Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Single-molecule Multi-color FRET:

from DNA to Protein

Sushi Madhira

aus

Hyderabad, Indien

2017

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Don C. Lamb, PhD betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 15. March 2017

.....

(Sushi Madhira)

Dissertation eingereicht am	17. March 2017
1. Gutachter	Prof. Don C. Lamb, PhD
2. Gutachter	Associate Prof. Promod R. Pratap, PhD
Mündliche Prüfung am	31. March 2017

Abstract

Single-molecule fluorescence techniques are a powerful tool that enable monitoring of the conformational dynamics of biomolecules in real time, whether in solution or tethered to the surface. Single-pair FRET (spFRET) has been extensively used in-vitro to obtain structural information and kinetics of the dynamics in the system. The combination of spFRET with total internal reflection fluorescence microscopy (TIRFM) allows monitoring the dynamic interconversion of states of labeled biomolecules, immobilized on the surface, for as long as the fluorophores do not bleach. The fluorescence intensities of the labeled molecules are extracted and the underlying sequence of states is obtained from an hidden markov model (HMM) analysis. However, spFRET limits the detection of dynamics to a single-dimension, while biomolecular systems usually involve multiple partners interacting in a three-dimensional manner. In this thesis, I extended an existing TIRFM setup to allow detection of multiple colors enabling the simultaneous observation of multiple distances. The dyes, excitation wavelengths and the optics were selected carefully to increase the detection efficiency and decrease the spectral cross-talk. However, the realisation of quantitative multi-color FRET is challenging owing to the various correction factors involved.

The multi-color TIRF system was further employed to study dynamics in different DNA origami based systems. Additionally, multiparameter fluorescence detection (MFD) in combination with pulsed interleaved excitation (PIE) was used to characterize the fluorescence properties (lifetime, anisotropy etc.) of the fluophores involved. In the first project, the autonomous switching of a surface tethered molecular construct performing a catalytic hairpin assembly reaction was monitored through spFRET measurements on TIRF and MFD-PIE. In the second project, the diffusive motion of a DNA roboarm on a DNA platform, synthesised by the DNA origami technique, was characterized by monitoring FRET between the donor dye at the end of the roboarm, and acceptor labeled docking strands placed on the opposite sides of the roboarm on the platform. Additionally, a staple strand on the roboarm (arm-dock) was extended to have certain number of nucleotides complementary to the docking strands. The three-color FRET experiments and intensity traces were extracted for the three dyes for the samples with different overlap lengths of the docking strand with the arm-dock. These intensity traces were further subject to an HMM analysis and dwell times of the states of each dye were determined. In the third project, the threading of a single stranded DNA through the meshwork of a hollow cylindrical DNA tube was monitored by FRET between the acceptor-labeled threading strand and the donor-labeled

non-threading duplex. The dynamics were analyzed with an HMM to determine the kinetics of switching of the underlying FRET states. The threading mechanism was further probed by introducing single stranded interfering strands of varying lengths complementary to the threading strand.

The last project involved investigation of the Na^+-K^+ ATPase with different single-molecule fluorescence techniques. The sodium pump was extracted from cellular membrane fragments, fluorescently labeled and solubilized in detergent. Subsequently, the monomer population was separated by employing fast protein liquid chromatography (FPLC) and reconstituted into lipid vesicles. The monomers in lipid vesicles were investigated with TIRF microscopy and MFD-PIE technique. The results suggest that the monomers exhibit kinetic cooperativity that could be supported by the crystal structures of the different conformations.

Contents

1. Introduction

The 21^{st} century has seen the rise of various interdisciplinary fields, notably Biophysics, that combines the perspectives and techniques of physics, chemistry, mathematical analysis and computer modeling to successfully solve complex biological problems. Development of optics and microscopy has vastly contributed to the study of biological systems, which cannot be probed with naked eye. Over the last few decades, techniques like transmission electron microscope (TEM), atomic force microscopy (AFM), and X-ray diffraction (XRD) have been developed to extend the resolution to atomic level. However, they have their limitations as biological samples require native environment to survive. The aim ultimately is to study the biological systems in their most unperturbed form and to quantify the dynamics of the various conformational states observed. This is where fluorescence microscopy and spectroscopy techniques give the necessary edge by extracting the time-resolved molecular properties from fluorescence of specifically labeled bio-molecules in their native state. The fundamentals of fluorescence have been explained in chapter 2.

Till '90s, ensemble fluorescence measurements were employed to study biomolecular interactions, which are limited to reporting the average molecular behavior. Advent of single-molecule fluorescence techniques allowed real-time probing of individual molecules labeled with fluorophores to detect the underlying sub-populations while observing dynamic interconversion of states. The use of single-molecule total internal reflection fluorescence microscopy (TIRFM) in combination with single-pair fluorescence resonance energy transfer (spFRET) and alternating laser excitation (ALEX) allowed access to the structural information and kinetics of dynamic systems through distance measurements.

However, spFRET limited the study of interactions to one dimension, while biological systems often involve three-dimensional interactions between multiple partners. The development of high quantum-yield photostable dyes spanning the entire visible spectra, high numerical aperture objectives, versatile labeling techniques and better optical filters have helped extend spFRET to multiple colors, enabling simultaneous observation of multiple distances. However, multiple colors bring multiple problems, especially in relation to the correction factors required for the multi-color FRET data, which have been addressed in this thesis in chapter 4. Even for spFRET data, individual molecular interactions can be very sophisticated and challenging to analyze, requiring statistical models like hidden markov model (HMM) analysis to extract the dynamic information of states hidden

in noise and understand the kinetics of the underlying FRET states. HMM analysis has been extensively used in this thesis work and the process of optimizing its parameters and the subsequent dwell time analysis has been dealt with in chapter 4.

The most exciting advancement made possible by biophysics is the discovery of the double-helical structure of DNA due to the efforts of Franklin, and Watson and Crick. The highlight of the helix structure is the complementary base sequences and their strict base pairing rules. DNA is the basic building block of DNA nanotechnology, which involves designing and manufacturing 'self-assembling' DNA structures that can be, in turn, used as a tool to solve problems in structural biology and biophysics. Since its conceptualization in the late 20^{th} century by Seeman, the field has grown manifold upon the development of DNA origami technique by Rothemund. This has paved way to building DNA based nano-machines, the pre-cursors for lab-on-chip applications, with the capability to provide a localized high throughput. However, these applications involve several dynamic processes, that must be monitored and understood to allow control over functioning of these DNA based systems. Single-molecule fluorescence techniques provide an easy way to monitor these dynamics with high spatial and temporal resolution. In chapter 5, I employed spFRET and multi-color FRET on a TIRFM along with HMM analysis to study dynamics in DNA based systems in three projects. In the first project, I studied the autonomous switching of a surface tethered molecular construct performing a catalytic hairpin assembly reaction. In the second project, I investigated the diffusive motion of a DNA Roboarm using multi-color FRET with TIRFM. The last project involved analysis of the spFRET measurements of the complex dynamics of a single stranded DNA threading through the wall of a hollow cylindrical DNA tube using HMM.

Biophysical methods have also contributed to protein biosynthesis and solving structure of proteins using X-ray crystallography and nuclear magnetic resonance (NMR). The sodium pump, the first ion pump to be discovered in 1957 by Skou has been extensively studied over several decades through several methods like fluorescence assays, patch clamp experiments and enzyme assays for kinetic studies. However, the crystal structures of several of its conformations are yet to be identified. To throw new light over this very old pump, I investigated it with different single-molecule fluorescence techniques in chapter 6. The sodium pump was extracted from cellular membrane fragments, fluorescently labeled and solubilized in detergent. Subsequently, the monomer population was separated by employing fast protein liquid chromatography (FPLC) and reconstituted into lipid vesicles. The monomers in lipid vesicles were investigated with TIRFM and multiparameter fluorescence detection pulsed interleaved excitation

(MFD-PIE) techniques.

2. Fluorescence

2.1 Light-Matter Interactions

In 1916, Einstein postulated that there are three ways in which light and matter can interact (Figure 2.1):

1. Absorption of a photon by an atom or molecule, promoting an electron from a lower to a higher energy level (Einstein coefficient B_{01}).

2. Spontaneous emission of photon when an electron de-excites to a lower energy level without any external influences (Einstein coefficient A_{10}).

3. Stimulated emission of a photon instigated by the presence of electromagnetic radiation at the frequency of the transition, that induces de-excitation of an electron to a lower energy state (Einstein coefficient B_{10}). Notably, this is the principle process that is utilized in lasers and super-resolution microscopy, in particular, Stimulated Emission Depletion (STED).

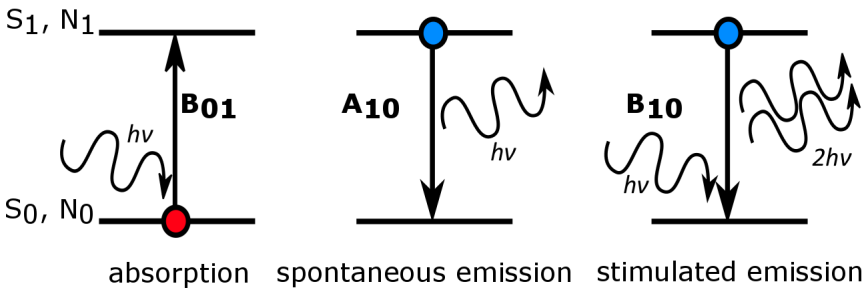

Figure 2.1: The light-matter interactions as postulated by Einstein in a two-level system, with the Einstein coefficients B_{01}, A_{10} and B_{10}. The ground state S_0 has N_0 molecules and excited state S_1 has N_1 molecules.

Assuming that N_0 is the number of molecules in the ground state S_0 and N_1 is the number of molecules in the excited state S_1, and applying detailed balance, one arrives at the following equation:

$$\frac{dN_0}{dt} = \frac{-dN_1}{dt} = -N_0 B_{01} U_\nu + N_1 A_{10} + N_1 B_{10} U_\nu \tag{2.1}$$

where $-N_0 B_{01} U_\nu$ describes the probability of absorption, $N_1 A_{10}$ is the probability of spontaneous emission, $N_1 B_{10} U_\nu$ is the probability of stimulated emission and U_ν denotes the spectral energy density of the isotropic radiation field at the frequency (ν) of the transition. In the upcoming sections, the processes of absorption and spontaneous emission will be discussed in greater detail.

2.2 Fluorescence

Fluorescence is the phenomenon of emission of light by a molecule that has been excited by absorption of a photon. The term Fluorescence was coined in 1852 by Sir George Gabriel Stokes and mentioned in his seminal works on "refrangibility" of light where he studied the properties of Fluor-Spar, a fluorite that absorbs in the UV region and emits blue light.

Various processes involved in excitation and de-excitation of a fluorophore (fluorescing molecule) can be visualized through a *Jablonski diagram* (Figure 2.2). The singlet ground, first and second electronic states are represented by S_0, S_1 and S_2 respectively. In each of these electronic states, the fluorophore can exist in a number of vibrational energy levels. The transitions between states are represented as vertical lines [1].

At room temperature, a fluorophore is found in the lowest vibrational level of the singlet electronic ground state (S_0) as thermal energy (kT) is not sufficient to populate higher electronic and vibrational states. On absorbing light, an electron is excited to a higher vibrational level of an electronic excited state (S_1 or S_2). The electronic transition causes a shift in the electron density of the molecule. This changes the coloumbic force field on the nuclear configuration causing them to rearrange. Since the mass of an electron is negligible compared to the nuclei, the electronic transitions take place much faster ($\sim 10^{-15}$ s) compared to the movement of the nuclei. The approximation that an electronic transition is most likely to occur without changes in the positions of the nuclei in the molecule is called as the *Franck-Condon principle*. However, an electronic transition stimulates vibrations in the previously stationery (*Born-Oppenheimer approximation*) nuclei. Thus, an electronic transition is accompanied by excitation of vibrations of the molecule.

Subsequently, the molecule rapidly relaxes to the lowest vibrational level of S_1. This is called Internal Conversion (IC) and occurs in the timescale of 10^{-12} s. The fluorescence lifetimes are typically in the order of 10^{-8} s, therefore internal conversion is achieved before emission, thus emission generally occurs from thermally equilibrated state, which is the lowest energy vibrational level of excited state S_1 [Lakowicz, 2007].

To bring the molecule back to ground state, there are several competing de-excitation processes: Fluorescence, quenching processes, phosphorescence, and other non-radiative processes.

[1] A vertical line represents a vertical transition, which is based on the the Franck–Condon principle.

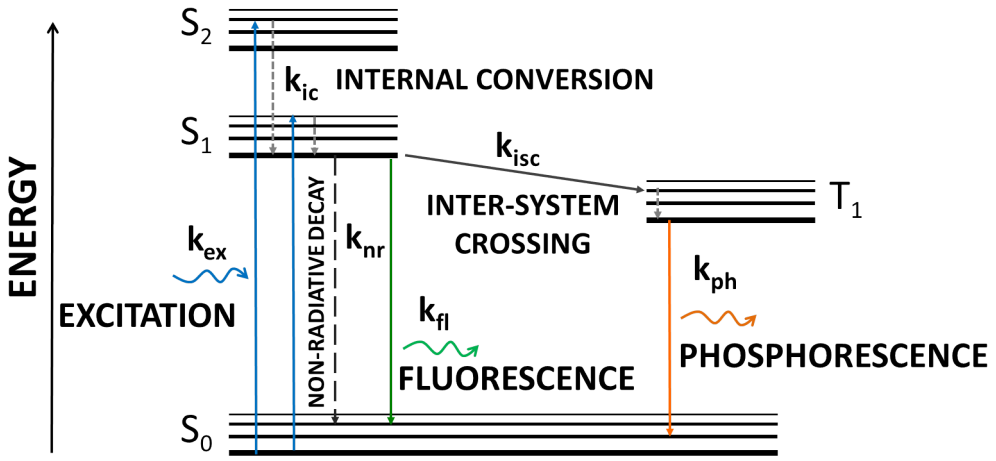

Figure 2.2: A typical Jablonski diagram. The singlet electronic ground state S_0, the excited singlet states S_1 and S_2 and the triplet state T_1 are depicted by thick black lines; the corresponding vibrational states are depicted by thinner black lines. The various excitation and de-excitation pathways are depicted by arrows.

Fluorescence is a radiative (releases energy in the form of a photon) de-excitation process occurring on the timescale of ∼10 ns wherein the molecule returns to a higher vibrational level of S_0, and reaches thermal equilibrium by quickly ($\sim 10^{-12}$ s) relaxing to the lowest vibration level of S_0.

Molecules in the S_1 state can also undergo Inter-System Crossing (ISC), wherein the molecule enters an excited Triplet state (T_1) by spin conversion. The transition from T_1 to S_0 is forbidden and hence the decay times are much longer (ms to hours) than for fluorescence. This usually occurs in molecules containing heavy atoms (e.g. Bromine, Iodine etc.) which have pronounced spin-orbit coupling that expedites ISC, thus increasing phosphorescence quantum yields.

2.2.1 Characteristics of Fluorescence

1. Stoke's Shift
   Due to vibrational relaxation (internal conversion), fluorescence typically occurs from ground vibrational level of S_1. Moreover, the molecule generally decays to higher vibrational levels of S_0. This results in the energy of emission being less than the energy of absorption, thus the wavelengths of emission are longer than wavelengths of absorption. Stoke's shift can be inferred from the Jablonski diagram (Figure 2.2), where it is seen clearly that the energy gap for excitation (blue transition) is greater than energy gap for fluorescence (green

transition). This phenomenon is called the Stoke's shift after Sir G.G. Stokes who first observed it in 1952. Stoke's shift is also deciphered as

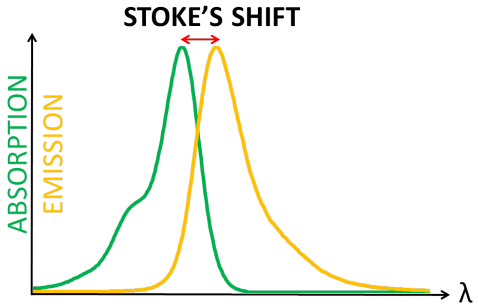

Figure 2.3: The absorption (green) and emission (yellow) spectra of Atto565 are plotted. The difference between the emission peak maximum and the absorption peak maximum illustrates the Stoke's Shift.

the spectral shift between the maxima of the absorption spectra and emission spectra of a fluorophore (as shown for Atto565 in figure 2.3). This spectral shift helps isolate emission of the dye from its absorption by employing optical filters; this is useful for microscopy as described in later sections.

2. Kasha's rule
On excitation to higher vibrational levels of higher electronic states, the molecule quickly relaxes to the vibrational ground level of S_1. Because this rapid relaxation precedes fluorescence, emission always takes place from vibrational ground level of S_1. This makes the emission spectrum independent of excitation wavelength. This is called Kasha's rule [Kasha, 1950]. By extension of this rule, it is observed that quantum yields are generally independent of excitation wavelength [Wawilow, 1924].

3. Mirror-image rule
A fluorophore is usually excited to the higher vibrational levels of a higher electronic state. This transition occurs in approx. 10^{-15} s, which is too fast for any significant movement of the nuclei (Franck-Condon principle). Subsequently, the molecule rapidly relaxes to ground vibrational level of S_1 in approx. 10^{-12} s. This relaxation could be a result of strong overlap among numerous states of nearly equal energy [Lakowicz, 2007]. Since the vibrational relaxation is complete prior to fluorescence, the emission generally happens from the lowest vibrational level of S_1. Furthermore, decay to the ground state occurs to some higher vibrational level of S_0, followed by quickly achieving thermal equilibrium (in approx. 10^{-12} s). This entire process results in emission spectrum being a mirror-image of the absorption spectrum when S_0 and

S_1 states have similar vibrational structure and similar transitions are involved. This is illustrated in the inset of figure 2.4. This similarity

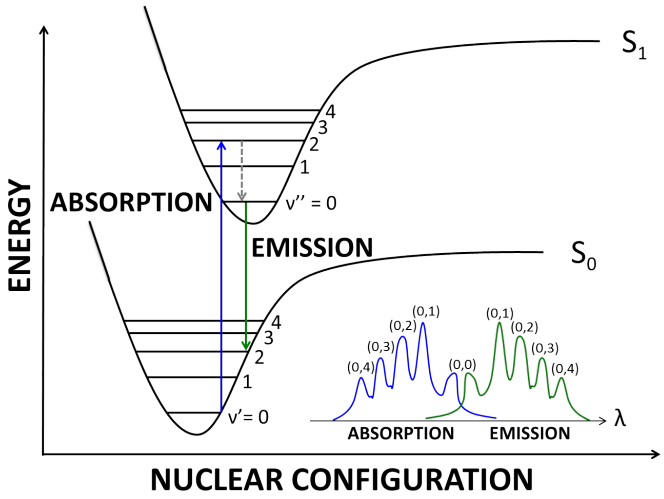

Figure 2.4: Jablonski diagram with nuclear configuration showing electronic transitions as vertical transitions according to the Franck-Condon principle.
Inset: The emission and absorption spectra are mirror images of each other, showing the Mirror-image rule. The transitions which make the absorption and emission spectra mirror-images are marked.

between spectral peaks occurs as a result of the excitation not changing the nuclear configuration, thus spacing between vibrational levels of excited electronic state remains similar to that of the ground state. Therefore, it is generally found that the emission spectrum is the mirror image of $S_0 \rightarrow S_1$ absorption spectrum.

2.2.2 Properties of Fluorophores

Fluorophores have distinct absorption, emission and excitation spectra. In addition to the spectral properties, the other parameters that characterize the photophysics of a fluorophore are:

1. The Extinction Coefficient or Molar Absorption Coefficient
   This is an indication of the strength of a transition. This coefficient depends on the wavelength of the incident radiation and is proportional to the square of the transition dipole moment, which is the theoretical basis of intensity of an electronic transition. When the extinction coefficient is known, the Absorbance[2] measured using a spectrometer

[2]Absorbance (optical density, OD) gives the strength of absorption of the sample and depends on the ratio of intensity of incident light $I_{incident}$ to intensity of transmitted light $I_{transmitted}$ as $A = log_{10} \frac{I_{incident}}{I_{transmitted}}$

allows one to determine the concentration of a fluorophore sample by using the *Beer-Lambert's law*:

$$A = \epsilon \cdot c \cdot l \tag{2.2}$$

where l is the path length (in cm), c is concentration (in M) and ϵ is the extinction coefficient (in $M^{-1}cm^{-1}$)

2. The Quantum Yield
Since fluorescence is one of the many de-excitation processes, every photon absorbed does not result in fluorescence. The fluorescence quantum yield is defined as the ratio of the number of photons emitted to the number of photons absorbed.

$$\Phi = \frac{N_{photons\ emitted}}{N_{photons\ absorbed}} \tag{2.3}$$

3. The Fluorescence Lifetime
In the absence of non-radiative pathways, the lifetime of a fluorophore is called the Intrinsic or Natural lifetime (τ_n), which is the inverse of the fluorescence decay rate k_{fl}.

$$\tau_n = \frac{1}{k_{fl}} \tag{2.4}$$

The fluorescence lifetime or lifetime of an excited state is thus defined as the average time spent by fluorophore in the excited state before returning to ground state. Under normal conditions, the measured lifetime (τ) will be reduced due to presence of other non-radiative pathways (k_{nr} being the non-radiative decay rate). Thus, taking the competing processes into consideration, the lifetime is given by:

$$\tau_n = \frac{1}{k_{fl} + \sum k_{nr}} \tag{2.5}$$

From the rate of the various processes, we can also define the quantum yield:

$$\Phi = \frac{k_{fl}}{k_{fl} + \sum k_{nr}} \tag{2.6}$$

2.3 FRET

Förster Resonance Energy Transfer (FRET) or Fluorescence Resonance Energy Transfer is a distance dependent radiationless transfer of energy from an electronic excited state of a fluorophore (donor) to another (acceptor) through dipole-dipole coupling interactions. The energy transfer pathway from the donor to the acceptor is illustrated in the jablonski diagram in figure 2.5.

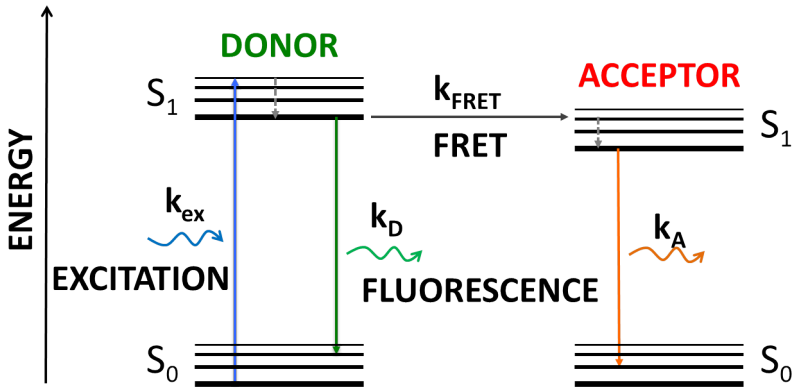

Figure 2.5: Jablonski diagram illustrating transfer of energy from S_1 of donor molecule (on left) to S_1 of acceptor molecule (on right), resulting in fluorescence of acceptor molecule after excitation of donor.

The efficiency of energy transfer is inversely proportional to the sixth power of distance between the fluorophores, making FRET extremely sensitive to small changes of distance in the range of 2-8 nm. FRET is an essential tool for investigating various biological phenomena that involve dynamics of moving components of a system (intra-molecular FRET) or involve multi-component interactions (inter-molecular FRET).

The original derivation of FRET by Perrin [Perrin, 1927] in 1927 modeled the fluorophores as identical oscillators that interact as dipoles only over a distance R between them. But the assumption of identical oscillators implied exact resonance of the line spectra, which is not possible because the absorption and emission spectra are broadened by the solvent over a large frequency range. Since energy must be conserved, it all boils down to calculating the probability that the donor and acceptor molecules will have the same energy (frequency) at the same time. *Theodor Förster* solved this by taking the spectral distribution into account along with the assumption that the interaction energy is lower than the energy of any spectral transition [Förster, 1948]. He showed that resonance between the electronic transitions of the donor and the acceptor after vibrational

relaxation of the donor to the lowest energy excited state brings about the energy transfer, provided some coupling is present [Förster, 1949]. The extent of resonance is gauged by the overlap of the emission spectrum of donor with the absorption spectrum of the acceptor. Moreover, the coupling is strongest when the corresponding electronic transitions are *allowed* (selection rules of spectroscopic transitions).

In near field, the dipole-dipole interactions are dependent on the cube of the distance between the dipoles. The energy transfer probability, being proportional to square of the interaction energy, is thus dependent on the sixth power of separation of the fluorophores. Accordingly, Förster used the probability arguments with the resonance condition and Fermi's golden rule [Dirac, 1927, Fermi, 1950] to derive the correct equation for rate of energy transfer as follows:

$$k_{FRET} = \frac{1}{\tau_D}(\frac{R_0}{R})^6 \tag{2.7}$$

where τ_D is the lifetime of donor, R_0 is the Förster radius for the donor-acceptor pair and R is the their separation distance. The Förster radius is defined as the distance at which the FRET efficiency is 50% and is given by

$${R_0}^6 = \frac{9ln10}{128\pi^5 N_A}\frac{\kappa^2 c^4 Q_D}{n^4}J(\nu) \tag{2.8}$$

where N_A is Avogadro's constant, Q_D is the quantum yield of donor (in absence of acceptor), κ^2 is the orientation factor of dipole, and $J(\nu)$ is the overlap integral and spectra ν is in s^{-1} (from the lecture notes of Robert Clegg).

The FRET Efficiency (E), therefore, depends on three significant factors:

1. The Distance between Donor and Acceptor
   As mentioned earlier, FRET is most sensitive in the distance range of 2-8 nm and FRET efficiency has a sixth power dependence on the distance:
   $$E = \frac{1}{1+(\frac{R}{R_0})^6} = \frac{R_0^6}{R^6+R_0^6} \tag{2.9}$$
   where R is the distance between the fluorophores and R_0 is the Förster radius for the fluorophore pair.

   Figure 2.6 depicts how the FRET efficiency varies with distance between the fluorophores for various dye pairs used in this thesis.

2. The Overlap Integral J
   The spectral overlap integral J incorporates the extent of overlap between the emission spectrum of donor and absorption spectrum of

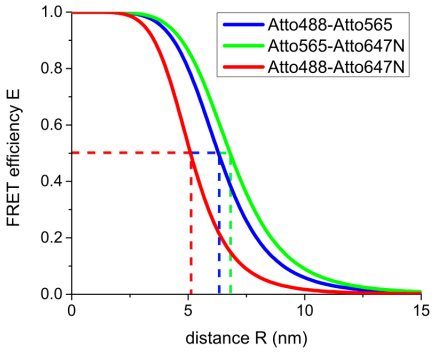

Figure 2.6: E vs R. Representation of how FRET efficiency varies with distance for different dye pairs. The R_0 for each dye pair is determined by the dotted lines. For Atto488 - Atto565 R_0 = 6.3 nm (blue), for Atto565 - Atto647N R_0 = 6.8 nm (green) and for Atto488 - Atto647N R_0 = 5.1 nm (red).

acceptor. The higher the overlap, the greater the probability of FRET between the fluorophores. Besides, the Förster radius (R_0) is also dependent on the overlap integral. J is calculated as:

$$J(\nu) = \frac{\int_0^\infty f_D(\nu)\varepsilon_A(\nu)\nu^{-4}d\nu}{\int_0^\infty f_D(\nu)d\nu} \tag{2.10}$$

where f_D is the donor quantum emission distribution (fluorescence), and ε_A is the acceptor molar extinction coefficient (from the lecture notes of Robert Clegg). If the spectra are taken on wavelength scales, then

$$J(\lambda) = \frac{\int_0^\infty f_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda}{\int_0^\infty f_D(\lambda)d\lambda} \tag{2.11}$$

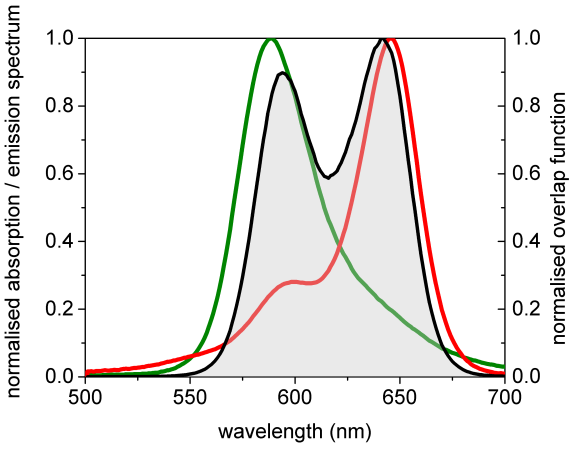

Figure 2.7: Overlap integral J(λ). Visualization of the overlap function (grey area) of the overlap of emission spectrum of the donor Atto565 (in green) and excitation spectrum of the acceptor Atto647N (in red)

3. The Orientation Factor κ^2
   Since FRET involves the interaction between dipoles of the fluorophore pair, the relative orientation between the donor's emission transition dipole moment and the acceptor's absorption transition dipole moment plays a major role. The orientation factor κ^2 is calculated as:

$$\kappa^2 = (cos\theta_T - 3cos\theta_D cos\theta_A)^2 \tag{2.12}$$

   where θ_D and θ_A are the angles made by the donor transition dipole moment (μ_D) and the acceptor transition dipole moment (μ_A) with vector $\overrightarrow{R}$ joining the dipoles respectively. θ_T is the angle between the transition dipole moments μ_D and μ_A as illustrated in Figure 2.8. The DR plane is the plane containing the donor transition dipole moment μ_D and vector $\overrightarrow{R}$, while AR plane is the plane containing the acceptor transition dipole moment μ_A and vector $\overrightarrow{R}$. ϕ is the angle between DR plane and AR plane.

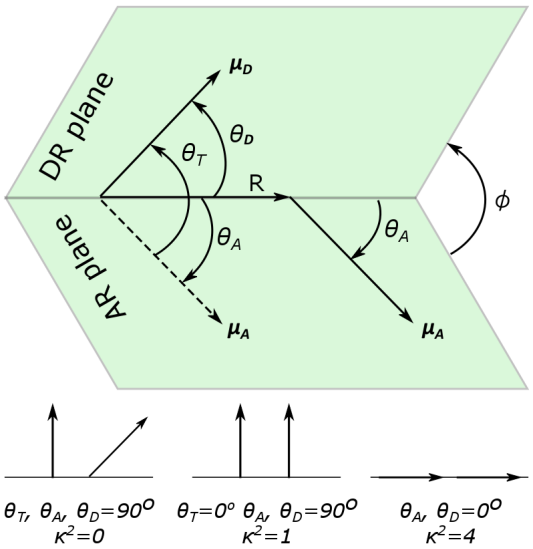

Figure 2.8: The orientation factor. Illustrated are the various angles required for the calculation of κ^2. The three special cases of κ^2 are also described with the corresponding values of the angles involved.

2.4 Fluorescence Anisotropy

Fluorescence anisotropy measurements are based on the principle of preferential excitation of the fluorophores whose transition dipole moments are aligned parallel to the electric field of the excitation light. Therefore, excitation with polarized light results in selective excitation of fluorophores, resulting in partially polarized fluorescence emission. Fluorescence

anisotropy measurements are useful for determining the shape and orientation of the molecule, and the rotational diffusion or the relative angular displacement of the fluorophore during the lifetime of the excited state.

The steady state fluorescence anisotropy (r), which gives the correlation between excitation and emission polarization is defined as:

$$r = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + 2F_{\perp}} \tag{2.13}$$

where $F_{\parallel}$ and $F_{\perp}$ are the fluorescence intensities of the parallel ($\parallel$) and perpendicular ($\perp$) polarized emission with respect to excitation beam. The steady-state anisotropy helps us separate free dye from labeled protein, since freely diffusing dye has steady-state anisotropy values close to 0, as they can rotate many times during the excited state lifetime while dyes bound to protein have reduced rotational diffusion, thus a longer rotational correlational time. One can also perform time-resolved anisotropy measurements, which are useful in determining various components that effect movement of the dye on the molecule. Here, the anisotropy is measured as a function of time t (obtained by using TCSPC) after an excitation laser pulse:

$$r(t) = \frac{F_{\parallel}(t) - F_{\perp}(t)}{F_{\parallel}(t) + 2F_{\perp}(t)} \tag{2.14}$$

In the case of labeled protein, one can clearly separate the motion of dye (which depends on linker length that controls mobility of the dye) from the much slower rotation of the protein alone. Here, the expected decay of the time-resolved anisotropy r is given by the Perrin equation :

$$r = \frac{r_0}{1 + \frac{\tau}{\theta}} \tag{2.15}$$

where r_0 is the measured anisotropy in the absence of rotational diffusion, θ is the rotational correlation time for the diffusion process and τ is the fluorescence lifetime. Therefore, any factor that changes the rate of rotational diffusion will show an effect on the measured anisotropy. This makes the fluorescence anisotropy measurements a useful tool for measuring reaction kinetics and binding constants that change the rotational correlational time of the molecule. It is important to note that the above analysis is valid only in absence of any energy transfer mechanism like homoFRET that lowers the expected anisotropy.

2.5 Single-molecule Fluorescence

Bulk or Ensemble fluorescence spectroscopy provides information about the average molecular behavior of a sample. Therefore, ensemble measurements

of a homogenous sample will give accurate information about the system. However, biological systems are generally non-homogenous, usually consisting of molecules in different states, that may evolve with time. Ensemble measurements of such a non-homogenous sample would not yield an accurate result; only an experimental average of the observable (e.g. fluorescence intensity, FRET etc.) would be obtained, which would not help resolve the molecular behavior of different sub-populations.

Also, the type of heterogeneity in the system matters. The case of static heterogeneity (where the molecules in different states do not interconvert in the time scale of detection of the measured signal) could be possibly resolved by separating the components (i.e. molecules in different states) prior to measurement, by methods like chromatography or gel electrophoresis, depending on the nature of these states. Dynamic heterogeneity, on the other hand, is simply undetectable in bulk measurements as the molecules stochastically switch between states during the course of the experiment. With synchronization, one could possibly detect the different sub-states initially, but progressively, the system will get desynchronized as there is no control over the dynamic interconversion of states, and one will end up detecting an average value. Moreover, short-lived states or sub-populations and rare events cannot be detected.

All the aforementioned problems can be resolved by switching to single-molecule (sm) detection. Single-molecule spectroscopy allows us to probe changes in individual molecules in real time and detect the underlying sub-populations while observing the dynamic interconversion of states. It provides a distribution of states, thus resolving sample heterogeneity and making even rare events and short-lived states discernible. Furthermore, it makes synchronization during measurement redundant as one can always post-synchronize the time trajectories during data analysis [Joo and Ha, 2012]. This makes sm spectroscopy a very powerful tool for investigating biological events.

However, detecting fluorescence from a single molecule (or even a single fluorophore) is technically challenging. The first experiment involving fluorescence detection (indirect) of a single molecule was performed by Rotman [Rotman, 1961], when detecting multiple products of a single enzymatic molecule. The first direct optical detection was performed by Hirschfeld [Hirschfeld, 1976] by labeling a single molecule with multiple fluorophores. This work demonstrated a proof-of-principle for detection of sm fluorescence in solution and furthered the development of the field. Many years later in 1990, Shera et al. [1990] were successfully able to detect single fluorescent molecules (Rhodamine-6G) for the first time. In 1995, Funatsu et al. [1995] demonstrated the first biological application

by detecting sm fluorescence in aqueous medium at room temperature. In 1996, Ha et al. [1996] detected FRET between a single pair of fluorophores, thus opening a plethora of avenues for single-molecule studies in the field of biology. Over the years, the field has benefited tremendously from the technological advancements in lasers, improvement in numerical aperture (NA) of objective lenses, increase in sensitivity of photon detectors, and synthesis of brighter, more photostable fluorophores. This has resulted in an exponential increase in the research papers published [Joo et al., 2008] in the last two decades.

It is of utmost importance to note here that detection of fluorescence signal from a single-molecule is only one half of the coin. The other essential aspect is NOT to detect anything else. This involves suppressing the background (elastic and non-elastic scattering of light by surrounding medium and fluorescence from unwanted sources [Enderlein, 2005]), which is achievable by reduction of the detection volume. This makes two microscopy techniques especially suited for sm detection: Total Internal Reflection Fluorescence Microscopy (TIRFM) and Confocal microscopy. TIRFM involves evanescent wave illumination of single-molecules up to a depth little short of 150-200 nm, thus detecting fluorescence only from the molecules in the immediate vicinity of the surface. The next chapter provides a detailed description of TIRFM and a detailed description of the data analysis follows in the chapter after that.

In contrast, confocal microscopy involves focusing of laser light through an objective to create a very small volume ($\sim$ few fl) in the sample for detection. Since the sample is very dilute (pM-nM range), only the fluorescence of molecules diffusing through the spot is detected and the out-of-focus light is eliminated by a pinhole. This technique is explained further in the Burst-Analysis with MFD-PIE section 2.7 of this chapter.

2.6 Multi-color FRET

FRET between a single donor-acceptor molecule pair has been extensively used to study biological systems [Deniz et al., 1999, Schuler et al., 2002, Zhuang et al., 2000, Zarrabi et al., 2007]. However, single-pair Fluorescence Resonance Energy Transfer (spFRET) limits our probing of proximities and dynamics to only one-dimension while the reactions in biological landscape are seldom so. This can be overcome by extending FRET to multiple fluorophores thus allowing us to monitor multiple distances simultaneously and helping us unravel the many complex dynamics involved in multi-component interactions.
Though advantageous and seemingly straightforward, the realization of

multi-color FRET is in reality tricky. It is not sufficient to have photostable dyes with high quantum yield; a clear separation of the emission spectra of the dyes is required for the detection of fluorescence with a good Signal-to-Noise Ratio (SNR) and to keep the crosstalk between the detection channels to a minimum. In addition, there must be substantial FRET between the dyes, which requires sufficient overlap between the emission spectrum of the donor and the absorption spectra of acceptors. Since these two conditions are difficult to be satisfied concomitantly with the dyes and excitation sources currently available, one must select the fluorophores and their labeling positions intelligently. Though advances have been made regarding specific labeling of proteins with multiple dyes by incorporating unnatural amino acids with bio-orthogonal handles [Noren et al., 1989, Lang and Chin, 2014], the procedure needs to be optimized specifically for the protein-fluorophores system. Moreover, quantitative analysis of multi-color FRET is a challenge as numerous correction factors must be incorporated to compute the precise FRET efficiency.

2.6.1 Multi-color FRET Schemes

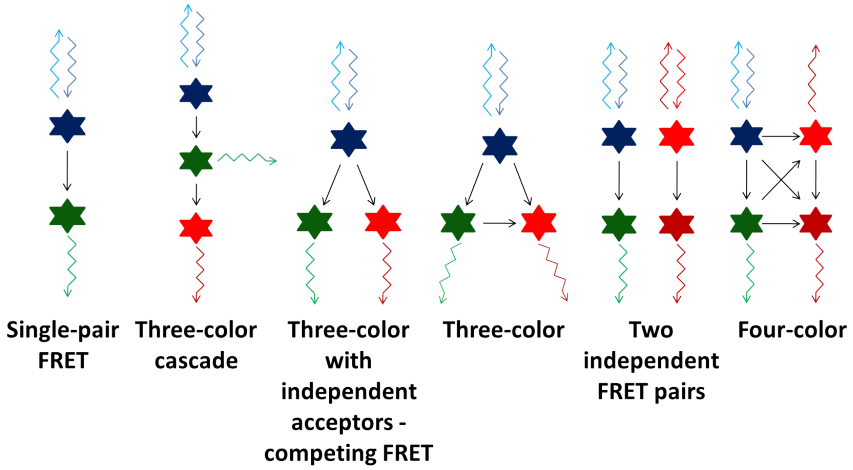

Figure 2.9: Single-molecule multi-color FRET schemes.

Fig. 2.9 illustrates a variety of multi-color FRET schemes. Depending on the application one can play around with the fluorophores and their positions. The various application ideas for the different FRET schemes are illustrated in fig. 2.10. Single-pair FRET is undoubtedly the simplest system and the most extensively used one. It can either be used to probe intramolecular FRET, where two different parts of a biomolecule are labeled. This has been widely applied to probe conformational dynamics of proteins undergoing folding, for example, cold-shock proteins [Schuler et al., 2002], Double Mutant-Maltose Binding Protein (DM-MBP) [Chakraborty

et al., 2010], Hsp 70s−DnaK and SSc1 [Mapa et al., 2010] etc. Another successful application of spFRET has been to study the bending mechanism of double-labeled DNA upon TATA Box Binding Protein (TBP) binding [Wu et al., 2001, Schluesche et al., 2007]. SpFRET has also been used to look at interactions between two labeled binding partners [Funatsu et al., 1995, Vale et al., 1996].

A three fluorophore system can be arranged in a variety of ways:

1. The Three-color Cascade
   A cascade of FRET is formed by placing the first acceptor in between the donor and second acceptor. The distance between the donor and a second acceptor is too large for any FRET to occur. Thus, on exciting the donor, the energy is transferred to first acceptor, which transfers some of the energy to the second acceptor.

2. The Three color Competing FRET
   Two acceptors are placed far apart, beyond FRET regime, and they compete for energy transfer from the excited donor, which is placed in the middle.

3. The Three-color Triangle
   The dyes are placed in a triangle fashion [Clamme and Deniz, 2005, Hohng et al., 2004, Lee et al., 2010b], with the donor transferring energy to both acceptors, and the first acceptor also transferring energy through FRET to the second acceptor.

For the case of four fluorophores, the arrangement can either involve two independent FRET pairs (RecA mediated strand displacement [Lee et al., 2010a] or a complete four-color FRET system with multiple interacting labeled partners (Multi component protein system [Ratzke et al., 2014], movement of $Mut - \alpha$ on DNA [DeRocco et al., 2010]) or multiple labels on the same molecule (Investigation of movement of Holliday junctions by labeling each arm [Lee et al., 2010a]). A four-color FRET system involving two competing three color-cascades, where the energy transfer path is determined by the position of the first acceptor has also been implemented [Stein et al., 2011].

2.6.2 Alternating Laser Excitation - ALEX

For spFRET experiments, usually only one laser is used to excite the donor and the fluorescence collected is spectrally separated into donor and acceptor detection channels. A decrease in FRET observed in such a case can occur due to different possibilities: dynamics of the system or acceptor bleaching or donor blinking (going into dark state or triplet state). This ambiguity can be clarified by direct excitation of the acceptor to confirm it

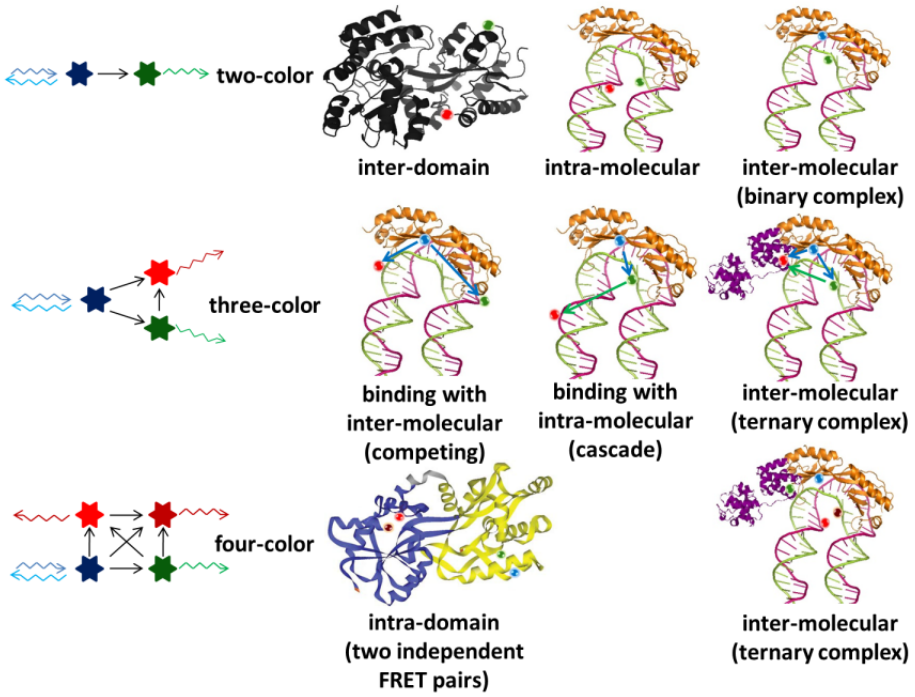

Figure 2.10: Realization of various multi-color FRET schemes.
Top panel: spFRET for probing intramolecular FRET between different domains of a protein (left), between double-labeled DNA and TBP (middle), and inter-molecular FRET between labeled TBP binding to single labeled dsDNA (right).
Middle panel: Three-color FRET for probing binding with competitive inter-molecular FRET (left), or in cascade combined with intra-molecular FRET (middle) and a ternary complex with three labeled partners (right).
Bottom panel: Four-color FRET probing intra-domain dynamics with two independent FRET pair, one in each domain of the protein (left) and a ternary complex with two single-labeled and one double-labeled interacting partners (right).

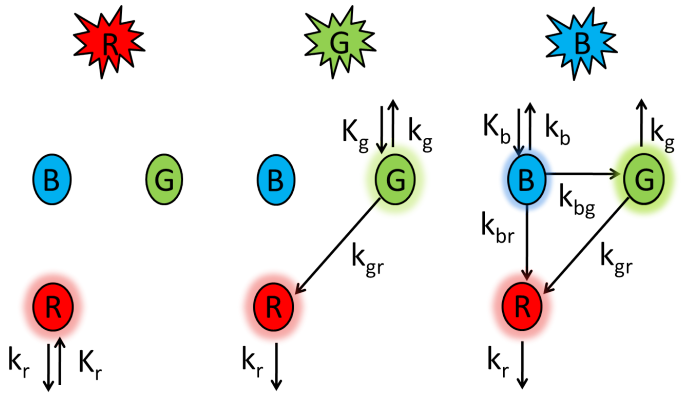

Figure 2.11: Alternating laser excitation with Red-Green-Blue lasers in a three-color FRET system. The fluorescence and FRET rates obtained from each excitation are illustrated.

is present and photoactive. Kapanidis et al. [2005] proposed the Alternating Laser Excitation (ALEX) mechanism, where two continuous wave lasers (one for exciting the donor and other for acceptor) are alternated on the μs timescale (hence, μsALEX), thus independently probing the state of the acceptor to confirm the true source of change in FRET, if any is observed.

ALEX can be extended to multi-color FRET by alternating between lasers that excite each fluorophore independently. This has the added advantage of assimilating additional information from a single experiment. This can be understood through the schematic representation of ALEX in a three-color FRET system shown in figure 2.11. The lasers would be alternatingly switched ON, one laser per frame (assuming TIRF excitation and simultaneous EMCCD detection of all dyes) and the sequence usually starting from the last acceptor to the donor, being $R-G-B$ in this case. On exciting with red laser, the red dye is probed, confirming its presence and its fluorescence is detected. Next the green laser is turned ON, thus probing the green dye, determining its presence and fluorescence signal. Additionally, in the case of FRET, fluorescence from the red dye will also be observed. This step determines the spFRET between the green and red dye. When the blue laser is switched ON, the fluorescence of blue dye is detected, and also FRET to the green and/or red dyes, from the signal in green and red detection channels respectively. Here, the FRET rate between green and red dye determined from the previous green excitation can be taken into account, thus allowing us to focus only on the blue-green FRET and blue-red FRET. In this way, through simultaneous continuous acquisition in all channels, we can determine the source of signal in each channel based on the excitation laser in the ALEX sequence. This helps in the calculation of the individual FRET efficiencies of each FRET pair involved. However, determining the actual FRET efficiencies (including the various correction factors involved) is a process wrought with other problems, which will be dealt with in the chapter 4.

