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

Single-molecule fluorescence studies of Protein Folding and Molecular Chaperones

Martin Sikor

aus

Weilheim i. OB

2011

Erklärung

Diese Dissertation wurde im Sinne von §13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 (in der Fassung der sechsten Änderungssatzung vom 16. August 2010) von Herrn Professor Don C. Lamb, PhD betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 08.11.2011

_____

(Unterschrift des Autors)

Dissertation eingereicht am: 15.11.2011
1. Gutachter: Prof. Dr. Don C. Lamb
2. Gutachter: Prof. Dr. Christoph Bräuchle
Mündliche Prüfung am: 22.12.2011

Abstract

Folding of newly synthesized proteins is an essential part of protein biosynthesis and misfolding can result in protein aggregation which can also lead to several severe diseases. Protein folding is a highly heterogeneous process and rarely populated intermediate states may play an important role. Single-molecule techniques are ideally suited to resolve these heterogeneities. In this thesis, I have employed a variety of single-molecule fluorescence spectroscopy techniques to study protein folding using model systems on different levels of complexity. The acidic compact state (A state) of Myoglobin is used as a model system of a protein folding intermediate and is studied by a combination of molecular dynamics (MD) simulations and several fluorescence spectroscopic techniques. Using two-focus fluorescence correlation spectroscopy (FCS), it is shown that the A state is less compact than the native state of myoglobin, but not as expanded as the fully unfolded state. The analysis of exposed hydrophobic regions in the acidic structures generated by the MD simulations reveals potential candidates involved in the aggregation processes of myoglobin in the acidic compact state. These results contribute to the understanding of disease-related fibril formation which may lead ultimately to better treatments for these diseases.

A huge machinery of specialized proteins, the molecular chaperones, has evolved to assist protein folding in the cell. Using single molecule fluorescence spectroscopy, I have studied several members of this machinery. Single-pair fluorescence resonance energy transfer (spFRET) experiments probed the conformation of the mitochondrial heat shock protein 70 (Hsp70), Ssc1, in different stages along its functional cycle. Ssc1 has a very defined conformation in the ATP state with closely docked domains but shows significantly more heterogeneity in the presence of ADP. This heterogeneity is due to binding and release of ADP. The nucleotide-free state has less inter-domain contacts than the ATP or ADP-bound states. However, the addition of a substrate protein decreases the interaction between the domains even further simultaneously closing the substrate binding lid, showing that substrate binding plays an active role in the remodeling of Ssc1. This behavior is strikingly different than in DnaK, the major bacterial Hsp70. In DnaK, complete domain undocking in the presence of ADP was observed, followed by a slight re-compaction upon substrate binding. These differences may reflect tuning of Ssc1 to meet specific functions, i.e. protein import into mitochondria, in addition to protein folding. Ssc1 requires the assistance of several cofactors depending on the specific task at hand. The results of spFRET experiments suggest that the cofactors modulate the conformation of Ssc1 to enable it to perform tasks as different as protein import and protein folding.

Downstream of Hsp70 in the chaperone network, the GroEL/ES complex is a highly specialized molecular machine that is essential for folding of a large subset of proteins. The criteria that distinguish proteins requiring the assistance of GroEL are not completely understood yet. It is shown here that GroEL plays an active role in the folding of double-mutant maltose binding protein (DM-MBP). DM-MBP assumes a kinetically trapped intermediate state when folding spontaneously, and GroEL rescues DM-MBP by the introduction of entropic constraints. These findings suggest that proteins with a tendency to populate kinetically trapped intermediates require GroEL assistance for folding. The capacity of GroEL to rescue proteins from such folding traps may explain the unique role of GroEL within the cellular chaperone machinery.

Contents

List of Figures

List of Tables

1 Introduction

Physics and biology coexist as individual disciplines only since the 19th century. Before, they were combined as two aspects of natural philosophy. Physics is the description of nature using a mathematical framework aimed at thorough understanding of the behavior of the entire universe. Biology has been successful in the phenomenological description of the immensely complex processes enabling life. Although distinguished disciplines, physics and biology have contributed to each other's development and interdisciplinary research is becoming more and more important again. Biophysical research and the application of physical models and methods to biological questions has developed into a thriving field that has led to significant advances in our knowledge about processes in living systems and has also sparked the development of several physical methods.

A major contribution of physics to the ability to study biological systems has been the development of microscopy. It allows direct observation and quantification of phenomena not resolvable with the human eye. Over the 20th century, techniques have been developed to further and further push down the observation length scale. Electron microscopy, X-ray crystallography and nuclear magnetic resonance (NMR) techniques have reached atomic resolution. However, they suffer from significant limitations like the need to put the sample into a non-native environment or the lack of dynamic information. This gap is filled by fluorescence microscopy and spectroscopy which can yield time-resolved information of specifically labeled molecules in their natural environment. For a long time, the spatial resolution of fluorescence microscopy was limited to 200 nm by the *Abbe limit*. Recent advances have pushed the resolution limit down to tens of nanometers (Huang, 2010). However, intramolecular length scales are still out of reach even for these *super-resolution* fluorescence microscopy techniques.

Contrary to fluorescence microscopy, *fluorescence spectroscopy* does not rely on the spatial resolution of molecules but derives molecular properties from their fluorescence emission. The molecule of interest is either fluorescent on its own, or is labeled by a genetically expressed or synthetic fluorophore. The fluorescence intensity, lifetime and spectrum then yield information about the molecule and its environment (Section 2.1). Structural information is accessible by distance measurements using *Förster resonance energy transfer* (FRET, section 2.2) experiments. FRET is sensitive at typical length scales of biomolecules and has been used successfully to study structural changes of biomolecules including the dynamics on timescales down to µs (Hertzog *et al.*, 2004) and on the single molecule level (Hamadani and Weiss, 2008; Hofmann *et al.*, 2010).

The technological advances naturally also resulted in countless new biological insights. The most exciting and influential advancement made possible by biophysical techniques may have been the discovery of the double-helical structure of DNA by Watson and Crick (Watson and Crick, 1953). This discovery immediately explained how genetic material can be inherited reliably during cell division. By now, biophysical methods have contributed not only to the research of DNA structure and protein biosynthesis but have played a key role in solving the structure of proteins using X-ray crystallography and NMR. For many proteins, the native structure is today known down to a few Å resolution.

The transition from the nascent, unfolded chain to the final, native, three-dimensional structure, however, is a highly dynamic and inhomogeneous process that has remained enigmatic for several decades. Anfinsen realized already in 1957 (Sela *et al.*, 1957) that the three-dimensional structure of a protein is uniquely determined by the one-dimensional sequence of amino acids in a protein. At the same time, it is impossible for a protein to reach its native state on a biologically relevant timescale by a purely random search (*Levinthals paradox*, Levinthal, 1969). Even today, there is still no clear picture of a unified protein folding theory (section 3.2). It has become clear in the last two decades that while most proteins can fold on their own in an isolated *in vitro* environment, in the crowded environment of the cell, a specialized class of proteins, molecular chaperones, has to aid them towards their native state.

The highly heterogeneous nature of the protein folding process stages a significant difficulty when using ensemble fluorescence spectroscopy methods. Subpopulations may be hidden in the ensemble average, and temporal synchronization of dynamic events is not always possible, especially due to the stochastic nature of the microscopic processes involved. These problems can be circumvented by the use of techniques that avoid ensemble averaging by the observation of one single molecule at a time. Fluorescence detection has reached the necessary sensitivity to detect single fluorophores in the 1990s, and since then, single-molecule fluorescence spectroscopy has become a common, powerful tool to study protein folding and molecular chaperones.

I employed single-molecule fluorescence spectroscopy, in particular single-pair FRET (spFRET) and fluorescence correlation spectroscopy experiments, to study protein folding and the role of molecular chaperons using different model systems. The basics of fluorescence spectroscopy (Chapter 2) and an overview of protein folding and molecular chaperones (Chapter 3) are given followed by a thorough description of the single-molecule methods employed in the research for this thesis (Chapter 4). Then, the acidic compact state of Myoglobin is studied as a model for the molten globule state, a common protein folding intermediate (Chapter 5). The acidic compact state of Myoglobin shows features that may help explaining the disease-related formation of amyloid fibrils. Chapter 6 presents the conformational dynamics of a the major Hsp70 in mitochondria, Ssc1, along its functional cycle on different timescales from milliseconds to several seconds. Using spFRET, it is shown that Ssc1 has important differences compared to bacterial Hsp70, possibly related to its more diverse functionality. In Chapter 7, spFRET burst analysis and Fluorescence Correlation Spectroscopy (FCS) are used to clarify how GroEL, the major molecular chaperone in bacteria, assists in protein folding by modulation of the energy landscape of a kinetically trapped folding intermediate.

2 Fluorescence spectroscopy

2.1 Fluorescence

Fluorescence was first described and documented in the middle of the 19th century by Sir John Frederic William Herschel (Herschel, 1845). It is a special category of *Luminescence* and characterized by the transition of a molecule from an optically excited singlet state to the ground state by the emission of a photon (Lakowicz and Masters, 1999). This transition is spin-allowed and can occur with fast rates on the order of $10^8\,\mathrm{s}^{-1}$. The processes involved in fluorescence are usually visualized by a *Jablonski diagram*. A typical Jablonski diagram is shown in Figure 2.1 with S_0, S_1 and S_2 denoting singlet ground and excited states and T_1 denoting an excited triplet state.

Each electronic state is divided into vibrational substates that decay typically by internal conversion to the lowest vibrational level within 1 ps. Due to the small mass of the electron compared to the mass of the nuclei, the electronic transitions are fast compared to the movement of the nuclei. The positions of the nuclei thus do not change during the electronic transition (*Born-Oppenheimer approximation*). Because the equilibrium positions of the nuclei are different in the same vibrational level of different electronic states, usually a molecule in the lowest vibrational state of S_0 is excited upon absorption of a photon to higher vibrational levels of S_1 or S_2. The probabilities for these transitions are given by the Franck-Condon principle. Excitation is followed by internal conversion to the lowest vibrational state of S_1. Internal conversion in most molecules is complete after 1 ps and fluorescence is emitted by a transition from the lowest vibrational state of S_1 to a higher vibrational level of S_0. This results in several important characteristics of typical fluorescence absorption and emission spectra. First, since the transition from the S_1 state to the S_0 state always originates from the lowest vibrational level, the emission spectrum of a fluorophore is independent of the excitation wavelength (*Kasha's rule*). Second, the vibrational levels of the S_0 and S_1 states have a similar structure. Thus, emission spectra are an approximate mirror image of the absorption spectra (*Mirror image rule*). Third, the emission maximum is shifted to longer wavelengths compared to the absorption maximum (*Stoke's shift*). Spectra for a typical fluorophore (Atto 532, Atto-Tec) are shown in Figure 2.2. The absorption and emission spectra are connected to the rate of photon emission Γ by the *Strickler-Berg* relationship (Strickler and Berg, 1962),

$$\Gamma = 2.88 \cdot 10^{-9} n^2 \frac{\int F(\overline{\nu}) d\overline{\nu}}{\int \frac{F(\overline{\nu}) d\overline{\nu}}{\overline{\nu}^3}} \int \frac{\epsilon(\overline{\nu}) d\overline{\nu}}{\overline{\nu}} \tag{2.1}$$

where $\overline{\nu}$ is the wavenumber, $F(\overline{\nu})$ is the emission spectrum of the dye, $\epsilon(\overline{\nu})$ is the absorption spectrum of the dye, and n is the refractive index of the medium. The integrals are calculated over the $S_0 \leftrightarrow S_1$ absorption and emission bands.

Figure 2.1 also shows intersystem crossing of the fluorophore to a triplet state T_1. This transition requires a change in the spin of the electron. Hence, it is spin-forbidden and takes place through weaker interaction and thus on slower timescales. De-excitation of T_1 to S_0 (*Phosphorescence*) is also spin-forbidden and takes place on timescales several orders of magnitude longer than fluorescence (10^{-3} s

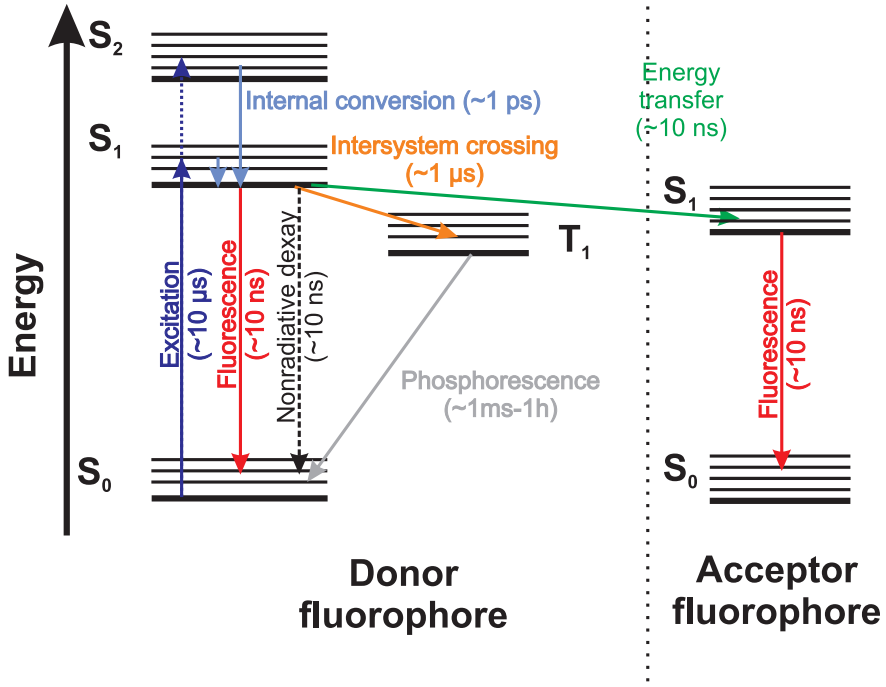


Figure 2.1: *Typical Jablonski diagram. The singlet ground and excited states S_0, S_1, and S_2 of the donor are shown on the left side with excitation and de-excitation pathways and their typical timescales. In the center, an excited triplet state T_1 is shown. On the right side of the figure, the ground state S_0 and first excited state S_1 of an acceptor for FRET is shown together with the energy transfer pathway from the donor S_1 state to the acceptor S_1 state.*

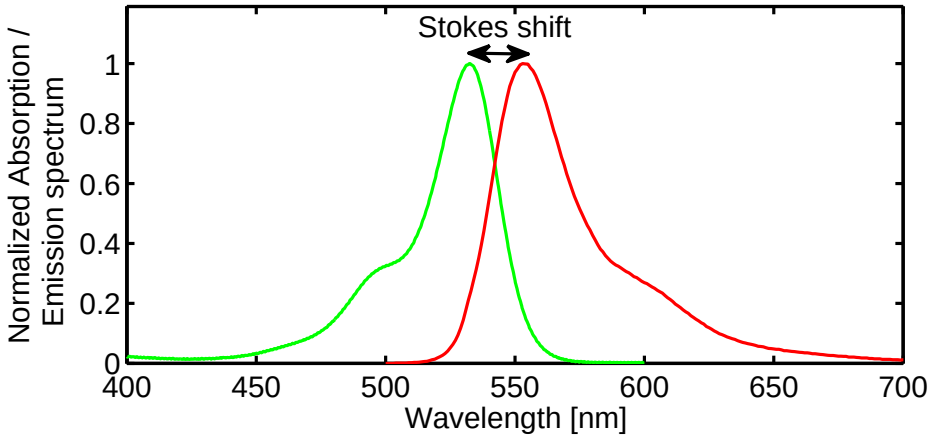


Figure 2.2: *Fluorescence absorption and emission spectra of a typical dye used in this thesis, Atto 532 (Atto-Tec, spectra downloaded from* www.atto-tec.de*).*

- 1 h). Another process depicted in Figure 2.1 is resonance energy transfer to another fluorophore. This process will be discussed in more detail in section 2.2. Finally, non-radiative processes can de-excite the fluorophore from the S_1 to the S_0 state without photon emission.

Because both radiative and non-radiative decay processes depopulate the S_1 state, not every absorbed photon will result in a fluorescence photon. The ratio of the number of photons emitted by a fluorophore to the total number of photons absorbed is called the fluorescence quantum yield. It can be calculated from the rates of non-radiative decay k_{nr} and fluorescence emission k_f,

$$\phi = \frac{k_f}{k_f + k_{nr}} \tag{2.2}$$

The quantum yield is closely related to another important characteristic of a fluorophore, the lifetime of the excited state $\tau = \frac{1}{k_f + k_{nr}}$. The lifetime τ is directly accessible in single-molecule burst analysis experiments (section 4.1.5.3) and thus reports on changes of the fluorescence quantum yield due to additional quenching processes.

2.2 Fluorescence resonance energy transfer

An excited fluorophore (*donor*) can transfer its energy to another fluorophore in close proximity (*acceptor*) by dipole-dipole interactions. This process is referred to as either *Fluorescence resonance energy transfer* or *Förster resonance energy transfer* (FRET). The FRET efficiency E is defined by the ratio of the energy transfer rate k_{FRET} to the sum of all rates depopulating the excited state of the donor,

$$E = \frac{k_{FRET}}{k_{FRET} + k_f + k_{nr}} \tag{2.3}$$

The rate of energy transfer k_{FRET} depends strongly on the distance between the donor and the acceptor. The near field of the donor decays proportional to the third power of the distance. The transition dipole of the acceptor is induced by the donor near field, resulting in a dependence of k_{FRET} on the square of the donor field and thus on the sixth power of the donor-acceptor separation R (Figure 2.3A).

$$k_{FRET} = \frac{1}{\tau_D} \left(\frac{R_0}{R} \right)^6 \tag{2.4}$$

where τ_D is the donor lifetime without energy transfer. The Förster radius R_0 combines the other factors affecting the energy transfer rate k_{FRET} and is given by

$${R_0}^6 = \frac{9 \cdot \ln 10 \cdot \kappa^2 \, \phi_{D_0} \, J}{128 \, \pi^5 \, n^4 \, N_A} \tag{2.5}$$

with the refractive index of the sample n, the quantum yield of the donor in the absence of the acceptor ϕ_{D_0}, and Avogadro's number N_A. The overlap integral J describes the overlap between the donor emission and the acceptor absorption spectra (Figure 2.3A) and is given by

$$J = \int f_{\mathrm{D}}(\lambda) \, \epsilon_{\mathrm{A}}(\lambda) \, \lambda^4 \, d\lambda \tag{2.6}$$

with the normalized donor emission spectrum $f_D(\lambda)$ and the wavelength dependent extinction coefficient of the acceptor $\epsilon_A(\lambda)$. When the wavelength is given in nm and the extinction coefficient is given in $\mathrm{M^{-1}cm^{-1}}$, the following simplified expression for R_0^6 in units of $\mathrm{Å}^6$ is obtained:

$${R_0}^6 = 8.79 \cdot 10^{-5} \cdot (\kappa^2 \, n^{-4} \, \phi_D \, J) \tag{2.7}$$

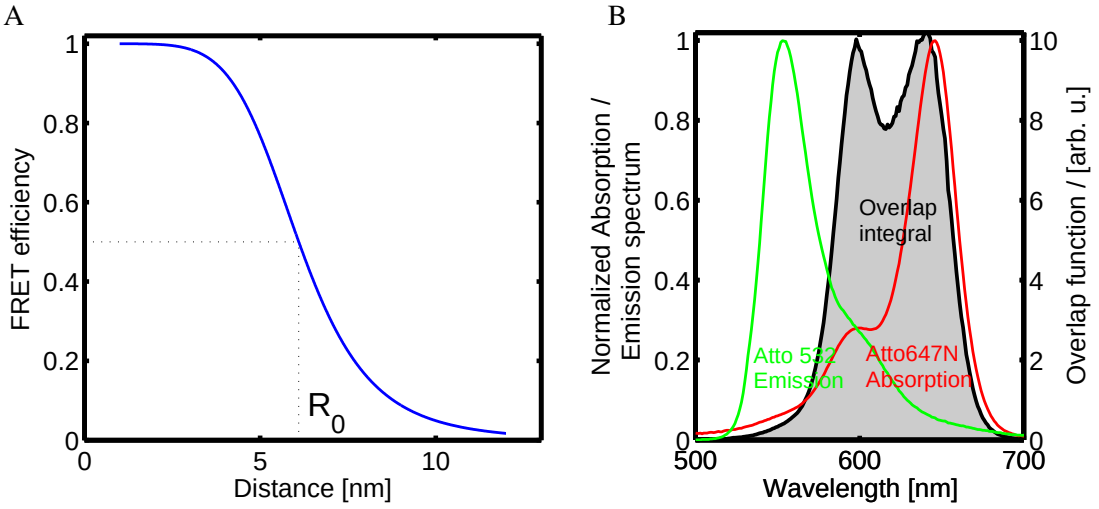


Figure 2.3: *Dependence of FRET on the distance between the fluorophores and their spectral overlap. (A) Distance dependence of the FRET efficiency E for the dye pair Atto 532 and Atto 647N with a Förster radius $R_0 = 6.1\,\mathrm{nm}$. The Förster radius is the distance where $E = 0.5$. (B) Visualization of the overlap integral J (grey area) of the emission spectrum of Atto 532 and the absorption spectrum of Atto 647N. J is defined in equation 2.6.*

Another important parameter for R_0 is κ^2 (Figure 2.4). Dipole-dipole interactions depend critically on the relative orientation of the donor emission and the acceptor excitation dipoles. For angles θ_{DA} between the planes defined by the donor emission dipole and the vector between the dipoles and defined by the acceptor excitation dipole and the vector between the dipoles, and θ_D and θ_A between the respective dipoles and the vector connecting donor and acceptor, κ^2 is given by

$$\kappa^2 = (\cos\theta_{DA} - 3\cos\theta_D \cos\theta_A)^2 \tag{2.8}$$

Theoretically, κ^2 can take values between 0 and 4. A common approach in fluorescence spectroscopy is to assume an isotropic sample with orientational averaging resulting in a value of $\kappa^2 = \frac{2}{3}$ (Dale *et al.*, 1979). However, this assumption may not be valid due to limited freedom of movement if the fluorophores are attached to a protein. The availability of additional time resolved anisotropy information from donor and acceptor using PIE allows the calculation of maximum errors of κ^2 (section 4.1.5.6).

Combining equations 2.3 and 2.4, the FRET efficiency E can now be expressed as a function of donor-acceptor separation R

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{2.9}$$

E can be calculated from properly corrected values (section 4.1.5.4) of the donor intensity I_D and acceptor intensity I_A. Additionally, it is clear from the Jablonski diagram (Figure 2.1) that FRET decreases the lifetime of the excited state of the donor by providing an additional deexcitation pathway. It is thus possible to use the donor lifetime in the presence of the acceptor τ_{D_A} and in the absence of the acceptor τ_{D_0} as an alternative way for the calculation of E.

$$E = \frac{I_A}{I_D + I_A} = 1 - \frac{\tau_D}{\tau_{D_0}} \tag{2.10}$$

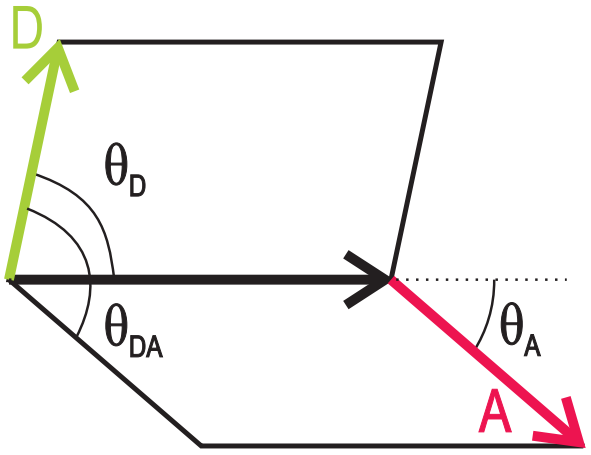


Figure 2.4: *The angles involved in the calculation of the orientation factor κ^2. The donor emission dipole D is given by the green arrow, the acceptor absorption dipole A is given by the red arrow. The vector connecting the donor emission and acceptor absorption dipoles is given in black. The angles θ_D and θ_A are the angles between the line connecting the dipoles and the dipoles D and A, respectively. θ_{DA} is the angle between the plane defined by D and the vector between the dipoles and the plane defined by A and the vector between the dipoles.*

2.3 Fluorescence anisotropy

As mentioned above, the measurement of quantitative distances using FRET depends critically on knowledge of the rotational mobility of the dyes to justify the assumption of $\kappa^2 = \frac{2}{3}$. The mobility of a dye can be measured by fluorescence anisotropy experiments. Absorption is maximum if the absorption dipole is parallel to the excitation light. During the lifetime of the excited state, the fluorophore orientation changes due to rotational diffusion. The polarization of the emitted photon will thus be different from the polarization of the excitation light. The correlation between excitation polarization and emission polarization is described by the fluorescence anisotropy. Averaged over the time of the emission, the steady-state anisotropy is defined by

$$r = \frac{I^{\parallel} - I^{\perp}}{I^{\parallel} + 2 \cdot I^{\perp}}, \tag{2.11}$$

where $I^{\parallel}$ and $I^{\perp}$ are the intensities of fluorescence light polarized parallel and perpendicular to the excitation light, respectively. The steady-state anisotropy contains valuable information about the sample. Freely diffusing dye molecules have rotational correlation times of less than a nanosecond. Thus, over the lifetime of the dyes, all orientational information is lost and the anisotropy is close to 0. For labeled proteins, however, typical rotational correlation times are on the order of tens or even hundreds of nanoseconds. Depending on the mobility of the dyes along their linkers, the anisotropy can be significantly higher allowing the discrimination of free dye and labeled protein.

Additional information is available from time-resolved anisotropy experiments. In time-resolved anisotropy experiments, the anisotropy is measured as a function of the lag time t after an excitation pulse,

$$r(t) = \frac{I^{\parallel}(t) - I^{\perp}(t)}{I^{\parallel}(t) + 2 \cdot I^{\perp}(t)} \tag{2.12}$$

The time-resolved anisotropy can be used to identify different components contributing to the movement of the dyes. In particular, for dye-labeled proteins, it is possible to separate the mobility of the

dye on the protein from slower movement of the protein itself. Knowledge of the mobility of both dyes in spFRET experiments then allows an error estimation for κ^2 as described in section 4.1.5.6.

2.4 Single-molecule fluorescence

Ensemble fluorescence spectroscopy can yield several important molecular properties of a homogeneous, static sample (Figure 2.5A). For samples which consist of several different subpopulations, however, only an average value for the experimental observables is available. Circumvention of this limitation is cumbersome, depending on the nature of the heterogeneity, and not always possible. Static heterogeneity, i.e. the coexistence of several substates without interconversion on the timescale of the measurement, may be resolved if it is possible to separate the different species before the spectroscopic measurement, for example by gel electrophoresis or chromatography (Figure 2.5B). Dynamic heterogeneity, i.e. the dynamic switching of a protein between different states over the course of the experiment, is also impossible to detect in an ensemble measurement (Figure 2.5C). Synchronization of all sample molecules at the beginning of the experiment may allow the detection of dynamics. However, synchronization is not always possible, and the stochastic nature of transitions between states will result in desynchronization after some time (Figure 2.5D). Reducing the observed "ensemble" to a single molecule immediately solves the above mentioned problems. Now, subpopulations can be easily resolved and also dynamic switching between different states can be observed.

The detection of the fluorescence from a single molecule (or even a single fluorophore) is technically difficult. The first experiments where fluorescence was used to detect the presence of single molecules was in 1961, when Rotman (1961) detected the fluorescent product of single enzymes encapsulated in separated droplets. Hirschfeld (1976) was the first to be able to detect labeled single molecules, antibodies directly tagged with 80-100 fluorophores. After that, it took another almost 15 years until the first detection of a single fluorophore in a biologically relevant environment (Brooks Shera *et al.*, 1990) and another five years to the first detection of FRET between a single pair of fluorophores (Ha *et al.*, 1996). Besides technical advances that increased the sensitivity of the detectors, the major step leading to the detection of the signal from a single fluorophore is efficient suppression of background by a reduction of the detection volume. Mainly two microscope schemes provide sufficiently small detection volumes and are thus especially well suited for single fluorophore detection. Total internal reflection microscopy uses an evanescent wave for fluorescence excitation thus reducing the detection volume to a thin layer of $\sim 100\,\mathrm{nm}$ thickness (Section 4.3). This allows simultaneous observation of a complete field-of-view but is limited to surface-immobilized samples. In contrast, confocal microscopy is limited to the detection of a single spot and requires scanning for acquisition of a complete image. This spot, though, can be placed anywhere in the sample. It is created by focusing of laser light by an objective and eliminating out-of-focus light by a pinhole (Section 4.1.1). Prior to detailed description of the single-molecule fluorescence methods and the corresponding data analysis in chapter 4, the next chapter gives a short overview of the biological content of this thesis, protein folding and molecular chaperons.

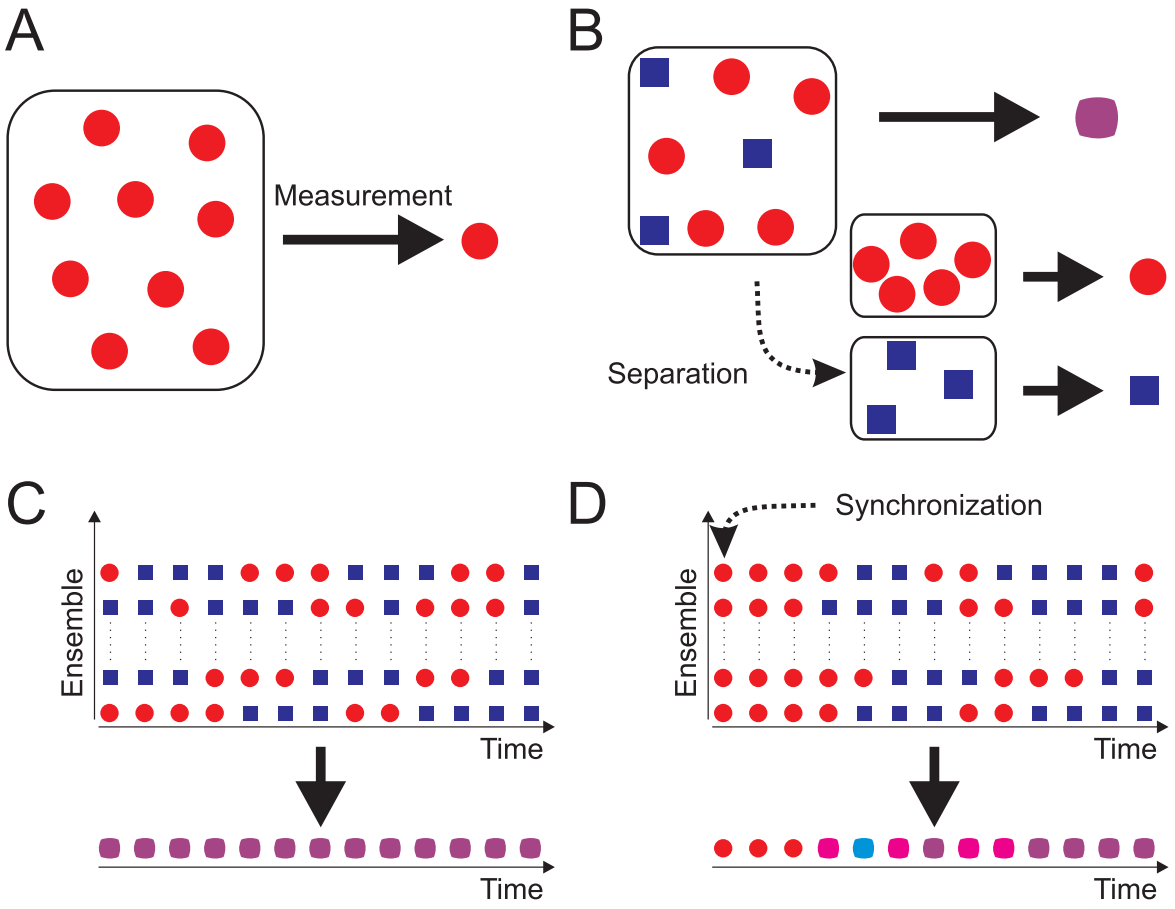


Figure 2.5: *Advantages of single-molecule experiments. (A) An ensemble measurement of a homogenous sample yields the correct result. (B) Static heterogeneity in the observed ensemble results in measurement of an average value. Separation before the spectroscopic measurement may be able to resolve the problem. (C) Dynamic heterogeneity in the observed ensemble is completely hidden in the bulk measurement. For all times, an average value of the ensemble is observed. (D) Synchronization may allow resolving dynamic heterogeneity. However, the stochastic nature of biological microscopic processes will result in de-synchronization after some time.*

3 Protein folding and molecular chaperones

3.1 Protein structure

Living organisms generally consist of four different classes of molecules. While lipids form membranes to provide specialized compartments, carbohydrates provide energy, and nucleic acids are used mainly for information storage, most of the work in a cell is done by proteins (Frauenfelder *et al.*, 2010). Proteins are polypeptides built from twenty different naturally occurring amino acids. Amino acids are chiral molecules having four different side groups on a central carbon atom. Two of these side groups, an amine group and a carboxylic acid group, form the peptide bonds in a protein and the third side group is a single hydrogen atom. The different natural amino acids differ by the forth substituent on the central carbon atom which can be charged, polar, or hydrophobic. Each protein has a specific, unique order of amino acids determining the *primary structure* of the protein (Figure 3.1A).

The peptide group has two resonance forms resulting in a partial double-bond between the C atom of the carboxyl group and the N atom of the next amino acid. Thus, the central C atoms of two neighboring amino acids are coplanar and can exist only in two states, *cis* and *trans* state. The *trans* state is prevalent in natural amino acids at a ratio of approximately 1000:1. The two other dihedral angles of the freely rotating single bonds determine the *secondary* and *tertiary* structure of the protein.

Local segments of a protein can form specific three-dimensional structures, usually referred to as *secondary structure*. The secondary structure is determined by the relative orientation of the intra-chain hydrogen bonds. The most common secondary structures in proteins are the α*-helix* (Figure 3.1B, left) in which every backbone N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues earlier, and the β*-sheet*, where separate parts of the chain form β strands and several β strands are then connected by two or more hydrogen bonds between neighboring strands (Figure 3.1B, right).

The three-dimensional arrangement of the secondary structure elements is called the *tertiary structure* (Figure 3.1C). The correct tertiary structure is essential for proper function of the protein, and is already determined by the primary structure of the protein (*Anfinsen's dogma*). Often, multiple protein subunits, each one an independent polypeptide chain, form a functional complex. This structural level is called *quaternary structure* (Figure 3.1D).

3.2 Protein biosynthesis and folding

Proteins perform a variety of functions and play a key role in essentially every aspect of cellular reproduction, survival, and even death. They can transport small molecules, ions, and other proteins, produce ATP, and of course they are also involved in the production of other proteins.

