentrations at $10\,\mathrm{nM}$ hairpin concentration, showing no effects by the detergent.

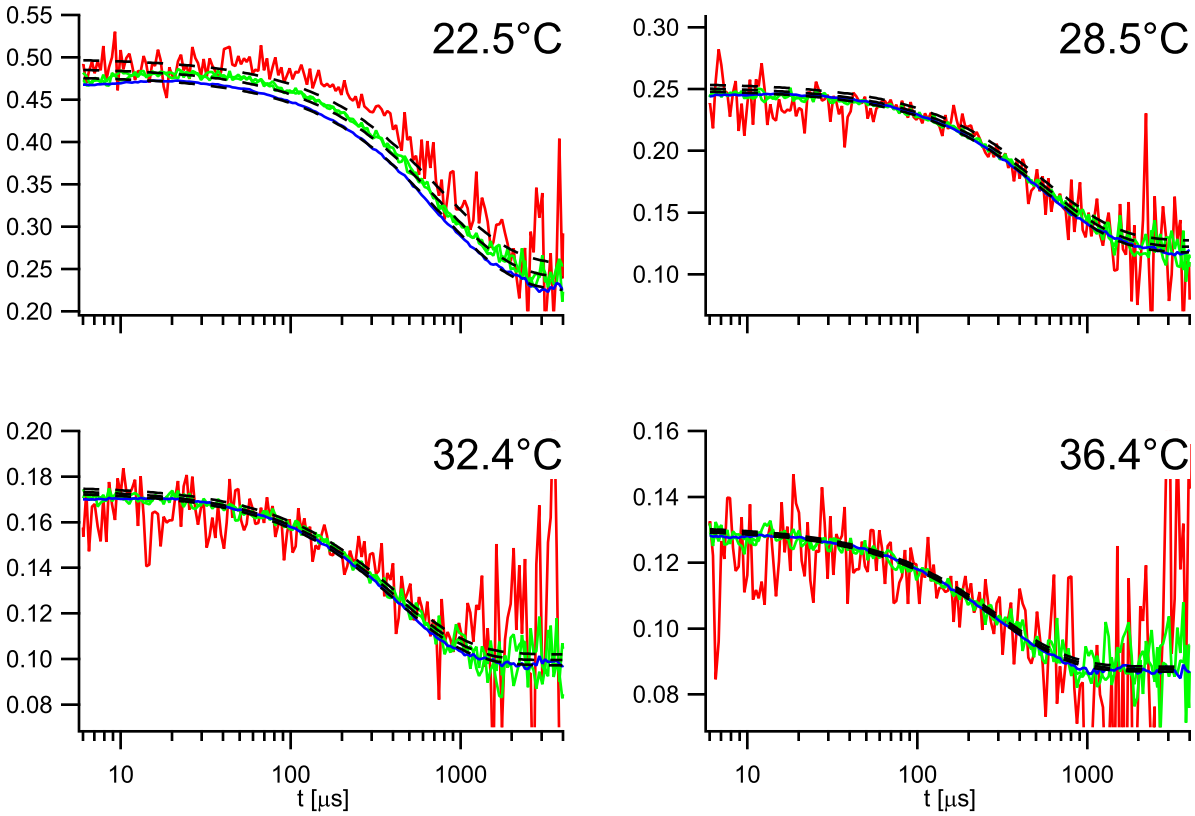

Figure 7.13: Sample RT21D: (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different temperatures.

Table 7.4: Sample RT21D: Global fit parameters directly measured by fitting to the higher-order reaction functions shown in Figure (7.13).The "linked" parameters are shared by the fit functions at all temperatures (global fit).

linked	N	$11.9(3)$
	q_2	$0.00372(35)$
	q_3	$1.12(1)$
$22.5°\mathrm{C}$	K_2	$7.34(3)$
	K_3	$0.517(16)$
	t_F	$749.8(5.1)$
$28.5°\mathrm{C}$	K_2	$2.55(12)$
	K_3	$0.280(19)$
	t_F	$577(3.7)$
$32.4°\mathrm{C}$	K_2	$1.30(9)$
	K_3	$0.225(35)$
	t_F	$423.7(3.7)$
$36.4°\mathrm{C}$	K_2	$0.599(78)$
	K_3	$0.098(74)$
	t_F	$297.8(3.2)$

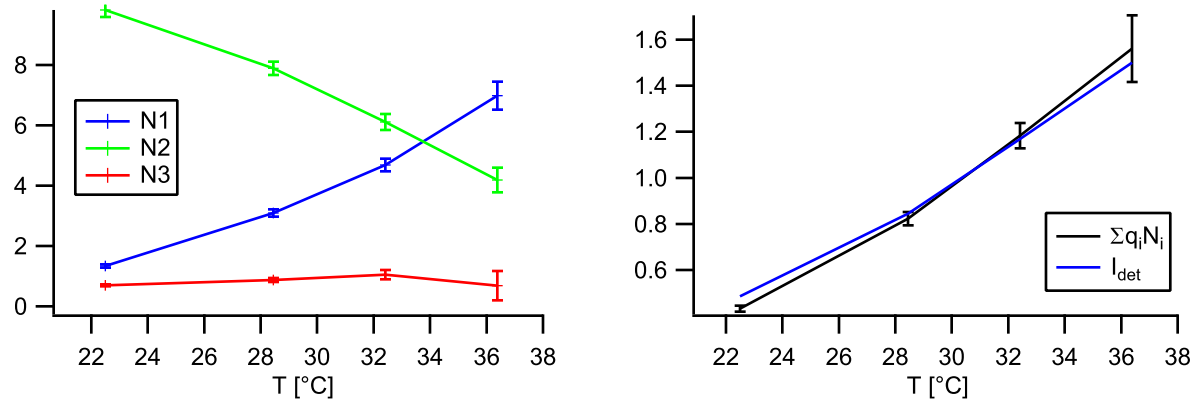

Figure 7.14: Sample RT21D: Left: Population numbers for the unfolded (N_3), intermediate (N_1) and folded (N_2) states as a function of temperature. Right: The overall brightness of the molecules as predicted by the fit parameters (black) compared against the measured detector counts (blue). Uncertainties in the latter are relatively small hence not shown. The curves are normalized by their average values.

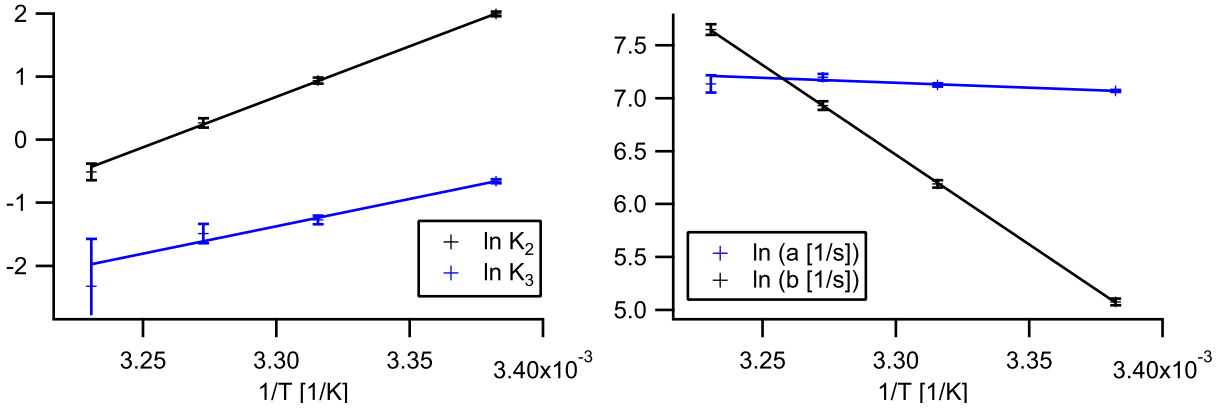

Figure 7.15: Sample RT21D: Left: Van 't Hoff plots of population ratios K_2 (folded/intermediate) and K_3 (unfolded/intermediate). Right: Arrhenius plots of the rate constants, a and b, of the fast step of the reaction (between folded and intermediate states). Linear fits on both graphs are obtained using all points.

Table 7.5: Sample RT21D: Enthalpy change, entropy change, activation energy, and the pre-exponential factors obtained from the van 't Hoff and Arrhenius plots in Figure 7.15.

K_2	$\Delta H_2[\mathrm{kcal\,mol^{-1}}]$	$-24.4(9)$
	$\Delta S_2[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.0786(29)$
K_3	$\Delta H_3[\mathrm{kcal\,mol^{-1}}]$	$-13.0(1.3)$
	$\Delta S_3[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	$-0.0454(45)$
a	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	$1.87(39)$
	$\ln(A[s^{-1}])$	$10.2(7)$
b	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	$33.7(7)$
	$\ln(A[s^{-1}])$	$62.4(1.2)$

Sample TT21G

Sample TT21G is labeled with TAMRA on the 5' end, but has a deoxyguanosine instead of dabcyl at the 3' end. The terminal guanine causes a much weaker quenching compared to dabcyl. The resulting reaction amplitudes are smaller and therefore more susceptible to concentration fluctuations, photochemical effects, and impurities. Figure 7.16 shows the second-order reaction functions (right) compared to the R6G-dabcyl-labeled counterpart (left). The terminal dG might have introduced a new component to the reaction responsible for the fast relaxation time observed around $10\,\mu$s, which seems larger and slower than typical photochemical effects (such effects are canceled out to a large extent by normalization to the reference sample). The uncertainty in this data set is consequently higher than others. The fitting results were found to be somewhat more consistent when the N values were individually fitted (unlinked) and the q values were linked instead (any possible temperature dependence in q's was found to be smaller than the uncertainty), which is the way presented in this section.