2.7 Burst Analysis with MFD-PIE

As described earlier, single-molecule burst analysis on a confocal microscope allows us to observe molecules as they freely diffuse in solution through the confocal spot. Time-Correlated Single Photon Counting (TCSPC) measurements described in this dissertation were performed on a confocal system based on an inverted Nikon Eclipse microscope using Pulsed Interleaved Excitation (PIE) and Multiparameter Fluorescence Detection (MFD). In this section, I give a brief introduction to confocal microscopy, a short description of the MFD-PIE setup, followed by the principles of TCSPC, data analysis using Burst search and important parameter

calculations for individual bursts employed specifically for sodium pump measurements described in Chapter 6. There is a depth of information that can be revealed by MFD-PIE measurements which is beyond the scope of this dissertation but has been discussed in great detail in papers and book chapters published from our group [Müller et al., 2005, Kudryavtsev et al., 2012, Hendrix and Lamb, 2012, Barth et al., 2014].

2.7.1 Confocal Microscopy

For the successful detection of single-molecules, an important issue to be addressed is suppression of scattered excitation light, which has two components - Rayleigh scattering and Raman scattering. Rayleigh scattering (elastic scattering) can be efficiently reduced by utilizing appropriate emission filters in the detection path. On the other hand, Raman scattering is difficult to eliminate due to its significant spectral overlap with the fluorescence signal. Though with the latest band pass filters, it is possible to eliminate it to an extent, but we end up losing precious signal. Since intensity of scattered light is proportional to the number of scattering molecules, the most effective way to suppress scattering is by decreasing the detection volume, which is the fundamental basis of a confocal setup. The general scheme of a confocal setup is depicted in figure 2.12.

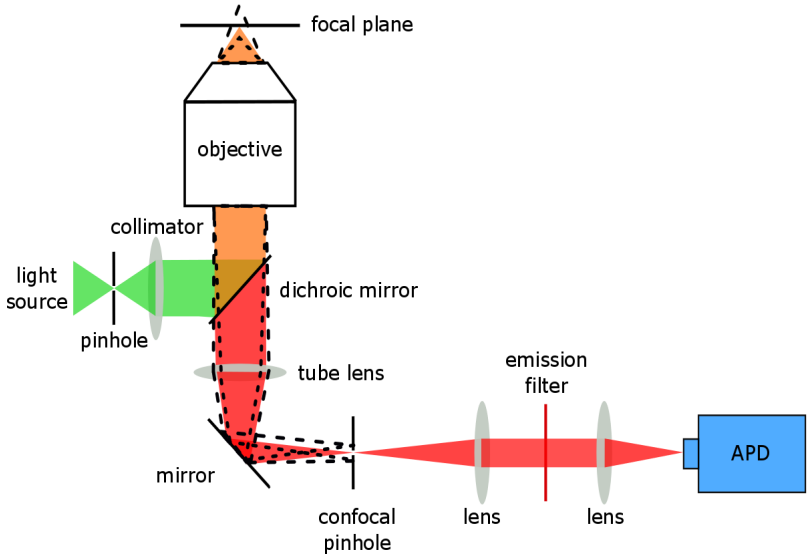

Figure 2.12: Schematic of a confocal setup. The excitation laser (green) goes through a pinhole and is reflected into the objective by a dichroic. Fluorescence from the sample plane passes through the confocal pinhole onto the detector (APD) while the light from above (dashed) or below (dotted) the plane is blocked.

Invented by Marvin Minsky in 1961, confocal microscopy involves focusing light from a light source (typically lasers) through a lens (typically objective with high numerical aperture (NA)) into a diffraction-limited spot in the

sample which has a very small volume (∼1 fl). If the light source is non-collimated, a pinhole is necessary in the excitation pathway to generate a good beam profile and ensure that the excitation is limited to a very small region in the focal plane (where the intensity is a maximum). The detection volume can be approximated as an ellipsoid, and, for single photon excitation, the point spread function (PSF) can be modeled by a 3D Gaussian:

$$PSF(x, y, z) = exp(-\frac{2(x^2 + y^2)}{\omega_r^2} - \frac{2z^2}{\omega_z^2}) \tag{2.16}$$

where ω_r and ω_z are the distances from center along lateral and axial axis, respectively where the detection efficiency drops to $\frac{1}{e^2}$ of that at the center. The fluorescence signal is collected by the same lens and separated from excitation light by a dichroic mirror and through the detection pinhole onto a detector, typically, an avalanche photodiode (APD). The pinhole in the detection pathway is placed in the same image plane as that of excitation spot focus, thus making the excitation and detection planes con-focal. This has the advantage of removing the fluorescence that does not originate from the excitation focal plane (see dotted lines in the figure 2.12), thus increasing the signal to background ratio by reducing the background effectively. Moreover, this method can be extended to samples immobilized on a surface by scanning the focused laser over surface. This is known as confocal laser-scanning microscopy (CLSM) where one can scan along lines, or get a two-dimensional image by scanning in both x and y directions or extend it into the third dimension by scanning planes at different locations along optical z axis. The evident downside is that acquiring a complete image takes much longer compared to wide-field (WF) or TIRF detection on a charge coupled device (CCD) camera.

2.7.2 Pulsed Interleaved Excitation - PIE

Although fluorophores can be excited with any light source, lasers are the ideal choice for single-molecule fluorescence experiments, providing a collimated and monochromatic excitation beam. For spFRET burst analysis experiments, usually only one laser is used to excite the donor and the fluorescence collected is spectrally separated into donor and acceptor detection channels. The disadvantage here is that one cannot determine the actual reason for change in FRET as the molecule diffuses through the confocal volume; whether it is because of dynamics or acceptor bleaching or donor blinking (going into a dark state or a triplet state). Also, it is difficult to separate the low and no FRET populations from donor only. One can overcome this by confirming the presence of an active acceptor by exciting it directly. Kapanidis et al. [2005] proposed the alternating laser excitation (ALEX) mechanism, where two continuous wave lasers (one for exciting the donor and other for the acceptor) are alternated on the μs timescale

(hence, μsALEX). Müller et al. [2005] used subnanosecond pulsed lasers, alternating on the ns timescale and called it Pulsed Interleaved Excitation (PIE). Though both ALEX and PIE provide information regarding the presence and environment of the fluorophores, PIE has the additional lifetime information.

In our lab, we have a two-color and a three-color MFD-PIE setup. In the two-color setup, we use a 532 nm solid state laser (frequency doubled 1064 nm, PicoTA 530, Toptica, München, Germany) and a 640 nm diode laser (LDH-D-C-640, PicoQuant, Berlin, Germany) emitting short laser pulses(∼300 ps) at a repetition rate of 26.66 MHz, which corresponds to a 37.5 ns duty cycle. The lasers are triggered alternatingly by a computer-controlled Multichannel Picosecond Diode Laser Driver 'Sepia I' (PicoQuant, Berlin, Germany), with the red laser being delayed by approx. 15 ns relative to the green laser. The Sepia provides the sync pulse for the TCSPC cards.

The lasers are fibercoupled to single mode fibers (Schäfter+Kirchhoff, Germany) and combined using a wavelength division multiplexer (WDM-12P-111-532/647-3.5/125-PPP-50-3A3A3A-3-1,1,2, OZ Optics) and then collimated (60FC-4-RGB11-47, Schäfter+Kirchhoff). The linearly polarized light is further cleaned by a Glan-Thompson polarizer (GTHM Polarizer, Thorlabs, Dachau, Germany) and reflected by the multiline dichroic mirror (DualLine z532/635, AHF Analysentechnik, München, Germany) into the objective (Plan Apo IR 60x A/1.27 WI, Nikon, Germany) which focuses the light onto the sample. The typical laser powers used are 100 μW after the polarizer.

In the three color setup, a 480 nm pulsed diode laser (LDH-D-C-485, PicoQuant), the tunable frequency-doubled fiber-based picosecond pulsed laser (FemtoFiber and SHG, Toptica Photonics) operated at 565 nm (at a fixed frequency of 27.4 MHz) and the the 641 nm pulsed diode laser (LDH-D-C-640, PicoQuant) are coupled into a polarization maintaining single-mode fiber (PM488PM-FC-5, Thorlabs), collimated (Schäfter und Kirchhoff), then reflected by polychroic mirror (zt405/488/561/633, AHF Analysentechnik) and finally focused into the sample by a 60x 1.27 NA water immersion objective (Plan Apo IR 60x 1.27 WI, Nikon).

2.7.3 Multiparameter Fluorescence Detection - MFD

The optimal way to get the maximum amount of information about several molecular properties simultaneously from detected photons is by using MFD. Just with a single experiment, one can obtain information regarding fluorescence lifetime, fundamental anisotropy, quantum yield,

FRET efficiency and the separation between the donor and acceptor fluorophores [Widengren et al., 2006]. Realization of MFD essentially involves separating the detected fluorescence according to polarization (parallel and perpendicular) and wavelength (colors) and using different detectors for each.

In both MFD-PIE setups, fluorescence from the sample is collected by the same objective that focuses the laser onto the sample (epifluorescence microscopy) and separated from the excitation beam by the same polychroic mirrors used in respective setups for reflecting the excitation beam into the objective.

In the two-color MFD-PIE setup, the collected fluorescence is then focused by an achromatic lens (f=200 mm) onto a 50 μm pinhole (OWIS GmBH, Germany) to remove out-of-focus light. The light is re-collimated by another achromatic lens and then split first by polarization using a polarizing beamsplitter (05FC16PB.3, Newport, Darmstadt, Germany) and then by color using dichroic mirrors (2x BS 650, AHF Analysentechnik) and cleaned up with emission filters (green: BrightLine HC 582/75 and red: ET 700/75, AHF Analysentechnik). The fluorescence is finally focused onto the active area of four single photon counting avalanche photodiodes ((2x SPCM-AQR-14, and 2x SPCM-AQR-16, Perkin Elmer, Massachusetts, USA).

In the three-color setup, the fluorescence is also focused onto a 50 μm pinhole and subsequently separated by polarization using a polarizing beamsplitter (PBS251, 420-680 nm, Thorlabs) and then into three color channels by dichroic mirrors (2x BS560 and 2x 640DCXR, AHF Analysentechnik) with additional filtering by emission filters (red: ET670/30; yellow: ET607/36; blue: ET525/50, AHF Analysentechnik) and eventually focused onto 6 single photon counting avalanche photodiodes (yellow: 2x SPCM-AQR-14; red: 2x SPCM-AQRH-14, Perkin Elmer, and blue: 2x COUNT-100B, LaserComponents).

2.7.4 Time Correlated Single Photon Counting - TCSPC

The photons detected by the APDs from both setups are registered by time correlated single photon counting (TCSPC) data collection cards (SPC-154, Becker & Hickl GmbH, Berlin, Germany). For the three-color setup, the three channels: red parallel, green parallel and green perpendicular are combined onto one TCSPC card using a router (HRT-82, Becker & Hickl). Synchronization of lasers and TCSPC cards is performed by the diode laser driver (SepiaII, PicoQuant) at the fixed frequency of the yellow laser of 27.4

MHz.

For the two-color setup, each detection channel has its own TCSPC card, which helps avoid artifacts arising from routing electronics [Müller et al., 2005]. To start the acquisition, the TCSPC modules are triggered simultaneously by an external trigger (USB-6008, National Instruments). The modules use constant fractions discriminators (CFD) to get the arrival times of the pulses (independent of pulse amplitude) [Sikor, 2012]. Afterwards, a time-to-amplitude converter (TAC) measures the time between APD pulse and sync pulse. The sync pulse is generated by 'Sepia I' and for each TCSPC module, the arrival time of photon relative to sync pulse (*Microtime*) is recorded, along with the number of sync pulses since the last photon in that channel (*Macrotime*). This is called the time-tagged time resolved (T3R) mode, and for each photon, the Macrotime and Microtime alongwith routing and overflow information are stored into a .spc file for each module [Becker, 2014].

2.7.5 MFD-PIE Data Analysis

The single-molecule burst data collected by the TCSPC modules was analyzed by a home built MATLAB program named PAM (PIE Analysis with MATLAB). Firstly, PIE is used to assign the correct excitation source to each photon. The microtime information is used to further sub-divide the red parallel and perpendicular channels in red detection after red excitation (RR) and red detection after green excitation (GR). Together with the parallel and perpendicular green detection after green excitation (GG) channels this results in a total of 6 independent photon streams available for data analysis for the two-color setup (RR$\parallel$, RR$\perp$, GR$\parallel$, GR$\perp$, GG$\parallel$ and GG$\perp$) as exemplarily depicted in fig.2.13. For the three-color setup, there are 12 independent photon streams (BB$\parallel$, BB$\perp$, BG$\parallel$, BG$\perp$, BR$\parallel$, BR$\perp$, GG$\parallel$, GG$\perp$, GR$\parallel$, GR$\perp$, RR$\parallel$ and RR$\perp$).

The next step in the analysis is distinguishing the photon stream of a burst, coming from a single molecule from the background (section 2.7.5). Burst detection paves the way for calculating the molecular properties from the parameters of the collected photons.

Burst Search

Since the freely diffusing molecules reside for a very short time (∼0.5-5 ms) in the confocal volume, the number of photons detected per molecule (typically ∼50-500 photons) and channel (typically ∼0-300 photons) is rather low. Hence, the detection of single-molecule events and distinguishing the burst originating from a fluorescent molecule from background is very

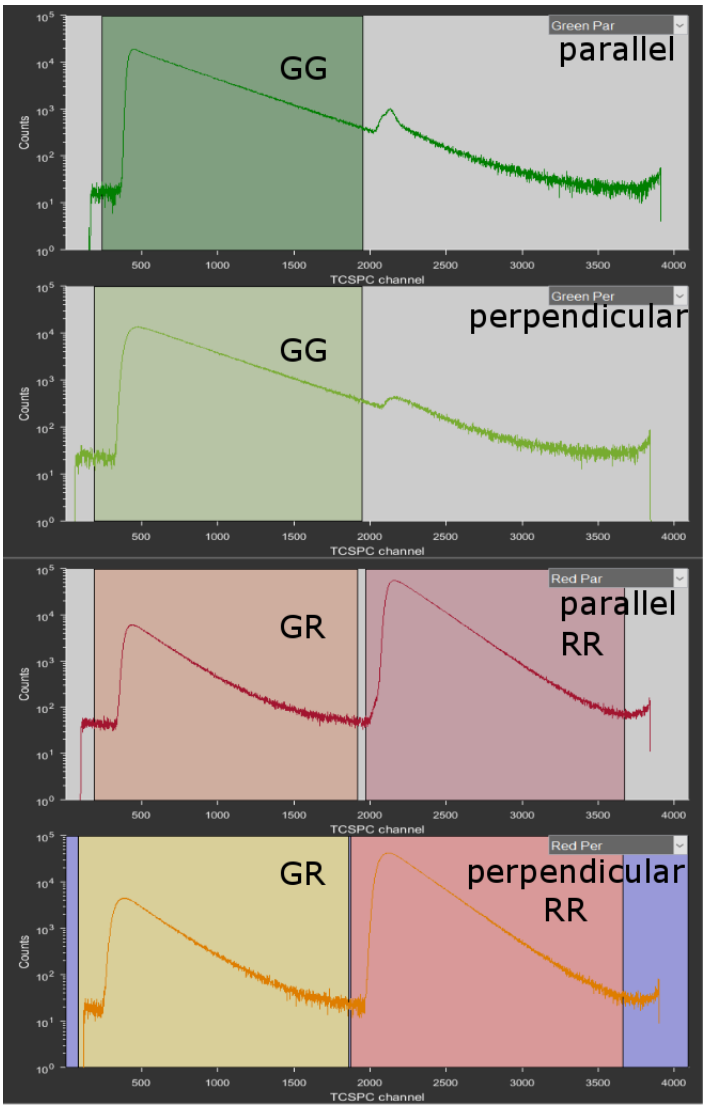

Figure 2.13: Microtime histogram for the four detection channels on the two-color MFD. An overlay of microtime ranges is used to differentiate between green detection after green excitation (GG), red detection after green excitation (GR) and red detection after red excitation (RR). The first channel (topmost) shows green parallel, the second channel shows green perpendicular, the third channel shows red parallel and the fourth channel (bottom most) shows red perpendicular. Data provided by Ganesh Agam.

important. Currently, there are three widely used algorithms for burst detection and all of them rely on the overall photon stream independent of the channel in which the photons were detected in.

The most straight forward approach is based on Thresholding [Müller et al., 2005]. This involves dividing the total photon stream into time bins with a length T (on the order of the burst duration). All photons in a time bin are summed up and bursts are selected by requiring a minimum number of

photons. Information can be optimized by applying a lower threshold to bins next to detected bursts, hence reducing the possible loss of data by partially filled bins.

In another approach, photon bursts are selected using an all-photons burst search (APBS) algorithm [Nir et al., 2006]. This method applies a sliding window of fixed duration T (instead of simply binning the whole data trace) with a minimum number of M photons to be detected in a time window as part of a burst. Additionally, only the bursts with a minimum of L photons are accepted.

A similar method has been suggested by [Enderlein et al., 1997] that involves applying a Lee Filter to smooth the data followed by a threshold of the inter-photon times. Compared to APBS, the Lee filter weighs the averaging with the standard deviation of the inter-photon times. The experiments in this thesis were analyzed by the Lee filter algorithm, with the burst search parameters usually set as follows: Inter-photon time threshold after filtering $\delta t = 0.01 - 1$ ms, size of averaging window for Lee filter m = 30, weighting parameter for averaging, $\sigma_0 = 10$, and the minimum number of photons per burst $L > 80$. Photons from each burst have three important properties: macrotime, microtime, and channel. This information can be used to calculate the various parameters explained in upcoming sections.

Correction Factors

To correct for background, crosstalk, direct excitation, and different detection sensitivities, different correction factors to the recorded raw photon counts need to be considered. The corrections for two-color setup were done according to Kudryavtsev et al. [2012] and extended to three-colors, as described in Barth [2013]. With proper corrections, the different parameters of the bursts can be accurately calculated.

FRET Efficiency

The FRET efficiency is calculated from the number of photons originating from the donor (after correction) F'_{DD} and from the acceptor after donor excitation F'_{DA} as:

$$E = \frac{F'_{DA}}{F'_{DA} + F'_{DD}} \tag{2.17}$$

Prior to this calculation, the number of actually detected photons F'_{DA} was corrected for spectral crosstalk (β, donor fluorescence detected in the acceptor channel) and direct excitation (α, excitation of the acceptor by donor laser) as follows:

$$F'_{DA} = F_{DA} - \beta \cdot F_{DD} - \alpha \cdot F_{AA} \tag{2.18}$$

Furthermore, the differences in quantum yields of donor ϕ_D and acceptor ϕ_A and the detection efficiencies of the detection channels η_D, η_A were accounted for through the gamma factor $\gamma = \frac{\phi_A \cdot \eta_A}{\phi_D \cdot \eta_D}$

$$F'_{DD} = \gamma \cdot F_{DD} \tag{2.19}$$

Therefore, the corrected FRET efficiency is calculated as

$$E = \frac{F_{DA} - \beta \cdot F_{DD} - \alpha \cdot F_{AA}}{F_{DA} - \beta \cdot F_{DD} - \alpha \cdot F_{AA} + \gamma \cdot F_{DD}} \tag{2.20}$$

These equations above are for spFRET and can be extended to three-color as described in [Barth, 2013]

Stoichiometry

The stoichiometry S is defined as the ratio of the total photons detected after donor excitation to the total number of detected photons. It is proportional to the ratio of number of photons absorbed by the donor to the total number of photons absorbed, and is calculated as:

$$S = \frac{\gamma \cdot F_{DD} + F_{DA} - \beta \cdot F_{DD} - \alpha \cdot F_{AA}}{\gamma \cdot F_{DD} + F_{DA} - \beta \cdot F_{DD} - \alpha \cdot F_{AA} + F_{AA}} \tag{2.21}$$

Thus, $S = 1$ for molecules with no active acceptor and $S = 0$ for no active donor. This allows separation of double-labeled molecules from single-labeled ones with $0.2 < S < 0.8$ as shown in fig. 2.14. There is no need, a priori, to correct for the S value. However, it has the advantage of standardizing the analysis which involves correcting the E value.

Even double-labeled molecules may undergo blinking or photobleaching during a burst, which results in falsified E and S values. To filter out such molecules from the analysis, the parameter $T_{DX} - T_{AA}$ is employed. T_{DX} is the mean macroscopic arrival time of all photons after donor excitation and T_{AA} is the mean macroscopic arrival time of all photons from the acceptor after direct excitation, and are given by

$$T_k = \frac{1}{N_k} \sum_{i=1}^{N_k} t_i \tag{2.22}$$

where t_i is the macrotime of photon i and the summation runs over all N_k relevant photons.

If the acceptor photobleaches during diffusion of molecule through detection volume, the mean arrival time of the photons after acceptor excitation will be shifted to earlier times compared to the average arrival time of the photons after donor excitation. Therefore, its possible to exclude such molecules from the analysis by applying the restriction $|T_{DX} - T_{AA}| < 1$ ms.

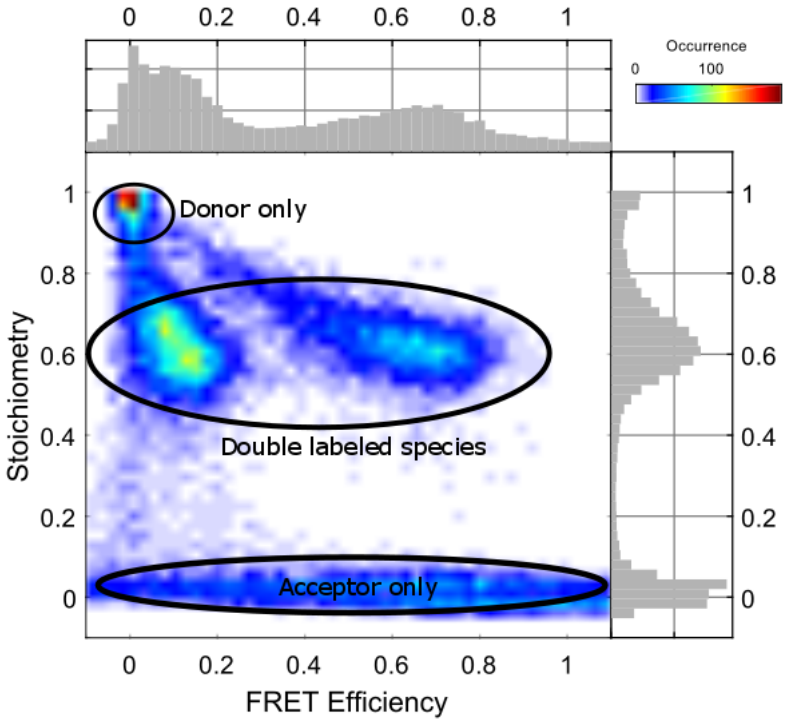

Figure 2.14: A two-dimensional Stoichiometry S vs FRET efficiency E histogram. The sample contains two dsDNA species with different interdye distances (one with $E = 0.3$ and other with $E = 0.7$). The sample clearly shows donor only impurities (top left) and acceptor only impurities (bottom). In addition, two double labeled species are visible in the middle, with high and low FRET efficiencies. Data provided by Ganesh Agam.

Fluorescence Lifetime

The donor fluorescence lifetime for individual bursts is determined from a maximum likelihood estimator (MLE) [Schaffer et al., 1999, Kudryavtsev, 2006] which assumes the fluorescence intensity convoluted with an experimentally determined instrument response function (IRF) decays as a single-exponential. The IRF is determined by measuring light scattered by a water sample. Given the low number of photons available for lifetime determination, the MLE yields a more accurate result than a least−squares fit. For MLE, it is assumed that the time-dependent model *M(t)* for the signal is composed of a fraction α of background signal from scattered light *B(t)*, and $1-\alpha$ of the signal from a single exponential fluorescence component $F(t)$, given by the convolution of the single exponential with the instrument response function (IRF), $F(t) = IRF \bigotimes e^{-t/\tau}$. The MLE determination of the lifetime is then based on the calculation of likelihood function:

$$L(F|M(x)) = \prod_{i=1}^{k} w(F_i|M_i(x)) \tag{2.23}$$

where F is the experimental decay histogram, x is a set of model parameters

for model M, w is the probability of detecting F_i photons in the ith TCSPC channel, and k is the number of the TCSPC channels. The detailed derivation is provided in [Kudryavtsev, 2006]. The acceptor lifetime τ_A is determined by the same MLE.

Fluorescence Anisotropy

The anisotropies are related to the rotational mobility of the dyes. The anisotropies do not influence the Förster radius. Rotational averaging allows us to get around the ambiguities of not knowing the relative separations and orientation of the two dipoles. Using anisotropy, we can decrease our uncertainty. The fluorescence anisotropies of the donor r_{DD}, acceptor r_{AA} and FRET r_{DA} are calculated from the respective intensity ratios of the photons polarized perpendicular or parallel to the excitation beam as:

$$r_i = \frac{G \cdot F_i^{\parallel} - F_i^{\perp}}{(1-3I_2)G \cdot F_i^{\parallel} + (2-3I_1) \cdot F_i^{\perp}}; i = \{DD, AA, DA\} \tag{2.24}$$

where G is an effective gamma factor for correcting the difference in detection efficiencies of the parallel and perpendicular channels and I_1 and I_2 are the correction factors for depolarization effects of high NA objective.

3. Single-molecule TIRF Microscopy

The major challenge in single-molecule (sm) spectroscopy is detecting the fluorescence from a single molecule alone and to separate it notably from the ambient background noise. The ability to do so is quantified by the factor called Signal-to-Noise ratio (SNR), which is defined as the ratio of fluorescent signal to the fluorescent background. Maximizing the SNR for sm fluorescence detection predominantly involves either increasing the photons of the signal or decreasing the fluorescent background or if possible both.

Total internal reflection fluorescence microscopy (TIRFM) improves the SNR for sm studies in comparison to confocal or wide-field microscopy by primarily using the evanescent wave from a totally internally reflected laser beam to limit the excitation volume to molecules that are tethered to the surface and close to interface (∼150-200 nm from surface). This minimizes the background as the bulk of sample is not illuminated. Moreover, immobilization of the single molecules allows one to monitor their dynamic conformational changes over longer timescales (seconds to minutes) compared to confocal microscopy, where the observation times are limited to the diffusion times of single molecules through the focal spot. For the majority of work in this thesis, I have employed smTIRF with one-color, two-color or multi-color detection to investigate the dynamics of protein and DNA-based systems.

3.1 TIRF Excitation

3.1.1 Principles of Excitation

Refraction is the phenomenon of bending of light when it passes through an interface between media of different refractive indices. The relation between the refractive indices (n_1 and n_2) and the angle of incidence (θ_1) and refractance (θ_2) at the optical interface is given by Snell's law and illustrated in figure 3.1:

$$n_1 sin\theta_1 = n_2 sin\theta_2 \tag{3.1}$$

When a laser beam traveling in a high refractive index n_1 medium, impinges on a planar interface with low refractive index medium n_2, it is refracted at the surface, with the the angle of refractance θ_r determined from the angle

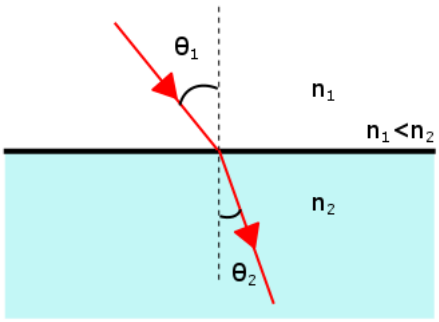

Figure 3.1: Refraction of light at the interface between two media of different refractive indices with $n_2 > n_1$. The angle of refraction θ_2 is therefore smaller than the angle of incidence θ_1

of incidence θ_i by Snell's law. As the angle of incidence increases, $\frac{n_2 sin\theta_r}{n_1}$ approaches 1 and thus θ_r approaches 90^o. Beyond this angle, refraction can no longer occur. The incident light gets totally internally reflected into the high refractive index medium when the incidence angle θ_i is greater than the critical angle $\theta_C = sin^{-1}\frac{n_2}{n_1}$. [1] This creates a surface-associated evanescent electromagnetic field in the lower refractive index medium, that propagates parallel to the interface and penetrates into the lower refractive index medium, decaying exponentially. This evanescent wave is capable of exciting fluorophores present near the interface. The evanescent electric field intensity *I(z)* as a function of perpendicular distance z from the interface is given by:

$$I(z) = I_0 exp(\frac{-z}{d}) \tag{3.2}$$

where I_0 is the intensity of beam at the interface and d is the characteristic penetration depth into lower refractive index medium. The penetration depth d is a function of the vacuum wavelength λ_o of incident light, the refractive indices of the media and the angle of incidence θ_i (for $\theta_i > \theta_C$):

$$d = \frac{\lambda_0}{4\pi(n_1^2 sin^2\theta_i - n_2^2)^{\frac{1}{2}}} \tag{3.3}$$

Penetration depth d is independent of the polarization of the incident light and increases as θ_i decreases from 90^o to θ_C . Except for $\theta_i \simeq \theta_C$ (where $d \to \infty$), d is on the order of λ_o or smaller [Axelrod et al., 1984]. Based on electromagnetic wave nature of light, one can understand the behavior of evanescent electric field. I_0, the intensity at $z = 0$, is proportional to the square of amplitude of the evanescent electric field at $z = 0$. I_0 depends

[1] The critical angle is the special case of refraction where the refracted light travels along the interface and thus obtained by substituting $\theta_r = 90^o$ in Snell's law above.

on both the incident beam polarization and the angle of incidence θ_i. Thus the parallel (p-polarized) and perpendicular (s-polarized) intensities of the evanescent wave at the interface ($z = 0$) are given by:

$$I_P(0) = \imath_P \frac{4cos^2\theta_i(2sin^2\theta_i - n^2)}{n^4cos^2\theta_i + sin^2\theta_i - n^2} \tag{3.4}$$

$$I_S(0) = \imath_S \frac{4cos^2\theta_i}{1 - n^2} \tag{3.5}$$

where $\imath_P$ and $\imath_S$ are the incident electric field intensities with polarizations parallel and perpendicular respectively to the plane of incidence and $n = \frac{n_2}{n_1} < 1$.

$I_P(0)$ is always more intense than $I_S(0)$ for all angles, and both tend to zero as the angle of incidence tends to 90^o. When the incident light is p-polarized, the evanescent wave is elliptically polarized in the plane of propagation. When the incident light is s-polarized, the evanescent field is also s-polarized. For both the polarizations, the wavefront of the evanescent field travels parallel to interface and is slightly phase shifted. These phase shifts depend on the angle of incidence and give rise to the Goos-Hanchen shift [McGuirk and Carniglia, 1977] - a phenomenon where linearly polarized totally internally reflected light undergoes a small lateral shift, perpendicular to the direction of propagation, i.e. in the plane of incidence. The Goos-Hanchen shift distances range from a fraction of wavelength at $\theta_i = 90^o$ to infinite at $\theta_i = \theta_C$[Axelrod et al., 1984].

3.1.2 Excitation Configurations

In general, there are two alignments employed for generating evanescent wave on an inverted microscope. The traditional scheme of Prism-type (PT) TIRF involves the laser beam being internally reflected by a prism (fused silica) on top of the sample chamber [Thompson and Axelrod, 1983]. The alternative arrangement of Objective-type (OT) TIRF involves using a high numerical aperture (NA) objective to shift the excitation laser off axis at an incident angle above that of the critical angle. In this configuration, the laser is focused on the back aperture of the objective such that the outcoming beam is collimated.

PT-TIRF has the advantage of higher SNR due to less background, as the incident beam is reflected away from objective. This makes it more suitable for in-vitro single-molecule measurements. Moreover, the alignment is relatively simpler and reproducible. Additionally, when a laser beam with Gaussian profile is internally reflected through a prism, the resultant evanescent wave is characteristically similar to it (unless incident beam is

tightly focused) [Burghardt and Thompson, 1984] making quantitative data more accurately interpretable [Thompson and Pero, 2005]. The primary disadvantage of this technique is that it requires a prism, thus the sample chamber is spatially unavailable from top.

OT-TIRF, on the other hand, due to its geometry, has the disadvantage of objective collecting the incident and scattered light along with the fluorescence, resulting in higher background. But this can be countered to some extent by using a notch filter to remove the excitation beam from detection path. Nevertheless, with technical advancements in optical lenses, there are specialized higher NA TIRF objectives (eg. Apo TIRF 60x, Nikon and NA 1.49 100x oil objective, Nikon) in market which can collect more photons and give better image quality due to proper index matching of objective with coverslip [Thompson and Pero, 2005]. And since no prism is involved, OT-TIRF is preferred for live-cell imaging and there are many commercial OT-TIRF systems available especially for in-vivo studies.

3.1.3 PT-TIRF Experimental Setup

For the single-molecule studies involved in this thesis, I used a PT-TIRF setup, which was initially setup by Peter Schluesche [Schluesche, 2007] and later modified by Gregor Heiss [Heiss, 2012]. During this thesis, I built a new detection path to allow for multi-color FRET experiments. The TIRF setup is based on an inverted microscope (TE $2000-U$, Nikon). As excitation sources, we have 4 continuous wave (cw) diode-pumped solid state (DPSS) lasers: 491nm 75mW (Cobolt Calypso), 532nm 100mW (Cobolt Samba), 561nm 75mW (Cobolt Jive) and 647nm 120mW (Cobolt MLD). The laser lines are combined through dichroic mirrors (as shown in schematic in figure and controlled via an acousto-optical tunable filter (AOTF: AOTFnC.$400-650-PV-TN$, Pegasus Optik) before being coupled into a polarization maintaining single-mode fiber (532/647 nm, OZ Optics) aided by the apochromatic laser coupler ($60SMS-1-4-RGBV11-47$, Schaefter und Kirchhoff).

At the output of the fiber, the laser beam is decoupled and collimated by the apochromatic collimator ($60FC-4-RGB11-47$, Schaefter und Kirchhoff). There is an additional IR laser (730nm, 40mW, iBeam Smart, Toptica) which is independently coupled (does not go through AOTF) into a polarization maintaining single mode fibre ($PMC-630$, Schaefter und Kirchhoff) through a coupler ($60SMS-1-4-M8-10$, Schaefter und Kirchhoff), and is collimated on the other side of the fiber by a collimator ($60FC-4-M8-10$, Schaefter und Kirchhoff) and combined along with the other laser beams with the help of mirrors. The combined laser beam

is focused[2] onto the prism through an achromatic lens (f = 200 mm). The focusing lens is adjusted to limit the excitation area (and thus the bleaching of fluorophores) to the observation area.

The fluorescence is collected by a water immersion objective (Apo 60x NA 1.27 WI objective, Nikon) and the tube lens focuses the fluorescence out of the side port, where it is spectrally separated by dichroic mirrors and focused by achromatic lenses onto EMCCD cameras. The details of detection part of setup and its evolution are discussed in section 3.2. The AOTF helps control the selection of laser lines, their intensity and duration. The AOTF itself is driven by a radio frequency (RF) modulator, which in turn is controlled by pulses from a field-programmable gated array (FPGA). The FPGA also manages the synchronization of the camera acquisition, the program for which was written by Philipp Messer in Labview.

Prism

Commonly, PT-TIRF setups involve a prism placed on top of a flow chamber formed by sandwiching a spacer (cut out from Nescofilm) between two coverslips. This method requires using an index matching fluid between the prism and flow chamber, which can introduce reflections due to additional interfaces in excitation path, as well as background fluorescence due to impurities in the index matching fluid. To circumvent this, we have custom made (from Zell Quarzglas und Technische Keramik Technologie) prisms with holes drilled into them, so that the flow chamber can be formed directly on prism surface (bottom part of prism). These prisms are made of Suprasil 2, a synthetic fused silica manufactured by flame hydrolysis of $SiCl_4$. It is seen in the schematic of PT-TIRF in figure 3.2 that the prism angle (72^o) is chosen such that when the incident laser beam is normal to incident surface, it will pass through unrefracted and then get internally reflected at prism surface (bottom surface where molecules are immobilized). The angle of the incident beam is adjusted to hit the angular prism surface perpendicularly (i.e. along the normal) so as to avoid refraction which causes the excitation areas (on the sample immobilized surface of prism) for different lasers to be shifted. Also, the prism angles are asymmetrical, to avoid back reflection of the excitation beam.

The flow chambers are formed by sandwiching a nesco film cut out (using a laser cutter) in the design of channels between the drilled holes (template shown in figure), between the quartz prism and the coverslip. This geometry allows preparation of two independent channels that can be used for separate experiments, provided the sealing worked properly. This is helpful since

[2]The focusing is essential to increase laser energy density in the excitation spot and minimize photobleaching outside of the field at view.

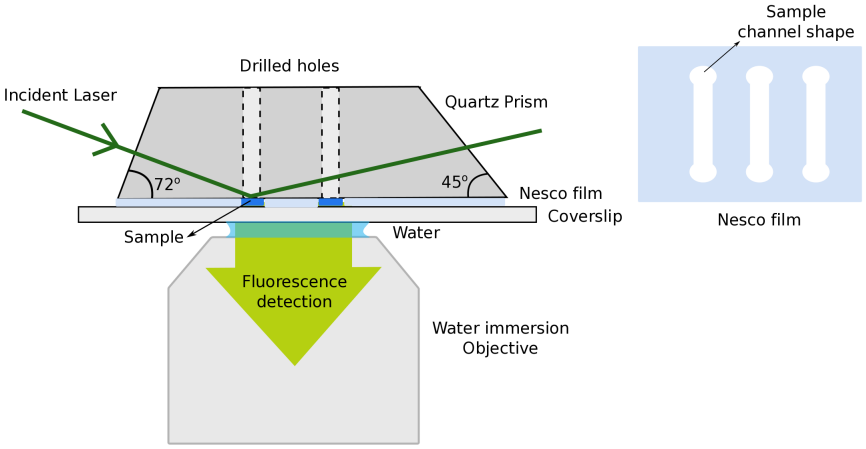

Figure 3.2: Schematic of the prism with the sample chamber and the excitation pathway for PT-TIRF experiments. The excitation beam is focused normal to the incident surface and is totally internally reflected at the prism surface. The sample chambers are formed by sandwiching the nesco film, cut out in the shape of the sample channels, between the holes drilled into the prism and the coverslip.

prisms need to be extensively cleaned and surface functionalized everytime before use, which is a time consuming process, as described in the next section.

Prism preparation for Surface Immobilization

The method for surface preparation of the prisms used in this work has been described in detail earlier[Schluesche, 2007]. Briefly, the prisms are first rinsed with ethanol and then with acetone to remove any organic impurities. This is followed by boiling them in 2% Hellmanex solution and rinsing them extensively with millipore water, followed by boiling in water and further rinsing. The silane layer from the previous preparation cycle is removed by treating the prisms to a 1 : 2 mixture of Hydrochloric acid (HCl, 37%) and Hydrogen peroxide (H_2O_2, 30%) for 45 min., under the chemical hood, till the chlorine gas (product) is dissipated. Then the prisms are washed again, and dried with a stream of nitrogen (N_2).

In order to prepare a fresh silane layer, the prisms are then put in a 2% aminosilane solution in acetone, for 30 min. This is again followed with extensive washing with water and drying with N_2. Next, to passivate the surface, the silanated prism surface is covered with poly-ethylene glycol (PEG) where a small percentage ($3-4\%$) of PEG has been modified with a biotin (biotinylated PEG or bPEG). This is accomplished by incubating the prisms with PEG-bPEG solution ($1\ mg/ml$) in Carbonate buffer ($NaHCO_3$, pH 9.0) for 30 min. Subsequently, the prisms are yet again washed and dried. Finally, the sample chambers on the prism surface are assembled

by sandwiching the nesco film spacer cut-out between a prism and a cleaned coverslip (washed with Ethanol, Acetone, Hellmanex and Water and N_2 dried) and pressed down on a hot plate at 110^oC. These prepared prisms are wrapped in parafilm and stored in falcon tubes at -20^oC until use.

As discussed earlier, the excitation volume in TIRFM is restricted to a thin layer close to surface of prism (∼150-200 nm). Though this reduces the background, it makes the immobilization of the molecules on the surface important. There are several procedures available for single-molecule surface immobilization, the standard one being a Biotin-Streptavidin-Biotin linkage. This basically involves coating the surface of prism with molecules that are modified to have biotin, like, biotinylated bovine serum albumin (BSA) or biotinylated PEG. This is followed by flushing with streptavidin or neutravidin, which bind to biotin with a K_d on the order of 10^{-14} M (very high affinity). Lastly, one adds the molecule of interest, also modified on one end with biotin, thus tethering the molecule to the surface.

Protein Encapsulation in Vesicles

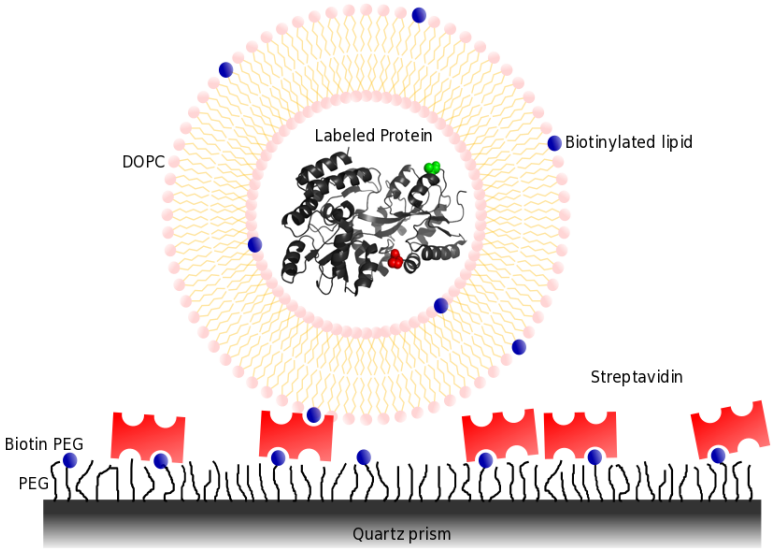

Figure 3.3: Vesicle encapsulation of labeled protein for immobilization on prism surface for single-molecule TIRF experiments. The vesicles are immobilized using biotin-streptavidin-biotin linkage between biotinylated lipids in vesicle and biotinylated PEG on the prism surface. Image not to scale.

Biotinylated proteins can be bound directly to the surface by using streptavidin, as described in the previous section. Nevertheless, this is not an efficient method as introducing a biotin label onto a protein may perturb its behavior. Moreover, unwanted effects could arise due to interaction of the protein with the surface. In addition, the surface may sterically hinder

the conformational changes of the protein. An alternative is to encapsulate proteins into liposomes with a fraction of biotinylated lipids and immobilize the vesicles on the prism surface [Boukobza et al., 2001]. In this way, the protein of interest can freely diffuse inside the vesicle, yet stay close enough to surface to be excited by the exponentially decaying evanescent illumination of TIRFM. Since even large proteins diffuse fairly fast inside the small volume of vesicles (∼200 nm diameter), any intensity fluctuations due to the exponential decay of the excitation profile is averaged out.

A lipid film mixture was prepared by dissolving powdered lipids in chloroform and mixing them in a test-tube to a final concentration of 300 μg 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC, Avanti Polar Lipids Inc., Alabaster, USA) and 6 μg of 1,2-dipalmitodyl-sn-glycero-3-phosphoethanolamine-N-(cap-biotinyl) (sodium salt) (Biotinylated PE, Avanti Polar Lipids) (∼30 μl of lipid mixture per test-tube). These lipid films were then evaporated under stream of N_2 and residual chloroform was removed by placing the test-tubes in a vacuum for an hour.

For preparing the vesicles, the protein sample was incubated with the lipid film test-tube for 30 min. at 4^oC, with occasional vortexing. After the first vortexing, the sample in test-tube appears milky due to formation of multilamellar vesicles (MLVs). These MLVs are then extruded 33 times through a polycarbonate membrane containing 0.2 μm pores, using an Avanti mini extruder (Avanti Polar Lipids). This procedure resulted in large unilamellar vesicles (LUVs) with a homogenous size distribution of 200 nm.

Generally, the proteins are Poissonian distributed within the vesicles. The sample concentration should be adjusted to maximize the number of vesicles encapsulating only a single protein. Post-experiment, one can always quantify the number of proteins per vesicle by counting the number of photo-bleaching steps per trace (assuming that every protein is labeled with exactly one dye molecule).

3.2 Multi-color Detection

The fluorescence from the sample is collected by a water immersion objective (Apo 60x NA 1.27 WI objective, Nikon) and the tube lens focuses the fluorescence out of the side port. We had two detection setups earlier, one at each side port, which was later combined into a single detection path. At each port, there is a rectangular aperture, which is used to adjust the size of detection channel on the EMCCD camera. On the right side, there was a two-color detection setup which is described in detail in [Heiss, 2012]. On

the left side, I built a multi-color detection path, which has evolved over time, as additional EMCCD cameras became available.

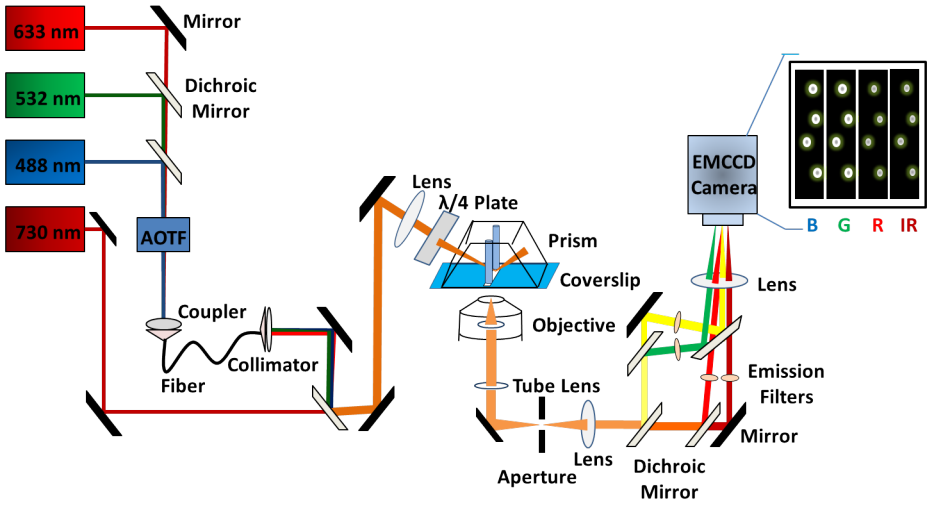

Figure 3.4: Schematic of the first four-color excitation prism-TIRF setup with a single EMCCD camera for four-color detection.

In figure 3.4 a schematic of the first multi-color detection path I built is shown. The aim was to expand the setup to simultaneously detect blue, green, red and IR dyes. As seen in the schematic, the tube lens focuses the fluorescence onto the rectangular aperture placed outside of the side port, exactly at the focal plane of tube lens (approx. 29 mm from the end of adapter tube of the left side port when tube lens of 200 mm is at 1x intermediate magnification).

An achromatic lens ($f = 100\ mm$) placed at its focal distance from the aperture then re-collimates the light which is subsequently split by the dichroic (BS 630 DCXR, AHF), with the blue and green emission being reflected, and red and IR emission being transmitted through. The blue and green emissions were then further split by a dichroic mirror (BS 530 DCXR), and cleaned by emission filters HQ 512/18 for blue detection and HQ 582/75 for green detection before finally being reflected by another 630 DCXR dichroic mirror onto an achromatic lens ($f = 200\ mm$, 500−800 nm) that focuses the detection channels onto the EMCCD camera. The red and IR emissions are split by a dichroic mirror (BS 730 DCXR) and then cleaned by emission filters HQ 685/40 for red detection and HQ 785/62, before they pass through the same BS 630 DCXR dichroic onto the same achromatic lens to be imaged onto the camera. The EMCCD camera was split into 4 channels, one for each detection color.

This setup had the disadvantage that each channel was quite narrow, only 512×128 pixels. Moreover, the molecules at the edges of channel are mostly distorted due to optical aberrations. The choice of dichroics and filters made the blue detection channel very narrow, thus most of the fluorescence from blue fluorophores (like Atto488) would be cut off and the signal was very dim. To solve these problems, a second camera was introduced, and the spectral range of blue detection channel was broadened and green detection was shifted more into spectral range of yellow. For this purpose, a 561 nm laser was introduced into the setup. Furthermore, it was decided to focus on just three colors, leaving out the IR for the moment. Thus the multi-color detection catered to blue, yellow and red detection, and the two-color detection was still available on the right side. In this way, I modified the setup for three-color two-camera detection, as illustrated in figure 3.5.