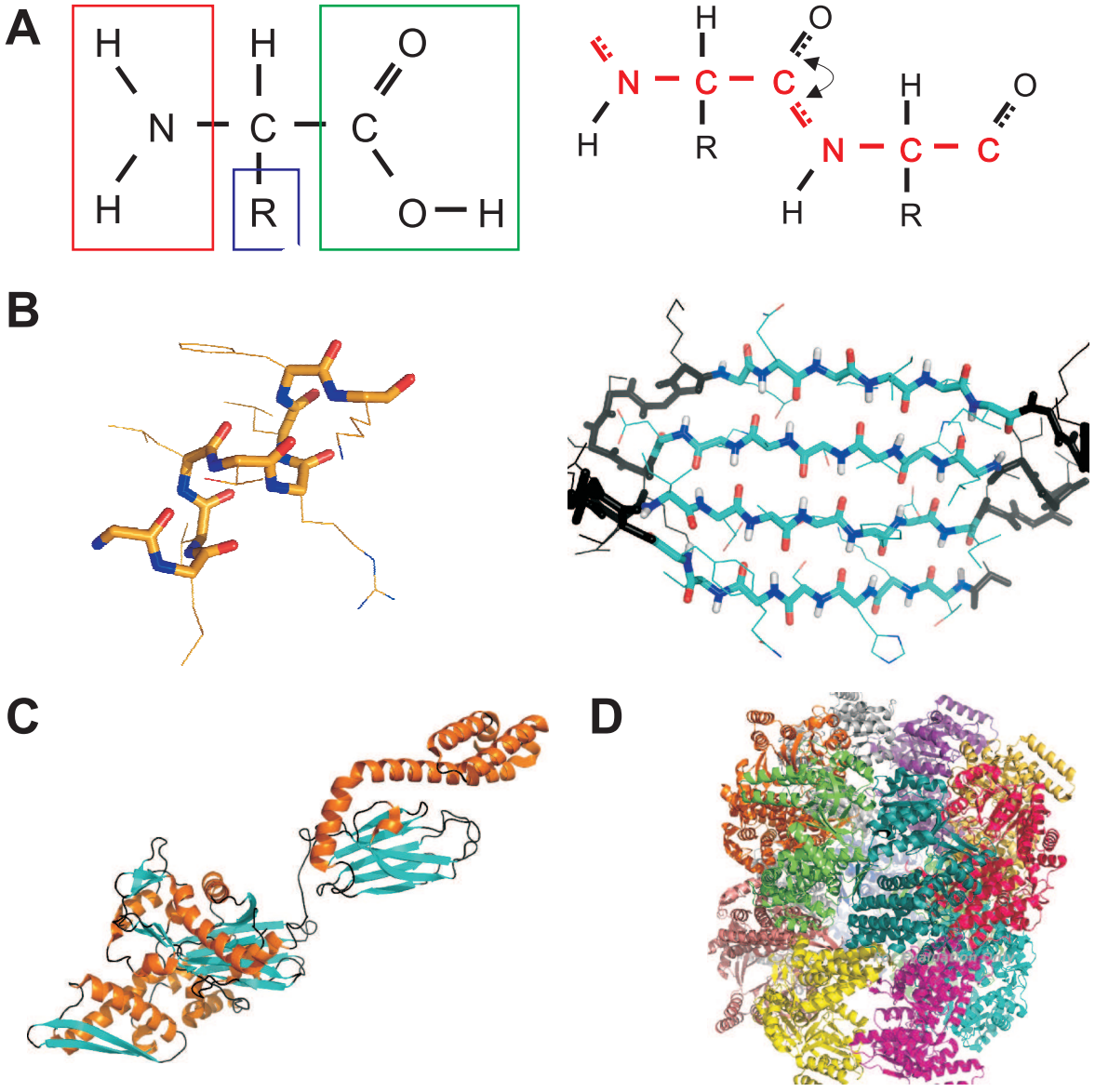


Figure 3.1: *Levels of structural organization in proteins. (A) Primary structure. Left: Structural formula of an α-amino acid with an amino group (red), a carboxyl acid group (green) and the substituent R (blue). Right: Peptide bond between two amino acids. The peptide backbone is marked in red. The two resonance structures of the peptide bond are indicated by dotted lines in the involved double bonds and an arrow. (B) Secondary structure. Left: A 9-residue α-helix with two complete turns. The peptide backbone is indicated as bold sticks while the side groups are shown as lines. Each oxygen (red) forms a hydrogen bond with the nitrogen (blue) four residues ahead. Right: Four β-strands of 5-6 residues forming an antiparallel β-sheet. Here, hydrogen bonds are formed between N and O atoms of neighboring strands. (C) Tertiary structure. Tertiary structure of the* e. coli *Hsp70 DnaK. α-helical secondary structure elements are shown in orange, β-helical secondary structure elements in cyan, and regions without secondary structure in black. (D) Quaternary structure. Arrangement of 14 identical subunits in the chaperonin GroEL. Different subunits are shown in different colors. (B-D) Structures were downloaded from the protein data bank (PDB,* www.pdb.org*), data files 1MYT (B left, Yellowfin Tuna Myoglobin), 2KHO (B right and C, DnaK) and 2NWC (D, GroEL) and visualized using PyMol (DeLano Scientific).*

Protein manufacturing or biosynthesis starts by the transcription of DNA to RNA by the *Polymerase* protein complex, followed by translation to a polypeptide chain in the *Ribosome*. Then, the nascent chain has to fold to a unique native structure characterized by a free energy minimum to be fully functional. Christian B. Anfinsen showed by experiments on Ribonuclease that the native structure is uniquely determined by the amino acid sequence (Sela *et al.*, 1957). More than 50 years later, however, it is still not clear how proteins can reach the native structure on a biologically relevant timescale starting from a random conformation. A random search of all possible conformations would take several orders of magnitude longer than the age of the universe, even for a small 150 residue protein and as little as 1 fs per search step (Levinthal, 1969, *Levinthal's paradox*). Faster folding can be explained by a funneling mechanism into the minimum of the free energy landscape. This transition can either occur by pure downhill folding or via one or several intermediate states. One model assumes an initial collapse to a *molten globule* state due to hydrophobic interactions of the side chains, followed by diffusion, collision and docking of the secondary structure elements finally resulting in the native state (Diffusion-Collision Model Baldwin, 1989; Karplus and Weaver, 1994)). Alternatively, the early stage of folding may be dominated not by hydrophobic collapse but formation of secondary structure elements. A second model, the nucleation-condensation model, proposes that secondary and tertiary structure are stabilized at the same time. There, most residues contribute towards the stability of the folding nucleus (Itzhaki *et al.*, 1995). Which model dominates folding for a particular sequence seems to depend on the balance between the amount and stability of secondary structure and the propensity of the polypeptide chain to undergo hydrophobic collapse (Daggett and Fersht, 2003).

3.3 Molecular chaperones

The models described in the last section are well suited to characterize protein folding in vitro, i.e. in an isolated environment. However, the crowded environment of the cell stages an additional problem for efficient protein folding. Unfolded proteins exhibit hydrophobic surfaces that can result in aggregation as an alternative pathway to productive folding with possibly higher yield. Uncontrolled aggregation of proteins can result in several severe diseases like Alzheimer's, Huntington's, or Parkinson's disease (Selkoe, 2004). Only small proteins can fold efficiently *in vivo*. To ensure high yields for folding even in this highly crowded environment, a specialized class of proteins has evolved that prevents aggregation and assists proper folding of nascent proteins. These proteins are called *molecular chaperones*.

The terms *heat shock proteins* and *molecular chaperones* are often used in the same context. Although many heat shock proteins are molecular chaperones and vice versa, the terms are not equivalent. Heat shock proteins were discovered 15 years prior to the realization that some of them are involved in the folding of other proteins. As the name already suggests, their original discovery was in *Drosophila* cells undergoing heat shock (Tissieres *et al.*, 1974) where they are expressed at high levels. For several years, it was believed that they play an important role in the stress response of cells but were not involved in regular "housekeeping", i.e. under normal environmental conditions. During the 80's, this changed gradually, e.g. due to the discovery that RuBisCo (Ribulose-1,5-bisphosphat-carboxylase/-oxygenase) binds to a previously unknown large protein complex before it reaches its functional state (Barraclough and Ellis, 1980). It was realized in 1989 that the heat shock protein 60 is required for the folding and assembly of other proteins in mitochondria (Ostermann *et al.*, 1989; Cheng *et al.*, 1989) and in *e. coli* (Goloubinoff *et al.*, 1989b,a). Only one year later, Kang *et al.* (1990) identified another protein necessary for protein import and folding in mitochondria, the mitochondrial heat shock protein 70.

In the following twenty years until today, huge research efforts were undertaken to understand the function of molecular chaperones. Molecular chaperones are a highly conserved class of proteins present in prokaryotes and in most organelles of eukaryotic organisms. They are classified according to their molecular weight and fulfill a variety of different functions.

A nascent polypeptide chain leaving the ribosome has to be treated differently depending on the protein being produced and its destination. On the one hand, it may be necessary for it to be transported to a target organelle like mitochondria or the endoplasmic reticulum (ER) . This is effected either cotranslationally into the ER (Brodsky, 1996) or posttranslationally into mitochondria (Kang *et al.*, 1990; Schneider *et al.*, 1994) by Hsp70s in the respective organelle which are essential parts of the translocation machineries. On the other hand, it may be required to fold the nascent protein in the cytosol. This may work spontaneously or with the help of chaperones or fail and result in a misfolded protein. In the case of terminal misfolding, the protein is transported by Hsp70s to the proteasome for degradation (Meacham *et al.*, 2001). Proper folding is facilitated also by Hsp70s, which bind unfolded proteins (Kang *et al.*, 1990), unfold misfolded proteins (Mayer and Bukau, 2005), and transport unfolded proteins to other molecular chaperones like GroEL (Langer *et al.*, 1992) or Hsp90s (Wegele *et al.*, 2004).

Hsp70 chaperones bind exposed hydrophobic patches of unfolded proteins in a nucleotide-dependent manner preventing aggregation and transporting them to destination organelles and other chaperones. The details of the functional cycle will be discussed in the introduction to chapter 6, which describes the results of single-molecule observations of this cycle.

The prototype Hsp60, the bacterial chaperone GroEL, has received by far the most attention and is probably the best understood molecular chaperone. GroEL is a tetradecamer of 57 kDa subunits arranged in two barrel-shaped heptamers sitting back to back. Each heptamer contains a cavity that is used to capture and bind unfolded proteins. GroEL encapsulates the substrate in cooperation with the cochaperone GroES in an ATP dependent manner, and the substrate is free to fold inside the cavity for the duration of ATP hydrolysis. The functional cycle of GroEL will be discussed in more detail in the introduction to chapter 7. Although many details of this functional cycle have been resolved, some fundamental questions are still subject of discussion. One of these, whether GroEL plays an active role or a passive role in protein folding, will be addressed in chapter 7. It is shown there that for the folding of DM-MBP, GroEL plays a more active role by rescuing DM-MBP from a kinetically trapped state.

4 Single-molecule methods

Detection of single molecules can be achieved by different methods, each of which has specific advantages and disadvantages. Single-molecule burst analysis on a confocal microscope (Section 4.1) allows the observation of molecules freely diffusing in solution. However, this restricts the observation time per molecule to the diffusion time of the molecules through the confocal volume on the order of 1-10 ms. Longer observation times are possible when immobilized molecules are observed. I have used total internal reflection (TIRF) microscopy to follow the FRET efficiency of immobilized proteins over timescales of seconds to minutes with a time resolution of down to 30 ms (Section 4.3).

4.1 Burst analysis with Pulsed Interleaved Excitation and Multi-parameter Fluorescence Detection

Single-pair FRET (spFRET) measurements of single molecules in solution were performed on a confocal system based on an inverted microscope (Nikon 200) using pulsed interleaved excitation (PIE) (Müller *et al.*, 2005) and Multi-parameter Fluorescence Detection (MFD) (Eggeling *et al.*, 2001). In this section, I will describe the principles of confocal microscopy, the excitation pathway including PIE, and the multi-parameter detection. The data that is taken on this setup contains an enormous amount of information and can be subjected to a wealth of different analysis methods described in Section 4.1.5. The results of burst analysis measurements to study the conformation of heat shock proteins 70 are then described in Chapter 6. In Chapter 7, I used spFRET burst analysis to characterize a kinetically trapped folding intermediate.

4.1.1 Confocal microscopy

Confocal microscopy was invented in the late 1950s by the American mathematician Marvin Minsky (Minsky, 1961). He used the combination of a lamp, a pinhole and a lens to illuminate a single spot in the sample. The light coming from this spot was detected by a point detector through a pinhole in the image plane of this spot (*confocal*) to reject light from different image planes. The sample was scanned over the detection spot to acquire an image. Of course, Minksy was limited by the technology of that time to slow scanning speeds and low detection efficency. Technological advances have increased the sensitivity and lowered the pixel dwell times in confocal microscopy, making it today an ideal tool for fluorescence microscopy. Of special importance was the invention of the laser providing a powerful, collimated, monochromatic light source. With non-collimated light sources, a pinhole in the excitation pathway had to be used to generate collimated light thus losing much of the excitation light.

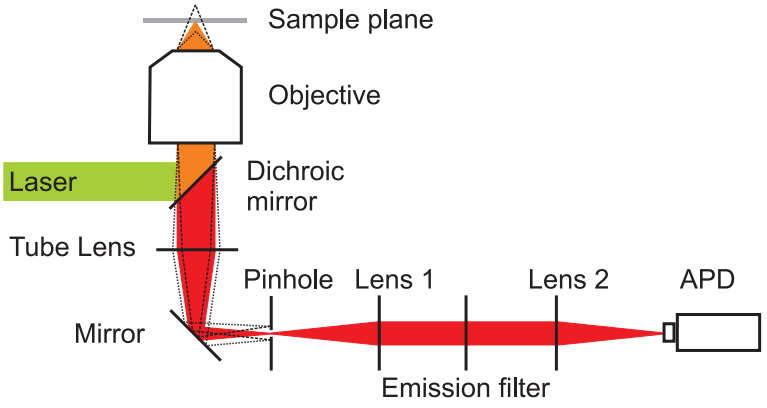


Figure 4.1: *Schematic of a general confocal setup. The excitation laser (green) is deflected into the objective by a dichroic mirror. Fluorescence from the sample plane (grey) passes the pinhole to the avalanche photo diode (APD) while light from above (dashed line) or below (dotted line) is blocked.*

The lateral resolution in confocal microscopy originates mainly from the focusing of collimated laser light by an objective. The refraction limit d for a focused laser beam of wavelength λ without a confocal pinhole is

$$d_{lat} = 0.61\frac{\lambda}{\mathrm{NA}} \tag{4.1}$$

where the numerical aperture $\mathrm{NA} = n \cdot sin(\alpha)$ describes the maximum angle of the light cone the objective can observe, with the refractive index of the immersion medium n and the maximum collection angle of the objective α. For the water immersion objectives used in this work, $n = 1.33$ and $\mathrm{NA} = 1.2$. The resolution can be increased at the cost of detection efficiency in the limit of an infinitely small pinhole to

$$d_{lat} = 0.37\frac{\lambda}{\mathrm{NA}} \tag{4.2}$$

Selection of an observation plane is achieved by inserting a pinhole in the detection pathway (Figure 4.1). The pinhole is placed in the same image plane as the laser focus (*confocal*) to reject light not originating from the focal plane. The axial resolution is given by the ability of the pinhole to reject out of focus light. Several different equations for the axial resolution have been suggested depending on the size of the confocal pinhole. Assuming that the pinhole size is similar to the distance between the maximum intensity and the first intensity minimum in lateral direction (1 Airy unit), the axial resolution is given by

$$d_{ax} = \frac{2n\lambda_0}{\mathrm{NA}^2} \tag{4.3}$$

where n is the refractive index of the sample and λ_0 is the wavelength of the emitted light. Typically, the axial resolution is a factor of 2-10 lower than the lateral resolution.

The detection volume or point spread function (PSF) for one photon excitation can be modeled by a three-dimensional gaussian,

$$\mathrm{PSF}(x, y, z) = exp\left(-\frac{2(x^2 + y^2)}{\omega_r} - \frac{2z^2}{\omega_z}\right) \tag{4.4}$$

where ω_r and ω_z are the distances from the center in lateral (x, y) and in axial (z) direction, respectively, where the detection efficiency has decreased to $1/e^2$.

Solution single-pair FRET (spFRET) measurements of single molecules are performed by placing the detection volume into a dilute solution of the sample molecules in buffer. The concentration has to be low enough to ensure that the chance of having more than one molecule in the focus at the same time is sufficiently low, usually below 1%. Using Poisson statistics, one can calculate the maximum allowed average number of molecules in the focus $\langle n \rangle$ where the probability of having 2 or more molecules in the focus is less than 1% of the probability of having exactly one molecule in the focus,

$$p_{mm} = \frac{P_{\langle n \rangle}(2) + \ldots}{P_{\langle n \rangle}(1)} = \frac{1 - P_{\langle n \rangle}(0) - P_{\langle n \rangle}(1)}{P_{\langle n \rangle}(1)} = \frac{1 - e^{-\langle n \rangle} - \langle n \rangle * e^{-\langle n \rangle}}{\langle n \rangle * e^{-\langle n \rangle}} < 0.01 \Rightarrow \langle n \rangle < 0.02 \tag{4.5}$$

where $P_\lambda(k)$ is the value of the Poisson distribution with average λ at k, i.e. the probability of having k molecules in the focus if the average number is λ. For a typical confocal volume used in this work of 2.5 fl, this corresponds to a concentration of 15 pM.

4.1.2 Excitation sources and pulsed interleaved excitation

Fluorophores in principle can be excited with any light source. Ideally, the excitation light is collimated and monochromatic for single-molecule fluorescence spectroscopy resulting in lasers being the most sensible choice. For spFRET burst analysis measurements, often only one laser is used to excite the donor and fluorescence from donor and acceptor is collected in spectrally separated detection channels. This yields data about the donor and the (apparent) FRET efficiency but lacks information on the presence and the environment of the acceptor. This can result in falsified FRET efficiencies due to e.g. acceptor bleaching during the transit of the molecule through the confocal volume.

One way to detect the presence of an active acceptor is using µs alternating laser excitation (µsALEX) introduced by (Kapanidis *et al.*, 2004). An additional laser for direct acceptor excitation is alternated on the µs timescale with the donor excitation using an electrooptical modulators. Pulsed interleaved excitation (PIE) (Müller *et al.*, 2005) uses pulsed lasers to alternatively excite the donor and the acceptor on the ns timescale. The use of pulsed lasers in PIE has the additional advantage that the donor and acceptor lifetimes carrying information about their local environment are also available.

The experiments on the folding of MBP and on the conformational dynamics of mtHsp70 (chapters 6 and 7) were performed using a frequency-doubled 1064 nm solid state laser ($\lambda = 532\,\mathrm{nm}$, PicoTA, Toptica, München, Germany) and a diode laser ($\lambda = 635\,\mathrm{nm}$, LDH-D-C-640, Picoquant, Berlin, Germany) as excitation sources. The lasers are triggered alternatingly by a computer-controlled Multichannel Picosecond Diode Laser Driver "Sepia I" from Picoquant (Berlin, Germany). The repetition frequency of the lasers was 27 MHz corresponding to a 37.5 ns duty cycle. The red laser was delayed relative to the green laser by approximately 15 ns. Typical excitation powers for burst anaysis measurements were 100 µW for both lasers. The light pulses are coupled into single-mode optical fibers and combined by a boxed dichroic mirror directly connected to the fibers (WDM-12P-111-532/647-3.5/125-PPP-50-3A3A3A-3-1,1,2, OZ Optics). The combined multi-color pulse train is collimated by an achromatic lens (60FC-4-RGB11-47, Schäfter+Kirchhoff) and directed by a multi-line dichroic mirror (DualLine z532/635, AHF Analysentechnik, München, Germany) into a 60x water immersion objective (Plan Apo VC 60x A/1.20 W, Nikon, Germany) with a numerical aperture (NA) of 1.2.

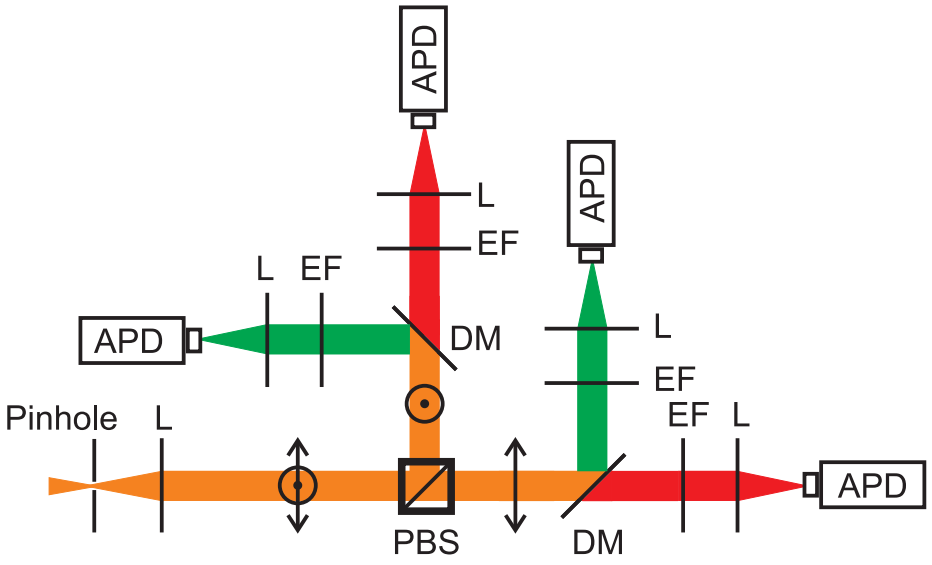


Figure 4.2: *Schematic of the detection pathway of the MFD setup. L: Achromatic lens ($f = 100\,\mathrm{mm}$); DM: Dichroic mirror, BS650; EF: Emission filter, HC582/75 (green), ET700/75 (red); PBS: Polarizing beamsplitter.*

4.1.3 Multi-parameter fluorescence detection

Fluorescence is excited by a laser focused through a high-NA objective (see previous section). The same objective collects the fluorescence from the sample (*epifluorescence microscopy*). After separation from the excitation light by the multi-line dichroic mirror, the fluorescence is focused by an achromatic lens on a 50 µm pinhole (OWIS GmbH) to reject out-of-focus light (Figure 4.2). The light is re-collimated by another achromatic lens and then split into four different detection channels by a polarizing beamsplitter cube (Thorlabs) and dichroic mirrors (640 DCXR, AHF Analysentechnik). Additional bandpass filters (BrightLine HC582/75, ET700/75, AHF Analysentechnik) were chosen to reduce the amount of scattered light by at least 6 orders of magnitude while having a transmission of >90% in the spectral range of donor and acceptor emission. Fluorescence is then focused on the active area of four avalanche photodiodes (2x SPCM-AQR-14, 2x SPCM-AQR-16, Laser Components).

With MFD, several molecular properties can be calculated simultaneously from the parameters of the detected photons (refer to Section 4.1.5.3). This allows the accurate and absolute determination of FRET efficiencies and donor-acceptor separations without the need for additional ensemble control measurements.

4.1.4 Time Correlated Single Photon Counting (TCSPC)

APDs produce one electrical TTL pulse for each detected photon. These pulses are registered by four time-correlated single photon couting (TCSPC) modules (SPC-154, Becker & Hickl, Berlin, Germany), one for each detection channel. Using individual modules for each APD avoids artifacts arising from routing electronics (Müller *et al.*, 2005). It is important to make sure that the TCSPC modules are triggered simultaneously by using an external trigger (USB-6008, National Instruments)

to start the acquisition. The TCSPC modules use constant fraction discriminators (CFD) to get arrival times of the pulses independent of the pulse amplitudes (for a more detailed description of the TCSPC cards and parameters refer to appendix A.1 and Becker, 2005a,b). A time-to-amplitude converter (TAC) then measures the time between the APD pulse and a synchronization pulse (Becker, 2005a). The synchronization pulse in this setup is generated by the laser driver "Sepia I". In addition to the arrival time relative to the laser pulse (*Microtime*), in time-tagged time resolved (TTTR,T3R) mode, the TCSPC modules also record the number of sync pulses between subsequent photons (*Macrotime*). For each photon, Macrotime and Microtime are saved together with routing and overflow information into one .spc file for each TCSPC module (Becker, 2005b). The data was taken using the program provided by Becker&Hickl. Typical settings are given in Table A.1 in appendix A.1.

4.1.5 Data analysis

After recording of the single-molecule bursts, the data collected by the TCSPC modules were analyzed either by a MATLAB program developed in our lab or with Labview programs developed in Claus Seidel's lab and adapted for PIE data in our lab by Volodymyr Kudryavtsev. Pulsed interleaved excitation was used to assign an excitation source to each photon (Section 4.1.5.1). Then, bursts in the collected photon stream were discriminated from background (Section 4.1.5.2), and molecular properties were calculated from the parameters of the collected photons (Section 4.1.5.3).

4.1.5.1 Pulsed interleaved excitation

Figure 4.3 shows typical TCSPC histograms of a MFD-PIE experiment. In the upper panel, light scattered by a water sample was collected to measure the instrument response function. The times of the laser pulses with respect to the sync pulse are indicated above the figure. The green laser is triggered approximately 3 ns after the sync pulse. The red laser is delayed with respect to the green laser by 14 ns. This delay allows to assign an excitation source to each photon. Photons detected between 3 and 17 ns were excited by the green laser, and the other photons were excited by the red laser.

The lower panel shows a measurement of a double-labeled DNA. In the donor detection channels (green), only fluorescence after donor excitation is observed with a characteristic monoexponential lifetime decay. The signal in the acceptor detection channels (red) is clearly divided in two parts. The left part (3-17 ns) originates from donor excitation and thus consists of crosstalk, direct acceptor excitation by the donor laser and FRET signal. The right part (>17 ns) consists of photons from direct acceptor excitation. Both parts show characteristic lifetime decays. The direct acceptor excitation signal decays almost monoexponentially. However, before emission of a FRET photon, energy transfer has to take place. Due to this two-step process, the FRET signal shows a clear double exponential behavior.

4.1.5.2 Burst search

There are different algorithms to distinguish a burst originating from a fluorescent particle in the confocal volume from background. The simplest algorithm uses a threshold. The photons are histogrammed into timebins with length T on the order of the burst duration. Then, a threshold of M_1 photons is applied to select bursts. For bins next to bins exceeding this threshold, one can use a

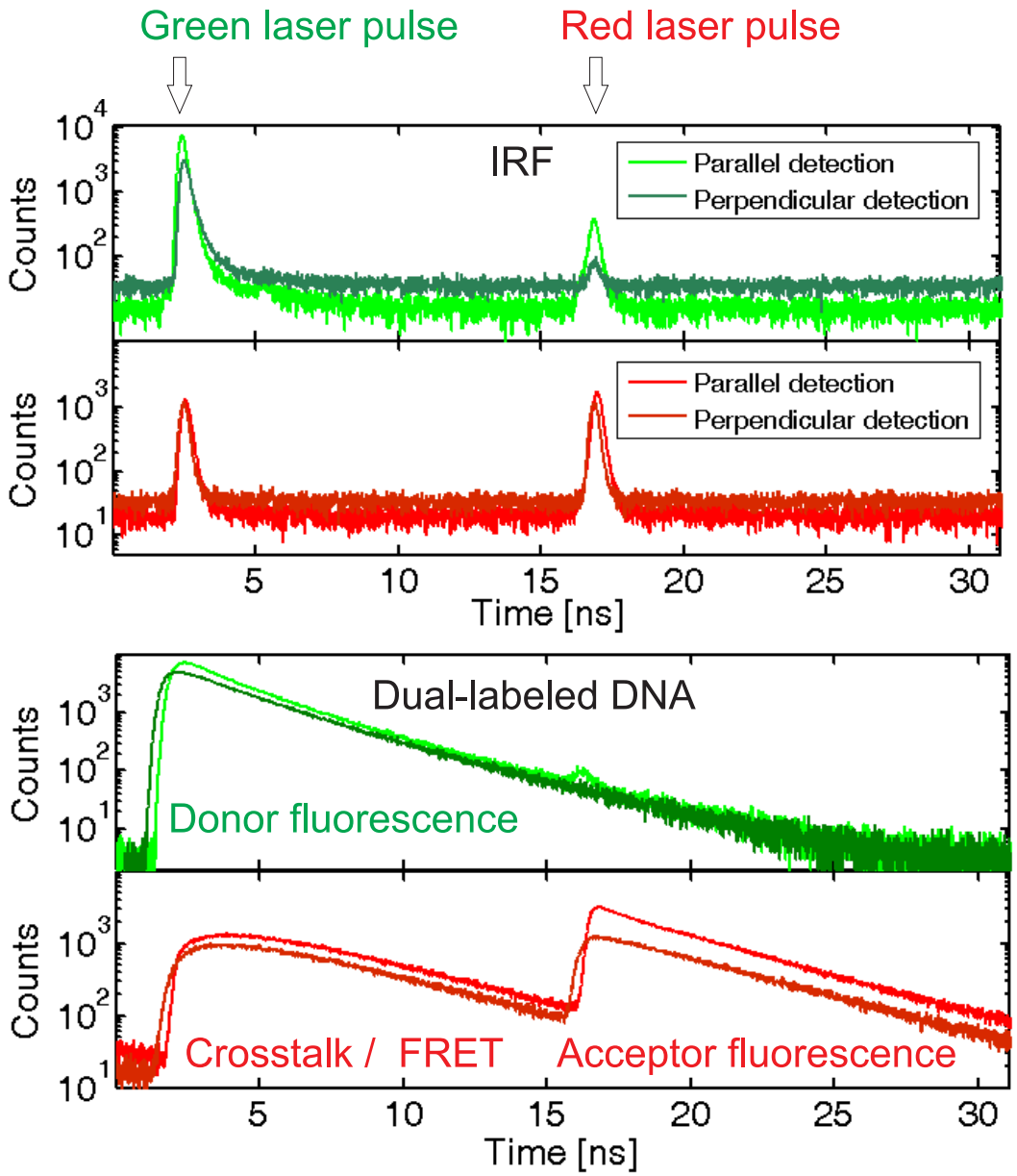


Figure 4.3: *Pulsed interleaved excitation. The upper panel shows the instrument response function of the MFD setup measured from scattered light from a water sample. For the four detection channels (upper graph: green parallel and green perpendicular; lower graph: red parallel and red perpendicular), a histogram of the number of counts per TCSPC channel is shown with arrows indicating the green and red laser pulse. The lower panel shows TCSPC histograms collected from a sample of double-labeled DNA. From the arrival time relative to the laser pulses, the excitation source of the photons in the red detection channels can be derived and the photons can be attributed to FRET (red detection after green excitation) and direct acceptor fluorescence (red detection after red excitation).*

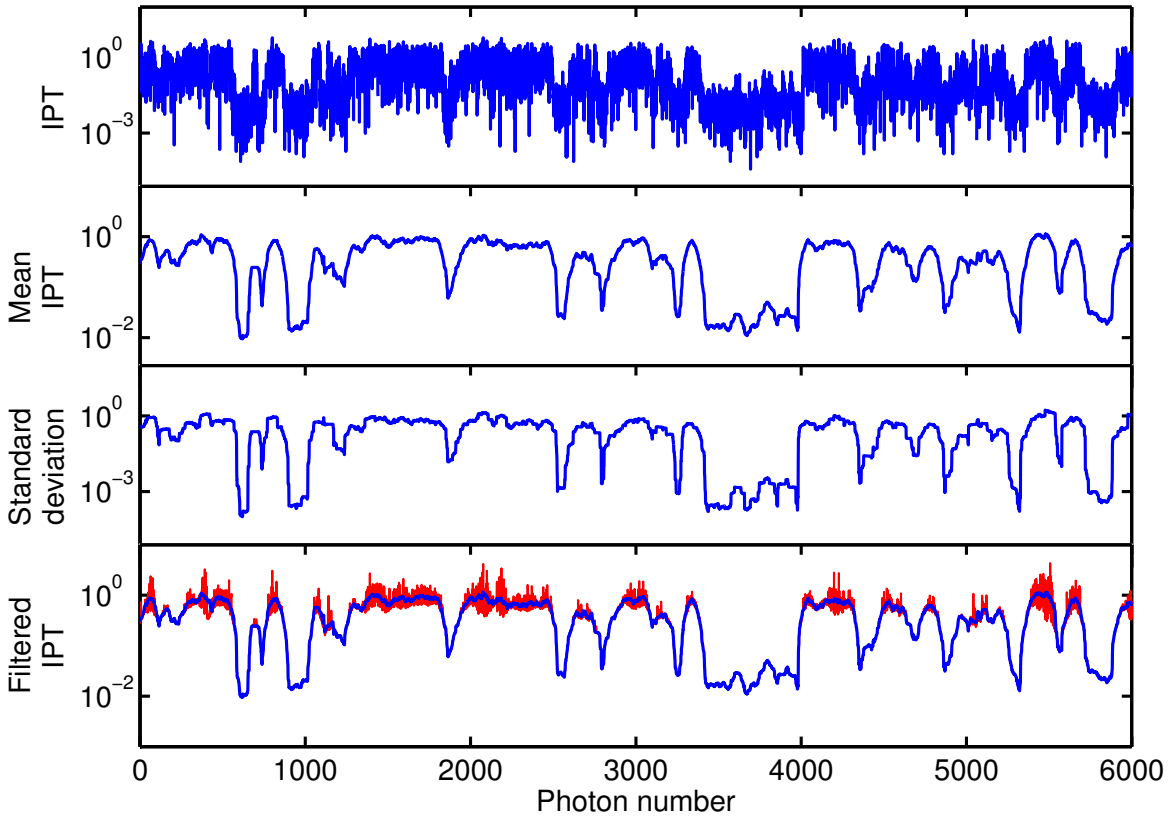


Figure 4.4: *Principle of the Lee filter. (1) Inter-photons times (IPT) from raw data without smoothing over photon number. (2) Moving average of inter-photon times with moving window of size $m = 30$ over photon number. (3) Moving standard deviation of inter-photon times with moving window of size $m = 30$ over photon number. (4) Filtered inter-photon times with Lee filter parameters $m = 30$ and $\sigma_0 = 10\,\mathrm{ms}$ (blue curve) or $\sigma_0 = 1\,\mathrm{ms}$ (red curve) aver photon number. For the lower σ_0, the preferential smoothing of the bursts is clearly visible.*

second lower threshold M_2 to avoid losing half-filled bins at the edges of the bursts (Müller *et al.*, 2005).

Instead of a fixed binning, the *all photon burst search (APBS)* method (Nir *et al.*, 2006) requires a minimum number of photons M in a sliding window of duration T. Only bursts with a minimum number of L photons are accepted. This method applies basically a smoothing procedure with a moving average filter of L photons and subsequently thresholds the time between two photons to T/L.

This method is similar to using a *Lee filter* with parameters m and σ_0 (refer to Table 4.1) to smooth the data followed by thresholding of the inter-photon times (Enderlein *et al.*, 1997). The main difference to the APBS is that the Lee filter weights the averaging with the standard deviation of the inter-photon times. Inter-photon times with moving standard deviation $\sigma < \sigma_0$ are averaged more than for $\sigma > \sigma_0$ resulting in preferential smoothing of the bursts (Figure 4.4). This burst search algorithm has been used to analyze the experiments described in this thesis.

Table 4.1: *Description of burst search parameters.*

Algorithm	Parameter	Description	Standard range
Binning	T	Length of time bin	0.2-5 ms
	M_1	Minimum number of photons per bin	$10 - 100$
	M_2	Threshold for adjacent bins	$5 - 20$
APBS	T	Length of time window	0.2-5 ms
	M	Minimum number of photons in time window	$10 - 100$
	L	Minimum number of photons per burst	> 80
Lee filter	Δt	Inter-photon time threshold after filtering	0.01 - 1 ms
	m	Size of averaging window for Lee filter	30
	σ_0	Weighting parameter for averaging.	10
	L	Minimum number of photons per burst	> 80

4.1.5.3 Parameter calculations for individual bursts

FRET efficiency

The parameters of the photons of the detected bursts can now be used to calculate several different properties of the fluorescent particles (refer to Table 4.2). The most prominent parameter is the *FRET efficiency*, E, which depends strongly on the distance between the donor and the acceptor fluorophore (refer to Section 2.2). It can be calculated from the number of photons originating from the donor $\tilde{F}_{D,D}$ and from the acceptor $\tilde{F}_{D,A}$ after donor excitation,

$$E = \frac{\tilde{F}_{D,A}}{\tilde{F}_{D,A} + \tilde{F}_{D,D}} \tag{4.6}$$

To calculate the number of photons originating from FRET, the number of actually detected photons $F_{D,A}$ has to be corrected for *crosstalk* (CT, donor fluorescence detected in the acceptor channel) and DE*, direct excitation* (excitation of the acceptor by the donor laser),

$$\tilde{F}_{D,A} = F_{D,A} - \mathrm{CT} \cdot F_{D,D} - \mathrm{DE} \cdot F_{A,A} \tag{4.7}$$

where $F_{A,A}$ is the photon count in the acceptor detection channel after excitation by the acceptor laser. Additionally, the differences in quantum yield of donor ϕ_D and acceptor ϕ_A and detection efficiencies of the detection channels η_D, η_A have to be taken into account by weighting the donor fluorescence with a factor $\gamma = \frac{\eta_A \cdot \phi_A}{\eta_D \cdot \phi_D}$.

$$\tilde{F}_{D,D} = \gamma \cdot F_{D,D} \tag{4.8}$$

The corrected FRET efficiency can thus be calculated by

$$E = \frac{F_{D,A} - \mathrm{CT} \cdot F_{D,D} - \mathrm{DE} \cdot F_{A,A}}{F_{D,A} - \mathrm{CT} \cdot F_{D,D} - \mathrm{DE} \cdot F_{A,A} + \gamma \cdot F_{D,D}} \tag{4.9}$$

Fluorescence lifetime

E can also be calculated from the fluorescence lifetime of the donor in the presence of the acceptor τ_{D_A} and in the absence of the acceptor τ_{D_0} (Eq. 2.10 in Section 2.2). To determine the *donor lifetime* for each individual burst, a maximum likelihood estimator (MLE) is used as described in more detail

in (Schaffer *et al.*, 1999; Kudryavtsev, 2006). The MLE assumes a single exponential decay[1] of the fluorescence intensity convoluted with an experimentally determined instrument response function (IRF). The IRF is determined from a measurement of light scattered by a water sample. The MLE yields more accurate results than a least squares fit for the low number of photons (down to 80) available for the lifetime determination.

For the MLE, the time-dependent model $M(t)$ for the signal is taken to consist of fractions α of background signal from scattered light, $B(t)$, and $(1-\alpha)$ of the signal from a single exponential fluorescence component $S(t)$. $S(t)$ is given by the convolution of the single exponential with the instrument response function (IRF), $F(t) = \mathrm{IRF} \otimes exp(-t/\tau)$. The MLE determination of the lifetime is then based on the calculation of the likelihood function

$$L(F|M(x)) = \prod_{i=1}^{k} w(F_i|M_i(x)) \tag{4.10}$$

where F is the experimental decay histogram, x is a set of model parameters for M, w is the probability of detecting F_i photons in the ith TCSPC channel, and k is the number of TCSPC channels. The lifetime is determined by minimization of the normalized fit quality parameter[2]

$$2I_r^* = -2\frac{L(F|M(x))}{k-\zeta-1} \tag{4.11}$$

where ζ is the number of model parameters. The derivation of the complete equation for $2I_r^*$ is given in (Schaffer *et al.*, 1999; Kudryavtsev, 2006) and only the final result is given here:

$$2I_r^* = \frac{2}{k-\zeta-1}\left[\left(F_\parallel ln\frac{F_\parallel}{S_\parallel} + F_\perp ln\frac{F_\perp}{M_\perp}\right) + \sum_i F_{\parallel,i} ln\frac{F_{\parallel,i}}{M_{\parallel,i}} + \sum_i F_{\perp,i} ln\frac{F_{\perp,i}}{M_{\perp,i}}\right] \tag{4.12}$$

The same MLE is used to determine the *acceptor lifetime* τ_A which contains information about changes of the local environment of the acceptor. These changes can be used to correctly determine the γ-factor in different conformations.