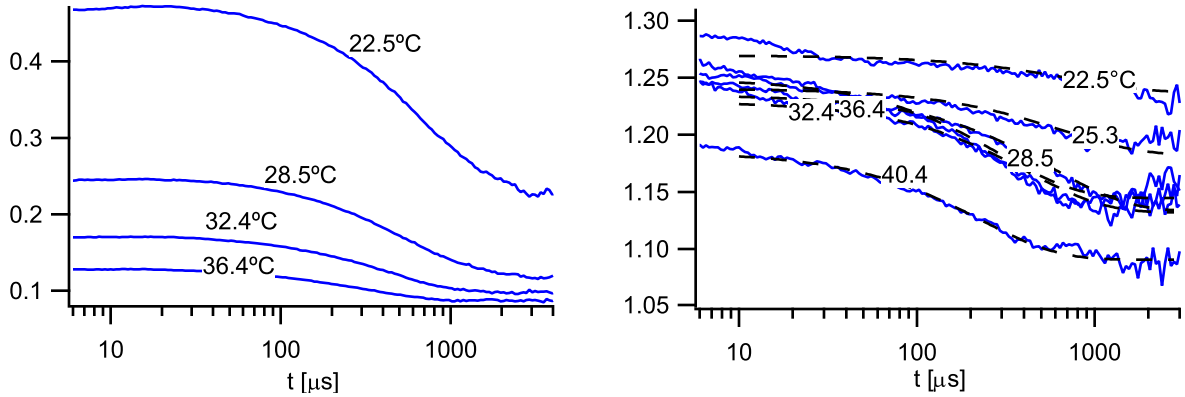

Figure 7.16: Samples RT21D and TT21G: The second-order reaction curves, $R_{11}(t)$, from Figures 7.13 (left) and 7.17 (right) shown together for better comparison.

It may be noted in Figure 7.17 that the separation between the higher-order curves gets smaller as they decay to their baselines. This feature is shared by both classes of models where the fast reaction is intermediate $\rightleftharpoons$ dimmest and brightest $\rightleftharpoons$ intermediate. (But the curves getting farther apart seems to be unique to the former class.) The q values obtained from fitting are the reliable identifiers of each class.

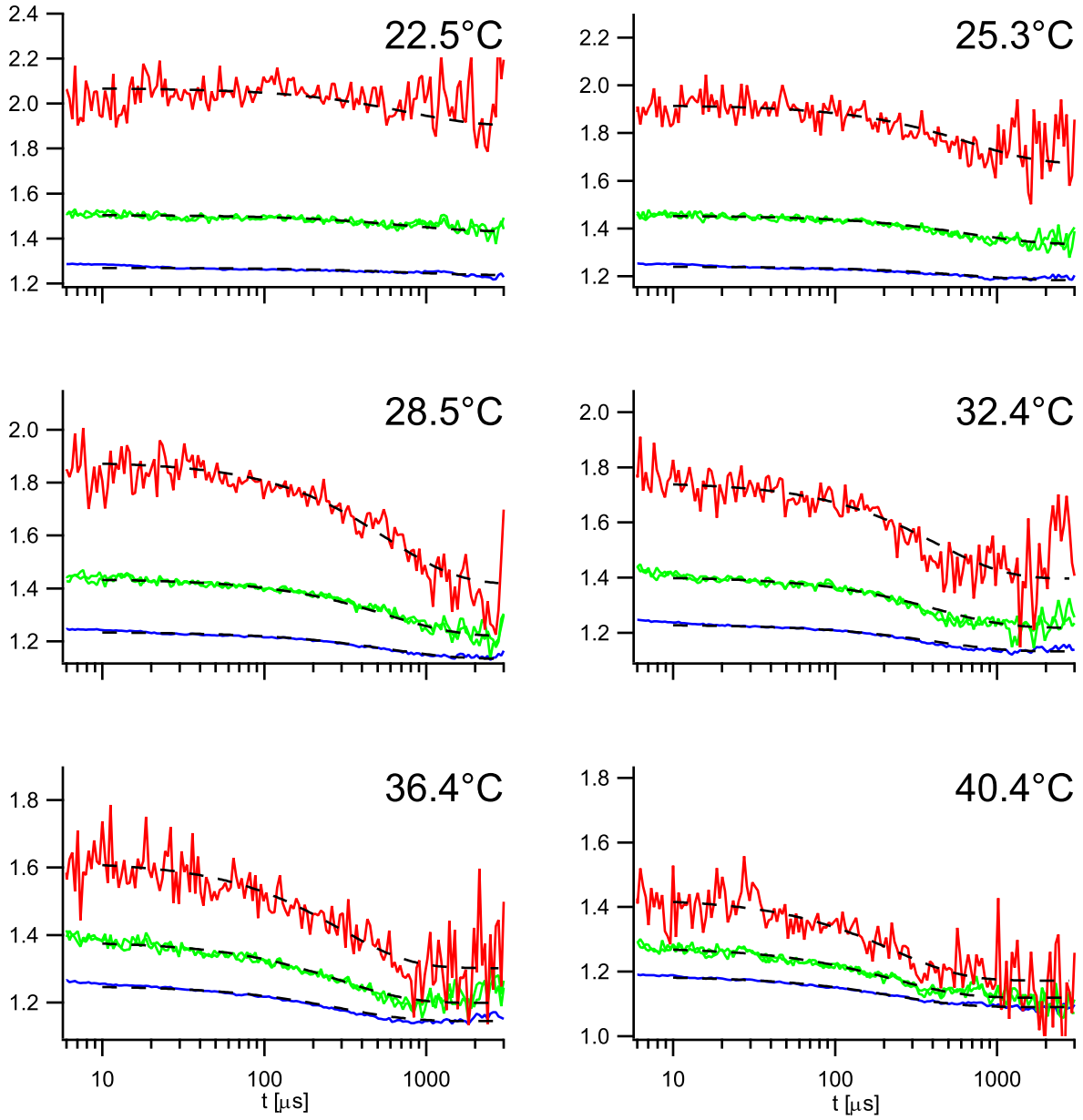

Figure 7.17: Sample TT21G. (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different temperatures.

Table 7.6: Sample TT21G. Global fit parameters directly measured by fitting to the higher-order reaction functions shown in Figure (7.17). The "linked" parameters are shared by the fit functions at all temperatures (global fit). Due to small reaction amplitudes and high uncertainties, t_{F} parameters were held at or between (by interpolation) the values obtained from sample RT21D, except at $40.4°\mathrm{C}$.

linked	q_2	0.471(3)
	q_3	1.11(1)
$22.5°\mathrm{C}$	N	0.907(4)
	K_2	24.3(1.7)
	K_3	3.33(26)
	t_{F}	748.7(held)
$25.3°\mathrm{C}$	N	0.945(2)
	K_2	9.48(43)
	K_3	1.86(11)
	t_{F}	663.6(held)
$28.5°\mathrm{C}$	N	0.937(3)
	K_2	5.40(20)
	K_3	0.537(50)
	t_{F}	578.3(held)
$32.4°\mathrm{C}$	N	0.951(2)
	K_2	3.57(13)
	K_3	0.786(59)
	t_{F}	418.8(held)
$36.4°\mathrm{C}$	N	0.921(2)
	K_2	1.85(7)
	K_3	0.614(55)
	t_{F}	293.6(held)
$40.4°\mathrm{C}$	N	0.947(2)
	K_2	1.03(4)
	K_3	0.445(50)
	t_{F}	229(7)

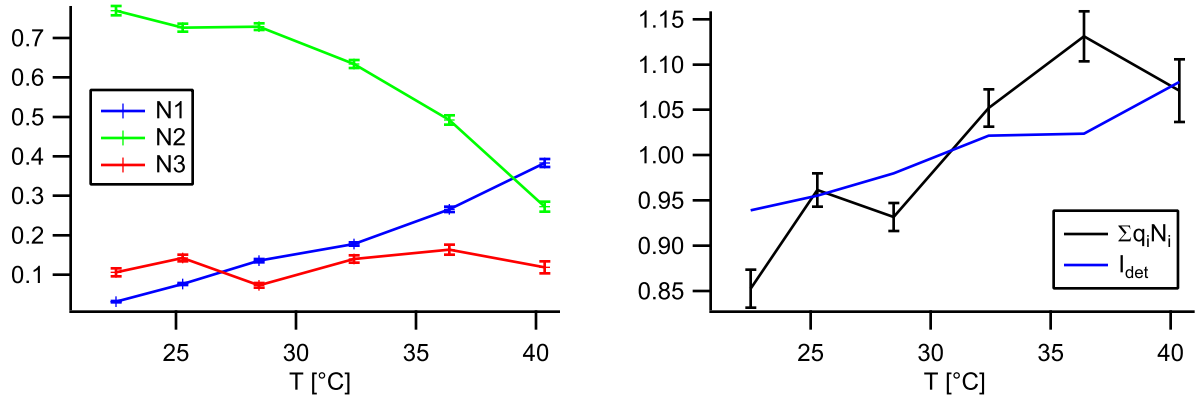

Figure 7.18: Sample TT21G. Left: Population numbers for the unfolded (N_3), intermediate (N_1) and folded (N_2) states as a function of temperature. Right: The overall brightness of the molecules as predicted by the fit parameters (black) compared against the measured detector counts (blue). Uncertainties in the latter are relatively small hence not shown. The curves are normalized by their average values.

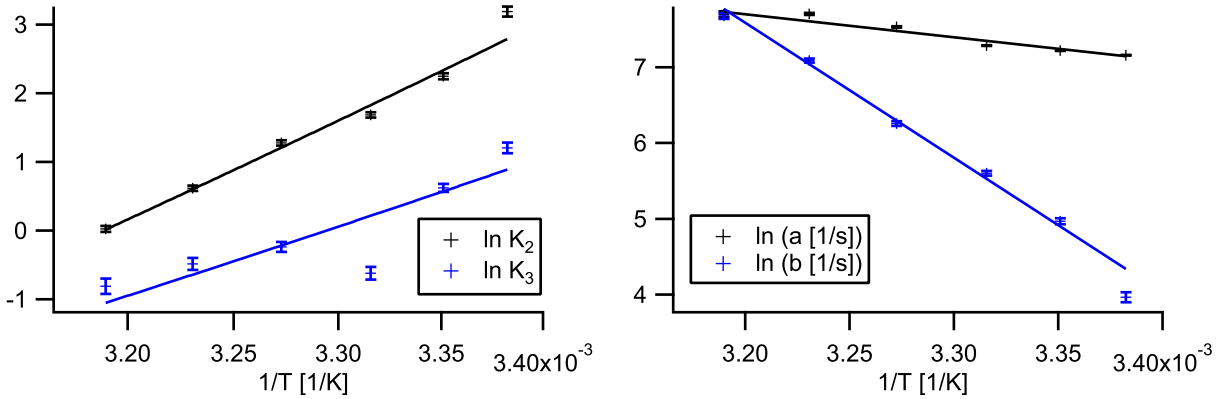

Figure 7.19: Sample TT21G: Left: Van 't Hoff plots of population ratios K_2 (folded/intermediate) and K_3 (unfolded/intermediate). Right: Arrhenius plots of the rate constants, a and b, of the fast step of the reaction (between folded and intermediate states). Linear fits on both graphs are obtained using all points.