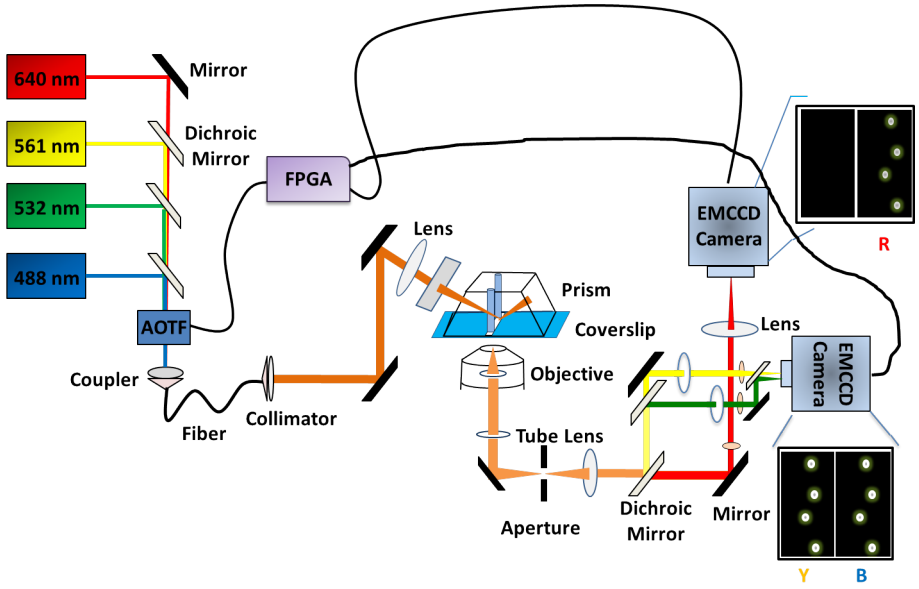

Figure 3.5: Schematic of prism-TIRF setup with two EMCCD cameras for three-color detection.

In three-color two camera detection, the fluorescence for blue and yellow detection is reflected by the dichroic mirror BS 630 DCXR, while the red fluorescence was transmitted. The red fluorescence was cleaned by the emission filter HQ 685/40 and focused onto the red camera by the achromatic lens ($f = 200$ mm). Initially, the blue and yellow fluorescence were separated by the dichroic mirror BS 560 DCXR, then filtered by HQ 525/50 for blue and HQ 595/50 for yellow before being focused by the same apochromatic lens ($f = 200$ mm) onto two halves of the blue-yellow camera. Thus each channel would have 512×256 pixels. The problem with using the same lens for blue and yellow detection channel was that it was very difficult to adjust the different optical components to have both the channels simultaneously in

focus on the camera. Usually, one of the channels would be a bit out-of-focus.

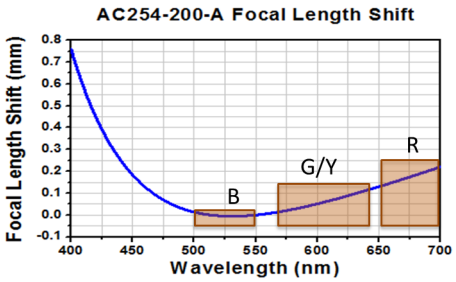

Figure 3.6: Graph of focal length shift in achromatic doublet lenses (AC254-200-A) for light of different wavelengths. Parts of spectra representing the blue, yellow and red fluorescence detected for our setup are highlighted. The underlying graph has been obtained from the website of Thorlabs GmbH, Dachau, Germany

The root of this problem was traced back to the focal length shift of the achromatic lens. As shown in figure 3.6, the focal length shift for the blue detection is negligible compared to yellow and red, thus making it difficult to focus both the yellow and blue detection channels by the same lens.

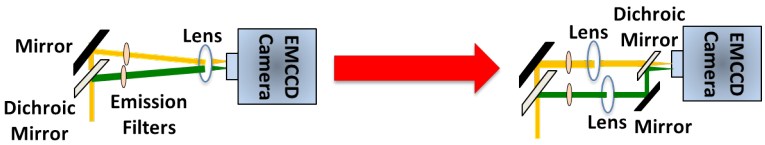

Figure 3.7: Changing the detection path of the blue and yellow fluorescence to have separate lenses for focusing them onto the same camera.

To circumvent this problem, I separated the paths of blue and yellow detection by introducing separate focusing lens for each, and 'combining' them through another 560 DCXR dichroic mirror before impinging on the camera chip. This change in paths is illustrated in figure 3.7.
Recently, a third camera was introduced into the setup and the the two-color and three-color detection paths were combined by Baessem Salem. This latest three-color three-camera detection is illustrated in schematic in figure 3.8. Here again, the red fluorescence is separated from blue and yellow fluorescence by the dichroic mirror 630 DCXR, cleaned by filter HQ 715/150 and focused by achromatic lens (f = 200 mm) onto the red camera. The blue and yellow fluorescence are next split by dichroic 560 DCXR. The blue fluorescence is cleaned by filter HQ 525/50 and focused onto the blue camera by an achromatic lens (f = 200 mm). The yellow fluorescence is cleaned by a HQ 595/50 emission filter and focused onto the green/yellow camera by an achromatic lens (f = 200 mm). For an experiment involving

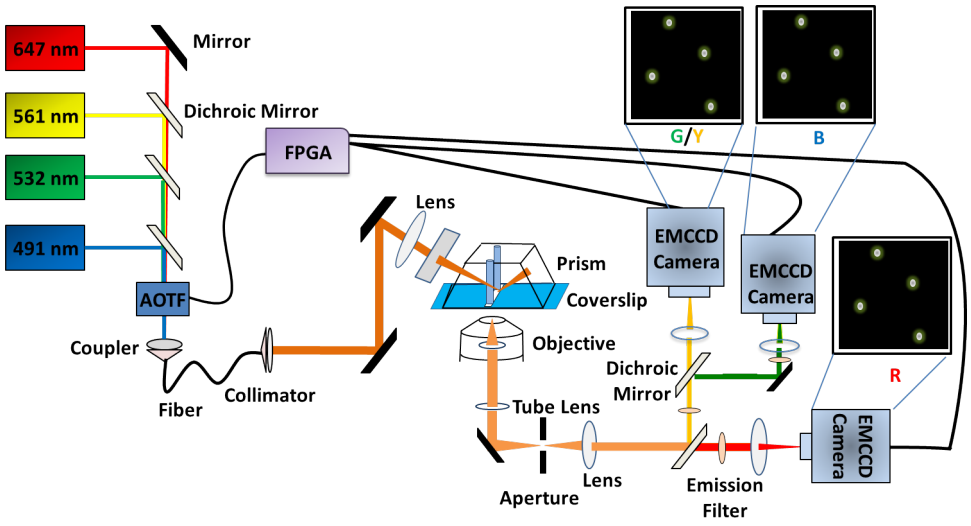

Figure 3.8: Schematic of four-color excitation prism-TIRF setup with three EMCCD cameras for three-color detection.

green detection, the dichroic 560 DCXR would be replaced by 530 DCXR and the filter HQ 595/50 by HQ 585/80 to detect the green fluorescence on the green/yellow camera. This setup has the advantage of having the entire 512×512 pixel area for detection of each color, thus more molecules can be observed at a time. Moreover, each color has its own focusing lens which can be adjusted individually to get a sharply focused image on each camera. Also, as shown in the figures 3.5 and 3.8, the cameras and the AOTF are connected to the FPGA for simultaneous triggering and acquisition.

The spectra of the various dyes measured on the setups, along with the transmission spectra of the various dichroics and emission filters to prove their suitability in detecting these dyes are shown in the figures 3.9, 3.10 and 3.11. The dichroics and emission filters were chosen in a manner to minimize the crosstalk between the channels, yet maximizing the detection efficiency and filtering out the Raman signal from water.

3.3 Electron Multiplying Charge Coupled Device - EMCCD

Detection of single molecules requires high sensitivity of the camera as the fluorescence signal is very weak. While the regular *Charge Coupled Device (CCD)* cameras are limited by the readout noise ($\sim$10 e^-) of the CCD chip, the EMCCD cameras have the advantage of an additional solid state *Electron Multiplying (EM)* register that amplifies the charge per pixel, raising it above the level of readout noise, thus enabling detection of even

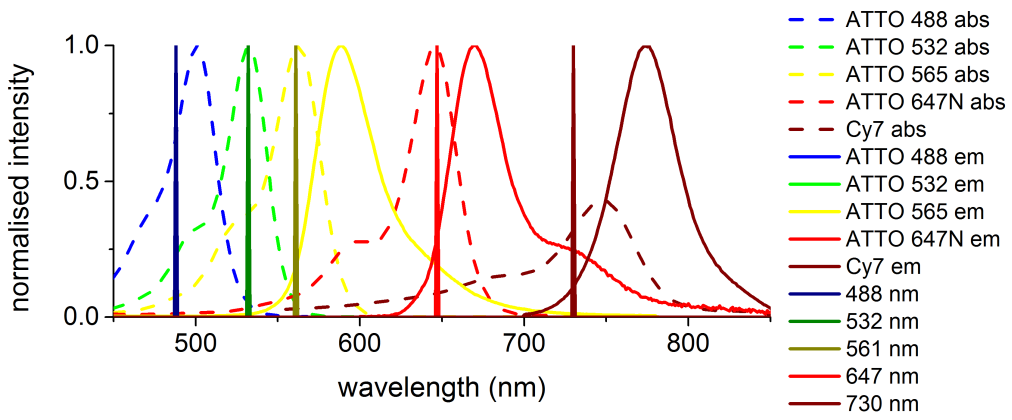

Figure 3.9: Normalized absorption and emission spectra of various dyes used along-with the laser lines used in the setup.

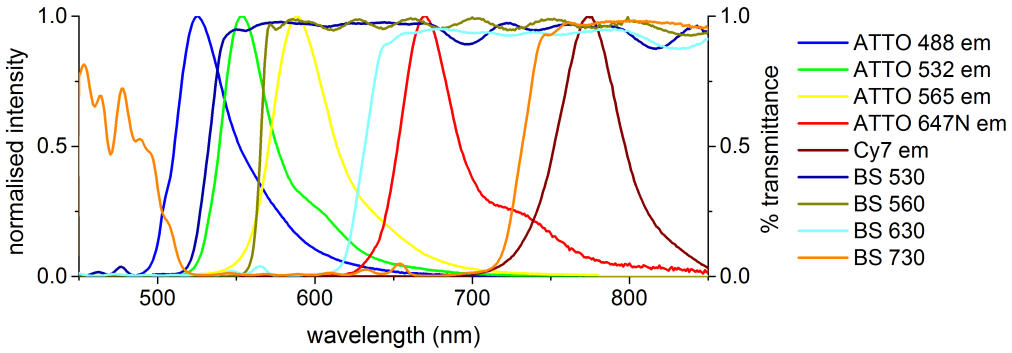

Figure 3.10: Normalized emission spectra of the dyes used superimposed with the transmission spectra of the various dichroics in the setups.

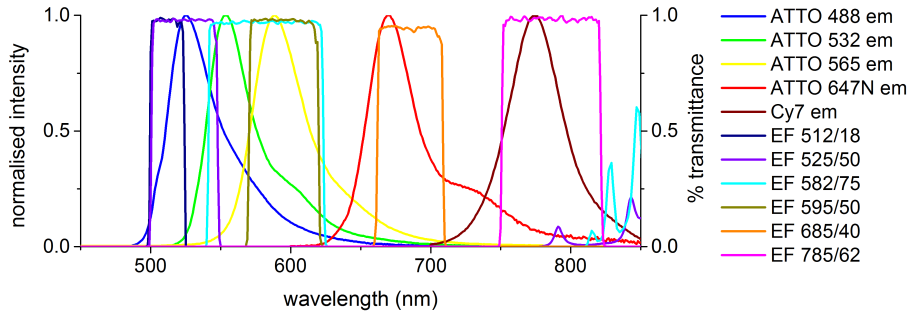

Figure 3.11: Normalized emission spectra of dyes used, superimposed with the transmission spectra of the various emission filters in the setups.

signals with few photons per pixel.

The videos in this thesis were acquired using *frame transfer* mode, which provides faster acquisition rates by using a two-part sensor: one for capturing the image and other for storing the image prior to readout. Both parts are identical in size and the storage area is covered by an opaque aluminium mask. During acquisition, an image is captured by exposing the light

sensitive part to fluorescence from the sample. This image is then shifted vertically downwards into the storage portion of the chip and then read out, while the mask acts like an electronic shutter. Simultaneously, the light-sensitive part is exposed to acquire the next frame. Thus acquisition is limited only by the vertical shift speeds, which are quite fast (on the order of ms).

Readout of the sensor involves shifting the charge through a readout register. The multiplication register amplifies the charge by first clocking it with higher voltage, followed by impact ionization. The probability of charge multiplication is temperature dependent, with lower temperatures contributing to higher gains. It also increases with the voltage applied to multiplication register. Assuming that each shift in the multiplication register applies a minor amplification gain g, the gain G after N shifts would be

$$G = g^N \tag{3.6}$$

After the multiplication register, the charge is shifted through a pre-amplifier, followed by on chip charge to voltage converter. Thus, the final output of the EMCCD is given as greylevels or Analog Digital Units (ADU).

Sensitivity of EMCCD

Sensitivity of any detector is determined by the system noise floor (the instrument detection limit) and quantum efficiency (ability to detect photons). The idea behind using EMCCDs for single-molecule experiments is to increase the signal-to-noise ratio (SNR) for the given experimental conditions. EMCCDs, the most sensitive detectors, help optimize the SNR by eliminating readout noise and dark current noise contributions to the detection limit.

3.3.1 Noise Sources in EMCCD

The noise sources in EMCCD are as follows:

1. Readout Noise ($\sigma_{readout}$)
   Readout noise is the charge generated during each readout process by the chip. Technically, this is the true detection limit of a CCD, contributing to the system noise floor. In CCDs, readout noise can be significant for short exposure times or fast pixel readout rates, but with the EM gain in an EMCCD, readout noise is sufficiently reduced to <1 electron/pixel readout and is independent of exposure time. Since the readout noise is caused by the analog electronics and the AD conversion, it is only amplified by the analog pre amplification gain A, and not the EM gain.

2. Shot Noise (σ_{SN})
   Owing to the particle nature of light, the photons are absorbed as distinct quanta. The fundamental trait of light is that arrival of photons is stochastic, and the probability of arrival is governed by a Poisson distribution. Shot noise is prominent when collecting a relatively small number of photons and reduces by collecting more photons (e.g. by increasing the exposure time). Shot noise depends on the number of photons N detected, and is defined as:

$$\sigma_{SN} = \sqrt{N} \tag{3.7}$$

3. Dark Counts
   Dark counts are randomly generated (shot-noise nature) charges in the absence of light, caused by thermionic emission of electrons. The typical value for $\sigma_{DC} = 10^{-3}$ $electron/pixel/sec$. The dark currents can be almost eliminated by cooling the chip to -80 ^oC.

4. Clock-Induced Charge (CIC)
   CIC, also called as 'Spurious Noise', are single electron events generated usually during charge shift, with the number of shifts required depending on the position of charge on the chip. As in the case of the readout noise, CIC is independent of exposure time. In CCDs, CIC is a part of readout noise. In the EMCCD, however, CIC is amplified by the EM register, unlike the readout noise. CIC can be minimized by conducting fast vertical shifts (μs). Typical values for $\sigma_{CIC} = 1$ $electron/line$.

5. Multiplicative Noise
   As a result of the probabilistic nature of impact ionization, the amplification factor in EM register varies. Based on this, the excess noise factor F can be calculated as a function of the number of shift steps N and the EM gain G [Robbins and Hadwen, 2003]. This noise factor is alternatively known as *multiplicative noise* and is defined by:

$$F^2 = \frac{\sigma_{out}^2}{G^2 \cdot \sigma_{in}^2} \tag{3.8}$$

   which can be refined to

$$F^2 = 2(G-1)G^{-((N+1)/N)} + \frac{1}{G} \tag{3.9}$$

   For the usual settings for our single-molecule experiments, $G = 300, N = 512$ and $F^2 = 1.98$

Hence, taking all noise factors into consideration, the total noise σ_{total} is given by:

$$\sigma_{total} = \sqrt{A^2\sigma_{readout}^2 + F^2G^2A^2(\sigma_{dark}^2 + \sigma_{signal}^2)} \tag{3.10}$$

where σ_{signal} is the noise of optically generated signal S and σ_{dark} is the noise of the dark signal S_{dark}.

The mean output signal is given by:

$$S_{out} = AG(S + S_{dark}) = AG(\sigma^2_{signal} + \sigma^2_{dark}) \tag{3.11}$$

Therefore, the SNR is given by:

$$SNR = \frac{S + S_{dark}}{\sqrt{\frac{\sigma^2_{readout}}{G^2} + F^2(\sigma^2_{signal} + \sigma^2_{dark})}} \tag{3.12}$$

4. Data Analysis

The prism-type TIRF setup described in the previous chapter was used to record movies of fluorescing single-molecules immobilized on the prism surface. The size of the fluorophores is much smaller than the point spread function of the setup, resulting in diffraction limited spots being detected on the EMCCD cameras. The EMCCD records the intensities of single-molecules from the area imaged onto it, in the form of Tagged Image File Format (TIFF) stacks.

For multi-color FRET experiments, the single-molecule intensities were extracted from these TIFF movies using the *Multi-Intensity Trace Analysis* (MITrA) program, that I have written. Further analysis of these intensity traces was performed with the *Trace Intensity Analysis* (TRACY) program, written by Gregor Heiss. Both MITrA and TRACY are based on functions and toolboxes available in MATLAB (Mathworks Inc.). The steps involved in trace-extraction and the analysis procedures used for determining quantitative FRET results and kinetic rates in dynamic systems are described in detail here.

4.1 Mapping

EMCCD cameras have the highest detection efficiencies available for detecting single-molecule fluorescence. However, EMCCDs have an inherent drawback that they operate in intensity mode, hence cannot distinguish between different wavelengths of light. This requires that the fluorescence from the molecules be spectrally separated to form spectra-specific detection channels, which are then imaged onto different parts of a single camera or onto different cameras. As seen from the setup schematics in the previous chapter, the fluorescence is separated into different channels by using the appropriate dichroic mirrors and detected on different cameras or on the different regions of the same camera.

As a result, the first step in post processing involves determining the position of the same molecule in all the detection channels. One could overlay the channels on top on each other. However, due to achromatic and optical aberrations introduced by the various optical elements in the optical path of each detection channel, the channels are spatially distorted with respect to each other. Moreover, the distortions cannot be solved by geometrical transformation alone, but usually require a combination of translational, rotational, magnification, and sometimes even shear transformations.

The process of generating an image transformation between two channels is termed as *Mapping*. Incase of more than two channels, it is sufficient to map all channels onto a single 'master' channel. Mapping is only used to identify the same molecule, but all further analysis is done with the raw data.

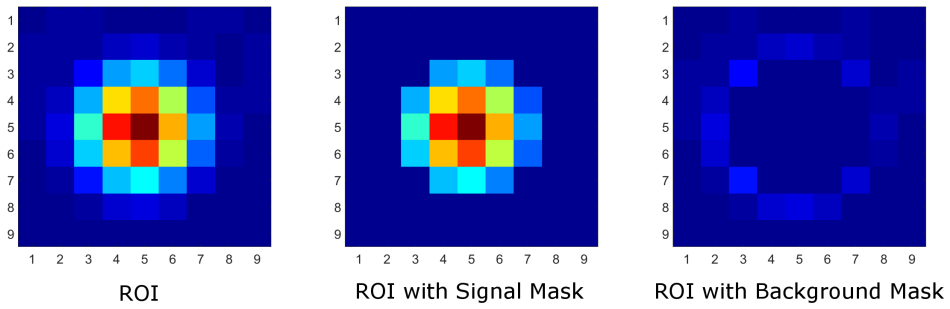

Figure 4.1: Scaled intensity image of a single zero-mode waveguide (ZMW) image with a region of interest (ROI) around the central brightest pixel. Left: ROI of 9 is selected around the brightest pixel (at center) Middle: ROI with signal mask applied. These pixels are averaged to obtain mean signal. Right: ROI with background mask applied. These pixels are averaged to obtain the mean background.

Prior to generating a map, it is required to detect the location of fluorescent spots (assumed to represent a single-molecule) in every channel. Analysis of the movies begins by loading the movies acquired simultaneously in the form of TIFF stacks into the memory. MATLAB can read the TIFF stack images into a three-dimensional matrix, where each image frame is a two-dimensional matrix of the intensity count values of the pixels and the third dimension represents the number of images in the stack.

To detect the spots in a movie, a sum intensity image of all the frames is used. The background of this intensity image is set to zero by applying a threshold. This threshold is calculated from the intensity histogram of the image such that only the brightest 5% of the pixels are displayed. Within this thresholded image, the brightest pixel (pixel with maximum intensity) is located. A region of interest (ROI) is selected around this pixel (as shown in figure 4.1). This ROI is multiplied with a circular signal mask and the pixel values are averaged to obtain the mean signal. Similarly, the ROI is multiplied with a doughnut shaped background mask and averaged to obtain the mean local background around the spot. The ratio of the mean signal to mean local background is used to judge if the spot is of diffraction limited size and has sufficient signal. If the signal originates from a source larger than the diffraction limit (for e.g. two single-molecules tethered too close to each other), the signal will bleed into the background pixels, therefore decreasing the ratio and leading to the spot getting discarded. Also, the distribution of the pixel values after multiplying with the signal mask is determined and compared with a Gaussian distribution of equal size

to determine if the spot is sufficiently Gaussian or not. Only the spots that fulfill these criteria are used for further analysis. Once an ROI fulfills all the requirements, it is removed from the threshold image and the process is repeated with the next brightest pixel, until no more are left. This way, positions of all spots are determined for each detection channel.

After localizing the spots in each channel, the group of spots corresponding to the same molecule need to be determined. As a starting approximation, the linear shift of the channels with respect to 'master' channel is determined by stepwise linear shifting of the coordinates of the spots found in them. The distance between the location of spots in the 'master' channel and the shifted spots in other channels is calculated. The number of pairs with distance below a certain tolerance is determined. The tolerance must be chosen appropriately to allow matching of pairs despite non-linear aberrations, yet avoiding pairing of spots that do not belong to the same molecule. The optimal linear shift is reached when the maximum number of matching pairs has been found. The location of spots at the optimal shift position are used as "control" points for the generation of the transformation map. It must be mentioned here that only the points where particles are found in both channels are considered for further analysis.

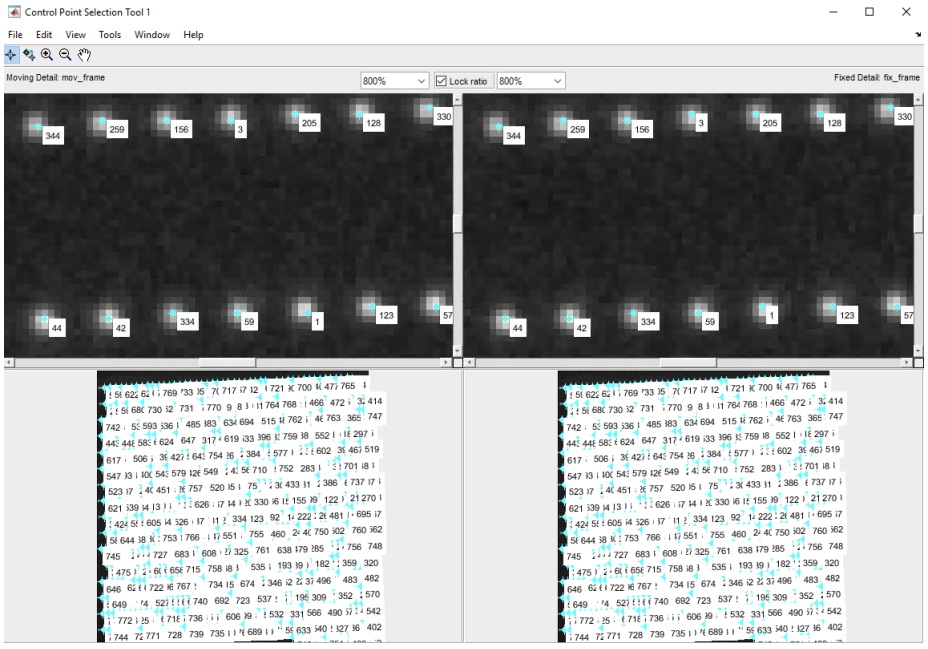

Figure 4.2: The control point selection tool showing the pre-determined "control" points in the two channels. The lower part shows all the matching molecules identified and numbered by the pairs. The upper part shows a zoomed in portion of the lower part. The tool allows selecting new points as well as changing the positions of the existing points.

The "control" points can also be determined manually by using the control point selection tool (by *cpselect* function in MATLAB) to open the intensity

frame of 'master' channel and another channel and then manually selecting the control point pairs of the spots meant to represent the same molecule. This is more tedious than using the automatic identification of spot pairs algorithm used earlier. There is, of course, a possibility to apply both together, opening the control point selection with pre-determined "control" points and then adding more spots that might not have been matched before, or even adjusting the position of the pairs detected if required.

The locations of the spots in the 'master' channel are used as base points and the locations of the corresponding spots in the other channels are used as input points. The image transformation is generated by mapping the input points onto base points. The image transformation was implemented using the built-in *cp2tform* and *imtransform* functions in MATLAB versions 2014a and earlier. The versions of MATLAB from 2014b onwards used the functions *fitgeotrans* and *transformpointsinverse* respectively. The fitgeotrans function can infer a spatial transformation between a set of vectors. The position (coordinates) of each pixel in an image can be considered as a vector. The fitgeotrans function was used to infer a spatial transformation between the pixels of two images based on the input and base points. A third order polynomial was used as the transformation function, which requires a minimum of 10 mapping points to work. Thus each of the other channels was mapped onto the "master" channel.

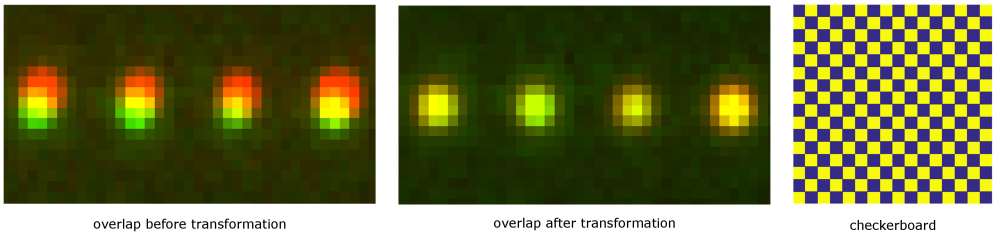

Figure 4.3: The map transformation is applied onto the images to correct for the distortions. Left: The overlay of the images is bad before applying mapping, showing the ZMWs to be spatially shifted. One image is scaled in green and the other is scaled in red. The overlap would be visible as yellow pixels. Middle: The overlap is better after the transformation map is applied, resulting in all yellow pixels. Right: The checkerboard with the transformation applied shows uniform sized squares and no apparent distortions in the map. The transformation map is applied to the images only to check the performance of the mapping. For intensity extraction, always the raw image data is used.

Alternatively, for mapping, the movies of Zero Mode Waveguides (ZMWs) under the lamp light of the microscope were used. ZMWs have the advantage of having a grid of wells that cover the entire camera chip and the grids are available in different patterns, the square grid with an asymmetric cross leaving blank spaces in the grid was mostly employed, as shown in figure 4.4. Alternatively, a movie from streptavidin coated fluorescent beads like Nile red beads from Kisker Biotech could be recorded and their locations

used. The disadvantage with using beads is that first of all, the quality of the transformation map would depend on the density of the beads in that part of image, which cannot be manually controlled. However, this could be partially overcome by taking multiple movies while translating the beads to cover the entire field of view. The movies could be combined together. The z-projection of the TIFF stack of the combined movie provided sufficient number of spots covering the entire field of view. Secondly, it is difficult to find multi-fluorescent beads that could be detected in all channels. The Nile red beads worked well for green and red channels but were hardly detectable in blue. To overcome this, I made my own multi-fluorescent beads by taking polystyrene beads functionalized with amine groups and labeled them with NHS-esters of the various dyes and also Biotin-NHS. This solution worked well but later we switched completely to ZMWs for the ease of use and because one frame was sufficient as the waveguides covered the entire camera area. When ZMWs are used to generate a map, each

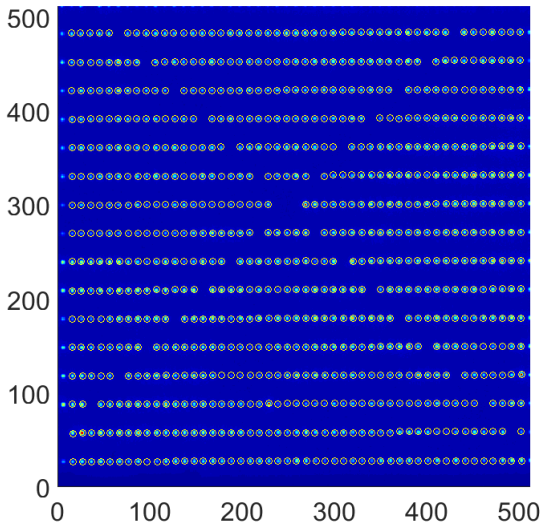

Figure 4.4: The intensity scaled image of a ZMW movie with the ZMWs marked by circles as identified by the automatic molecule detection algorithm.

waveguide as a diffraction limited spot will be visible in all channels (as lamp light is used for excitation) and no mismatch can occur. If instead a movie of molecules undergoing FRET is used for mapping, the locations of non-FRETing molecules would not correlate. Though the number of random mismatches would be smaller compared to the number of correctly matched pairs, the wrongly matched pairs can distort the mapping function locally. The quality of the map is always checked by applying the map to the mapped channel and overlaying the image on the 'master' channel. The transformation is also applied to a checkerboard to visualize any inherent distortions in the map. The transformation map is saved in the data

directory. Prior to extracting traces, the presence of a map is confirmed and if no map is found, only then is a new one generated.

4.2 Intensity Extraction

For extracting the single-molecule intensities in all channels, the molecules were first localized in all channels and then the inverse image transformation (*transformpointsinverse* function in MATLAB) was applied to locations in the 'master' channel to transform them into coordinates of other channels. Fluorescent spots for which the transformed locations matched with the localized locations in the other channel were considered to belong to the same molecule. Thus, for each single-molecule, the original coordinates of the spots in all the channels were determined and written into a table. For multi-color FRET, only the spots that were found in all channels were considered. Another approach could be the transformation of the whole image, but it is not recommended as one should stay with the raw data when calculating intensity.

For extracting the single-molecule intensities, the raw data was used. Using the coordinates of the spots in all channels belonging to same molecules, ROIs were determined around them. The signal and local background intensities were then extracted by multiplying them with the circular signal mask and doughnut-shaped local background mask as performed earlier during spot detection, and the mean signal and mean background was determined for the pixels in every frame. The intensities of the donor and acceptors were saved into a structure in *.mat* file for further analysis.

4.3 Correction Factors

The single-molecule fluorescence intensity traces, obtained as described in the previous section, must be corrected using various correction factors including Background subtraction, relative detection efficiency γ, spectral crosstalk β are described in the upcoming sections. The correction factors explained here are for spFRET.

4.3.1 Background Correction

Background correction is the first correction factor to be applied on the extracted intensities. All other correction factors are applied on background corrected intensities. The local background of a molecule, as calculated during intensity extraction process, can be either directly subtracted or smoothed by a moving average prior subtraction. One should be wary

of direct subtraction of a noisy signal as it adds additional noise to the signal. To minimize this additional noise, the background can be smoothed with a moving average filter (usually 10 frames). Alternatively, a user defined number of frames at the end of each trace can be averaged, subject to the condition that the fluorescent signal has photobleached by then. This single averaged intensity value is then subtracted from the whole signal intensity trace. This process does not add any noise factor to the signal. However, if the background is changing, this will not be accounted for.

4.3.2 Alpha

Ideally, in smFRET experiments, each laser should excite only one fluorophore, which when excited behaves as the donor. However, as shown in the figure 4.8, the absorption spectra of fluorophores have a long tail that extends into shorter wavelengths. This results in the direct excitation of the acceptors by the laser for donor excitation. Therefore, a small amount of signal is detected in the acceptor channel despite there being no FRET. The α factor accounts for direct-excitation of the acceptor dyes.

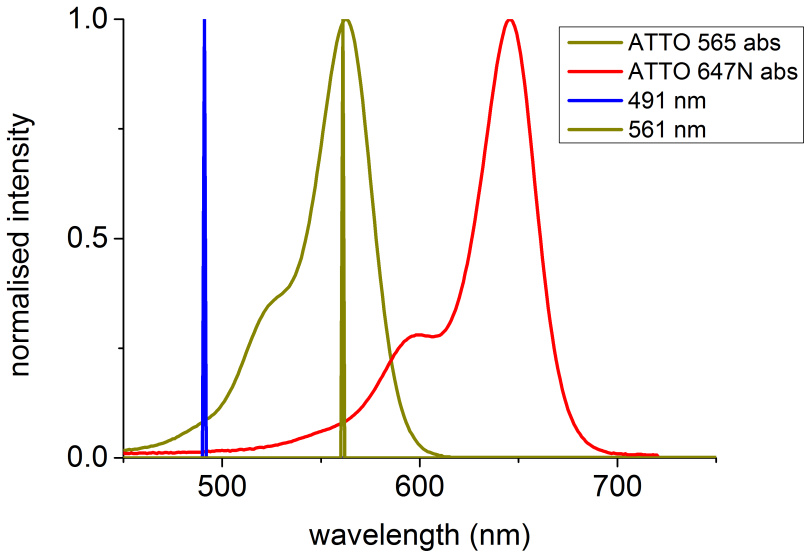

Figure 4.5: The direct excitation of Atto565 and Atto647N by the blue laser (491 nm) and of Atto647N by the yellow laser (561 nm).

The direct excitation depends on the laser power. The α factor can be theoretically calculated from the absorption spectrum of the acceptor as the ratio of intensities at the donor laser to the intensity at acceptor laser. For Atto565, $\alpha = 0.08$ on 491 nm excitation. For Atto647N, $\alpha = 0.01$ on 491 nm excitation and $\alpha = 0.08$ on 561 nm excitation.
Experimentally, the direct-excitation can be obtained by measuring an acceptor-only construct with donor laser excitation. Alternatively, α can be

estimated as the ratio of intensity of acceptor on excitation with donor laser to excitation with acceptor laser, after donor bleaching. When determining experimentally, it is important to remember that the α depends on the laser power, so it must be determined in each measurement.

4.3.3 Beta

The spectral crosstalk is accounted for by the β factor. The β factor accounts for crosstalk from the donor to the acceptor channel and is calculated as the ratio of acceptor to the donor fluorescence detected after acceptor photobleaches and is written as:

$$\beta = |\frac{\langle I_A \rangle}{\langle I_D \rangle}|_{afteracceptorbleaching} \tag{4.1}$$

Without any spectral crosstalk, no fluorescence should be observed in the acceptor channel after acceptor bleaching. However, since the tail of emission spectrum of donor overlaps with the acceptor emission spectrum, some of the donor fluorescence bleeds into the acceptor channel and need to be accounted for calculating the correct FRET efficiency.

4.3.4 Gamma

The γ factor accounts for the relative detection efficiencies (η_A and η_D in acceptor and donor channels respectively and for differences in the quantum yields ϕ_A and ϕ_D of acceptor and donor fluorophores respectively:

$$\gamma = \frac{\eta_A \cdot \phi_A}{\eta_D \cdot \phi_D} \tag{4.2}$$

The factor is different for every molecule and can be determined experimentally by comparing the fluorescent intensities before and after acceptor bleaching as shown in figure as:

$$\gamma = \frac{\langle I_A \rangle_{before} - \langle I_A \rangle_{after}}{\langle I_D \rangle_{after} - \langle I_D \rangle_{before}} = \frac{\Delta I_A}{\Delta I_D} \tag{4.3}$$

On the other hand, if the trace shows dynamics, the γ factor can be calculated by minimizing the variations (standard deviation) in the total intensity ($I_A + \gamma I_D$). This approach can be implemented practically by minimizing the standard deviation of the total intensity as a function of the γ factor as:

$$\frac{d}{d\gamma}(\sigma_{I_A+\gamma I_D}) = 0 \tag{4.4}$$

It is presumed that different states within the dynamic FRET traces have the same γ factor. The dynamic γ factor changes with fluctuations in local

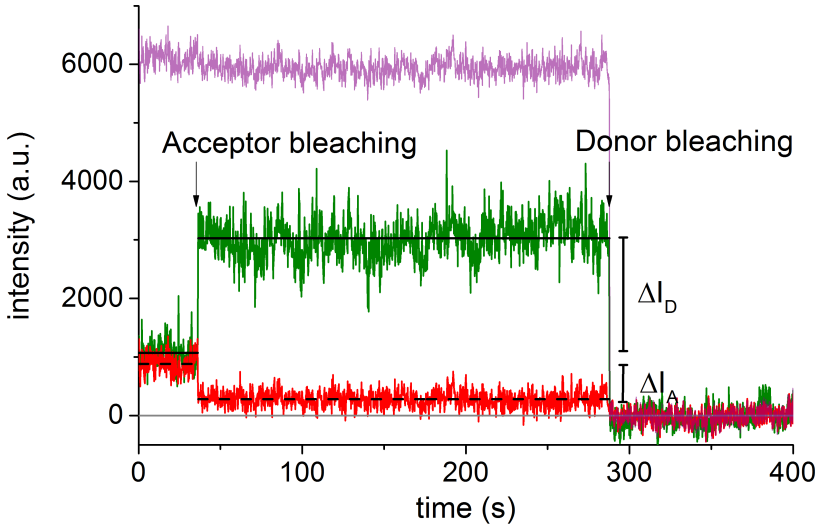

Figure 4.6: The determination of β and γ factors for static FRET traces. The acceptor bleaching and donor bleaching steps are marked. The thick black line shows the mean intensity of the donor before and after acceptor bleaching. The mean intensity of the acceptor before and after acceptor bleaching is shown in dashed black line. After acceptor bleaching, the residual intensity observed in acceptor channel is the crosstalk from the donor channel. After donor bleaching, intensities of both donor and acceptor go to zero. The total intensity (purple) has been increased by a factor of 2 to show it separately from the donor and acceptor intensities

environment of the fluorophore like during protein binding. If the γ factor for the different FRET states differs significantly, the γ-corrected total intensity would correlate with the FRET trace.

Since not all traces have acceptor bleaching, the γ obtained from static or dynamic traces is averaged and applied to molecules where no gamma could be determined. Sometimes the median of gamma distribution is a better estimate of gamma than the mean. The γ factor can be adjusted at any point during or after the analysis without the need to reanalyze the data.

4.3.5 Correction Factors for Three-color FRET

The intensities of the fluorophores for three-color FRET are also corrected for background, direct excitation, cross-talk and detection efficiencies. However, the correction factors have to be determined for each dye and dye-pair and combined together.

Given a three-color triangle FRET (all three dyes interacting with each other) system, there are 3 FRET efficiencies to be determined. For the sake of convenience, let the dyes, detection channels and lasers be represented as Blue, Green and Red. The identities of the symbols used in the calculations are as follows:

M, total intensity detected in channel M after background subtraction.
M', intensity of the dye M detected in channel M.
M", total intensity of dye M detected in all channels.
M^*, gamma corrected intensity of M.
α_{MN}, the direct excitation factor of dye N on excitation with laser M.
β_{MN}, the bleedthrough of dye M into the detection channel N, on excitation with laser M.
γ_{MN}, the gamma factor of dye M with respect to dye N, on excitation with laser M, calculated as $\frac{\Delta I_N}{\Delta I_M}$ during bleaching of dye N.
where M and N are B, G or R.

Assuming no crosstalk of longer wavelength dyes into shorter wavelength detection channels, the intensities of the dyes detected in the respective detection channels can be written as:
$B' = B$
$G' = G - \beta_{BG}B$
$R' = R - \beta_{BR}B' - \beta_{GR}G'$
$= R - \beta_{BR}B - \beta_{GR}(G - \beta_{BG}B)$

Therefore, the total intensity of each dye from all channels obtained by adding back the crosstalk into proper detection channels:
$B" = B' + \beta_{BG}B + \beta_{BR}C = (1 + \beta_{BG} + \beta_{BR})B$

$G" = G' + \beta_{GR}G' = (1 + \beta_{GR})(G - \beta_{BG}B)$

$R" = R' = R - \beta_{GR}G - (\beta_{BR} - \beta_{BG}\beta_{GR})B$

To correct for difference in detection efficiencies, the gamma factors are incorporated as follows:
$B^* = B"$
$G^* = G"/\gamma_{BG}$
$R^* = R"/\gamma_{BR}$

And finally, the intensities were corrected for direct excitation as follows:
$G^{**} = G^* - \alpha_{BG}G^*$

$R^{**} = R^* - \alpha_{BR}R^*$

The individual FRET efficiencies can thus be calculated as:
$E_{BG} = G^{**}/(B" + G^{**} + R^{**})$

$E_{BR} = R^{**}/(B" + G^{**} + R^{**})$

$$E_{GR} = R^{**}/(R^{**} + G^{**})$$

In the case of ALEX experiments, the ALEX data can be used to determine the FRET efficiencies follows. Referring to figure 2.11 in chapter 3, the symbols used there represent:
K_m, the excitation rate of dye m.
k_m, intrinsic fluorescence decay rate of dye m.
k_{mn}, rate of FRET from dye m to dye n.
R_{mn}, distance between dye m and dye n.
I_{mn}, intensity of dye n after m excitation.

Assuming all the intensities have been corrected as given before, the FRET efficiency can be expressed as:

$$E_{mn} = \frac{k_{mn}}{k_m + k_{mn}} = \frac{1}{1 + (R_{mn}/R_{0,mn})^6}$$

From rate equations, fluorescence intensities at green excitation is:

$$I_{gg} = \frac{k_g}{k_g + k_{gr}} K_g$$

$$I_{gr} = \frac{k_{gr}}{k_g + k_{gr}} K_g$$

which gives $E_{gr} = \dfrac{I_{gr}}{I_{gg} + I_{gr}}$

Similarly, at blue excitation, the intensities of the dyes are:

$$I_{bb} = \frac{k_b}{k_b + k_{bg} + k_{br}} K_b$$

$$I_{bg} = \frac{k_{bg}}{k_b + k_{bg} + k_{br}} \frac{k_g}{k_g + k_{gr}} K_b$$

$$I_{br} = \left(\frac{k_{br}}{k_b + k_{bg} + k_{br}} + \frac{k_{gr}}{k_g + k_{gr}} \frac{k_{bg}}{k_b + k_{bg} + k_{br}}\right) K_b$$

Using above equations, we get the FRET efficiencies as:

$$E_{BG} = \frac{I_{bg}}{I_{bb}(1 - E_{gr}) + I_{bg}}$$

$$E_{BR} = \frac{I_{br} - E_{gr}(I_{bg} + I_{br})}{I_{bb} + I_{br} - E_{gr}(I_{bb} + I_{bg} + I_{br})}$$

Knowing the FRET efficiencies, the distances can be calculated too, by using the E-R relation.

4.4 The Hidden Markov Model - HMM

A *Markov Model* is a stochastic model where the transitions between discrete observable states of a system are governed by a *Markov process* - a stochastic process without memory. This implies that the probability of transitioning to a state depends only on the present state, and is independent of the past sequence of states. As a result, knowing the present state of the system, any future time point distribution of states can be calculated. Named after the Russian mathematician Andrey Markov who first studied them in 1906, the Markov Models have been applied to a variety of fields, from modeling population processes [Watterson, 1996] and simulating DNA sequences [Pratas et al., 2011] to generating texts [Hartman, 1996] and for advanced sports analysis [Bukiet et al., 1997].

A *Hidden Markov Model* is a markov model where the underlying state sequence is 'hidden' by the noise in the system. The transition probabilities between the states, though unknown, are still governed by a Markov process. Additionally, there is a second stochastic process involved, that generates a series of *observable* outputs. The distribution of these outputs from each state is given by the *emission function*.

Developed by Baum and Petrie in 1966 [Baum and Petrie, 1966], HMM was originally developed for speech and pattern recognition [Rabiner and Juang, 1986, Rabiner, 1989]. With the advent of DNA sequencing, HMM extended its influence to the field of bioinformatics [Lukashin and Borodovsky, 1998, Pedersen and Hein, 2003]. In biophysics, HMM was first implemented in ion-channel studies [Qin et al., 2000] to model current through ion channels, followed by optical trapping studies [Smith et al., 2001]. In single-molecule studies, HMM was initially tested for robustness in determining kinetic rates from simulated photon trajectories with incorporated noise factors [Andrec et al., 2003]. McKinney et al. [2006] were the first to apply HMM to single-molecule FRET trajectories. They used HMM to determine the states and the transition rates in Holliday junction and RecA binding-dissociation. Since then HMM has proven to be a robust and efficient method for analyzing smFRET data for example in studying rotation of F0F1-ATP synthase [Zarrabi et al., 2007], binding of klenow fragments on dideoxynucleic acid [Munro et al., 2007],and the dynamics of DNA bending on binding TBP and other transcriptional factors [Gietl et al., 2014].

For the work in this thesis, HMM was employed on single-molecule intensity traces (collected through TIRFM) to extract dynamics on the timescale of several seconds. The HMM analysis was performed as recommended by McKinney et al. [2006], utilizing the MATLAB HMM toolbox by Kevin Murphy [Murphy, 1998]

4.4.1 HMM Model Parameters

As explained previously, an HMM model is regulated by two stochastic processes: one for describing the probability of transitioning between the discrete states and other for describing the distribution of observable outputs for each of the given states. In this thesis, the HMM algorithm is specifically being applied to single color intensity traces, though the procedure is same for applying HMM to any discrete time series of states. For a single-molecule fluorescence intensity trace, the intensity at a given time point is the observable output and the objective is to determine the underlying sequence of states that generates this intensity time trace. An intensity trace can be assumed to be a time series $Q = \{\ q_t\ \}$, comprising of a sequence of N discrete states S_i. As a result of memorylessness of a Markovian process, the state q_t at time t depends only on the previous state q_{t-1}.

$$P\left(q_{t+1} = S_i \mid Q\right) = P\left(q_{t+1} = S_i \mid q_t\right) \tag{4.5}$$

The transition probability matrix $P_{ij} = p_{i \to j}$ gives the probability of transitions between two states and is constant over time.

$$P\left(q_{t+1} = S_i \mid q_t\right) = P\left(q_{t+2} = S_i \mid q_t + 1\right) \tag{4.6}$$

The emission function $f_i(x)$ gives the distribution of observable outputs $X = \{x_t\}$ from the states S_i. In essence, the emission function provides us the probability of observing an output, given that the system is in a particular state. In our analysis, the emission function f_i was modeled by a multivariate Gaussian distribution and the intensity values of each state S_i were modeled as a single Gaussian distribution with mean μ_i and variance σ_i, as given by:

$$f(x|\mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{4.7}$$

The prior π embodies the knowledge about the system (the initial state). In smTIRF experiments, no knowledge of the initial fluorescence state is available. Therefore, equal probability is assigned to all states at the beginning of the analysis. This is known as a *Flat Prior*.

To summarize, the HMM model M for single-molecule trace data consists of the following four parameters:

1. The number of states N and definition of each state S_i
   $\Rightarrow$ N Gaussian distributions with mean μ_i and variance σ_i

2. The transition matrix P that incorporates the transition probabilities:

$$p_{ij} = p(q_t = S_i | q_{t+1} = S_j) \tag{4.8}$$

3. The emission function f_i containing the distribution of observable outputs for each state (described by the given Gaussian distribution)

$$f_i(x) = p(x|q_t = S_i) = f(x|\mu_i, \sigma_i) \tag{4.9}$$

4. The prior π which contains the prior knowledge about the system

$$\pi_i = p(q_{t=0} = S_i) \tag{4.10}$$

Thus the states transition through a 'hidden' sequence and generate a sequence of observable outputs through the emission function. In our case, the intensity at each time point in the trace is the observable output. The HMM deciphers the hidden sequence of states and the molecule transitions that most likely generate this intensity trace.