Fluorescence anisotropy

From the intensity ratio of photons polarized parallel and perpendicular to the excitation light, we can calculate the *fluorescence anisotropy* of donor, FRET and acceptor fluorescence, r_d, r_f and r_a.

$$r_i = \frac{G\cdot F_i^{\parallel} - F_i^{\perp}}{(1-3I_2)G\cdot F_i^{\parallel} + (2-3I_1)\cdot F_i^{\perp}},\ i = \{D,D;D,A;A,A\} \tag{4.13}$$

$G_i = \frac{\eta_{\perp,i}}{\eta_{\parallel,1}}$ is the correction factor for different detection efficiencies of the parallel and perpendicular detection channels and I_1 and I_2 are the correction factors for depolarization effects of the high-NA objective (Koshioka *et al.*, 1995). The anisotropies are related to the rotational mobility of the donor and acceptor dyes which in turn influences the Förster radius by the κ^2 factor. We can use the three

[1] A single exponential decay may not describe the data properly depending on the experimental conditions, e.g. for acceptor fluorescence with FRET or for fluorophores with a more complex decay like GFP . However, given the low number of photons per burst available for the lifetime determination, more complex models cannot be fit and a single exponential decay is a good first approximation.

[2] The fit quality parameter $2I_r^*$ is related to the Kullback Leibler divergence, which is used to measure the difference of two probability distributions. Here, the likelihood function $L(F|M(x))$ formally is compared to the likelihood of the experimental data given the experimental data $L(F|F) = 1$.

anisotropies accessible by PIE to calculate upper and lower bounds for κ^2 (Kudryavtsev *et al.*; Ivanov *et al.*, 2009).

Stoichiometry
Another important parameter in PIE experiments is the *stoichiometry* S defined as

$$S = \frac{\tilde{F}_{D,D} + \tilde{F}_{D,A}}{\tilde{F}_{D,D} + \tilde{F}_{D,A} + \tilde{F}_{A,A}} \tag{4.14}$$

and with corrections for CT, DE and γ

$$S = \frac{\gamma \cdot F_{D,D} + F_{D,A} - \mathrm{CT} \cdot F_{D,D} - \mathrm{DE} \cdot F_{A,A}}{\gamma \cdot F_{D,D} + F_{D,A} - \mathrm{CT} \cdot F_{D,D} - \mathrm{DE} \cdot F_{A,A} + F_{A,A}} \tag{4.15}$$

It is important to notice that the stochiometry will assume a value of $S = 1$ for molecules having no active acceptor and $S = 0$ for molecules having no active donor. This allows the separation of double-labeled molecules with $0.2 < S < 0.8$ from single-labeled ones.

$\mathbf{T_{D,x} - T_{A,A}}$
Even double-labeled molecules might undergo photobleaching or blinking during a burst. This will result in falsified E and S values (Figure 4.5B and Section 4.1.5.5). A valuable parameter to filter out these molecules is from the analysis is

$$T_{D,x} - T_{A,A} \tag{4.16}$$

the difference between the mean macroscopic arrival time of all photons after donor excitation $T_{D,X}$ and the mean macroscopic arrival time of all photons from the acceptor after red excitation $T_{A,A}$, with

$$T_y = \frac{1}{N_y} \sum_{i=1}^{N_y} t_i \tag{4.17}$$

where t_i is the *macrotime* of photon i and the summation runs over all N_y relevant photons, e.g. over all photons after donor excitation for $T_{D,X}$.

If photobleaching of the acceptor occurs during transition of the molecule through the 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 (Figure 4.5A). It is thus possible to exclude these molecules from the analysis by requiring $|T_{D,X} - T_{A,A}| < 1\,\mathrm{ms}$.

In Figure 4.5A, different average arrival times for an example trace with acceptor photobleaching are shown. $T_{D,x} - T_{A,A}$ is over 20 arbitrary time units in this example. However, other differences are equally high ($T_{D,x} - T_{D,A}$) or even higher ($T_{D,D} - T_{A,A}$, $T_{D,D} - T_{D,A}$). These may thus be even better parameters for filtering acceptor bleaching bursts. However, the main advantage of $T_{D,X} - T_{A,A}$ becomes clear from Figure 4.5C. Conformational dynamics during the burst result in a change in FRET efficiency and shift also $T_{D,D}$ and $T_{D,A}$. The only difference of average arrival times that is independent of structural dynamics of the sample is $T_{D,X} - T_{A,A}$, and using other differences could result in filtering of bursts containing conformational dynamics (Kudryavtsev *et al.*).

4.1.5.4 Correction Factors

An accurate calculation of most fluorescence parameters relies on the application of correction factors. These correction factors can be determined either by separate measurements of control samples or directly from other molecules or parameters in the same measurement.

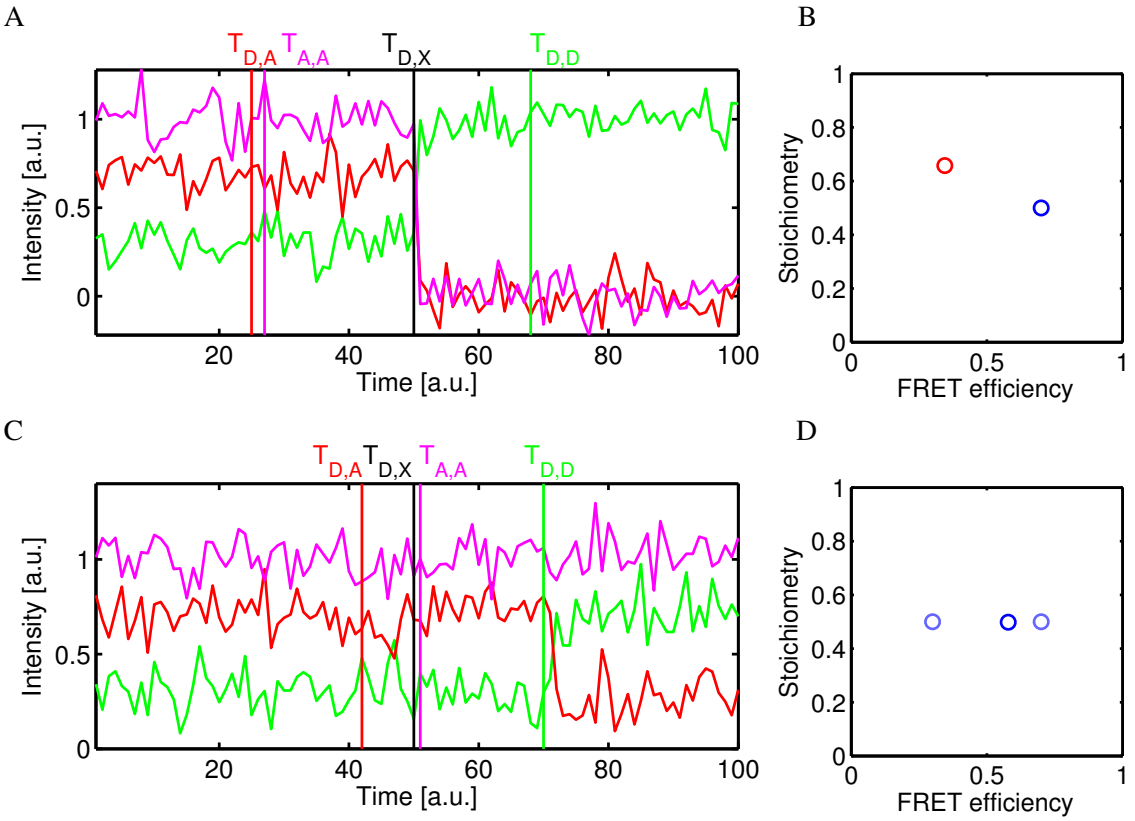


Figure 4.5: *Using the parameter $T_{D,X} - T_{A,A}$ to detect photobleaching events. (A) Idealized time trace of a burst where the acceptor photobleaches during transition of the molecule through the detection volume. The time traces show the donor intensity (green) and acceptor intensity (red) after donor excitation and the acceptor intensity after acceptor excitation (magenta). The vertical lines represent the average arrival times of the donor photons after donor excitation (green, $T_{D,D}$), the acceptor photons after donor excitation (red, $T_{D,A}$), all photons after donor excitation (black, $T_{D,X}$), and the acceptor photons after acceptor excitation (magenta, $T_{A,A}$). (B) Localization of the burst in the E-S-histogram. The true position for the molecule with active donor and acceptor is shown in blue, the measured position for the bleaching molecule is shown in red. (C) Idealized time trace of a burst where the molecule undergoes a transition to a different FRET state while inside the detection volume. The time traces and vertical lines are the same as in A. (D) Localization of the burst in the E-S-histogram. The positions of the molecule's substates are shown in light blue. The averaged position (dark blue) is on a line between the two substates.*

Table 4.2: *List of parameters determined from a MFD-PIE experiment*

Property	Description	Equation
FRET efficiency E	Efficiency of fluorescence energy transfer between donor and acceptor.	4.9
Fluorescence lifetimes τ_i	Lifetime of donor fluorescence (τ_D), FRET (τ_F) and acceptor fluorescence (τ_A).	4.12
Anisotropies r_i	Anisotropy of donor fluorescence τ_D, FRET τ_F and acceptor fluorescence τ_A.	4.13
Stoichiometry S	Ratio of fluorescence after donor excitation over total fluorescence. It is proportional to the ratio of the number of photons absorbed by the donor to the total number of absorbed photons and can thus be used to select double-labeled molecules.	4.14
$T_{D,X} - T_{A,A}$	Time between center of mass of photon arrival times after donor excitation and after acceptor excitation. Values significantly different from 0 ms indicate photobleaching, -blinking or a multimolecule event.	4.16

For the calculation of the FRET efficiency, three correction factor are needed (see previous section). The correction factor for spectral crosstalk (CT) is determined by requiring that the FRET efficiency of a donor-only labeled sample is 0. For a donor-only labeled sample, $F_{A,A} = 0$ and thus one can calculate CT as

$$E^{D_{only}} = \frac{F_{D,A} - \mathrm{CT} \cdot F_{D,D}}{F_{D,A} - \mathrm{CT} \cdot F_{D,D} + \gamma \cdot F_{D,D}} \stackrel{!}{=} 0 \Rightarrow \mathrm{CT} = \frac{F_{D,A}}{F_{D,D}} \tag{4.18}$$

CT can be related to the uncorrected FRET efficiency of the donor-only population $E_{PR}^{D_{only}}$,

$$E_{PR}^{D_{only}} = \frac{F_{D,A}}{F_{D,D} + F_{D,A}} = \frac{\mathrm{CT}}{1 + \mathrm{CT}} \Rightarrow \mathrm{CT} = \frac{E_{PR}^{D_{only}}}{1 - E_{PR}^{D_{only}}} \tag{4.19}$$

On the other hand, we can use an acceptor-only labeled sample with $F_{D,D} = 0$ and require that $S = 0$ to calculate the correction factor for direct acceptor excitation by the donor laser DE.

$$S = \frac{F_{D,A} - \mathrm{DE} \cdot F_{A,A}}{F_{D,A} - \mathrm{DE} \cdot F_{A,A} + F_{A,A}} \stackrel{!}{=} 0 \Rightarrow \mathrm{DE} = \frac{F_{D,A}}{F_{A,A}} \tag{4.20}$$

Again, DE can be related to the uncorrected stoichiometry of the acceptor-only population $S_{PR}^{A_{only}}$ by

$$\mathrm{DE} = \frac{S_{PR}^{A_{only}}}{1 - S_{PR}^{A_{only}}} \tag{4.21}$$

Since the fluorescence signals from the acceptor in the above equation origin from different excitation lasers, DE has to be determined using the same ratio of donor and acceptor excitation powers as in the actual experiment. This is automatically the case if it is determined as described above from the acceptor-only population of the same experiment.

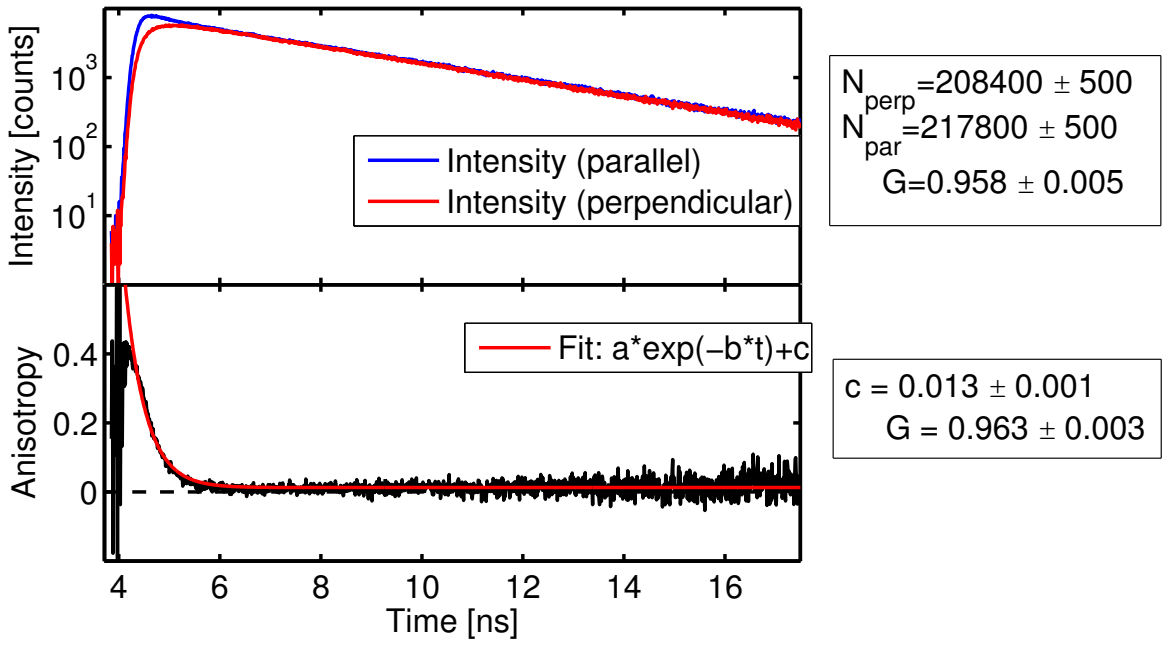


Figure 4.6: *Determination of the correction factor G by taking the ratio of the number of photons in the perpendicular and parallel detection channels in the tail of the lifetime decay (upper panel) and by fitting of the time-resolved anisotropy decay (lower panel).*

The third and maybe most important correction factor for the FRET efficiency is the detection correction factor γ. γ depends on the ratio of the detection efficiencies of the donor and acceptor detection channels and the relative quantum yields of the two dyes. The former can change due to alignment and due to spectral changes of the used dyes, while the latter describe the number of excitations resulting in photon emission, which can be affected e.g. by quenching of the dyes. Since the alignment should be optimized and usually no significant spectral shifts of the emission spectra are induced by the attachment to a biological sample, the main cause of changes in γ are changes in the quantum yields. In principle, these can be observed due to the associated changes in fluorescence lifetime. The donor lifetime can be used as an γ - independent way to calculate the FRET efficiency. Matching the lifetime-calculated efficiency and the intensity-calculated efficiency can thus be another way to calculate γ. Finally, for correct γ, E and S are uncorrelated. By minimizing the correlation between E and S, we can also determine the correct γ factor (Kapanidis *et al.*, 2004; Müller *et al.*, 2005).

In the calculation of the anisotropy, $G_i = \frac{\eta_{\perp,i}}{\eta_{\parallel,1}}$ corrects for differences in the detection efficiencies of the parallel and perpendicular detection channels. For the determination of G, a measurement of a fast rotating molecule such as a free dye can be used. The correct value for G is obtained by requiring that the time-resolved anisotropy for a freely rotating molecule falls off to 0 quickly after excitation, either by taking the ratio of the number of photons in the perpendicular and parallel detection channels in the tail of the lifetime decay (Figure 4.6, upper panel), or by direct fitting of the uncorrected time-resolved anisotropy and calculation of $G = \frac{1-c}{1+2c}$ from the fitted offset c (Figure 4.6, lower panel). Both methods yield a very similar value for G.

4.1.5.5 Molecule selection

Based on the fluorescence parameters calculated above, we can now select bursts for the analysis. Figure 4.7 shows how the S-E histogram is changed when only selecting molecules with stoichiometry values of $0.3 < S < 0.8$ (Figure 4.7C) and without photobleaching ($|T_{FRET} - T_{RED}| < 1\,\mathrm{ms}$, Figure 4.7B).

Additional quenching of the donor not due to FRET or acceptor quenching can be observed by changes in the fluorescence lifetime in 2D histograms of the FRET efficiency versus fluorescence lifetime of the donor after donor excitation or the fluorescence lifetime of the acceptor after acceptor excitation (Figure 4.8A-B).

The donor lifetime depends on the FRET efficiency. The theoretical relationship between donor lifetime in the presence of the acceptor $\tau_{D,A}$ and E is given by equation 2.10. Gansen *et al.* (2009) have shown that due to the flexibility of the dye linkers, this relationship changes slightly and is described by the following empirically determined third order equation [3]:

$$E = 1 - (-0.0421 + 0.5080 \cdot \tau_{D_A} + 0.26914 \cdot \tau_{D_A}^2 - 0.035889 \cdot \tau_{D_A}^3)/\tau_{D_0} \tag{4.22}$$

where τ_{D_0} is the lifetime of the donor in the absence of the acceptor given in ns.

Additional quenching of the donor results in a shift of the species' population to lower lifetimes than predicted by equation 4.22. In Figure 4.8A, the average donor lifetime of the high FRET population is higher than predicted by theory, indicating that the γ factor used for the analysis was to low. In contrast, for the low FRET population the average donor lifetime is lower than predicted, indicating that for this measurement the γ factor used for the analysis was to high. The reason for the population-dependent γ factor is quenching of the acceptor in the low FRET state, observed by a 7% decrease of the acceptor lifetime (Figure 4.8B). Acceptor lifetime and quantum yield are proportional as well as acceptor quantum yield and γ. Using this information, the correct γ for each population can be calculated.

The 2D histograms of anisotropy versus lifetime allow the detection of the subspecies with a different rotational mobility (Figure 4.8C-D). The theoretical relationship between lifetime τ and steady-state anisotropy r for a spherical molecule is given by the Perrin equation (Perrin, 1934, 1936)

$$r(\tau) = \frac{r_0}{(1 + \frac{\tau}{\rho})} \tag{4.23}$$

with the rotational correlation time ρ and the fundamental anisotropy r_0. While the fundamental anisotropy r_0 depends on the angle between the absorption and emission dipole β [4] and is constant for a given fluorophore, $\rho = 1/(6D_r)$ depends on the rotational mobility of the molecules described by the rotational diffusion coefficient D_r. Molecules with different molecular weight, e.g. unbound dyes, will thus have an anisotropy described by a different rotational diffusion time.

In Figure 4.8C and D, free dye with a different lifetime was added to the DnaK sample. The free dye species can be separated in the 2D histograms and used for the determination of the G factor as described in Section 4.1.5.4.

[3] The coefficients of this equation are determined by a numerical simulation of all possible values of the donor-acceptor distance R and calculation of the corresponding values of E and τ_{D_A}. For the calculation, the Förster radius R_0, the donor-only lifetime τ_{D_0}, and the width of the distance distribution resulting from linker movement is necessary. The latter is determined by a fit of the TCSPC decay of a FRET sample (Sindbert *et al.*, 2011).

[4] $r_0 = 0.2 * (3cos^2(\beta) - 1)$, thus for parallel absorption and emission dipoles $r_0 = 0.4$.

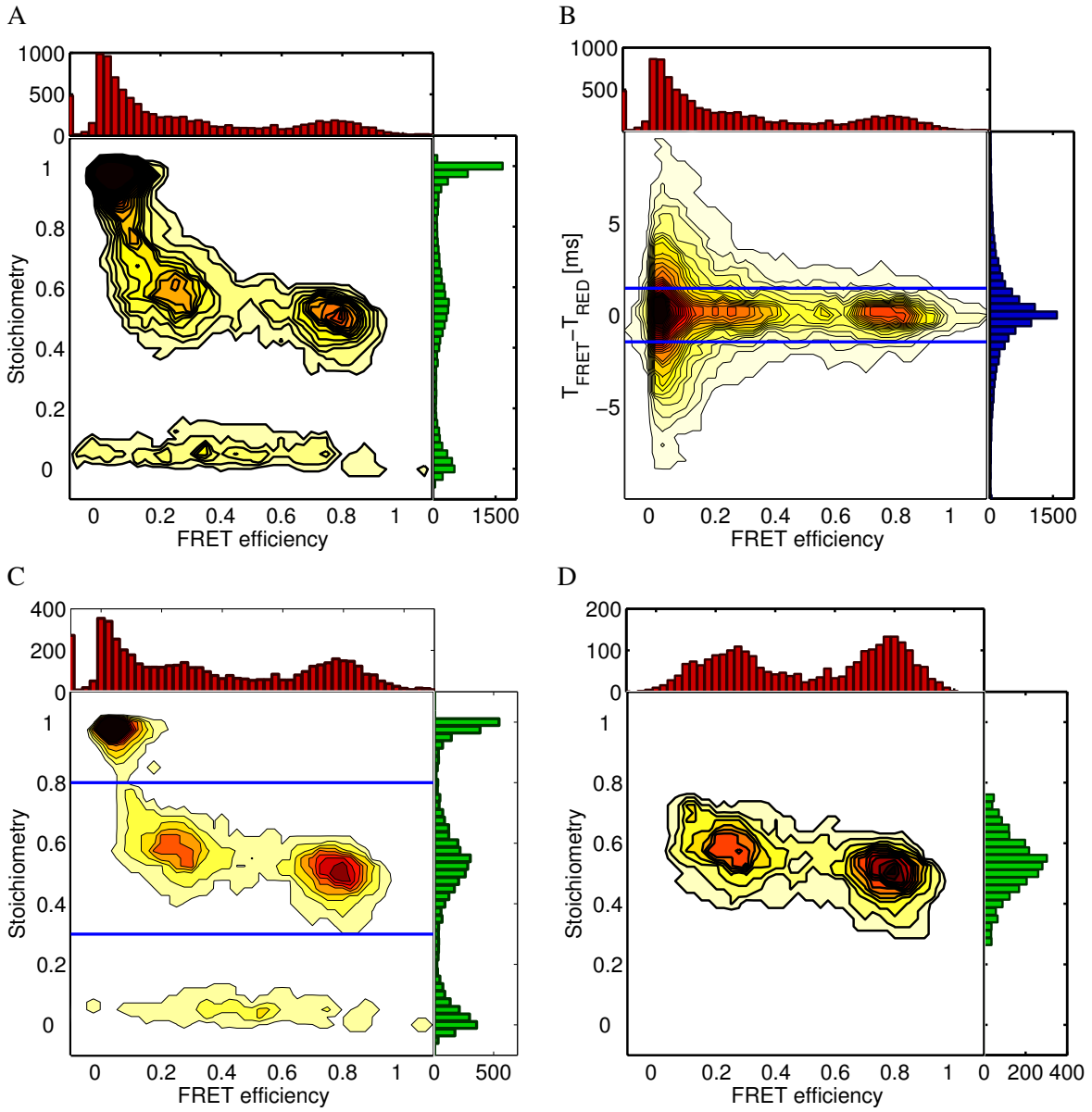


Figure 4.7: *Selection of double-labeled molecules without photobleaching based on 2D histograms of different fluorescence parameters. 1D histograms of the parameters are shown above and to the right of 2D histograms. (A) All bursts in the measurement displayed in a 2D S-E histogram. Trailing is visible mainly between donor-only molecules and double-labeled molecules. The 1D histogram for E is dominated by the donor-only molecules with $E \approx 0$. (B) All bursts in the measurement displayed in a 2D $(T_{D,x} - T_{A,A})$-E histogram. The bursts are distributed around $(T_{D,x} - T_{A,A}) = 0$. Only bursts with $|T_{D,x} - T_{A,A}| < 1\,\mathrm{ms}$ are included in the analysis. (C) 2D S-E histogram after cut in (B). Trailing is strongly reduced. The stoichiometry is then used to select only double-labeled molecules with $0.3 < S < 0.8$. (D) Final S-E histogram.*

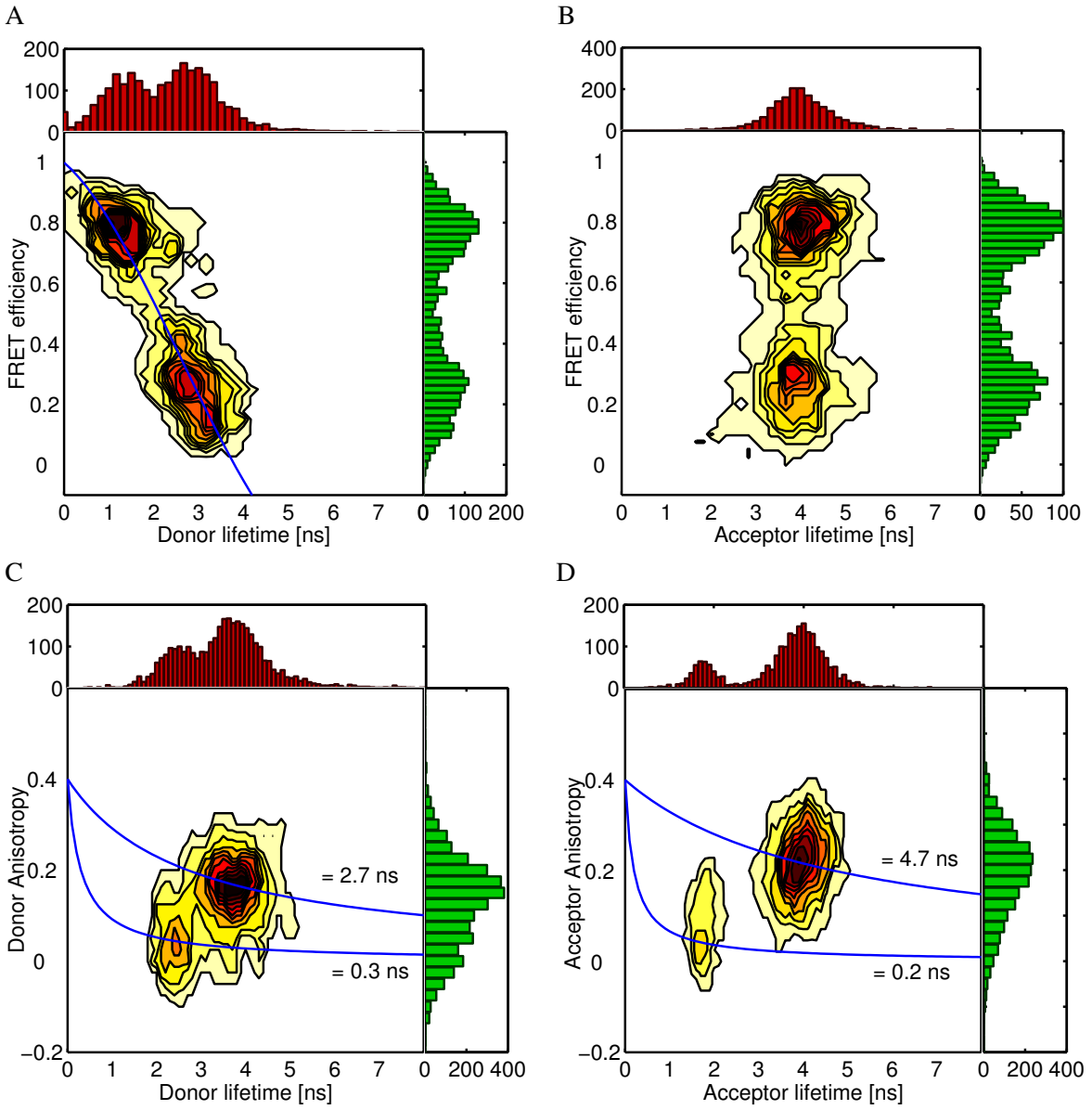


Figure 4.8: *Quenching of donor or acceptor and contamination by free dye detected in 2D histograms of DnaK with ATP and substrate. (A) Histogram of E versus donor lifetime after donor excitation for the double-labeled species. The blue line corresponds to the theoretical FRET curve (equation 4.22). (B) Histogram of E versus acceptor lifetime after acceptor excitation for the double-labeled species. (C) Histogram of anisotropy versus lifetime for donor fluorescence after donor excitation for the donor-only species. Clearly, two species with different rotational mobilities can be distinguished. The blue lines describe the Perrin equations for different ρ values. (D) Histogram of anisotropy versus lifetime for acceptor fluorescence after acceptor excitation for the acceptor-only species. Again, two species with different rotational mobilities are distinguishable. The blue lines describe the Perrin equations for different ρ values.*

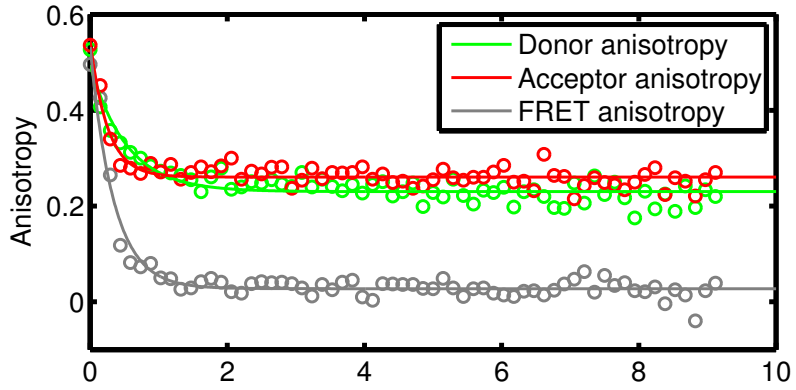


Figure 4.9: *Error estimation of the Förster radius. Time resolved anisotropies of donor signal (green), FRET signal (grey), and acceptor signal (red). Fits to equation 4.24 are shown as lines.*

4.1.5.6 Error estimation of the Förster radius

The Förster radius R_0 depends on the relative orientation of the donor emission dipole and the acceptor excitation dipole (Section 2.2). The relative orientation is described by the parameter κ^2 given in equation 2.8. κ^2 can theoretically lie between 0 and 4. Ivanov *et al.* (2009) have proposed an approach based on an earlier study (Dale *et al.*, 1979) to narrow down the possible range of κ^2 based on the steady-state anisotropies of donor emission, FRET signal, and acceptor emission. For the experiments presented in this thesis, the time-resolved anisotropy data for donor emission, FRET signal, and acceptor emission is also available from MFD-PIE measurements. We used it to estimate the possible values of the orientation factor κ^2 and calculated the minimum and maximum values possible for the Förster radius. If the rotational correlation time of the whole protein is long compared to the fluorescence lifetime of the dyes, the time-resolved anisotropy decays can be described by

$$r(t) = (r_0 - r_\infty)e^{-\left(\frac{t}{\theta_{dye}}\right)} + r_\infty \tag{4.24}$$

where r_0 is the fundamental anisotropy given by the angle between excitation and emission dipole, r_∞ is the residual anisotropy and θ_{dye} is the rotational correlation time of the the rapid, restricted motion of the dye along the linker (Figure 4.9). The residual anisotropy of the directly excited donor and acceptor molecules characterize the extent of the mobility of the dyes while the residual anisotropy of the FRET signal allows calculation of the angle between the distributions of the donor and acceptor transition dipoles. The fit values for the fundamental anisotropy have a high error due to the low timing performance of the APDs used in the experiment. For the correct determination of the fundamental anisotropies, measurements with faster detectors like PMTs would be necessary. For the calculation of uncertainties in R_0, we thus used a fundamental anisotropy of 0.385 assuming almost parallel excitation and emission dipoles.

For the experiment shown in Figure 4.9, we calculated a minimum value for κ^2 of 0.19 and a maximum value of 2.52. These values correspond to a Förster radius in the range of $0.81R(\kappa^2 = 2/3) < R_0 < 1.25R(\kappa^2 = 2/3)$, where $R(\kappa^2 = 2/3)$ is the Förster radius assuming rapid orientational averaging and $\kappa^2 = 2/3$.

4.1.5.7 Probability distribution analysis (PDA)

The single-molecule bursts originating from double-labeled molecules were subjected to probability distribution analysis to extract the maximum available information. This method derives the width and position of the distance distribution underlying the FRET efficiency histogram by taking into account the inherent broadening caused by photon shot noise and the effects of histogramming discrete photon values (Antonik *et al.*, 2006; Kalinin *et al.*, 2007, 2008).

The FRET efficiency describes the probability of energy transfer from the excited donor dye to an acceptor dye. This is directly related to the ratio of actually measured photons in the donor detection channel S_G to the number of detected photons in the acceptor channel S_R. Only in the limit of an infinite number of total detected photons, $N \rightarrow \infty$, a single number is resolved for S_G/S_R. For finite N, however, the observed ratio S_G/S_R will be distributed due to the stochastic nature of photon emission. This phenomenon is called *shot noise*. With signals $S_i = F_i + B_i$ consisting of fluorescence F_i and background photon counts B_i, the probability to observe a specific ratio S_G/S_R is given by

$$P\left(\frac{S_G}{S_R}\right) = \sum P(F_G, F_R, B_G, B_R) \tag{4.25}$$

where the summation runs over all combinations of F_G,F_R,B_G, and B_R that result in a given value of $\frac{S_G}{S_R}$.
$P(F_G, F_R, B_G, B_R)$ can be factorized since the probabilites to detect B_G background photons in the donor channel, $P(B_G)$, and to detect B_R background photons in the acceptor channel, $P(B_R)$, are independent of the probability to detect a combination of F_G and F_R fluorescence photons, $P(F_G, F_R)$. Additionally, $P(F_G, F_R)$ can be separated into the probability to detect F total fluorescence photons combined with the probability to detect F_A acceptor photons given a total photon count of F, $P(F_G, F_R) = P(F)P(F_R|F)$.

Assuming that the data is binned in equal time bins, the probability distributions for background photons are Poissonian distributions. The means of the distributions are the average number of photons per bin $\langle B_i \rangle$,

$$P_{\langle B_i \rangle}(B_i) = \frac{\langle B_i \rangle^{B_i} e^{-\langle B_i \rangle}}{B_i!} \tag{4.26}$$

$\langle B_i \rangle$ can be calculated from the countrate of a clean buffer measurement.

The probability of observing F_R acceptor photons given F total fluorescence photons is described by a binomial distribution of drawing F times with an *apparent transfer efficiency* ϵ,

$$P(F_R|F) = \frac{F!}{F_R!(F-F_R)!}\epsilon^{F_R}(1-\epsilon)^{(F-F_R)} \tag{4.27}$$

$P(F)$, the distribution of the number of fluorescence photons per burst, is not as easily modeled as the other probability distributions. Analytical descriptions exist in the literature (Fries *et al.*, 1998), but due to the complexity of these models using an experimentally accessible quantity for $P(F)$ would be easier. Assuming that B_G and B_R are given and $F = N - B_G - B_R$, one can normalize the measured distribution of the total number of photons per time bin N to get $P(N)$. This approach is valid if the fluorescence signal is much stronger than the background $F \gg (B_G + B_R)$ (Kalinin *et al.*, 2007). This condition was fulfilled for all experiments in this thesis. For weaker signals, $P(F)$ is determined by deconvolution of $P(N)$ using a maximum entropy method to model the distribution of photons coming from different volume elements in the sample (Kalinin *et al.*, 2007).

Putting equations 4.26 and 4.27 into equation 4.25 we get

$$P\left(\frac{S_G}{S_R}\right) = \sum P(N)P(F_R|N - B_G - B_R)P_{\langle B_G\rangle}(B_G)P_{\langle B_R\rangle}(B_R) \tag{4.28}$$

where the summation now runs over all combinations of N, F_R, B_G, and B_R yielding a given value of S_G/S_R. ϵ is now the only free parameter and is found by χ^2 minimization fitting of a model histogram resulting from equation 4.28 to the data histogram.

ϵ is related to the FRET efficiency E and therefore to the distance between donor and acceptor by equation 4.29.

$$\epsilon = 1 - \frac{1}{1 + \gamma\frac{E}{1-E} + \mathrm{CT}} = 1 - \frac{1}{1 + \gamma\frac{R_0^6}{R^6} + \mathrm{CT}} \tag{4.29}$$

Due to linker flexibility, dye photophysics, or conformational changes of the sample, usually a single value for ϵ is not enough to describe the data but a distribution of ϵ values is required. This distribution is usually composed of at least one transfer efficiency accounting for donor-only molecules ($\epsilon = \mathrm{CT}$) and one or more Gaussian distributions of distances with means $\overline{R}_{DA,i}$ and standard deviations $\sigma_{DA,i}$. Equation 4.28 has to be integrated over all possible values of ϵ weighted by the probability of ϵ as a function of donor-acceptor distance, $p(\epsilon(R_{DA}))$.

$$P\left(\frac{S_G}{S_R}\right) = \int p(\epsilon(R_{DA})) \sum P(N)P(F_R|N - B_G - B_R)P_{\langle B_G\rangle}(B_G)P_{\langle B_R\rangle}(B_R)d\epsilon \tag{4.30}$$

4.2 Fluorescence correlation spectroscopy

4.2.1 Basic theory

The single-molecule burst analysis measurements described in the last section rely on sufficient temporal separation of two molecules diffusing through the detection volume to calculate fluorescence parameters for each single molecule individually. A more concentrated sample will result in a constant average count rate with fluctuations resulting from diffusion of particles through the confocal volume. Fluorescence correlation spectroscopy (FCS) analyzes these fluctuations to extract molecular properties of the sample.

Correlation is a mathematical operation that measures the time dependence of the similarity of two signals. The correlation function of two time-dependent signals $f(t)$ and $g(t)$ is given by $(f*g)(\tau) = \langle f(t)g(t+\tau)\rangle$ where $\langle \ldots \rangle$ denotes averaging over time. Often, the correlation function is calculated for the fluctuations $\delta f = f - \langle f\rangle$ and $\delta g = g - \langle g\rangle$,

$$(\delta f * \delta g)(\tau) = \langle \delta f(t)\delta g(t+\tau)\rangle = \langle f(t)g(t+\tau)\rangle - 1 \tag{4.31}$$

FCS, in its most straightforward form, uses the autocorrelation of the fluctuations in the fluorescence signal $I(t)$ caused by molecules freely diffusing through the detection volume. f and g in equation 4.31 are then substituted by the same signal $I(t)$ and normalized by the squared average signal intensity.