Table 7.7: Sample TT21G: Enthalpy change, entropy change, activation energy, and the pre-exponential factors obtained from the van 't Hoff and Arrhenius plots in Figure 7.19.

K_2	$\Delta H_2[\mathrm{kcal\,mol^{-1}}]$	-28.6(6)
	$\Delta S_2[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	-0.0913(19)
K_3	$\Delta H_3[\mathrm{kcal\,mol^{-1}}]$	-20.0(1.1)
	$\Delta S_3[\mathrm{kcal\,mol^{-1}\,K^{-1}}]$	-0.0660(35)
a	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	6.06(11)
	$\ln(A[s^{-1}])$	17.5(2)
b	$E_\mathrm{a}[\mathrm{kcal\,mol^{-1}}]$	35.4(7)
	$\ln(A[s^{-1}])$	64.6(8)

7.2.4 Salt dependence measurements

These measurements are performed with sample RT21D (reference RT21).

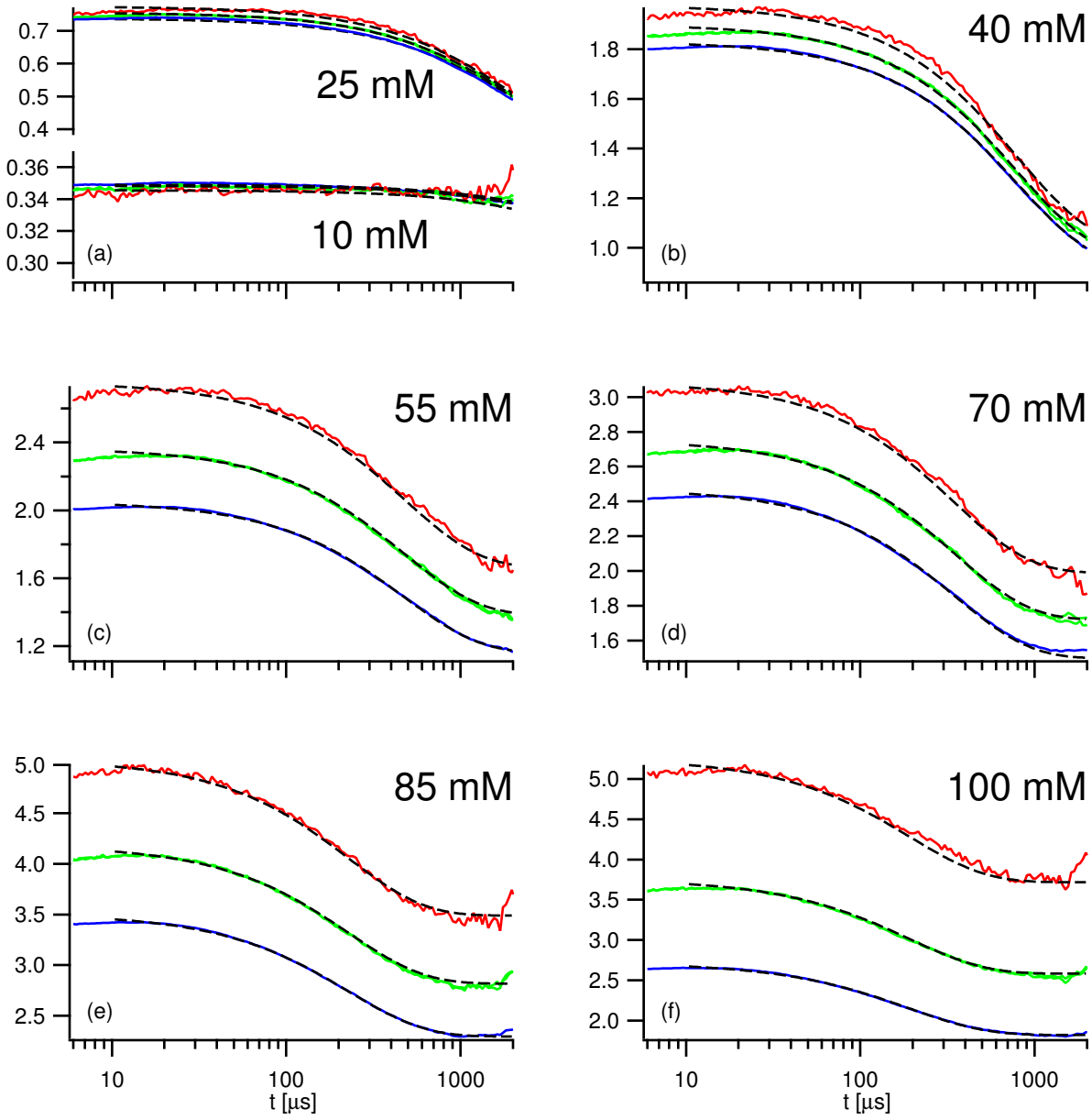

Figure 7.20: (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different NaCl concentrations. Trial 1.

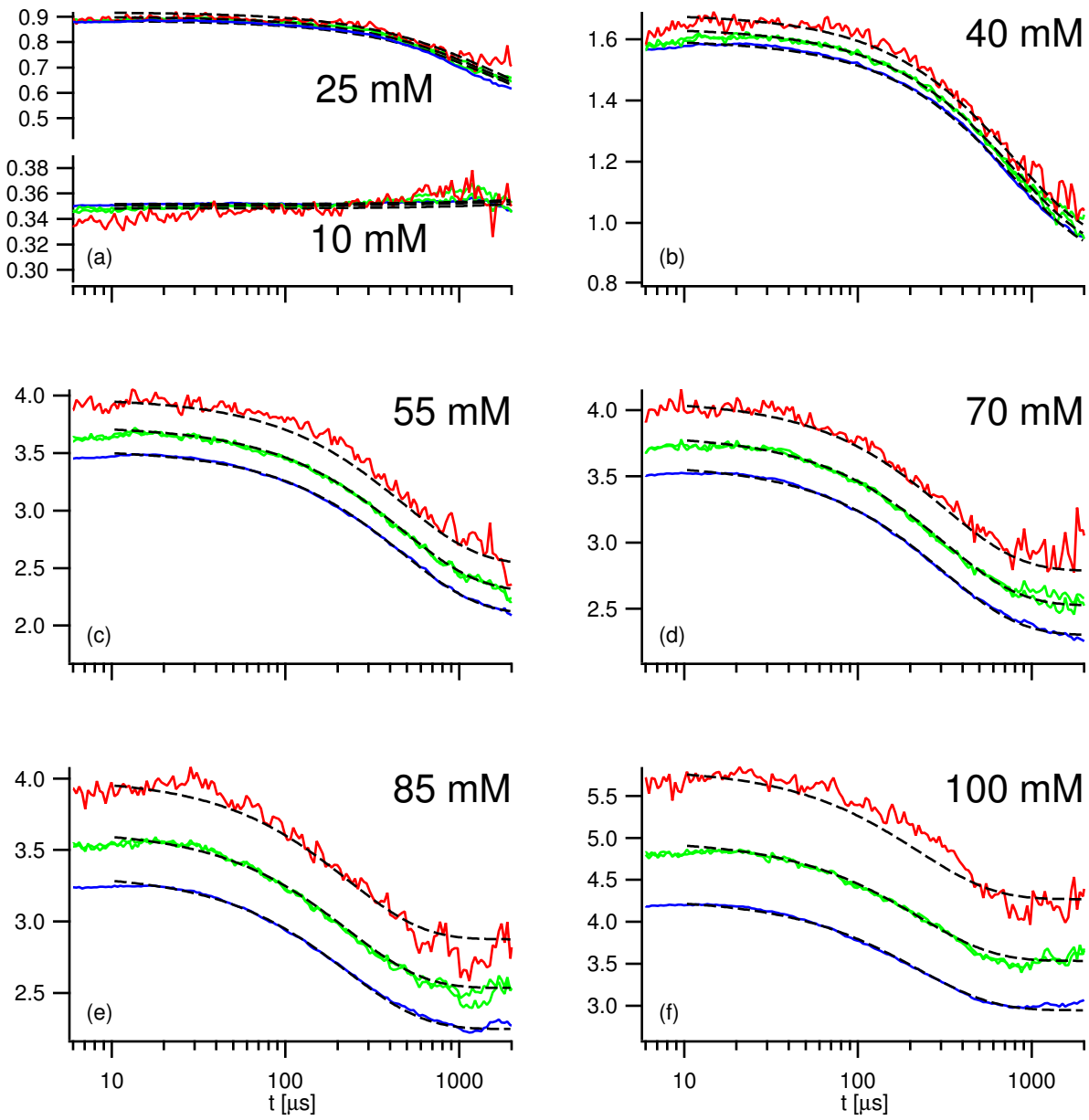

Figure 7.21: (a–e) $R_{11}(t)$ (blue), $R_{21}(t)$ and $R_{12}(t)$ (green), $R_{22}(t)$ (red), and the global fit curves (dashed black) at different NaCl concentrations. Trial 2.

Table 7.8: Sample RT21D: Global fit parameters directly measured by fitting to the higher-order reaction functions shown in Figures 7.20, 7.21, and 6.9. The "linked" parameters are shared by the fit functions at all NaCl concentrations (global fit). t_{F} at 10 mM NaCl is outside the accessible time scale, therefore it was held at a guessed value.