4.4.2 Training of HMM

Given the HMM model M and observations X, the training of HMM entails modifying the parameters of M such that the likelihood that M describes X is maximized. Precisely, parameters of M are trained with X to maximize the probability of model M, given the observations X. When the HMM model M has been optimized, the underlying hidden sequence of states Q can be recovered. Therefore, training of an HMM involves two steps:

1. Calculating the likelihood that the observation X can be described by model M, i.e. $P(X|M)$
   The probability to observe X given a hidden state sequence Q is given by:

$$P(X|Q, M) = p(x_1|q_1) \cdot p(x_2|q_2) \cdots p(x_T|q_T) = \prod_{t=1}^{T} p(x_t|q_t) = \prod_{t=1}^{T} f(q_t) \tag{4.11}$$

   The probability of certain hidden sequence S given the HMM model M is given by:

$$P(Q|M) = p(q_1, q_2, ..., q_T|M) = p(q_{t-1}|q_t) \cdots p(q_{T-1}|q_T) = p(q_1) \prod_{t=2}^{T} p(q_t|q_{t-1}) \tag{4.12}$$

   Thus, the probability of X given the model M can be written as a joint probability of 4.11 and 4.12:

$$P(X|M) = \sum_{Q_i} P(X|Q, M) \cdot P(Q|M) \tag{4.13}$$

A direct evaluation would require $2TN_T$ calculations to compute which is not feasible. A better technique is to use the *Baum-Welch* algorithm [Baum et al., 1970, Welch, 2003] that employs the *Forward-Backward* algorithm [Baum, 1972] to calculate the probabilities at each step independently, by taking advantage of the markovian property of the transition probabilities and the emission probabilities at any time point.

2. Modifying parameters of M to maximize $P(X|M)$
   This is an iterative two-step process involving estimation of the likelihood of the model (E step) by calculating the maximum likelihood estimates of the parameters, followed by modifying the parameters of the model to maximize the expected log-likelihood (M step). The traditional *Expectation-Maximization* (EM) algorithm [Dempster et al., 1977, Wu, 1983, Moon, 1996] is based on using the derivative of the likelihood function to optimize the model parameters. The Estimation and Maximization steps are repeated alternatively until convergence.

Once the parameters of the model are optimized, the last step in HMM analysis is to determine the most-likely sequence of hidden states producing the observations or the Viterbi path. The Viterbi algorithm [Viterbi, 1967] which is based on the Forward algorithm was used to determine the Viterbi path. Briefly, the final probability of the most likely sequence to terminate at a state is calculated and then back tracked to determine the most-likely sequence of states. The Viterbi algorithm also provides the probability of the Viterbi path for a given set of parameters (model). Thus, this information can be used in the optimization step to improve the model.

4.4.3 Application of HMM to Single-molecule Data

In single-molecules studies, HMM has been extensively applied to FRET data to study dynamics of conformational changes in the system, which are characterized by changes in distance between the fluorophores. Generally, the conformational states of a system can be identified by fitting the frame-wise FRET histogram of shot-noise limited FRET distribution with Gaussian distribution. It is possible to separate the different FRET states manually, but not recommended as it is prone to user bias. The better alternative is to employ a thresholding algorithm but only if the separation of FRET states is significantly larger than the inherent noise. This is where HMM saves the day. By means of the emission function, HMM takes the distribution of individual states into account, thus boosting its ability to assign the hidden states and the state sequence accurately.

In this thesis, HMM has been mostly applied to individual fluorescence intensity distributions of the fluorophores. Broadening of the fluorescence intensity distribution occurs due to presence of background fluorescence,

shot-noise and the additional noise factor from EMCCD detection. The fluorescence signal is Poisson distributed and can be approximated by a Gaussian distribution. In the case of FRET, the signal is best approximated by a Beta distribution, though a Gaussian approximation suffices and is extensively used.

Since HMM takes the emission function into account, its parameters would contribute additional degrees of freedom to the model, hence there is a requirement to initialize and train it appropriately. Also, when training, some parameters can be fixed to an initial value, like the σ, to prevent its broadening owing to fluorophore blinking. When applying HMM to FRET signal, one could fix the FRET states to the distribution obtained from FRET histogram, thus requiring the HMM only to train its transition matrix. This considerably decreases the computational time and cost. In essence, there are two approaches to training the HMM: either by training an individual HMM on each trace in the dataset which is referred to as Local HMM analysis, or by training one HMM model on all traces in dataset which is called a Global HMM analysis. For the work in this thesis, I mostly employed a local HMM on single-color traces (section 5.2 and chapter 6), as there is generally a broad distribution in the exact value of intensities of the traces. In section 5.3, I applied local as well as global HMM to FRET data.

The other important aspect to be considered is the number of states present. The simplest way to determine the number of states is by glancing through the traces. For FRET, one can estimate it from the peaks in the frame-wise FRET histogram. For example, the frame-wise FRET histogram in figure 4.15 shows three peaks, therefore it is a good estimate for the minimum number of states in the system. If the number of states is underestimated, the Viterbi path would show clear deviation from the intensity distribution in a trace. If they are overestimated, several states could have similar value, and not all states might be visited in a Viterbi path. This can be best visualized through a Transition Density Plot (TDP) as explained later. The empirical way to determine number of states is through Bayesian Information Criterion (BIC). BIC is defined as:

$$BIC = -2LL + kln(N) \tag{4.14}$$

where LL is the log-likelihood of the HMM model, k is the number of states and N is the number of data points, which is the total number of frames used for obtaining the HMM model BIC compares the likelihoods of different models by associating each with a 'penalty' factor, that accounts for the additional states. The solution then reduces to finding a global minima in BIC as a function of number of states, which can be quite cumbersome for experimental data.

4.4.4 Transition Density Plot (TDP) and Dwell Time Analysis

Once the traces have been "digitalized", the HMM model generated from the dataset can be visualized through a Transition Density Plot (TDP). Each transition in the Viterbi path is plotted as a two-dimensional Gaussian in a TDP, with the intensity before and after transition as the x and y center coordinates respectively. The standard deviation of noise found in HMM is plotted as the width of the Gaussian. The distribution of the transitions from each of m molecules from an initial state q_i^m to a final state q_i^f is thus given by:

$$TDP(I_i, I_f) = \sum_{transitions,m} \frac{1}{2\pi\sigma_i^m\sigma_f^m}\, e^{-(\frac{(I_i - q_i^m)^2}{2(\sigma_i^m)^2} + \frac{(I_f - q_f^m)^2}{2(\sigma_f^m)^2})} \qquad (4.15)$$

Furthermore, the volume under the curve of 2D Gaussian of each transition is normalized to one. Thus similar transitions would cluster on the TDP. In figure 4.7, the TDP of the global 3 state HMM (right) shows that 6 clusters of transitions are distinguishable, which correspond to transitions between 0.1, 0.4 and 0.7 FRET states. The TDP is symmetric about the diagonal, with the diagonal elements being zero, as there are understandably no transitions from the state to itself. In the case that transitions between all n states is possible, there would be $n^2 - n$ clusters on the plot. If only transitions between neighbors occur, then only $2n - 2$ clusters would be present. In other cases, certain clusters might be missing, if transitions between states are not possible.

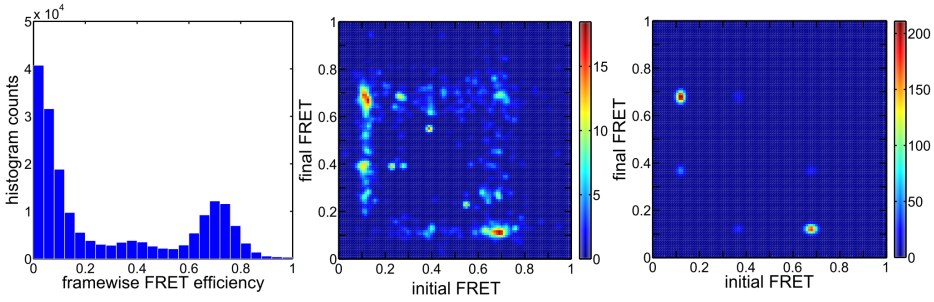

Figure 4.7: The frame-wise FRET efficiency histogram (left) for the experiments shows presence of atleast 3 FRET states. The TDP of local HMM with 3 states is shown in the middle. The TDP of global 3 state HMM is shown on the right.

Within the TDP, a particular cluster of transitions can be selected by an elliptical selection tool (*imellipsis* in MATLAB), and a cumulative dwell time histogram (number of molecules in initial state) can be built from them. Fitting this dwell time histogram with a mono-exponential decay curve (as

shown in figure 4.8) provides the rate of transition and subsequently the dwell time in the initial state.

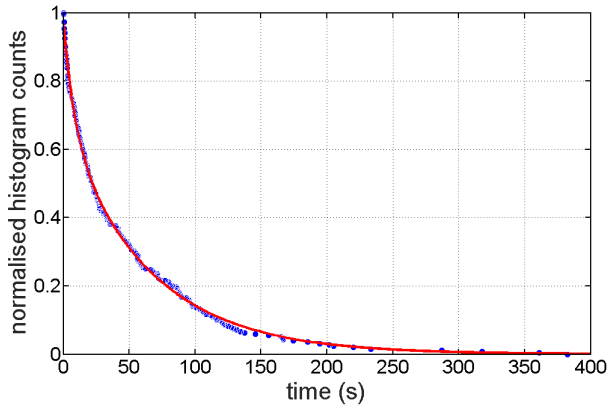

Figure 4.8: The dwell time histogram of the cluster of transitions from FRET 0.1 to 0.7 from the TDP of 3 state global HMM in figure 4.7 is fitted to a mono-exponential decay. The dwell time obtained from the fit is 47.6 s.

Theoretically, the transition rates K between states can be calculated directly from transition probability matrix of the HMM model.

$$K = \frac{(P - \mathbf{1})}{\Delta t} \tag{4.16}$$

where $\mathbf{1}$ is the identity matrix and δt is the duration of a step in the sequence, i.e., the frame time of camera.
The dwell times need a more elaborate approach. By definition, the dwell time of a state is the average time the HMM model stays in that state before transition to a specific other state, which can be calculated by using a Taylor series expansion as follows:

$$\tau_i = \frac{1}{1 - p_{ij}} \delta t \tag{4.17}$$

where p_{ij} is the transition probability from state $i \rightarrow j$ obtained from the transition probability matrix P.

Nevertheless, the dwell time analysis is helpful as it provides a direct visualization of kinetics.

5. Single-molecule Analysis of DNA Origami Systems

Owing to the Watson-Crick base pairing, the unique self-assembly properties of DNA have made it the molecule of choice for building increasingly complex structures on the nanoscale. Pioneered by Seeman [Seeman, 1982, 2003], the field of DNA nanotechnology [Pinheiro et al., 2011] utilizes a variety of bottom-up molecular self-assembly methods to build synthetic DNA nanostructures with precisely controlled features. The most notable and versatile of these is the DNA origami [Rothemund, 2006] technique, that involves hybridization of a long single-stranded DNA "scaffold" (generally a viral genome, e.g. 7249 nt long M13mp18) with multiple short synthetic DNA "staple" strands causing the scaffold strand to fold into a two-or three-dimensional structure whose shape can be "programmed" by the staple strand sequences.

This enables the fabrication of almost arbitrary molecular structures with typical sizes of 10-100 nm, with a precise molecular "addressability" on a length scale of 1 nm. This makes the DNA origami structures the molecular "breadboards" of choice for the precise arrangement of various functional entities (e.g. gold nanoparticles [Ding et al., 2010], proteins [Chhabra et al., 2007, Kuzyk et al., 2009], quantum dots [Bui et al., 2010]etc.). Such structures pave way for applications in molecular computing, nanorobots and molecular assembly lines [Gu et al., 2010]. Furthermore, the 3 and 5-prime ends of the staple strands can be also modified with functional molecules such as biotin, streptavidin or simply extended to act as docking strands as in the case of DNA-PAINT [Jungmann et al., 2010].

The structural information of DNA origami systems is generally revealed through Atomic Force Microscope (AFM) imaging and Transmission Electron Microscope (TEM)imaging. However, these applications involve several dynamic processes and for their precise control, it is essential to monitor them in real time with high temporal resolution.

Single-molecule fluorescence techniques provide a great variety of tools to probe these structures by labeling the DNA strands and detecting the fluorescence. The dynamics in DNA-based systems can be observed through changes in FRET efficiency between the donor and acceptor labeled DNA strands, chosen appropriately to probe the system. The dynamics in the timescales of sub-milliseconds to seconds can be identified through measurements in solution (MFD-PIE) and/or on the surface (surface-MFD,

TIRFM). Moreover, these systems usually involve multiple interacting components. Thus, multi-color FRET can be extremely valuable in gaining insight into the overall picture.

The projects in this chapter are collaboration with the group of Prof. Friedrich Simmel, TUM, Garching. The individual contributions are described in the respective introduction or experimental design sections for each project. These projects mainly involve studying the specially designed DNA origami systems or a specific dynamic functionality within, using two-color and/or three-color FRET.

5.1 Catalytic Hairpin Assembly - CHA

Catalytic hairpin assembly (CHA) is an enzyme-free nucleic acid circuit. It has previously proven useful in amplifying and transducing signals from nucleic acid analytes [Jiang et al., 2013]. CHA provides the advantage of tunability, thereby yielding manifold catalytic amplification. The binding of analytes can be read out as fluorescence or electrical signals [Li et al., 2011, Ren et al., 2011], thus making CHA the first choice for several sensor-based applications [Zheng et al., 2012, Li et al., 2013].

This project was collaboration with Anna Kostina (AK Simmel). The aim of the project was to study autonomous switching of a surface tethered molecular construct performing a catalytic hairpin assembly reaction.

5.1.1 Experimental Design

Experiments were designed to utilize real-time CHA detection by smFRET to determine important reaction parameters (such as temperature, salt concentration, length) that influence the binding kinetics. These would be used as a basis for the development of DNA-based walkers [He and Liu, 2010] utilizing the CHA reaction. The strand sequences were designed by Anna Kostina and measured and analyzed by me and Anders Barth.

The basic elements of the CHA construct are shown in figure 5.1. The different nucleotide domains are numbered and the complementarity is denoted by asterisks (*). The CHA construct consist of 4 ssDNA strands - the anchor strand A (20 nt), the catalyzing strand C (44 nt), and the hairpins H1 (51 nt) and H2 (46 nt). The 5' end of A is modified with biotin, to immobilize it on the PEGylated (PEG/Biotin−PEG) prism surface by biotin-streptavidin-biotin linkage. The catalyzing strand C is tethered to the surface on hybridization of its 0 domain (20nt) with A that consists of the complementary 0^* domain. The 3' end of A is labeled with the acceptor

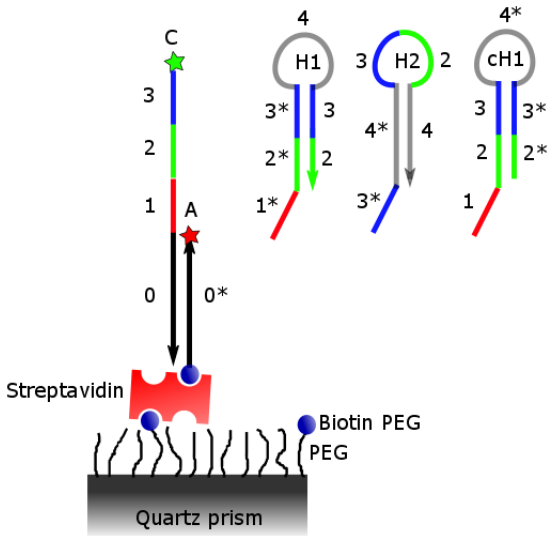

Figure 5.1: Basic elements of the CHA construct. Different nucleotide domains (highlighted in various colors) are numbered and complementary domain strands are marked with asterisks (∗). A is the anchor strand, C is the catalyzing strand and, H1 and H2 are the hairpins.

dye, while 5' end of C is labeled with donor dye.

The hairpin H2 contains domains complementary to H1 (domains 2,3,4* in figure 5.1), hence the hairpins could spontaneously hybridize to form a H1:H2 duplex. However, this is kinetically hindered as the complementary domains are preserved inside the respective hairpin secondary structures.

The steps of the CHA cycle are illustrated in the schematic in figure 5.2. Initially, the *0* domain of C is hybridized with A, and tethered to surface. The 24 nt long unhybridized domains of C (domains *1*, *2* and *3*, each 8 nt long) persist as a random coil, thus the distance between the donor and acceptor is expected to be ∼5 nm (using 0.6 nm/nt for ssDNA, and random coil assumption). When H1 is added, the domain *1* of C serves as a "toehold" for initiating interactions with domain 1* of H1. Hairpin H1 is thus opened by branch migration, resulting in formation of AC:H1 intermediate, and exposing the domains *2*, *3* and *4* of *H1*. Hybridization of the complementary domains of C and H1 results in increase of the donor-acceptor distance to 24 bp or ∼8 nm (using 0.34 nm/bp for dsDNA). On adding H2, the domain 3* of H2 can bind to domain *3* of H1, initiating a branch migration to form AC:H1:H2 complex (not shown here), that being unstable dissociates into AC and H1:H2 duplexes. Therefore, the distance between the donor-acceptor pair again reduces to ∼5 nm. This completes one reaction cycle of CHA and the AC construct is ready to catalyze the hybridization of next pair of H1 and H2 hairpins. Each cycle results in formation of more base pairs. This

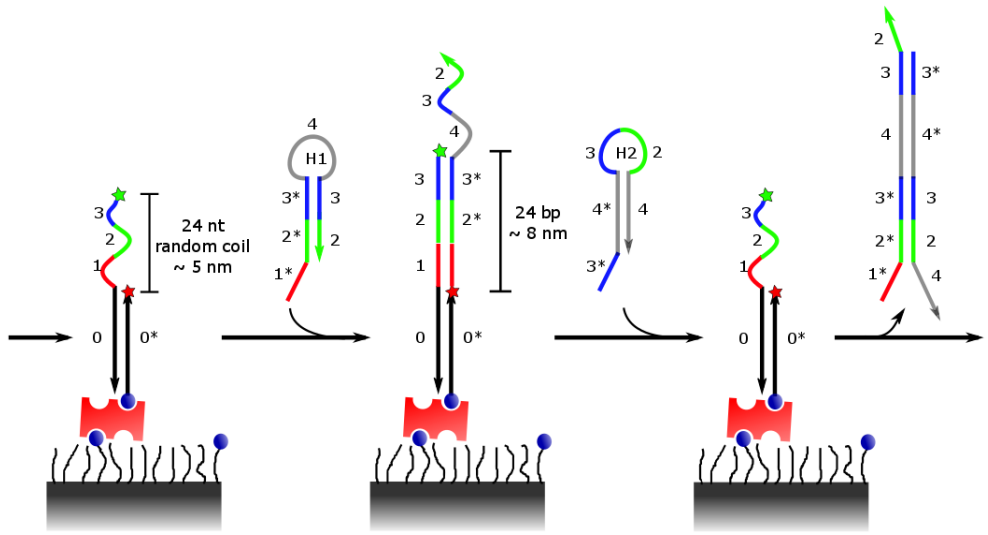

Figure 5.2: Scheme of the CHA cycle. The single-stranded domains of C (1,2 and 3) catalyze the hybridization of hairpins H1 and H2 through a series of toe-hold mediated strand displacement reactions.

causes the enthalpy of overall reaction to decrease and therefore could be partially responsible for driving it [Li et al., 2011].

5.1.2 Experimental Procedure

Experiments to test the CHA cycle and optimize its parameters were performed in solution (three-color MFD-PIE) and on the surface (prism-type TIRF). Two CA constructs were measured, one labeled with Cy3 and Cy5 (C1A1) and the other labeled with Atto565 and Atto647N (C2A2). For C1A1, the hairpins H11 and then H12 were added to see the working of CHA cycle. For C2A2, the hairpins H21 and cH21 (complementary hairpin H21) were added. The sequences are given in Appendix. FRET efficiencies of the donor-acceptor pairs were measured at each step in the CHA reaction cycle: 1.CA alone 2. After addition of H1, followed by 3. Addition of H2. Both the solution and surface measurements were performed with pM concentrations of the CA constructs, followed by the addition of μM concentrations of the hairpins. For TIRF measurements, 10-12 videos with 561 nm laser excitation were recorded for 50 s (1500 frames at 30 ms exposure time) and the fluorescence was simultaneously detected on two cameras, one each for donor and acceptor fluorescence detection.

5.1.3 Results

The solution measurements of the C1A1 construct are shown in fig 5.3. Panel A shows the burst-wise FRET efficiency histogram of the initial

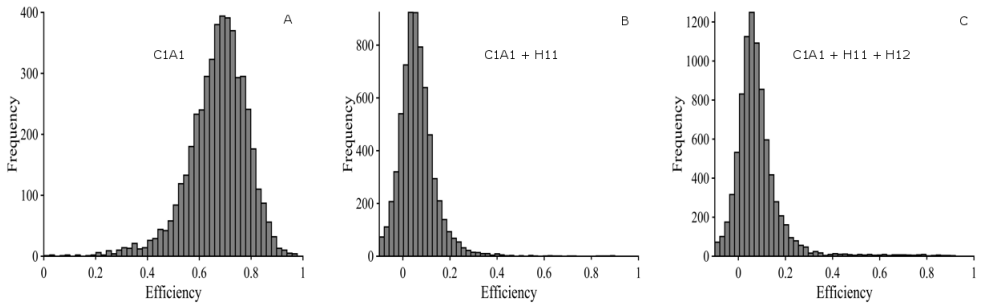

Figure 5.3: The burst-wise FRET efficiency histograms of the C1A1 construct after each step in the CHA cycle. A: The initial FRET efficiency of C1A1. B: The FRET efficiency of C1A1 upon adding H11. C:The FRET efficiency of C1A1 upon further addition of H12. The results were analyzed by Anders Barth.

C1A1 construct. The C1A1 in its "native" form has a FRET efficiency of 0.7. Upon addition of 1 μM H11, the FRET efficiency decreased to 0.05 (panel B) and no change was observed over 5 h of measurement. This was expected from the experimental design, since the distance between the dye pair would be $\sim$8 nm in the fully extended conformation of A1C1:H11. Finally, upon addition of 2 μM of H12 to the previous mixture, the FRET efficiency remains unchanged at 0.05 even after measuring for 5 h. This was contrary to the expectation that on formation of the H1:H2 duplex, the FRET efficiency would increase back to 0.7 as C1A1 returns to its initial state.

To understand why the CHA cycle did not function as expected, and to verify that it was due to inability of H12 to initiate toehold-mediated branch-migration of H11 from the C1A1 construct, the C2A2 construct was measured on TIRF, but with H21 and cH21, instead of H21 and H22.

The single-molecule intensity traces obtained from TIRF measurements showed no switching of FRET state, implying that there were no dynamics involved in the time scales of the experiment. The histogram of mean FRET efficiency of each molecule until bleaching was plotted as the molecule-wise FRET efficiency histogram. Further, the histogram of FRET efficiencies of each frame of every molecule until bleaching was plotted as the frame-wise FRET efficiency histogram. The molecule-wise and frame-wise FRET efficiency histograms obtained from the TIRF measurements of the CHA cycle with C2A2, H21 and cH21 are shown in figure 5.4. The top panel shows the FRET efficiency histograms for C2A2 initially. The C2A2 construct in its initial form has a FRET efficiency of 0.7. Upon the addition of 1 μM H21, the FRET efficiency decreased to 0.5 (figure 5.4, middle panel). Finally, when 2 μM of cH21 was added to the previous mixture in the prism, the FRET efficiency stayed at 0.5 (figure 5.4, bottom panel). The molecule-wise

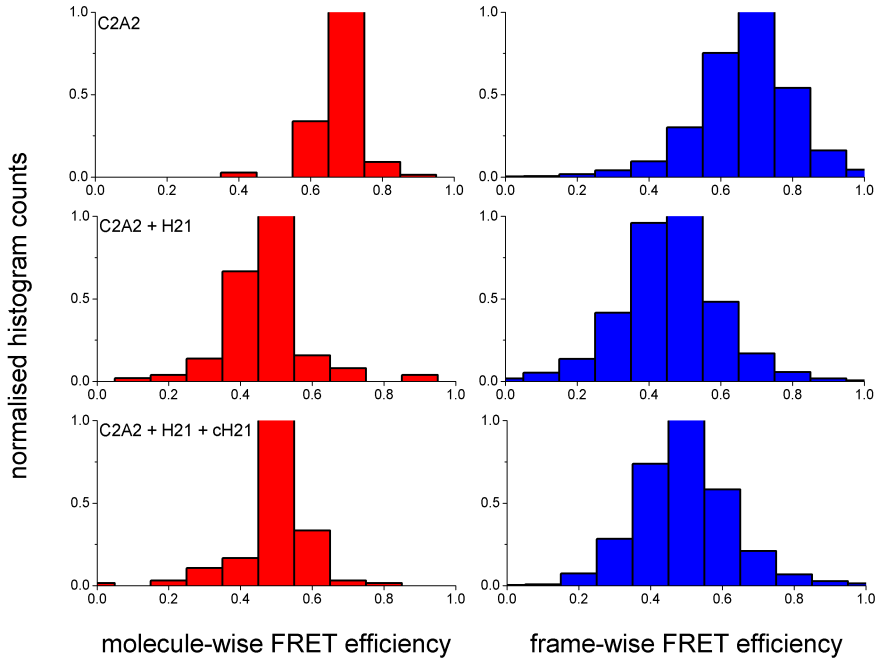

Figure 5.4: Normalized FRET efficiency histograms molecule-wise (red) and frame-wise (blue) for C2A2 construct after each step in the CHA cycle. Top: The initial FRET efficiency of C2A2. Middle: The FRET efficiency of C2A2 upon addition of H21. Bottom: The FRET efficiency of C1A1 upon further addition of cH21.

FRET histograms show a single Gaussian peak, indicating only a single population of molecules. The frame-wise FRET histograms show a single Gaussian FRET distribution, indicating the absence of any dynamics or change in FRET efficiency of a single molecule.

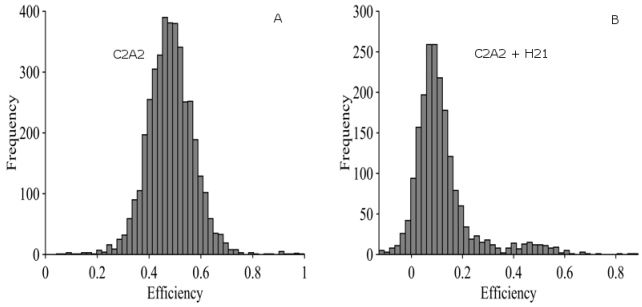

Figure 5.5: The burst-wise FRET efficiency histograms of the C2A2 construct during the first two steps of the CHA cycle. A: The initial FRET efficiency of C2A2. B: The FRET efficiency of C2A2 upon addition of H21. The results were analyzed by Anders Barth.

As a last experiment, the C2A2 construct was measured in solution with Atto488 labeled cH21. The C2A2 in its initial form has a FRET efficiency of 0.48 (figure 5.5A). Upon addition of 1 μM H21, the FRET efficiency decreased

to 0.1 (figure 5.5B) and remained so even after 5 h of measurement. An additional small peak at 0.48 was still noticeable, which could indicate the presence of a small amount of unbound C2A2 present in its random-coil state.

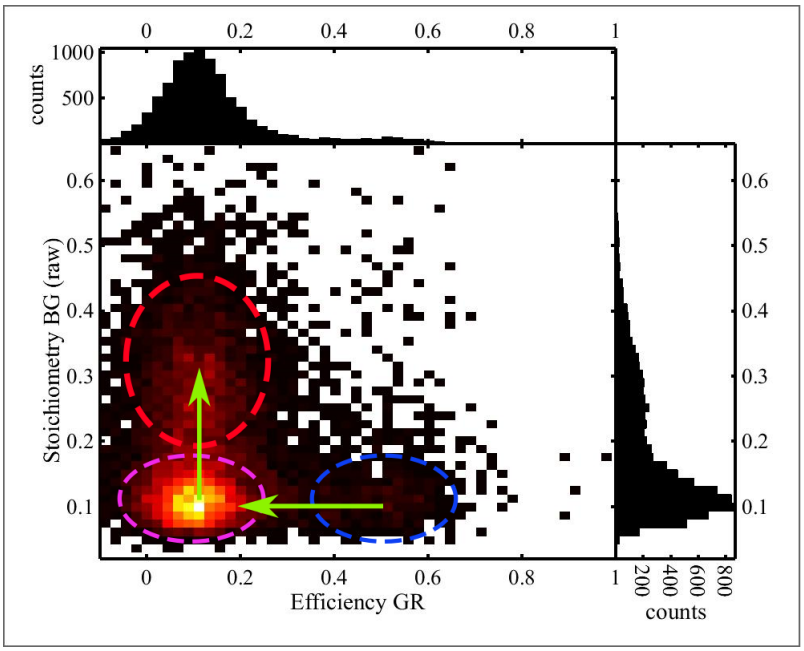

Figure 5.6: The BG stoichiometry vs GR FRET efficiency plotted for the reaction mixture of C2A2 with H21 and cH21 labeled with Atto488. The results were analyzed by Anders Barth.

Finally, 2 μM of Atto488-cH21 was added to the previously incubated mixture of C2A2 (100 pM) and 1 μM H21. This mixture was then diluted to pM concentration and measured on the three-color MFD-PIE setup. The results are shown in figure 5.6 as a BG stoichiometry vs GR FRET efficiency plot. The green arrows indicate the progression of the reaction as a whole. The blue dashed circle points to the coiled C2A2 still in its initial form. Upon addition of H21, A2C2:H21 intermediate is formed (green dashed circle) which notably has lower GR FRET efficiency as the distance between the dyes increases. Upon further addition of cH21, the BG stoichiometry shifts from 0.1 to 0.3 indicating that the sample has both fluorophores (Atto488 and Atto565) together, which points to the fact that cH21 binds to H21 but does not manage to remove it from C2A2. Thus, all the components of the CHA cycle are stuck together and the cycle therefore cannot proceed further. This would explain why we do not observe an increase in the FRET efficiency of C1A1 upon addition of H12. In the case that C1A1, H11 and H12 have all stuck together, the FRET efficiency of C1A1 will not change and the cycle cannot continue.

5.1.4 Discussion

SmFRET measurements were performed to monitor the progress of the CHA cycle by measuring the FRET efficiency between the donor labeled catalyzing strand C and the acceptor labeled anchor strand A at each step in the reaction cycle. Two constructs were prepared with different dye pairs; C1A1 labeled with Cy3 and Cy5, C2A2 labeled with Atto565 and Atto647N. Both the constructs were measured in solution and C2A2 was also measured on TIRFM. In both the cases, the FRET efficiencies decreased upon addition of the first hairpin; from 0.7 to 0.05 for C1A1 (figure 5.3) and from 0.48 to 0.1 for C2A2 (figure 5.5). This result is in agreement with the CHA cycle design, as the toe-hold mediated branch migration of the first hairpins onto the CA constructs would increase the distance between the dye pairs. However, further addition of the second hairpins did not result in the FRET efficiencies returning to their original values. Instead, the FRET efficiencies remained unchanged. This was in contradiction to the assumption from the CHA cycle that the second hairpins initiate a toe-hold mediated branch migration of the first hairpins, thus removing them from the CA constructs. This should have actually resulted in a return of the CA constructs to their initial states and the catalysis process can start all over again. This discrepancy was confirmed by TIRF measurements of C2A2 too (figure 5.4).

In order to understand the reason for this discrepancy, it was essential to probe the process of binding of the second hairpin to the CA bound first hairpin. Three-color MFD-PIE experiments were performed with Atto488 labeled cH21 as the second hairpin. The results (in figure 5.6) showed that upon addition of Atto488 labeled cH21, populations with two different BG stoichiometries, 0.1 and 0.3, were observed. The larger population with BG stoichiometry 0.1 and GR FRET efficiency 0.1 can be explained as the complex A2C2:H21. These complexes have no cH21 bound, as verified by the low BG stoichiometry, pointing to a green-only population. This could just be the result of the how the experiment was performed; upon addition of 2 μM Atto488-cH21, the sample was diluted to picomolar concentration before putting on the MFD setup.

Alternatively, the presence of a large population of A2C2:H2 can be explained by the strand sequence design. The toe-hold domain of cH21 (domain 1) does not have the complementary domain 1^* available in H21 for binding (strand domains shown in figure 5.1); 1^* of H21 is already bound to domain 1 of C2. Moreover, the domains in cH21 that are complementary to the freely available domains of H21 (2, 3 and 4) are involved in the hairpin secondary structure, making it difficult to initiate binding. These structural factors could be contributing to the inability of cH21 to initiate a toe-hold

mediated branch migration of H21, thus explaining the presence of a large population of A2C2:H21 complex.

The smaller population with BG stoichiometry 0.3 indicating blue-green double-labeled molecules, and GR FRET efficiency of 0.1 indicating a larger distance between the fluorophores compared to the initial state of C2A2 can be explained as formation of a complex of A2C2:H21:cH21, where all the strands are stuck together. Interestingly, this intermediate was assumed to be quite unstable in the CHA cycle, thereby dissociating into the more stable H21:cH21 and C2A2 duplexes. Another plausible explanation could be the formation of a A2C2:cH21 complex. There is a possibility that, instead of a branch migration, the cH21 initiates displacement of H21 from the C2A2 construct. Additionally, there is a very small population with BG stoichiometry 0.1 and GR FRET efficiency 0.48, which is the C2A2 construct alone, with no H21 strand bound.

The three-color experiment was clearly advantageous in explaining the various possibilities that could provide the results as seen in two-color experiments. Multi-color FRET is especially helpful in resolving multi-component interactions as seen in the CHA cycle. Through a single three-color MFD experiment with PIE, the different ways in which the components of CHA cycle could interact were deciphered.

These possibilities were further verified by strand equilibrium analysis using NUPACK [Zadeh et al., 2011] package at a temperature of $\sim 18^oC$ and using the concentrations applied in the experiments. NUPACK analysis of the given strand sequences for C1A1 with H11 and H12 predicted that, at equilibrium, the mixture would consist of about 32% each of H12 and A1C1:H11:H12, about 16% each of H11:H12 and A1C1:H12 and less than 1% of A1C1. Similarly, the equilibrium analysis for C2A2, H21 and cH21 resulted in about 40% each of cH21 and A2C2:H21:cH21 and about 6% each of H21 and A1C1:H12, and H21:cH21 making up the rest. The possible drawback of NUPACK analysis is that it assumes that all the strands are present together from the beginning. This is however not true for our case, where the experiments were performed by sequentially adding the hairpins to CA constructs. A more precise step-by-step computational modeling could be helpful in this case.

Overall, these results suggest that the constructs did not function as expected. In principle, the sequences could be redesigned to be better suited for CHA. The lengths of the domains could be altered to change the thermodynamic and kinetic properties. Especially, the toe-hold binding sequence must be optimized to ensure efficient strand displacement and yet dissociate easily to retrieve the catalyst. Also, any secondary structure

formations or strand dimerizations should be avoided. Further, special attention must be paid to maintaining an appropriate amount of GC content in each domain; around 50% is beneficial for effective binding.
Alternatively, the same experiments could be repeated at a higher temperature (37^oC) in an incubator with an objective heating ring on an oil-objective. This could perhaps ease the removal of the hairpin pair from the CA construct. Moreover, CHA has been successfully implemented for the temperatures starting from 37^oC upwards [Jiang et al., 2013].
Nevertheless, this project shows how the application of the multi-color FRET approach to multiple interacting components labeled with different fluorophores is an efficient way to obtain maximum information about the system. The visualization of the processes and the species involved was simplified through the BG stoichiometry vs GR FRET efficiency plot.

5.2 The DNA Roboarm

DNA Nanotechnology has paved the way for the realization of nanomachines made out of DNA [Lund et al., 2010, Douglas et al., 2012]. The long-term goal would be to assemble nano-factories with molecular assembly lines. Though assembly lines with molecular walkers has been attempted before [Gu et al., 2010], here a more classical approach is taken. The pre-requisite for any assembly line would be mechanical realization of a positioning arm, that picks up the cargo from a location and drops it onto another location. In contrast to classical molecular walkers, here only the start and end points would matter and not the exact route taken by the cargo.

Using the DNA origami technique, a "robotic-arm" or "roboarm" structure could be designed, where the arm angles are controlled through design, leaving diffusion to do the rest. The advantage of using diffusion as the driving force is that it enables faster transport than the current artificial walker systems. Typical hybridization-based DNA walkers move at $\sim$0.01 nm/s, while the arm switches between left and right ($\sim$50 nm) in tens of ms. Moreover, tethering the arm to the platform allows it to stay in the vicinity of the load, resulting in faster reaction kinetics (binding to the load is easier). Previous results have shown that diffusive motion on a molecular tether has a highly efficient mechanism for fast transfer of cargos over long distances [Kopperger et al., 2015].

The long term goal of this project is to have a one-pot reaction for constructing a robust DNA roboarm with addressable positions on arm and the platform that can be used to study different kinds of dynamic interactions between molecules that are fixed at those positions. The dynamic interactions were studied by labeling the arm with a donor dye and the tethering sites on the platform with acceptor dyes. A schematic of the roboarm design and functioning is shown in figure 5.7.

The DNA roboarm construct consists of a square platform $\sim$55 nm x 55 nm with a roboarm consisting of six-helix bundles in the middle of the platform. The structure is a continuous DNA block, folded in a one-pot reaction. The roboarm is $\sim$25 nm long and is labeled with the donor Alexa488 at the end of the arm. The roboarm also has an extended staple strand at the end of the arm, with certain number of bases complementary to the docking strands on the platform, which is hereafter referred to as the "arm-dock". The two single stranded docking strands are present on the platform, approx. 43 nm apart, at an angle of $\sim 165^o$ to the roboarm hinge. These docking strands are labeled with Atto565 and Atto647N. The roboarm would be free to diffuse around and could temporarily bind to the docking strands with the arm-dock.

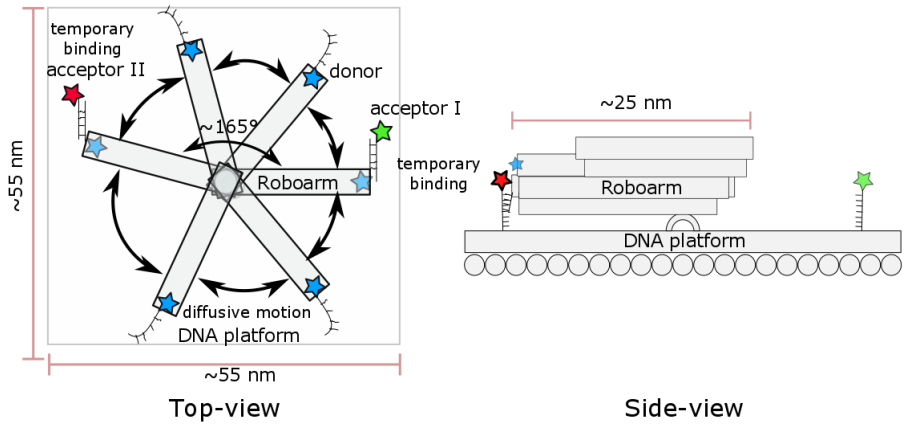

Figure 5.7: The DNA Roboarm construct. The top-view with the arm labeled with donor Alexa488 diffusing around, and transiently binding to the acceptor Atto565 and Atto647N labeled docking strands on the platform. The side-view shows the six-helix arm labeled with the donor, and the acceptors labeled docking strands on the origami platform.

5.2.1 Experimental Procedures

Sample Preparation

The DNA Roboarm was designed and prepared by Enzo Kopperger (AK Simmel). A detailed description will be made available in the paper to be published on this project. Hence, only a concise explanation is given here. Four samples were prepared, difference being the absence of arm-dock, or presence of arm-dock with varying overlap length (number of complementary nucleotides on the arm-dock on Roboarm with respect to the docking strands on the platform labeled with the two acceptor fluorophores) - 8 nt, 9 nt and 10 nt. Each of the samples was prepared in a similar fashion. Briefly, modified M13mp18 of 7705 nt length (provided by Florian Praetorius, AK Dietz, TUM) was folded with 2-fold molar excess of staples that take part specifically in the structure folding. The special purpose staple strands, such as those with fluorophores, biotin anchors and the docks from arm to the platform, were incorporated during origami folding in 5-fold excess. All oligos were ordered from MWG Biotech AG (Eurofins Genomics, Ebersberg, Germany).

Folding was performed in Assembly buffer: 1xTAE (40 mM Tris, 20 mM Acetic Acid, 1 mM EDTA) + 20 mM $MgCl_2$) using a temperature ramp from 70^{o}C to 20^{o}C over 16 h, and then held at 40^{o}C for 5 h. This was followed by the precipitation of structures by adding 11% PEG (PEG 8000 molecular biology grade, PromegaGmbH, Mannheim, Germany) in 1xTAE + 1 M NaCl and then resuspension in 1xTAE + 1 M NaCl. Further purification steps were performed in buffer with 1xTAE + 1 M NaCl and the final concentration of the sample adjusted to 10 nM.

SmTIRF Measurements

Each of the assembled Roboarm samples (8 nt, 9 nt, 10 nt and no arm-dock) were diluted to 100 pM in Working Buffer: 1xTE (10 mM Tris, 1 mM EDTA, pH 8.0) + 12.5 mM $MgCl_2$ and put into the surface-prepared prism flushed with streptavidin solution (0.3 mg/ml in PBS), to tether the structures onto the surface through a biotin-streptavidin-biotin linkage. After 5 min., the prism was again flushed with working buffer to remove the untethered origami structures. This was followed by flushing with Imaging Buffer (1xTE, 12.5 mM $MgCl_2$, 10 $\%v/v$ Glycerol, 10 $\%w/v$ Glucose) with 1 mM Trolox, pre-activated under a UV lamp for 10 min. and subsequently adding fresh GODCAT (Glucose oxidase + Catalase, 1:10 dil. of stock). Finally the same Imaging Buffer with Trolox and GODCAT was filled into both channels of the prism and the channels were sealed to facilitate oxygen removal.

Several movies were recorded with simultaneous acquisition on 3 EMCCD cameras on the three-color TIRF setup with exposure time of 30 ms (frame time ∼33 ms) for 3000 frames, in both the ALEX mode (647 nm: 3 mW, 561 nm: 5 mW, 491 nm: 11 mW) and only excitation with the blue (491 nm: 11 mW) laser .

5.2.2 Three-color Traces

The single-molecule fluorescence intensity traces were extracted from the movies measured using a home-written MATLAB program (MITrA). Each fluorescent spot in the movie is presumed to represent a labeled Roboarm structure. Using ZMW movies, the EMCCD cameras were mapped onto one another to detect the same molecule on all three cameras. The program then extracts the intensity traces from respective cameras and constructs a unified trace for each molecule, showing the intensities of the three fluorophores as a function of time. Further, each color trace of every molecule was individually subjected to an HMM analysis to determine the Viterbi path.

Owing to the presence of photococktail with oxygen scavenging system, the fluorophores usually survived for the entire duration of the video (∼300 s). An exemplar trace for 9 nt docking overlap Roboarm with only blue excitation is shown in figure 5.8. Since only blue excitation is used in this case, the fluorescence signal from green or red dye is due to FRET from the blue dye when it is in proximity to the respective acceptor. As evident from the top panel of the figure, it was observed that the intensities of each of the fluorophores was fluctuating between two levels - a high fluorescence intensity level and a low fluorescence intensity level. Moreover, at any given point of time, only one of the dyes is fluorescing in its high intensity level,

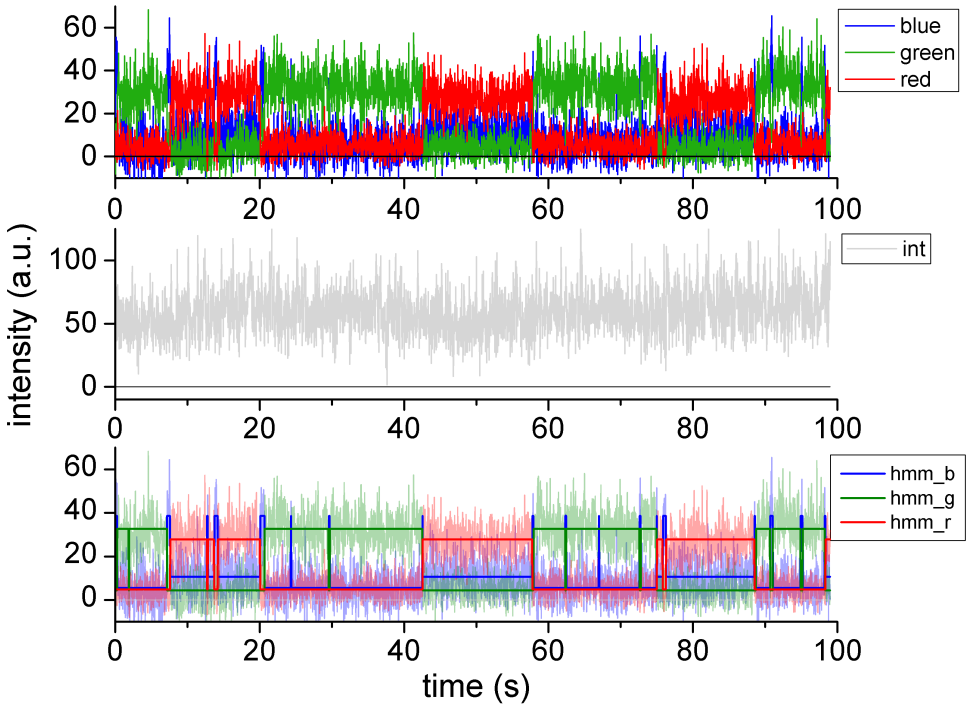

Figure 5.8: DNA Roboarm dynamics example trace for 9 nt docking overlap with blue excitation. Top panel: The intensities in the blue, green and red channels represented in their respective colors. Middle panel: The total fluorescence intensity = sum of intensities in all three channels is plotted in grey. Bottom panel: The three-color trace shown in the top panel with the individual HMMs. The blue trace is fit with three-state HMM (cyan), the green trace is fit with a two-state HMM (fluorescent green) and the red trace is fit with two-state HMM (black).

while the other two are in their respective low fluorescence intensity levels.

The high fluorescence intensity level of the blue dye is an unquenched state of the blue dye. This depicts the case when the blue dye is not in the vicinity of either the acceptors, therefore its fluorescence in not quenched by FRET. When the blue dye is in an unquenched state, the fluorescence intensities of both the acceptors are in their low fluorescence intensity levels.

An acceptor is in its high fluorescence intensity level when there is FRET to that particular acceptor. This depicts the case when the roboarm is docked at that particular docking strand, thus the blue dye is in vicinity of the acceptor for FRET to occur. This corresponds to the quenched state of the blue dye. The other acceptor is in its low fluorescence intensity level as there is no FRET between the acceptors, as expected from the structure, since a separation of ∼50 nm is beyond the FRET of range.

In order to simplify the working of the system, the high fluorescence intensity level would be referred to as the "ON" state and the low

fluorescence intensity level as the "OFF" state of the dyes. Thus, the observations can be summarized as follows:

1. At any point of time, only one dye is in the ON state, while rest are in OFF state. This is supported by the total intensity trace in the middle panel (figure 5.8), which is almost constant, irrespective of which dye is in ON state.

2. The blue ON state refers to the unquenched state of the blue. This depicts the case when the blue dye is not in the vicinity of either the acceptors, thus the Roboarm is diffusing somewhere in between the two docking strands. Hereafter, the blue ON state is also referred to as blue "blips", as generally they are short-lived compared to green ON or red ON states.

3. The green ON state refers to the green dye fluorescing as a result of FRET from the blue dye. This depicts the case when the Roboarm is docked at the green dye labeled docking strand.

4. The red ON state refers to the red dye fluorescing as a result of FRET from the blue dye. This depicts the case when the Roboarm is docked at the red dye labeled docking strand.

Furthermore, most of the time, only green or red is in the ON state, interspersed with blue blips. Two type of blue blips are visible:

1. Type I blip - Blue blip between green ON and red ON (or vice versa)
   These blips indicate the traversal of Roboarm from one docking site to the other.

2. Type II blip - Blue blip between green ONs or between red ONs
   These blips can be inferred as the arm moving away from the docking site and returning, instead of traversing to the other side. Interestingly, these types of blips are much shorter in duration.