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} \tag{4.32}$$

The detection volume or point spread function (PSF) for one photon excitation is assumed to be modeled by a three-dimensional gaussian and is given in equation 4.4.

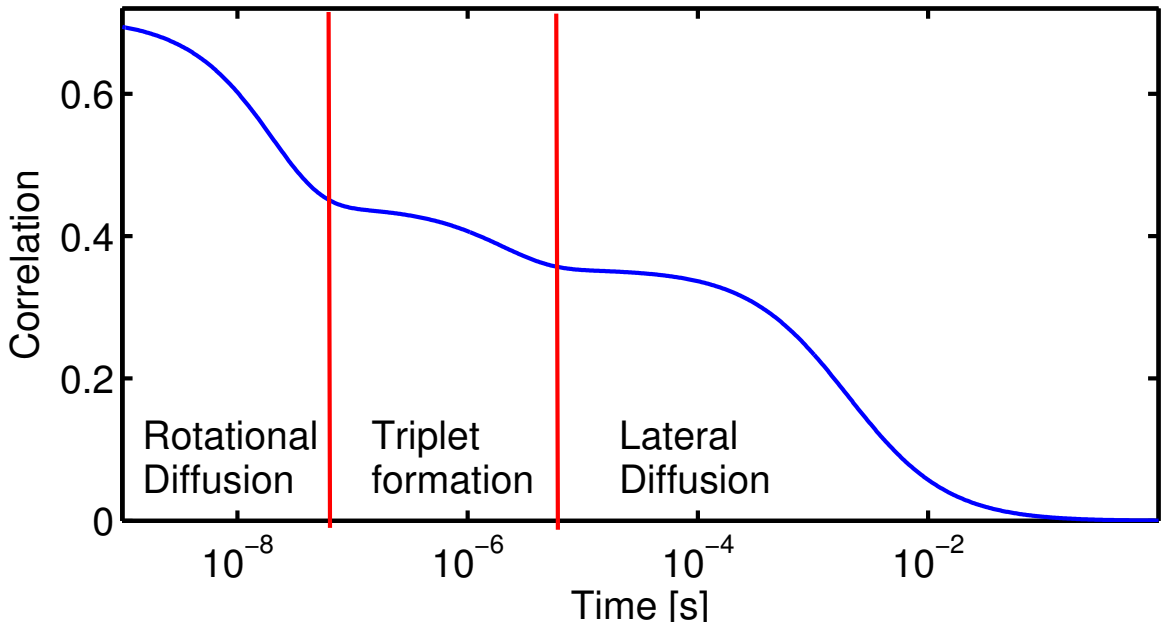


Figure 4.10: *Model graph for an FCS curve with lateral diffusion, triplet formation and rotational diffusion (Equations 4.35, 4.36, and 4.33*

For a single diffusing species with diffusion coefficient D and average concentration $\langle c \rangle$, the autocorrelation function $G(\tau)$ of the signal fluctuations $\delta I(t)$ is

$$G(\tau) = \frac{1}{\langle c \rangle \omega_r^2 \omega_z \pi^{3/2}} \left(\frac{1}{1 + \frac{4D\tau}{\omega_r^2}} \right) \sqrt{\frac{1}{1 + \frac{4D\tau}{\omega_z^2}}} \tag{4.33}$$

This equation is often used with the following substitutions

$$\tau_D = \frac{\omega_r^2}{4D}, \quad p = \frac{\omega_z}{\omega_r}, \quad \langle N \rangle = \langle c \rangle V_0 = \langle c \rangle \left(\frac{\pi}{2} \right)^{3/2} \omega_r^2 \omega_z \tag{4.34}$$

with the translational diffusion time τ_D, the structure parameter p, the average number of molecules in the PSF $\langle N \rangle$, the effective volume of the PSF, $V_0 = \int \mathrm{PSF}(x, y, z) \mathrm{d}x\mathrm{d}y\mathrm{d}z$, and the factor[5] $2^{-3/2}$

$$G_D(\tau) = \frac{2^{-3/2}}{\langle N \rangle} \left(\frac{1}{1 + \frac{\tau}{\tau_D}} \right) \sqrt{\frac{1}{1 + \frac{1}{p^2} \frac{\tau}{\tau_D}}} \tag{4.35}$$

Fitting this model to experimental data yields information about the concentration of the sample molecules in the solution and on the diffusion coefficient D (Figure 4.10). Assuming spherical particles, D is related to the hydrodynamic radius of the molecules by the Einstein-Stokes relation $D = \frac{kT}{6\pi\eta R}$. This makes it a valuable tool e.g. to measure binding of ligands to proteins, aggregation of proteins (Chapter 7) and conformational changes that affect the compactness of a protein (Chapter 5).

[5] This factor is often denoted with γ. This is omitted here to avoid confusion with the detection correction factor for FRET. The factor of $2^{-3/2}$ is due to the gradual decrease of the intensity at the edges of the gaussian confocal volume and is defined by $\frac{\int \mathrm{PSF}^2(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z}{\int \mathrm{PSF}(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z}$

Translational diffusion is not the only process accessible to FCS. Basically, every process causing fluctuations in the fluorescence signal is reflected by a change in the autocorrelation function. For example, many fluorophores can enter a triplet state resulting in a temporary dark state. This process gives an additional exponential decay term in the autocorrelation function. The model equation describing a single diffusing fluorescent species with rates of excitation k_{01}, emission k_{10}, triplet formation k_{1T}, and relaxation from the triplet state k_{T0} (see Figure 2.1 in Chapter 2) was derived in (Widengren *et al.*, 1994, 1995) and given by equation 4.36.

$$G(\tau) = G_T(\tau) \cdot G_D(\tau) = \left(1 + \frac{T}{1-T} e^{-\tau/\tau_{trip}}\right) \cdot G_D(\tau) \tag{4.36}$$

There,

$$T = \frac{k_{1T} k_{01}}{k_{01}(k_{1T} + k_{T0}) + k_{T0}(k_{10} + k_{1T})} \tag{4.37}$$

is the equilibrium fraction of molecules in the triplet state and

$$\tau_{trip} = \left(-k_{T0} - \frac{k_{01} k_{1T}}{k_{01} + k_{10}}\right)^{-1} \tag{4.38}$$

is the characteristic triplet state lifetime on the order of 10 µs.

Detection of other sources of fluctuations than diffusion and triplet formation may require specific experimental setups and the cross-correlation of different signals in contrast to the autocorrelation discussed above. Fluctuations on fast timescales, for example, are obscured by detector dead time and afterpulsing. During the detector dead time of 50 ns, no additional photon can be detected resulting in a strong anti-correlation. However, for hundreds of nanoseconds after the detection of a photon, the probability for a dark count in the APD is increased. This effect is called afterpulsing. The timescale and amplitude of afterpulsing depends on the specific detector and on the count-rate and is typically between 100 ns and 1 µs. These detector-dependent effects can be filtered out by using a 50:50 beamsplitter to split the signal on two independent detectors. In the cross-correlation function of the signals from the two detectors, only fluctuations of the actual fluorescent signal will be observed. An example for a fast process that makes this experimental setup necessary is rotational diffusion, which happens on a timescale of $\approx 100\,\mathrm{ns}$ for proteins and $\approx 1 - 10\,\mathrm{ns}$ for small molecules like dyes. A polarizer in the detection pathway allows selective detection of photons in a plane parallel to the excitation laser polarization. Rotation of the dye results in an exponential decay of the correlation function with the rotational diffusion time $\tau_R = 1/(6D_R)$. The rotational diffusion is more sensitive to changes in the molecule's size than translational diffusion, $D_R = \frac{kT}{6\pi\eta R^3}$. However, the rotational motion of the protein is overlaid with faster tumbling of the dyes on the linkers connecting it to the protein of interest (Widengren *et al.*, 1999).

4.2.2 Two focus Fluorescence Correlation Spectroscopy (2fFCS)

The determination of hydrodynamic radii from FCS data requires the calibration of the size of the confocal volume, i.e. ω_r and ω_z, which is typically done by measuring a dye with known diffusion coefficient. This constraint can be overcome by the introduction of an additional length scale into the system. Enderlein and coworkers achieved this by creating two spatially separated foci by combining pulsed interleaved excitation of two perpendicularly polarized lasers with a Nomarski prism. The prism, positioned between the dichroic separating excitation and detection and the objective, deflects the pulses of different polarization at slightly different angles, resulting in a lateral shift of the two foci. The distance D between the foci for a given wavelength depends only on the Nomarski prism

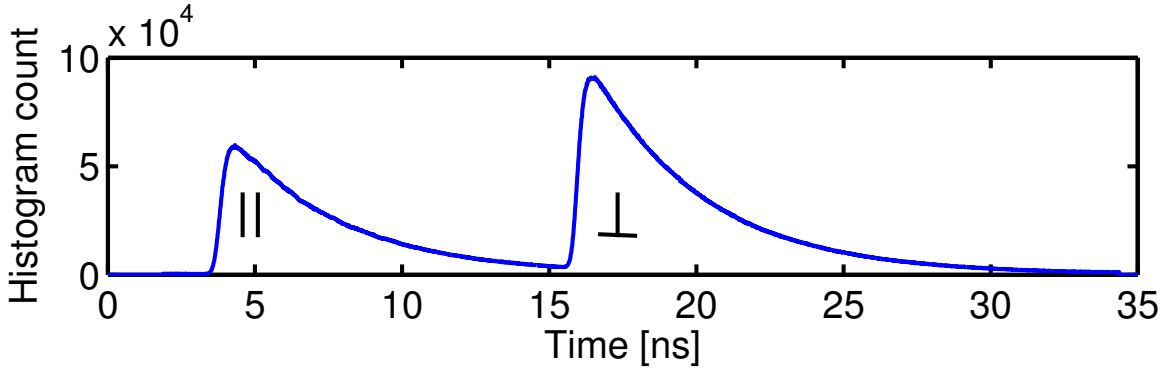


Figure 4.11: *Typical TCSPC histogram for a 2fFCS measurement. The photons originating from the two different foci can be clearly distinguished by their* microtimes. *Photons arriving between 3 ns and 16 ns originate from the focus formed with parallel polarized light while photons arriving between 16 ns and 35 ns originate from the focus formed with perpendicular polarized light.*

and the objective in the setup offering the desired absolute length scale. With known d, a global fit of the autocorrelation curves for the signal coming from the two foci and the cross-correlation curve of the two signals yields accurate values for the diffusion coefficient of the sample (Dertinger *et al.*, 2007). Fluorescence from the two different foci is distinguished by the TCSPC arrival time as in PIE (Figure 4.11).

2fFCS has been realized in the MFD setup described in Section 4.1 (Figure 4.12). The interleaved pulses for excitation are produced by one laser and a polarizing beamsplitter cube directly coupled to the optical fibres (PBS, OZ Optics, FOBS-22P-1111-3.5/125-MPPP-532-PBS-40-3A3A3A3A-3-2,1.5,1,2). The laser light enters the beamsplitter through a multimode fiber, thus loosing most of its polarization. The PBS reflects light perpendicular to the beam plane[6] directly to the fiber leading to the microscope. Light having the opposite polarization is transmitted into a second fiber to delay the pulses before entering the PBS from the fourth side and is transmitted to the output fiber (Figure 4.12). This produces a train of pulses with alternating polarization that is directed into the microscope. In front of the objective, the Nomarski prism splits the two polarizations, resulting in two spatially separated foci.

Now, autocorrelation curves for each focus and a cross-correlation function between the two foci can be calculated. The spatial shift between the two foci results in an additional exponential term in the cross-correlation function depending on the separation between the two foci d resulting in the following model function:

$$G(\tau,d) = \frac{1}{\langle c\rangle}\sqrt{\frac{16}{\pi^3 D\tau}}\left[\int_{-\infty}^{\infty} dz_1 \int_{-\infty}^{\infty} dz_2 \frac{\kappa(z_1)\kappa(z_2)}{8D\tau+\omega^2(z_1)+\omega^2(z_2)}\cdot exp\left[-\frac{(z_2-z_1)^2}{4D\tau}-\frac{2d^2}{8D\tau+\omega^2(z_1)+\omega^2(z_2)}\right]\right]\cdot\left[\int_{-\infty}^{\infty} k(z)dz\right]^{-2} \tag{4.39}$$

[6]The plane through all four fibers connected to the cube.

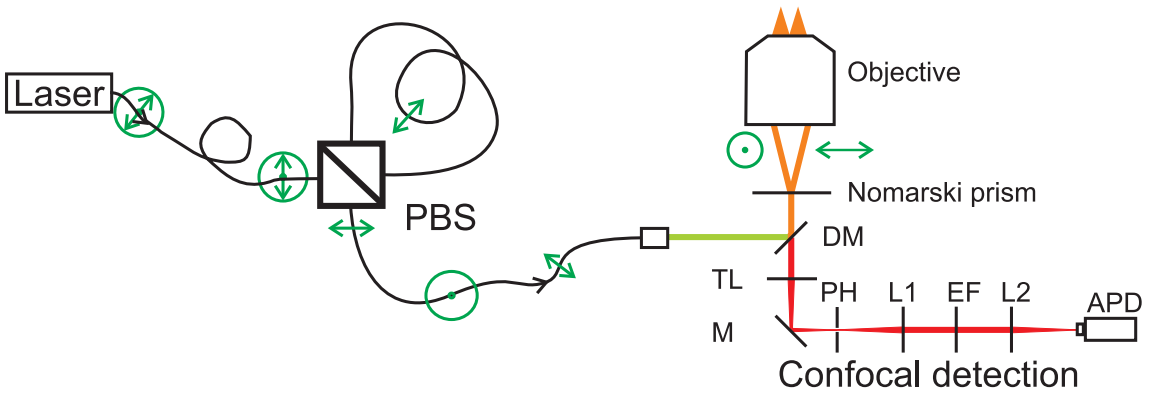


Figure 4.12: *Schematic of the confocal setup with the additional changes necessary for 2fFCS. The left part shows how the pulse train of alternating polarizations is generated with a polarizing beam splitter (PBS) and an optical delay fiber. In the right part, the Nomarski prism splits the two polarizations and creates the two spatially separated foci as shown. The drawing is not to scale. L1,L2: lenses; TL: tube lens ; DM: dichroic mirror; M: mirror; PH: pinhole; EF: emission filter; APD: avalanche photo diode.*

The functions $\kappa(z)$ and $\omega(z)$ are used to describe the detection volume instead of the three-dimensional gaussian used for one-focus FCS to empirically account for the z-dependent width of the focus. They are given by

$$\begin{aligned} \omega(z) &= \omega_0 \left[1 + \left(\frac{\lambda_{ex} z}{\pi \omega_0^2 n}\right)^2\right]^{1/2} \\ \kappa(z) &= 1 - exp\left(-\frac{2a^2}{R^2(z)}\right) \end{aligned} \tag{4.40}$$

with

$$R(z) = R_0 \left[1 + \left(\frac{\lambda_{em} z}{\pi \omega_0^2 n}\right)^2\right]^{1/2} \tag{4.41}$$

ω_0 is the lateral beam waist at the focus of the laser, λ_{ex} and λ_{em} are the wavelengths of excitation and emission, respectively, n is the index of refraction of the sample, a is the confocal pinhole radius in the sample plane, and R_0 is a model parameter.

A global fit of the three correlation functions yields the most accurate values for the diffusion coefficient of the sample. The fit was performed using a custom MATLAB function generously provided by Jörg Enderlein.

The spatial separation between the two foci can be determined by several methods. Either a dye with known diffusion coefficient D is used to calibrate d using the fit model described above with a fixed diffusion coefficient D and the distance d as fitting parameter. A reference diffusion coefficient of $D = 414\,\mu\mathrm{m}^2/\mathrm{s}$ was chosen for the dye Atto 532[7] yielding a foci separation of $d = 415\,\mathrm{nm}$ for the setup described above (Figure 4.13A). This method relies on prior knowledge of the diffusion coefficient similar to standard FCS, and may therefore produce systematic errors in the measured foci separation.

[7] This is the diffusion coefficient of Rhodamine 6G (Müller *et al.*, 2008). The same diffusion constant is assumed for Atto 532 since it is a derivative of Rhodamine 6G

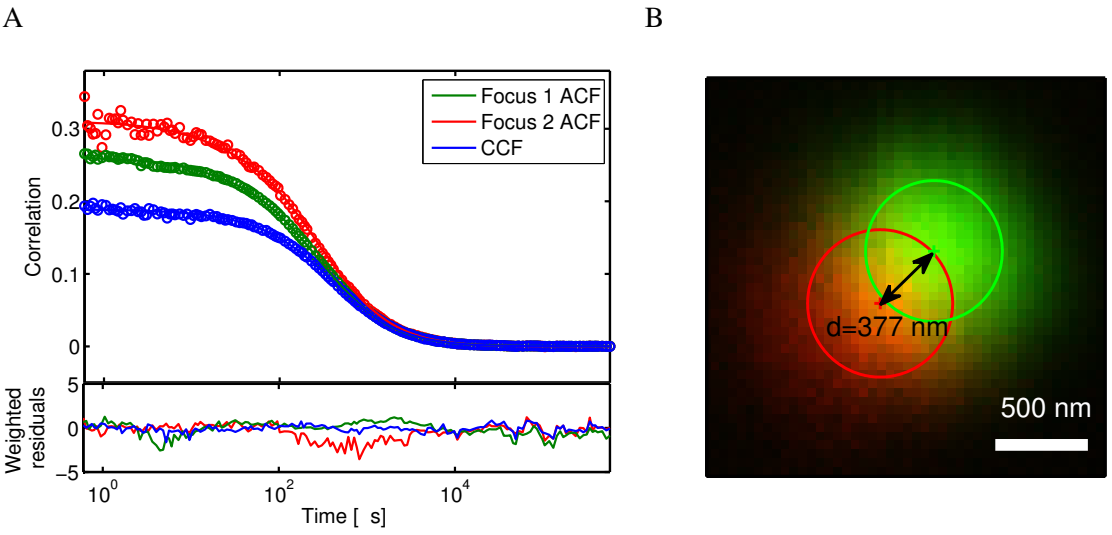


Figure 4.13: *Determination of the foci separation in 2fFCS. (A) Fit of the ACFs and CCF for 2fFCS. The raw data is shown as circles, and fits using equation 4.39 are shown as lines. Weighted residuals of the fit are shown in the lower graph. (B) Image of the two-focus point spread function obtained by scanning a bead. The two foci are shown in red and green together with the 1/e radius of a two-dimensional gaussian fit. The distance between the centers of the foci is 377 nm.*

An independent value for d can be found by measuring the point spread functions of the two foci directly, usually by scanning a bead that is smaller than the confocal volume. Figure 4.13B shows bead scans performed on the setup described above together with the $1/e$ line of gaussian fits to the PSFs. These measurements resulted in a foci separation of $d = 377\,\mathrm{nm}$. The deviation from the value determined from the fit may have multiple reasons. The diffusion coefficient assumed for Atto 532 may not be correct due to different experimental conditions (temperature, viscosity of buffer, functional groups on dye,...). For example, when Müller *et al.* (2008) determined the diffusion coefficient of Rhodamine 6G using 2fFCS, they found significant variation of the diffusion coefficients of Rhodamine 6G purchased from different manufacturers. On the other hand, there may be a systematic error in the value for d determined by the bead scans due to an error in the calibration of the pixel size for the confocal scanning.

A third method for the determination of d that may resolve the issues discussed above was not used in this work. It is possible to calculate d by z-scan FCS using a two-dimensional sample, e.g. a lipid bilayer of an immobilized giant unilamellar vesicle (Dertinger *et al.*, 2007). However, for this work, only measurements of the changes in the diffusion coefficient were necessary and it was thus not necessary to perform an unambiguous determination of d. For the determination of relative changes in the diffusion coefficient of a molecule, 2fFCS is superior to one focus FCS due to its insensitivity to optical artifacts introduced e.g. by changes in refractive index of the buffer or varying coverslip thickness (Enderlein *et al.*, 2004).

4.2.3 Two color Fluorescence Cross-Correlation Spectroscopy

In the biological or pharmaceutical sciences, it is often of particular interest to determine whether two molecules interact and if so, to what extend. The binding efficiency is often evaluated by fluorescence methods usually based on the change in mobility of the smaller ligand when bound to the protein compared to the free ligand. Typically, measurements of the fluorescence anisotropy detect changes

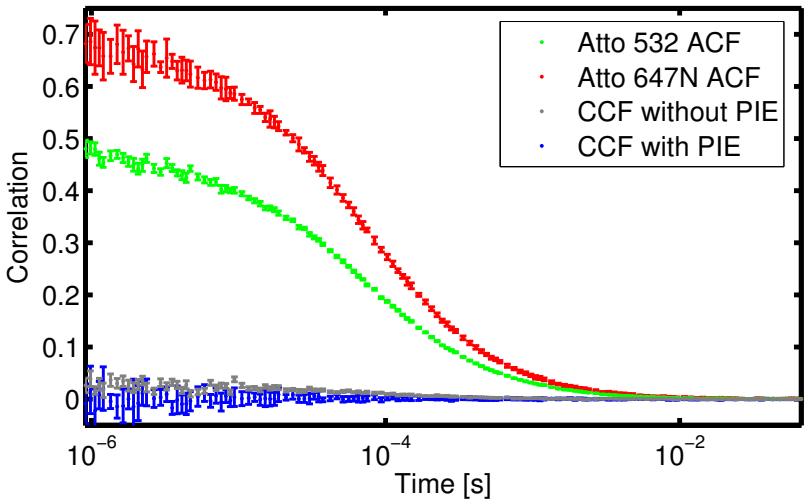


Figure 4.14: *Increased sensitivity of FCCS with PIE. Autocorrelation curves and cross-correlation curves with and without PIE for a mixture of Atto 532 and Atto 647N. Without PIE, a residual artifical cross-correlation is observed for the independently diffusing dyes due to spectral crosstalk of Atto 532 signal in the Atto 647N detection channel. Using PIE, the crosstalk photons can be removed resulting in increased sensitivity for weakly interacting molecules.*

of the rotational mobility on a timescale of the fluorescence lifetime while FCS detects both changes in the rotational mobility on longer timescales and the translational mobility. These measurements, however, often lack the sensitivity to measure low binding efficiencies or binding of molecules of approximately the same size. To deal with such situations, two-color fluorescence cross-correlation spectroscopy (2cFCCS) is an ideal complementary method.

In 2cFCCS, the two binding partners are labeled with two fluorophores that can be spectrally separated into different detection channels. The ACFs of the two signals will then carry information about the concentrations and diffusion coefficients of the two species. The cross-correlation of two detection channels, however, detects only fluctuations that are present in both detection channels, i.e. diffusing molecules that carry both labels at the same time.

The cross-correlation function for two differently labeled species i and j with average concentrations of unbound molecules $\langle c_i \rangle$ and $\langle c_j \rangle$ and bound molecules $\langle c_{ij} \rangle$ observed with a three-dimensional gaussian point spread function is given by

$$G(\tau) = \frac{\langle c_{ij} \rangle}{(\langle c_i \rangle + \langle c_{ij} \rangle)(\langle c_j \rangle + \langle c_{ij} \rangle)} \frac{1}{\omega_r^2 \omega_z \pi^{3/2}} \left(\frac{1}{1 + \frac{4D\tau}{\omega_r^2}} \right) \sqrt{\frac{1}{1 + \frac{4D\tau}{\omega_z^2}}} \tag{4.42}$$

In combination with PIE, the sensitivity of 2cFCCS for weak binding is significantly enhanced. When using simultaneous continuous excitation with both colors, a fraction of the photons in the detection channel at longer wavelength originates from the species labeled with the short-wavelength dye due to spectral crosstalk. Of course, these photons are correlated with the photons detected from the short-wavelength dye in the short-wavelength detection channel, resulting in a residual cross-correlation even for non-interacting species. These photons can be removed by selecting only photons originating from acceptor excitation (Figure 4.3) resulting in a flat cross-correlation curve (Figure 4.14 and Müller *et al.*, 2005).

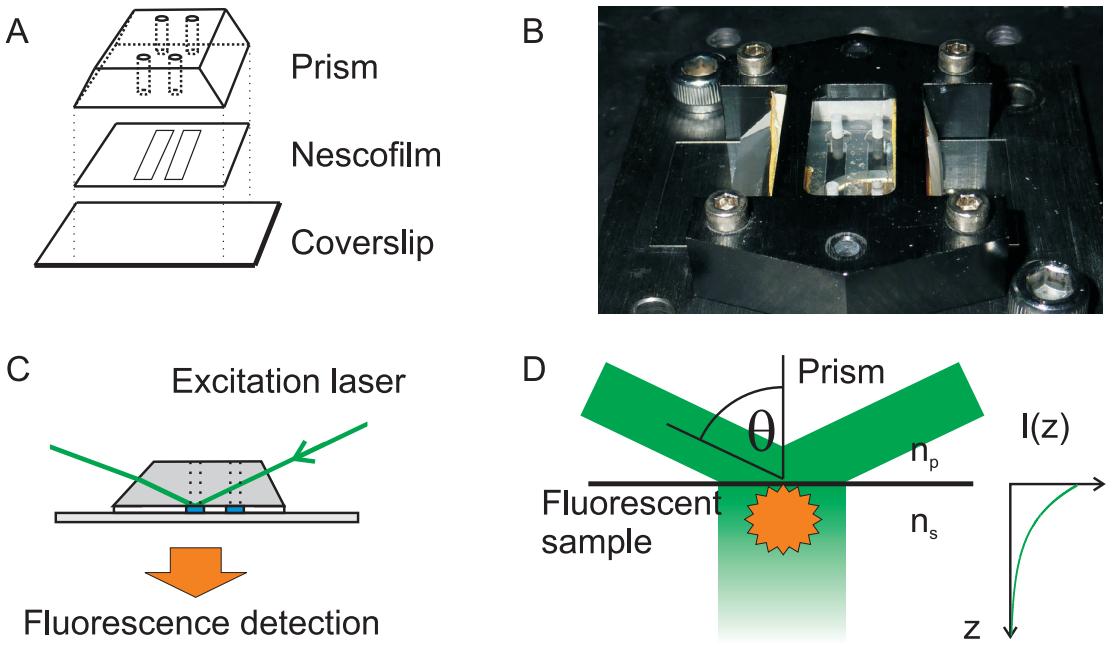


Figure 4.15: *Sample chamber and excitation pathway for single-molecule TIRF experiments. (A) Schematic drawing of the sample chambers. Each prism has two pairs of holes to flush the sample channels cut into the Nescofilm. (B) Photograph of the sample chamber on the microscope. (C) Schematic drawing of the sample chamber and excitation pathway. (D) TIRF excitation. The excitation laser hits interface between the prism (refractive index n_p) and the sample solution (n_s) at a flat angle θ and is totally reflected. The sample is excited by the exponentially decaying evanescent field.*

4.3 Single-molecule measurements using a total internal reflection (TIRF) microscope

The observation time for one molecule in confocal microscopy is limited by its diffusion time through the detection volume. To follow the dynamic conformational changes of a single molecule over longer timescales, immobilization of the molecule is necessary. For these investigations, I have used spFRET measurements on a total internal reflection fluorescence (TIRF) microscope . TIRF uses the evanescent wave of a totally internally reflected laser beam to reduce the detection volume.

4.3.1 TIRF excitation

Two experimental realizations of TIRF microscopy are commonly employed. Objective-type TIRF is usually preferred for measurements on living cells while single-molecule measurements benefit of the higher signal-to-noise ratio of prism-type TIRF. For both, only fluorophores in a thin ($\approx 100\,\mathrm{nm}$) layer next to the surface are excited by the evanescent field of a totally reflected laser beam. In the setup used in this thesis, the sample was immobilized on a quarz prism coated with a polyethylene glycol (PEG) layer. The prism has two holes for sample injection and is joint with a cover glass by a Nesco film (Karlan) containing a channel as sample chamber (Schlüsche, 2007). Figure 4.15 shows a schematic (A) and a photograph (B) of the prism and sample chamber.

The excitation lasers are selected by an acousto-optical tunable filter (AOTF) and coupled into a single-mode optical fiber to achieve a gaussian beam profile. The laser is then directed into the

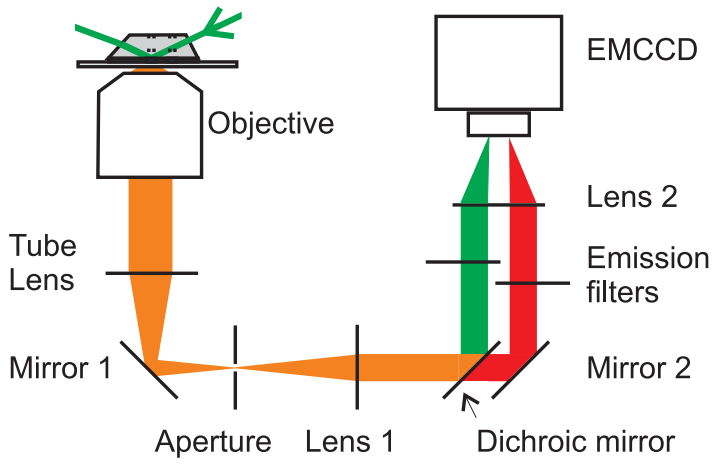


Figure 4.16: *Detection pathway of the TIRF setup. The fluorescence is collected by the objective, focused on an aperture by the microscope tube lens, re-collimated and split into the spectral regions of donor and acceptor fluorescence, respectively. The spectrally distinct light beams are then focused on different regions of an EMCCD camera.*

prism at a flat angle (Figure 4.15C). The lasers enter the prism perpendicular to the prism surface to avoid refraction. At the bottom surface of the prism, the beam is totally reflected resulting in an exponentially decaying evanescent field in the sample chamber (Figure 4.15D). The intensity $I(z)$ as a function of distance from the surface z is given by

$$I(z) = I_0 \exp\left(-\frac{z}{d}\right) \tag{4.43}$$

The penetration depth d is a function of excitation wavelength λ_{ex}, refractive indices of the prism n_p and the sample chamber n_s, and the angle of the incident laser light from the normal of the prism surface θ (equation 4.44) and is approximately $d = 100\,\mathrm{nm}$..

$$d = \frac{\lambda_{ex}}{4\pi}\left(n_p^2 \sin^2(\theta) - n_s^2\right)^{-1/2} \tag{4.44}$$

4.3.2 Fluorescence detection

Fluorescence is collected by a high N.A. objective (Nikon Plan Apochromat 60x NA 1.2) and directed through an aperture in the focal plane of the microscope tube lens ($f = 200\,\mathrm{mm}$). An achromatic lens ($f = 100\,\mathrm{mm}$) again collimates the light and then a dichroic mirror (630DCXR, AHF) splits donor and acceptor emission. Individual emission filters (HQ585/80 and HQ715/150) reject scattered laser light. The spectrally split fluorescence from the donor and from the acceptor is imaged by another achromatic lens ($f = 200\,\mathrm{mm}$) on different regions of an electron multiplying charged coupled device (EMCCD) camera (iXon DV887 ECS-BV, Andor). A scematic drawing of the detection pathway is shown in Figure 4.16.

The EMCCD camera is controlled by a computer running the software provided by Andor (Solis). A more detailed description of the EMCCD camera is given in appendix A.1 and the settings of the EMCCD collection parameters are given in Table A.2.

The spatial overlap between the two spectrally distinct images on the EMCCD chip is optimized optically by adjustment of the dichroic mirror and mirror 2 in Figure 4.16. A MATLAB program was developed to automatically acquire images using dll files provided by the Andor Software Development Kit and to calculate the image cross- correlation between the two halfs of the image. The position of the maximum cross-correlation corresponds to the shift between the images.

4.3.3 Surface preparation and immobilization

The restriction of the excitation volume to a thin layer reduces the background from scattered light but makes immobilization of the molecules of interest on the surface necessary. Immobilization by Biotin-Streptavidin-Biotin linkage is one of the most common immobilization procedures. Surface preparation is described in detail in (Schlüsche, 2007) and only a short description is given here.

The prisms were cleaned by first boiling them in 3% Hellmanex solution followed by extensive washing with Millipore water. The old silane layer is removed by incubation in a 1:2 mixture of HCl (37%) and H_2O_2 (30%) for 30 min. The prisms are then washed again with Millipore water, dried under a steam of Nitrogen, and singed. For a new silanization layer, the prisms were put in a 2% solution of Aminosilane in Acetone for 40 min, again followed by extensive washing with Millipore water and drying under Nitrogen. A layer of Polyethylene glycole (PEG) is bound to the silane by incubation of the prisms with a 45% PEG solution in Carbonate buffer ($NaHCO_3$, pH 9.0) containing 2% biotinylated PEG for 1 hour. After another washing-and-drying step, the sample chambers are assembled by heating the prism-Nescofilm-coverslip sandwich to $150\ ^\circ\mathrm{C}$, covered in Parafilm and stored in Falcon tubes at $-20\ ^\circ\mathrm{C}$.

Using Streptavidin, proteins can be bound directly to the biotinylated surface by the introduction of a biotin label in the protein. However, this can result in unwanted effects due to the interaction of the protein of interest with the surface (e.g. by steric hindrance of conformational changes). These problems can be circumvented by encapsulation of the protein into liposomes and immobilization of the liposomes (Boukobza *et al.*, 2001; Okumus *et al.*, 2004). The protein then freely diffuses inside the liposome but is still confined for TIRF microscopy. The diffusion inside the vesicle is sufficiently fast even for large proteins to average out intensity fluctuations due to the exponential excitation profile.

4.3.4 Vesicle encapsulation for TIRF microscopy

The sample has to be immobilized to allow observation by single-molecule TIRF microscopy over long (seconds to minutes) timescales. To avoid artifacts from the immobilization, vesicle encapsulation and subsequent immobilization of the vesicles was used (Figure 4.17, Boukobza *et al.*, 2001; Okumus *et al.*, 2004). Lipid films containing 300 µg 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC, Avanti Polar Lipids, Alabaster, USA) and 6 µg biotinylated lipid (1,2-Dipalmitodyl-sn-Glycero-3-Phosphoethanolamine-N-(Cap Biotinyl) (Sodium Salt)) were prepared by mixing dissolved lipid in chloroform, evaporation of the chloroform under a stream of nitrogen and subsequent removal of residual chloroform in vacuum for at least one hour.

The lipid films were hydrated with buffer containing the sample and all other components necessary for the experiment for 30 minutes at $4\ ^\circ\mathrm{C}$. This procedure results in the formation of multilamellar vesicles (MLVs). The MLVs were then extruded 31 times through polycarbonate membranes with a pore diameter of 200 nm using an extruder (Liposofast Basic, Avestin Europe GmbH, Mannheim,

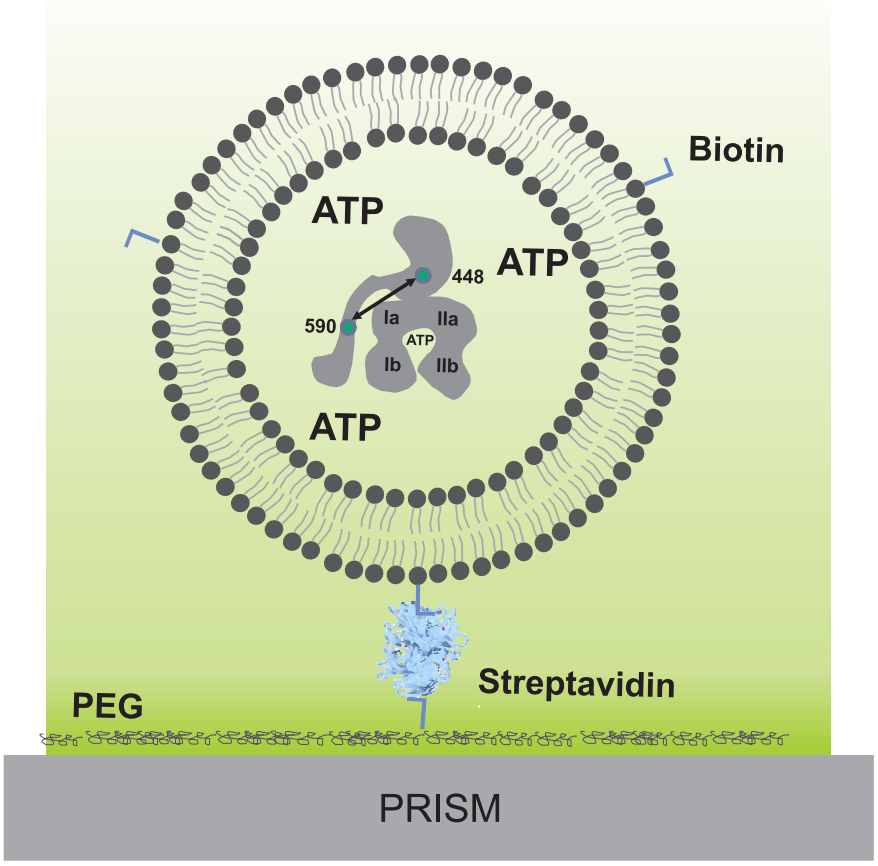


Figure 4.17: *Immobilization for single-molecule TIRF experiments by vesicle encapsulation. Double-labeled Ssc1 is encapsulated together with nucleotide, substrate, or cochaperones in lipid vesicles. The vesicles are immobilized using a biotin-streptavidin-biotin linkage between biotinylated lipids in the vesicle and biotinylated PEG on the surface, and viszualized using TIRF microscopy.*

Germany). This procedure breaks up the MLVs to large unilamellar vesicles (LUVs) with a homogeneous size distribution of 200 nm diameter.

The vesicles were immobilized on a quartz prism covered with a PEG/3% Biotin-PEG layer using a Biotin-Streptavidin-Biotin linkage (Section 4.3.3). SpFRET TIRF measurements were then performed as described in Section 4.3.