[NaCl]		Trial 1	Trial 2	Trial 3
linked	q_2	0.0239(8)	0.0215(15)	0.0144(8)
	q_3	1.141(4)	1.15(1)	1.29(1)
10 mM	N	3.22(1)	2.74(2)	3.04(4)
	K_2	0.115(4)	-0.033(6)	0.084(16)
	K_3	-0.145(8)	-0.075(15)	0.063(13)
	t_{F}	7000(held)	7000(held)	7000(held)
25 mM	N	2.66(4)	1.93(5)	2.93(19)
	K_2	1.06(3)	0.824(59)	1.51(16)
	K_3	-0.006(12)	0.083(32)	0.0291(16)
	t_{F}	1863(44)	1520(75)	2714(238)
40 mM	N	1.31(2)	1.25(3)	2.54(11)
	K_2	1.85(7)	1.21(7)	3.30(24)
	K_3	0.191(14)	0.128(25)	0.289(19)
	t_{F}	797(5)	798(12)	939(7)
55 mM	N	2.94(8)	0.956(51)	1.27(6)
	K_2	12.5(5)	4.68(45)	2.94(27)
	K_3	0.789(12)	0.714(36)	0.460(30)
	t_{F}	475(2)	472(3)	594(8)
70 mM	N	2.01(6)	0.989(55)	2.55(14)
	K_2	9.81(41)	5.78(55)	13.3(1.0)
	K_3	0.913(16)	0.956(40)	1.13(2)
	t_{F}	347(3)	314(3)	320(2)
85 mM	N	2.00(6)	1.42(8)	3.01(17)
	K_2	20.1(8)	10.4(9)	20.4(1.5)
	K_3	1.24(2)	1.26(4)	1.41(3)
	t_{F}	228(2)	229(3)	227(3)
100 mM	N	3.64(11)	1.63(10)	3.36(18)
	K_2	43.2(1.6)	20.5(1.7)	27.1(1.9)
	K_3	1.381(14)	1.45(4)	1.42(2)
	t_{F}	191(1)	222(4)	185(2)

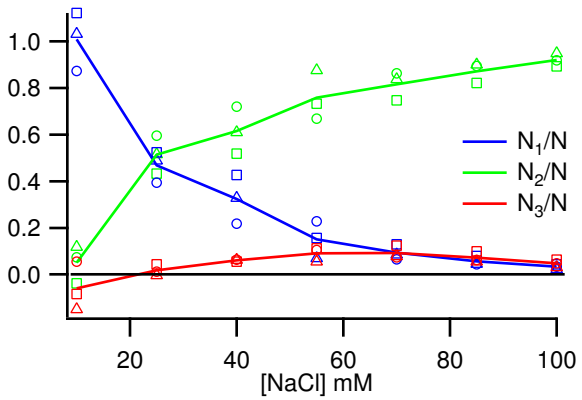

Figure 7.22: Population numbers for the brightest (N_3), intermediate (N_1) and dimmest (N_2) states as a function of NaCl concentration. Solid lines connect the average points of trials 1 (triangle), 2 (square), 3 (circle).

7.3 A hypothesis for the extended conformation

Strong support to our observations was found in recent studies of the conformational distributions of poly$(\mathrm{T})_{30}$ and poly$(\mathrm{A})_{30}$ single-stranded DNA using small-angle x-ray scattering (SAXS) [126, 127]. These reports by Plumridge et al. show a distinct extended population with properties remarkably similar to the brightest state found in our higher-order FCS studies. Details of their data merit attention and discussion, hence are reproduced in Figure 6.14.

Notable similarities can be listed between the extended populations observed by the two independent methods, as described in Section (6.6): the size of the population (maximum $\sim 10\%$), its conformation (consistent with the highest brightness and entropic measurements in higher-order FCS), its salt-dependence behavior (in very good agreement with Figure 6.10 from higher-order FCS), and the energy barrier separating it from the more compact random-coil conformations (consistent with the slow transitions inferred by higher-order FCS).

Two more observations are also noteworthy in the SAXS data: First, a similar extended conformation is also observed for poly$(\mathrm{A})_{30}$ (Figure 6.14, right column), though the population is smaller than poly$(\mathrm{T})_{30}$. Second, at very low salt concentrations where the extended population is absent, the extended conformation is manifestly inaccessible by the hairpin, i.e. there is a void in

the corresponding region of the conformation space. This latter observation is exclusive to SAXS; higher-order FCS would probably not be able to distinguish it from a scenario in which the energy barrier is nonexistent and the random-coil and extended conformations are merged into one bright state. Instead, the SAXS data suggests that the mechanism causing the energy barrier persists in all measurements.

The collection of observations point to a combination of attractive and repulsive forces as a possible explanation. We propose base stacking as the source of the attractive potential. Base stacking is known to be weaker in poly(T) than poly(A) [132], and sometimes even assumed to be absent in poly(T) [35, 82, 104] . However, the stacking free energy of T|T is estimated to be at least half of that of A|A [29, 133–135]. Computational studies have also found significant free energies for stacked T|T and U|U (~10kcal/mol, inherent values without solvent effects) [136–139]. NMR measurements have shown that stacking is significant not only in $\mathrm{Ap_2A}$, but also in $\mathrm{Tp_2T}$ and all other tested dinucleoside diphosphates [140]. Stacking in short hairpin loops makes the hairpins more stable [94, 95, 120] including in poly(T) hairpin loops [141].

The stronger stacking free energy in poly(A) does not necessitate a larger extended population in poly(A) than poly(T), since only the relative populations of the random coil and extended states are relevant. Therefore, the free energy difference between the random coil and extended conformations should be considered in each type of single-stranded DNA, then this difference can be compared between the two types of DNA. This energy difference largely depends on the shape and extent of the repulsive potential function, which is also responsible for the potential barrier between the extended and random coil states. We ascribe the repulsive force to the electrostatic interactions between the adjacent phosphate residues on the DNA backbone.

Example potential functions shown in Figure 7.23 help explain the idea. Attractive stacking free energy functions are selected to approximately match those computed by Mak [29] for two T|T or A|A bases as they slide parallel to each other. A hypothetical electrostatic repulsive function is added to the stacking interactions. At a low salt concentration (e.g. $20\,\mathrm{mM}$ NaCl), the electrostatic repulsion is strong, preventing full alignment of the bases and the formation of an extended

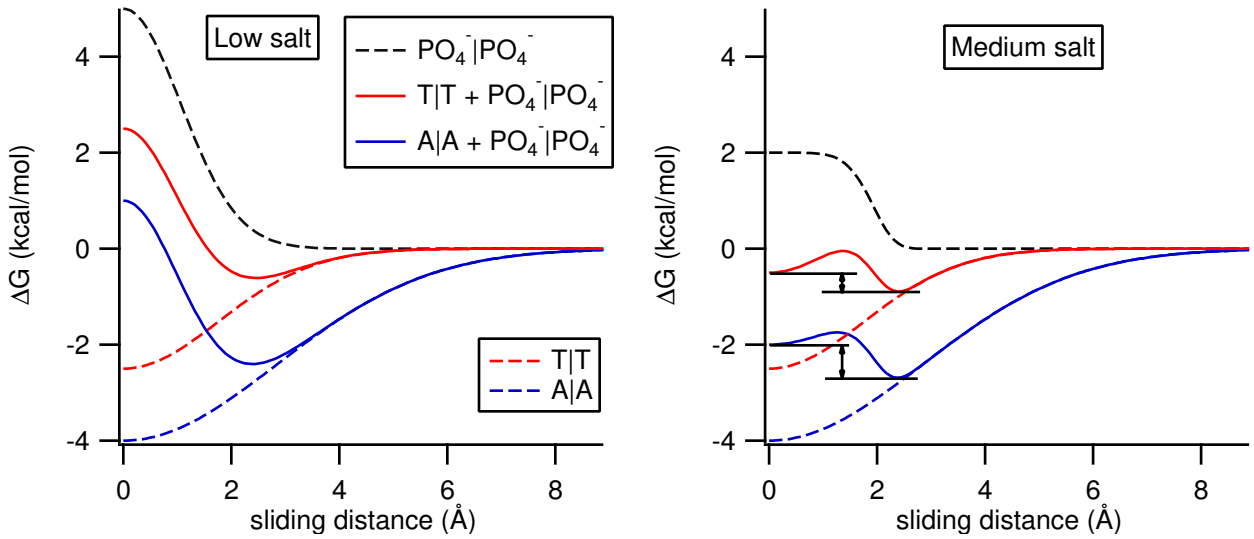

Figure 7.23: Schematic free energy profiles for two bases as a function of sliding distance (or the corresponding angle). Realistic sizes of T|T and A|A stacking free energies are adopted from Mak [29]. A single hypothetical electrostatic interaction between the phosphate residues is added to both functions. At low salt concentration (left), a local minimum corresponding to the random coil state exists, but the extended conformation is unstable. Increasing the salt concentration (right) primarily changes the electrostatic repulsion, creating a new local minimum corresponding to the extended conformation. Arrows indicate the the difference between the random coil and extended free energies.

conformation. A local minimum exists at a misaligned location and favors the intermediate (random coil) conformations (Figure 7.23, left). Increasing the salt concentration to a medium level (e.g. $100\,\mathrm{mM}$ NaCl) makes the electrostatic repulsion weaker, but does not significantly alter the stacking potential [142, 143]. This may allow a new local minimum to appear at the zero sliding position, corresponding to the extended conformation. The stacking interaction is still stronger in poly(A) than in poly(T), particularly affecting the bases in the random (not free) coil state and making poly(A) appear more "rigid" than poly(T) [82, 124]. However, poly(T) shows a larger population of the extended conformation because the extended and intermediate energy levels are closer to each other (Figure 7.23, right).

For a more complete analysis, the entropic advantage of the random coil state, and the long-range electrostatic repulsion between non-neighbor phosphate residues should also be considered. When the ionic strength is further increased beyond the intermediate levels, the weakening of the

long-range electrostatic repulsion may shift the balance back toward the intermediate state (and the folded state if present).

According to our hypothesis, the extended conformation is formed by highly stacked and optimally aligned bases. The molecular origins of stacking interaction are being actively investigated. Recent study suggests solvent entropic forces as the dominant factor, with dispersive and solvent hydrogen-bonding forces producing only secondary perturbations. [29] Solvent screening is thought to fully compensate for any electrostatic components. [30] The stacking free energy is thus found to be nearly constant over two or three orders of magnitude in salt concentration. [142, 143] On the other hand, the electrostatic repulsion between the negatively charged phosphate residues remains susceptible to ionic conditions, explaining how the extended state may become more stable by increasing [NaCl], but outcompeted by the folded state if [NaCl] is too high.