The third panel in the figure 5.8 shows the three colored traces, each fit individually with an HMM model. For fitting the green and red traces, two states were sufficient - which clearly coincided with the fluorophores being ON or OFF. On the other hand, the blue dye was best fit with three states - ON, OFF when green is ON and OFF when red is ON.

Traces with similar behavior were observed for 8 nt and 10 nt docking overlap with blue excitation. For the Roboarm with no arm-dock strand, the traces were instead characterized by a mostly blue ON state as shown in the exemplar trace in figure 5.9. Only residual signal was observed in the

green and red channels, which can be accounted for by the direct excitation of dyes and cross-talk from blue channel.

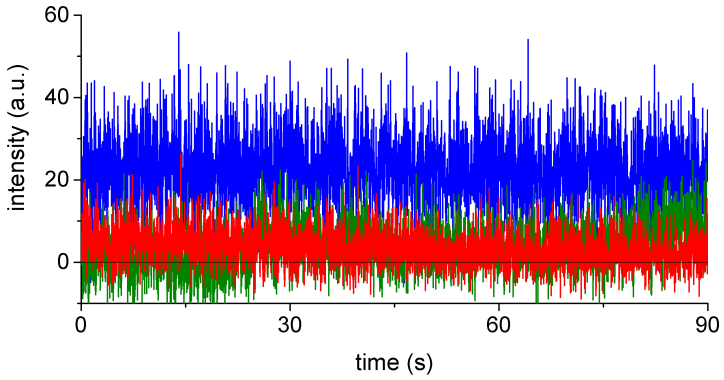

Figure 5.9: An exemplar trace of Roboarm without arm-dock strand.

ALEX

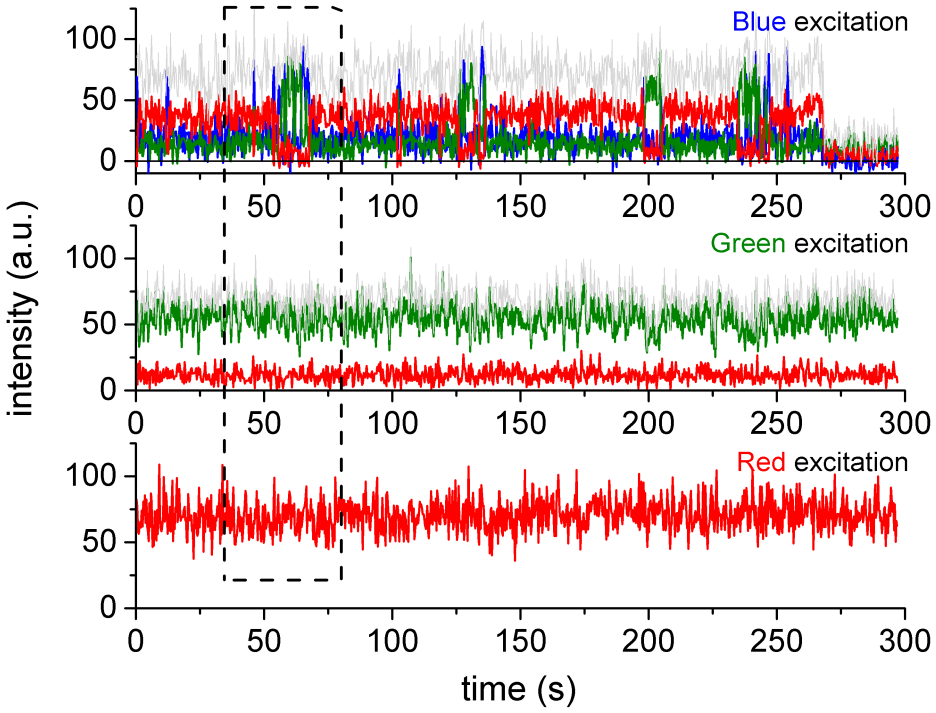

Figure 5.10: Intensity traces of a Roboarm molecule with 9 nt docking overlap and on R-G-B ALEX excitation scheme with simultaneous detection on all three cameras. Top panel: Intensity traces of all three dyes from the respective channels, upon blue excitation. Middle panel: Intensity traces of the red and green dyes from the respective channels, upon green excitation. Bottom panel: Intensity trace of red dye as detected in the red channel (camera) upon red excitation. The part of the trace highlighted in the dashed box is expanded in the figure 5.11

Each sample was also recorded using alternating laser excitation (ALEX), with the three lasers in the RGB scheme, and simultaneous detection on all three EMCCD cameras. An ALEX trace for the 9 nt docking overlap is shown in figure 5.10, with the top panel showing the intensity in the red channel upon red excitation, the middle panel showing the intensity in red and green channels on green excitation and the bottom panel showing all three channels upon blue excitation. The total intensity after each excitation is shown in grey. Since an RGB excitation scheme is followed, starting from the first frame, every third frame would give the intensity upon red excitation, and starting from the second frame, every third frame would give the intensity in the respective channels upon green excitation. Similarly, starting from the third frame, every third frame would give the intensity in the respective channels upon blue excitation. Since shorter wavelength fluorophores are not excited by longer wavelengths, traces were plotted in this manner.

It is clearly seen in the ALEX trace (figure 5.10) that all the three dyes were present and initially active, though the blue dye bleached around 275 s into the experiment. Furthermore, there is no FRET between the green and red dyes, as seen from the traces upon green excitation (figure 5.10, middle panel). The residual intensity observed in the red channel upon green excitation is due to the spectral cross-talk from the green dye and a small amount of direct excitation of the red dye by the green laser.

Upon blue excitation, there is FRET between the blue and green dyes and the blue and red dyes, as seen in the bottom panel of figure 5.10. This is further supported by the fact that, once the blue dye bleaches, there is no signal left in the green or red channels and the total intensity signal (in grey) drops to zero.

This is better visualized in the fig 5.11 where the part of the ALEX trace in dashed box has been zoomed in for better clarity. The total intensity (in grey) upon blue excitation (figure 5.11, bottom panel) remains constant despite the dyes switching between their respective ON and OFF states. This shows that the switching is the result of change in FRET from the blue dye. Also, the two types of blue blips discussed earlier are clearly seen here. Around 46 s, the roboarm moves away from the red dye labeled docking strand and soon after returns back to it, as evident from the type II blue blip. Around 56 s, Roboarm again moves away from the red dye labeled docking strand, but this time it traverses to the green-dye labeled docking strand, as evident from the type I blue blip.

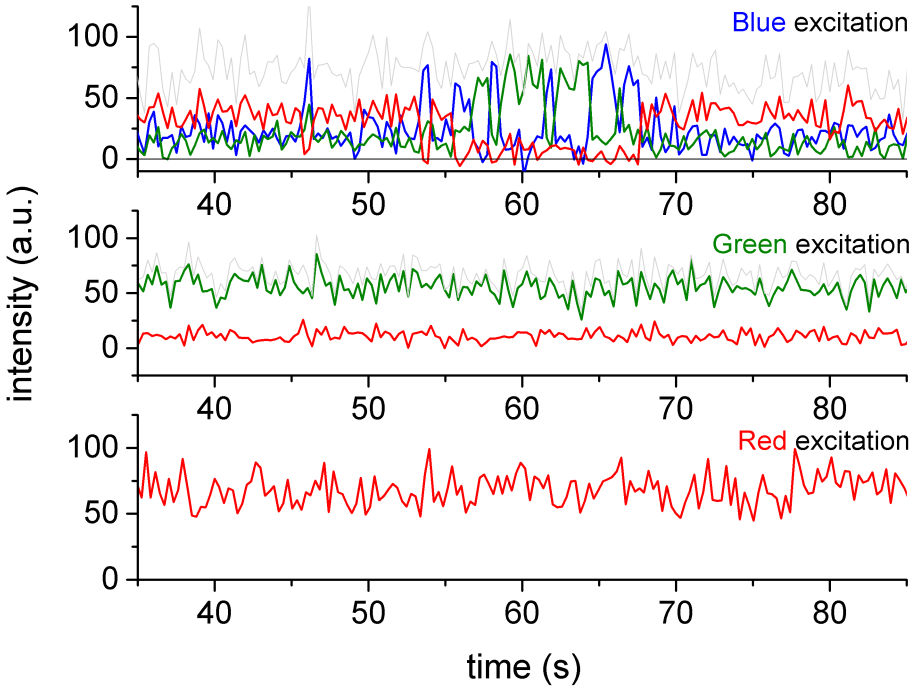

Figure 5.11: Zoom in of the portion of the ALEX trace in figure 5.10 inside the dashed box. The three panels (top to bottom) show intensity traces of the dyes upon blue, green and red excitation respectively.

5.2.3 Dwell Time Analysis

The traces of Roboarm involve two FRET interactions - FRET between the blue and the green dyes, or FRET between the blue and the red dyes, depending on which site the arm is docked at. Since multiple dyes and different number of states for the dyes are involved, the direct dwell time analysis from the HMMs was not possible. Hence I developed a new approach to determine the dwell times of the different states of each dye. The analysis was automated through a self-written code in MATLAB.

Figure 5.12 illustrates the steps involved in this new approach. First of all, every color trace for each molecule was subject to an individual HMM analysis, with the blue trace being fit with a 3 state HMM and the red and the green traces being fit with 2 state HMMs. This is followed by generating a color-ON trace from these three HMMs, which is essentially a sequence of Rs, Gs and Bs, with the R or G or B assigned to each frame according to which dye is in ON state according to the HMMs. Basically, the color-visible trace behaves like a cumulative HMM, showing which dye is ON at any time point. The color-ON trace is shown for a part of the trace in the inset in figure 5.12.

In the next step, the color-ON trace is shortened by replacing the consecutive

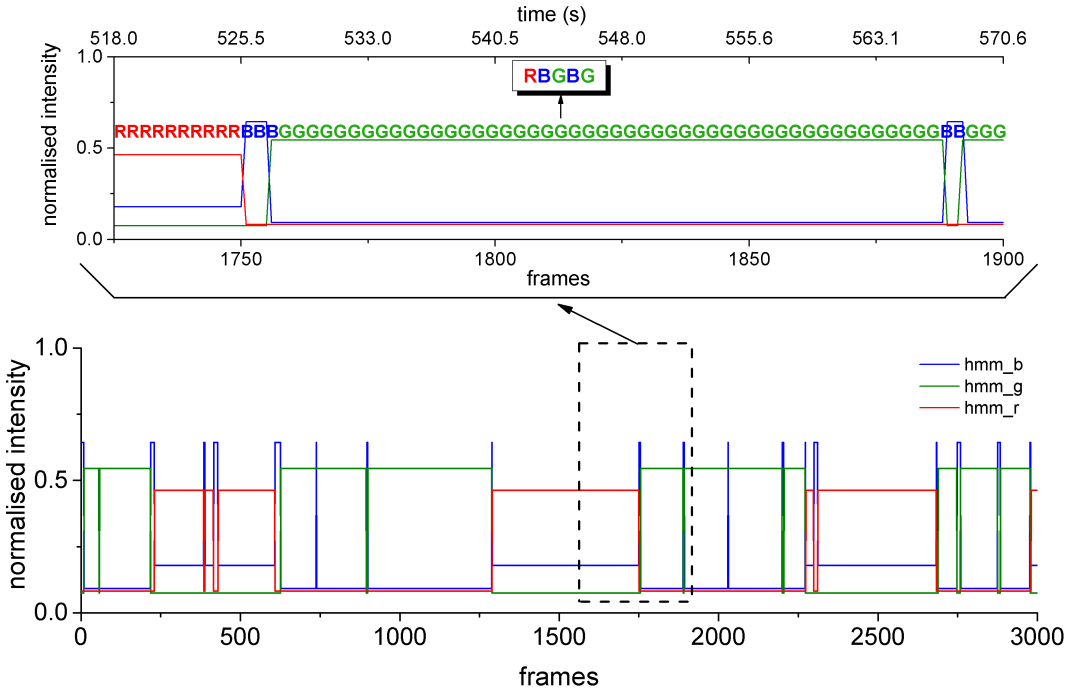

Figure 5.12: The Dwell Time analysis approach. The individual HMMs for the different dyes are plotted with a 3 state HMM for blue dye in cyan, a 2 state HMM for the green dye in fluorescent green and a 2 state HMM for the red dye in black. Inset shows a zoom-in of the part of the trace highlighted with the dashed box, with the color-ON trace and the shortened color-ON trace.

same color frames with a single frame. In principle, a shortened color-ON trace depicts the sequence of transitions between the ON states of different dyes. This makes it easier to observe the different types of blue blips, which would be visualized by the sequences: RBG or GBR for the blue blips of type I and RBR or GBG for blue blips of type II.

Furthermore, counting the number of frames for which a color is ON in the color-ON trace and multiplying with the frame time gives the dwell time of the ON state for that dye. In this way, the average ON dwell time of each of the dyes, for the different types of blue blips and for specific sequences for each blip type were calculated and plotted in figure 5.13. The error bars represent the standard error of mean from 3 experiments, where each experiment consisted of 10-12 videos, with approx. 500 traces and several thousand transitions. The values are given in table 5.1.

In essence, the DNA roboarm is free to diffuse in any direction around its hinge. Once the roboarm unbinds from a docking strand, it is assumed to have an equal probability to traverse to the other docking site or return to the same site. However, when the average probabilities for each case of transition sequences: RBG vs RBR and GBR vs GBG were calculated, the

overlap	Blue	Green	Red
8 nt	0.13±0.01	1.01±0.10	1.00±0.07
9 nt	0.24±0.02	2.55±0.07	2.50±0.19
10 nt	0.38±0.07	3.61±0.36	3.09±0.39

Table 5.1: The average ON dwell times (in s) for the different docking overlap lengths for the different dyes.

overlap	RBG	GBR	RBG+GBR
8 nt	0.17±0.01	0.16±0.001	0.16±0.004
9 nt	0.37±0.02	0.22±0.02	0.30±0.02
10 nt	0.8±0.18	0.39±0.05	0.61±0.12

Table 5.2: The average ON dwell times (in s) for the different docking overlap lengths for the different blue blips.

overlap	RBR	GBG	RBR+GBG
8 nt	0.08±0.003	0.19±0.02	0.12±0.01
9 nt	0.16±0.02	0.28±0.02	0.21±0.02
10 nt	0.14±0.002	0.41±0.08	0.24±0.03

Table 5.3: The average ON dwell times (in s) for the different docking overlap lengths for the different blue blips.

overlap	P(RBG)	P(GBR)	P(RBR)	P(GBG)
8 nt	0.37±0.03	0.53±0.01	0.72±0.02	0.61±0.003
9 nt	0.59±0.01	0.65±0.01	0.69±0.003	0.68±0.01
10 nt	0.73±0.01	0.76±0.02	0.75±0.01	0.77±0.02

Table 5.4: The average probability of a type I blue blip vs a type II blue blip: RBG vs RBR sequence and GBR vs GBG sequence, for different docking overlap lengths.

overlap	P(RBG)	P(GBR)
8 nt	0.81±0.001	0.81±0.01
9 nt	0.90±0.01	0.91±0.01
10 nt	0.95±0.003	0.94±0.002

Table 5.5: The average probability of a GBR sequence type I blue blip vs GR sequence and a RBG sequence type I blue blip vs RG sequence for different docking overlap lengths.

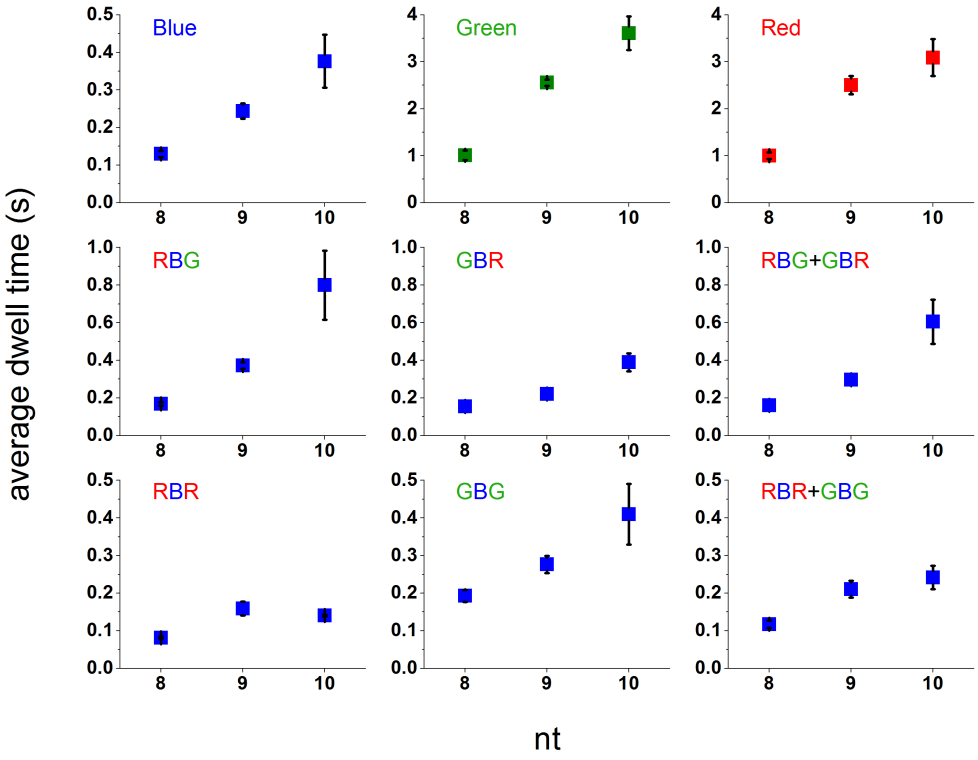

Figure 5.13: The average ON dwell time plot for different docking overlap lengths. Top row: Average ON Dwell time of the blue dye (left column), green dye (middle column) and red dye (right column). Middle row: The average ON dwell time of different type I blue blips - RBG blips (left), GBR blips (middle) and both together (right). Bottom row: The average ON dwell time of different type II blue blips - RBR blips (left), GBG blips (middle) and both together (right). The error bars represent the standard error of the mean from 3 experiments.

results showed a different trend, as shown in figure 5.14. The probability of finding a sequence of type I blue blip vs the corresponding type II blue blip was calculated for each molecule and then averaged over all molecules, hence the probabilities shown in the figure do not add up to 1.

In the case of an ideal Roboarm with arm-dock, one would expect to always observe a blue ON state between a green ON and a red ON state, as the blue dye should not be quenched when traversing from one docking site to the other. However, it was observed that this was not the case sometimes. Occasionally, no blue blip is observed and there is a direct RG or GR transition in the reduced color-ON trace. This could happen if the type I blue blip is shorter (the traversal from one docking strand to the other is faster) than the frame time, thus the camera is unable to detect the signal from the blue dye.

To determine the significance of such rare events, the probability of finding

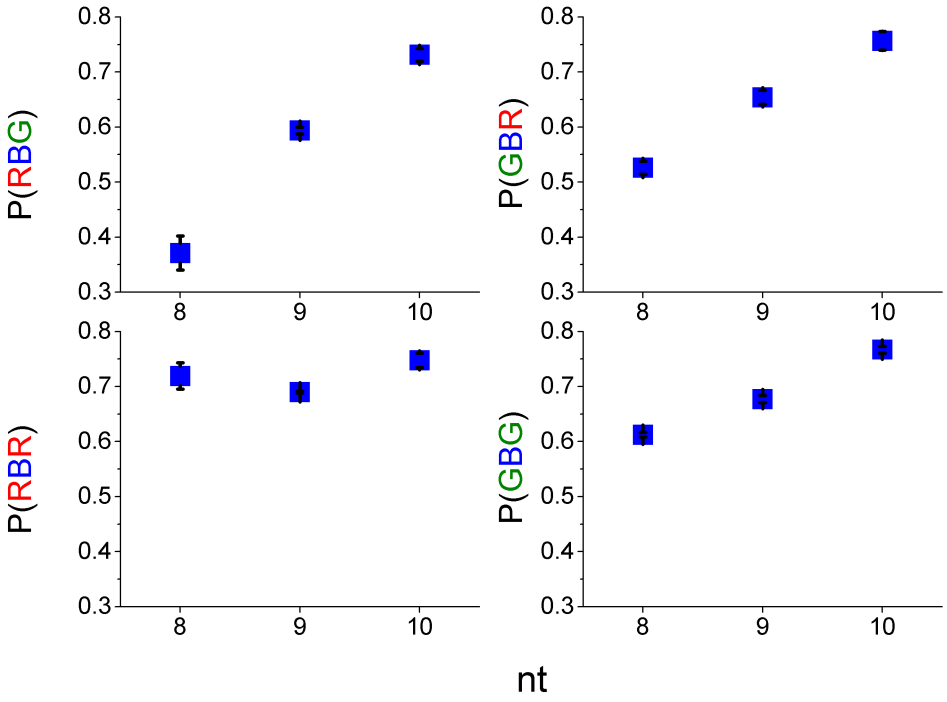

Figure 5.14: The average probability of a type I blue blip compared to a type II blue blip and vice versa. The average probability of a RBG sequence type I blue blip (top-left) compared to a RBR sequence type II blue blip (bottom-left), and a GBR sequence type I blue blip (top-right) compared to a GBG sequence type II blue blip (bottom-right) for different docking overlap lengths.

a RBG or GBR type I blue blips compared to all traversal transitions (GR, RG, GBR, RBG) was calculated and plotted in figure 5.15 for each docking overlap length.

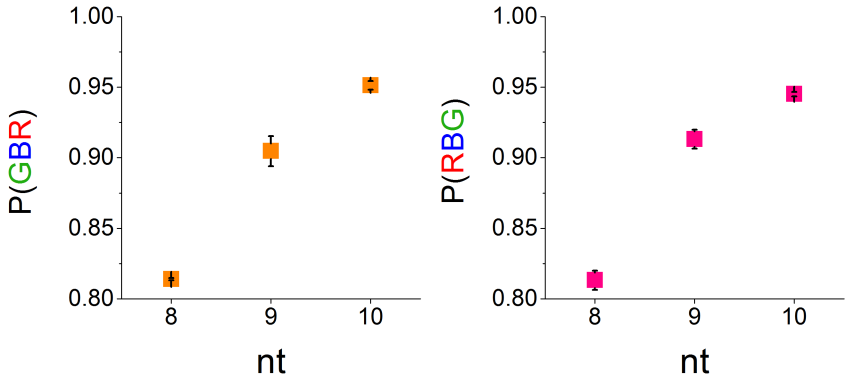

Figure 5.15: The average probability of a distinct type I blue blips. Probability of a GBR sequence type I blue blip (left) compared to a GR sequence and a RBG sequence type I blue blip (right) compared to a RG sequence for different docking overlap lengths.

5.2.4 Special Cases

In the analysis of different traces for different docking overlap lengths, there were two special types of traces observed.

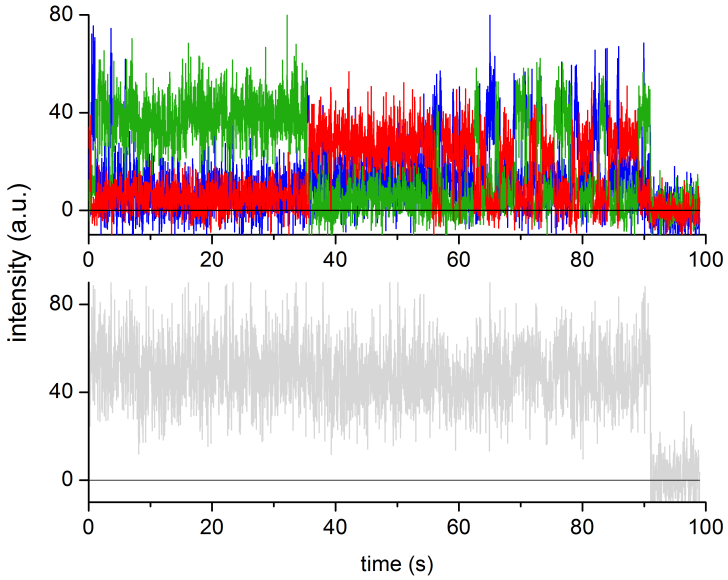

Figure 5.16: Slow-fast trace for 9 nt docking overlap.

Firstly, there were a few 'Slow-fast' traces, as the one depicted above in figure 5.16, where the transitions between the ON states for the dyes are slow initially and then suddenly become visibly faster after sometime, as seen at approx. 60 s in the exemplar trace. The total intensity (bottom panel, grey) remains constant throughout, ruling out any dye photophysics.

Secondly, there were traces that could be characterized by unique blue quenching behavior, seen as two bleaching steps for the blue dye. From the exemplar trace in figure 5.17, it is seen clearly that initially the blue dye shows three states and higher fluorescence intensity, which drops to two states and lower fluorescence intensity around 50 s into the measurement.

This allegedly could be pointing to the presence of two blue dyes, as indicated by the step in total intensity trace (bottom panel, grey) but it is highly unlikely as about 1% traces show this behavior, irrespective of docking overlap length. However, this could be attributed to the spectral red-shift of Alexa488, which has been addressed earlier for other blue dyes in the supplementary information in [Lee et al., 2010a]. This drop in intensity of the blue dye is mirrored in the drop of intensities of green and red dyes, which can be explained on the basis of FRET efficiency being a function of

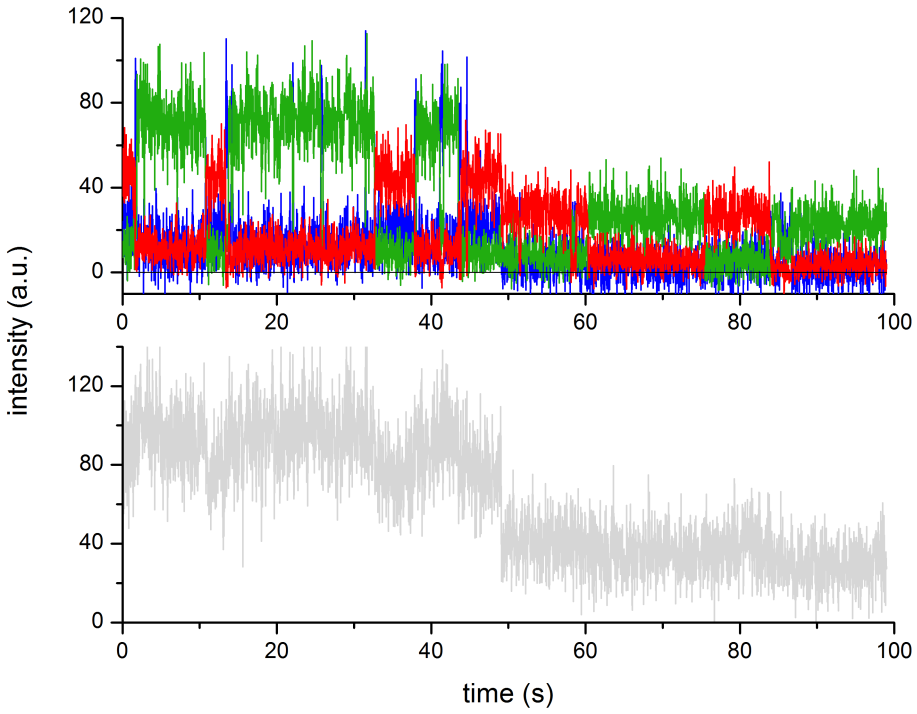

Figure 5.17: An exemplar trace (9 nt docking overlap) showing unique blue quenching behavior. The total intensity (lower panel, grey) drops around 50 s.

donor intensity. Therefore, the decrease in the intensity of the blue dye due to the spectral shift leads to decrease in FRET, thus causing the intensities of the acceptors to decrease.

5.2.5 Discussion

In this project, a DNA roboarm on a rectangular platform was designed and constructed in a one-pot reaction. The end of roboarm was labeled with Alexa488 (donor). Two staple strands on the platform with the same sequence, on the opposite sides of the roboarm hinge, were extended. These "docking" strands were terminally labeled with Atto565 and Atto647N (figure 5.7). Another staple strand, the "arm-dock", at the end of roboarm was extended to have few bases complementary to the docking strands on the platform. Hinged to the center of the platform, the roboarm is free to rotate by diffusion and can temporarily bind to the docking strand with the arm-dock. The diffusion of the roboarm was characterized by FRET between the donor dye on the roboarm and the acceptor dyes on the docking strands on the platform. Different roboarm samples were prepared with varying number of nucleotides on the arm-dock that are complementary to the docking strands. A few staples on the bottom side

of the platform were labeled with biotin to immobilize the sample on prism surface for multi-color TIRF experiments. Movies were recorded with only blue excitation or with millisecond alternating laser excitation (msALEX) mode. The measurements were made in buffer with GODCAT and activated Trolox. Furthermore, oxygen removal was aided by sealing the prism, which resulted in very long surviving traces with minimal blinking.

Since the two acceptors are far apart (∼50 nm), there is no FRET between them. This was verified using ALEX where only cross-talk was seen in the red channel upon green excitation (figure 5.11). Upon excitation with blue laser (491 nm), it was expected that the acceptor signal would be observed only when the donor and hence the roboarm was in close proximity. However, when the roboarm sample with no arm-dock was measured with blue laser, only a steady signal from the blue dye was observed (figure 5.9). The residual signal in the green and red channels can be accounted for by direct-excitation by the blue laser and bleed-through from the blue channel. This showed that, in the absence of arm-dock strand, the diffusion was at least faster than the time resolution of the camera (∼33 ms).

When roboarm samples with arm-dock strands with varying docking overlap lengths (8 nt, 9 nt and 10 nt) were measured, the traces showed clear indication of the movement of the roboarm from one docking strand to the other (figure 5.8). At any time point in the trace, only one dye among the three was in ON state (high fluorescence intensity), while the other two would be in OFF states (low fluorescence intensity). The only exception was upon blinking or bleaching of blue dye and then, of course, no signal was detected in any channel. This proved that the intensity seen in the acceptor channels was solely due to FRET from donor, which in turn could only be possible when donor is in proximity. This was verified by ALEX with all three lasers (figure 5.10, 5.11). Thus, any role of dye photophysics in the observed dynamics was ruled out.

Overall, the traces comprised of three states, either the green or red dye is ON, implying that the roboarm is temporarily docked at that particular docking site, or the blue dye is ON, implying the roboarm is in transit, diffusing somewhere away from either of the docking sites.

Subsequently, the intensity traces of each dye were independently subject to an HMM analysis. A cumulative "ON trace" (figure 5.13) was constructed for each molecule by combining the Viterbi paths of the respective dyes. The ON trace determined which dye was in the ON state for each time point. The dwell time information for the dyes was determined from this ON trace and average ON dwell times were calculated for each dye (table 5.1 and 5.2).

As seen in the top row of figure 5.13, the dwell times of the blue ON, green ON and red ON increased with increasing docking overlap length. This indicates that the overlap length affects both the binding to and the dissociation from the docking strands. This is different from the DNA-PAINT experiments where only the dissociation rate k_{off} is strongly dependent on the length of the duplex formed by the imager and docking strands [Jungmann et al., 2010]. However, this can be explained because in the roboarm experiments the binding rate cannot be considered as a pseudo first order rate, since neither the docking strand nor the arm-dock strands are in excess. The binding kinetics are of second-order. Assuming a second-order reaction, $rate = k_{on}[docking - strand][roboarm]$.

Since the concentrations of the roboarm and the docking strand can be assumed to be the same, the equation can be written as: $rate = k_{on}[roboarm]^2$. Assuming for the sake of convenience that the roboarm probes a hemispherical volume when diffusing, the concentration of the roboarm, calculated using the length of the roboarm (∼50 nm) as the radius of the hemisphere, is 6.34 μM. Assuming that this reduces the dependency to a first-order kinetics similar to DNA-PAINT experiments, using the relation $k_{binding} = k_{on}[roboarm] = 1/(\tau_{blueON})$, the k_{on} is $1.2 \times 10^6\ M^{-1}s^{-1}$ for 8 nt, $0.66 \times 10^6\ M^{-1}s^{-1}$ for 9 nt and $0.41 \times 10^6\ M^{-1}s^{-1}$ for 10 nt overlap length. For the roboarm, the dwell times for the blue dye are much shorter (<0.4 s) than for green and red dyes, which are similar (within error) and varying between 1-4 s for different overlap lengths. For a first-order reaction, the dissociation rate is given by $k_{off} = \frac{1}{\tau_{ON(G/R)}}$. k_{off} is approx. 1 s^{-1} for 8 nt, 0.4 s^{-1} for 9 nt and 0.3 s^{-1} for 10 nt. Comparing these value to the results from DNA-PAINT experiments (1.6 s^{-1} for 9 nt and 0.2 s^{-1} for 10 nt), it is obvious that the free diffusion of the roboarm is influencing the kinetics of dissociation from the docking strands.

The average dwell ON time for the blue dye is a good estimate of the time taken for the roboarm to diffuse from one side to the other. However, the blue ON state does not necessarily imply that the arm traverses to the other docking strand. From the traces, two types of blue blips were identified. The type I blue blip is where blue ON is observed between a green ON and red ON states, represented by the transition sequences RBG and GBR. This corresponds to the arm moving from one docking strand to the other. The type II blue blip is where the blue ON state is observed between only green ON states or only red ON states, represented by the transition sequences RBR and GBG. This corresponds to the arm returning to the same docking strand. These blips were separately analyzed and the individual dwell times were determined (figure 5.13, middle and bottom rows).

The average dwell times of the individual type I blue blips RBG and GBR and the cumulative dwell time of all type I blue blips are shown in the middle row of figure 5.13. The dwell times of type I blue blips is a real indicator of the diffusion times from one docking site to the other. Interestingly, the diffusion time increases with the length of the overlap of the docking strands. This could indicate that the larger the number of nucleotides required to hybridize, the higher are the chances of fraying, thus lower the probability of binding. Alternatively, it could also be that the arm-dock is transiently binding to other small staples on the DNA platform, so a longer arm-dock has a greater probability of unspecific binding.

The average dwell times for RBG are longer than GBR (for 9 and 10 nt), implying that, on average, the roboarm takes longer to traverse from the red dye labeled docking strand to the green dye labeled docking strand than the other way around. This can be explained by the asymmetric position of the strands ($\sim 165^o$) on the platform. Moreover, from the structure, it is expected that the resting position of the arm is asymmetric, which could be contributing to this bias.

The dwell time analysis of the type II blue blips (bottom row figure 5.13) shows that the type II blue blips are much shorter (by 50 %) compared to the type I blips. These blips represent the roboarm returning to the same docking strand instead of traversing across to the other one. Interestingly, the RBR blips have a similar dwell time across all overlap lengths, while the GBG blips show a strong correlation between dwell time and overlap length. This could also be attributed to the system bias, by the resting position of the arm being closer to the green dye. As suggested before, there could be also be interactions from other staples on the platform that could be interfering with the free movement of the roboarm.

The average probability of the roboarm to diffuse from one docking strand to the other compared to the roboarm returning to the same docking site was compared for different overlap lengths (figure 5.14). There is generally a higher probability of the roboarm to return to the same docking site, clearly seen for 8 nt overlap length where P(RBR) and P(GBG) are higher than P(RBG) and P(GBR) respectively. However, the probability changed with increasing overlap length and for 10 nt overlap length, all the transitions (RBG, RBR, GBR, GBG) became equally probable. Interestingly, the probability of a RBR blip remains unaffected by overlap length while the other blips show a trend of increase in probability.

Lastly, the probability of finding a RBG or GBR type I blue blips compared to the instances of missing it (only RG and GR) was calculated (figure 5.15)

for different overlap lengths. Interestingly, the increase in overlap lengths seem to be increasing the probability of detecting a type I blue blip. This could possibly be an effect of the overlap lengths increasing the dwell time of the blue blips overall and hence reducing the chance of missing it. Generally a blue blip would be missed only if its duration was shorter than the time resolution of the experiment, which is ∼33 ms in our case. This again could be the reason why the presence of an arm-dock is essential for seeing the acceptor signal, otherwise the roboarm seems to be diffusing very fast.
Thus, multi-color FRET on TIRFM was employed to characterize the diffusion of the DNA Roboarm. In future, roboarm with different arm lengths could be designed and similarly characterized.

5.3 DNA Threading

The results of this project will be published in the paper - Staple threading in a DNA origami tube. As discussed in the introduction, DNA origami involves folding a long single-stranded "scaffold" strand into a specific 2-D or 3-D shape by multiple shorter "staple" strands. This results in multiple DNA crossovers that form a meshwork-like structure. Thus, the commonly used single-layered 2-D origami designs have a critical design feature in the form of holes, which are adequately large (1.1 ± 0.3 nm deep and 7.2 ± 0.4 nm wide) to allow the passage of a sufficiently long ssDNA (cross-section is roughly 0.8 x 1 nm^2) [Wu et al., 2013]. Extended staple strands, protruding from one side of the origami, "thread" through the origami meshwork to the other side. This process has been shown to affect the binding of streptavidin to biotinylated staple strands [Wu et al., 2013, Aghebat Rafat et al., 2014, Mallik et al., 2015]. Furthermore, it is hypothesized that this "threading" phenomenon is responsible for incongruities observed in the DNA PAINT super-resolution technique; the docking sites become inaccessible for hybridization by threading through the mesh to the wrong side of the origami [Jungmann et al., 2010, Scheible et al., 2014].

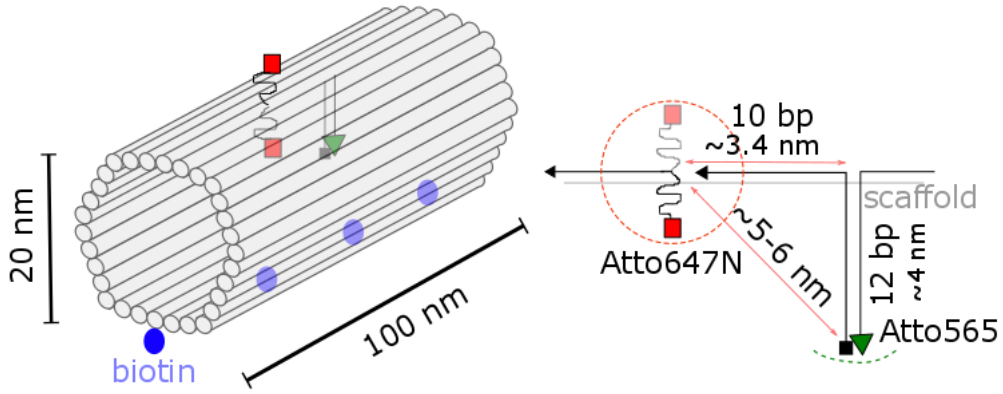

Figure 5.18: The hollow tubular DNA structures designed to study the threading process through smFRET. Left: The hollow tubular DNA origami structures with a hexagonal geometry. The structures are 20 nm in diameter and 100 nm in length. The bottom wall of the structure in labeled with four biotins (blue spheres), which enables immobilization on the surface for TIRF measurements. The arrangement of extended and labeled strands is shown on the top wall with the transparency implying the strands are inside the tube. The single stranded staple labeled with Atto647N is free to thread to outside of wall. Right: Two staple strands in the top wall of the structure are extended by 12 nt to point inside of the tube. One is labeled with Atto565 (donor; green triangle) and other with Atto647N (acceptor; red square). The donor labeled strand is hybridized with another staple strand similarly extended from the wall to inside, thus restricting movement of the donor. The acceptor labeled single strand is free to thread through the structure. The dashed circle/arc represent the possible movement of the respective labeled strands.

Since the functionality of the DNA origami system depends on the availability

of the staple strands, understanding how threading works is crucial. In this project, the kinetics of the threading phenomenon were probed using smFRET on a TIRFM. Threading of an extended staple strand labeled with Atto647N (acceptor) on a hollow tubular origami structure through the tube wall, was studied by monitoring the FRET efficiency from another extended staple strand labeled with Atto565 (donor), fixed on the inside of the structure (see figure 5.18). The designing, folding of all DNA origami structures, AFM and TEM imaging, and spFRET experiments were performed by Anna Kostina (AK Simmel). The HMM analysis of the intensity traces, dwell time analysis and determination of FRET distributions were my contributions to the project.

5.3.1 Experimental Procedures

Sample Preparation

A hollow tubular DNA structures of 100 nm length and 20 nm diameter were designed and folded. The folding process was a one pot reaction with the scaffold (genome of phage M13mp18) and staple strands mixed in a ratio of 1:5 . The mixture was cooled in steps of 0.3^oC in a buffer with 12.5 mM $MgCl_2$ from 70^oC to 20^oC. The structure and proper formation were confirmed by AFM and TEM imaging.

SmTIRF Measurements

To study threading phenomenon using smFRET with TIRF, two staple strands (10 nt or ∼3.4 nm apart) extended to point inside were labeled with Atto565 and Atto647N (figure 5.18, right panel). The donor-labeled staple was locked in position by hybridizing with a similarly extended neighboring staple of complementary sequence. The acceptor-labeled staple was single stranded and thus free to thread. These tubes were immobilized via streptavidin bound to the biotinylated BSA-coated glass slides by the four biotinylated staple strands on the opposite wall of the cylinder. The sample was excited with a 532 nm laser on an objective type TIRF setup (as described in [Scheible et al., 2014]) and the collected fluorescence was separated with a split viewer and imaged onto two parts of an EMCCD camera. The movies were recorded for 2000 frames with an exposure time of 200 ms. The imaging buffer contains GODCAT and Trolox for improving the signal and lengthening the lifetime of the dyes (as described in 5.2.1). Subsequently, the intensity traces of the donor and the acceptor molecules were extracted and analyzed.

The FRET efficiency and the corresponding distance distribution were obtained as follows. The FRET efficiency was calculated from the donor and

acceptor trace intensities, corrected for the gamma and beta appropriately (as described in chapter 4) using the equation:

$$E = \frac{I_A - \beta \cdot I_D}{I_A + \gamma \cdot I_D - \beta \cdot I_D} \tag{5.1}$$

The distances were calculated from FRET efficiency by using equation:

$$R = R_0(\frac{1}{E} - 1)^{1/6} \tag{5.2}$$

where R is the distance between the fluorophores, R_0 is the förster radius of the Atto565-Atto647N FRET pair, which is 6.8 nm and E is the calculated FRET efficiency. The FRET traces (traces showing FRET efficiency over time) were further treated to an HMM analysis and the dwell times of the various states were determined.

5.3.2 Kinetics of Threading of Single-stranded Staple Strand

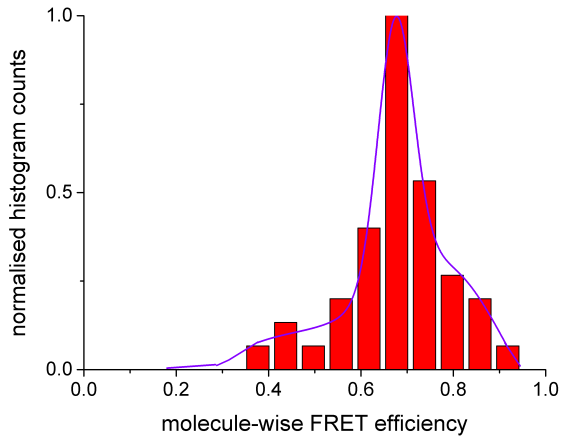

Figure 5.19: Molecule-wise FRET efficiency histogram of static traces obtained for the two-color construct. The blue line shows the fit with 3 Gaussians.

Among the traces selected for analysis, 60% showed no dynamics in the FRET traces. These traces were classified as static traces. The molecule-wise histogram of FRET efficiency for the static traces is shown in figure 5.19. The molecule-wise FRET histogram could be fit using three Gaussians, indicating presence of at least three populations of static molecules, with FRET efficiencies at 0.45 and 0.68 respectively and maybe a third population with E = 0.8.

On the other hand, the dynamic traces showed the FRET signal to be switching between 2 or 3 states with the transitions between the states being

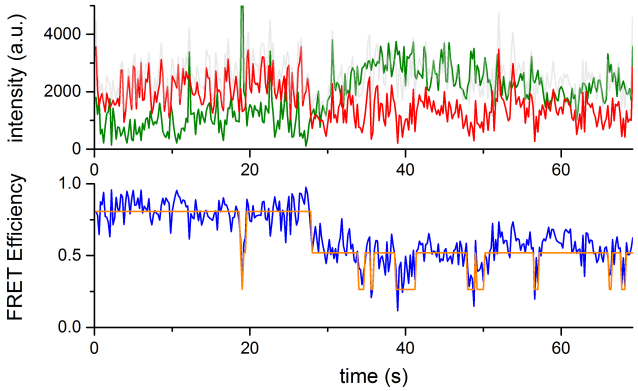

Figure 5.20: An exemplar dynamic trace of DNA tube construct with the single-stranded acceptor labeled staple strand as obtained from two-color spFRET experiments. The top panel shows the intensity traces of the donor Atto565 in green and the acceptor Atto647N in red and the total intensity in grey. The bottom panel shows the FRET efficiency trace calculated in blue, overlaid with the Viterbi path obtained from HMM analysis in orange.

characterized by slow and fast fluctuations (figure 5.20). As seen in the top panel of the figure, the donor and acceptor intensity traces seem to be very noisy. There is an anti-correlated change in the donor and acceptor intensities around 28 s. This change is more strongly visible in the FRET signal (blue) in bottom panel, as the molecule switches from a high FRET state (E = 0.8) to a middle FRET state (E = 0.5). In the example trace shown, the FRET signal is switching between 3 states. The fast fluctuations observed could be a result of the fast conformational changes of the single stranded staple or acceptor photophysics, which is difficult to confirm in the absence of direct excitation of acceptor. The slow transitions could be attributed to threading in and out of the tube wall.

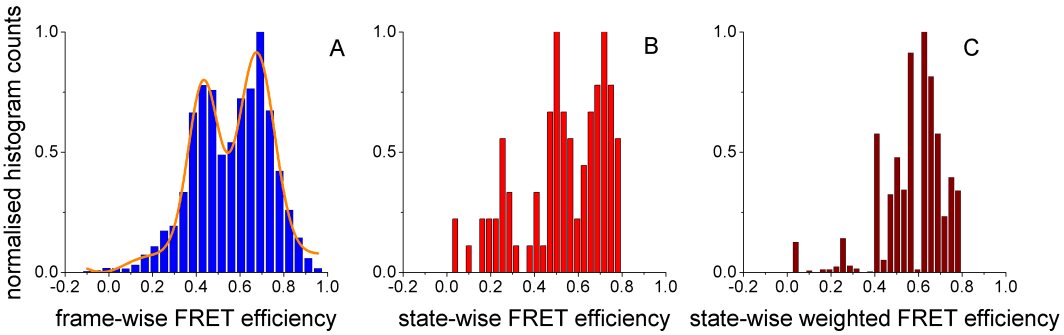

Figure 5.21: The FRET efficiency histograms from an HMM analysis of the dynamic traces obtained for the two-color construct are shown. A: The frame-wise FRET efficiency histogram. B: The state-wise FRET efficiency histogram. C: The state-wise weighted FRET efficiency histogram of dynamic traces obtained for the two-color construct.

The frame-wise FRET efficiency histogram of the dynamic traces and the corresponding fit is shown in figure 5.21 A, which indicates dynamics between at least two states, one at 0.7 and a second at 0.45 FRET efficiency.

To elaborately study these threading dynamics, the FRET signal of dynamic traces was analyzed using an HMM algorithm with 3 states (as indicated by the frame-wise FRET histogram) and the Viterbi path for each trace was determined. From the Viterbi path, a histogram of the mean FRET efficiency of each state for all dynamic traces was plotted as the state-wise FRET efficiency histogram (5.21 B). The state-wise weighted FRET efficiency histogram (5.21 C) was then plotted by weighing each state in state-wise histogram by duration of the state (number of frames). The state-wise histogram shows which FRET states are more often occupied. The state-wise weighted histogram elaborates on how long the states were occupied. In figure 5.21 B, it is seen that 3 states are most likely occupied by the traces, corresponding to E of 0.7, 0.5 and 0.25. While figure 5.21 C shows that molecules spend most time in high FRET state around 0.7 and very less time in the low FRET states around 0.25.