The number of encapsulated fluorescent sample molecules per vesicle is described by a Poisson distribution. To maximize the number of vesicles encapsulating exactly one sample molecule, a sample concentration corresponding to an average of one sample molecule per vesicle (400 nM) was chosen. The number of sample molecules per vesicle was quantified in the analysis by counting the number of bleaching steps per trace. All intensity traces showing more than one distinct bleaching step were discarded from the analysis.

4.3.5 Data analysis

The raw data taken in spFRET TIRF measurements consists of movies of images taken from a single field-of-view. One half of each image consists of fluorescence in the spectral region of donor emission while the other half contains the fluorescence in the spectral region of acceptor emission.

Even if great care is taken to optically overlap the spectrally distinct images, a small shift on the order of one pixel cannot be circumvented (Figure 4.18). Complete overlap then can only be achieved by application of a coordinate transformation consisting of translation and dilation operations. The correct coordinate transformation can be derived by measurement of a sample that is clearly visible in both spectral regions at the same time, e.g. fluorescent beads with a broad emission spectrum. This allows the selection of control points in the images that can be mapped on top of each other using the MATLAB function *cp2tfrom*. This transformation is then applied to extract the intensity traces from the correct transformed coordinates in the raw data movies. The coordinate transformation was applied to the image in Figure 4.18A to visualize the better overlap after the transformation (Figure 4.18B).

A peak finding algorithm now detects the positions of single molecules in the movies. Intensity time traces of donor and acceptor fluorescence are then extracted from the movie by averaging of the counts from a circular area of 3 pixel radius in each frame of the movie. The averaged counts from a ring around the molecule (3 pixel inner radius and 5 pixel outer radius) are used as background and subtracted from the intensity.

The FRET efficiency E can be calculated from the background-corrected fluorescence intensities of the donor I_D and the acceptor I_A after donor excitation as

$$E = \frac{I_A - \mathrm{CT} \cdot I_D}{\gamma \cdot I_D + I_A - \mathrm{CT} \cdot I_D} \tag{4.45}$$

Depending on the trace, different possibilities for determining the correction factors exist (Figure 4.19). The correction factors for crosstalk of donor fluorescence in the acceptor detection channel CT and for different quantum yields and detection efficiencies γ can be determined directly for a specific molecule if the acceptor bleaches before the donor (Figure 4.20). The corrected total intensity has to stay constant until the donor bleaches. Thus, the average corrected total intensities before bleaching of the acceptor $\gamma\langle I_D\rangle_{before} + \langle I_A\rangle_{before} - \mathrm{CT}\langle I_D\rangle_{before}$ and after bleaching $\gamma\langle I_D\rangle_{after} + \langle I_A\rangle_{after} - \mathrm{CT}\langle I_D\rangle_{after}$ have to be equal, hence

$$\gamma - \mathrm{CT} = \frac{\langle I_A\rangle_{before} - \langle I_A\rangle_{after}}{\langle I_D\rangle_{after} - \langle I_D\rangle_{before}} \tag{4.46}$$

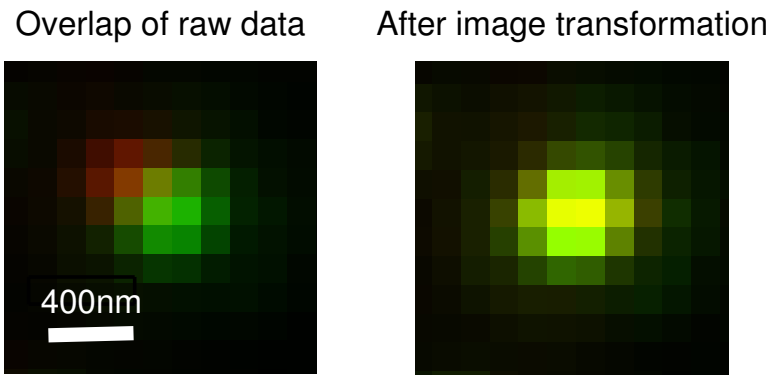


Figure 4.18: *Overlay of the green and red spectral parts of one molecule. A clear shift between the detection channels exists before the application of the image transformation. After image transformation, the two peaks are perfectly overlapped.*

CT can be calculated from the intensities in the donor and acceptor detection channels after bleaching of the acceptor.

$$\mathrm{CT} = \frac{\langle I_A \rangle_{after}}{\langle I_D \rangle_{after}} \tag{4.47}$$

For molecules showing dynamic switching between states of different FRET efficiency but no acceptor bleaching, γ was determined by minimizing the variance of the overall intensity. Special care has to be taken when this method is applied that γ does not depend on the conformational state of the molecule. This can easily be the case since the quantum yields of the dyes depend on their environment. For all measurements shown in this work, however, complementary MFD burst analysis measurements made it possible to rule out any dependence of γ on the conformational state.

For molecules where no direct determination of the correction factors was possible, average values for γ and CT determined from all directly accessible values were applied.

4.3.6 Hidden Markov Modeling

A Markov process is defined as a time-dependent random process where the transition probability between different, observable states is memoryless, i.e. does not depend on its past history. Consequently, the transition probabilities are the only unknown variables when modeling a Markov process to experimental data. A Markov process becomes hidden if the states are not observable anymore, e.g. by overlaid noise. In a *Hidden Markov Model* (HMM), not only the transition probabilities between the states but also the states themselves are unknown.

Hidden Markov Modeling was developed in the 1960s (Baum and Petrie, 1966) and was used mostly for speech recognition applications for two decades (Rabiner and Juang, 1986). After the rise of rapid, reliable DNA sequencing techniques in the 1980s, HMM analysis was also frequently applied to bioinformatic questions (Durbin, 1998). A typical example of a biophysical application of HMM

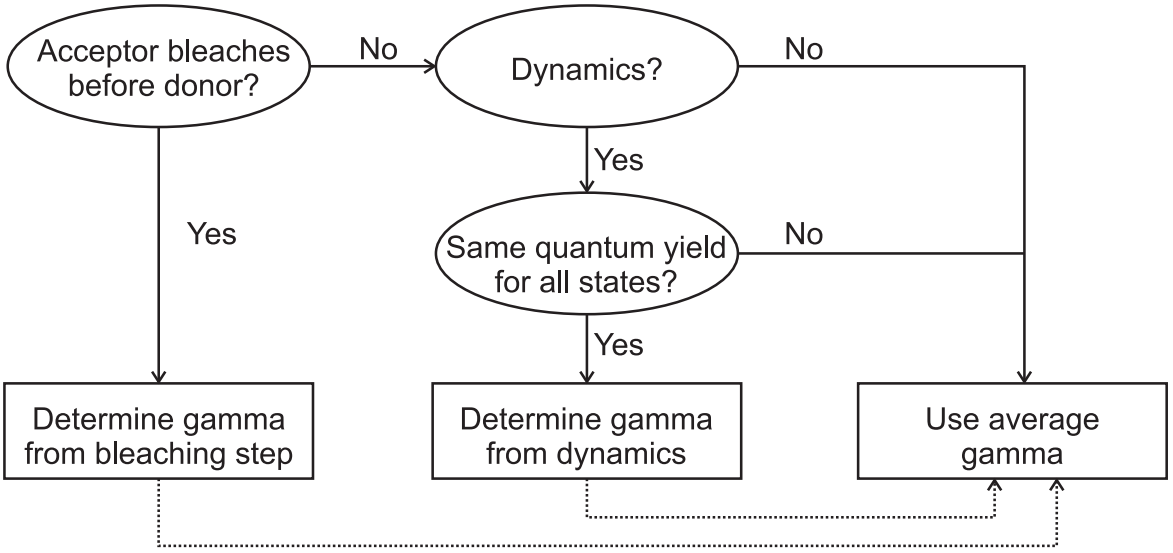


Figure 4.19: *Decision process of the gamma correction to be applied to a specific trace.*

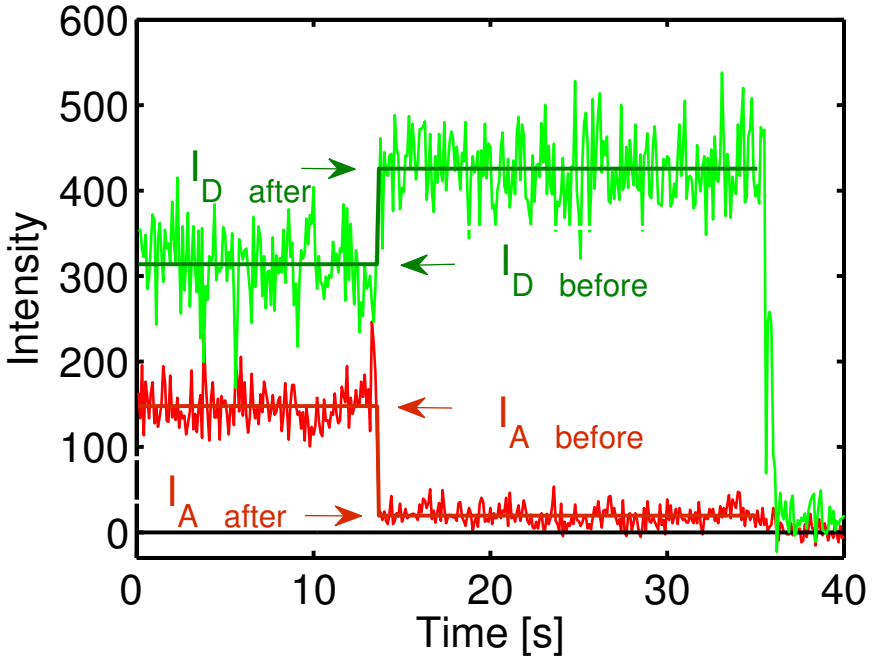


Figure 4.20: *Determination of the γ factor from acceptor photobleaching. The dashed lines show the intensity traces in the donor (green) and acceptor (red) channels. Acceptor photobleaching occurs at $t = 14\,\mathrm{s}$, donor photobleaching after $t = 36\,\mathrm{s}$. The average intensities of donor and acceptor before and after photobleaching of the acceptor are indicated by solid lines. For this trace, $\gamma = 1.15$ and $\mathrm{CT} = 0.045$ was determined.*

is modeling of a time trace of the current through an ion channel switching between an *open* and a *closed* state observed by electric single channel recordings (Qin *et al.*, 2000). Single-molecule fluorescence experiments of a dynamic system, e.g. a protein switching between two or more states, is a very similar problem and HMM analysis has consequently also been applied to this type of data. Several authors used HMM to model the sub-ms dynamics hidden in single-photon streams recorded from individual molecules using TCSPC detection (Yang *et al.*, 2003; Andrec *et al.*, 2003; Schröder and Grubmüller, 2003). In this thesis, however, I used HMM to extract dynamics on the timescale of several seconds from single-molecule traces collected by TIRF microscopy as proposed by McKinney *et al.* (2006). The HMM analysis was performed using a MATLAB HMM toolbox written by Kevin Murphy (Murphy, 1998).

For a molecule switching dynamically between different conformational states, the HMM model parameters are the underlying FRET states $\{E_1, E_2, \ldots\}$ that are overlaid by noise. The noise is assumed to be gaussian with standard deviations $\{\sigma_1, \sigma_2, \ldots\}$, an assumption that greatly simplifies the analysis and that has been shown to be sufficiently accurate (Dahan *et al.*, 1999; McKinney *et al.*, 2006). Transitions between the states are described by the transition probability matrix

$$P = \begin{pmatrix} p_{1\to1} & p_{1\to2} & \cdots \\ p_{2\to1} & p_{2\to2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{4.48}$$

where $p_{i\to j}$ denotes the probability of a transition to state j in the next time step $t_0 + \Delta t$ if the molecule is in state i at time t_0. The time step Δt is defined by the exposure time of the experiment. Of course, the probabilities for the transitions originating from state i have to sum up to one, $\sum_j p_{i\to j} = 1$. The transition probabilities are related to the transition rates between states by $K = (P - I)/\Delta t$, where I is the identity matrix[8]. Consequently, the transition rate matrix K has the following form:

$$K = \begin{pmatrix} -\sum_{j\neq i} k_{1j} & k_{12} & \cdots \\ k_{21} & -\sum_{j\neq i} k_{2j} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{4.49}$$

Assuming a given set of underlying parameters, the *Viterbi* algorithm (Viterbi, 1967) calculates the most likely sequence of states for an experimental FRET trace based on step-by-step probabilities (McKinney *et al.*, 2006). The Viterbi algorithm also provides the total probability of the most likely state sequence for a given parameter set. Thus, optimization procedures can be used to find the set of model parameters with maximum probability, i.e. the model that best describes the experimental data. The software used in this work employs the Baum-Welch algorithm (Baum *et al.*, 1970; Bilmes, 1998) to perform the optimization. Figure 4.21 shows an exemplary dynamic FRET trace with the Viterbi path calculated by the HMM analysis overlayed in orange.

It is possible to perform a global HMM analysis using one parameter set to describe all traces. However, the exact values of the FRET efficiencies in different traces can vary, e.g. due to variations in the γ factor. Therefor, it is difficult to model all traces with the same FRET efficiency states, and the HMM analysis was performed individually for all FRET traces in one dataset. Analysis of simulated FRET time traces of a two-state system showed that individual and global HMM analysis yield similar results (Table 4.3).

[8] This relation is based on the first term a Taylor expansion of the transition probabilities in terms of the transition rates and is only approximately true for transition probabilities $p_{i\to j} \ll 1$

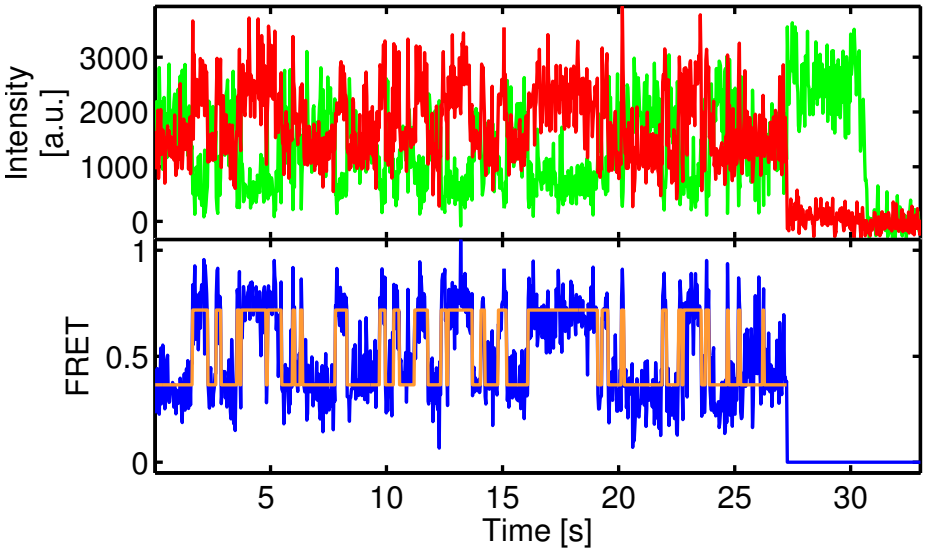


Figure 4.21: *Hidden Markov Modeling of spFRET time traces. Time traces of donor intensity (upper panel, green), acceptor intensity (upper panel, red), and FRET efficiency (lower panel, blue) of a molecule dynamically switching between two FRET efficiencies. The idealized trace found by the HMM analysis is overlaid with the FRET efficiency (lower panel, orange).*

Each transition is then plotted as a two-dimensional gaussian in a *transition density plot* (TDP, Figure 4.22) with the FRET efficiency before and after the transition as x and y center coordinates, respectively. The standard deviation of the noise found by the HMM is plotted as width of the two-dimensional gaussian. The volume under the curve of the two-dimensional gaussian is normalized to one for each transition. Similar transitions then form a cluster in the transition density plot. In Figure 4.22, two clusters of transitions are clearly distinguished, corresponding to transitions between a 40% FRET state and a 80% FRET state. The transitions in one cluster can now be selected for a dwell time analysis. For all selected transitions, the dwell times of the state before the transition are plotted in a cumulative dwell time histogram (Figure 4.23). There, the number of molecules still in the initial state are plotted over time. This avoids potential artifacts from binning of the data for a histogram of the dwell times. The cumulative dwell time histograms are then fitted to extract rates of the transitions using MATLAB.

Theoretically, the kinetic rates of the system are directly modeled by the HMM analysis. However, the cumulative dwell time histograms provide an additional way for the determination of the rates and allow direct visualization of the kinetics. HMM analysis of simulated single-molecule time traces were performed to show that the rates determined by the HMM and from fitting of the cumulative dwell time histograms are equal within the experimental error (Table 4.3). Additionally, although the precision of the mean error of the HMM rates for the individual traces is high, the individual error is high compared to the other methods due to trace-to-trace variations of the optimized parameters.

Table 4.3: *Kinetic rates given in units of s^{-1} determined by different analysis methods from simulated data (100 traces of 100 s with 100 ms time resolution).*

	k_{12}	k_{21}
Simulation parameters	1.000	1.000
Local HMM	1.001 ± 0.199[1]	0.976 ± 0.226
Global HMM	0.982[2]	1.005
Cumulative dwell time histogram fit	1.059 ± 0.012[3]	1.008 ± 0.008

[1] Standard deviation of rates of individual HMM analysis
[2] No error estimate available from global HMM analysis
[3] Error from Jacobian of monoexponential least-squares fit

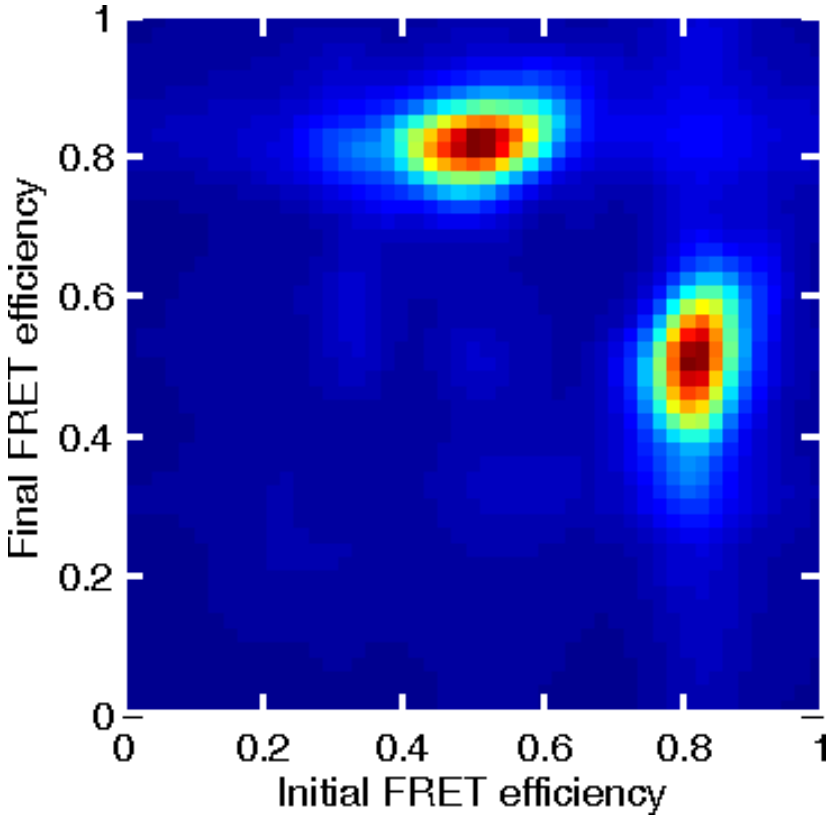


Figure 4.22: *Transition density plot for the transitions detected by the HMM. Each transition is represented by a two-dimensional gaussian. The FRET efficiency before the transition is plotted on the x axis, the FRET efficiency after the transition on the y axis. The width of the gaussian is given by the standard deviation of the noise determined by the HMM analysis. The amplitude of the sum over the two dimensional gaussians for all transitions is color-coded from blue to red.*

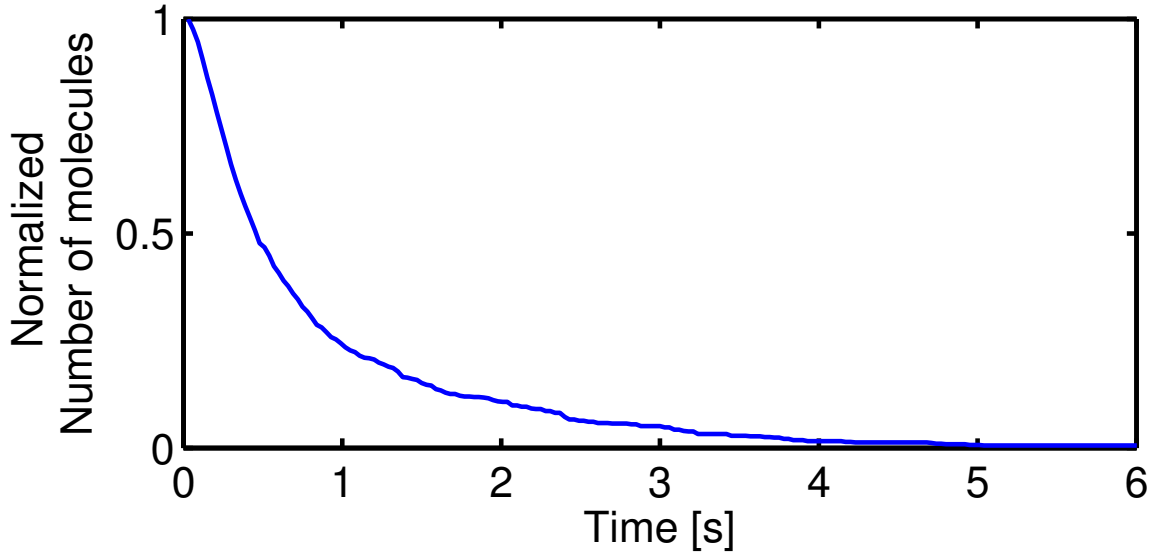


Figure 4.23: *Cumulative dwell time histogram for a cluster of transitions selected in the transition density plot. The number of molecules remaining in the state before the transition are plotted against time.*

5 The acidic compact state of apomyoglobin from yellowfin tuna

The results from this chapter have been published previously in (Bismuto *et al.*, 2009).

5.1 Background

A protein has to fold into its correct three-dimensional structure to be functional. A complete, detailed description of this process is still lacking due to experimental difficulties. The processes involved in protein folding are highly heterogeneous and often completed in very fast timescales down to µs. These difficulties complicate the observation of folding pathways and possible protein folding intermediates. Non-native, partially unfolded or denatured protein conformations may have similarities to protein folding intermediates and have thus been studied to get a better understanding of the protein folding process (Baldwin, 1994; Fink, 1995; Dinner *et al.*, 2000; Mayor *et al.*, 2000; Plotkin and Onuchic, 2002a,b; Muñoz, 2007). By now, it has become clear that the study of these systems can also yield important insights into a range of other topics, including protein translocation, misfolding, and aggregation (Chiti and Dobson, 2006; Morozova-Roche and Malisauskas, 2007). Partially unfolded states with exposed hydrophobic residues can be precursors for amyloid fibril formation (Uversky and Fink, 2004) which plays an important role in the pathogenesis of neurodegenerative diseases like Alzheimer's or Parkinson's disease.

Some proteins show a tendency for aggregation or fibril formation without relation to any disease. One of these proteins is myoglobin (Fändrich *et al.*, 2001), an iron- and oxygen-binding protein present in the muscle tissue of mammals. It is closely related to hemoglobin, the iron- and oxygen-binding protein in blood. It is a single-chain globular protein of 154 amino acids and a molecular weight of 17.8 kDa, containing a prosthetic heme group as catalytic center around which the remaining apoprotein folds into seven or eight[1] alpha helices (Kendrew *et al.*, 1958). Myoglobin is a well known model system for protein dynamics (Frauenfelder *et al.*, 1991; Parak and Nienhaus, 2002) and forms fibrils containing β-strands under experimental conditions that are known to favor the formation of partially folded states (Sirangelo *et al.*, 2004).

Under certain pH and ionic strength conditions, myoglobin, similar to numerous other proteins, preferentially assumes a partially folded state called the acidic compact state or A-state. The A-state has several similarities to the molten globule state, a common protein folding intermediate (Section 3). First, its hydrodynamic radius is higher than in the native state but lower than in the unfolded state, with the hydrodynamic radius in the A-state being closer to the native state than the unfolded state. Second, it has a similar amount of secondary structure compared to the native state, but is lacking tertiary structure. Third, as a consequence of the lacking tertiary structure, it forms a hydrophobic core but has remaining hydrophobic patches on the surface that can lead to aggregation or fibril formation (Fink, 1995; Irace *et al.*, 1986; Barrick and Baldwin, 1993).

[1] Myoglobin has eight alpha helices in most species. In yellowfin tuna, however, helix D is unfolded in the native state.

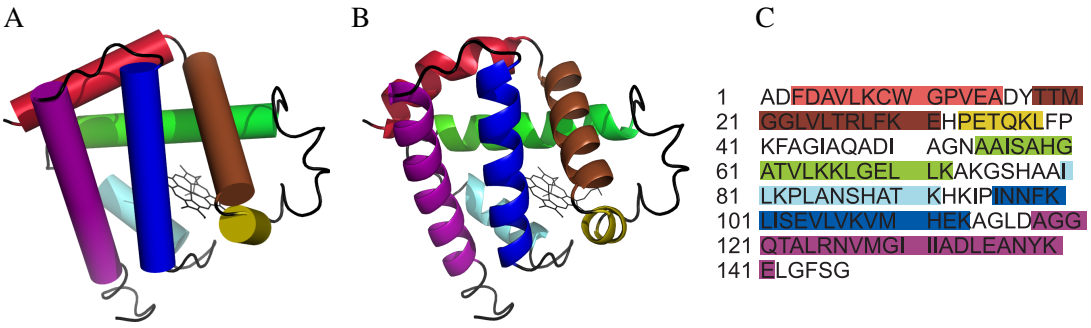


Figure 5.1: *Structure of myoglobin in its native state at pH 7 (PDB model 1MYT), represented in PyMol in* cylinder style *(A) and* cartoon style *(B). The α-helix segments indicated in the amino acid sequence in (C) are depicted in different colors (Helix A: red, B: brown, C: yellow, E: green, F: light blue, G: dark blue, H: purple).*

Compact protein folding intermediates differ depending on the protein and on the experimental conditions (Goto and Fink, 1990). They may consist of either a native state-like topology with regions of secondary structure similar to the native state connected by disordered regions of the polypeptide chain, or of a core of native-like structure surrounded by totally or partially unfolded parts of the polypeptide chain as in acidic apomyoglobin, which has a native-like core formed by a subset of the A, G, and H-helices (Hughson *et al.*, 1990; Jennings and Wright, 1993; Eliezer *et al.*, 1998).

In (Bismuto *et al.*, 2009), the acidic compact state of apomyoglobin from yellowfin tuna was characterized in molecular detail as a model of the molten globule state of the protein. The best studied myoglobin is from sperm whale, however, myoglobin from yellowfin tuna offers several advantages for MD simulations and fluorescence spectroscopy studies. It has a lower unfolding free energy reducing the computational effort for MD unfolding studies (Bismuto *et al.*, 1984), and its single Tryptophan simplifies the interpretation of the intrinsic fluorescence emission and decay rate (Alcala *et al.*, 1987; Bismuto *et al.*, 1988; Bismuto and Irace, 1994). Our collaborators generated an ensemble of non-native structures of apomyoglobin by molecular dynamics (MD) simulations at higher temperature (500 K) and acidic conditions (pH 3.0) in the presence of 0.15 M NaCl. The high temperature MD simulation was followed by short re-equilibrium steps at intermediate temperatures until 300 K was reached. The ensemble of final states from 25 different MD simulations was considered to represent the distribution of conformers constituting the A-state (Figure 5.2).

Frequency domain fluorometry on the single Tryptophan (Gratton and Limkeman, 1983), acrylamide fluorescence quenching (Eftink and Ghiron, 1981; Zelent *et al.*, 1998) experiments, and far UV circular dichroism were performed to test the results of the MD simulations. The results of these measurements compared well with values obtained by the MD simulations for the acidic apomyoglobin structures. Two-focus fluorescence correlation spectroscopy (2fFCS) experiments (Dertinger *et al.*, 2007) provide an additional independent measurement of the hydrodynamic radius of the protein in native, acidic and unfolding environments showing that the protein is indeed more expanded in the A-state than in the native state. FCS is advantageous for experiments of aggregation-prone protein states due to the low (nanomolar) concentrations suitable for these experiments.

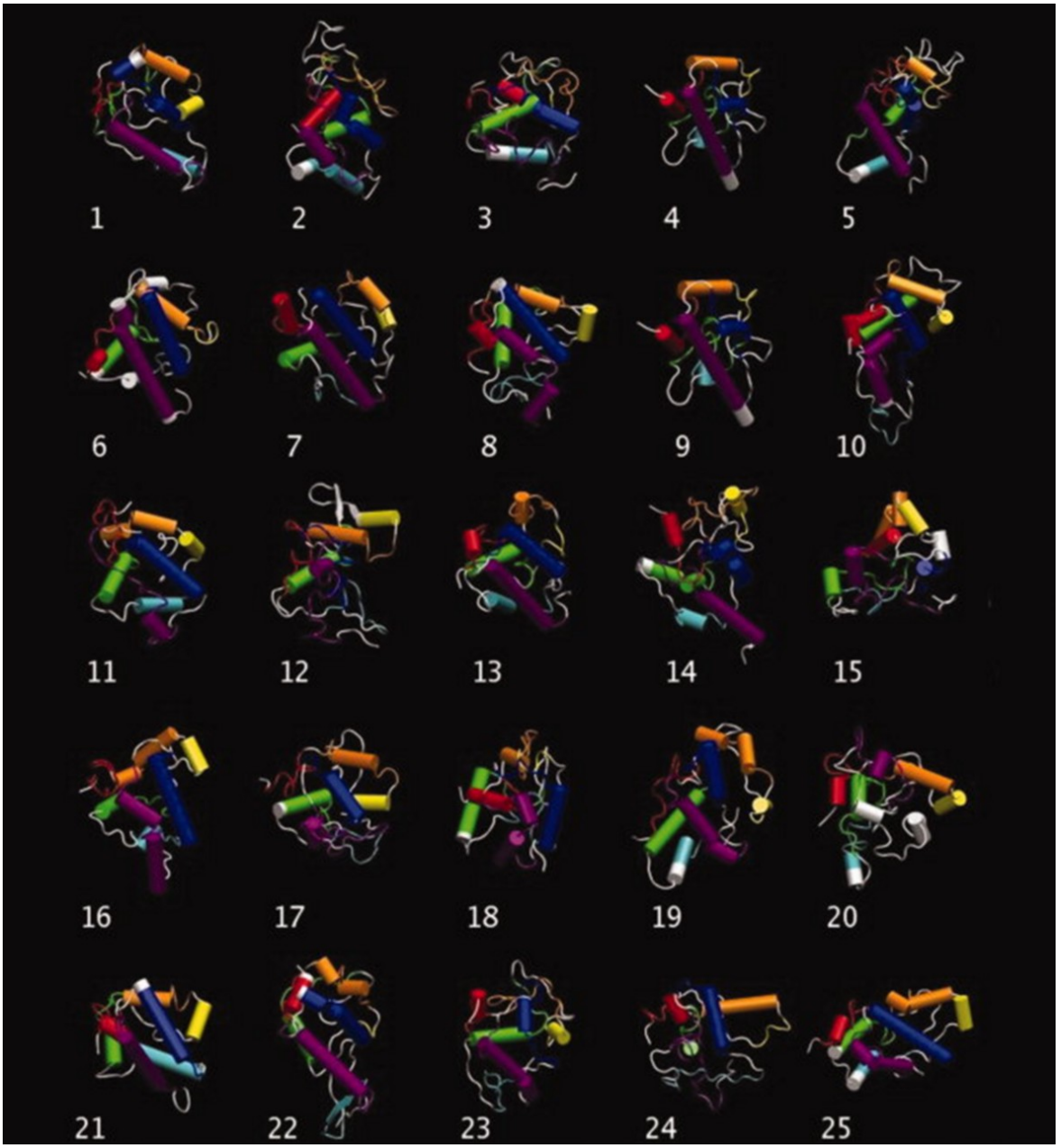


Figure 5.2: *Output structures of the 25 apomyoglobin molecular dynamics simulations at pH 3, represented in* cylinder style*. The former α-helix segments of the native structure, as indicated in the right part of Figure 5.1, are depicted in the same color coding as before (Helix A: red, B: brown, C: yellow, E: green, F: light blue, G: dark blue, H: purple). Figure reproduced from (Bismuto et al., 2009).*

5.2 Experimental procedures

5.2.1 Myoglobin purification and apomyoglobin extraction

The main component of tuna myoglobin was prepared by Ettore Bismuto and Emiddio di Maggio according to the method described by Bismuto *et al.* (1988). The homogeneity of the preparations was controlled by sodium dodecyl sulfate/polyacrylamide gel electrophoresis with 15% acrylamide gels and 5% acrylamide stacking gels. The heme was removed from myoglobin by the 2-butanone extraction procedure (Teale, 1959). Contamination of the apoprotein by heme-containing myoglobin was assessed spectrophotometrically. In all cases, no significant absorption was observed in the Soret region. The myoglobin concentration was determined spectrophotometrically at 409 nm in 10 mM phosphate, 150 mM NaCl buffer (pH 7.0) by using 139,000 $M^{-1}cm^{-1}$ as the extinction coefficient at 407 nm (Bismuto *et al.*, 1985). For apomyoglobin, the molar extinction at 280 nm was calculated from the Tryptophan and Tyrosine content by using molar extinction coefficients of 5500 $M^{-1}cm^{-1}$ and 1250 $M^{-1}cm^{-1}$, respectively (Wetlaufer, 1962). All common chemicals were reagent grade and were purchased from Sigma-Aldrich.

5.2.2 Protein labeling

Purified yellowfin tuna myoglobin was labeled stochastically on the amine groups of the numerous lysines or the N-terminus of the protein according to the coupling procedure recommended by Atto-tec. Different dyes were tried for the coupling reaction. Figure 5.3 shows one-focus FCS curves of Atto 532-NHS , myoglobin labeled with Atto 532-NHS and myoglobin labeled with Alexa Fluor 546-NHS. From the diffusion time of the samples, it is clear that the labeling reaction with Atto 532 did not work as expected. Even after removal of free dye using Biorad Micro Bio-Spin 6 spin columns, after dilution to nanomolar concentrations the diffusion time has changed only by 42% compared to the free dye Atto 532, although, from the molecular weight difference between the free dye (1 kDa) and myoglobin (16.7 kDa), an increase of a factor of 2.56 is expected. This suggests that the dye was not coupled covalently to the protein but was bound by unspecific, weaker interactions during the purification step and released after the dilution. For myoglobin labeled with Alexa Fluor 546-NHS, the diffusion time changed by a factor of 2.34, showing that in this case the labeling reaction did work as expected. The remaining experiments were thus conducted with myoglobin labeled with Alexa Fluor 546.

For the labeling reaction, myoglobin was incubated with equimolar (150 µM) amounts of Alexa Fluor 546 succinimidyl ester (Invitrogen) in 25 mM HEPES, 75 mM NaCl (pH 8.2) for 2 h at room temperature. The labeled protein was separated from the remaining free dye using Biorad Micro Bio-Spin 6 spin columns at 1000 rcf for 4 min. This purification step was repeated three times. The protein was stored in 10 mM sodium phosphate buffer (with 150 mM NaCl) at pH 8 at micromolar concentration and diluted directly before the measurements with buffer at the respective pH values (3.19, 3.86, 4.42, 5.08, 5.50, 5.98, 6.35, and 7.02), or buffer containing 4 M GuHCL, 10 mM sodium phosphate and 150 mM NaCl (pH 5.0), to concentrations appropriate for the FCS measurements ($\sim$ 2 nM).