Chapter 8

Conclusion

We have developed techniques to compute higher-order correlation functions free of artifacts down to $0.1\,\mu$s and $1\,\mu$s resolution for the third- and fourth-order functions, respectively. The sub-binning technique for artifact removal does not rely on the specifics of any particular experimental setup or a specific modeling of the detector artifacts: no tuning of the detector parameters is needed. This makes the technique easy to add to an existing conventional FCS setup and correlator design; even prerecorded photon data in conventional experiments can be reanalyzed to extract higher-order correlation functions. Signal-to-noise ratio in higher-order correlation functions is dramatically improved by selecting the bin size optimally. We have shown that as long as the bin size is significantly smaller than the time scale of the targeted reactions, it can be safely increased to reduce the shot noise in higher-order correlation functions.

We introduce methodology to analyze a broad range of molecular kinetics from the slowest limit of time scales, i.e. the static mixtures, to the fastest accessible times scales, or any combination of these time scales in multi-step reactions, within a single framework of higher-order reaction function analysis. We employ the factorized form of higher-order correlation functions, which follows from the cumulant-based formulation, to eliminate any dependence of the analysis on the properties of the illumination and detection profiles. This yields both simplicity and precision: the approximate models for molecular detection functions are usually too inaccurate for higher-order analysis.

Although higher-order FCS still has limits in reaching nanosecond time scales and below, its capability to characterize molecular kinetics over all other time scales with single-molecule resolution is unprecedented. This is best displayed in systems with multiple components of fast and slow kinetics, such as the hairpin conformational dynamics. To view this broad range of capabilities, we summarize the performance of higher-order FCS in the three application cases studied in this dissertation. Figure (8.1) shows the results in the format overviewed in the Introduction section

Figure (1.7). Correlation functions of order $(1,1)$, $(2,1)$, $(1,2)$, and $(2,2)$ are included on each graph in Figure (8.1). A distinct qualitative pattern of higher-order reaction functions is revealed for each category of systems, most notably the hairpin dynamics. In all cases, the third-order functions (orders $(2,1)$ and $(1,2)$, both green) are identical due to the reversible nature of the studied mechanisms. This points to the unique capability of higher-order FCS to characterize irreversible reactions as well, as was shown earlier by Melnykov and Hall [17]. The analysis of fast two- and multi-state reversible kinetics using fourth-order correlations and resolving static mixtures within the same reaction formalism are new applications reported by us [71].

Figure (8.1), top, shows that in the slow time scale limit, a two-component mixture can be resolved using the correlation functions of second, third, and fourth orders combined. Each of these functions is theoretically constant (static) in lag time. The values of the three constants, obtained by fitting, yield a system of equations which can be solved to find the three underlying parameters of the system: N_{bright}, N_{dim}, and the relative brightness, q. Although Palmer and Thompson [15, 16] showed the applicability of higher-order FCS to resolve mixtures in their pioneering work, our analysis using the reaction function formulation which is independent of setup properties is simpler and more precise. Moreover, it showcases the success of the reaction function analysis in higher-order FCS in a critically important limit of time scales. Also, the static mixture analysis helps us understand the components of a faster reacting system from the behavior of higher-order reaction functions in time scales much shorter than the characteristic reaction time, analogous to studying the heterogeneity in a "snapshot" of a fast reacting system.

Figure (8.1), middle, shows that in a fast two-state reaction, all higher-order reaction functions decay to a common baseline, as supported by theory. The amplitudes of the reaction functions and the common baseline, as well as the overall relaxation time, are independent measurable parameters from this data. To resolve the four underlying reaction parameters, only the second- and third-order reaction functions provide enough measurable parameters, although we are showing the fourth-order correlation function for verification as well.

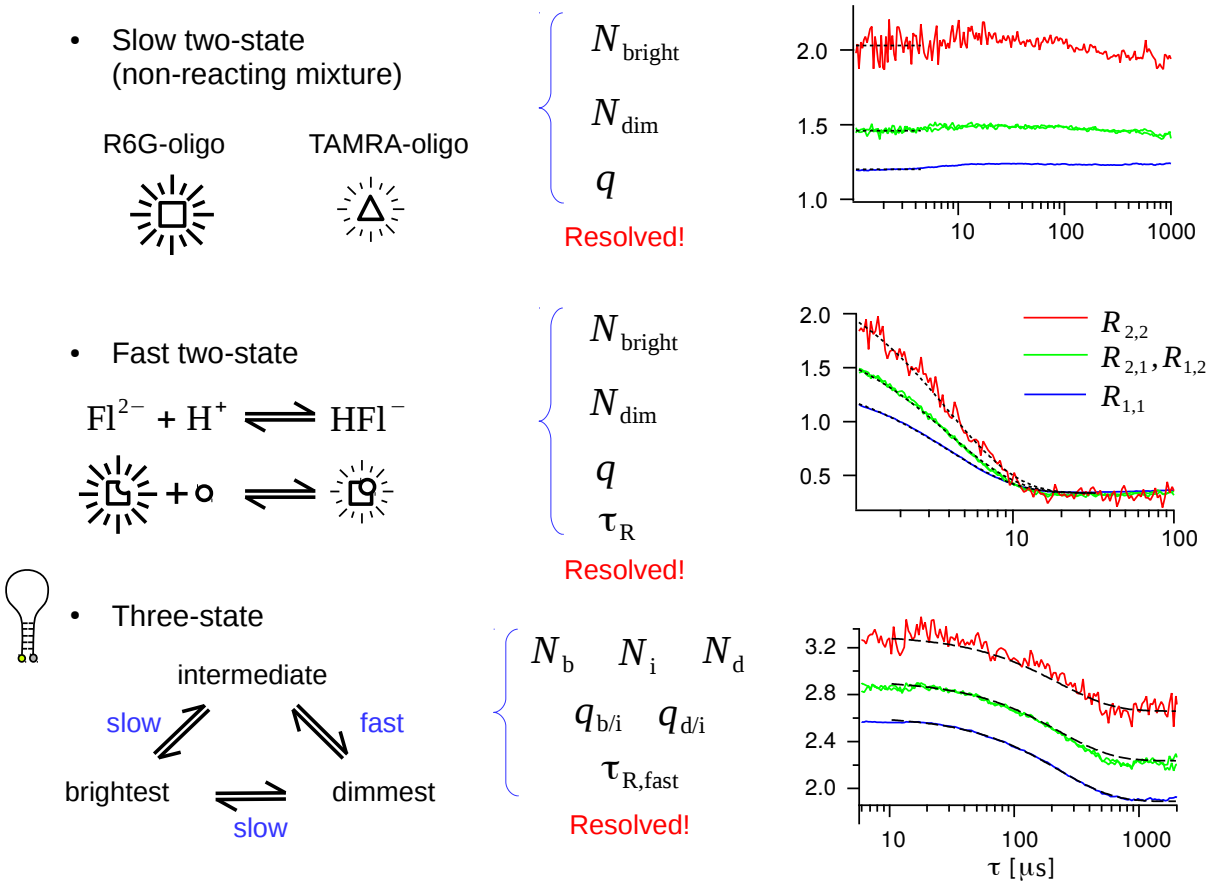

Figure 8.1: Higher-order reaction functions and their performance in resolving two-state reactions in the slow and fast limits; also in the analysis of hairpin folding dynamics. The parameters obtained from fitting to the reaction functions in each case are enough to solve for the parameters that describe the system (middle column). Additionally, the reaction functions show the three-state hairpin mechanism and uniquely identify the fast reaction step.

The hairpin dynamics show a distinct pattern of reaction functions in the bottom graph in Figure (8.1): the baselines stay separate, unlike the two-state reaction. This reveals that the mechanism is more complex than two states, and at least one component of the mechanism is slow (static) on FCS time scale. A three-state model reasonably fits the data, and the static state is found to be the brightest state. The fast reaction step is between the intermediate and dimmest states. This rules out the earlier assumptions that the static state was the dimmest [37]. Using this data, we can now measure the populations and relative brightnesses of all three states, as well as the fast relaxation times. These also yield the forward and reverse rate constants between the intermediate and dimmest states. The slow rate constants remain unattained because the slow relaxation time is out of the FCS time frame. Despite this limitation, the static population can be precisely characterized by higher-order FCS.

To understand the nature of the three hairpin states, we have measured the dependence of the parameters obtained using higher-order FCS on temperature and salt concentration. The thermodynamical analysis reveals an energy landscape as shown in Figure (8.2). The dimmest state has a low entropy and a low enthalpy, consistent with the folded hairpin conformation. The intermediate state has a high entropy and a high enthalpy, consistent with the unfolded random-coil conformations of a hairpin. The salt dependence measurements confirm these conclusions. The brightest state has a low entropy, similar to the folded state, and a slightly higher enthalpy than the folded state. These, together with its brightness, lead us to believe that the brightest state is associated with anomalously extended conformations. These conformations reach a maximum population of nearly 10% near physiological salt concentrations. The analysis measures little or no energy barrier between the intermediate and folded conformations (the fast reaction step). The energy barrier between the intermediate and extended conformations is inferred to be large (due to the slow reaction step), although not directly characterized.

The finding of the new extended conformation was unexpected to us; however, we later found both theoretical [107] and experimental [126, 127] support for it in the literature. In particular, the experimental evidence is in the conformational distribution of single-stranded DNA, with length

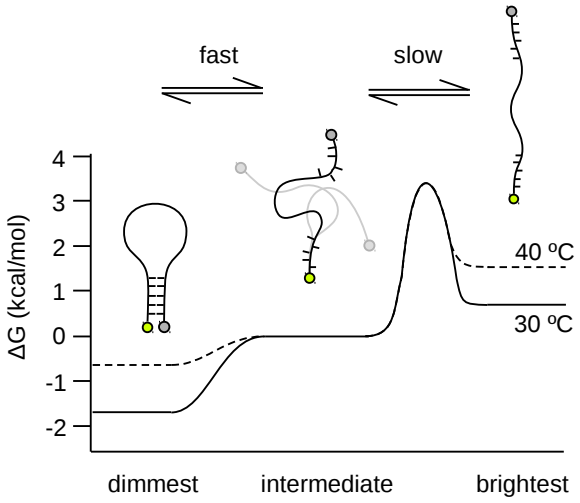

Figure 8.2: The free energy landscape of DNA hairpin conformational dynamics as revealed by higher-order FCS through thermodynamical analysis. The hairpin conformations corresponding to each state are inferred from the brightness, entropy, enthalpy, and ionic dependence of each state. The energy barrier between the folded and random-coil (intermediate) conformations has been measured to be negligibly small. The energy barrier between the random-coil and extended (brightest) conformations has been found to be large, though not specifically characterized. Data is shown for hairpins of type RT8D described in Section 6.4.

and structure similar to our hairpins, characterized using small-angle x-ray scattering (SAXS). In those data, a distinct population of extended DNA conformation is observed, with size and salt concentration dependence remarkably similar to our observations. Also in agreement is an apparent barrier that separates this population from the more compact unfolded conformations, and is responsible for the slow transitions.