5.3.3 Effect of Different Interfering Strand Lengths on Threading

The above experiments were repeated where "interfering" strands of different lengths (5, 6, 7, 8, 9, 10 and 12 nt), which are complementary (starting from the tube wall) to the 12 nt acceptor labeled staple strand. Thus, these freely diffusing strands could hybridize with the acceptor strand and lock it in place, for the duration of binding. The "interfering" strands behave similar to the imager strands in DNA-PAINT experiments [Jungmann et al., 2010], but without the label, thus being governed by similar kinetics. Given the excess amount of the interfering strand present compared to the threading construct, the association reaction can be assumed to be a pseudo first-order reaction. Therefore, the binding rate of the interfering strands depends on their concentration; the pseudo first-order rate constant is given by: $k_{bind} = k_{on} \times c$ where c is the concentration. In these experiments, with an interfering strand concentration of $c = 100$ nM and assuming a typical on-rate k_{on} of approx. 10^6 $M^{-1}s^{-1}$ [Jungmann et al., 2010], the binding rate k_{bind} obtained is on the order of 0.1 s^{-1}.

This implies that there is 1 binding event every 10 s, which transpires to ∼40 binding events during every movie recorded of 2000 frames at exposure time of 200 ms. The unbinding rate k_{off} is dependent on the length and the sequence of the nucleotides involved in the duplex formation. From the DNA-PAINT experiments it is known that k_{off} is approx. 1.6 s^{-1} for an overlap of 9 nt and 0.2 s^{-1} for an overlap of 10 nt [Jungmann et al., 2010]. Therefore, with these acquisition settings, the dynamics of the duplexes of the threading strand and the interfering strand could be probed.

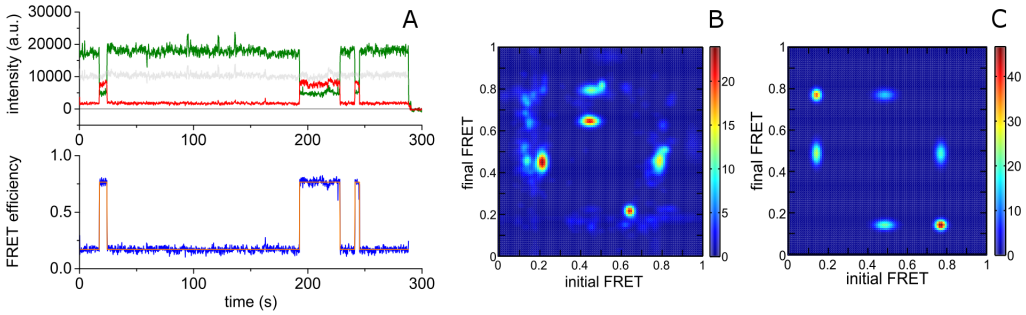

Figure 5.22: An exemplar trace upon addition of the 5 nt interfering strand to the two-color construct. A: The top panel shows the donor intensity (red), acceptor intensity (green) and β and γ corrected total intensity (grey). The bottom panel shows the FRET efficiency trace (blue) overlaid with the Viterbi path (orange) obtained from the HMM analysis. B: The transition density plot (TDP) of the traces upon addition of the 5 nt interfering strand, after fitting with a 3 state local HMM. C: The transition density plot (TDP) of the traces upon addition of the 5 nt interfering strand, after fitting with a 3 state global HMM.

On adding the interfering strand, the traces had remarkably less noise, with the intensities of the donor and acceptor switching in an anti-correlated manner, and the transition being clearly observed in the intensities of the donor and acceptor molecules as well as in the FRET efficiency. This can be seen in the exemplar trace in figure 5.22 of the two-color construct in the presence of 5 nt interfering strands. Similar dynamic traces were obtained for other interfering strand lengths. For the longer strand lengths of 10 and 12 nt, very few transitions were observed, with most of the traces showing only one transition during the entire experiment. The fraction of the dynamic traces observed for each interfering strand length is plotted in figure 5.23.

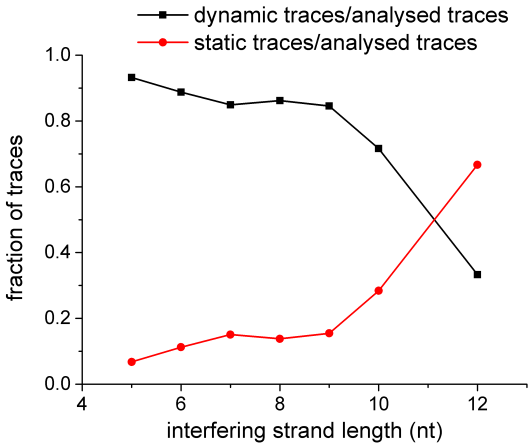

Figure 5.23: Fraction of the dynamic (black squares) and static (red circles) traces, plotted for different lengths of interfering strands. The data were provided by Anna Kostina.

Most dynamic traces showed transitions only between the high (around 0.75)

and the low FRET (around 0.2) state and transitions to the middle FRET state (0.4-0.5) were few as evident from the figure 5.22 C. This can be seen in the frame-wise FRET histograms in figure 5.24, where the different interfering strand lengths show three peaks. A better visualization is provided by the three-dimensional plot in figure 5.25. In the absence of the interfering strand (plotted in black), the peak at the low FRET is hardly existent. Adding the 5 nt strand dramatically shifts the FRET from high and middle to the low FRET state. For different interfering strand lengths, the low FRET peaks are of similar FRET efficiency values (around E = 0.01). The middle FRET peaks vary a bit in their values from 0.35-0.5, while the high FRET peaks are all centered around 0.75. Only the 8 nt histogram (in dark green) stands out a bit, with a prominent middle FRET peak. The negative values of FRET are a result of the β correction factor being very high (0.1). In the absence of ALEX, this makes it difficult to distinguish between low FRET state and the case when the acceptor has bleached and there is no FRET.

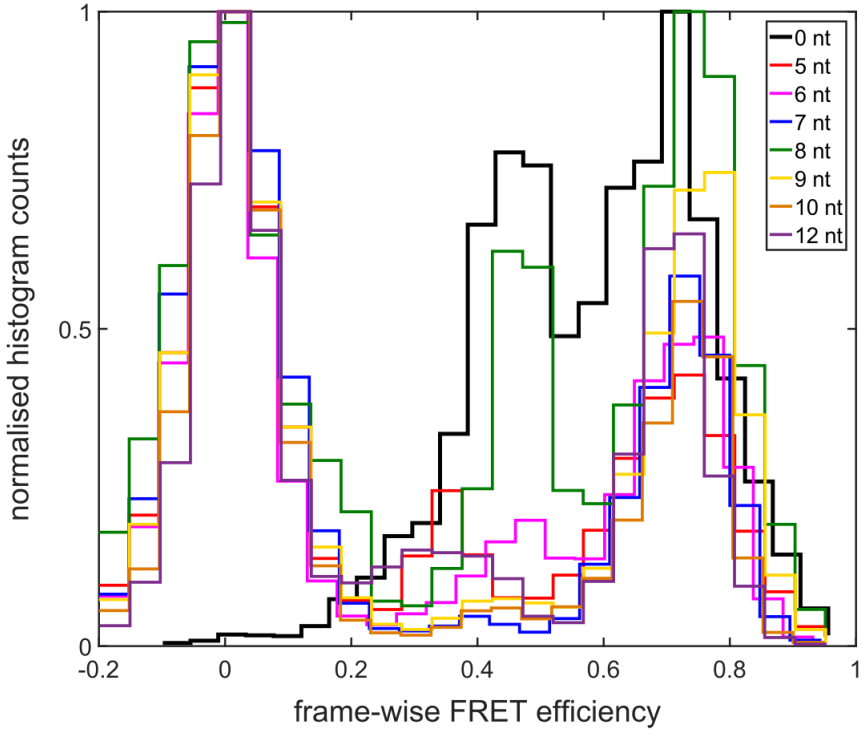

Figure 5.24: Frame-wise FRET efficiency histograms of dynamic traces obtained for the two-color construct in absence and presence of interfering strands of various lengths as a two-dimensional plot.

To understand the influence of the length of the interfering strand on the switching dynamics of the FRET states, the dynamic traces were subject to an HMM analysis followed by determination of the dwell times, reported in figure 5.26 A.

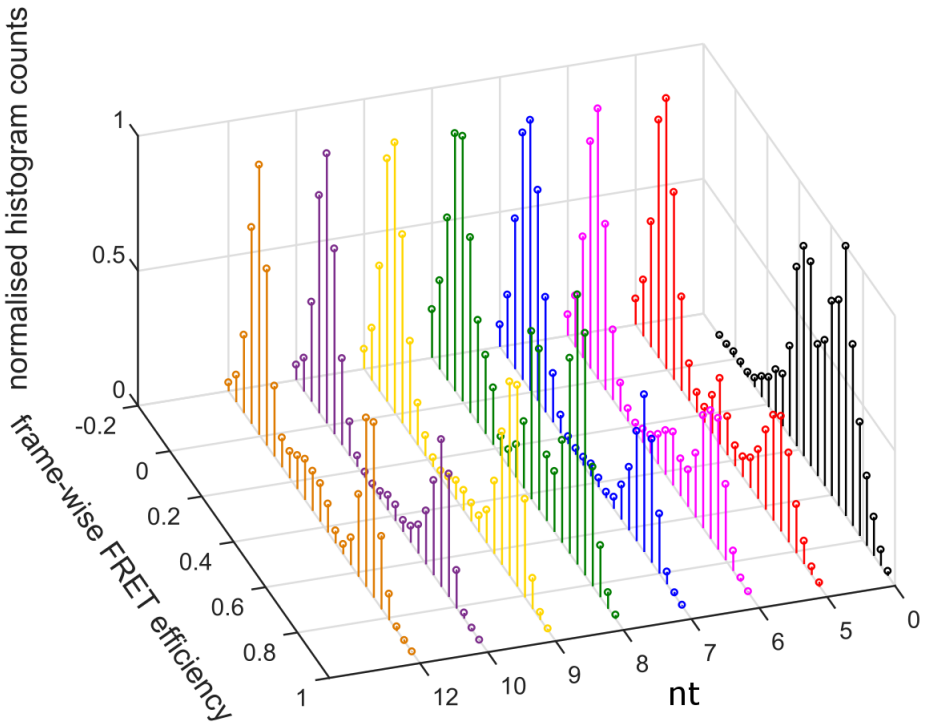

Figure 5.25: Frame-wise FRET efficiency histograms of dynamic traces obtained for the two-color construct in absence and presence of interfering strands of various lengths as a 3-dimensional plot.

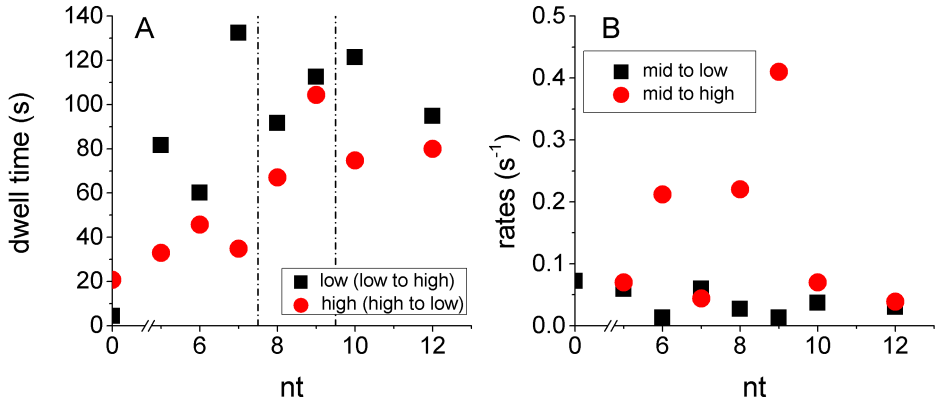

Figure 5.26: Dwell time analysis of the dynamic traces. A: The dwell times for high FRET state (red circles) and low FRET state (black squares) in absence (0 nt) and in presence of interfering strands of different lengths. The dotted lines are for guiding the different behaviors as discussed in the text. B: The rates (s^{-1}) from middle to the high FRET state (red circles) and middle to the low FRET state (black squares).

In the absence of the interfering strand, the threading strand spends more

time in the high FRET state. On adding interfering strands of any length, this is reversed. The time spent in the high FRET state for 0, 5, 6 and 7 nt are comparable. For the 8 and 9 nt strands, both dwell times increase considerably. For 10 and 12 nt strands, the dwell times are biased by the limited time resolution as most traces show only a single transition. This is supported by the huge increase in percentage of static traces seen for these lengths in figure 5.23. The dotted lines in 5.26 A separate these three categories of the threading strand lengths.

5.3.4 Discussion

The process of threading was observed through dynamic switching of FRET states of a two - color construct. The construct consists of a DNA origami hollow tube where two staple strands about 3.4 nm (∼10 bp) apart were extended from the tube wall to point inside. One strand was hybridized with another neighboring strand similarly extended from tube wall and labeled with Atto565. The second strand remained single-stranded and was labeled with Atto647N.

The two-color construct was also measured in the presence of interfering strands of different lengths. In the absence of direct excitation of the acceptor by ALEX, it is difficult to rule out switching of the FRET signal as a result of acceptor blinking. This could be of significance especially since the low FRET states observed in the traces, after correction for gamma and beta, center around 0.01. However, this could be easily verified by either performing ALEX experiments or repeating experiments with a different acceptor dye like Alexa647. Also, the middle FRET state observed in many traces could be attributed to dye photophysics or presence of multiple dyes. Since the size of the hollow tube is around 100 nm, the possibility of multiple origamis in the same fluorescence spot, though rare, cannot be ruled out. This also can be verified through ALEX experiment, as one can differentiate between multiple labels through the stoichiometry. Furthermore, the flexibility of the fixed strand labeled with donor needs to be verified. This can be accomplished by placing the acceptor on the wall of the DNA tube.

Assuming that the donor strand is fixed to one side and the absence of any dye photophysics, the data can be interpreted as follows. On measuring this two-color construct on TIRF, about 40 % traces exhibit dynamical switching of FRET states. On fitting the frame-wise FRET histogram of the dynamic traces (figure 5.21), two Gaussian peaks were deciphered, with E = 0.7 and E = 0.4. Using equation 5.3.1, these correspond to distances of 5.9 nm and 7.3 nm respectively between the donor-acceptor pair. Considering the lengths of the donor and acceptor strands on the DNA tube, these distances

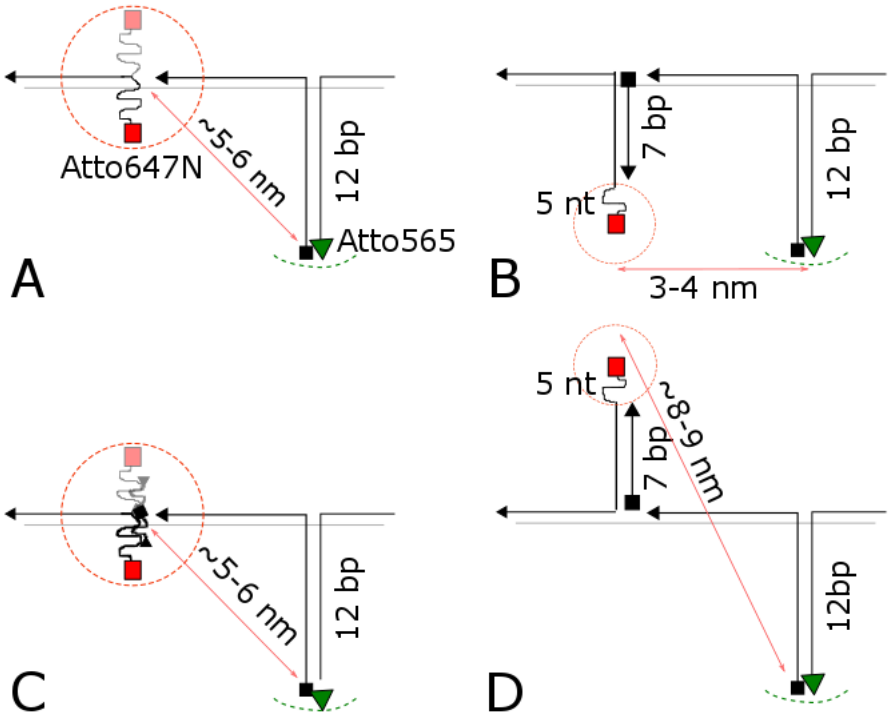

Figure 5.27: The various stages of threading with the distance between the fluorophores are illustrated. A: The single-stranded threading strand is near to the wall of the tube or stuck at the hole, with the acceptor (Atto647N) about 5-6 nm from the donor. B: The interfering strand of 7 nt forms a duplex with the threading strand and is present inside the tube. The distance between the fluorophores is expected to be about 4 nm. C: The interfering strand of 7 nt forms a duplex with the threading strand and is in the process of threading to outside, near to the hole. The distance between the fluorophores is about 5-6 nm. D: The interfering strand of 7 nt forms a duplex with the threading strand outside of the DNA tube. The distance between the fluorophores is about 8-9 nm

correspond to the acceptor strand being either in a random coil on the inner side of the wall or just threading to the other side but still staying close to the wall as shown in figure 5.27 A. The static trace populations show FRET values around 0.8, 0.68 and 0.45 (see figure 5.19). These correspond to distances of 5.4 nm, 6 nm and 7 nm, respectively. These values for the static traces suggest that the strand could be stuck at the hole (as shown in figure 5.27 A), probably due to the positively charged Atto647N dye interacting with the negatively charged DNA tube.

Upon addition of interfering strands of different lengths, the FRET states shift dramatically to low values. For different interfering strand lengths, three prominent FRET states of approx. 0.01, 0.4 and 0.7 FRET efficiency were discerned (figure 5.24 A). These correspond to distances of 14 nm, 7.3 nm and 5.9 nm, respectively. The distance greater than 9 nm indicates that the acceptor strand has threaded to other side, as shown in figure

5.27 D. The high FRET state of 0.7 and the middle FRET state of 0.4 correspond to the strands being on the inside of the tube (5.27 B) or very close to the hole (5.27 C). Since the middle FRET state is distinct, it indicates that the strand could be stuck at the hole, as suggested previously.

Considering the distances obtained, the dwell times can be seen in new light. The dwell time of high FRET state would be representative of time spent by the acceptor strand inside the tube and the low FRET state dwell time would correspond to time spent by the strand outside the tube. This implies that the dwell time distribution shown in figure 5.26 indicates that the strands prefer staying outside the tube, once they have threaded. The reason behind this could be the higher entropy outside, with more space for the strand to move around. Interestingly, the dwell time outside increases with strand length. This could imply that, the interfering strand binds to the threading strand after it has threaded to the other side, thus the duplex is stable outside and does not prefer to thread back to the inside. However, this would not effect the duplexes formed by the shorter interfering strands (5-7 nt) as they are quite unstable and would readily unbind. The rate for going from high to low FRET state for shorter interfering strands could then be taken as an estimate of the threading time, which is 37.8 $\pm$ 3.2 s.

However, this does not take the middle state into consideration. Alternatively, it can be assumed that the middle state corresponding to the strand being near the hole is an intermediate state. Given that most traces do not show middle state, it could be that this state is very short lived. This is seen from the dwell times in middle states, which are much lower compared to the dwell times of low and high states. This points to the middle state representing a thermodynamic transient high energy state, corresponding to the strand being stuck at the hole or to the origami. In this case, the rate of threading would be appropriately represented by the rate from middle to low FRET, which as seen in figure 5.26 B remains almost constant around 0.05 s^{-1}, which corresponds to a threading time of 20 s. This would imply that the duplex nature of the threading strand, which is important for the case of longer interfering strands, would hardly effect the rate of threading. However, the rate of threading back to inside (figure 5.26 B, rate of mid to high, red squares) seems to vary a lot with the interfering strand length, the rates being remarkably faster for the 6, 8 and 9 nt interfering strands.

Furthermore, the effect of increase of stability of duplexes of threading and interfering strand with increase in length of interfering strand is reflected in the increase in fraction of static traces (figure 5.23). Also, for static traces, FRET states correspond to the threading strand staying inside the tube or near the hole. In contrast, the fraction of dynamic traces decreases with

increasing interfering strand length, with the threading strand spending more time outside the tube.

The 12 bp duplexes formed by binding of the 12 nt interfering strand to the threading strand are too stable to allow any unbinding of the strand, thus should not be dissociating in the timescale of the experiment. Nevertheless, dynamic traces are observed for 12 nt interfering strand experiments, with switching between high and low FRET states, indicating threading to the other side. This could be possible either if these traces observed actually have no interfering strand hybridized (hence similar to no interfering strand case) or if the proximity of the origami wall to the duplex decreases the stability of the duplex, perhaps due to collisions with the wall, causing the duplex to dissociate. In both these cases, these results have important ramifications for the threading process. Further analysis through molecular dynamics simulations on the timescales of threading events ($\sim$20-30 s) would be helpful. Also, the experiments suggested could make the interpretation more clearer.

6. Single-molecule Studies of Na^+-K^+ ATPase

This project is collaboration with Prof. Promod Pratap of the University of North Carolina at Greensboro (UNCG), USA. The protein was extracted from nasal glands of juvenile ducks and separated into membrane fragments by the collaborator. Labeling in membrane-fragments and subsequent single-molecule measurements on TIRF and MFD setup and, analysis of TIRF data were performed by me. Solution measurements were analyzed by Anders Barth. The results of this project would be published in the paper - Single-molecule studies on the Sodium pump: Effects of ATP.

For the sake of convenience, the Na^+-K^+ ATPase has sometimes been referred to as "pump", "protein" or "enzyme".

6.1 Introduction

Na^+-K^+ adenosine triphosphatase (ATPase) is a transmembrane ion pump found in plasma membrane of all animal cells. It utilizes the energy obtained from ATP hydrolysis to maintain the electrochemical gradient of sodium and potassium ions across the cell membrane which is, crucial for nerve impulse transmission. It is also pivotal for muscle contraction, nutrient reabsorption, regulation of cellular volume and intracellular pH balance [Morth et al., 2007].

Na^+-K^+ ATPase was the first ion pump to be discovered [Skou, 1957] and belongs to the family of P-type ATPases [Palmgren and Nissen, 2011]; fittingly named due to presence of a phosphorylated intermediate in their reaction cycle. The phosphorylated intermediate is formed by transfer of the $\gamma-$phosphate from ATP to the pump, under physiological conditions. The intermediate, possessing higher energy, relaxes through conformational changes that induce release of ions to the other side of membrane. This facilitates transport of ions against the electrochemical gradient [Clarke, 2009]. Other important members of this family include the extensively studied sarcoplasmic reticulum Ca^{2+} ATPase, which regulates skeletal muscle relaxation [Toyoshima and Mizutani, 2004], the $H^+ - K^+$ ATPase responsible for stomach acidification [Shin et al., 2009], and the $H+$ ATPase [Palmgren, 2001], which is a Na^+-K^+ ATPase analogue in the plant cells.

The crystal structure of the Na^+-K^+ ATPase (figure 6.1) shows that the pump is a heterodimer comprising an $\alpha-$subunit (113 kDa) and a highly glycosylated $\beta-$subunit (45-58 kDa)[Kanai et al., 2013]. The $\alpha-$subunit

consists of 10 transmembrane (TM) segments (numbered from M1 to M10 starting at the amino end), and a large intracellular loop between M4 and M5. This loop has three major domains: the *nucleotide binding* (N) domain, the *actuator* (A) domain, and the *phosphorylation* (P) domain. The TM region also houses the ion-binding sites [Nyblom et al., 2013].

The $\beta-$subunit has a single transmembrane segment and is thought to facilitate correct insertion of the $\alpha-$subunit into the membrane [Noguchi et al., 1990, Geering, 1991]. It also plays a peripheral role in K^+ transport [Jaisser et al., 1994].

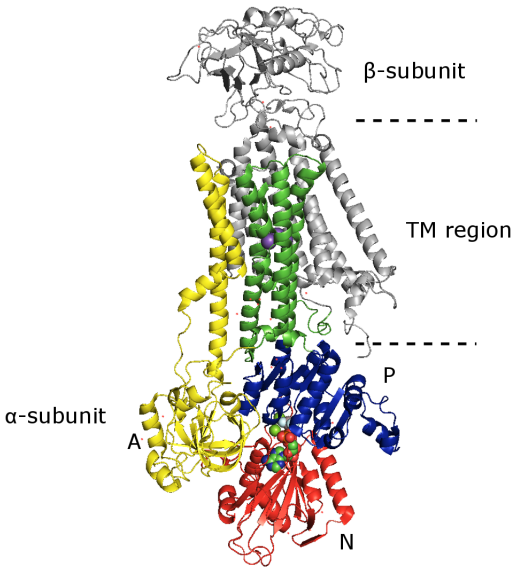

Figure 6.1: The crystal structure of Na^+-K^+ ATPase in the Na^+ bound form clearly showing the $\alpha-$subunit with the *Actuator* (A) domain in yellow, the *Phosphorylation* domain (P) in blue and *Nucleotide* (N) domain in red. The 3 Na^+ ions bound are shown in the transmembrane (TM) region as purple spheres. The $\beta-$subunit with one TM helix is shown in grey. The TM helices with the ion binding sites are shown in green. The image has been prepared from the PDB structure (3WGU) of the Na^+ bound pump preceding the E1P state with ADP and aluminum fluoride (colored spheres near P domain), from Kanai et al. [2013].

Since 2007, various crystal structures of the pump in different conformations have been ascertained [Morth et al., 2007, Ogawa et al., 2009, Kanai et al., 2013, Nyblom et al., 2013]. All of them point to the existence of an ion transport pathway along the axis of a single monomer ($\alpha\beta$ protomer) of the pump. Also, the phosphorylation capacity measurements [Martin and Sachs, 1999] and other kinetics studies [Jørgensen and Andersen, 1988, Ward and Cavieres, 1993, Brotherus et al., 1983] demonstrate that the $\alpha\beta$ protomer is capable of functioning as a single catalytic unit. Thus, there is no structural or functional basis for the pump to aggregate into dimers or multimers, for ion translocation.

Yet intriguingly, the pump functions as oligomers under certain conditions [Taniguchi et al., 2001, Clarke and Kane, 2007, Clarke et al., 2007]. Native membrane environment shows clear evidence of protein aggregation [Blanco et al., 1994, Donnet et al., 2001, Palladino et al., 2003], forming dimers or higher oligomers. Under saturating conditions, no significant difference in the molecular activity of the protomers and the oligomers was found, indicating that the maximal turnover is unaffected by protein oligomerization [Martin et al., 2000]. However, under non-saturating conditions (relevant under starvation conditions), the protein-protein interactions could affect the enzyme kinetics of the pump; an aspect shown for the Ca^{2+} ATPase by Mahaney et al. [2004]. It is these protein-protein interactions that we wish to study. However, it is first essential to understand the pump cycle and the various conformations involved, in detail.

6.2 Pump Cycle

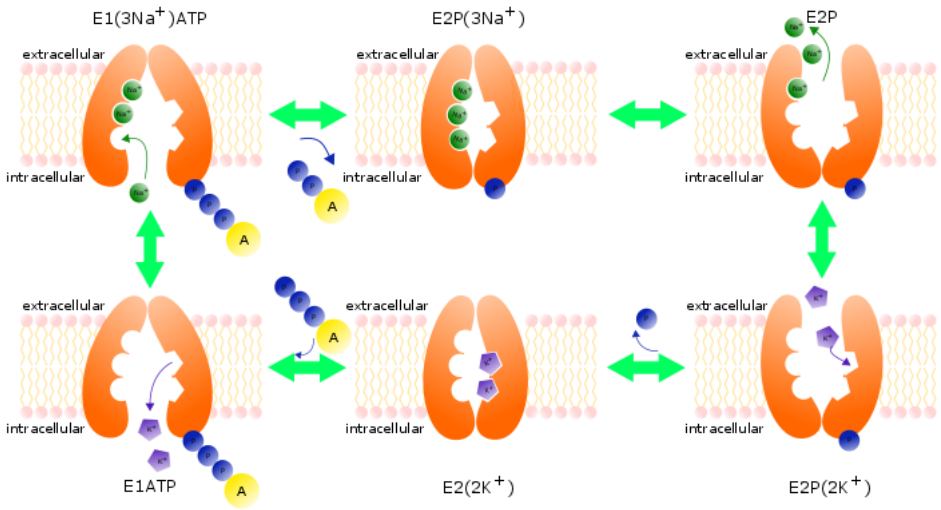

Figure 6.2: The Albers-Post reaction cycle of Na^+-K^+ ATPase involving the E1 and E2 conformations and six intermediates. The reaction sequence is described by equation 6.2. The image has been modified from Scheiner-Bobis [2002].

The pump cycle has been extensively studied and the most widely accepted model is the *Albers-Post model* [Kaplan, 2002], based on the seminal works of Albers [Albers, 1967] and Post [Post et al., 1972]. Also called the E1-E2 model, it describes the pump functioning by assuming two conformations of the pump:

1. The E1 state
In the E1 state, the pump's ion binding sites in TM region are accessible from the cytoplasmic side of the plasma membrane. The pump in E1 state has higher affinity for Na^+ ions (K_d = 0.19-0.26 μM) and ATP (K_d = 0.1-0.2 μM)

2. The E2 state
In the E2 state, the ion binding sites are accessible from the extracellular side and the pump has higher affinity to K^+ ions.

Both E1 and E2 exist in the ATP bound, phosphorylated and non-phosphorylated configurations. In the E1 state, the pump binds $3Na^+$ ions and ATP with high affinity. Upon subsequent phosphorylation of an aspartate residue[1] in the presence of Mg^{2+} ions, the pump occludes Na^+ ions and undergoes a conformational change into the E2 state. This is followed by release of Na^+ ions into the extracellular environment against the ion gradient. Thus, ATP hydrolysis is coupled to ion translocation. The E2 state binds $2K^+$ ions that induce rapid dephosphorylation of the pump which is followed by occlusion of K^+ ions. On further binding of ATP from cytoplasmic side, the pump changes its conformation to E1 and releases the $2K^+$ ions into the cell. Therefore, the sequence around the cycle with all the reaction intermediates can be summarized as:

$$\text{E1}(3Na^+)\text{ATP} \rightarrow \text{E2P}(3Na^+) \rightarrow \text{E2P} \rightarrow \text{E2P}(2K^+) \rightarrow \text{E2}(2K^+) \rightarrow \text{E1}(2K^+) \tag{6.1}$$

Remarkably, the step involving release of K^+ ions into the cytosol was the first reaction discovered that demonstrated the allosteric effect of ATP. Being the rate determining step (slowest reaction) of the entire reaction cycle, this reaction: $E2(2K^+) + 3Na^+ \rightarrow E1(3Na^+) + 2K^+$ is of great physiological significance. It occurs very slowly (on the timescale of seconds) in the absence of ATP but is dramatically accelerated in the presence of ATP (k = 65-90 s^{-1} [Clarke, 2009]). Moreover, this acceleration is exclusive to ATP binding to the E2 state, as it happens even in the absence of Mg^{2+} ions [Pratap et al., 1996], a pre-requisite for any phosphorylation reaction with ATP. Consequently, ATP not only provides energy for ion pumping, but also plays a pivotal role in stimulating the pump and optimizing its pumping rate. There is evidence of similar allosteric roles of ATP in mechanisms of other P-type ATPases [Cable and Briggs, 1988, Reenstra et al., 1988].

[1]The particular phosphorylation site (Asp) depends on the source of the pump. For the duck protein used here, the phosphorylation site is Asp^{376}.

6.3 Understanding the Allosteric Behavior

In spite of all these observations of the allosteric roles of ATP (see previous paragraph), there has been no definitive mechanism proposed that could account for this behavior. There could be two possible ways for this to occur: 1. The existence of two different ATP binding sites on the pump; one catalytic site and one regulatory site or, 2. The presence of only a single ATP site that changes its property from allosteric to catalytic depending on the where the pump stands in its pumping cycle.

The former can be mostly invalidated based on the crystal structure evidence available for sarcoplasmic reticulum $Ca^{2+}-$ATPase from skeletal muscle (SERCA) [Sørensen et al., 2004] and the $H^{+}-$ATPase of plants [Palmgren, 2001], which show no evidence whatsoever for multiple nucleotide binding sites. Since the overall structure of the α subunit of theNa^{+}-K^{+} ATPase is very similar to that of SERCA [Kanai et al., 2013], it is likely that the sodium pump too has a single ATP site per α subunit. The results from activity measurements on theNa^{+}-K^{+} ATPase from the duck nasal gland by Martin and Sachs [2000] are also consistent with this assumption.

Furthermore, the crystal structure evidence shows that nucleotide binding site and phosphate binding site in P-type ATPases are not mutually exclusive [Sørensen et al., 2004]. Additionally, there is kinetic evidence of ATP and phosphate binding simultaneously to the Na^{+}-K^{+} ATPase [Askari and Huang, 1982]. As a result, a single ATP binding site on pump monomer is strongly favored.

Recently, kinetic experiments by Clarke and Kane [2007] have authenticated this assumption by demonstrating that the high affinity ATP site can act as purely catalytic or purely regulatory, depending on the reaction being studied (phosphorylation or the E1 to E2 conformational transition). Thus, there is neither a functional necessity nor structural evidence for separate sites. However, this still does not eliminate the possibility of different protomers of an oligomer being in different states of pump cycle, consequently exhibiting different ATP affinities.

In the biochemical context, the discussion about 'allosteric behavior' is incomplete without considering the classic example of Hemoglobin, an allosteric globular protein with four subunits, that exhibits positive cooperativity in binding with oxygen, as evident from the sigmoidal binding curve. On the other hand, Myoglobin, a monomer, displays no such behavior as substantiated by its hyperbolic binding curve. Thus, the allosteric binding of Hemoglobin is ascribed to protein-protein interactions between its neighboring subunits, with these interactions increasing the affinity of

the neighbors to oxygen when oxygen binds to the first subunit. Likewise, the allosteric effects of ATP observed in the Na^+-K^+ ATPase could be associated with protein-protein interactions. It must be pointed out here that it is well established that a monomer ($\alpha\beta$ protomer) is the minimal functional unit of the pump (with the α subunit as its catalytic center). Also, it is neither structurally nor functionally necessary for the pump to exist as dimers or higher oligomers to pump ions. However, in the native membrane environment, the pump aggregates into oligomers (as discussed in 6.1), possibly fostering protein-protein interactions. Although the original Albers-Post scheme involves only monomers, there have been modifications to account for dimers and multimers, like the two-gear bicyclic model for dimers by Clarke and Kane [2007]. Previously, Pratap et al. [2009] measured

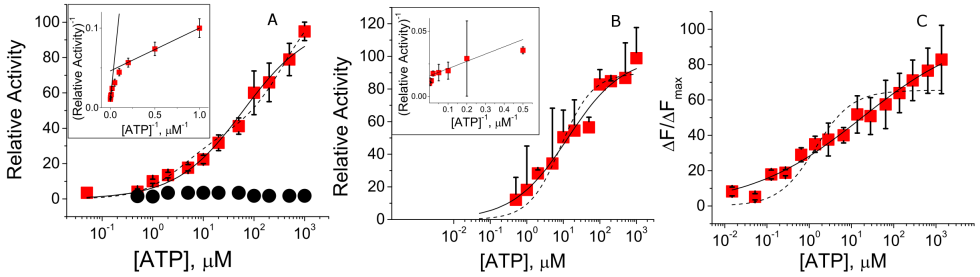

Figure 6.3: Previous bulk measurements published in [Pratap et al., 2009]. A. The Na^+-K^+ ATPase activity as a function of ATP concentration in Na^+-K^+ buffer (30 mM HEPES, pH 7.5, 1 mM EDTA, 4 mM $MgCl_2$, 130 mM NaCl, 20 mM KCl) in the absence (squares) or presence (circles) of 1 mM ouabain. Both activities were normalized to the maximal activity in the absence of ouabain (57 (μmol Pi) $(mgprotein)^{-1}$ min^{-1}). The solid line represents a fit to the Hill equation ($n = 0.66 \pm 0.04$, $K_{0.5} = 61 \pm 7$ μM), while the dashed line is a fit to the sum of two rectangular hyperbolae ($K_1 = 4.05 \pm 0.01$ μM, $K_2 = 413 \pm 1$ μM; the high-affinity component is 31% of the maximum and the low-affinity component is 69% of the maximum). The inset is a Lineweaver-Burke plot of the data with the lines representing the two components of the rectangular hyperbolae. B. Activity of the Na^+-K^+ ATPase isolated from duck supraorbital salt glands as a function of ATP concentration in Na^+ buffer (30 mM HEPES, pH 7.5, 1 mM EDTA, 4 mM $MgCl_2$, 150 mM NaCl). Activity was normalized to the maximal activity (3 (μmol Pi)$(mg\ protein)^{-1}$ min^{-1}). The solid line represents a fit to the Hill equation (n_H =0.58 $\pm$ 0.05, $K_{0.5}$ = 13 $\pm$ 2 μM), while the dashed line is a fit to a rectangular hyperbola ($K_{0.5}$ =8.5 $\pm$ 0.1 μM). The inset is a Lineweaver-Burke plot of the data, where the solid line represents the fit to a rectangular hyperbola. C. Titration of Cy3-ATPase with ATP plot, in buffer containing 30 mM HEPES, pH 7.5, 1 mM EDTA, and 150 mM NaCl. Enzyme was incubated with Cy3-maleimide for 2 h. The data are expressed as a percentage of ΔF_{max} (% Δ F_{max} was 16%). The normalized data were fitted with the Hill equation (solid line, $K_{0.5}$ = 20$\pm$2 μM, and $n = 0.33\pm0.02$) and with a rectangular hyperbola (dashed line, Δ F_{max} = 65 $\pm$ 4 % and $K_{0.5} = 1.4 \pm 0.5$ μM). The quality of fit, expressed as the normalized χ^2 is significantly better for the Hill equation than for the rectangular hyperbola. Data points represent an average of three experiments.

the kinetics of ATPase activity of the Na^+-K^+ ATPase isolated from duck supraorbital salt glands in buffer without and with K^+ ions, although Na^+ and Mg^{2+} ions were present in both cases. The change in enzyme

activity as a function of ATP concentration could be fit with the sum of two rectangular hyperbolae - representing the two affinities for ATP (dotted line, panels A and B figure 6.3). However, the data had a better fit to the Hill equation (solid line, panels A and B), with $n < 1$ in both cases, indicating negative cooperativity. They also titrated Cy3 labeled enzyme against unlabeled ATP in buffer with only Na^+ ions (pump in the E1 state) and here too the dependence of fluorescence quenching on ATP concentration showed negative cooperativity (figure 6.3, solid line, panel C, n = 0.33). These experiments linked the negative cooperativity observed in the enzyme activity to the negative cooperativity in ligand binding. This could probably be explained as the protein in their preparation was multimeric, with the protein-protein interactions between monomers affecting the ATP binding kinetics.

This project was started to study whether the pump exhibited cooperativity under monomeric conditions. The main goal is to determine whether the ATP binding kinetics is solely due to interaction between protomers by determining the intrinsic properties of monomers.

6.4 Experimental Procedures

In the project, Cy3 labeled ATPase was investigated at the single-molecule level. For this purpose, the pump was isolated, purified, labeled with Cy3 maleimide, solubilized in detergent, to run over a size-exclusion column to separate monomers from multimers. Finally, the monomers were encapsulated into lipid vesicles and probed with different fluorescence spectroscopy techniques, in presence of varying ATP concentrations. Each of these steps are elaborated upon further in the subsections.

Protein Isolation

The pump was isolated from the supraorbital salt glands of juvenile Pekin ducks by Promod Pratap according to the method described earlier [Pratap et al., 2003, Martin and Sachs, 1999]. Concisely, one day old ducklings were purchased from Metzer Farms (Gonzales, CA, USA). Ducklings (12-13 per batch) were grown initially on feed designed for chicks (Southern States Cooperative, Inc., Richmond, VA) and tap water. At the end of the 14 days, all water that the ducklings come in contact with was replaced by 1% (w/v) saline. After a further 10 days of salt adaptation, the ducks were euthanized, the supraorbital salt glands were removed and flash-frozen with liquid nitrogen.

The salt glands were then minced and homogenized in homogenization

buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 250 mM sucrose) with a Polytron PT10/35 homogenizer (Kinematica, Cincinnati, OH). The homogenate was filtered through three layers of cheesecloth. The filtrate was centrifuged for 15 min at 7,100 rpm in a Beckman clinical centrifuge (Model J2-21) with a JA20 rotor (Beckman-Coulter USA, Fullerton, CA). The supernatant was saved and the pellet was resuspended in homogenization buffer and centrifuged as described above. The supernatants from both these spins were combined and centrifuged at 17,000 rpm for 1.5 h with a Beckman Ti 50.2 rotor. The resulting microsomal pellet was resuspended in homogenization buffer and stored at -80^oC. The protein concentration in the microsomal suspension was determined using a modified Lowry protein assay (Sigma Aldrich, St. Louis, MO, USA).

Furthermore, the pump was purified from microsomes by partial extraction with SDS [Pratap et al., 2003, Martin and Sachs, 1999]. Microsomes (1 mg/ml) were incubated in buffer (14 mM Tris-HCl, pH 7.5, 1.7 mM EDTA, 2.2 mM Na-ATP) at room temperature. A stock solution of SDS (2 mg/ml) was added slowly (over a period of 20 min) using a peristaltic pump (Mini variable pump, Fisher Scientific, Pittsburg, PA) to a final concentration of 0.55 mg/l. The SDS protein suspension was incubated at room temperature for 40 min. The suspension was then layered on a sucrose step gradient (% w/v sucrose of 10:15:29.4:55) in Beckman Ultraclear tubes. The tubes were centrifuged at 25,000 rpm for 14–16 h. The band at the boundary between 29.4% and 55% was aspirated with a syringe. This solution was diluted with buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA) and subsequently centrifuged at 50,000 rpm for 1 h using a Beckman Ti 50.2 rotor. The pellet, consisting of pure pump protein in membrane fragments, was resuspended in homogenization buffer and stored at -80^oC. Protein was assayed using the modified Lowry assay as described previously. ATPase activity was assayed using a coupled-enzyme assay [Pratap et al., 2003, Banik and Roy, 1990]. Typically, 3-6 mg of purified pump was obtained in each preparation, and the activity of the enzyme used in these experiments was > 55 (μmole Pi)$(mgprotein)^{-1}(min)^{-1}$.

Enzyme Labeling with Cy3-Maleimide: Maximal Labeling (for MALDI)

Enzyme was maximally labeled with Cy3-maleimide (GE Healthcare Life Sciences, GE Healthcare Europe GmbH, Freiburg, Germany) as described earlier in Pratap et al. [2009]. 1 mg of enzyme (in membrane fragments) in 30 mM HEPES (pH 7.5), 130 mM NaCl, 20 mM KCl, and 1 mM EDTA, was incubated with 200 nmoles of Cy3-maleimide at room temperature for two hours. The labeled membrane fragments were separated from free dye by centrifugation at 150,000 x g in a Beckman Airfuge for 15 min. The pellet

was resuspended in buffer and centrifuged again. After the second spin, the pellet was resuspended to a concentration of 1 mg/ml in buffer (30 mM HEPES, 150 mM NaCl, 1 mM EDTA, pH 7.5). This maximal labeled protein sample was used to identify the cysteines on the protein that were labeled by Cy3-maleimide using Matrix-assisted laser desorption/ionization (MALDI).

Enzyme Labeling with Cy3-Maleimide: Minimal Labeling

Labeling positions that are easily accessible were investigated by minimally labeling the enzyme with Cy3-maleimide as follows: flakes of Cy3-maleimide were dissolved in 10 μl DMSO. 1 μl of this stock dye solution was dissolved in 9 μl buffer (Buffer I: 30 mM HEPES, 1 mM EDTA, 150 mM NaCl, pH 7.5). The concentration of this dye solution was determined from its absorbance as measured on a Nanodrop 1000 (Thermo Fisher Scientific, Wilmington, DE, USA), using a molar extinction coefficient of 150,000 $M^{-1}cm^{-1}$.

Protein was added to the stock dye solution at a molar ratio of dye:protein of 2:1; molar concentration of the Na^+-K^+ ATPase was calculated from the protein concentration in mg/ml and a molecular weight of 158 kD [Martin and Sachs, 2000]. The dye-protein suspension was incubated at room temperature for 30 min. and then centrifuged at 85,000 rpm for 15 min. in a table-top ultracentrifuge (Beckman Max-E with a TLA 100.2 rotor, Beckman-Coulter, Inc., USA in the group of Prof. Ladurner/Dr. Mokranjac, Physiological Chemistry, LMU Munich and in the group of Prof. Beckmann, Gene Center, LMU, Munich). The pellet was resuspended in Buffer II (Buffer I containing 100 $\mu g/ml$ BSA) and centrifuged a second time to remove residual free dye. Finally, the pellet was resuspended to a concentration of 1 mg/ml in Buffer II.

6.5 Deciphering the Labeling Position using MALDI-MS

For protein labeling, Cy3-maleimide was used, which binds to the Cysteine residues in the protein. In the duck pump, none of the cysteines in the $\beta-$subunit are available for maleimide binding; 6 out of 7 cysteines are involved in disulfide bridges while the 7th one, Cys46 is contained within the TM segment. This leaves only the cysteines in $\alpha-$subunit available for labeling with Cy3 maleimide. However, from the 23 cysteines in the $\alpha-$subunit, cysteines located in transmembrane segments within the protein cannot be labeled by the method of labeling in our study. These are known to be Cys93, Cys145, Cys809, Cys937, and Cys990. Four other cysteine residues are located very close to these transmembrane bound segments,

and are most likely not labeled . These are Cys111, Cys343, Cys918 and Cys971. These 9 cysteines were not analyzed for maleimide binding using the mass spectrometry data obtained.

To determine the labeling position of the Cy3-maleimide, Matrix-assisted Laser Desorption Ionization (MALDI) was used as a mass spectrometry (MS) imaging technique. MALDI-MS was performed commercially by David Kirchner at the DHMRI, North Carolina Research Campus (NC, USA). The MALDI spectra were recorded for the unlabeled protein, minimally labeled protein (used for our measurements) and saturation-labeled protein (used in Pratap et al. [2009]). In the unlabeled protein, the cysteines were alkylated before digestion; alkylation prevents the formation of disulfide bonds, and improves the digestion and the detection of the fragments. However, labeled proteins were not alkylated because alkylation could possibly cleave Cy3 off the protein. Trypsin digestion was performed on all samples prior to MALDI analysis by the addition of Rapigest for delipidation and solubilization.

For minimally labeled sample, one cysteine containing tryptic peptide LIIVEGCQR [699-707] in the duck pump $\alpha-$subunit sequence was shown likely to be labeled by Cy3-maleimide. The predicted unlabeled, unmodified, non-alkylated mass of this peptide is 1030.57 m/z. The predicted Cy3-maleimide labeled mass is 1609.87 m/z. In the MALDI-MS spectrum of trypsin digested minimally labeled protein, a peak appeared with a monoisotopic mass of 1608.76 m/z. The loss of H (-1.0073 m/z) or H and an electron (-1.0078 m/z) or the possibility of the conversion of Glu to Gln during labeling reaction would perhaps account for this difference in mass for the measured LIIVEGCOR peptide. This cysteine (Cys705) falls in the large intracellular loop between membrane-spanning regions M4 and M5, and it contains the ATP binding site and the protein phosphorylation site (P domain). Hence, it is likely to be sensitive to ATP binding.

Saturation labeling offered evidence for three possible labeling sites: Cys249, Cys374, and Cys705. Other possibilities for the fourth labeling site include the peptide CIELCCGSVK [459-468]. However, further experiments would be necessary to positively determine this. All the possible labeling sites are marked in the figure 6.4. It is important to note that all these labeling sites are around the ATP binding site (figure 6.4, shown in cyan spheres), so they are likely to be sensitive to ATP binding.