5.2.3 Two-focus FCS setup

Two-focus FCS experiments (Dertinger *et al.*, 2007) were performed in part by Stefan Pleus in Ulrich Nienhaus' lab in Ulm on a custom designed confocal microscope (Schenk *et al.*, 2004; Wiedenmann

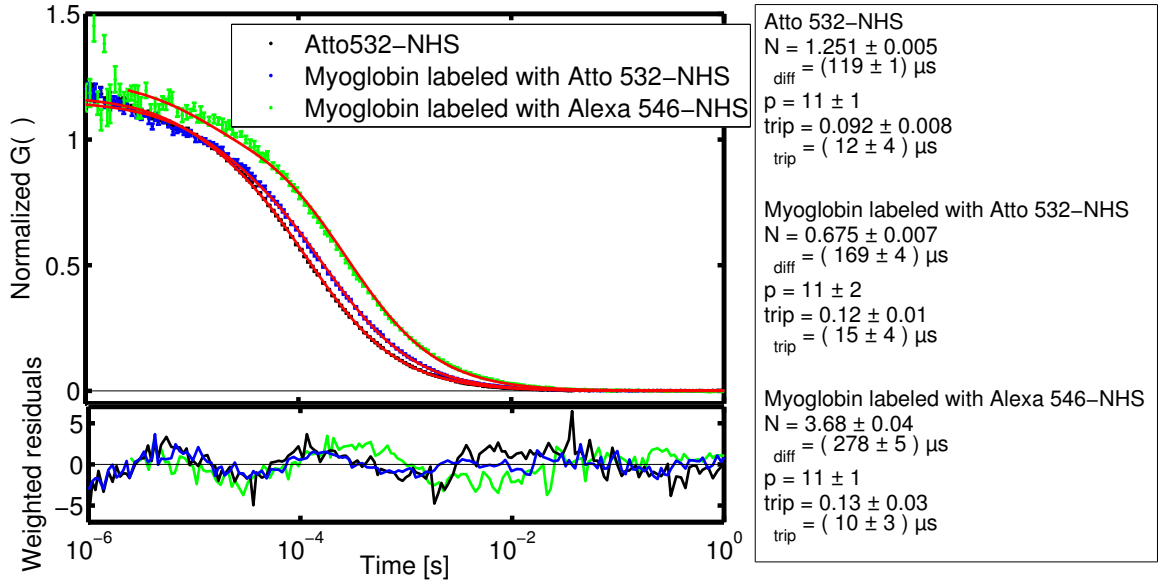


Figure 5.3: *Dye comparison for Myoglobin labeling. FCS curves for Atto 532-NHS (black), myoglobin labeled with Atto 532-NHS (blue) and myoglobin labeled with Alexa Fluor 546-NHS (green) with corresponding fits (red) and weighted residuals (lower panel). The fit values are shown to the right.*

et al., 2004) built around an inverted epi-fluorescence microscope (Axiovert 35, Carl Zeiss, Göttingen, Germany). For pulsed interleaved excitation (PIE) of two separated but overlapping volumes, excitation light pulses of alternating perpendicular polarizations were generated and passed through a Nomarski prism (DIC 1033-911, Carl Zeiss) before entering the microscope objective (UPLAPO 60x/1.2w, Olympus, Hamburg, Germany). A train of pulses at 10 MHz was generated from a single mode-locked frequency-doubled solid-state laser of 532 nm (GE-100, Time-Bandwidth Products, Zürich, Switzerland) and a pulse picker (Pulse select, APE, Berlin, Germany). Alternating polarizations were achieved using a half-wave plate (AHWP05M-630, Thorlabs, Dachau, Germany) followed by separation and recombination of the light pulses with a pair of polarizing beam splitters (48571, Edmund Optics, Barrington, NJ). The S-polarized pulses were delayed by 30 ns with respect to the P-polarization with an optical delay line. The recombined pulse train was fed into a polarization maintaining single mode fiber (QPMJ, OZ Optics, Ottawa, Canada) and reflected by a dichroic beam splitter (z532/633 xr, AHF, Tübingen, Germany) before entering the Nomarski prism. The emitted fluorescence photons were focused with an achromatic lens ($f = 150\,\mathrm{mm}$) onto a confocal pinhole ($d = 100\,\mu\mathrm{m}$), passed through an emission filter (HQ 582/50, AHF) and detected by an avalanche photodiode (SPCM-AQR-14, Perkin-Elmer, Fremont, CA). As in PIE (Müller *et al.*, 2005), the photon arrival time with respect to the laser trigger was measured with time-correlated single-photon counting (Timeharp 200, Pico-Quant, Berlin, Germany) and used to attribute each recorded photon to the corresponding excitation pulse and therefore to the respective detection volume. Data analysis was carried out as described in more detail in Section 4.2.2. Prior to the determination of diffusion coefficients, the separation of the foci was determined as 237 nm, using the diffusion coefficient of rhodamine 6G in water ($D = 285\,\mu\mathrm{m}^2/\mathrm{s}$) given by the average of two independent investigations (Magde *et al.*, 1974; Gell *et al.*, 2001). The two-focus measurements were performed with an excitation power of 35 µW per focus in a temperature-controlled laboratory at 22°C.

5.3 Results

2fFCS was applied to compare the hydrodynamic radius of apomyoglobin in the native and acidic compact states. A thorough description of 2fFCS is given in the methods section 4.2.3. In short, FCS measures generally the fluctuations in the fluorescence signal coming from the small ($\sim 1\,\mathrm{fl}$) focus of a confocal microscope (Magde *et al.*, 1972; Elson and Magde, 1974; Magde *et al.*, 1974). The main source of fluctuations is typically the translational diffusion of the sample through the confocal volume. The decay time of the correlation function then yields information about the diffusion coefficient D and hence the hydrodynamic radius R_H of the sample. These two quantities are related by the Stokes-Einstein relation:

$$D = \frac{k_B T}{6\pi\eta R_H} \tag{5.1}$$

where k_B is the Boltzmann constant, T is the temperature, and η is the viscosity of the buffer. FCS measurements on a typical confocal microscope may be affected by experimental artifacts. FCS measures the diffusion time of the sample, $\tau_D = \frac{\omega_r^2}{4D}$, and D can be calculated only after careful calibration of the volume with a dye with known diffusion coefficient. However, even if the volume is calibrated for a given set of experimental conditions, the focus shape may change during a set of experiments due to refractive index changes of the buffer or changes in coverslip thickness (Enderlein *et al.*, 2004).

These shortcomings can be circumvented by the introduction of an additional, fixed length scale in the experimental setup. Dertinger *et al.* (2007) have introduced the method of two-focus FCS (2fFCS, Section 4.2.2), where separate foci are created and a global fit of the autocorrelation functions and the cross-correlation function of the signals from these foci is used to get absolute numbers for the diffusion coefficient. The probe volume was given using the expressions in equations 4.39 - 4.41 in Chapter 4.2.2.

A representative 2fFCS measurement of freely diffusing tuna apomyoglobin labeled with Alexa Fluor 546 in 10 mM phosphate, 150 mM NaCl buffer (pH 3.2) and the corresponding global fit are shown in the upper panel of Figure 5.4. The global fit yielded a diffusion coefficient of $D = 71 \pm 5\,\mu\mathrm{m}^2 s$. With changing pH, we observed a corresponding change of the diffusion coefficient of myoglobin (Lower panel of Figure 5.4). Control experiments showed no pH dependence of the diffusion coefficient of Alexa Fluor 546 ruling out viscosity changes of the buffer or optical effects as cause for the pH dependent diffusion coefficient of myoglobin. The unchanged value measured for the diffusion coefficient of Alexa Fluor 546 is also a good indication for the stability of the experimental setup over the course of the experiments.

The correlation curves for Alexa Fluor 546 could be fit using only a single diffusion component (Equation 4.33). However, the ACF and CCF from yellowfin tuna myoglobin at intermediate pH values contained an additional component on the time scale of 15-25 µs. To achieve a reasonable fit to the data, it was necessary to include an exponential relaxation term in the fit function. This additional exponential term may be explained by different models. Increased population of the fluorophore triplet state may be able to explain the observed triplet term. However, it is not obvious why the triplet state of the dye should be considerably more populated at intermediate pH or due to the proximity of the protein.

A more reasonable assumption is that interactions of the dye with parts of the protein may influence the emission properties of the dye. Possible candidates as quenchers are aromatic residues throughout the protein. Lamb *et al.* (2000) have shown using time-gated FCS that even in the absence of the Heme-group, stochastically labeled myoglobin exhibits a significant amount of (on the

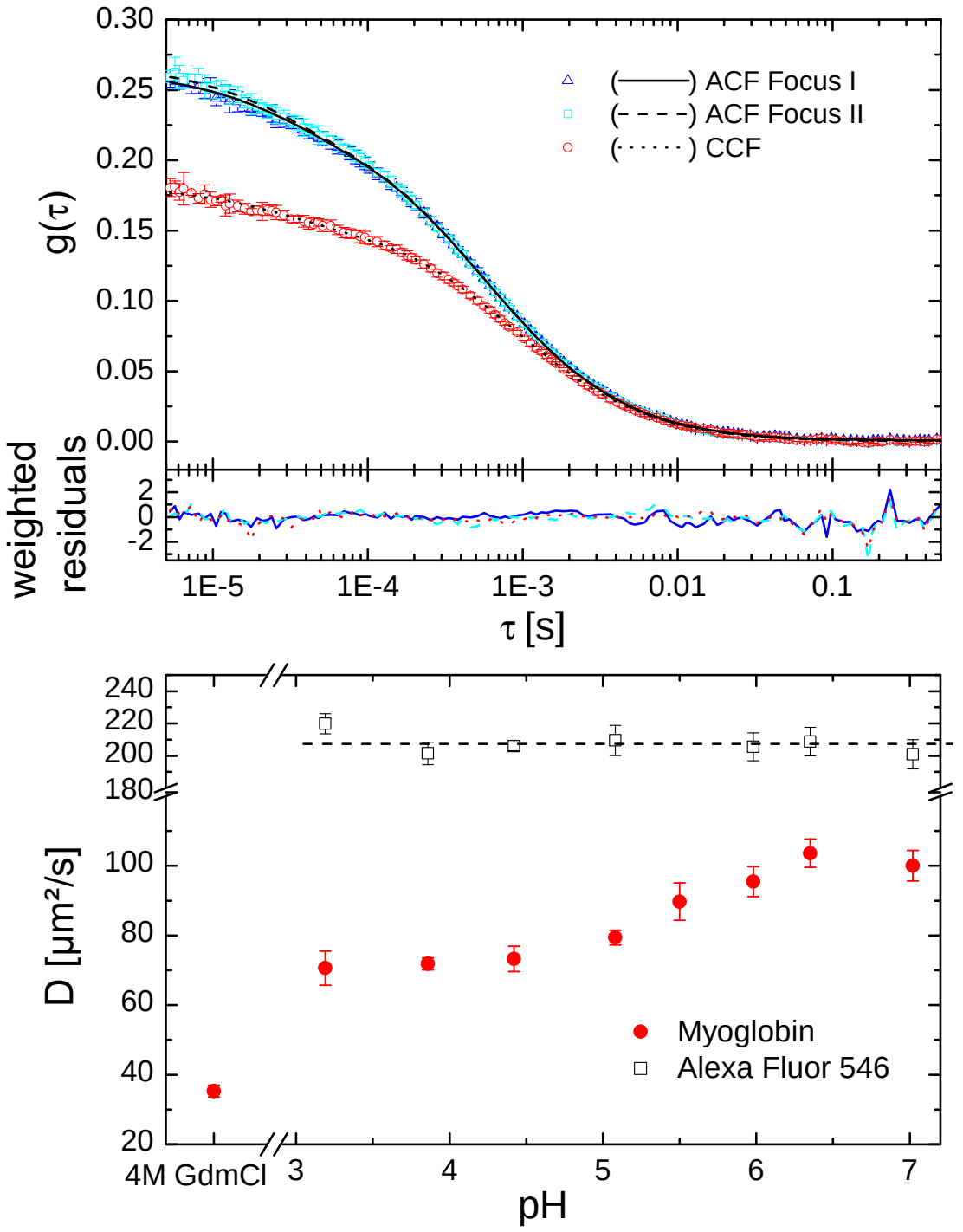


Figure 5.4: *Two-focus FCS experiments of Tuna myoglobin. Upper panel: The ACFs and CCF of the two foci for a 2fFCS experiment on fluorescently labeled Tuna myoglobin at pH 3.19 and 150 mM NaCl. The error bars represent the standard deviation of multiple measurements. The fit and reduced residuals are shown as lines in the appropriate graphs. Lower panel: The diffusion coefficient of Alexa Fluor 546 and Tuna myoglobin as a function of pH determined using 2fFCS. To the left, a measurement of the diffusion coefficient of fully denatured myoglobin in 4M GuHCL at pH 5.0 is shown for comparison. The error bars represent the standard deviation of at least 4 measurements. Figure adapted from Bismuto et al. (2009).*

>10 ms timescale) static heterogeneity. Chen *et al.* (2007) have observed fluctuations on a timescale of 3-200 µs in myoglobin that were attributed to quenching by the aromatic residues. However, the fluctuations we have observed were maximum for intermediate pH, while the fluctuations observed by Chen *et al.* (2007) were maximum for low pH.

The determination of the diffusion coefficients was not disturbed by the additional relaxation term. The diffusion coefficient of yellowfin tuna myoglobin changes by 30% from $100 \pm 4\,\mu\mathrm{m}^2/\mathrm{s}$ at pH 7.0 to $71 \pm 5\,\mu\mathrm{m}^2/\mathrm{s}$ at pH 3.2. The hydrodynamic radius can be calculated from Eq. 5.1 treating myoglobin as a hard sphere assuming that the viscosity of the buffer is the same as the viscosity of water at the same temperature. Then, the measured diffusion coefficient corresponds to a hydrodynamic radius of $2.3 \pm 0.1\,\mathrm{nm}$ at pH 7.0 and $3.3 \pm 0.2\,\mathrm{nm}$ at pH 3.2. The hydrodynamic radius is not directly accessible from the MD simulations. However, the radius of gyration can be calculated since it is simply the mean squared distance of the atoms from the center of mass. The radius of gyration can be related to the hydrodynamic radius assuming that the protein behaves like a hard sphere:

$$R_G = \sqrt{\frac{3}{5}}\, R_H = 0.775\, R_H \tag{5.2}$$

The experimentally determined radius of gyration thus changes from $1.8 \pm 0.1\,\mathrm{nm}$ to $2.5 \pm 0.2\,\mathrm{nm}$ between pH 7.0 and 3.2. These values are based on the assumption of a diffusion coefficient of Rhodamine 6G of $D = 285\,\mu\mathrm{m}^2 s$. Shortly after publishing of the results presented in this chapter, Enderlein and coworkers (Müller *et al.*, 2008) measured a 45% higher value for the diffusion coefficient of Rhodamine 6G, $D = 414\,\mu\mathrm{m}^2 s$, using 2fFCS calibrated by z-scanning (Section 4.2.2). This affects the absolute values of the hydrodynamic radii determined in this study, increasing the measured diffusion coefficients by 45% and hence decreasing the hydrodynamic radii by the same factor to $1.2 \pm 0.1\,\mathrm{nm}$ and $1.7 \pm 0.2\,\mathrm{nm}$ at pH 7.0 and 3.2, respectively.

The absolute value of the radius of gyration from the MD simulations is 1.4 nm and 1.6 nm at pH 7 and 3, respectively. Thus, the experimental values correspond well with the values predicted by the MD simulation considering the possible inaccuracies both in the experimental and simulation data.

In a next step, we compared the hydrodynamic radii of myoglobin in the native and in the acidic compact state with the hydrodynamic radius of unfolded myoglobin. Myoglobin was denatured in 4 M guadinium hydrochloride (GuHCl) and the diffusion coefficient was determined again by 2fFCS. To calculate the hydrodynamic radius, the viscosity change of the GuHCl solution compared to the buffer without GuHCl was estimated from the diffusion coefficient change of free dye Atto 532 to be $\eta_{GuHCl}/\eta_{Buffer} = 1.57$.[2] For unfolded myoglobin, we measured a hydrodynamic radius of $4.2 \pm 0.3\,\mathrm{nm}$. The size of the acidic compact state is thus between the native state and the unfolded state in agreement with the MD simulations.

5.4 Discussion

Protein folding intermediates are usually more compact than the unfolded state but less compact than the native state. They often have hydrophobic surface patches which can result in an increased tendency to aggregate. Detailed understanding of the effects that lead to the aggregation of these partially

[2] This is more than expected for a 4 M GuHCl solution (1.27x, Kawahara and Tanford, 1966). This deviation may be caused by a change in focus size due to the higher refractive index of the GuHCl solution. Using the dye as a reference allows taking both viscosity change and refractive index change into account (Enderlein *et al.*, 2005).

unfolded proteins may play a fundamental role in the development of therapies for neurodegenerative diseases in the future. These intermediates are usually populated only for very short timescales (<1 µs) during folding which can make observation difficult. However, stable folding intermediates have often been observed at acidic conditions and moderate ionic strength (*A-states*, Fink, 1995). One model system for the investigation of the A-state is the acidic compact state of apomyoglobin from sperm whale. An increasing number of spectroscopic techniques, particularly NMR (Barrick and Baldwin, 1993; Cocco and Lecomte, 1990), has been employed to understand the nature of the acidic compact state. However, the partially unfolded ensemble of proteins is actually comprised of a large number of structurally different states, and ensemble spectroscopic techniques like NMR will provide measurements only for the average properties of the system.

MD simulations have been employed to characterize the states comprising the acidic compact state of yellowfin tuna myoglobin in molecular detail (Bismuto *et al.*, 2009). To validate the results of the MD simulations, several spectroscopic methods were employed. The presence of a single Tryptophan in yellowfin tuna apomyoglobin allows an easier interpretation of its intrinsic fluorescence. The lifetime distribution of the Tryptophan residue measured by Frequency Domain Fluorometry broadens for lower pH suggesting more heterogeneity of myoglobin in the A-state compared to the native state. The acrylamid quenching data is in good agreement with the collisional quenching constant of the intrinsic fluorescence emission of the 25 MD-generated acidic conformers of myoglobin. Also, the α-helical content measured by circular dichroism is in very good agreement with the MD simulations. The diffusion coefficients obtained by 2fFCS measurements are compatible with the calculated ones taking the protein hydration shell into account. While the agreement at pH 7 is very good, the measured values at pH 3 are somehow larger than expected from the MD simulations. However, taking the uncertainties of both the MD simulations (approximation of the force field parameters, cut off, periodic conditions, and a limited number of generated structures) and the 2fFCS measurements (reference diffusion coefficient of Rhodamine 6G, possible aggregation at acidic pH) into account, these differences are within the experimental errors.

Chen *et al.* (2007) have studied the fast fluctuations of myoglobin in various pH caused by quenching of an N-terminal dye by aromatic residues using FCS. As a byproduct of these measurements, they also report on changes of the diffusion coefficient with varying pH. They also observed a pH-dependent 40% change of D, however, in their measurements, D is essentially pH independent between pH 4.5 and 7 and the transition midpoint is lower than pH 3.5. The difference to the pH dependence in the 2fFCS experiments presented here may be explained by the use of sperm whale myoglobin by Chen *et al.* (2007) or the ionic strength differences between the two studies.

In conclusion, the good agreement of the experimental results, in particular the 2fFCS results, show the validity of the structures derived from the MD simulations. Combined, the simulation and experimental results support the point of view that folding intermediates cannot be described by a single conformational state but rather by a number of structurally different conformations with similar energetic characteristics. A subset of these states shows exposed hydrophobic surface patches that possibly can trigger the aggregation of the protein.

6 Single-molecule investigation of the conformational cycle of mitochondrial Hsp70

Some results of this chapter have been published previously in (Mapa *et al.*, 2010).

6.1 Introduction

6.1.1 The Hsp70 chaperones

Heat shock proteins 70 (Hsp70s) are a highly conserved and ubiquitous class of molecular chaperones present in prokaryotic organisms and in all major organelles of eukaryotes. They are involved in folding of newly synthesized proteins, prevention of protein aggregation, remodeling of protein complexes and transport of proteins across membranes. These functions are performed by ATP-regulated binding and release of substrate proteins.

All Hsp70s share a similar structure (Figure 6.1). They consist of an N-terminal 44 kDa nucleotide binding domain (NBD) connected by a flexible linker to the C-terminal 25 kDa substrate binding domain (SBD). The SBD is divided into a β-sandwich sub-domain with a hydrophobic cleft for substrate binding and an α-helical lid. The SBD binds hydrophobic regions of substrate peptides. The affinity of the SBD for substrate is regulated by the binding of different nucleotides to the NBD. In the ATP-bound state, the lid is open and the affinity for substrate is low. ATP hydrolysis and substrate binding is stimulated by J proteins. Then, release of ADP and subsequent ATP binding is modulated by nucleotide exchange factors (NEF), allowing substrate release and the beginning of a new cycle (Bukau *et al.*, 2006; Frydman, 2001; Genevaux *et al.*, 2007; Mayer and Bukau, 2005; Saibil, 2008; Young *et al.*, 2004).

The structures of the substrate-bound state and the ATP-bound state of different Hsp70s have been recently solved using NMR and X-ray crystallography (for an overview, refer to Bertelsen *et al.*, 2009, and references therein). However, these structures are only static pictures of specific states of the Hsp70 cycle and do not give insight into the processes involved in the functional cycle. The dynamic processes connecting these states can be resolved by single-molecule fluorescence studies using spFRET. However, only few studies have used single-molecule fluorescence to get more insight on the function of Hsp70s. Besides the work on mtHsp70 presented in this thesis and in (Mapa *et al.*, 2010), another paper (Marcinowski *et al.*, 2011) studied the only Hsp70 in the endoplasmatic reticulum, the Immunoglobulin Binding Protein (BiP). Marcinowski *et al.* (2011) used a spFRET assay to study the conformation of BiP in different stages of its conformational cycle. The labeling positions were designed in a similar way as in this work and one additional sensor monitored the distance between the tip of the substrate binding lid and the NBD. In the apo (nucleotide-free) state and in the ADP state, BiP has a high conformational flexibility. ATP binding results in domain

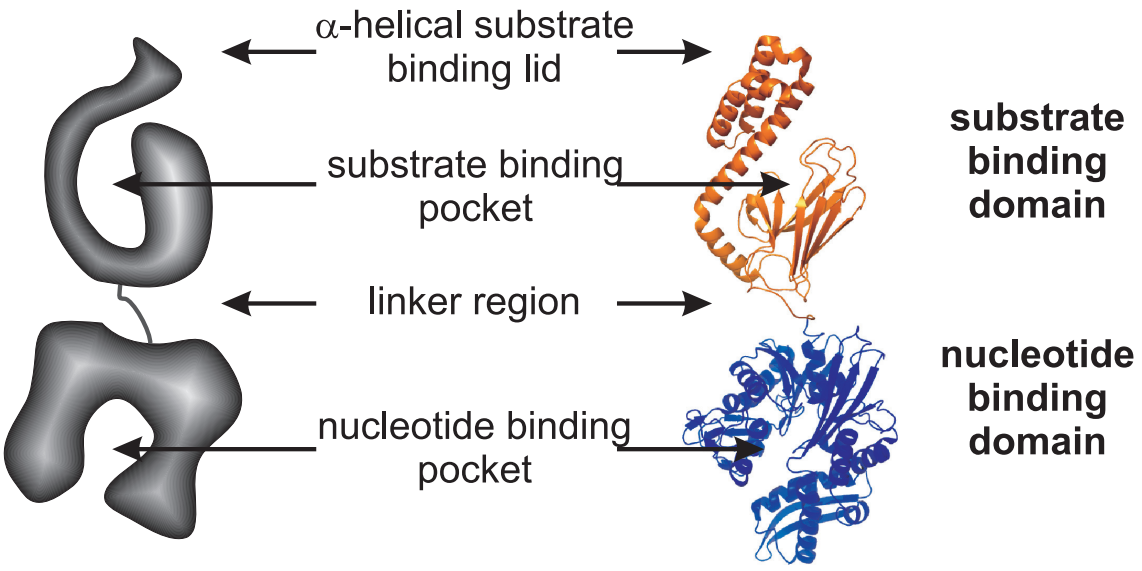


Figure 6.1: *The structure of Hsp70 chaperones. The two domains are denoted in a sketch of a general Hsp70 (left) and the structure of DnaK (PDB: 2KHO).*

docking and opening of the substrate binding lid. Substrate binding does not result in further domain undocking and lid closing compared to the ADP state. The conformation of BiP changes significantly depending on the substrate bound. It is similar to the ATP state for BiP bound to the complete first domain of the IgG heavy chain C_H1 and similar to the ADP state for a small peptide derived from C_H1 suggesting substrate discrimination by the lid communicated to the NBD.

6.1.2 Mitochondrial Hsp70

In the mitochondrial matrix of yeast, three different species of Hsp70s with significant sequence homology to the bacterial Hsp70 DnaK are present. Ssq1 is involved in the biogenesis of Fe-S clusters, and Ecm10 is of low abundance and has been suggested to have a similar function as Ssc1 (Baumann *et al.*, 2000; Craig and Marszalek, 2002; Voos and Röttgers, 2002). The major mitochondrial Hsp70, known as Ssc1, mediates folding of proteins in the mitochondrial matrix and is involved in the import of preproteins into the mitochondrial matrix. Both the protein folding and the preprotein translocation activities of Ssc1 rely on its affinity for unfolded proteins and its ATPase activity. Ssc1 makes up 1-3 % of mitochondrial protein and is essential for cell viability (Craig *et al.*, 1987; Craig and Marszalek, 2002; Yoneda *et al.*, 2004). A specialized chaperone, Hep1, is required to maintain the structure and function of Ssc1 (Sichting *et al.*, 2005).

6.1.2.1 Role of Ssc1 in protein import into mitochondria

Only few of the proteins present in mitochondria are actually produced there. The vast majority is produced by cellular ribosomes and has to be imported into mitochondria and targeted to the correct sub-compartment. Several translocation machineries exist in mitochondria to correctly deliver the preprotein to its destination (Neupert and Brunner, 2002; Mokranjac and Neupert, 2009). Preproteins are imported through the outer mitochondrial membrane to the inter-membrane space by the TOM complex. Then, the translocase of the inner mitochondrial membrane 23 (TIMM23) complex

transports preproteins from the inter-membrane space between the inner and outer mitochondrial membranes to the mitochondrial matrix by an energy-driven process. The energy comes from the membrane potential across the inner mitochondrial membrane and from ATP hydrolysis.

The translocation channel is formed by the channel-forming membrane proteins Tim23 and Tim17. Targeting of the preprotein to the outer part of the channel is mediated by the receptor protein Tim50. The protein Tim44 resides associated to the membrane close to the translocation channel and is responsible for the recruitment of the soluble components of the import motor. The actual import motor of the translocation complex is the mitochondrial Hsp70 Ssc1 present in the mitochondrial matrix in conjunction with the J protein Tim14, the J-like protein Tim16 (Bolender *et al.*, 2008; Craig *et al.*, 2006; Mokranjac *et al.*, 2006), and the nucleotide exchange factor Mge1 (Neupert and Brunner, 2002).

The positively charged targeting sequence of precursor proteins is transported through the translocation pore by the potential across the inner mitochondrial membrane (Neupert and Brunner, 2002). Then, Ssc1 can bind to hydrophobic parts of the preprotein chain and an inward motion of the protein is generated. Generally, motor proteins can promote this motion according to two different principles. First, brownian motion may be the driving force and the motor protein prevents backsliding passively by binding to the preprotein (*brownian ratchet model*). Second, the force responsible for the inward motion can be generated by an active conformational change of the motor protein, with ATP hydrolysis as the source of the energy (*power stroke model*). For Ssc1, a third model has been proposed combining both ideas into an *entropic pulling model*. This model suggests that the limited amount of freedom of motion for Ssc1 close to the membrane induces an entropic force in the inward direction (De Los Rios *et al.*, 2006; Goloubinoff and De Los Rios, 2007).

6.1.2.2 Role of Ssc1 in protein folding in the mitochondrial matrix

Like the majority of Hsp70 chaperones, Ssc1 supports folding of proteins and prevents their aggregation, in addition to its function as an import motor (Horst *et al.*, 1997; Kang *et al.*, 1990; Rowley *et al.*, 1994; Westermann *et al.*, 1996). In this function, it is supported by the J protein Mdj1 nd the nucleotide exchange factor Mge1. The chaperone activity of Ssc1 has been demonstrated by in vitro experiments, where a reconstituted system including Ssc1, Mdj1 and Mge1 was able to prevent the aggregation and increase the refolding efficiency of heat-denatured luciferase (Westermann *et al.*, 1996; Kubo *et al.*, 1999).

Although the protein folding cycle of Ssc1 is closely related to the cycle of the bacterial Hsp70 DnaK and Ssc1 and DnaK share over 50% sequence identity, yeast Ssc1 and DnaK from *E. coli* cannot substitute each other, neither in bacteria nor in yeast mitochondria (Deloche *et al.*, 1997; Moro *et al.*, 2002). In contrast, the respective nucleotide exchange factors, Mge1 for Ssc1 and GrpE for DnaK, and J proteins, Mdj1 for Ssc1 and DnaJ for DnaK, can substitute for each other in both organisms, at least to a certain extent (Deloche *et al.*, 1997; Lisse and Schwarz, 2000). Similarily, specifity of Hsp70s from different subcellular compartments of the same organism was shown which is surprising due to the high level of sequence conservation (Brodsky, 1996; Genevaux *et al.*, 2007; Hennessy *et al.*, 2005; Walsh *et al.*, 2004).

6.1.3 Scope of this project

Due to its wide range of different functions, Ssc1 is an essential protein for cell viability. Although many studies have elucidated the function of Ssc1 and the relation to other Hsp70s, a detailed mech-

anistic understanding of the mechanisms of Ssc1-mediated protein folding and protein import is still lacking. The inherent instability of Ssc1 prohibits the study of its conformation using crystallization methods. To this end, FRET studies can give new insights into the different conformations of Ssc1 along its functional cycle. In particular spFRET experiments may give additional insights into possible substates and inhomogeneities as well as dynamic changes.

To this end, we developed two FRET-based sensors and used them to follow the effects of nucleotides and interacting proteins on the relative distance between the two domains of Ssc1 and the opening of the SBD. The spFRET data collected by MFD-PIE experiments was analyzed using PDA to extract information about the number, mean and width of the distance distributions underlying the spFRET histograms. The dynamics of the system were followed by spFRET measurements of immobilized molecules using TIRF microscopy.

Cochaperones modulate the function of Ssc1 depending on the specific task at hand. The FRET sensors are used to study the effects of different cochaperones from the folding cycle and from the import cycle on the conformation of Ssc1. Additionally, we use similar FRET-based sensors for DnaK, the major bacterial Hsp70, to study the differences between Hsp70s with different functions.

6.2 Experimental Procedures

6.2.1 Protein expression, purification, and labeling

Protein expression, purification and labeling were performed by our collaborators Koyeli Mapa and Dejana Mokranjac. Only a short overview will be given here. For a more detailed description, refer to (Mapa *et al.*, 2010) and (Mapa, 2009).

The structure of Ssc1 or its domains has not been solved yet. However, Ssc1 shares sequence similarities with other Hsp70s. Residues of Ssc1 suitable for positioning of the fluorophores were identified by modeling Ssc1 on the structure of the SBD of DnaK with bound substrate (PDB accession number 1DKX) and of the ATP state of Sse1 (PBD accession number 2QXL). The selected residues were D341, I448, and D590 in Ssc1. According to the sequence alignment, these residues correspond to E318, V425, and A564 of *E. coli* DnaK and E319, R435, and K600 of yeast Sse1. Figure 6.2A shows the structure of the SBD of DnaK with bound substrate (Zhu *et al.*, 1996). The labeling positions are indicated as red and green spheres. The corresponding positions in Ssc1 (residues 448 and 590) were found by alignment of the sequences of Ssc1 and DnaK. The sensor with these labeling positions, Ssc1(448,590) or lid sensor, is sensitive to the conformation of the substrate binding lid. Similarly, the labeling positions of the domain sensor are sensitive to the inter-domain distance (residues 341 and 448). The labeling positions of the domain sensor are shown in the structure of Sse1 in Figure 6.2B.

The selected positions were mutated to Cysteines and the resulting constructs were expressed in BL21(DE3) cells as described in (Sichting *et al.*, 2005). Ssc1 was then purified over a His6-Mge1 column according to a protocol similar to (Weiss *et al.*, 2002). Residues with similar properties were identified in DnaK and mutated to Cysteines (E318 and V425 for the domain sensor, N458 and T563 for the lid sensor). Before the introduction of the Cysteines, the endogenous Cysteine residue of DnaK was mutated to Alanine. Overexpressed DnaK was purified in the same way as Ssc1. Mdj1 and Mge1 were expressed recombinantly and purified as described previously (Horst *et al.*, 1997). The expression and purification of Tim44 is described in (Mapa, 2009).

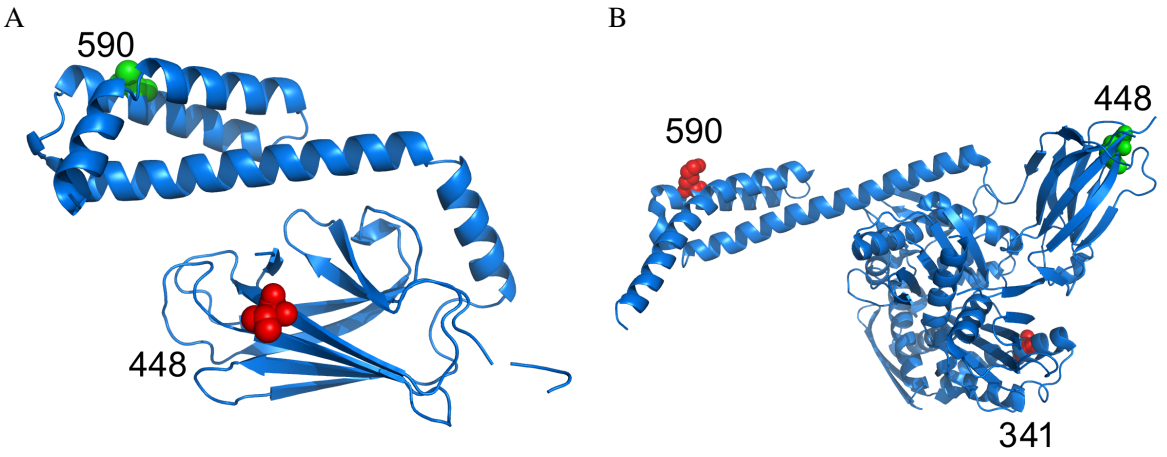


Figure 6.2: *Labeling positions for the double-labeled Ssc1 FRET sensors. (A) The structure of the SBD of DnaK with bound substrate (PDB 1DKX). The residues in DnaK corresponding to the labeling positions of Ssc1 are shown as colored spheres. The fluorophores were attached stochastically at the two locations. (B) The structure of Sse1 (PDB 2QXL). The residues in Sse1 corresponding to the labeling positions of Ssc1 are shown as colored spheres. The fluorophores were attached stochastically either at the sites 341 and 448 or 448 and 590.*

Double-cysteine variants of Ssc1 and DnaK were labeled for spFRET experiments at both positions simultaneously with Atto 532-maleimide as donor fluorophore and Atto 647N-maleimide (both Atto-Tec GmbH) as acceptor fluorophore according to the protocol provided by Atto-tec. Introduction of Cysteine residues at these positions had no effect on the viability of yeast cells, and labeling with Atto 532 and Atto 647N neither affected the basal ATPase activity of Ssc1 nor its stimulation by cochaperones (Mapa *et al.*, 2010).

6.2.2 Calculation of the error in the Förster radius

Using these double-labeled FRET sensors, spFRET experiments in solution were performed using PIE-MFD burst analysis as described in detail in Chapter 4. The major source of uncertainty of distances measured using spFRET originates from the orientational factor κ^2 in the Förster radius. To estimate the maximum error resulting from limited rotational freedom of the dyes, the residual anisotropies determined by a fit of the time-resolved anisotropies were used as described in Section 4.1.5.6. The residual anisotropies together with the maximum deviation of the Förster radius from the Förster radius determined for $\kappa^2 = \frac{2}{3}$ are given in Table 6.1.

6.3 Results

6.3.1 The conformation of Ssc1 in the presence of different nucleotides studied by spFRET in solution

To test whether FRET is observable in the Ssc1 constructs, the FRET-based sensors of the Ssc1 conformation were first examined by our collaborators in ensemble by measurement of the fluorescence spectra upon donor excitation in the presence of either ATP or ADP. There, the FRET efficiency was characterized using the ratio of the peak intensity in the spectral region of acceptor emission to the

Table 6.1: *Residual anisotropies and maximum error in the Förster radius for Ssc1 and DnaK*

	Residual anisotropies			Uncertainty in R_0	
	$r_{\infty,D}$	$r_{\infty,FRET}$	$r_{\infty,A}$	$R_{0,min}$	$R_{0,max}$
Ssc1(341,448)/ADP	0.12	0.03	0.13	-9 %	+18 %
Ssc1(341,448)/ATP	0.12	-0.02	0.14	-11 %	+16 %
Ssc1(341,448)/AMP-PNP	0.11	0.01	0.21	-12 %	+19 %
Ssc1(448,590)/ADP	0.13	-0.05	0.13	-12 %	+14 %
Ssc1(448,590)/ATP	0.10	-0.05	0.13	-12 %	+13 %
Ssc1(448,590)/AMP-PNP	0.08	0.03	0.21	-10 %	+19 %
Ssc1(341,448)/ATP + Mdj1	0.15	-0.03	0.16	-13 %	+18 %
Ssc1(341,448)/ADP + Tim44	0.07	0.02	0.20	-10 %	+18 %
Ssc1(341,448)/AMP-PNP + Tim44	0.10	0.00	0.21	-12 %	+19 %
Ssc1(341,448)/ATP + Mdj1/P5	0.13	0.01	0.15	-10 %	+18 %
Ssc1(448,590)/ATP + Mdj1	0.10	-0.03	0.14	-11 %	+15 %
Ssc1(448,590)/ADP + Tim44	0.12	-0.01	0.16	-11 %	+18 %
Ssc1(448,590)/AMP-PNP + Tim44	0.08	0.03	0.21	-10 %	+19 %
Ssc1(448,590)/ATP + Mdj1/P5	0.10	0.02	0.14	-9 %	+17 %
DnaK (318,425)/ATP	0.15	0.05	0.20	-12 %	+22 %
DnaK (318,425)/ADP	0.18	0.16	0.20	-12 %	+25 %
DnaK (318,425)/ADP + P5	0.05	0.02	0.16	-8 %	+16 %
DnaK (458,536)/ATP	0.15	0.10	0.16	-11 %	+23 %
DnaK (458,536)/ADP	0.16	0.13	0.20	-11 %	+24 %
DnaK (458,536)/ADP + P5	0.10	0.01	0.13	-9 %	+17 %

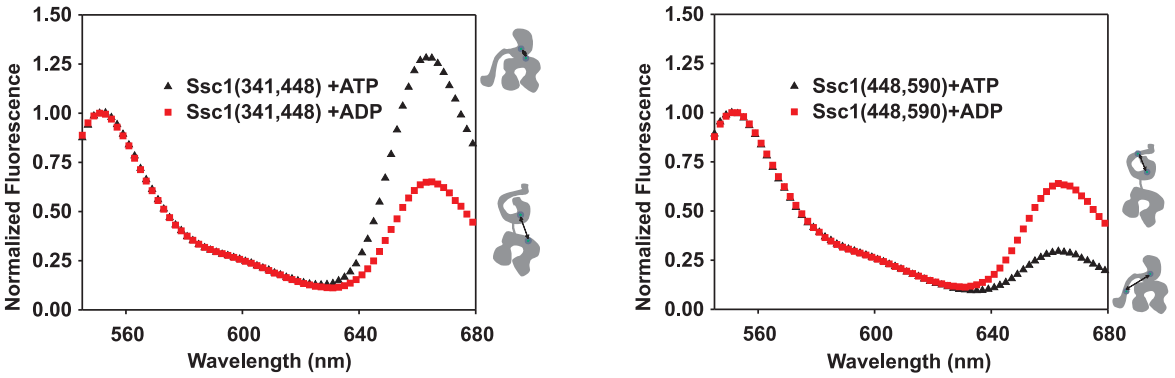


Figure 6.3: *Ensemble FRET experiments with Ssc1 in the presence of ATP or ADP. Fluorescence spectra recorded after exciting the donor fluorophore at 530 nm of double-labeled Ssc1(341,448) (A) and Ssc1(448,590) (B) with either 2 mM ATP (black curve) or 2 mM ADP (red curve). The fluorescence at 550 nm was normalized to one. Figure reproduced from Mapa et al. (2010).*

peak intensity in the spectral region of donor emission, hereafter referred to as the acceptor-to-donor ratio (ADR). Although this ratio does not provide a quantitative value for the FRET efficiency mainly due to the unknown amount of donor-only molecules, it allows an estimate of changes in FRET efficiency in one sample of proteins. This is especially useful in ensemble kinetics experiments, where changes in one sample are monitored over time.