These findings call for new experimental and theoretical studies to further confirm the presence of the extended hairpin conformation and help understand the nucleic acid interactions responsible for it. Although we have ruled out most other possibilities within our current capacity, and the independent fluorophore-free SAXS measurements are very compelling, more measurements are needed to eliminate any chances of errors or misinterpretation. If confirmed, the compartmentalization of the unfolded conformational space and the involved energy barriers will be of critical importance in forming our understanding of nucleic acid structures, interactions, and biological functions.

Whether the brightest state is due to extended conformations or something else, higher-order FCS has taken us one step closer to a clear understanding of the earlier FCS results in nucleic acid studies. Regardless of the nature of the brightest state, higher-order FCS remains the first method to independently measure the populations of the random-coil and folded conformations, as well as the forward and revers rate constants of the transitions between them, with a single measurement of the test and reference samples. The technique holds immense potential to be applied to a broad range of chemical and biochemical kinetics in vitro and vivo.

Bibliography

[1] Nils G Walter, Cheng-Yen Huang, Anthony J Manzo, and Mohamed A Sobhy. Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nature methods*, 5(6):475, 2008.

[2] Shimon Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676–1683, 1999.

[3] 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, 2007.

[4] William J Greenleaf, Michael T Woodside, and Steven M Block. High-resolution, single-molecule measurements of biomolecular motion. *Annu. Rev. Biophys. Biomol. Struct.*, 36:171–190, 2007.

[5] Samuel J Lord, Hsiao-lu D Lee, and WE Moerner. Single-molecule spectroscopy and imaging of biomolecules in living cells. *Analytical chemistry*, 82(6):2192–2203, 2010.

[6] Elliot L Elson and Douglas Magde. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers*, 13(1):1–27, 1974.

[7] Oleg Krichevsky and Grégoire Bonnet. Fluorescence correlation spectroscopy: the technique and its applications. *Reports on Progress in Physics*, 65(2):251, 2002.

[8] Alan Van Orden, Keir Fogarty, and Jaemyeong Jung. Fluorescence fluctuation spectroscopy: a coming of age story. *Applied spectroscopy*, 58(5):122A–137A, 2004.

[9] Rudolf Rigler and Jerker Widengren. Fluorescence-based monitoring of electronic state and ion exchange kinetics with FCS and related techniques: from T-jump measurements to fluorescence fluctuations. *European Biophysics Journal*, 47(4):479–492, May 2018.

[10] V Vukojević, A Pramanik, T Yakovleva, R Rigler, L Terenius, and G Bakalkin. Study of molecular events in cells by fluorescence correlation spectroscopy. *Cellular and molecular life sciences*, 62(5):535, 2005.

[11] Ernst-Ludwig Florin, Vincent T Moy, and Hermann E Gaub. Adhesion forces between individual ligand-receptor pairs. *Science*, 264(5157):415–417, 1994.

[12] Michael T. Woodside, Peter C. Anthony, William M. Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, and Steven M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001–1004, 2006.

[13] Sanaz Sadegh, Jenny L Higgins, Patrick C Mannion, Michael M Tamkun, and Diego Krapf. Plasma membrane is compartmentalized by a self-similar cortical actin meshwork. *Physical Review X*, 7(1):011031, 2017.

[14] Yu Chen, Dehong Hu, Erich R Vorpagel, and H Peter Lu. Probing single-molecule T4 lysozyme conformational dynamics by intramolecular fluorescence energy transfer. *The Journal of Physical Chemistry B*, 107(31):7947–7956, 2003.

[15] A.G. Palmer and N.L. Thompson. Molecular aggregation characterized by high order auto-correlation in fluorescence correlation spectroscopy. *Biophysical Journal*, 52(2):257–270, 1987.

[16] Arthur G Palmer and Nancy L Thompson. High-order fluorescence fluctuation analysis of model protein clusters. *Proceedings of the National Academy of Sciences*, 86(16):6148–6152, 1989.

[17] Artem V Melnykov and Kathleen B Hall. Revival of high-order fluorescence correlation analysis: generalized theory and biochemical applications. *The Journal of Physical Chemistry B*, 113(47):15629–15638, 2009.

[18] Gabriele Varani. Exceptionally stable nucleic acid hairpins. *Annual review of biophysics and biomolecular structure*, 24(1):379–404, 1995.

[19] David Bikard, Céline 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, 2010.

[20] Sergei M Mirkin. Expandable DNA repeats and human disease. *Nature*, 447(7147):932, 2007.

[21] Arturo López Castel, John D Cleary, and Christopher E Pearson. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature reviews Molecular cell biology*, 11(3):165, 2010.

[22] Richard R Sinden, Vladimir N Potaman, Elena A Oussatcheva, Christopher E Pearson, Yuri L Lyubchenko, and Luda S Shlyakhtenko. Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *Journal of biosciences*, 27(1):53–65, 2002.

[23] Guoqi Liu, Xiaomi Chen, John J Bissler, Richard R Sinden, and Michael Leffak. Replication-dependent instability at (CTG)·(CAG) repeat hairpins in human cells. *Nature chemical biology*, 6(9):652, 2010.

[24] A Marquis Gacy, Geoffrey Goellner, Nenad Juranić, Slobodan Macura, and Cynthia T McMurray. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, 81(4):533–540, 1995.

[25] Amalia Ávila-Figueroa, Douglas Cattie, and Sarah Delaney. A small unstructured nucleic acid disrupts a trinucleotide repeat hairpin. *Biochemical and biophysical research communications*, 413(4):532–536, 2011.

[26] SH Gellman, TS Haque, and LF Newcomb. New evidence that the hydrophobic effect and dispersion are not major driving forces for nucleotide base stacking. *Biophysical journal*, 71(6):3523, 1996.

[27] Ray Luo, Hillary SR Gilson, Michael J Potter, and Michael K Gilson. The physical basis of nucleic acid base stacking in water. *Biophysical Journal*, 80(1):140–148, 2001.

[28] O. Sínanoĝlu and S. Abdulnur. Hydrophobic stacking of bases and the solvent denaturation of DNA. *Photochemistry and Photobiology*, 3(4):333–342, 1964.

[29] Chi H. Mak. Unraveling base stacking driving forces in DNA. *The Journal of Physical Chemistry B*, 120(26):6010–6020, 2016.

[30] Jiří Šponer, Judit E Šponer, Arnošt Mládek, Petr Jurečka, Pavel Banáš, and Michal Otyepka. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*, 99(12):978–988, 2013.

[31] Branislav Milovanović, Marko Kojić, Milena Petković, and Mihajlo Etinski. New insight into uracil stacking in water from ab initio molecular dynamics. *Journal of Chemical Theory and Computation*, 14(5):2621–2632, 2018.

[32] Asaminew H Aytenfisu, Aleksandar Spasic, Alan Grossfield, Harry A Stern, and David H Mathews. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. *Journal of chemical theory and computation*, 13(2):900–915, 2017.

[33] Dazhi Tan, Stefano Piana, Robert M Dirks, and David E Shaw. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proceedings of the National Academy of Sciences*, page 201713027, 2018.

[34] Changsheng Zhang, Chao Lu, Zhifeng Jing, Chuanjie Wu, Jean-Philip Piquemal, Jay W Ponder, and Pengyu Ren. AMOEBA polarizable atomic multipole force field for nucleic acids. *Journal of chemical theory and computation*, 14(4):2084–2108, 2018.

[35] Grégoire Bonnet, Oleg Krichevsky, and Albert Libchaber. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proceedings of the National Academy of Sciences*, 95(15):8602–8606, 1998.

[36] Grégoire Bonnet, Sanjay Tyagi, Albert Libchaber, and Fred Russell Kramer. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proceedings of the National Academy of Sciences*, 96(11):6171–6176, 1999.

[37] 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.

[38] Jaemyeong Jung and Alan Van Orden. Folding and unfolding kinetics of DNA hairpins in flowing solution by multiparameter fluorescence correlation spectroscopy. *The Journal of Physical Chemistry B*, 109(8):3648–3657, 2005.

[39] 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.

[40] Yandong Yin, Peng Wang, Xin Xing Yang, Xun Li, Chuan He, and Xin Sheng Zhao. Panorama of DNA hairpin folding observed via diffusion-decelerated fluorescence correlation spectroscopy. *Chem. Commun.*, 48(59):7413–7415, 2012.

[41] Jerker Widengren, Johannes Dapprich, and Rudolf Rigler. Fast interactions between Rh6G and dGTP in water studied by fluorescence correlation spectroscopy. *Chemical Physics*, 216(3):417–426, 1997.

[42] Thorsten Wohland, Kirstin Friedrich, Rudolf Hovius, and Horst Vogel. Study of ligand-receptor interactions by fluorescence correlation spectroscopy with different fluorophores: evidence that the homopentameric 5-hydroxytryptamine type 3As receptor binds only one ligand. *Biochemistry*, 38(27):8671–8681, 1999.

[43] Yan Chen, Joachim D. Müller, Sergey Y. Tetin, Joan D. Tyner, and Enrico Gratton. Probing ligand protein binding equilibria with fluorescence fluctuation spectroscopy. *Biophysical Journal*, 79(2):1074–1084, 2000.

[44] Alena M Lieto, Randall C Cush, and Nancy L Thompson. Ligand-receptor kinetics measured by total internal reflection with fluorescence correlation spectroscopy. *Biophysical Journal*, 85(5):3294–3302, 2003.

[45] Stephen J Briddon and Stephen J Hill. Pharmacology under the microscope: the use of fluorescence correlation spectroscopy to determine the properties of ligand–receptor complexes. *Trends in pharmacological sciences*, 28(12):637–645, 2007.

[46] Yan Chen, Joachim D. Müller, Peter T.C. So, and Enrico Gratton. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophysical Journal*, 77(1):553–567, 1999.