ÄKTA Micro Calibration for Pump Monomer Separation

The Superose 6 size-exclusion column in the FPLC ÄKTA micro (GE Healthcare Life Sciences) was initially calibrated using a set of calibration

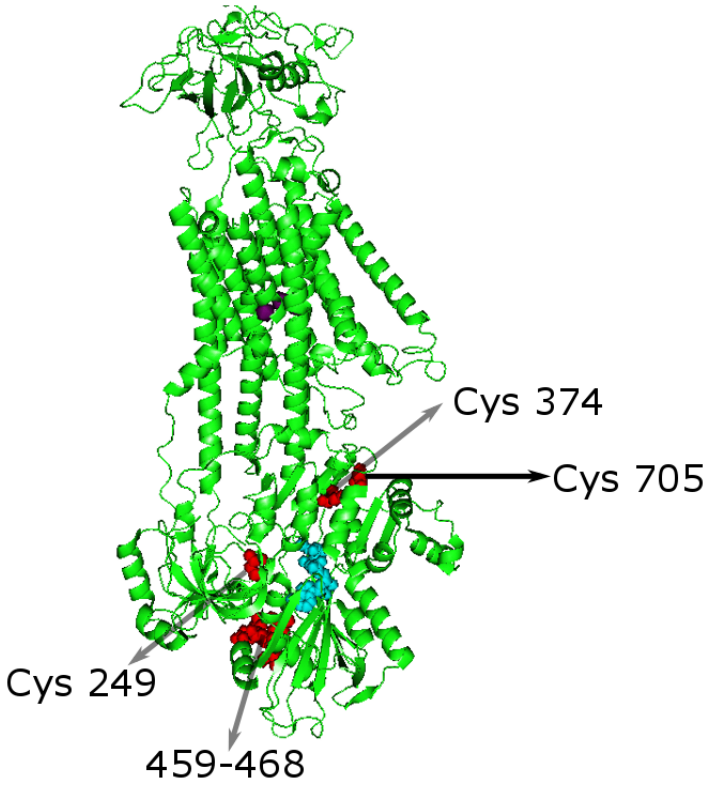

Figure 6.4: The crystal structure of the Na^+-K^+ ATPase from PDB (3WGU) with the Na^+ ions in purple and the ADP, aluminium fluoride and Mg^{2+} in cyan spheres with the possible labeling sites evidenced by MALDI-MS in red spheres.

standards (Gel Filtration HMW Calibration Kit, GE Healthcare Life Sciences), which notably contains Aldolase (158 kD) and Thyroglobulin (669 kD). The calibration standard solution was prepared in Buffer III, was run at a flow rate of 0.02 ml/min and collected in a serpentine fashion in a 96 well-plate. The monomer fraction should be obtained in the same well of a 96 well-plate as the Aldolase fraction when keeping the flow and buffer conditions same, since the mol. wt. of the pump monomer is the same as Aldolase (158 kD). This was identified as well B-9 from the sharp Aldolase peak in figure 6.5. Similarly, the tetramer fraction would coincide with the Thyroglobulin peak, procured from well B-12 as seen in figure 6.5.

Enzyme Solubilization and Reconstitution of Protein Monomers

Protein was solubilized with the nonionic detergent $C_{12}E_8$ (octaethylene glycol monododecyl ether, Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) as described in Brotherus et al. [1983]. Briefly, labeled enzyme at 1 mg/ml in Buffer II was mixed with 2 mg/ml $C_{12}E_8$ in Buffer I for a final ratio of enzyme to $C_{12}E_8$ of $1:1(w/w)$. The enzyme was incubated at room

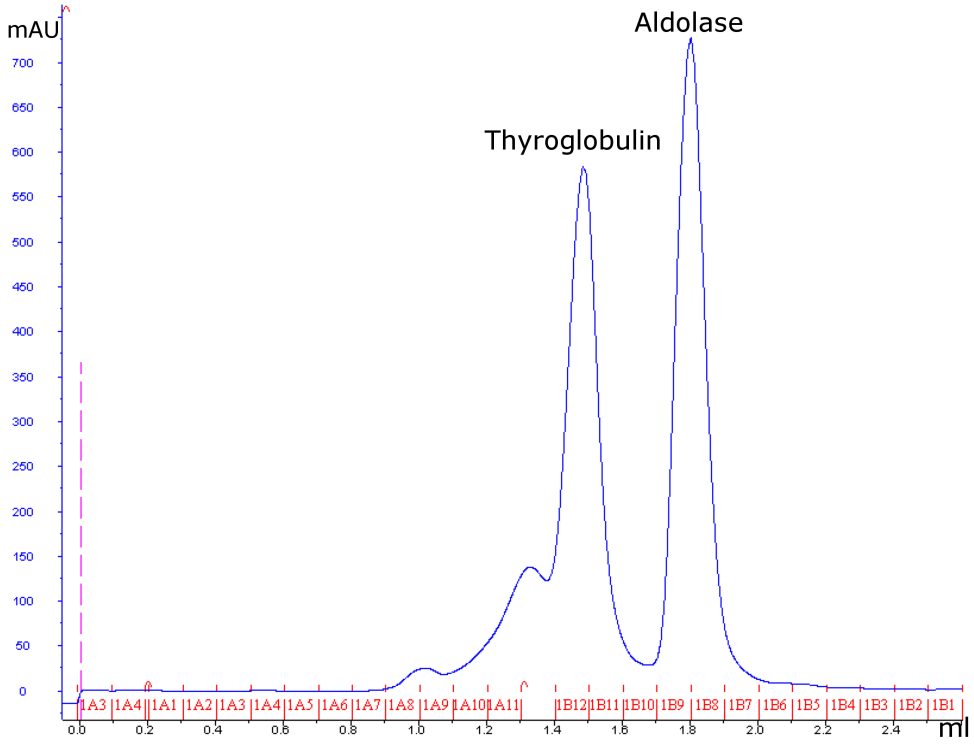

Figure 6.5: The ÄKTA micro calibration curve with the Aldolase and Thyroglobulin standards. The blue curve is the protein absorbance at 280 nm in mAU. On the x-axis is the volume collected (in ml) and the wells are demarcated in red. The magenta dotted line depicts the injection of the sample from the sample loop into the column. The standards were prepared in buffer III, ran at 0.02 ml/min and collected in a 96-well plate in a serpentine fashion. The Aldolase peak appears in well B-9 and the major Thyroglobulin peak appears in well B-12.

temperature for 30 min., with frequent mixing by gentle pipetting. Excess unsolubilized membrane was removed by ultracentrifugation (Beckman Max-E) at 85,000 rpm for 15 min. with a TLA 100.2 rotor. The supernatant was filtered with an Amicon Ultra 0.5 ml 100k centrifugal filter (Merck Chemicals GmbH, Darmstadt, Germany) by spinning in the centrifuge at 14000 rpm for 10 min. and resuspended in Buffer III (Buffer I with 50 $\mu g/ml$ BSA) to a final volume of 120 μl.

This sample was then injected into a Superose 6 size-exclusion column in the FPLC ÄKTA micro which had been equilibrated (2 column volumes = 2 x 2.4 ml) with Buffer III. A typical FPLC curve for the solubilized pump is shown in figure 6.6. The 100 μl fraction (well B-9) that represented the protomeric ATPase (see 6.5) was isolated and diluted to a volume of 500 μl in Buffer II. This protein solution was incubated in a pre-prepared test tube containing liposomes consisting of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC, Avanti Polar Lipids Inc., Alabaster, USA) and 1,2-dipalmitodyl-sn-glycero-3-phosphoethanolamine-N-(cap-biotinyl)

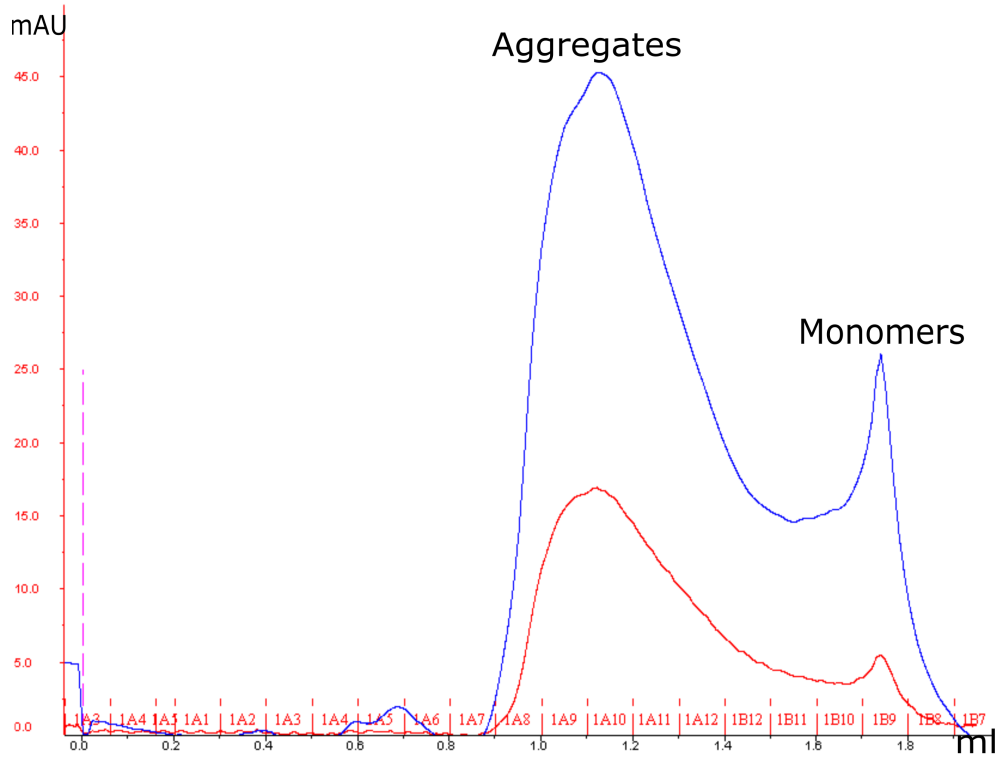

Figure 6.6: The ÄKTA micro curve of solubilized Cy3-labeled pump showing the Monomers peak in well B-9 and a broad Aggregates peak that includes the Tetramers expected to appear in well B-12. The blue curve is the protein absorbance at 280 nm in mAU and the orange curve is the absorbance at 550 nm (absorption maximum for Cy3) in mAU. On the x-axis is the volume collected (in ml) and the wells are demarcated in red. The magenta dotted line depicts sample injection from the sample loop into the column.

(sodium salt) (Biotinylated PE, Avanti Polar Lipids) for 30 min. at 4^oC with occasional vortexing, resulting in the formation of multilamellar vesicles (MLVs). This vesicle suspension was extruded 33 times through a polycarbonate membrane containing 0.2 μm pores using an Avanti mini extruder (Avanti Polar Lipids). This procedure resulted in unilamellar vesicles with a homogenous size distribution of 200 nm.

SmTIRF Experimental Conditions

Single-molecule TIRF microscopy experiments were performed on a home-built prism-type TIRF setup (as described in chapter 3). Since the enzyme was labeled with Cy3, the measurements were carried out by exciting with the 532 nm laser [2] and detected on the EMCCD camera in the two-color detection at the right exit of the microscope as described in Heiss [2012]. Since fluorescence from a single fluorophore was being detected, the size of the aperture was adjusted so that the detection channel covered the

[2] A 561 nm laser, a better choice for Cy3 excitation, was added much later to the setup.

entire EMCCD chip.

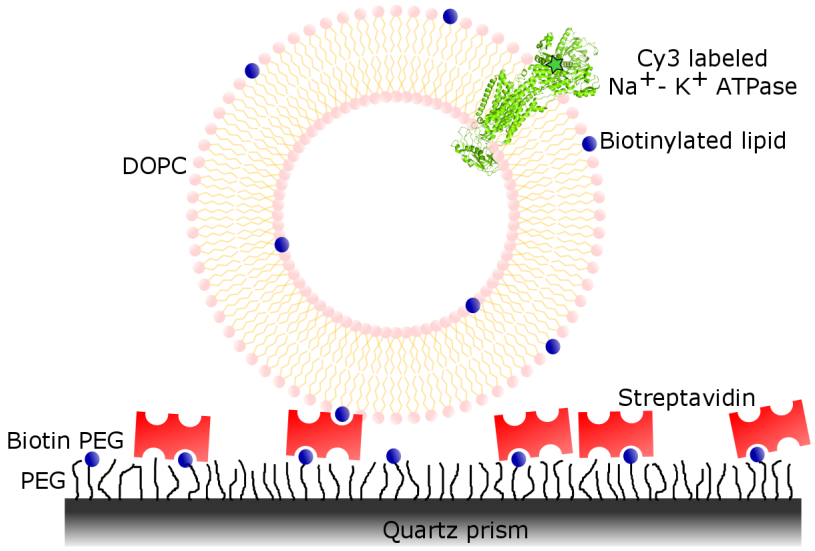

Figure 6.7: Cy3-labeled Na^+-K^+ ATPase inserted into the lipid membrane of the liposome vesicle immobilized on prism surface for single-molecule TIRF experiments. The vesicles are immobilized using biotin-streptavidin-biotin linkage between biotinylated lipids in the vesicle and biotinylated PEG on the prism surface. Image not to scale.

The sample was immobilized on a custom-made quartz prism covered with $PEG/3 - 4\%Biotin - PEG$ layer through a biotin-streptavidin-biotin linkage to the biotinylated lipids in the vesicles. It was then excited by the evanescent field of the total internally reflected beam at the prism – buffer interface and the fluorescence was collected through an objective (Plan Apo 60x WI, NA 1.2, Nikon, Germany) and imaged onto an EMCCD camera (iXon+, Andor Technology, Ireland).

Movies were recorded at a frame rate of 30 Hz (exposure time = 30 ms) for 1000 frames. In each movie, the fluorescent spots were identified and the intensity vs time traces were extracted using a home-built MATLAB program (see chapter 4). Each fluorescent spot was assumed to represent a vesicle with labeled proteins (within the field of view). Subsequently, each time trace was visually analyzed to identify and separate those with single bleaching step. These traces represent a single fluorophore indicating the presence of one singly labeled protein in the vesicle. The experiment was repeated in the presence of different concentrations of ATP (5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2 μM, 5 μM, 10 μM, 20 μM, 50 μM, 100 μM, 200 μM, 500 μM, 1 mM and 2 mM) and 15-20 videos were recorded for each condition. All the experiments were repeated three times. For each experimental condition, approx. 12,000-14,000 traces were obtained, of which $\sim$ 10% exhibited single bleaching steps. Single-molecule traces were analyzed

with an Hidden Markov Model (HMM, described in chapter 4) to determine the most-likely state sequence for each molecule (Viterbi path). The number of states in the sample was determined by over-fitting with higher number of states followed by a cluster analysis [McKinney et al., 2006]. Two states were found sufficient for the pump data and increasing the number of states did not improve the likelihood of the model.

MFD-PIE Experimental Conditions

Solution measurements of the pump monomer sample in the absence of ATP and presence of 1 mM ATP were performed on a two-color MFD-PIE setup (as described in chapter 2) in a 8 well Lab-Tek I slide (VWR) which was pre-incubated with 1 mg/ml BSA for 10 min. and washed with Buffer III. The sample was excited with 532 nm laser (frequency doubled 1064 nm solid state laser, PicoTA 530, Toptica, München, Germany) at 75 μW for 10 min. and Burst Analysis (as described in chapter 2) was performed on the photons detected with the two green APDs to retrieve information about the Fluorescence Lifetime and Anisotropy properties. Since the pump was single labeled, PIE was not used.

6.6 Effect of ATP on Pump Monomer Intensity Traces

The fluorescence of Cy3-labeled pump monomers in vesicles was measured on TIRFM in the absence and the presence of varying concentrations of ATP from 5 nM to 2 mM as described in 6.5 and the fluorescence intensity traces were obtained. Each of the fluorescence intensity traces was examined manually and the molecules that photo-bleached in a single step were identified and further subject to an HMM analysis, which yielded the Viterbi path.
For each of the experimental conditions, traces with two types of Viterbi paths were observed as seen in the exemplar traces in figure 6.8:
1. Non-Dynamic traces (figure 6.8 A) - traces where HMM found only one state. These were the traces with no dynamics.
2. Dynamic traces (figure 6.8 B) - traces which showed transitions between an unquenched (high-fluorescent) state and a partially quenched (low-fluorescence) state.

These transitions for the dynamic traces were plotted in transition density plots (TDP in figure 6.8 C) and clusters of transitions (approx. 2000-5000 transitions per cluster) were selected and subject to a dwell time analysis, resulting in dwell time histograms. The TDP graphically represents transitions between the two states (unquenched and partially

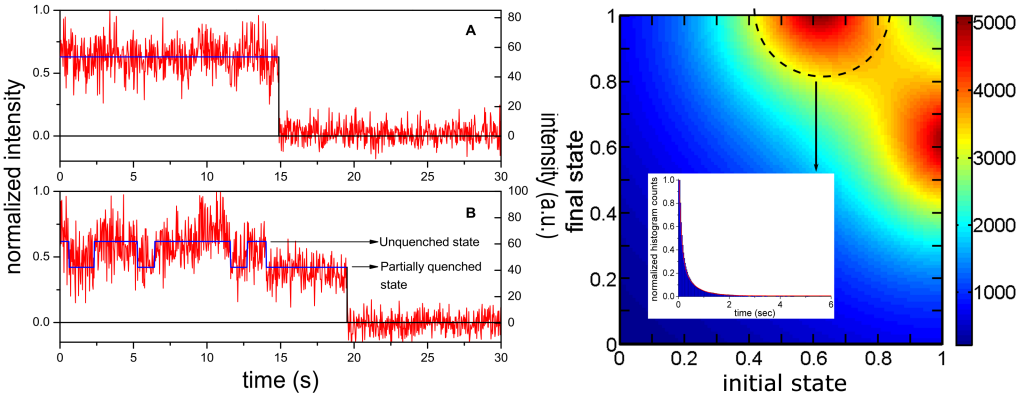

Figure 6.8: Analysis of the different types of single-step bleaching traces obtained in TIRFM measurements of Cy3 labeled pump monomers in vesicles. A. Non-dynamic trace with the intensity (red) plotted over time (s). The single state found by HMM analysis is shown in blue. B. Dynamic trace with intensity (red) plotted over time (s). Two states, corresponding to an unquenched and partially quenched fluorophore, are identified by the local HMM analysis as seen in the Viterbi path (blue). C. Transition density plot (TDP) of the transitions in dynamic traces for the measurement in the presence of 100 nM ATP. The color map represents the density of transitions. The cluster (dashed semi-circle) representing transitions from the partially quenched state to the unquenched state is selected and the normalized dwell time histogram (blue bars) is plotted in inset and fitted with mono-exponential decay curve (red).

quenched), weighted by the no. of transitions. These histograms were fit with a mono-exponential decay (red curve) and the dwell times and the corresponding rates were obtained.

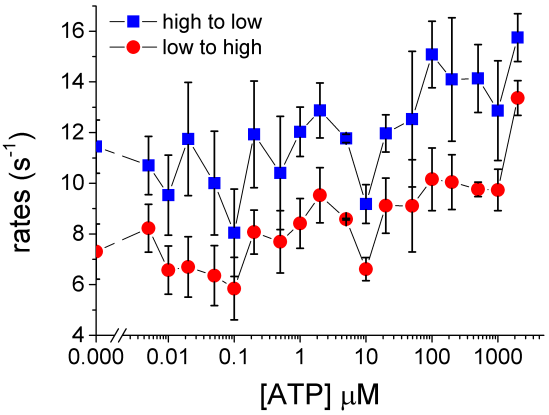

Figure 6.9: Dependence of the average rates for transitions on ATP concentration. Rates obtained for the high-to-low (blue squares) and low-to-high (red squares) transitions are plotted vs ATP concentration.

Figure 6.9 shows the average rates obtained from a dwell time analysis of the clusters of transitions that go from high-to-low fluorescence state

and vice versa as a function of ATP concentration. The error bars are the standard error of mean from three experiments. It is evident from the graph that rates of transitions are increasing overall with increase in ATP concentration. Furthermore, it was clearly seen that, irrespective of ATP concentration, the pump transited faster from the high state to low state. This implies that the pump prefers to stay longer in the low fluorescent state, regardless of whether ATP is present.

For a better understanding of the effect of ATP on quenching dynamics, the equilibrium constants (K) were plotted against ATP concentration (figure 6.10). The equilibrium constants were calculated as the ratio of the forward and the backward rates, which in this case was the ratio of the rate of the high-to-low to the rate of the low-to-high transitions. The equilibrium constants were mostly unaffected by the increase in the concentration of ATP. This indicates that ATP affects the forward and backward rates by the same amount.

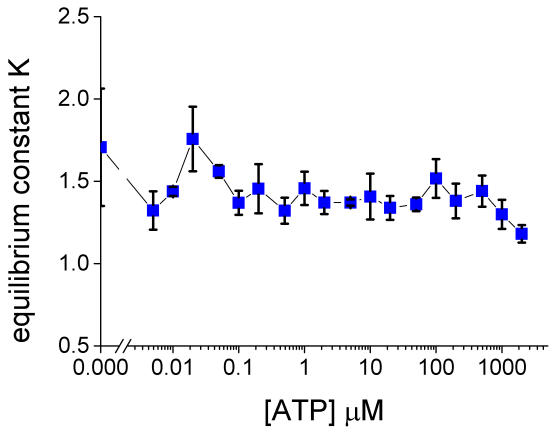

Figure 6.10: Equilibrium constants vs ATP concentration. Each point represents the average of three experiments with the standard error of mean plotted as the error bar.

The influence of ATP concentration on the dynamic behavior of the traces is presented in figure 6.11. The fraction of dynamic traces increased with ATP concentration, signifying that ATP binding to the ATPase resulted in an increase in the rates of transition probed by the fluorophore. The fit with the Hill equation (thick line) yielded a $K_{0.5}$ of 64 μM and n of 0.4. These results reveal that this response of the protein monomer to ATP demonstrates negative cooperativity; the kinetics being similar to those observed for oligomeric sample previously in bulk measurements [Pratap et al., 2009]. Similar to the bulk measurements, the Hill equation gave a better fit here than rectangular hyperbola (dashed line). This suggests that the monomers might be similar to multimers.
Lastly, the single-molecule traces were subjected to Pseudo-Ensemble analysis as follows. The average intensity (average of intensities at all time

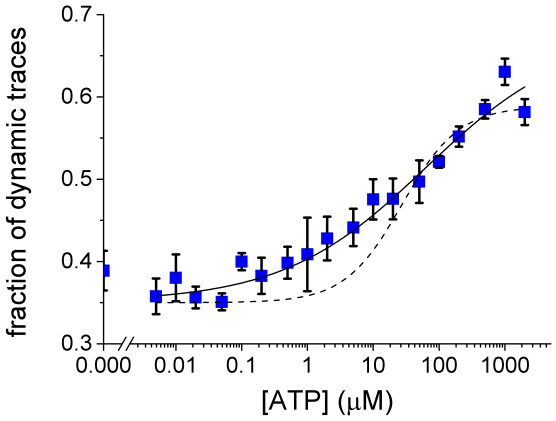

Figure 6.11: The fraction of dynamic traces recorded is plotted as a function of ATP concentration. The dotted line shows fit to a rectangular hyperbola and the solid line shows fit to the Hill equation. Each point represents the average of three experiments with the standard error of mean plotted as the error bar.

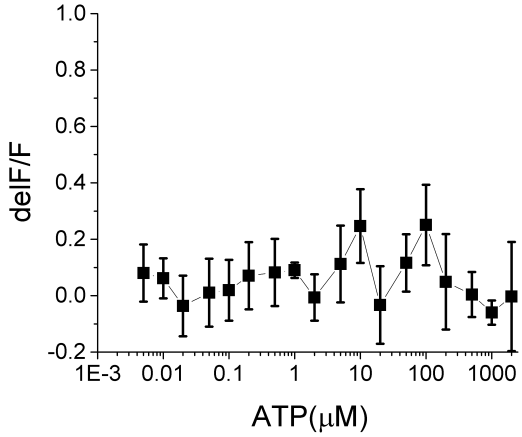

Figure 6.12: Relative change of ensemble average intensity (w.r.t. no ATP condition) plotted as a function of ATP concentration. The error bars represent standard error of mean from three experiments.

points until bleaching) of each single-step bleaching trace was determined. The mean of these average intensities (for a given experiment) was assumed to represent the ensemble average of the intensities of monomers for that specific condition. The relative change (with respect to no ATP condition) in ensemble average intensities is plotted against the ATP concentration in figure 6.12. As evident, there is no significant effect of ATP on the ensemble average fluorescence. This is in stark contrast to ensemble measurements of Cy3 labeled pump in membrane fragments, where ATP induced quenching. However, it is important to remember that, while the latter is a bulk fluorescence measurement in cuvette, the former is a single-molecule analysis of about a thousand traces in each case. Since there is a huge difference in no. of molecules involved in averaging, a direct comparison may not be

reasonable.

For an appropriate comparison, bulk fluorescence measurement of the pump monomers in vesicles was performed on a spectrophotometer (LAMBDA 1050, PerkinElmer in the group of Prof. Thomas Bein, LMU, Munich), at different ATP concentrations. Unfortunately, this proved to be a difficult experiment to reproduce reliably; the change in fluorescence upon adding ATP hardly noticeable after correcting for dilution. This was mainly due to the limitation of the yield of Cy3 labeled pump monomers from the ÄKTA. Thus, the fluorescence signal was already very weak for bulk measurement, though sufficient for single-molecule measurements.

6.7 Effect of ATP on Cy3 Fluorescence

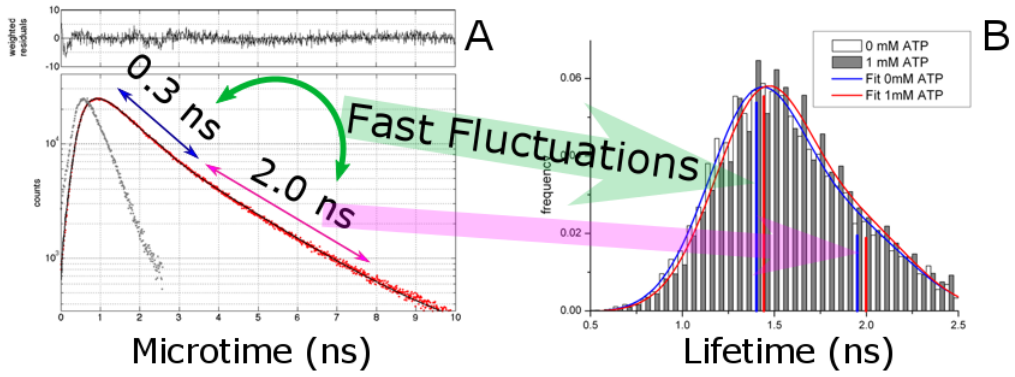

Figure 6.13: Ensemble TCSPC and Burst analysis of Cy3-labeled pump monomers in vesicles. A: Microtime histogram of the sample without ATP (red) is fitted with two lifetime components - 0.3 ns (free dye) and 2.0 ns (stuck state). The IRF is shown in grey and the weighted residuals of the fit are shown in the top part. B: The lifetime distribution obtained from a burst analysis of the sample in the absence (white bars) and presence (grey bars) of 1 mM ATP. The lifetime distributions were fit with two Gaussians, shown in blue for 0 ATP and in red for 1 mM ATP. The mean lifetimes are 1.4 ± 0.15 ns and 2.0 ± 0.2 ns. The shorter lifetime is attributed to fast fluctuations between the free and stuck state of the Cy3 dye (green arrow). The longer lifetime is purely due to the the dye being in stuck state (magenta arrow). The data were analyzed by Anders Barth.

In ensemble fluorescence measurements, it was observed that the pump labeled with Cy3 in the membrane was quenched upon adding ATP (figure 6.3C). However, the precise mechanism for quenching remained obscure. In order to understand the dye photophysics, solution measurements of the Cy3 labeled pump monomer in vesicles were performed on the MFD setup (as described in chapter 2).

The ensemble TCSPC of the collected photons is shown in figure 6.13A. The microtime histogram could be fit with two lifetime components; one with a lifetime of 0.3 ns and the other with a longer lifetime of 2.0 ns. The shorter lifetime is similar to the reported lifetime of free Cy3 dye in water (0.26 ns) [Muddana et al., 2014], hence indicating that the dye is present in an environment resembling its "free" state. The longer lifetime of the dye could be due to the dye being unable to fluoresce like in a free state, as a result of its environment. Longer lifetime is generally associated with higher brightness, implying that Cy3 is not undergoing cis-trans isomerization as in its free state, but is rather in a "stuck" state where the isomerization is being prohibited probably due to steric hindrance or effect of a nearby residue.

The burst analysis lifetime distributions (figure 6.13 B) of the Cy3-labeled pump monomers in vesicles showed no prominent difference in the absence (white bars) or presence of 1 mM ATP (grey bars). The lifetime distributions could be fitted to a sum of two Gaussians, one with a mean lifetime of 1.40 $\pm$ 0.15 ns and the second with a mean lifetime of 2.0 $\pm$ 0.2 ns. The burst analysis measurements indicate that the dye could be undergoing fast fluctuations between the free state (τ = 0.3 ns) and the stuck state (τ = 2.0 ns). The ensemble average of these fast fluctuations is reflected in the lifetime distribution by the Gaussian with the mean lifetime around 1.4 ns. The Gaussian with a mean lifetime around 2.0 ns thus arises directly from the dye being in the 'stuck' state.

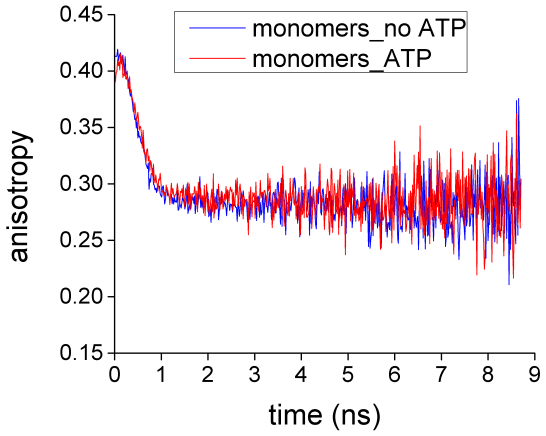

Figure 6.14: Time-resolved anisotropy of Cy3-labeled pump monomers in vesicles in the absence (blue) and presence (red) of 1mM ATP. No change is observed in the anisotropy. The data were analyzed by Anders Barth.

Anisotropy information of the dye was also collected in the MFD measurement. As evidenced in the figure 6.14, the calculated anisotropy of the monomers was unaffected by the presence of ATP. This signifies that ATP binding to pump monomers does not restrict the movement of dye any

more than before.

6.8 Effect of Aggregation on Cy3

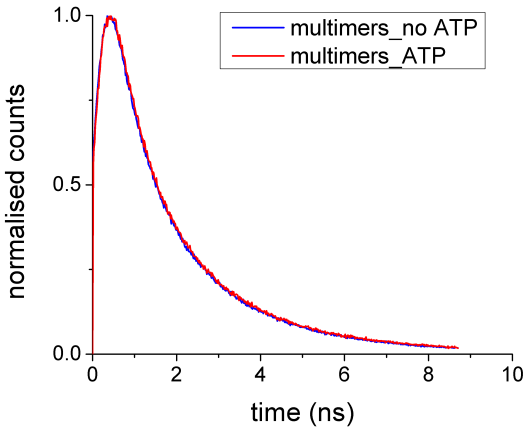

Figure 6.15: Fluorescence Lifetime histograms of Cy3 labeled pump multimers in vesicles, in absence of ATP (blue) and in presence of 1 mM ATP (red). The data were analyzed by Anders Barth.

It is difficult to measure a diluted sample of Cy3-labeled pump in membrane fragments on the MFD setup. The diluted sample, however finely pipetted, still contains huge aggregates that are too bright for the APDs (detectors) and, when detected, would usually trigger the automatic shut off to avoid damage to the detector. This made it cumbersome to measure even for a few minutes.

Therefore, to study the effect of pump aggregation on the dye properties, Cy3-labeled tetramers of the pump (obtained from well B-12, as explained in 6.5) were reconstituted into vesicles (similar to the monomers). The tetramer sample was measured on the MFD setup and the fluorescence lifetime and anisotropy information obtained. Hereafter, the tetramer sample is referred to as multimers, for the sake of convenience.

The fluorescence lifetime histograms of monomers and multimers are comparable, as seen in figure 6.6; the lifetimes being unaffected by ATP. In the individual case of multimers too, ATP produced no change in the fluorescence lifetime (seen in figure 6.15).

Likewise, the calculated anisotropy of multimers is unaffected by ATP (shown in figure 6.17). Moreover, as shown in figure 6.18, the anisotropies of the monomers and multimers do not differ significantly; almost invariable to the presence of ATP.

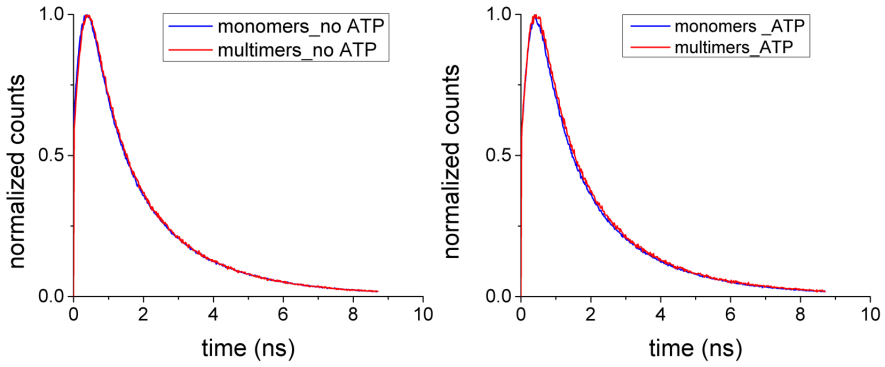

Figure 6.16: Lifetime histograms of monomers (blue) and multimers (red) in the absence (left) and presence (right) of 1 mM ATP. No change is observed due to ATP or aggregation. the data were analyzed by Anders Barth.

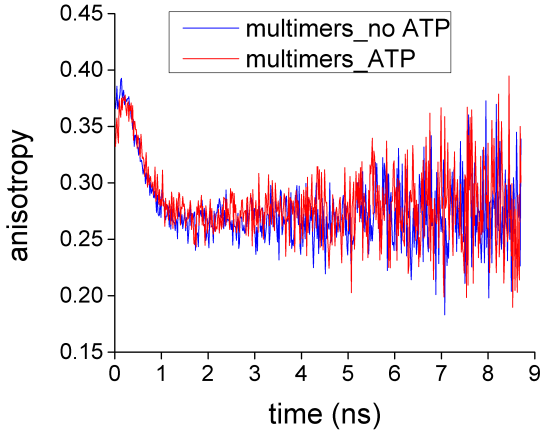

Figure 6.17: Time-resolved anisotropy of Cy3-labeled pump multimers in vesicles in the absence (blue) and presence (red) of 1 mM ATP. No change is observed in anisotropy. The data were analyzed by Anders Barth.

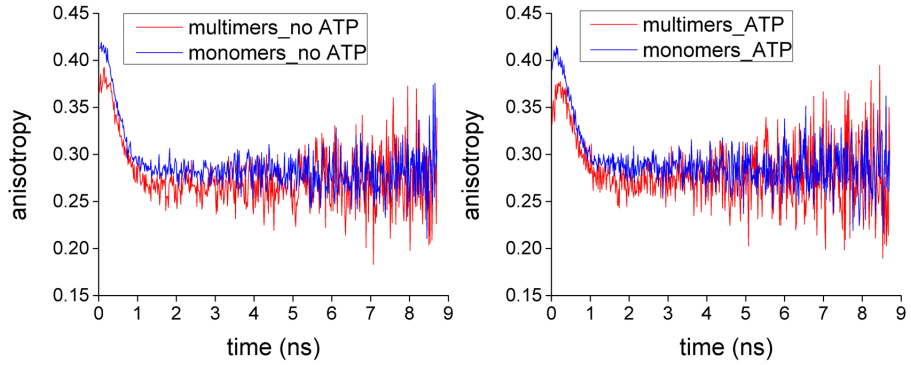

Figure 6.18: Time-resolved anisotropy of monomers (blue) and multimers (red) in the absence (left) and presence (right) of 1 mM ATP. No significant change is observed due to ATP or aggregation. The data were analyzed by Anders Barth.

6.9 The Curious Case of Cy3-labeled Pump

The Cy3 dye is known to undergo a cis-trans isomerization, which causes blinking [Jia et al., 2007]. Since in our experiments we mostly observe long lived traces (10-20 s long), without the addition of any photo-cocktail or oxygen scavenging system, the question arises whether the transitions we observe in the dynamic traces are actually due to changes in ATP concentration or it is just a protein-induced fluorescence enhancement (PIFE) effect [Hwang et al., 2011]. It could be that, being in the proximity of the protein, the blinking appears as a quenching dynamics of the dye. To test this hypothesis, the sample was labeled with Cy3b-maleimide. Cy3b does not undergo isomerization and does not exhibit PIFE.

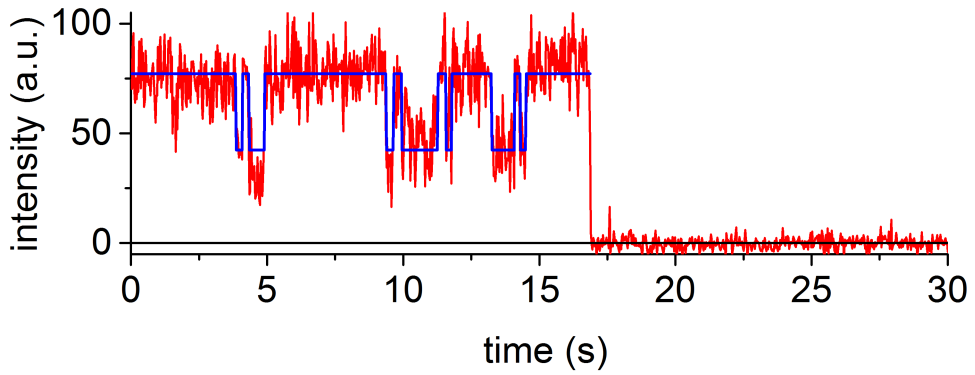

Figure 6.19: An exemplar trace of Cy3b-labeled pump monomer in vesicles measured on TIRFM showing dynamics between an unquenched state and a quenched state as seen earlier for Cy3-labeled pump monomer in figure 6.8 B

Even with Cy3b traces, fluorescence quenching dynamics were observed (figure 6.19). Thus, this ensures that the observed dynamics is not PIFE but actually represents changes in the dye environment. Similar to Cy3, Cy3b also shows an increase in the percentage of dynamic traces from 40% to 80% upon addition of 1 mM ATP.

6.10 Discussion

Na^+-K^+ ATPase was the first membrane-associated primary active ion transporter (ion pump) to be discovered [Skou, 1957]. Being one of the most important members of the P-type ATPase family, the pump has been extensively investigated over several decades using techniques ranging from stopped-flow with spectroscopic detection for kinetic studies to steady state fluorescence detection and even isothermal titration calorimetry. The pump has been purified into membranes, endothelial cells and even microsomes. Yet, all these experiments have been based on bulk measurements of the

pump. Experiments performed on solubilized pump; [Martin and Sachs, 1999] reported that protomers in microsomes were mostly monomers, yet no attempt was made to isolate monomers and study them.

The first ever single-molecule measurements of the pump on the TIRF and MFD setups are reported in this chapter, aiming to throw new light on an old pump. In our study, the pump was labeled in the membranes. The activity of the pump when labeled in the membrane is around 80% [Pratap et al., 2009]. After labeling, the pump in the membrane was solubilized in the surfactant $C_{12}E_8$. Yoneda et al. [2016] identified 7 different populations after chromatography purification of solubilized pump, with monomers and tetramers amounting to 55 % of the sample and a dynamic equilibrium between monomers, tetramers and higher order oligomers. They reported that higher-order oligomers could be removed with a 220 nm pore filter.

In our studies, specifically the monomer fraction (as determined from the aldolase standard measurement) was taken, and extruded through 200 nm pore membrane, which sufficiently removed higher aggregates. The orientation of pump in vesicles is assumed to be stochastic. Half of the monomers have the α subunit outside with the ATP binding site accessible and the other half have it oriented inside the vesicle. This implies that only the proteins with the α subunit outside could bind ATP.

Approximately, 20% of the single-molecule traces extracted from the TIRF measurements showed multiple steps; $2-7$ steps. For further analysis, I focused on the traces that exhibited a single bleaching step. These traces presumably represented a single labeled protein in one vesicle. There is the possibility that the vesicle contains an unlabeled protein as well but it is assumed to be a rare event.

The HMM analysis classified the single-molecule single bleaching step traces into two categories; non-dynamic traces and dynamic traces (figure 6.8). In dynamic traces, the protein existed in a dynamic equilibrium between 2 states (represented by the unquenched and the partially quenched states) (figure 6.8 B). The kinetics were characterized by slow transitions between these two states. ATP shifted the population distribution from those exhibiting no transitions (non-dynamic) to those exhibiting slow transitions (dynamic) (figure 6.11).Furthermore, the effect of ATP on the fraction of dynamic traces followed a negative cooperativity, with the data being better fit by the hill equation.

Interestingly, these two types of traces also existed in the absence of ATP. However, the presence of the ATP only increased the fraction of traces exhibiting dynamics, with no effect on the equilibrium constants

whatsoever (figure 6.10). Moreover, the rate from high-to-low stayed higher than the rate from low-to-high states, irrespective of ATP concentration, implying that the pump preferred staying in the partially quenched state. Moreover, the average bleaching time of the traces was around 10-20 s, whereas Cy3 commonly would bleach faster (3-5 s) and blink vigorously (private communication with Evelyn Ploetz). Despite the absence of any photococktails or ROXS system, the traces were stable. Therefore, to rule out any photophysics of Cy3 being responsible for the dynamics observed, monomers of the pump were labeled with Cy3b-maleimide and measured on the TIRF. These traces too exhibited similar behavior, with ATP increasing the fraction of dynamic traces (figure 6.19). Thus we could rule out the dynamics being a PIFE effect.

The ensemble TCSPC measurements showed that the fluorophore lifetime had two components (figure 6.13), one arising out of the fluorophore being in a "free" state as the lifetime was similar to that of free dye in solution (0.3 ns), and the other corresponding to a longer lifetime (2.0 ns) indicating the dye being in a sterically hindered conformation in a "stuck" state. The effect of these lifetimes is reflected in the burst analysis, where the mean lifetime could be fit with two Gaussians. One lifetime corresponded to the dye being in the stuck state (2.0 ns) while the other suggested the dye possibly undergoing fast fluctuations between the free and stuck state thus resulting in an ensemble average lifetime of 1.4 ns. Also in the burst analysis experiments, ATP concentration did not effect the lifetimes of the dye, indicating that a static quenching mechanism might be in operation. In order to connect these observations with the TIRF results, a surface-MFD would be required. It would help resolve dynamics on the timescale of μs, which would be missed in TIRF. Also, the fluorescence lifetimes of each state in the dynamics could be measured, thus giving a direct correlation between intensity and lifetime.

Comparison of the monomers with multimers yielded similar results. Neither the pump aggregation nor ATP or a combination thereof had any effect on the fluorophore lifetime or its anisotropy (figures 6.18, 6.16). These results indicate that the pump behaves the same in a monomer as it does in membrane fragments. Thus, negative cooperativity is an intrinsic property of the monomer.

Cooperativity is traditionally supposed to involve interactions between subunits. However, cooperativity in monomeric enzymes with single ligand binding sites, eg. Glucokinase [Storer and Cornish-Bowden, 1976, Cornish-Bowden and Cárdenas, 2004], is also well documented [Porter and Miller, 2012].

The pump monomer has a single ligand-binding site and exhibits kinetic cooperativity that deviates from hyperbolic response of a steady state enzyme response. Instead, the data fits with a Hill equation with $n < 1$ indicating negative cooperativity. Here, n (the Hill coefficient) is only an easy measure of the relative magnitude of cooperative effect. A non Michaelis-Menten kinetics response of monomeric enzymes is observed. All of these are the classic characteristics attributed to cooperativity in monomeric enzymes [Porter and Miller, 2012].

The proposed mechanism for cooperativity involves ATP induced conformational changes in the nucleotide binding domain of pump as previously reported by Hilge et al. [2003] and Petrushanko et al. [2014]. The latter reported that the ATP binding to Na^+-K^+-ATPase, in contrast to ADP binding, generates a structural transition in the enzyme, which is consistent with the movement of a significant portion of the surface area to a solvent-protected state. They proposed that ATP binding led to a convergence of the nucleotide-binding and phosphorylation domains, transferring the enzyme from the "E1-open" to "E1-closed" conformation, and prepare the enzyme for phosphorylation and showed supporting molecular modeling data from the crystal structure of the pump in Na^+ bound form, preceding E1P state (PDB: 3WGU). The pump monomers studied here are also assumed to be in a similar conformation, hence their results are applicable to our study. Moreover, such hinge-bending motions deciphered from the crystal structure information have been reported to be responsible for 'slow' transformational changes that cause monomeric cooperativity [Kamata et al., 2004].

From the MALDI-MS data, the minimally labeled pump was found to be labeled at Cys705, in the P domain, near to the ATP binding pocket in the N domain. This labeling site is most likely sensitive to any hinge-bending motions that bring about a change from the "E1-open" conformation - where N and P domains are a bit farther apart, to the "E1-closed" conformation - where N and P domains move closer together and lock the ATP in a pocket, making the pump ready for phosphorylation .

Our proposed model for monomeric cooperativity is similar to the ligand induced slow transition model [Ainslie et al., 1972, CÁRDENAS et al., 1984, Cornish-Bowden and Cárdenas, 1987, Porter and Miller, 2012] or LIST (illustrated in figure 6.20, left panel). LIST model involves the existence of two distinct enzyme conformations - E and E*. In our proposed model (figure 6.20, right panel), the E conformation is defined to be the state that exhibits non-dynamic traces, while the E* conformation of the pump exhibits dynamic traces. In LIST, E and E* have a pre-equilibrium in the absence of substrate, which in our case is in agreement with the fact that the two

types of traces (non-dynamic and dynamic) are present even in the absence of ATP. In LIST, the two enzyme conformations have different affinities for the substrate and the equilibrium between the states is controlled by the substrate concentration. This is clearly seen in our case where by the increasing the ATP concentration, the fraction of dynamic traces increases. This indicates that our substrate, the ATP, controls the "slow" transitions of the non-dynamic molecules into the dynamic molecules. These transitions, however, could not be directly observed in TIRF measurements as it was difficult to determine if the the traces switched from dynamic to non-dynamic in the duration of the experiment. However, it could also be that the non-dynamic traces actually exhibit transitions much faster than the time resolution of the experiment.

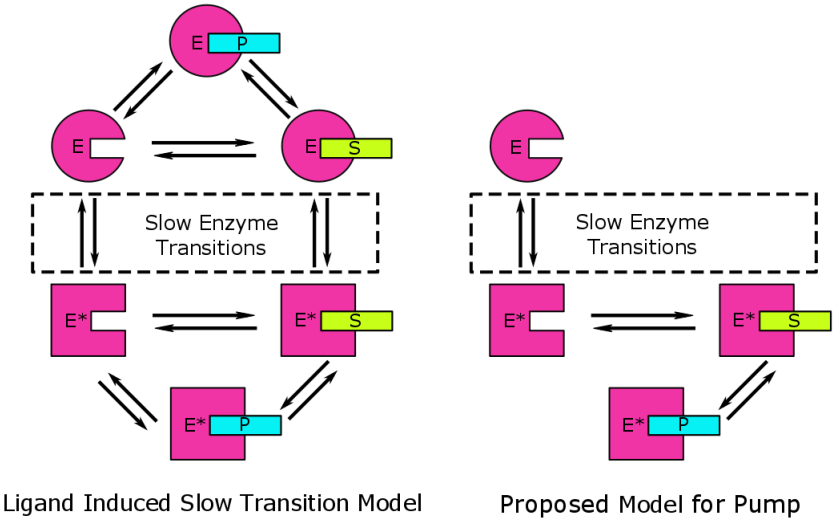

Figure 6.20: Left: The schematic of Ligand Induced Slow Transition model which explains kinetic cooperativity in monomeric enzymes by existence of two conformations E and E* that interconvert slowly and have different catalytic properties. Right: The proposed model for pump functioning in E1 state binding to ATP.

In LIST, each conformation can undergo its own catalytic cycle. This is where our model (Figure 6.20, right panel) differs. We assume that ATP effects the interconversion of the E and E*, but only one of these forms is phosphorylation-ready, which is necessary for the pump cycle to continue to next step. Since we are only looking at ATP binding but not phosphorylation, experiments in the presence of Mg^{2+} will help prove our model.