The ADR of the domain sensor is higher in the presence of ATP than in the presence of ADP (Figure 6.3A). This suggests that the SBD and the NBD of Ssc1 are closer to each other in the presence of ATP than in the presence of ADP. In contrast, the ADR for the lid sensor in the presence of ADP is higher than in the presence of ATP (Figure 6.3B). The SBD appears to be more closed in the presence of ADP than in the presence of ATP.

The ensemble measurements suggest large conformational changes between ATP and ADP-bound states of Ssc1. However, without excessive control experiments, they do not provide a quantitative estimate of the FRET efficiency and thus of the distance between donor and acceptor. Moreover, they inherently represent the average of a number of conformational states potentially present but are not able to resolve possible inhomogeneities or subpopulations in the different nucleotide conditions. Both shortcomings of the ensemble FRET experiments can be overcome by single-pair FRET (spFRET) experiments using pulsed interleaved excitation (PIE) and multi-parameter detection (MFD) (refer to Section 4.1).

SpFRET histograms of Ssc1(341,448) and Ssc1(448,590) in the presence of ATP are displayed in Figure 6.4A-B. The single-molecule histograms of FRET efficiencies for the domain sensor in the presence of ATP showed a narrow distribution of FRET efficiencies E peaking at E=89 % (Figure 6.4A). The single-molecule histograms were analyzed using probability distribution analysis (PDA, Antonik *et al.*, 2006; Kalinin *et al.*, 2007, 2008). PDA allows to determine the number of species present in the experiment and the homogeneity of the individual states (refer to Section 4.1.5.7). The PDA fit required at least two subpopulations with a gaussian distribution of distances to describe the FRET efficiency histograms of Ssc1(341,448) in the presence of ATP. Each gaussian distance distribution was described by a mean interdye distance and a width (given as standard deviation σ). The values determined by PDA are given in Table 6.2.

The PDA analysis of both Ssc1 sensors in the presence of ATP required two gaussian distance distributions. The dominant population of 80 % of the molecules in the ATP-bound configuration was narrow with a standard deviation of σ=0.55 nm of the Gaussian distribution with a mean distance

Table 6.2: *Distance distributions for Ssc1 in the presence of different nucleotides*

	Major subpopulation			Minor subpopulation			D-only	
	%	d/nm	σ/nm	%	d/nm	σ/nm	%	reduced χ^2
Ssc1(341,448)/ATP	80	4.3	0.6	17	6.1	0.6	3	1.04
Ssc1(341,448)/AMP-PNP	77.9	4.3	0.5	21.6	6.0	1.3	0.6	1.90
Ssc1(341,448)/ADP	83	6.2	1.4	15	4.4	0.5	2	1.72
Ssc1(448,590)/ATP	84	7.7	0.9	13	5.3	0.6	3	0.65
Ssc1(448,590)/AMP-PNP	77.9	9.0	0.9	21.9	5.8	0.9	0.2	2.23
Ssc1(448,590)/ADP	90	6.8	1.4	10	4.3	0.5	0.0	1.00

between donor and acceptor of 4.2 nm. For the lid sensor, a dominant peak (84% of the molecules) at E=12 % was observed (Figure 6.4B). This corresponds to a distance of 7.7 nm demonstrating that the SBD is open when ATP is bound (Figure 6.4B). The width of this distribution was 0.9 nm, slightly wider than for the domain sensor.

Although the spontaneous hydrolysis rate of Ssc1 in the absence of substrate and J proteins is slow with a decay time of approximately 500 s (Mapa *et al.*, 2010), the spFRET histograms measured in the presence of ATP may include Ssc1 molecules where ATP has already been hydrolyzed to ADP. Adenosine 5'-(β,γ-imido)triphosphate (AMP-PNP), a non-hydrolyzable analog of ATP, was used to further investigate the conformation of Ssc1 bound to ATP. The spFRET histograms for Ssc1 in the presence of AMP-PNP are shown in Figure 6.4C-D. They closely resemble the histograms measured in the presence of ATP. Detailed analysis using PDA for the domain sensor showed that the mean and width of the major subpopulation are essentially unchanged (Table 6.2). The width of the minor subpopulation increased compared to ATP, however, the fitting accuracy for the fit of the width is impaired by the overlap between the two populations and the low amplitude of the minor subpopulation.

Analysis of the lid sensor in the presence of AMP-PNP using PDA showed small differences to the measurements in the presence of ATP. The lid is more open in the presence of AMP-PNP (9.0 nm) than in the presence of ATP (7.7 nm). This may be due to less contamination of the ATP-bound conformation by ADP due to the absence of hydrolysis. However, it is also possible that the structural differences between ATP and AMP-PNP result in a different conformation of Ssc1.

The FRET efficiency distributions measured for the domain and the lid sensor in the presence of ADP were more heterogeneous than in the presence of ATP with at least two subpopulations observable (Figure 6.4E-F). For the domain sensor in the presence of ADP, 83% of the molecules were in a very broadly distributed (σ=1.4 nm) population with an average donor-acceptor separation of 6.2 nm. The spFRET histograms of the SBD measured using the Ssc1(448,590) sensor was also described by at least two subpopulations. In the presence of ADP, the spFRET histogram of the lid sensor was very broad (Figure 6.4B). The majority of molecules (90%) was described by a distance distribution peaking at 6.8 nm and a width of 1.4 nm.

The subpopulations observed in the presence of ATP and ADP are similar with regard to their mean and width and differ mainly by their amplitudes. The presence of the two different species may be due to biochemical impurities within the sample (e.g. hydrolyzed ATP, inactive protein). Another explanation may be that the protein can exist in two conformations and binding of the ATP or ADP shifts the equilibrium between these conformations.

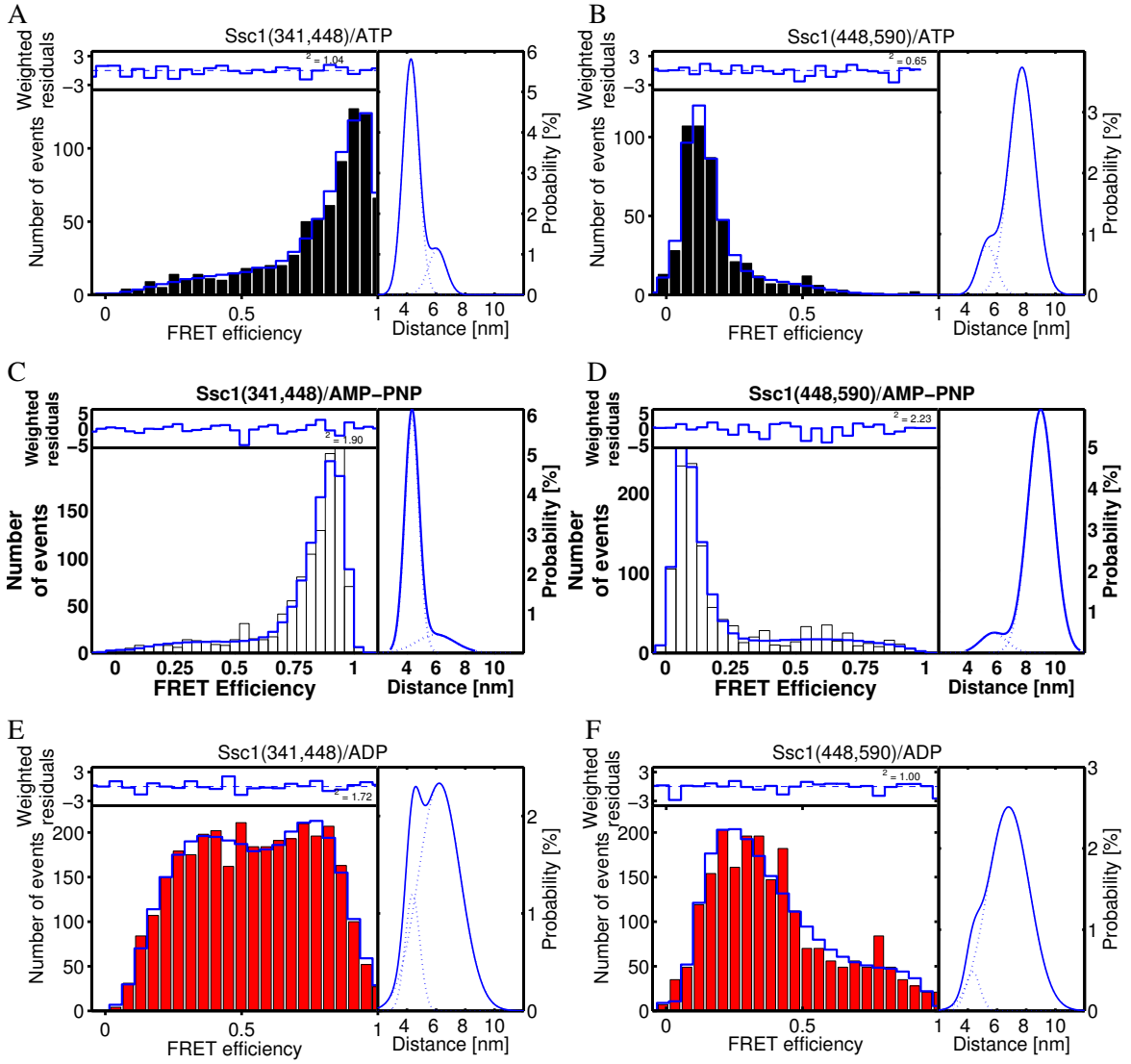


Figure 6.4: *SpFRET histograms of Ssc1 in the presence of different nucleotides. (Left panels) SpFRET histograms with PDA fits (blue line) of Ssc1(341,448) in the presence of ATP (black, A), AMP-PNP (white, C) and ADP (red, E), and of Ssc1(448,590) in the presence of ATP (black, B), AMP-PNP (white, D) and ADP (red, F). Weighted residuals from the PDA fits are shown above the histograms, and the corresponding distance distributions determined using PDA are shown in the right panels.*

The single-molecule data show a less well defined conformation of Ssc1 in the presence of ADP than in the presence of ATP. By analyzing the FRET efficiency of the bursts as a function of the burst-integrated fluorescence lifetime of the donor τ_{D_A} for the domain sensor (Figure 6.5, left panels), we found that the lifetime of bursts with intermediate FRET efficiency is longer than expected for a static sample (equation 4.22).

This effect can be explained by the presence of dynamic switching of a molecule between states of different FRET efficiency during the burst. Due to the low number of available photons for the determination of the burst-integrated fluorescence lifetime, only a single monoexponential lifetime is calculated per burst. Less photons are detected from the donor in the high-FRET state and thus more weight is given to the low-FRET state in the determination of the burst-integrated donor lifetime. This results in the observed longer donor lifetimes τ_{D_A} for bursts where the molecule undergoes a conformational change, i.e. with intermediate FRET efficiency. Gansen *et al.* (2009) have recently described the relationship between E and τ_{D_A} for a sample dynamically switching between different FRET efficiency states. E as a function of τ_{D_A} is then not described by the relationship in equation 4.22 anymore but by

$$E = 1 - \frac{\tau_1 \tau_2}{\tau_{\tau_{D_0}}(\tau_1 + \tau_2 - \tau_{D_A})} \tag{6.1}$$

where τ_1 and τ_2 are the donor lifetimes in the two different conformations and τ_{D_0} is the lifetime of the donor in the absence of the acceptor. For Ssc1(341,448), the lifetimes of the subpopulations were $\tau_1 = 3.1\,\mathrm{ns}$ and $\tau_2 = 0.4\,\mathrm{ns}$.

This analysis supports dynamic fluctuations between the different conformations observed in the domain sensor in the presence of ADP occurring within the duration of the bursts (5 ms on average). The major subpopulation of Ssc1(341,448)/ATP showed a static behavior, and only a minor subpopulation was described by the dynamic relationship. This minor subpopulation may be attributed to a contamination by ADP-bound molecules. In contrast, the E-τ_{D_A} histograms for the lid sensor in the presence of ADP or ATP did follow the relationship given for a static sample (Figure 6.6).

To determine the timescale of the observed dynamics, PDA with different time bins and fluorescence correlation spectroscopy were used. No evidence for conformational dynamics of Ssc1 on the µs to ms time scale (i.e. the time scale of the diffusion time of the protein through the confocal volume) was found using these methods for any of the above observed states. To check the presence of dynamics on longer timescales of 100 ms to minutes, we now employed single-molecule TIRF experiments of immobilized Ssc1.

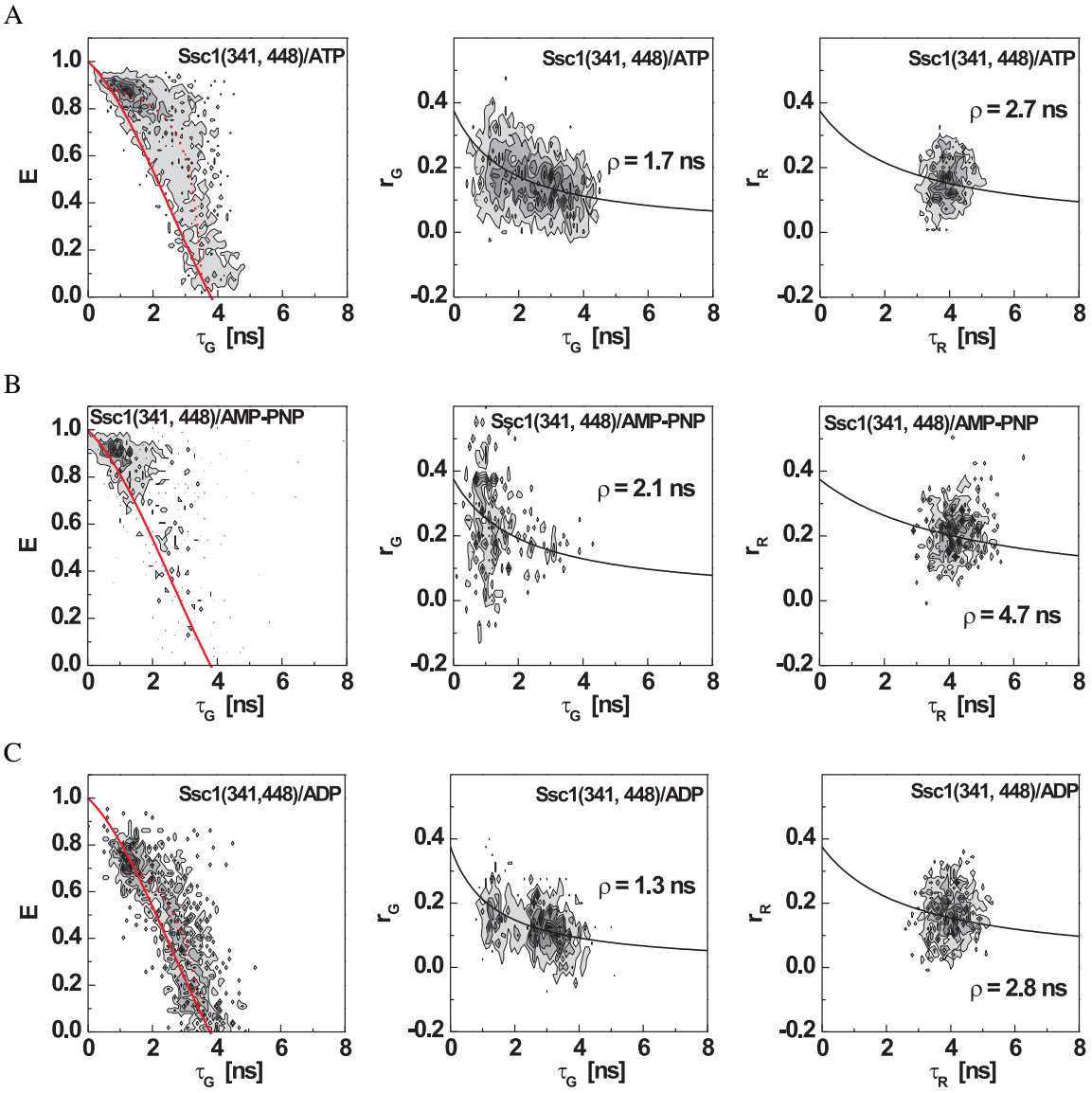


Figure 6.5: *Fluorescence lifetime and anisotropy histograms of Ssc1(341,448) in the presence of different nucleotides. 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for Ssc1(341,448) in the presence of ATP (A), AMP-PNP (B), and ADP (C). (Left panels) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). For Ssc1(341,448)/ADP and for a minor subpopulation of Ssc1(341,448)/ATP, the data does not fit this relation but the relation for a dynamic sample (equation 6.1, dashed line). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*

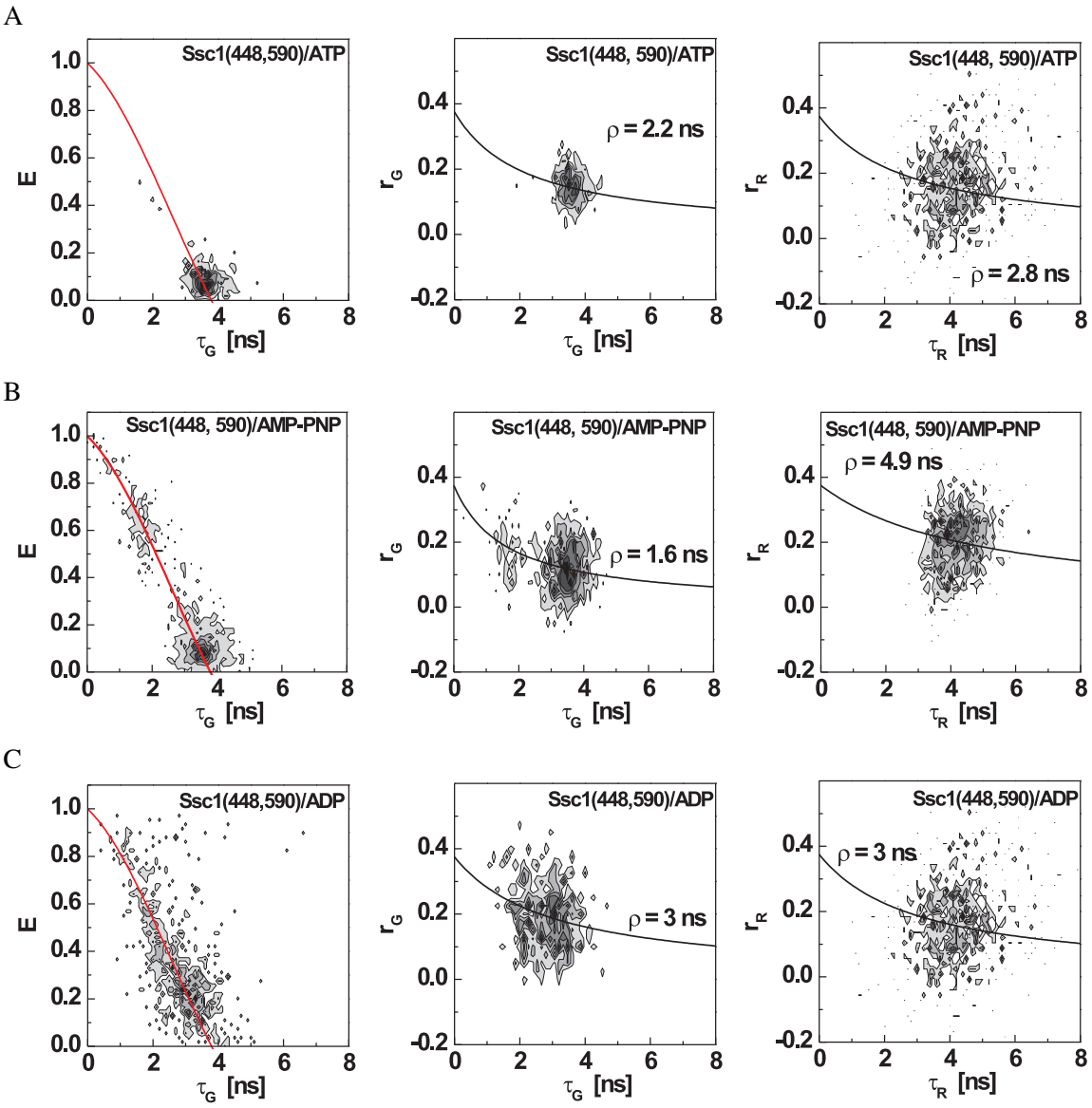


Figure 6.6: *Fluorescence lifetime and anisotropy histograms of Ssc1(448,590) in the presence of different nucleotides. 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for Ssc1(341,448) in the presence of ATP (A), AMP-PNP (B), and ADP (C). (Left panels) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*

6.3.2 Influence of Mdj1 and substrate on the conformation of Ssc1

The influence of J proteins and substrates on the conformation of Ssc1 during the folding cycle in the mitochondrial matrix was studied by addition of Mdj1, the J protein partner of Ssc1, alone or together with the model substrate peptide P5 (Sequence CALLLSAPRR, Metabion, Munich, Germany), to ATP-bound Ssc1 in the presence of free ADP (single turnover conditions). The spFRET histograms of both sensors bound to ATP in the presence of Mdj1 were similar to the ones obtained in the presence of ADP in both the ensemble (Mapa *et al.*, 2010) and single-molecule measurements (Figure 6.7). Mdj1 stimulates the ATPase activity of Ssc1, and the ATP-bound conformation of the chaperone is converted to the conformation in the presence of ADP. Ensemble-kinetics experiments performed by our collaborators showed that Mdj1 binds only transiently to Ssc1 and is already released at the time of the spFRET experiments (Mapa *et al.*, 2010).

In contrast, the spFRET histograms observed for both the domain and the lid sensor in the presence of Mdj1 and the substrate P5 were strikingly different. The Ssc1(341,448) sensor showed a shift to a single peak with a lower FRET efficiency of $E \approx 27\%$ (Figure 6.7A). The average distance between donor and acceptor increased to 7.5 nm in the presence of substrate (Table 6.2) indicating that substrate binding is required for full undocking of the two domains of Ssc1. The spFRET histograms of the lid sensor in the presence of Mdj1 and P5 showed a dominant peak ($E \approx 75\%$) in the presence of substrate corresponding to an average donor-acceptor separation of 5.2 nm and a width of 0.43 nm in the single-molecule measurements (Figure 6.7B). This width is comparable to the width determined for doubled-stranded DNA, implying that the protein is in a well defined conformation with the residual distribution in distance originating from acceptor photophysics or the flexibility of the linkers that attach the fluorophores to the protein.

A recent investigation has found significant differences in the conformation of BiP, the major Hsp70 from the endoplasmic reticulum, depending on the type of substrate bound (Marcinowski *et al.*, 2011). While the protein assumes an ADP-like conformation in the presence of a model substrate peptide similar to the peptide P5 used here, BiP assumes a very different, ATP-like conformation when bound to an authentic substrate, the C_H1 domain of the immunoglobulin heavy chain. Reduced carboxymethylated α-lactalbumin (RCMLA) is an intrinsically disordered protein and an authentic substrate of Ssc1. It was used to study the dependence of Ssc1 conformation on the type of bound substrate (Figure 6.7C-D). For the Ssc1(341,448) in the presence of ADP and RMCLA, a narrow FRET efficiency peak was observed corresponding to a narrow distance distribution with an average donor-acceptor separation of 8.0 nm (width 0.6 nm, Figure 6.7C). Thus, in the presence of the complete substrate protein RCMLA, the domains are even further undocked than in the P5-bound state. The lid of the SBD in the presence of RCMLA measured by the Ssc1(448,590) sensor was slightly more open (5.9 nm) than in the presence of P5 with a similarly narrow width (0.5 nm, Figure 6.7D)). Taken together, these data demonstrate that the conformation of Ssc1 is actively influenced by the binding of substrates and that different substrates are discriminated by Ssc1 by adopting slightly different conformations, but to a lesser extend as observed for BiP (Marcinowski *et al.*, 2011).

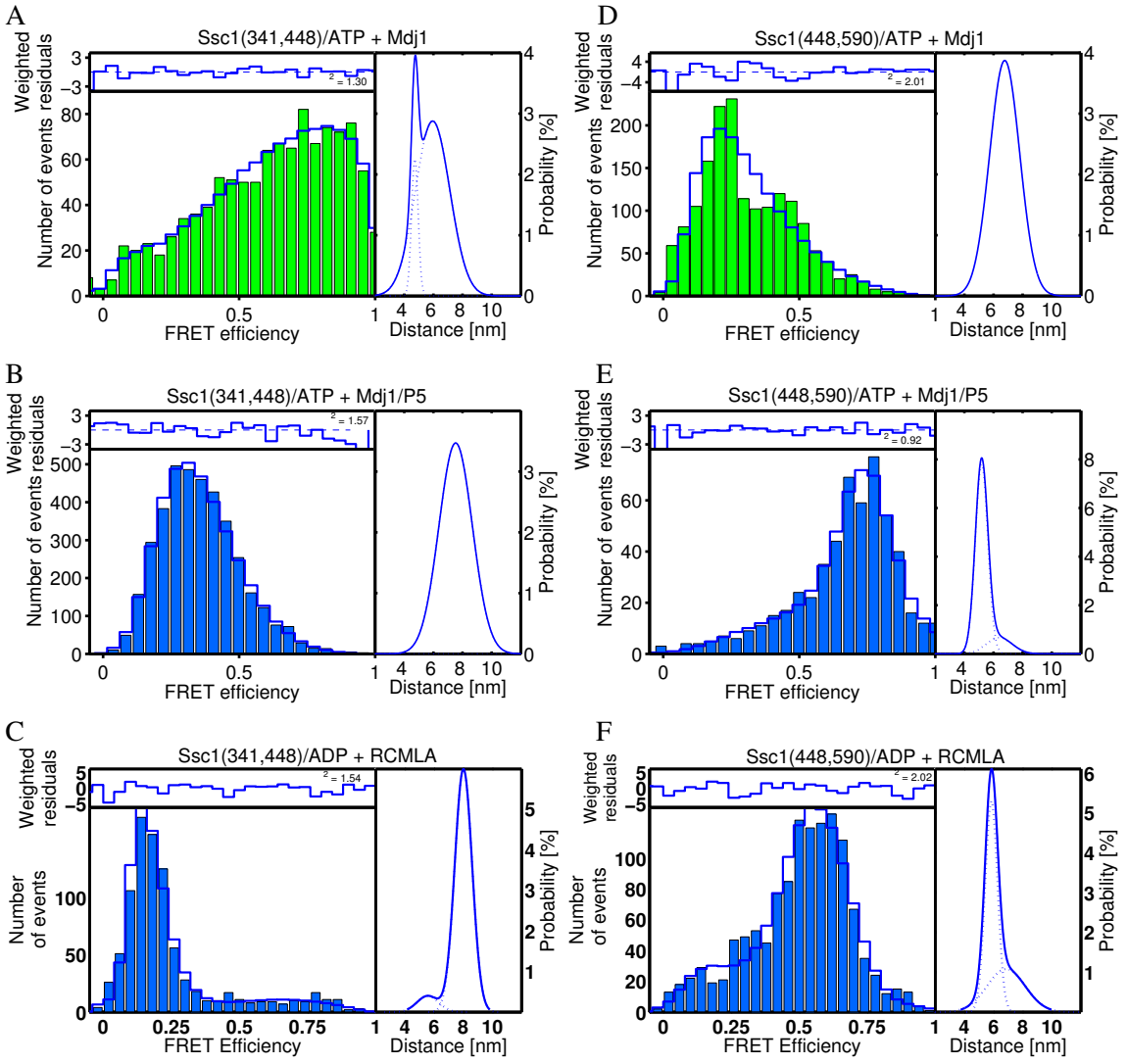


Figure 6.7: *SpFRET histograms of Ssc1 in the presence of Mdj1 and substrates. (Left panels) SpFRET histograms with PDA fits (blue line) of Ssc1(341,448) in the presence of ATP and Mdj1 (green, A), ATP and P5 (blue, B) and ADP and RCMLA (blue, C), and of Ssc1(448,590) in the presence of ATP and Mdj1 (green, D), ATP and P5 (blue, E) and ADP and RCMLA (blue, F). Weighted residuals from the PDA fits are shown above the histograms, and the corresponding distance distributions determined using PDA are shown in the right panels. The experiments with ATP were performed under single turnover conditions.*

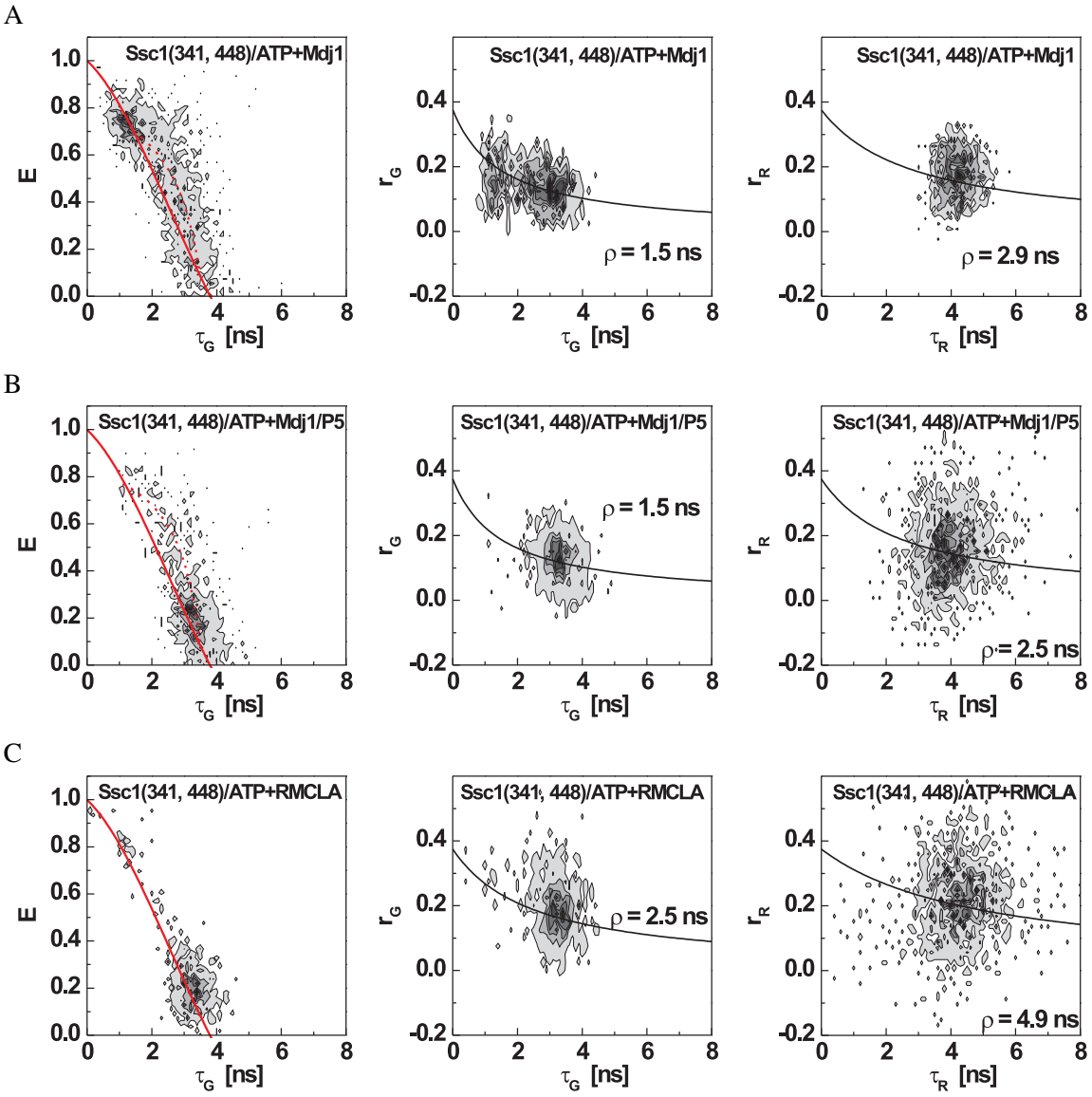


Figure 6.8: *Fluorescence lifetime and anisotropy histograms of Ssc1(341,448) in the presence of Mdj1 and substrates. 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for Ssc1(341,448) in the presence of ATP and Mdj1 (A), ATP and P5 (B), and ADP and RCMLA (C). (Left panels) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). For Ssc1(341,448)/ATP+Mdj1 and a minor subpopulation of Ssc1(341,448)/ATP+Mdj1/P5, the data does not fit this relation but the relation for a dynamic sample (equation 6.1, dashed line). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*

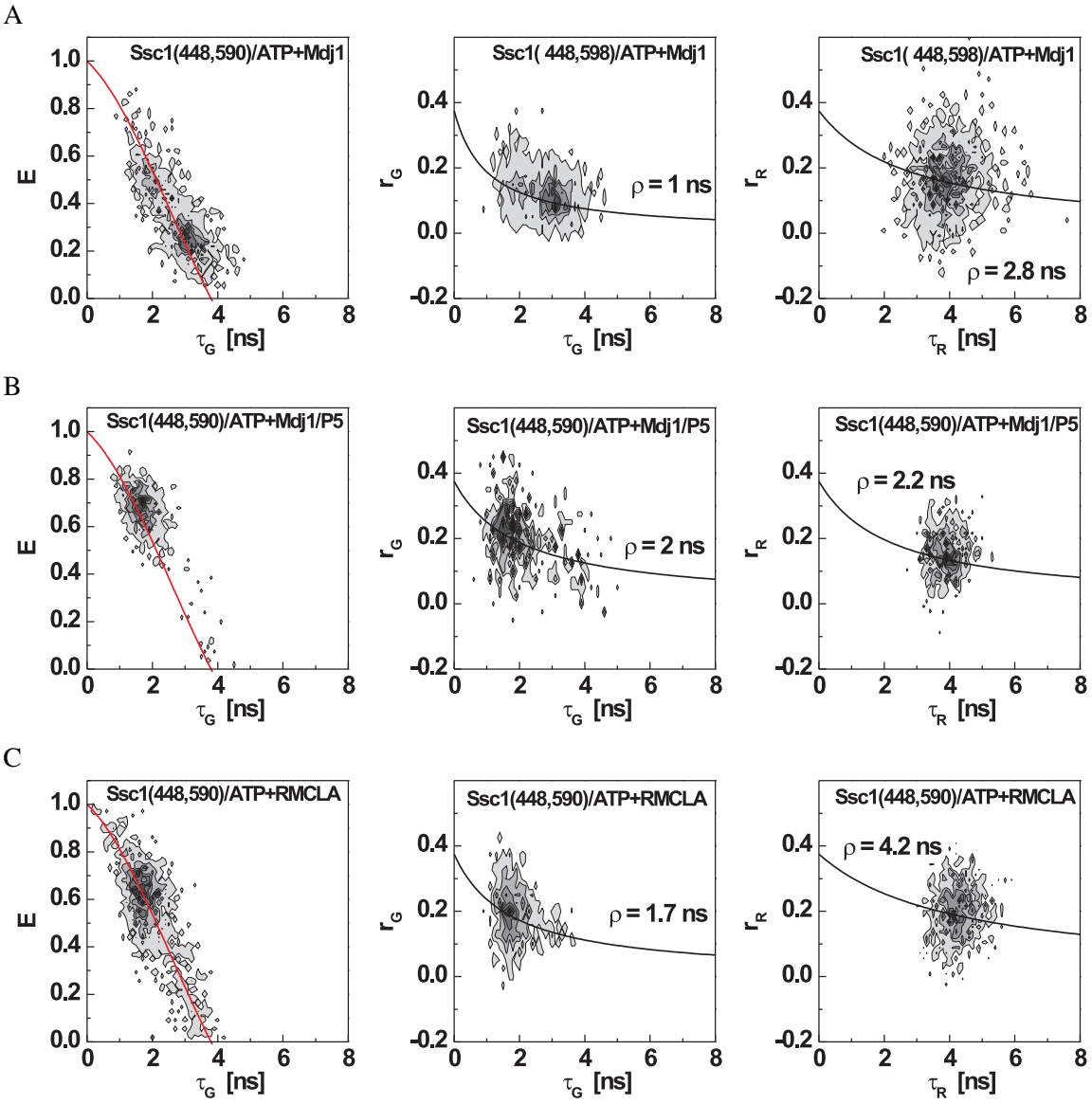


Figure 6.9: *Fluorescence lifetime and anisotropy histograms of Ssc1(448,590) in the presence of Mdj1 and substrates. 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for Ssc1(448,590) in the presence of ATP and Mdj1 (A), ATP and P5 (B), and ADP and RCMLA (C). (Left panels) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*

Table 6.3: *Distance distributions for Ssc1 in the presence of Mdj1 and substrates*

	Major subpopulation			Minor subpopulation			D-only	
	%	d/nm	σ/nm	%	d/nm	σ/nm	%	red. χ^2
Ssc1(341,448)/ATP[1]+ Mdj1	85	6.0	1.2	13	4.7	0.2	2	1.30
Ssc1(341,448)/ATP + Mdj1/P5	100	7.5	1.2	-	-	-	0	1.57
Ssc1(341,448)/ADP[2]+ RCMLA	84.1	8.0	0.6	14.4	5.6	0.7	1.5	1.54
Ssc1(448,590)/ATP + Mdj1	95	6.8	10.3	-	-	-	5	2.01
Ssc1(448,590)/ATP + Mdj1/P5	84	5.2	0.4	15	6.4	0.9	1	0.92
Ssc1(448,590)/ADP + RCMLA	69.3	5.9	0.5	30.4	6.9	1.3	0.4	2.02

[1] The experiments with Mdj1 and with Mdj1 and P5 were performed under single-turnover conditions with ATP-bound Ssc1 and ADP but no free ATP in solution.