[47] Peet Kask, Kaupo Palo, Dirk Ullmann, and Karsten Gall. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proceedings of the National Academy of Sciences*, 96(24):13756–13761, 1999.

[48] Joachim D. Müller, Yan Chen, and Enrico Gratton. Resolving heterogeneity on the single molecular level with the photon-counting histogram. *Biophysical Journal*, 78(1):474–486, 2000.

[49] Hong Qian and Elliot L Elson. Distribution of molecular aggregation by analysis of fluctuation moments. *Proceedings of the National Academy of Sciences*, 87(14):5479–5483, 1990.

[50] Hong Qian and Elliot L Elson. On the analysis of high order moments of fluorescence fluctuations. *Biophysical Journal*, 57(2):375, 1990.

[51] Joachim D Müller. Cumulant analysis in fluorescence fluctuation spectroscopy. *Biophysical Journal*, 86(6):3981–3992, 2004.

[52] Alvaro Sanchez-Andres, Yan Chen, and Joachim D. Müller. Molecular brightness determined from a generalized form of Mandel's Q-parameter. *Biophysical Journal*, 89(5):3531–3547, 2005.

[53] Kaupo Palo, Ülo Mets, Stefan Jäger, Peet Kask, and Karsten Gall. Fluorescence intensity multiple distributions analysis: concurrent determination of diffusion times and molecular brightness. *Biophysical Journal*, 79(6):2858–2866, 2000.

[54] Irina V. Gopich and Attila Szabo. Photon counting histograms for diffusing fluorophores. *The Journal of Physical Chemistry B*, 109(37):17683–17688, 2005.

[55] Bin Wu and Joachim D Müller. Time-integrated fluorescence cumulant analysis in fluorescence fluctuation spectroscopy. *Biophysical Journal*, 89(4):2721–2735, 2005.

[56] Bin Wu, Yan Chen, and Joachim D Müller. Dual-color time-integrated fluorescence cumulant analysis. *Biophysical Journal*, 91(7):2687–2698, 2006.

[57] William K. Ridgeway, David P. Millar, and James R. Williamson. The spectroscopic basis of fluorescence triple correlation spectroscopy. *The Journal of Physical Chemistry B*, 116(6):1908–1919, 2012.

[58] Ted A. Laurence, Achillefs N. Kapanidis, Xiangxu Kong, Daniel S. Chemla, , and Shimon Weiss. Photon arrival-time interval distribution (PAID): a novel tool for analyzing molecular interactions. *The Journal of Physical Chemistry B*, 108(9):3051–3067, 2004.

[59] Arthur G Palmer III and Nancy L Thompson. Intensity dependence of high-order autocorrelation functions in fluorescence correlation spectroscopy. *Review of Scientific Instruments*, 60(4):624–633, 1989.

[60] Luc Devroye. Chapter 4 Nonuniform random variate generation. In Shane G. Henderson and Barry L. Nelson, editors, *Simulation*, volume 13 of *Handbooks in operations research and management science*, pages 83–121. Elsevier, 2006.

[61] Perry G. Schiro, Christopher L. Kuyper, and Daniel T. Chiu. Continuous-flow single-molecule CE with high detection efficiency. *ELECTROPHORESIS*, 28(14):2430–2438, 7 2007.

[62] L Mandel. Fluctuations of photon beams and their correlations. *Proceedings of the Physical Society*, 72(6):1037, 1958.

[63] Farshad Abdollah-Nia. Cumulant-based formulation of higher-order fluorescence correlation spectroscopy, 2018. arXiv:1612.05388v2. arXiv.org e-Print archive. https://arxiv.org/abs/1612.05388v2 (accessed Oct 26, 2018).

[64] Jerome Wenger, Davy Gerard, Heykel Aouani, Herve Rigneault, Bryan Lowder, Steve Blair, Eloïse Devaux, and Thomas W Ebbesen. Nanoaperture-enhanced signal-to-noise ratio in fluorescence correlation spectroscopy. *Analytical Chemistry*, 81(2):834–839, 2009.

[65] Christian Kurtsiefer, Patrick Zarda, Sonja Mayer, and Harald Weinfurter. The breakdown flash of silicon avalanche photodiodes-back door for eavesdropper attacks? *Journal of Modern Optics*, 48(13):2039–2047, 2001.

[66] Wolfgang Becker. *Advanced time-correlated single photon counting techniques*, volume 81. Springer Science & Business Media, 2005.

[67] Daniel Nettels, Armin Hoffmann, and Benjamin Schuler. Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *The Journal of Physical Chemistry B*, 112(19):6137–6146, 2008.

[68] S. Felekyan, R. Kühnemuth, V. Kudryavtsev, C. Sandhagen, W. Becker, and C. A. M. Seidel. Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Review of Scientific Instruments*, 76(8), 2005.

[69] Lindsey N Hillesheim and Joachim D Müller. The dual-color photon counting histogram with non-ideal photodetectors. *Biophysical Journal*, 89(5):3491–3507, 2005.

[70] K Schatzel. Noise on photon correlation data. I. autocorrelation functions. *Quantum Optics: Journal of the European Optical Society Part B*, 2(4):287, 1990.

[71] Farshad Abdollah-Nia, Martin P. Gelfand, and Alan Van Orden. Artifact-free and detection-profile-independent higher-order fluorescence correlation spectroscopy for microsecond-resolved kinetics. 2. Mixtures and reactions. *The Journal of Physical Chemistry B*, 121(11):2388–2399, 2017.

[72] N Balakrishnan, Norman L Johnson, and Samuel Kotz. A note on relationships between moments, central moments and cumulants from multivariate distributions. *Statistics & probability letters*, 39(1):49–54, 1998.

[73] Alan Stuart and Keith Ord. *Kendall's advanced theory of statistics*, volume 1. Arnold, a member of the Hoddder Headline Group, 4th edition, 1994.

[74] Jerker Widengren, Bob Terry, and Rudolf Rigler. Protonation kinetics of GFP and FITC investigated by FCS — aspects of the use of fluorescent indicators for measuring pH. *Chemical Physics*, 249(2–3):259–271, 1999.

[75] Stefan Highsmith. Solvent accessibility of the adenosine 5'-triphosphate catalytic site of sarcoplasmic reticulum calcium ATPase. *Biochemistry*, 25(5):1049–1054, 1986.

[76] Horst Leonhardt, Larry Gordon, and Robert Livingston. Acid-base equilibriums of fluorescein and 2',7'-dichlorofluorescein in their ground and fluorescent states. *The Journal of Physical Chemistry*, 75(2):245–249, 1971.

[77] Robert Sjöback, Jan Nygren, and Mikael Kubista. Absorption and fluorescence properties of fluorescein. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 51(6):L7–L21, 1995.

[78] Michael J. Geisow. Fluorescein conjugates as indicators of subcellular pH. *Experimental Cell Research*, 150(1):29–35, 1984.

[79] Wuliang Sun and G.Nigel Godson. Structure of the Escherichia coli primase/single-strand DNA-binding protein/phage G4oric complex required for primer RNA synthesis. *Journal of Molecular Biology*, 276(4):689–703, 1998.

[80] Xing Dai, Miriam B. Greizerstein, Krisztina Nadas-Chinni, and Lucia B. Rothman-Denes. Supercoil-induced extrusion of a regulatory DNA hairpin. *Proceedings of the National Academy of Sciences*, 94(6):2174–2179, 1997.

[81] Angela K Kennedy, Anjan Guhathakurta, Nancy Kleckner, and David B Haniford. Tn10 transposition via a DNA hairpin intermediate. *Cell*, 95(1):125–134, 1998.

[82] Noel L. Goddard, Grégoire Bonnet, Oleg Krichevsky, and Albert Libchaber. Sequence dependent rigidity of single stranded DNA. *Phys. Rev. Lett.*, 85:2400–2403, Sep 2000.

[83] 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, 2000.

[84] 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*, 98(10):5584–5589, 2001.

[85] Haitao Li, Xiaojun Ren, Liming Ying, Shankar Balasubramanian, and David Klenerman. Measuring single-molecule nucleic acid dynamics in solution by two-color filtered ratiometric fluorescence correlation spectroscopy. *Proceedings of the National Academy of Sciences*, 101(40):14425–14430, 2004.

[86] 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, 2008.

[87] Liming Ying, Mark I. Wallace, and David Klenerman. Two-state model of conformational fluctuation in a DNA hairpin-loop. *Chemical Physics Letters*, 334(1):145–150, 2001.

[88] 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*, 98(14):7771–7776, 2001.

[89] Yiqing Shen, Serguei V. Kuznetsov, and Anjum Ansari. Loop dependence of the dynamics of DNA hairpins. *The Journal of Physical Chemistry B*, 105(48):12202–12211, 2001.

[90] Ranjani Narayanan, Li Zhu, Yogambigai Velmurugu, Jorjethe Roca, Serguei V. Kuznetsov, Gerd Prehna, Lisa J. Lapidus, and Anjum Ansari. Exploring the energy landscape of nucleic acid hairpins using laser temperature-jump and microfluidic mixing. *Journal of the American Chemical Society*, 134(46):18952–18963, 2012.

[91] 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.

[92] Mark E. Polinkovsky, Yann Gambin, Priya R. Banerjee, Michael J. Erickstad, Alex Groisman, and Ashok A. Deniz. Ultrafast cooling reveals microsecond-scale biomolecular dynamics. *Nature Communications*, 5:5737, Dec 2014. Article.

[93] Dietmar Pörschke. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophysical chemistry*, 1(5):381–386, 1974.

[94] CW Hilbers, CAG Haasnoot, SH De Bruin, JJM Joordens, GA Van Der Marel, and JH Van Boom. Hairpin formation in synthetic oligonucleotides. *Biochimie*, 67(7-8):685–695, 1985.

[95] 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.

[96] Rajesh K. Nayak, Olve B. Peersen, Kathleen B. Hall, and Alan Van Orden. Millisecond time-scale folding and unfolding of DNA hairpins using rapid-mixing stopped-flow kinetics. *Journal of the American Chemical Society*, 134(5):2453–2456, 2012.

[97] Artem V. Melnykov, Rajesh K. Nayak, Kathleen B. Hall, and Alan Van Orden. Effect of loop composition on the stability and folding kinetics of RNA hairpins with large loops. *Biochemistry*, 54(10):1886–1896, 2015.