From these observations, one can conclude that the protein exists in at least two conformations and that ATP binding stabilizes one form. It is likely that this conformation may be the form of the pump that participates in

ion transport and ATP hydrolysis.

Single-molecule fluorescence techniques and smFRET are excellent ways to monitor slow conformational changes in biomolecules. In the future, it would be interesting to label the N and P domains and directly observe the kinetics of the conformational change upon ATP binding. Also, other conformations of pump (E1P, E2P etc.) could be studied at single-molecule level. Furthermore, it might help to replace vesicles with nanodiscs [Bayburt et al., 2002, Bayburt and Sligar, 2010] to make the ATP binding site available irrespective of the orientation of the pump in the lipid membrane.

7. Conclusions

Single-molecule techniques enable real-time probing of individual molecules, allowing characterization of molecular dynamics and nanoscale interactions. Single-pair fluorescence resonance energy transfer (spFRET) measurements with hidden markov model (HMM) analysis of the FRET time traces has been extensively used to determine the sequence of underlying states and calculation of the dwell times. However, the complex biological systems involving multiple interacting components requires simultaneous monitoring of different regions of the system. In this thesis, a multi-color total internal reflection fluorescence(TIRF) microscope setup was assembled and optimized. TIRF microscopy in combination with alternating laser excitation and multi-color FRET detection was utilized to study different types of DNA based systems. Additionally, Multiparameter fluorescence detection (MFD) in combination with pulsed interleaved excitation (PIE) was used to characterize the fluorescence properties (lifetime, anisotropy , etc.) of the fluorophores involved.

In the first project, the autonomous switching of a surface tethered molecular construct performing a catalytic hairpin assembly (CHA) reaction was monitored through spFRET measurements on TIRF and MFD-PIE. CHA reaction cycle involves a series of toe-hold mediated branch migrations, initiated by the complementary nucleotide domains on the different components. At the end of each cycle, a hairpin duplex is formed and the catalyzing strand is refurbished. CHA reaction cycle was studied for a set of strands designed by monitoring FRET between the donor labeled catalyzing strand and acceptor labeled anchor strand after each step. Upon addition of the first hairpin, binding of the first hairpin to the catalyzing strand to form a duplex resulted in an increase in the distance between the fluorophores, which was seen as a decrease in FRET (from 0.48 to 0.1). However, the addition of the second hairpin did not result in the formation of hairpin duplex and recovery of the catalyzing strands the FRET efficiency remained unchanged. The results were similar when the experiments were repeated with the second hairpin being complementary to the first hairpin.

Moreover, when the fluorophore pair was changed, only the value of FRET efficiencies changed (0.7; 0.05) but the results were similar. To understand the reason for non-functioning of the cycle, three-color FRET measurements were performed on MFD-PIE setup with Atto488 labeled second hairpin. The plot of BG stoichiometry vs GR FRET efficiency clarified that the addition of the second hairpin resulted in the formation of a complex where all the strands were stuck together, thus stalling the CHA reaction cycle. Alternatively, the second hairpin could be displacing the first hairpin and

forming a duplex with the catalyzing strand. The multiple interactions involved can be better studied by also labeling the first hairpin and monitoring FRET between different fluorophore pairs, to get the overall picture. Also, the spFRET experiments could be repeated at a higher temperature, which is more suitable for the CHA reaction cycle. Lastly, the strand sequences can be better optimized for the CHA cycle, especially the toe-hold nucleotide domains, to assist the branch migration of the first hairpin from the catalyzing strand more easily.

In the second project, the diffusive motion of a DNA roboarm on a DNA platform, synthesized by the DNA origami technique, was characterized by monitoring FRET between the Alexa488 at the end of the roboarm, and Atto565 and Atto647N labeled docking strands placed on the opposite sides of the roboarm on the platform. Additionally, a staple strand on the roboarm (arm-dock) was extended to have certain number of nucleotides complementary to the docking strands. The three-color FRET experiments with ALEX were performed on the multi-color TIRF setup. In the absence of the arm-dock, the traces showed steady blue signal with only residual signal in the green and red channels from the bleedthrough from blue channel. This indicated that the roboarm was diffusing around faster than the time resolution of acquisition ($\sim$33 ms). In the presence of the arm-dock, the signals from the dyes fluctuated between a high- and a low-fluorescence intensity level, with only one dye being in the high fluorescence intensity level(ON state) at a time. The fluorescence detected after excitation with blue laser (491 nm) indicated that signal in red or green channel was solely due to FRET and was evidenced by the anti-correlated behavior with the blue intensity detected.

These intensity traces were further subject to an HMM analysis and dwell times of the ON states of each dye were determined by constructing a cumulative ON trace from the Viterbi paths of each color trace. The dwell times indicated that the overlap length affected the binding as well as the dissociation rates. Different types of blue ON states, or blue blips were observed. The blue blips were segregated according to the transition they represented - RBG, GBR, RGR, GBG. The dwell times of each of these was determined separately. The type I blue blips (GBR, RBG) represented traversal of the roboarm from one docking strand to the other, while type II blue blips (RBR and GBG) represented the roboarm unbinding from a docking strand and diffusing back to the same strand. Type II blue blips were found to be 50% shorter than type I. Overall, the different diffusive behaviors could be characterized using multi-color FRET with simultaneous detection. In future, the kinetics of the roboarm movement could be tweaked accordingly for application in nanoscale assembly lines. Also, this showed that multi-color FRET at single-molecule level is a powerful tool to

characterize dynamic behavior of nanoscale systems.

In the third project, the threading of a single stranded DNA (ssDNA) through the meshwork of a hollow cylindrical DNA tube was monitored by FRET between the acceptor on threading strand and the donor on a non-threading duplex. The dynamics were analyzed with an HMM to determine the kinetics of switching of the underlying FRET states. The threading mechanism was further probed by introducing single stranded interfering strands with varying lengths complementary to the threading strand. The frame-wise FRET distributions indicated the presence of at least 3 FRET states, that shifted to lower values upon addition of interfering strands. Also, the presence of the interfering strand altered the fraction of static traces for each case, that exhibited a dependence on the length of the interfering strands. Thus, the dynamics involved in threading could be determined in relation to the extent of duplexity of the threading strand. The results could be explained by a three-state model for threading. However, in the absence of ALEX data, dye photophysics could not be ruled out. Overall, the HMM analysis provided a thorough examination of the FRET states involved in each case, along with the rates of dynamic switching. Measuring with different dyes and performing ALEX experiments could help strengthen the model. The final results could help explain the discrepancies seen in DNA-PAINT and other experiments involving this threading behavior.

The last project involved investigation of the Na^+-K^+ ATPase with different single-molecule fluorescence techniques. The sodium pump was extracted from cellular membrane fragments, labeled with Cy3 and solubilized in detergent $C_{12}E_8$. Subsequently, the monomer population was separated by employing fast protein liquid chromatography (FPLC) and reconstituted into lipid vesicles. The monomers in lipid vesicles were investigated with TIRF microscopy and pulsed interleaved excitation multiparameter detection (MFD-PIE) techniques. The traces obtained from TIRF measurements exhibited dynamics in the absence and presence of varying concentrations of ATP, in the timescale of hundred milliseconds. The fraction of dynamic traces increased with ATP concentration and the trend could be fit with hill equation exhibiting negative cooperativity.

Moreover, the MFD measurements showed that the fluorophore exhibits two lifetimes, one corresponding to free dye (0.3 ns) and a second longer lifetime (2 ns), which could be associated with the dye being in an environment that inhibits the cis-trans isomerization native to Cy3. The burst analysis showed an average lifetime of 1.4 ns which could be the effect of fast fluctuations of the dye between its free and non-free state, and a lifetime of 2.0 ns which corresponds directly to the non-free state. However, the

fluorescence lifetime and the anisotropy were unaffected by the presence of ATP. Moreover, the tetramers of the pump reconstituted into vesicles in the same manner, showed similar lifetime and anisotropies. The results seem to point to the inherent kinetic cooperative behavior of the monomer, which could be explained on the basis of presence of two slowly interconverting conformations of the monomer pump, which have been reported earlier in literature. There could also be structural evidence supporting this. However, the confirmation of the existence of the two conformations and their interconversion dynamics would require spFRET measurements of specifically labeled phosphorylation and nucleotide domains of the alpha subunit of the pump, which would be a topic of dissertation on its own.

Overall, single-molecule fluorescence techniques provide a powerful toolbox for studying dynamics in complex biological systems. The limitations of spFRET in probing multiple dimensions and multiple interacting partners can be overcome by utilizing multi-color FRET. Combining multi-color detection with ALEX on TIRF or MFD-PIE for solution measurements provides additional information about the individual FRET efficiencies and stoichiometries between the different fluorophore pairs. This allows monitoring the system as a whole, rather than sum of spFRET measurements of individual parts. Moreover, for TIRF measurements, the HMM analysis of traces helps determine the underlying FRET or intensity states, as well as a dwell time analysis to determine the dwell time of different states and their interconverting rates. In this thesis, HMM has been applied to data from labeled protein and various DNA based systems. This proves the versatility of FRET measurements, with multi-color FRET increasing the data collected from a single-measurement and the various analytical tools like HMM, transition density plot (TDP), Stoichiometry vs FRET efficiency plots providing an in-depth visualization of the underlying dynamics in these complex systems.

Bibliography

A. Aghebat Rafat, T. Pirzer, M. B. Scheible, A. Kostina, and F. C. Simmel. Surface-assisted large-scale ordering of DNA origami tiles. *Angewandte Chemie International Edition*, 53(29):7665–7668, 2014.

G. R. Ainslie, J. P. Shill, and K. E. Neet. Transients and cooperativity a slow transition model for relating transients and cooperative kinetics of enzymes. *Journal of Biological Chemistry*, 247(21):7088–7096, 1972.

R. Albers. Biochemical aspects of active transport. *Annual review of biochemistry*, 36(1):727–756, 1967.

M. Andrec, R. M. Levy, and D. S. Talaga. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden markov models. *The Journal of Physical Chemistry A*, 107(38):7454–7464, 2003.

A. Askari and W. Huang. Na+, k+-atpase: evidence for the binding of ATP to the phosphoenzyme. *Biochemical and biophysical research communications*, 104(4):1447–1453, 1982.

D. Axelrod, T. P. Burghardt, and N. L. Thompson. Total internal reflection fluorescence. *Annual review of biophysics and bioengineering*, 13(1): 247–268, 1984.

U. Banik and S. Roy. A continuous fluorimetric assay for ATPase activity. *Biochemical Journal*, 266(2):611, 1990.

A. Barth. Three-color multiparameter fluorescence detection with pulsed interleaved excitation. *Masters Thesis*, 2013.

A. Barth, L. V. von Voithenberg, and D. C. Lamb. Mfd-pie and pie-fi: Ways to extract more information with tcspc. In *Advanced Photon Counting*, pages 129–157. Springer, 2014.

L. E. Baum. An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3: 1–8, 1972.

L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6): 1554–1563, 1966.

L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.

T. H. Bayburt and S. G. Sligar. Membrane protein assembly into nanodiscs. *FEBS letters*, 584(9):1721–1727, 2010.

T. H. Bayburt, Y. V. Grinkova, and S. G. Sligar. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Letters*, 2(8):853–856, 2002.

W. Becker. *The bh TCSPC handbook*. Becker & Hickl, 2014.

G. Blanco, J. C. Koster, and R. W. Mercer. The alpha subunit of the Na, K-ATPase specifically and stably associates into oligomers. *Proceedings of the National Academy of Sciences*, 91(18):8542–8546, 1994.

E. Boukobza, A. Sonnenfeld, and G. Haran. Immobilization in surface-tethered lipid vesicles as a new tool for single biomolecule spectroscopy. *The Journal of Physical Chemistry B*, 105(48):12165–12170, 2001.

J. R. Brotherus, L. Jacobsen, and P. L. Jørgensen. Soluble and enzymatically stable (na++ k+)-ATPase from mammalian kidney consisting predominantly of protomer $\alpha\beta$-units: Preparation, assay and reconstitution of active na+, k+ transport. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 731(2):290–303, 1983.

H. Bui, C. Onodera, C. Kidwell, Y. Tan, E. Graugnard, W. Kuang, J. Lee, W. B. Knowlton, B. Yurke, and W. L. Hughes. Programmable periodicity of quantum dot arrays with DNA origami nanotubes. *Nano letters*, 10(9): 3367–3372, 2010.

B. Bukiet, E. R. Harold, and J. L. Palacios. A markov chain approach to baseball. *Operations Research*, 45(1):14–23, 1997.

T. P. Burghardt and N. L. Thompson. Evanescent intensity of a focused gaussian light beam undergoing total internal reflection in a prism. *Optical Engineering*, 23(1):230162–230162, 1984.

M. Cable and F. Briggs. Allosteric regulation of cardiac sarcoplasmic reticulum Ca-ATPase: a comparative study. *Molecular and cellular biochemistry*, 82(1-2):29, 1988.

M. L. CÁRDENAS, E. RABAJILLE, and H. NIEMEYER. Suppression of kinetic cooperativity of hexokinase d (glucokinase) by competitive inhibitors. *European Journal of Biochemistry*, 145(1):163–171, 1984.

K. Chakraborty, M. Chatila, J. Sinha, Q. Shi, B. C. Poschner, M. Sikor, G. Jiang, D. C. Lamb, F. U. Hartl, and M. Hayer-Hartl. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell*, 142(1):112–122, 2010.

R. Chhabra, J. Sharma, Y. Ke, Y. Liu, S. Rinker, S. Lindsay, and H. Yan. Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures. *Journal of the American Chemical Society*, 129(34):10304–10305, 2007.

J.-P. Clamme and A. A. Deniz. Three-color single-molecule fluorescence resonance energy transfer. *ChemPhysChem*, 6(1):74–77, 2005.

R. J. Clarke. Mechanism of allosteric effects of ATP on the kinetics of P-type ATPases. *European Biophysics Journal*, 39(1):3–17, 2009.

R. J. Clarke and D. J. Kane. Two gears of pumping by the sodium pump. *Biophysical journal*, 93(12):4187–4196, 2007.

R. J. Clarke, H.-J. Apell, and B. Y. Kong. Allosteric effect of ATP on Na+, K+-ATPase conformational kinetics. *Biochemistry*, 46(23):7034–7044, 2007.

A. Cornish-Bowden and M. L. Cárdenas. Co-operativity in monomeric enzymes. *Journal of theoretical biology*, 124(1):1–23, 1987.

A. Cornish-Bowden and M. L. Cárdenas. Glucokinase: a monomeric enzyme with positive cooperativity. In *Glucokinase and glycemic disease: From basics to novel therapeutics*, volume 16, pages 125–134. Karger Publishers, 2004.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

A. A. Deniz, M. Dahan, J. R. Grunwell, T. Ha, A. E. Faulhaber, D. S. Chemla, S. Weiss, and P. G. Schultz. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of förster distance dependence and subpopulations. *Proceedings of the National Academy of Sciences*, 96(7):3670–3675, 1999.

V. C. DeRocco, T. Anderson, J. Piehler, D. A. Erie, and K. Weninger. Four-color single molecule fluorescence with noncovalent dye labeling to monitor dynamic multimolecular complexes. *Biotechniques*, 49(5):807, 2010.

B. Ding, Z. Deng, H. Yan, S. Cabrini, R. N. Zuckermann, and J. Bokor. Gold nanoparticle self-similar chain structure organized by DNA origami. *Journal of the American Chemical Society*, 132(10):3248–3249, 2010.

P. A. Dirac. The quantum theory of the emission and absorption of radiation. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, pages 243–265. The Royal Society, 1927.

C. Donnet, E. Arystarkhova, and K. J. Sweadner. Thermal denaturation of the Na, K-ATPase provides evidence for α-α oligomeric interaction and γ subunit association with the c-terminal domain. *Journal of Biological Chemistry*, 276(10):7357–7365, 2001.

S. M. Douglas, I. Bachelet, and G. M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834, 2012.

J. Enderlein. Single molecule spectroscopy: basics and applications. In *Fluorescence Spectroscopy in Biology*, pages 104–130. Springer, 2005.

J. Enderlein, D. L. Robbins, W. P. Ambrose, P. M. Goodwin, and R. A. Keller. The statistics of single molecule detection: an overview. *Bioimaging*, 5(3):88–98, 1997.

E. Fermi. Nuclear physics. In *A course given by Enrico Fermi at the University of Chicago, 1949, Chicago: University of Chicago Press, 1950, Rev. ed.*, volume 1, 1950.

T. Förster. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der physik*, 437(1-2):55–75, 1948.

T. Förster. Experimentelle und theoretische untersuchung des zwischenmolekularen übergangs von elektronenanregungsenergie. *Zeitschrift für naturforschung A*, 4(5):321–327, 1949.

T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, and T. Yanagida. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature*, 374(6522):555, 1995.

K. Geering. The functional role of the β-subunit in the maturation and intracellular transport of Na, K-ATPase. *FEBS letters*, 285(2):189–193, 1991.

A. Gietl, P. Holzmeister, F. Blombach, S. Schulz, L. V. von Voithenberg, D. C. Lamb, F. Werner, P. Tinnefeld, and D. Grohmann. Eukaryotic and archaeal TBP and TFB/TF (ii) B follow different promoter DNA bending pathways. *Nucleic acids research*, 42(10):6219–6231, 2014.

H. Gu, J. Chao, S.-J. Xiao, and N. C. Seeman. A proximity-based programmable DNA nanoscale assembly line. *Nature*, 465(7295):202–205, 2010.

T. Ha, T. Enderle, D. Ogletree, D. Chemla, P. Selvin, and S. Weiss. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences*, 93(13):6264–6268, 1996.

C. O. Hartman. *Virtual muse: experiments in computer poetry*. Wesleyan University Press, 1996.

Y. He and D. R. Liu. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nature nanotechnology*, 5 (11):778, 2010.

G. Heiss. Single-molecule microscopy study of nano-systems from synthetic photo-switchable nano-devices to the dynamics of naturally occurring transcription factors. *Ph.D. Thesis*, 2012.

J. Hendrix and D. C. Lamb. Pulsed interleaved excitation: principles and applications. *Fluorescence Fluctuation Spectroscopy (FFS)*, 518:205, 2012.

M. Hilge, G. Siegal, G. W. Vuister, P. Güntert, S. M. Gloor, and J. P. Abrahams. ATP-induced conformational changes of the nucleotide-binding domain of Na, K-ATPase. *Nature Structural & Molecular Biology*, 10(6): 468–474, 2003.

T. Hirschfeld. Optical microscopic observation of single small molecules. *Applied optics*, 15(12):2965–2966, 1976.

S. Hohng, C. Joo, and T. Ha. Single-molecule three-color FRET. *Biophysical journal*, 87(2):1328–1337, 2004.

H. Hwang, H. Kim, and S. Myong. Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proceedings of the National Academy of Sciences*, 108(18):7414–7418, 2011.

F. Jaisser, P. Jaunin, K. Geering, B. C. Rossier, and J.-D. Horisberger. Modulation of the Na, K-pump function by beta subunit isoforms. *The Journal of General Physiology*, 103(4):605–623, 1994.

K. Jia, Y. Wan, A. Xia, S. Li, F. Gong, and G. Yang. Characterization of photoinduced isomerization and intersystem crossing of the cyanine dye Cy3. *The Journal of Physical Chemistry A*, 111(9):1593–1597, 2007.

Y. Jiang, B. Li, J. N. Milligan, S. Bhadra, and A. D. Ellington. Real-time detection of isothermal amplification reactions with thermostable catalytic hairpin assembly. *Journal of the American Chemical Society*, 135(20): 7430–7433, 2013.

C. Joo and T. Ha. Single-molecule FRET with total internal reflection microscopy. *Cold Spring Harbor Protocols*, 2012(12):pdb–top072058, 2012.

C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.*, 77:51–76, 2008.

P. L. Jørgensen and J. P. Andersen. Structural basis for E1–E2 conformational transitions in Na, K-pump and Ca-pump proteins. *The Journal of membrane biology*, 103(2):95–120, 1988.

R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, and F. C. Simmel. Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano letters*, 10(11):4756–4761, 2010.

K. Kamata, M. Mitsuya, T. Nishimura, J.-i. Eiki, and Y. Nagata. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure*, 12(3):429–438, 2004.

R. Kanai, H. Ogawa, B. Vilsen, F. Cornelius, and C. Toyoshima. Crystal structure of a Na+-bound Na+, K+-ATPase preceding the E1P state. *Nature*, 502(7470):201–206, 2013.

A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, and S. Weiss. Alternating-laser excitation of single molecules. *Accounts of chemical research*, 38(7):523–533, 2005.

J. H. Kaplan. Biochemistry of Na, K-ATPase. *Annual review of biochemistry*, 71(1):511–535, 2002.

M. Kasha. Characterization of electronic transitions in complex molecules. *Discussions of the Faraday society*, 9:14–19, 1950.

E. Kopperger, T. Pirzer, and F. C. Simmel. Diffusive transport of molecular cargo tethered to a DNA origami platform. *Nano letters*, 15(4):2693–2699, 2015.

V. Kudryavtsev. *Development and applications of new fluorescence techniques to characterize single biomolecules in solution and living cells*. PhD thesis, Citeseer, 2006.

V. Kudryavtsev, M. Sikor, S. Kalinin, D. Mokranjac, C. A. Seidel, and D. C. Lamb. Combining MFD and PIE for accurate single-pair förster resonance energy transfer measurements. *ChemPhysChem*, 13(4):1060–1078, 2012.

A. Kuzyk, K. T. Laitinen, and P. Törmä. DNA origami as a nanoscale template for protein assembly. *Nanotechnology*, 20(23):235305, 2009.

J. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer US, 2007. ISBN 9780387463124.

K. Lang and J. W. Chin. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chemical reviews*, 114(9):4764–4806, 2014.

J. Lee, S. Lee, K. Ragunathan, C. Joo, T. Ha, and S. Hohng. Single-molecule four-color FRET. *Angewandte Chemie International Edition*, 49(51): 9922–9925, 2010a.

S. Lee, J. Lee, and S. Hohng. Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PloS one*, 5(8): e12270, 2010b.

B. Li, A. D. Ellington, and X. Chen. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic acids research*, 39(16):e110–e110, 2011.

F. Li, H. Zhang, Z. Wang, X. Li, X.-F. Li, and X. C. Le. Dynamic DNA assemblies mediated by binding-induced DNA strand displacement. *Journal of the American Chemical Society*, 135(7):2443–2446, 2013.

A. V. Lukashin and M. Borodovsky. Genemark. hmm: new solutions for gene finding. *Nucleic acids research*, 26(4):1107–1115, 1998.

K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, et al. Molecular robots guided by prescriptive landscapes. *Nature*, 465(7295): 206–210, 2010.

J. E. Mahaney, D. D. Thomas, and J. P. Froehlich. The time-dependent distribution of phosphorylated intermediates in native sarcoplasmic reticulum ca2+-atpase from skeletal muscle is not compatible with a linear kinetic model. *Biochemistry*, 43(14):4400–4416, 2004.

L. Mallik, S. Dhakal, J. Nichols, J. Mahoney, A. Dosey, S. Jiang, R. Sunahara, G. Skiniotis, and N. Walter. Electron microscopic visualization of protein assemblies on flattened DNA origami. *ACS nano*, 9(7):7133, 2015.

K. Mapa, M. Sikor, V. Kudryavtsev, K. Waegemann, S. Kalinin, C. A. Seidel, W. Neupert, D. C. Lamb, and D. Mokranjac. The conformational dynamics of the mitochondrial Hsp70 chaperone. *Molecular cell*, 38(1):89–100, 2010.

D. W. Martin and J. R. Sachs. Preparation of Na+, K+-ATPase with near maximal specific activity and phosphorylation capacity: evidence that the reaction mechanism involves all of the sites. *Biochemistry*, 38(23): 7485–7497, 1999.

D. W. Martin and J. R. Sachs. Ligands presumed to label high affinity and low affinity ATP binding sites do not interact in an ($\alpha\beta$) 2 diprotomer in duck nasal gland na+, k+-atpase, nor do the sites coexist in native enzyme. *Journal of Biological Chemistry*, 275(32):24512–24517, 2000.

D. W. Martin, J. Marecek, S. Scarlata, and J. R. Sachs. $\alpha\beta$ protomers of na+, k+-ATPase from microsomes of duck salt gland are mostly monomeric: Formation of higher oligomers does not modify molecular activity. *Proceedings of the National Academy of Sciences*, 97(7): 3195–3200, 2000.

M. McGuirk and C. Carniglia. An angular spectrum representation approach to the goos-hänchen shift. *JOSA*, 67(1):103–107, 1977.

S. A. McKinney, C. Joo, and T. Ha. Analysis of single-molecule FRET trajectories using hidden markov modeling. *Biophysical journal*, 91(5): 1941–1951, 2006.

T. K. Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6):47–60, 1996.

J. P. Morth, B. P. Pedersen, M. S. Toustrup-Jensen, T. L.-M. Sørensen, J. Petersen, J. P. Andersen, B. Vilsen, and P. Nissen. Crystal structure of the sodium–potassium pump. *Nature*, 450(7172):1043–1049, 2007.

H. S. Muddana, S. Sengupta, A. Sen, and P. J. Butler. Enhanced brightness and photostability of cyanine dyes by supramolecular containment. *arXiv preprint arXiv:1410.0844*, 2014.

B. K. Müller, E. Zaychikov, C. Bräuchle, and D. C. Lamb. Pulsed interleaved excitation. *Biophysical journal*, 89(5):3508–3522, 2005.

J. B. Munro, R. B. Altman, N. O'Connor, and S. C. Blanchard. Identification of two distinct hybrid state intermediates on the ribosome. *Molecular cell*, 25(4):505–517, 2007.

K. Murphy. Hmm toolbox for matlab. *Internet: http://www. cs. ubc. ca/˜ murphyk/Software/HMM/hmm. html,[Oct. 29, 2011]*, 1998.

E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, and S. Weiss. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *The Journal of Physical Chemistry B*, 110(44):22103–22124, 2006.

S. Noguchi, K. Higashi, and M. Kawamura. A possible role of the beta-subunit of (na, k)-ATPase in facilitating correct assembly of the alpha-subunit into the membrane. *Journal of Biological Chemistry*, 265 (26):15991–15995, 1990.

C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, 244(4901):182, 1989.

M. Nyblom, H. Poulsen, P. Gourdon, L. Reinhard, M. Andersson, E. Lindahl, N. Fedosova, and P. Nissen. Crystal structure of Na+, K+-ATPase in the Na+-bound state. *Science*, 342(6154):123–127, 2013.

H. Ogawa, T. Shinoda, F. Cornelius, and C. Toyoshima. Crystal structure of the sodium-potassium pump (na+, k+-atpase) with bound potassium and ouabain. *Proceedings of the National Academy of Sciences*, 106(33): 13742–13747, 2009.

M. J. Palladino, J. E. Bower, R. Kreber, and B. Ganetzky. Neural dysfunction and neurodegeneration indrosophila Na+/K+ ATPase alpha subunit mutants. *The Journal of neuroscience*, 23(4):1276–1286, 2003.

M. G. Palmgren. Plant plasma membrane H+-ATPases: powerhouses for nutrient uptake. *Annual review of plant biology*, 52(1):817–845, 2001.

M. G. Palmgren and P. Nissen. P-type atpases. *Annual review of biophysics*, 40:243–266, 2011.

J. S. Pedersen and J. Hein. Gene finding with a hidden markov model of genome structure and evolution. *Bioinformatics*, 19(2):219–227, 2003.

J. Perrin. Fluorescence et induction moleculaire par resonance. *CR Hebd. Seances Acad. Sci*, 184:1097–1100, 1927.

I. Y. Petrushanko, V. A. Mitkevich, A. A. Anashkina, E. A. Klimanova, E. A. Dergousova, O. D. Lopina, and A. A. Makarov. Critical role of γ-phosphate in structural transition of Na, K-ATPase upon ATP binding. *Scientific reports*, 4:5165, 2014.

A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan. Challenges and opportunities for structural DNA nanotechnology. *Nature nanotechnology*, 6(12):763–772, 2011.

C. M. Porter and B. G. Miller. Cooperativity in monomeric enzymes with single ligand-binding sites. *Bioorganic chemistry*, 43:44–50, 2012.

R. L. Post, C. Hegyvary, and S. Kume. Activation by adenosine triphosphate in the phosphorylation kinetics of sodium and potassium ion transport adenosine triphosphatase. *Journal of Biological Chemistry*, 247(20): 6530–6540, 1972.

P. R. Pratap, A. Palit, E. Grassi-Nemeth, and J. D. Robinson. Kinetics of conformational changes associated with potassium binding to and release from Na+ K+-ATPase. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1285(2):203–211, 1996.

P. R. Pratap, O. Dediu, and G. U. Nienhaus. Ftir study of ATP-induced changes in na+/k+-atpase from duck supraorbital glands. *Biophysical journal*, 85(6):3707–3717, 2003.

P. R. Pratap, L. O. Mikhaylyants, and N. Olden-Stahl. Fluorescence measurements of nucleotide association with the Na+/K+-ATPase. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1794(11): 1549–1557, 2009.

D. Pratas, C. A. Bastos, A. J. Pinho, A. J. Neves, and L. M. Matos. DNA synthetic sequences generation using multiple competing markov models. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 133–136. IEEE, 2011.

F. Qin, A. Auerbach, and F. Sachs. A direct optimization approach to hidden markov modeling for single channel kinetics. *Biophysical Journal*, 79(4): 1915–1927, 2000.

L. Rabiner and B. Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

C. Ratzke, B. Hellenkamp, and T. Hugel. Four-colour FRET reveals directionality in the hsp90 multicomponent machinery. *Nature communications*, 5, 2014.

W. W. Reenstra, J. D. Bettencourt, and J. G. Forte. Kinetic studies of the gastric H, K-ATPase. evidence for simultaneous binding of ATP and inorganic phosphate. *Journal of Biological Chemistry*, 263(36): 19618–19625, 1988.

J. Ren, J. Wang, L. Han, E. Wang, and J. Wang. Kinetically grafting g-quadruplexes onto DNA nanostructures for structure and function encoding via a DNA machine. *Chemical Communications*, 47(38): 10563–10565, 2011.

M. S. Robbins and B. J. Hadwen. The noise performance of electron multiplying charge-coupled devices. *IEEE Transactions on Electron Devices*, 50(5):1227–1232, 2003.

P. W. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

B. Rotman. Measurement of activity of single molecules of β-d-galactosidase. *Proceedings of the National Academy of Sciences*, 47(12):1981–1991, 1961.

J. Schaffer, A. Volkmer, C. Eggeling, V. Subramaniam, G. Striker, and C. Seidel. Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *The Journal of Physical Chemistry A*, 103(3):331–336, 1999.

M. B. Scheible, G. Pardatscher, A. Kuzyk, and F. C. Simmel. Single molecule characterization of DNA binding and strand displacement reactions on lithographic DNA origami microarrays. *Nano letters*, 14(3):1627–1633, 2014.

G. Scheiner-Bobis. The sodium pump. *European Journal of Biochemistry*, 269(10):2424–2433, 2002.

P. Schluesche. Untersuchung der funktion und dynamik von DNA-transkriptionfaktoren mittels einzelmolekuele-fluorezenzmikroskopie. *Ph.D. Thesis*, 2007.

P. Schluesche, G. Stelzer, E. Piaia, D. C. Lamb, and M. Meisterernst. Nc2 mobilizes tbp on core promoter TATA boxes. *Nature structural & molecular biology*, 14(12):1196–1201, 2007.

B. Schuler, E. A. Lipman, and W. A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002.

N. C. Seeman. Nucleic acid junctions and lattices. *Journal of theoretical biology*, 99(2):237–247, 1982.

N. C. Seeman. DNA in a material world. *Nature*, 421(6921):427–431, 2003.

E. B. Shera, N. K. Seitzinger, L. M. Davis, R. A. Keller, and S. A. Soper. Detection of single fluorescent molecules. *Chemical Physics Letters*, 174 (6):553–557, 1990.

J. M. Shin, K. Munson, O. Vagin, and G. Sachs. The gastric HK-ATPase: structure, function, and inhibition. *Pflügers Archiv-European Journal of Physiology*, 457(3):609–622, 2009.

M. Sikor. Single-molecule fluorescence studies of protein folding and molecular chaperones. *Ph.D. Thesis*, 2012.

J. C. Skou. The influence of some cations on an adenosine triphosphatase from peripheral nerves. *Biochimica et biophysica acta*, 23:394–401, 1957.

D. A. Smith, W. Steffen, R. M. Simmons, and J. Sleep. Hidden-markov methods for the analysis of single-molecule actomyosin displacement data: the variance-hidden-markov method. *Biophysical journal*, 81(5): 2795–2816, 2001.

T. L.-M. Sørensen, J. V. Møller, and P. Nissen. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science*, 304(5677):1672–1675, 2004.

I. H. Stein, C. Steinhauer, and P. Tinnefeld. Single-molecule four-color FRET visualizes energy-transfer paths on DNA origami. *Journal of the American Chemical Society*, 133(12):4193–4195, 2011.

A. C. Storer and A. Cornish-Bowden. Kinetics of rat liver glucokinase. co-operative interactions with glucose at physiologically significant concentrations. *Biochemical Journal*, 159(1):7–14, 1976.

K. Taniguchi, S. Kaya, K. Abe, and S. Mårdh. The oligomeric nature of na/k-transport ATPase. *Journal of biochemistry*, 129(3):335–342, 2001.

N. Thompson and J. Pero. Total internal reflection fluorescence microscopy: Applications in biophysics. In *Fluorescence spectroscopy in biology*, pages 79–103. Springer, 2005.

N. L. Thompson and D. Axelrod. Immunoglobulin surface-binding kinetics studied by total internal reflection with fluorescence correlation spectroscopy. *Biophysical journal*, 43(1):103, 1983.

C. Toyoshima and T. Mizutani. Crystal structure of the calcium pump with a bound ATP analogue. *Nature*, 430(6999):529–535, 2004.

R. D. Vale, T. Funatsu, D. W. Pierce, L. Romberg, Y. Harada, and T. Yanagida. Direct observation of single kinesin molecules moving along microtubules. *Nature*, 380(6573):451, 1996.

A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, 13(2):260–269, 1967.

D. G. Ward and J. D. Cavieres. Solubilized alpha beta Na, K-ATPase remains protomeric during turnover yet shows apparent negative cooperativity toward ATP. *Proceedings of the National Academy of Sciences*, 90(11): 5332–5336, 1993.

G. Watterson. Motoo kimura's use of diffusion theory in population genetics. *Theoretical population biology*, 49(2):154–188, 1996.

S. Wawilow. Die fluoreszenzausbeute von farbstofflösungen. *Zeitschrift für Physik A Hadrons and Nuclei*, 22(1):266–272, 1924.

L. R. Welch. Hidden markov models and the baum-welch algorithm. *IEEE Information Theory Society Newsletter*, 53(4):10–13, 2003.

J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken, and C. A. Seidel. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical chemistry*, 78(6): 2039–2050, 2006.

C. J. Wu. On the convergence properties of the EM algorithm. *The Annals of statistics*, pages 95–103, 1983.

J. Wu, K. M. Parkhurst, R. M. Powell, M. Brenowitz, and L. J. Parkhurst. DNA bends in TATA-binding protein· TATA complexes in solution are DNA sequence-dependent. *Journal of Biological Chemistry*, 276(18): 14614–14622, 2001.

N. Wu, D. M. Czajkowsky, J. Zhang, J. Qu, M. Ye, D. Zeng, X. Zhou, J. Hu, Z. Shao, B. Li, et al. Molecular threading and tunable molecular recognition on DNA origami nanostructures. *Journal of the American Chemical Society*, 135(33):12172–12175, 2013.

J. S. Yoneda, G. Scanavachi, H. G. Sebinelli, J. C. Borges, L. R. Barbosa, P. Ciancaglini, and R. Itri. Multimeric species in equilibrium in detergent-solubilized Na, K-ATPase. *International journal of biological macromolecules*, 89:238–245, 2016.

J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce. NUPACK: analysis and design of nucleic acid systems. *Journal of computational chemistry*, 32(1):170–173, 2011.

N. Zarrabi, M. Düser, S. Ernst, R. Reuter, G. Glick, S. Dunn, J. Wrachtrup, and M. Börsch. Monitoring the rotary motors of single FoF1-ATP synthase by synchronized multi channel tcspc. In *Optics East 2007*, pages 67710F–67710F. International Society for Optics and Photonics, 2007.

A.-X. Zheng, J. Li, J.-R. Wang, X.-R. Song, G.-N. Chen, and H.-H. Yang. Enzyme-free signal amplification in the DNAzyme sensor via target-catalyzed hairpin assembly. *Chemical Communications*, 48(25): 3112–3114, 2012.

X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, and S. Chu. A single-molecule study of RNA catalysis and folding. *Science*, 288(5473):2048–2051, 2000.

Appendix

A.1 Multi-Intensity Trace Analysis (MITrA)

The Multi-Intensity Trace Analysis (MITrA) was programmed in MATLAB, using mainly the image processing toolbox. MITrA allows one to perform the following tasks:

1. Extract single-molecule intensity traces from TIFF image stacks.
The different parameters for the extraction can be set on the right side of the main GUI (figure 7.1). The size of each channel, the size of signal and background masks for extracting intensities, the excitation sequence (ALEX or single-wavelength), the exposure time for acquisition and the mapping mode (the master channel for mapping) can be specified.

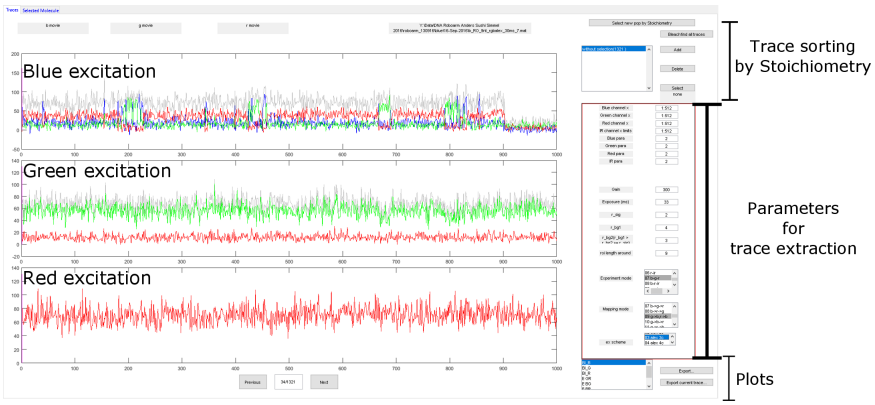

Figure 7.1: The main GUI of MITrA with all the various controls.

2. Selection of traces on the basis of Stoichiometry.
Once the intensity traces are extracted, they can be displayed according to the excitation source. An exemplar ALEX trace is shown in figure 7.1. The traces are colored according to the dye. The trace in grey represents total intensity detected after that excitation. One can scan through the traces and run an automatic bleach-step finding algorithm on them. Each molecule can be "tagged" with a user defined property and sorted into different populations. There is also provision for sorting traces according to Stoichiometry.

3. Plotting various FRET efficiency histograms.
Different FRET efficiency plots can be made by selecting a population and clicking on the required plot, for e.g. FRET Efficiency BG, FRET efficiency GR etc. Both molecule-wise and frame-wise FRET histograms are plotted for each case.

A.2 DNA Sequences for the Catalytic Hairpin Assembly

C1: 5'-Cy3-CGACATCTAACCTAGCTCACTGACTCAAAC TACCTAACCTAAAC-3'
A1: 5'-Biotin-GTTTAGGTTAGGTAGTTTGA-Cy5-3'
H11: 5'-GTCAGTGAGCTAGGTTAGATGTCGCCATGTGTAG ACGACATCTAACCTAGC-3'
H12: 5'-AGATGTCGTCTACACATGGCGACATCTAACCTAG CCCATGTGTAGA-3'
C2: 5'-Atto565-CGACATCTAACCTAGCTCACTGACTCAAACT ACCTAACCTAAAC-3'
A2: 5'-Biotin-GTTTAGGTTAGGTAGTTTGA-Atto647N-3'
H21: 5'-GTCAGTGAGCTAGGTTAGATGTCGCCATGTGTAG ACGACATCTAACCTAGC-3'
cH21: 5'-Atto488-GCTAGGTTAGATGTCGTCTACACATGGCGACA TCTAACCTAGCTCACTGAC-3'

A.3 Important DNA Sequences for the DNA Roboarm

Docking strand left: TT TT ATACAACTAT T [Atto647N]
Docking strand right: TT TT ATACAACTAT T [Atto565]
Arm-dock 8 nt: TT ATAGTTGT
Arm-dock 9 nt: TT ATAGTTGTA
Arm-dock 10 nt: TT ATAGTTGTAT

A.4 Models of Enzyme Kinetics

Michaelis-Menten Kinetics

Michaelis-Menten kinetics is the most well known model for enzyme kinetics. The model involves a monomeric enzyme E binding to a substrate S, to form the complex ES, which in turn releases the product P. The reaction rate V is given by the Michaelis-Menten equation :

$$V = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{7.1}$$

where V_{max} is the maximum rate achieved by the system, at saturating substrate concentration. The Michaelis constant K_m is the substrate concentration at which the reaction rate is half of V_{max}.
Taking the reciprocal, gives the Lineweaver-Burk plot:

$$\frac{1}{V} = \frac{K_m}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}} \tag{7.2}$$

Hill Equation

The Hill equation is used to model cooperative binding, when the binding curves are sigmoidal. The Hill equation is written as:

$$V = \frac{V_{max}[S]^n}{K_{0.5}^n + [S]^n} \tag{7.3}$$

where V is the reaction rate, *[S]* is the substrate concentration, $K_{0.5}$ is the substrate concentration at which reaction rate is half of V_{max}, and n is the Hill coefficient describing the cooperativity.
The cooperativity is defined from the value of n as follows:
1. $n > 1$ - positive cooperativity with respect to substrate binding
2. $n = 1$ - non-cooperative binding or Michaelis-Menten kinetics
3. $n < 1$ - negative cooperativity

The data in figure 6.11 was fit with 'Hill1' fit in Origin, with the equation for fit as:

$$y = start + \frac{(end - start)x^n}{k^n + x^n} \tag{7.4}$$

The rectangular hyperbola (dotted line) in figure 6.11 has the equation:

$$y = C + \frac{A \cdot x}{B + x} \tag{7.5}$$

.

Abbreviations

ADP	Adenosine Diphosphate
AFM	Atomic Force Microscope
ALEX	Alternating Laser Excitation
AOTF	Acousto-optic Tunable Filter
ATP	Adenosine Triphosphate
BIC	Bayesian Information Criterion
CHA	Catalytic Hairpin Assembly
CIC	Clock Induced Charge
DM-MBP	Double Mutant-Maltose Binding Protein
DNA	Deoxyribonucleic acid
EMCCD	Electron Multiplying Charge Coupled Device
FCS	Fluorescence Correlation Spectroscopy
FPGA	Field Programmable Gated Array
FPLC	Fast Protein Liquid Chromatography
FRET	Förster Resonance Energy Transfer
GODCAT	Glucose Oxidase/Catalase
HMM	Hidden Markov Model
IC	Internal Conversion
IRF	Instrument Response Function
ISC	Inter-System Crossing
MFD	Multiparameter Fluorescence Detection
MLE	Maximum Likelihood Estimate
NA	Numerical Aperture
PIE	Pulsed Interleaved Excitation
PIFE	Protein Induced Fluorescence Enhancement
PDB	Protein Database
ROI	Region of Interest
SmFRET	Single-molecule Fluorescence Resonance Energy Transfer
SNR	Signal-to-Noise Ratio
SpFRET	Single-pair Fluorescence Resonance Energy Transfer
STED	Stimulated Emission Depletion
TBP	TATA-Box Binding Protein
TCSPC	Time-Correlated Single Photon Counting
TDP	Transition Density Plot
TEM	Transmission Electron Microscope
TIFF	Tagged Image File Format
TIRFM	Total Internal Reflection Fluorescence Microscopy
ZMW	Zero Mode Waveguide

Acknowledgements

Looking back at my journey as a PhD student, I am very grateful for all the support, guidance and motivation that I received on the way.

I thank my advisor, Prof. Don C. Lamb, for the opportunity to work in his group and his support throughout my work here. The freedom to plan projects independently and the trust he put into me have helped me learn and grow as a scientist. His dedication to science and his insight into the tiniest details without losing focus of the bigger picture continue to inspire me.

The work in this thesis is the result of many wonderful collaborations. I thank Promod Pratap for introducing me to the Sodium pump and being my co-advisor. Without his support, this project would not have been possible. His unique biologist's perspective, his ever smiling demeanor and his bag-full of anecdotes made it wonderful to work with him. He always found the right words at the right time and kept his patience and optimism high despite the project being quite challenging with many hurdles on the way.

I thank Prof. Friedrich Simmel, my collaborator for the DNA origami based projects, for providing me insight into this very exciting and interesting field. It was wonderful working with Anna Kostina who designed the constructs for the CHA project and trusted me with her precious data in the Threading project. I am enormously thankful to Enzo Kopperger for bringing me on-board the fantastic DNA Roboarm project, which proved to be an exciting application to showcase my work on multi-color FRET. All the discussions and brain-storming sessions were worthwhile in making this project a success.

I am very grateful to Prof. Braeuchle, Prof. Hartschuh and Prof. Biel for agreeing to be on my thesis committee.

I thank Cindy from AK Bein and, AK Ladurner and AK Beckmann for letting me use their instruments.

I gratefully acknowledge the International Doctorate Program - Nanobiotechnology (IDK-NBT) of the Elitenetzwerk Bayern, the Center for Nanoscience (CeNS) and the SFB 1032 and SFB 1035 for funding, and networking opportunities. I am especially thankful to Marilena Pinto, program manager for IDK-NBT, who helped me settle in Munich and guided me through all the bureaucratic procedures. I also want to thank Prof.

Raedler and Dr. Susanne Hennig for the support from CeNS.

I felt very lucky to have an amazing working environment and peers from whom I learned a lot. Settling into a new country, new culture, new workplace is a bitter-sweet experience and I am thankful to everyone who helped me feel at home. I thank Gregor for introducing me to the TIRF system and smFRET experiments. His TRACY program has been very valuable for analysis. I thank Martin for answering all my questions and keeping patience with me. I thank Walde for the umpteen discussions that have helped me enormously improve. I am very thankful to Anders for all the help with the MFD-PIE data analysis and our many discussions to make the multi-color analysis routines better. I am very thankful to Lena for all the support, motivation and discussions, and for keeping the Biolab running. I am extremely thankful to Ellen for all the support and motivation during thesis writing and especially her super delicious cakes that brighten up even the gloomiest days. I thank Ganesh for all the late evening discussions about science and life. I am thankful to Evelyn for her insight into dye photophysics and for all the motivation. I thank Giulia and Wehne for the best times we had in our 'cool-office'. I thank Aurelie for insight into french culture and our shared interest in Nutella-substrate adsorption. I thank Jens, Daniela and Ivo for sharing their office with me. I thank Baessem for helping with the setup and taking care of the printers, and Philipp for writing the synchronisation software. I thank Adriano for making the lunch and coffee times fun. I would have totally missed out on bavarian culture and Star Wars if not for Niko. It was a pleasure to mentor Sebastian during his F-pratikum, for which I am very thankful. I thank Maria, Kira, Chen and Nader for all the lively discussions. I thank Frau Steger, Kasia, Viola, Doro, Kulpreet, Jelle, Matthias, Alvaro and Alex for support on various occasions. Special thanks to Moritz for guiding me through all the bureaucratic regulations, taking care of the contracts and giving me 'sound' advice. I am thankful to all the people in AK Lamb, AK Braeuchle and AK Michaelis for a great working environment. I thank also the people in the metal workshop and electronics workshop for building parts for the setup.

I am superbly thankful to Tanvi, Sneha, Anirudh, Aalekhya, Ramya and Aradhana for being my cheer-leaders and constantly motivating me to do my best.

Lastly, I feel extremely grateful for having a loving and caring family who have supported me throughout. I thank my parents, grandparents, uncle and aunt for being there always. This would not have been possible without them.