[98] 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, 2001.

[99] Roman Tsukanov, Toma E. Tomov, Rula Masoud, Hagai Drory, Noa Plavner, Miran Liber, and Eyal Nir. Detailed study of DNA hairpin dynamics using single-molecule fluorescence assisted by DNA origami. *The Journal of Physical Chemistry B*, 117(40):11932–11942, 2013.

[100] Roman Tsukanov, Toma E. Tomov, Yaron Berger, Miran Liber, and Eyal Nir. Conformational dynamics of DNA hairpins at millisecond resolution obtained from analysis of single-molecule FRET histograms. *The Journal of Physical Chemistry B*, 117(50):16105–16109, 2013.

[101] 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.

[102] Krishna Neupane, Dustin B. Ritchie, Hao Yu, Daniel A. N. Foster, Feng Wang, and Michael T. Woodside. Transition path times for nucleic acid folding determined from energy-landscape analysis of single-molecule trajectories. *Phys. Rev. Lett.*, 109:068102, Aug 2012.

[103] Megan C. Engel, Dustin B. Ritchie, Daniel A. N. Foster, Kevin S. D. Beach, and Michael T. Woodside. Reconstructing folding energy landscape profiles from nonequilibrium pulling

curves with an inverse Weierstrass integral transform. *Phys. Rev. Lett.*, 113:238104, Dec 2014.

[104] Daniel P. Aalberts, John M. Parman, and Noel L. Goddard. Single-strand stacking free energy from DNA beacon kinetics. *Biophysical Journal*, 84(5):3212–3217, 2003.

[105] Majid Mosayebi, Flavio Romano, Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. The role of loop stacking in the dynamics of DNA hairpin formation. *The Journal of Physical Chemistry B*, 118(49):14326–14335, 2014.

[106] Yu-Jane Sheng, Han-Jou Lin, Jeff Z. Y. Chen, and Heng-Kwong Tsao. Effect of the intermediate state on the loop-to-coil transition of a telechelic chain. *The Journal of Chemical Physics*, 118(18):8513–8520, 2003.

[107] Debayan Chakraborty, Rosana Collepardo-Guevara, and David J. Wales. Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins. *Journal of the American Chemical Society*, 136(52):18052–18061, 2014.

[108] Guillem Portella and Modesto Orozco. Multiple routes to characterize the folding of a small DNA hairpin. *Angewandte Chemie International Edition*, 49(42):7673–7676, 2010.

[109] SergueiÂăV. Kuznetsov and Anjum Ansari. A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophysical Journal*, 102(1):101–111, 2012.

[110] 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.

[111] Krishnarjun Sarkar, Konrad Meister, Anurag Sethi, and Martin Gruebele. Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe. *Biophysical Journal*, 97(5):1418–1427, 2009.

[112] Jiho Kim, Sören 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.

[113] Katherine Truex, Hoi Sung Chung, John M. Louis, and William A. Eaton. Testing landscape theory for biomolecular processes with single molecule fluorescence spectroscopy. *Phys. Rev. Lett.*, 115:018101, Jul 2015.

[114] Wenbing Zhang and Shi-Jie Chen. RNA hairpin-folding kinetics. *Proceedings of the National Academy of Sciences*, 99(4):1931–1936, 2002.

[115] Farshad Abdollah-Nia, Martin P. Gelfand, and Alan Van Orden. Artifact-free and detection-profile-independent higher-order fluorescence correlation spectroscopy for microsecond-resolved kinetics. 1. Multidetector and sub-binning approach. *The Journal of Physical Chemistry B*, 121(11):2373–2387, 2017.

[116] Zhenqin Wu, Huimin Bi, Sichen Pan, Lingyi Meng, and Xin Sheng Zhao. Determination of equilibrium constant and relative brightness in fluorescence correlation spectroscopy by considering third-order correlations. *The Journal of Physical Chemistry B*, 120(45):11674–11682, 2016.

[117] P-O Gendron, F Avaltroni, and KJ Wilkinson. Diffusion coefficients of several rhodamine derivatives as determined by pulsed field gradient–nuclear magnetic resonance and fluorescence correlation spectroscopy. *Journal of fluorescence*, 18(6):1093, 2008.

[118] CB Müller, A Loman, V Pacheco, F Koberling, D Willbold, W Richtering, and J Enderlein. Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL (Europhysics Letters)*, 83(4):46001, 2008.

[119] Salvatore A. E. Marras, Fred Russell Kramer, and Sanjay Tyagi. Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Research*, 30(21):e122, 2002.

[120] C. A. G. Haasnoot, S. H. de Bruin, C. W. Hilbers, G. A. van der Marel, and J. H. van Boom. Loopstructures in synthetic oligonucleotides. Hairpin stability and structure studied as a function of loop elongation. *Journal of Biosciences*, 8(3):767–780, Aug 1985.

[121] Paulius Vaitiekunas, Colyn Crane-Robinson, and Peter L Privalov. The energetic basis of the DNA double helix: a combined microcalorimetric approach. *Nucleic Acids Research*, 43(17):8577–8589, 2015.

[122] Kaushik Gurunathan and Marcia Levitus. FRET Fluctuation Spectroscopy of Diffusing Biopolymers: Contributions of Conformational Dynamics and Translational Diffusion. *The Journal of Physical Chemistry B*, 114(2):980–986, 2010.

[123] Tedman Torres and Marcia Levitus. Measuring Conformational Dynamics: A New FCS-FRET Approach. *The Journal of Physical Chemistry B*, 111(25):7392–7400, 2007.

[124] Xiaojuan Wang and Werner M. Nau. Kinetics of end-to-end collision in short single-stranded nucleic acids. *Journal of the American Chemical Society*, 126(3):808–813, 2004.

[125] Kiyohiko Kawai, Hiroko Yoshida, Akira Sugimoto, Mamoru Fujitsuka, and Tetsuro Majima. Kinetics of Transient End-to-End Contact of Single-Stranded DNAs. *Journal of the American Chemical Society*, 127(38):13232–13237, 2005.

[126] Alex Plumridge, Steve P. Meisburger, and Lois Pollack. Visualizing single-stranded nucleic acids in solution. *Nucleic Acids Research*, 45(9):e66, 2017.

[127] Alex Plumridge, Steve P. Meisburger, Kurt Andresen, and Lois Pollack. The impact of base stacking on the conformations and electrostatics of single-stranded DNA. *Nucleic Acids Research*, 45(7):3932–3943, 2017.

[128] Yusdi Santoso and Achillefs N. Kapanidis. Probing biomolecular structures and dynamics of single molecules using in-gel alternating-laser excitation. *Analytical Chemistry*, 81(23):9561–9570, 2009.

[129] W. Patrick Ambrose, Peter M. Goodwin, James H. Jett, Alan Van Orden, James H. Werner, and Richard A. Keller. Single molecule fluorescence spectroscopy at ambient temperature. *Chemical Reviews*, 99(10):2929–2956, 1999.

[130] Jerker Widengren, Uelo Mets, and Rudolf Rigler. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. *The Journal of Physical Chemistry*, 99(36):13368–13379, 1995.

[131] Jerker Widengren, Rudolf Rigler, and Ülo Mets. Triplet-state monitoring by fluorescence correlation spectroscopy. *Journal of fluorescence*, 4(3):255–258, 1994.

[132] Mary M Senior, Roger A Jones, and Kenneth J Breslauer. Influence of loop residues on the relative stabilities of DNA hairpin structures. *Proceedings of the National Academy of Sciences*, 85(17):6242–6246, 1988.

[133] Eric T Kool. Preorganization of DNA: Design principles for improving nucleic acid recognition by synthetic oligonucleotides. *Chemical reviews*, 97(5):1473–1488, 1997.

[134] Jacob S Lai, Jin Qu, and Eric T Kool. Fluorinated DNA bases as probes of electrostatic effects in DNA base stacking. *Angewandte Chemie International Edition*, 42(48):5973–5977, 2003.

[135] Reid F. Brown, Casey T. Andrews, and Adrian H. Elcock. Stacking free energies of all DNA and RNA nucleoside pairs and dinucleoside-monophosphates computed using recently revised AMBER parameters and compared with experiment. *Journal of Chemical Theory and Computation*, 11(5):2315–2328, 2015.

[136] Ruairidh S Hunter and Tanja Van Mourik. DNA base stacking: The stacked uracil/uracil and thymine/thymine minima. *Journal of computational chemistry*, 33(27):2161–2172, 2012.

[137] Matthew L Leininger, Ida MB Nielsen, Michael E Colvin, and Curtis L Janssen. Accurate structures and binding energies for stacked uracil dimers. *The Journal of Physical Chemistry A*, 106(15):3850–3854, 2002.

[138] Jiande Gu, Jing Wang, Jerzy Leszczynski, Yaoming Xie, and Henry F Schaefer III. To stack or not to stack: performance of a new density functional for the uracil and thymine dimers. *Chemical Physics Letters*, 459(1-6):164–166, 2008.

[139] Tanja van Mourik. Comment on 'To stack or not to stack: performance of a new density functional for the uracil and thymine dimers' [Chem. Phys. Lett. 459 (2008) 164]. *Chemical Physics Letters*, 473(1):206–208, 2009.

[140] Noa Stern, Dan Thomas Major, Hugo Emilio Gottlieb, Daniel Weizman, and Bilha Fischer. What is the conformation of physiologically-active dinucleoside polyphosphates in solution? Conformational analysis of free dinucleoside polyphosphates by NMR and molecular dynamics simulations. *Org. Biomol. Chem.*, 8:4637–4652, 2010.

[141] Rajesh K. Nayak and Alan Van Orden. Counterion and polythymidine loop-length-dependent folding and thermodynamic stability of DNA hairpins reveal the unusual counterion-dependent stability of tetraloop hairpins. *The Journal of Physical Chemistry B*, 117(45):13956–13966, 2013.

[142] Janet T. Powell, E. G. Richards, and W. B. Gratzer. The nature of stacking equilibria in polynucleotides. *Biopolymers*, 11(1):235–250, 1972.

[143] Dustin B. McIntosh, Gina Duggan, Quentin Gouil, and Omar A. Saleh. Sequence-dependent elasticity and electrostatics of single-stranded DNA: signatures of base-stacking. *Biophysical Journal*, 106(3):659–666, 2014.