[2] The experiments with RCMLA were performed with ADP-bound Ssc1 and ADP in solution.

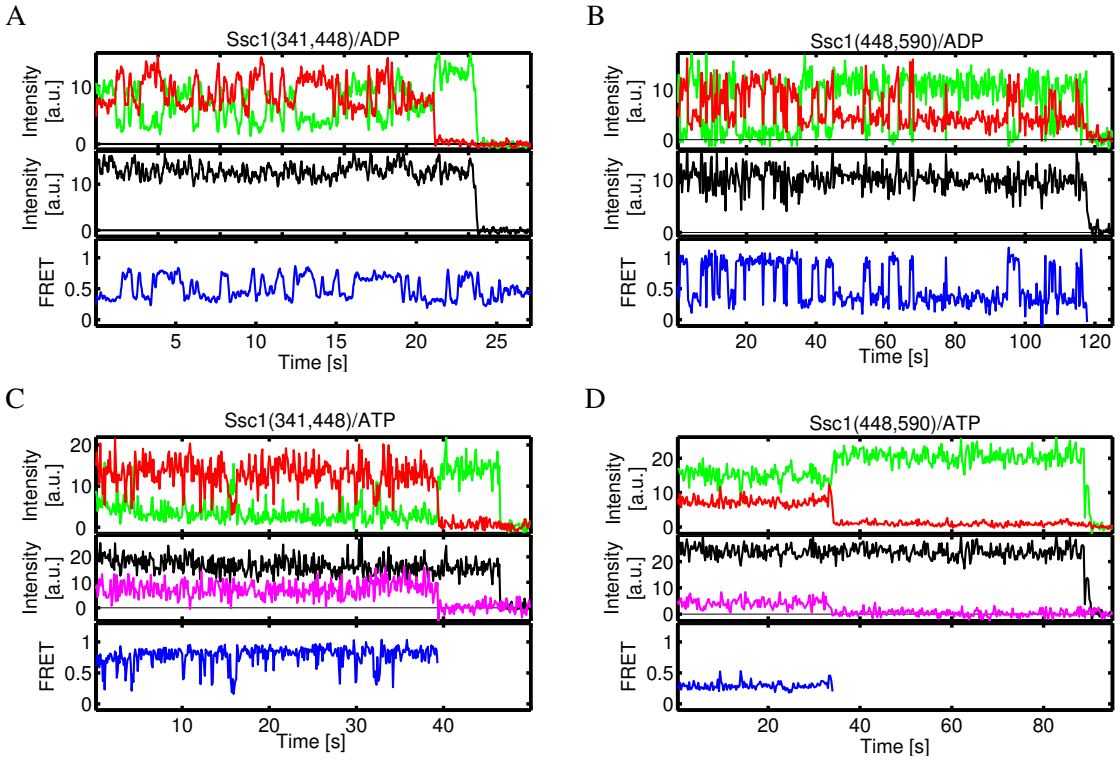


Figure 6.10: *Single-molecule TIRF time traces of the donor and acceptor intensity (upper panels), total intensity (middle panels) and FRET efficiency (lower panels) of Ssc1(341,448) in the presence of 1 mM ADP (A) and in the presence of 1 mM ATP (C), and of Ssc1(448,590) in the presence of 1 mM ADP (B) and in the presence of 1 mM ATP (D). The magenta line in the middle panel of (C) and (D) is the acceptor intensity after direct acceptor excitation measured using ALEX.*

6.3.3 Heterogeneity of Ssc1 is due to dynamic conformational changes

To identify the timescale of the dynamics that were observed in the MFD-PIE burst analysis experiments (Section 6.3.1), we immobilized the two FRET-based sensors to measure the time course of the distance between different residues in Ssc1 with TIRF microscopy. Ssc1 was encapsulated in lipid vesicles (Boukobza *et al.*, 2001; Okumus *et al.*, 2004) to avoid direct immobilization of the protein to the surface. The vesicles were immobilized using biotin-streptavidin-biotin linkage and visualized using TIRF microscopy. Depending on the presence of different nucleotides and of substrate peptide in the vesicles, we observed either static traces or dynamic switching between different FRET efficiencies in both FRET sensors. Figure 6.10A shows a typical trace for Ssc1(341,448) in the presence of ADP taken with 30 ms time resolution. The anti-correlation between the donor and acceptor signals (upper panel) with a constant overall intensity (lower panel) is a clear sign for FRET changes. The FRET signal is shown in the center panel and displays switching between an ∼50% FRET state and an ∼80% FRET state.

A similar behavior was observed for Ssc1(448,590) in the presence of ADP (Figure 6.10B). The FRET efficiency changes dynamically between ∼90% and ∼35% FRET. Here, we have expanded the time until the fluorophores bleach to 120s by choosing a lower excitation power and respectively longer exposure times of 250ms. Accordingly, the FRET changes of Ssc1(448,590)/ADP shown in the middle panel of Figure 6.10B occur on slower timescales than those in Ssc1(341,448)/ADP (Figure 6.10A).

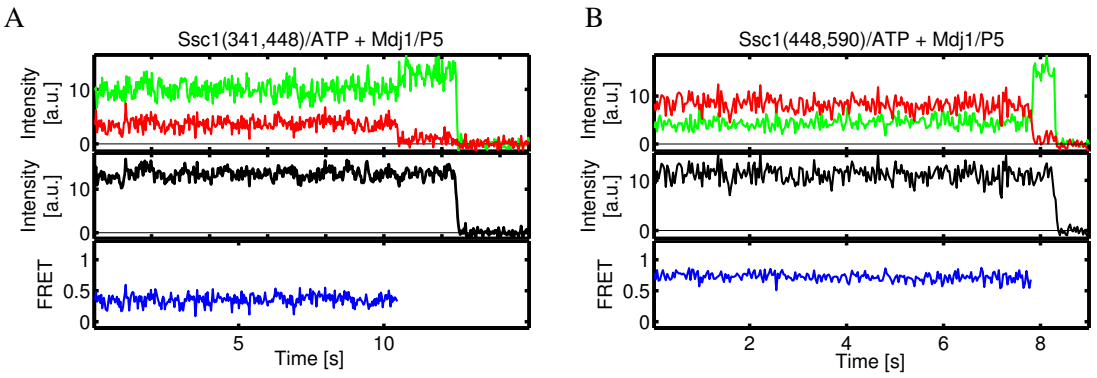


Figure 6.11: *Time traces of the donor and acceptor intensity (upper panels), total intensity (middle panels) and FRET efficiency (lower panels) of Ssc1(341,448) (A) and Ssc1(448,590) (B) in the presence of 1 mM ADP and 5 µM P5.*

In Ssc1(341,448) in the presence of ATP, we observed cycling of the inter-domain distance between a ∼90% FRET state and a ∼50% FRET state (Figure 6.10C, 100 ms time resolution). The dwell time of the 50% FRET state is short and transitions to the ∼90% FRET state occur after few frames. To exclude photophysical effects of the acceptor as the cause for these fast fluctuations, we performed experiments using alternating laser excitation (ALEX) (Kapanidis *et al.*, 2005) to directly probe the acceptor. The acceptor intensity after acceptor excitation is shown in Figure 6.10C (magenta line). No dynamics in the acceptor signal were observed until acceptor bleaching, excluding photophysics of the acceptor as cause of the FRET fluctuations. Contrary to Ssc1(341,448), the lid sensor Ssc1(448,590) did not show any dynamics in the presence of ATP (Figure 6.10D, 100 ms time resolution).

To check if substrate-bound Ssc1 shows slow dynamics, TIRF experiments with Ssc1 encapsulated together with ADP and P5 were performed[1]. Static FRET traces with FRET efficiencies of ∼30% for the domain sensor (Figure 6.11A) and ∼75% for the lid sensor (Figure 6.11B) were observed. These FRET values correspond very well to the FRET efficiencies measured by MFD-PIE experiments (Table 6.4).

6.3.4 Determination of underlying FRET efficiency states and transitions by Hidden Markov Modeling

To extract details of the conformational changes observed in the spFRET time traces for Ssc1, different analysis methods can be applied. A thresholding algorithm (Mickler *et al.*, 2009) can be used based on the transition of the FRET signal above or below a threshold chosen between the two states. However, for a FRET signal overlaid with noise, this could in theory result in unwanted false-positive transitions if the FRET efficiency crosses the threshold due to noise without an actual distance change, and could thus result in too fast observed rates [2]. Hidden Markov Modeling (HMM) has been developed in the 1960s (Baum and Petrie, 1966) to recover the signal of a Markov process

[1] For these experiments, no single turnover conditions were used. The same conformation was adopted under single turnover conditions after ATP hydrolysis and by direct addition of substrate in the presence of ADP.

[2] For each frame, an additional probability for an observed transition is introduced by the noise in addition to the actual transition probability. For example, if the threshold is chosen at one standard deviation of the distribution of framewise FRET efficiencies, for each frame the noise level will cross this threshold with a probability of 16%, or for a threshold at 2σ with 2.3%, etc. Additionally, for these noise-induced transitions, a high probability exists to return to the first state (e.g. 84% for threshold at 1σ).

overlaid by noise, and was applied to speech recognition problems for several decades (Rabiner and Juang, 1986). A typical example for a biophysical application of HMM is modeling of a time trace of the current through an ion channel switching between an open and a closed state observed by electric single channel recordings (Qin *et al.*, 2000). Analysis of single-molecule fluorescence data using HMM has been studied theoretically (Andrec *et al.*, 2003), and Ha and coworkers have applied HMM to extract the number of states, their FRET efficiencies and rates from spFRET time traces (McKinney *et al.*, 2006). The number of states can either be determined by overfitting with a high number of states and a cluster analysis or by using Bayesian maximum evidence (Bronson *et al.*, 2009). For the data of Ssc1(341,448) in the presence of ADP or ATP and of Ssc1(448,590) in the presence of ADP, both methods demonstrated that two states are sufficient to model the data.

Thus, each trace is modeled using an individual parameter set comprised of two FRET efficiencies and two transition rates to account for molecule-to-molecule variations of the conformational states. Then, the end points of all idealized transitions found by the HMM algorithm (Figure 6.12A) are plotted over the transition start points in transition probability plots (Figure 6.12B-D). The transitions in the Ssc1(341,448) sensor in the presence of ADP are clustered in two regions of the plot in Figure 6.12B suggesting the existence of two conformations with distinct FRET efficiencies. The cluster in the right part of the transition probability plot represents transitions from a conformation with closely docked domains of $\sim$80% FRET to a conformation with more undocked domains with $\sim$50% FRET, and the other cluster represents transitions in the opposite direction. Similarily, we observed two clusters of transitions in Ssc1(341,448) in the presence of ATP between one conformation with closely docked domains (E$\approx$90%) and a more expanded conformation (E$\approx$50%)[3]. For Ssc1(448,590) in the presence of ADP, we observed two clusters of transitions between one conformation with a tightly closed lid (E$\approx$90%) and a more open lid (E$\approx$40%).

The idealized traces available from the HMM analysis can be used to determine the FRET efficiencies of the different states underlying the spFRET traces (Section 6.3.5) and the kinetic rates between the states (Sections 6.3.6 and 6.3.7).

6.3.5 FRET efficiency distributions derived from HMM

The FRET efficiencies in the different conformations of the domain sensor in the presence of ADP and ATP resolved by the HMM analysis are compared to FRET efficiencies of substrate-bound Ssc1 in Figure 6.13. For the spFRET measurement of substrate-bound Ssc1(341,448), the protein was encapsulated together with ADP and P5, a short peptide acting as a model substrate for Ssc1 (Section 6.3.2). It was shown before using single-molecule burst analysis (Section 6.3.1) that Ssc1 assumes a defined, narrow distance distribution with undocked domains and a closed lid in the substrate-bound state. The spFRET TIRF assay using the encapsulated Ssc1 domain sensor yielded a FRET efficiency of 26% with bound substrate (Figure 6.13A and Tab. 6.4). This can be related to the distance between the dyes using a Förster radius of 6.1 nm (Mapa *et al.*, 2010), resulting in a 7.3 nm separation in the substrate-bound state in excellent agreement with the previous measurements in solution (Table 6.4). This is a good indication that there are no artifacts introduced by possible interactions with the lipids or other encapsulation-related effects.

[3]The increased width of the $E \approx$50% FRET state in the measurement of Ssc1(341,448) with ATP is probably due to the short dwell time of this state (refer to Figures 6.10 and 6.17) and the resulting low statistics for this state. The correlation beteen the initial and final FRET values (lower initial FRET ($\sim$90%) transitions to lower final FRET ($\sim$40%) and vice versa) may be due to erroreous determination of the γ factor. The γ factor was determined for most traces by the dynamic method (refer to Section 4.3.5). The performance of the dynamic γ determination may suffer from the low percentage of time in the $\sim$50% FRET state.

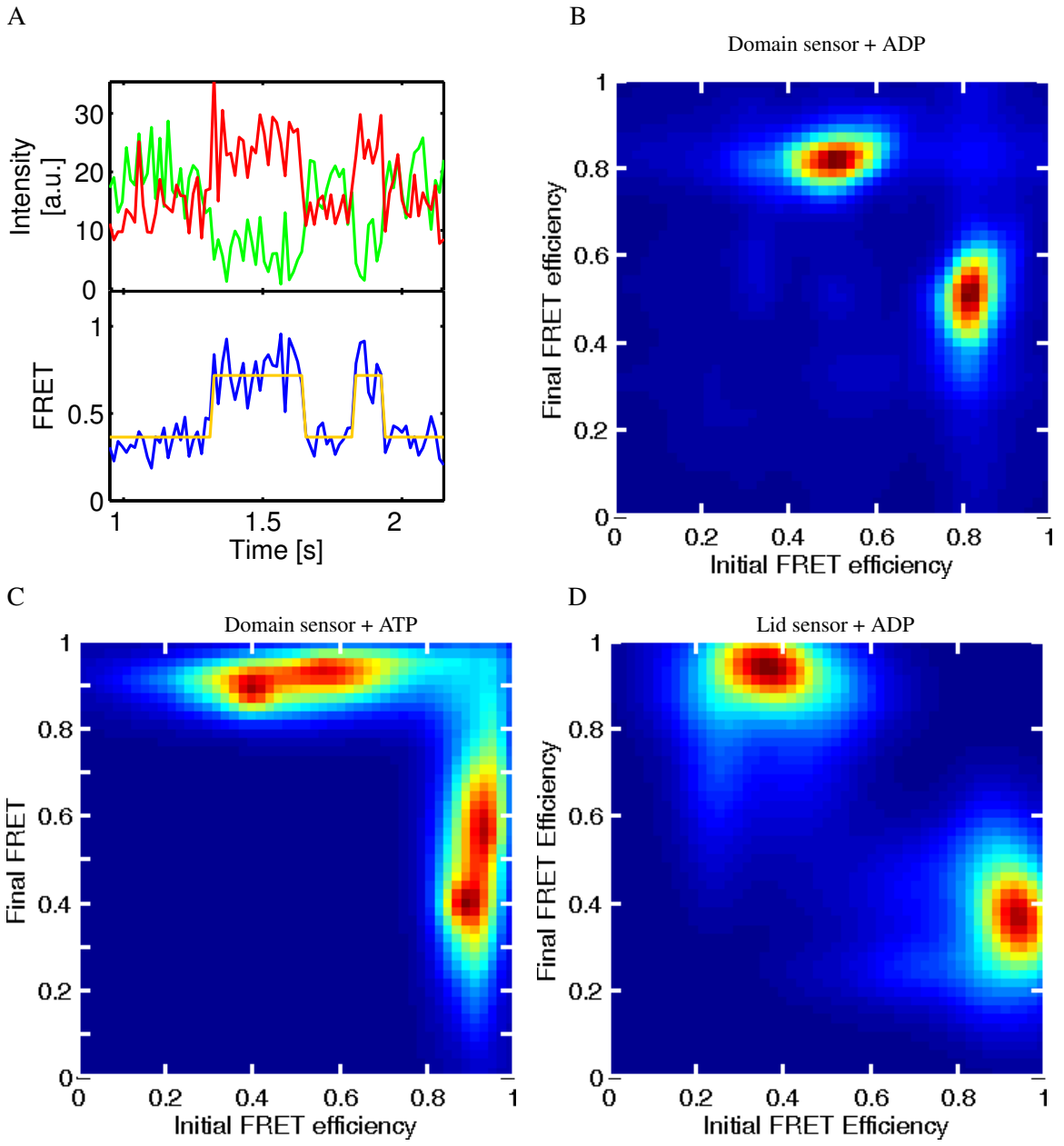


Figure 6.12: *Hidden markov modeling and transition density plots of smTIRF time traces of Ssc1. (A) FRET efficiency trace (blue) with corresponding idealized trace (red) from the HMM analysis to identify the FRET efficiencies and transition rates underlying the experimental data. Donor and acceptor intensity (green and red, upper panel) and FRET efficiency (lower panel) are shown over time. The experimentally measured FRET efficiency (blue) is overlaid with the result of the HMM analysis (red). (B-D) Transition density plots for Ssc1(341,448)/ADP (B), Ssc1(448,590)/ADP (C), and for Ssc1(341,448)/ATP (D) constructed from the transitions found by the HMM analysis.*

The two conformations observed for the domain sensor in the presence of ADP as well as the two conformations observed in the presence of ATP are clearly different from the substrate-bound conformation. While the respective high FRET populations have clearly different mean FRET efficiencies, the low FRET subpopulations observed in the presence of ATP and ADP have surprisingly similar average FRET values. The greater width of the low-FRET state with ATP than with ADP is due to the increased statistical uncertainty because of the low number of frames the molecule spends in this conformation. At this point, it could be speculated that it is indeed the same conformation which is assumed both in the presence of ADP and ATP, and Ssc1 cycles between this conformation and another nucleotide dependent conformation.

For the Ssc1 lid sensor, we measured a FRET efficiency of 13% in the ATP state and 73% in the substrate-bound state, corresponding to 8.4 nm and 5.2 nm, respectively (Figure 6.13B). This is also in good agreement with our previous results (Table 6.4). In the experiments with bound substrate, we observed a subpopulation (19%) of molecules showing identical switching behavior as in the presence of ADP alone. Although the bulk concentration of substrate peptide was high enough to ensure the presence of several substrates in each vesicle, the local concentration inside the vesicles may have been lower due to incorporation of the hydrophobic substrate peptide into the lipid bilayer. We thus attribute this subpopulation to vesicles containing no available substrate and excluded them from the histogram for the substrate-bound state. Apart from these molecules, we did not observe any dynamics in these measurements as expected from the narrow distance distributions in the ATP and ADP/P5 states (Mapa *et al.*, 2010).

The two conformations of the SBD observed by the lid sensor resolved by the HMM analysis in the presence of ADP are clearly distinct from the conformations observed in the presence of ATP and when substrate is bound (Figure 6.13B). While the presence of ATP results in a widely open lid ($E \approx 13\%$) and substrate binding triggers lid closure ($E \approx 74\%$), in the presence of ADP the SBD cycles between an intermediate state with $\sim 37\%$ FRET efficiency and a tightly closed state displaying a FRET efficiency even higher than in the substrate-bound state ($E \approx 90\%$).

Table 6.4: *FRET efficiencies and distances compared for TIRF and MFD-PIE experiments.*

		FRET efficiency [%]	Distance [nm]	
		TIRF[1]	TIRF[2]	MFD-PIE
Ssc1(341, 488)	ADP High-FRET	81.1±0.2	4.8 (4.2;5.6)	4.4
	ADP Low-FRET	50.2±0.9	6.1 (5.3;7.1)	6.2
	ATP High-FRET	90.6±0.8	4.2 (3.6;4.8)	4.3
	ATP Low-FRET	47.8±3.1	6.2 (5.3;7.3)	6.1
	ADP/P5	26.5±0.6	7.2 (6.3;8.4)	7.5
Ssc1(488, 590)	ADP High-FRET	88.7±0.9	4.3 (4.1;4.8)	4.3
	ADP Low-FRET	37.1±1.3	6.7 (6.2;7.3)	6.8
	ATP	12.8±0.9	8.4 (7.8;9.4)	7.7[3]
	ADP/P5	73.5±0.8	5.1 (4.8;5.7)	5.2

[1] Error estimated from covariance matrix of the Fit)

[2] Maximum error from κ^2 and E error)

[3] The relatively large deviation between the MFD-PIE data and the TIRF data for Ssc1(448,590)/ATP is possibly due to the high FRET efficiency in this state. The high E results in large errors in the distance determination due to the $\frac{1}{R^6}$ dependence.

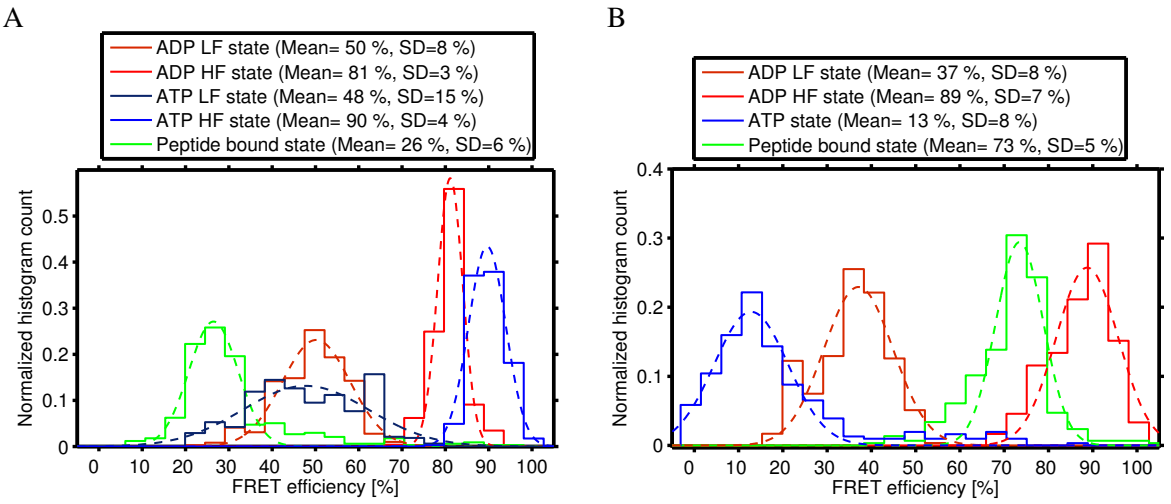


Figure 6.13: *SpFRET histograms derived from the HMM analysis of Ssc1. (A) FRET efficiency histograms of the two inter-domain conformations identified by the HMM analysis in the presence of ADP (dark/light red) and in the presence of ATP (dark/light blue), together with the molecule-wise FRET histogram of the domain sensor in the presence of ADP and P5 (green). (B) FRET efficiency histograms of the two SBD conformations identified by the HMM analysis in the presence of ADP (dark/light red) and molecule-wise FRET histograms of the SBD conformation in the presence of ATP (blue) and in the presence of ADP and P5 (green). (A,B) Single Gaussian fits to the data are shown by dashed lines. The average FRET value and the standard deviation are given in the figure legend.*

6.3.6 Dynamics of the Ssc1 substrate binding domain

The HMM analysis does not only allow the quantification of the FRET efficiencies of the states but also the determination of the dwell time distributions. Thus, the rate of the conformational changes was quantified as a function of ADP concentration. The dwell time distributions are presented as cumulative distributions in Figure 6.14, i.e. the number of molecules remaining in the respective state over time, to avoid data reduction by binning of the dwell times.

From the transition density plot in Figure 6.12C, it is clear that the SBD of Ssc1 cycles between at least two different conformations in the presence of ADP. HMM analysis was used to construct cumulative dwell time histograms for the closed lid ($E \approx 89\%$) and open lid ($E \approx 37\%$) states (Figure 6.14). In the presence of 1 mM ADP, the average dwell time of the lid opening transition is 3.8 s, and the average dwell time of the lid closing transition is 4.2 s. Both transition rates do not change significantly with ADP concentration.

An additional interesting feature of the observed SBD dynamics is that the dwell time distribution of the closed lid state does not fit to a single exponential decay. This suggests that indeed there is more than one closed lid state present in the SBD and the Ssc1(448,590) sensor is not sensitive to this conformational change. Different fit models were used to fit the data (Figure 6.15). A stretched exponential could account for an ensemble of states with a closed lid. However, the decay time distribution resulting from this fit is extremely wide ranging over ~3 orders of magnitude. Another possible explanation may be the existence of three states with two different states with FRET efficiencies indistinguishable for the lid sensor (Figure 6.14C). This results in double-exponential kinetics. After lid closure, the SBD can switch either back to an open lid conformation, giving rise to the fast

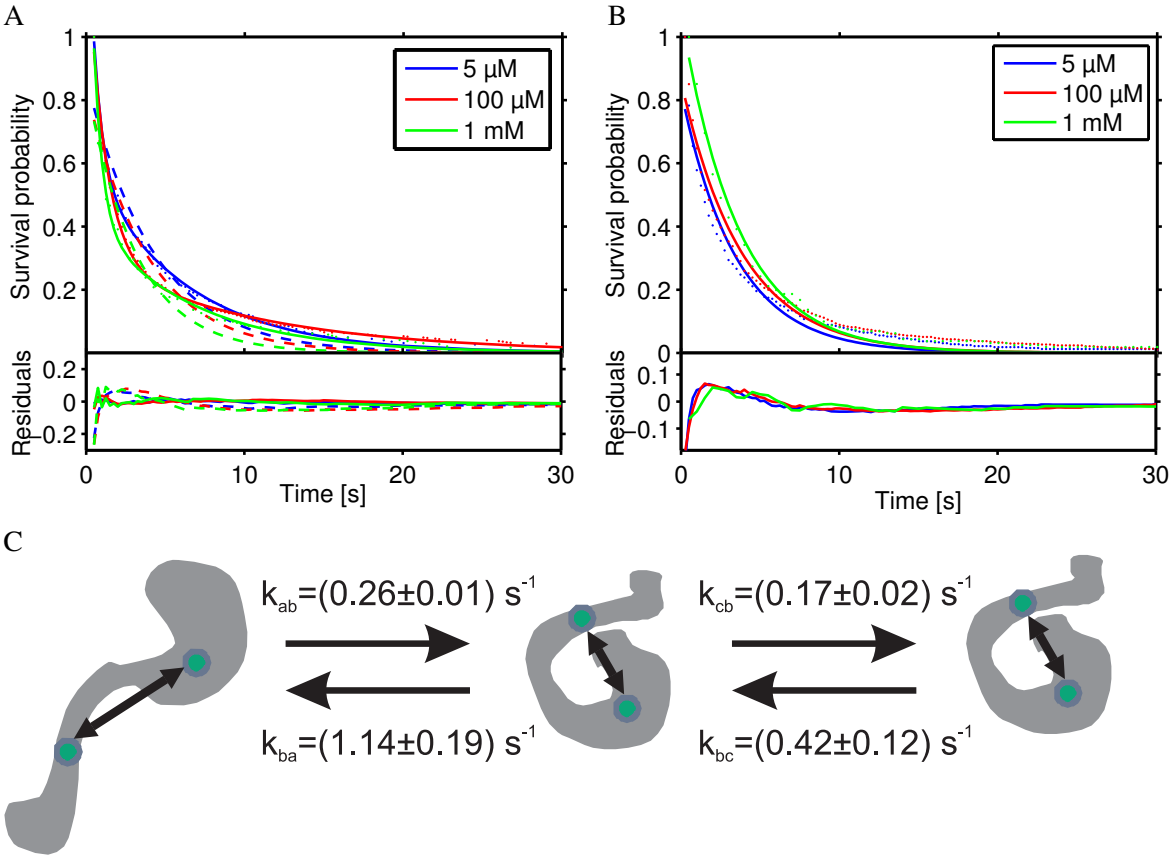


Figure 6.14: *Single-molecule kinetics of the substrate binding domain of Ssc1. (A) Cumulative dwell time histogram (dots) and single-exponential (dashed lines) and double-exponential fits (solid lines) of the lid opening transitions. (B) Cumulative dwell time histogram (dots) and single-exponential fits (solid lines) of the lid closing transitions. (A,B) The cumulative dwell time histograms calculated from the Viterbi paths from the HMM analysis for Ssc1(448,590) in the presence of 1 mM (green), 100 µM (red), and 5 µM ADP (blue). (C) Model of Ssc1 lid dynamics in the ADP state derived from a double-exponential fit of the lid-opening transition and a monoexponential fit of the lid closing transition. The rates are given as average values with the standard error for the measurements at different ADP concentration.*

component of the dwell time decay, or to the other closed lid conformation. In this case, it will take longer to return to the open lid conformation as observed in the slower component of the fit. Reports in the literature suggest three states in the ADP state of the SBD of DnaK (Schlecht *et al.*, 2011). Thus , we used the double-exponential fit to extract the rates k_{ba}, k_{bc}, and k_{cb} in the model in Figure 6.14C. The rate from the open to the closed lid state k_{ab} was determined by a monoexponential fit to the dwell time histograms in Figure 6.14.

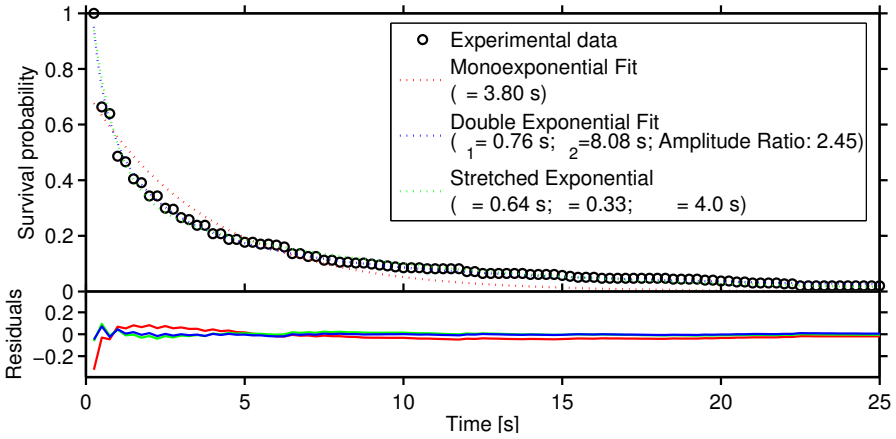


Figure 6.15: *Comparison of a single-exponential fit (red), a double-exponential fit (blue), and a stretched exponential fit (green) to the cumulative dwell time histogram for the lid opening transition. The fit results are given in the figure legend. The residuals for the fits are shown below the graph. Although the double-exponential and the stretched exponential fit the data very well, the double-exponential fit was preferred due to similar models in the literature (Schlecht et al., 2011).*

Table 6.5: *Kinetic rates[1] of Ssc1 determined from TIRF experiments*

[ADP]	5 µM	100 µM	1 mM	100 µM + 5 µM Mge1	1 mM ATP
Domain docking	1.08 ± 0.02	1.65 ± 0.03	1.78 ± 0.04	1.61 ± 0.03	4.12 ± 0.20
Domain undocking	1.35 ± 0.04	1.25 ± 0.02	1.37 ± 0.03	1.13 ± 0.03	0.30 ± 0.01
Lid k_{ab}	0.27 ± 0.02	0.24 ± 0.01	0.28 ± 0.02		
Lid k_{ba}	1.49 ± 0.10	1.11 ± 0.05	0.84 ± 0.05		
Lid k_{bc}	0.64 ± 0.06	0.32 ± 0.03	0.28 ± 0.03		
Lid k_{cb}	0.21 ± 0.01	0.14 ± 0.01	0.18 ± 0.01		

[1] The rates are given in units of s^{-1}. The errors were estimated from the Jacobian of the Fits.

6.3.7 Inter-domain dynamics of Ssc1

The cumulative dwell time distributions for the domain sensor at different ADP concentrations are displayed in Figure 6.16. Surprisingly, the dynamics of the inter-domain separation show two important differences compared to the SBD dynamics. First, the rates are more than a factor of five faster for the inter-domain dynamics compared to the intra-SBD dynamics. Second, while the rates for the SBD dynamics were not affected by the ADP concentration, the rate of domain docking is increased by 50% from 1.08 s^{-1} at 5 µM ADP to 1.78 s^{-1} at 1 mM ADP. At the same time, the rate of domain undocking was 1.35 s^{-1} in the presence of 1 mM ADP and remained essentially unchanged for ADP concentrations down to 5 µM. A straightforward hypothesis to explain the ADP dependence of the domain docking rate is to assume that the 50% FRET state indeed is a nucleotide-free state, and the conformational change is induced by binding and release of ADP.

This hypothesis is further supported by the fact that the average FRET efficiency of the minor subpopulation present in measurements with Ssc1(341,448) and ATP is very similar to the average FRET

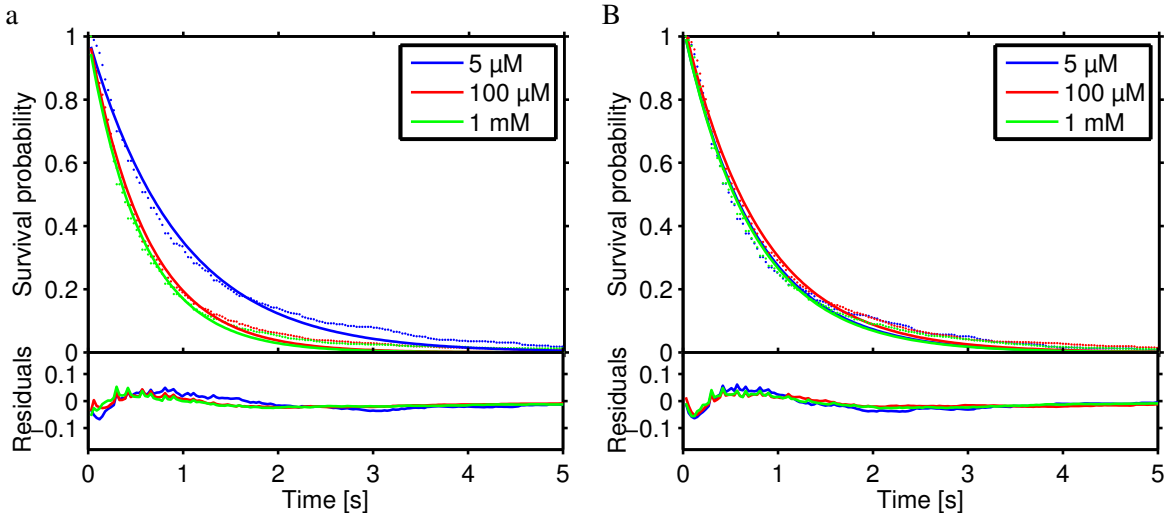


Figure 6.16: *Single-molecule kinetics of the inter-domain distance of Ssc1. (A,B) Cumulative dwell time histograms (dots) and single-exponential fits (lines) of the domain docking transitions (A) and the domain undocking transitions (B) using the Viterbi paths from the HMM analysis for Ssc1(341,448) in the presence of 1 mM (green), 100 µM (red), and 5 µM ADP (blue).*

efficiency of the low-FRET subpopulation of Ssc1(341,448) in the presence of ADP both in MFD-PIE and TIRF experiments (Figure 6.13 and Table 6.4). The increased width of the ∼50% FRET state in the measurement with ATP is probably due to the high rate out of this state (refer to Figures 6.10 and 6.17) and the resulting low statistics for this state.

6.3.8 Influence of Mge1 on the inter-domain distance dynamics

To further investigate the hypothesis that the 50% FRET state is the nucleotide-free state of Ssc1, the influence of the nucleotide exchange factor Mge1 on the observed dynamics was analyzed. These experiments have the appealing additional benefit that they could allow the direct study of the effect of Mge1 on the nucleotide exchange during the functional cycle of Ssc1.

Figure 6.17A-B show the cumulative dwell time histograms for the domain docking and undocking transitions for Ssc1(341,448) with 100 µM ADP in the presence and absence of Mge1. Clearly, Mge1 does not have an effect on the rates between the compact 80% FRET state and the conformation with more undocked domains. In contrast, Mge1 had an interesting effect on the behavior of Ssc1 in the presence of ATP. Without Mge1, transitions were observed between the 50% FRET state and the 90% FRET state with a domain undocking rate slower than in the presence of ADP and faster domain docking (Table 6.5). These transitions vanished completely in the presence of ATP and Mge1 (Figure 6.17C), and a static FRET efficiency of ∼90% was observed. Assuming that the 50% FRET state is a nucleotide-free form of Ssc1, this observation can be explained in two ways. If the on rate of ATP (domain docking rate) is unchanged by Mge1, the off rate of ATP (domain undocking rate) is affected by Mge1. It is required to be lower than 1.8×10^{-4} s^{-1} to explain that no transition was observed for the entire observation time. If, on the other hand, Mge1 affects the on rate instead of the off rate of ATP, the on rate has to be more than 250 s^{-1} to make the dwell times short enough to be undetected. The latter hypothesis is more reasonable, since binding times of Mge1 to Ssc1 of thousands of seconds would be orders of magnitude longer than the typical cycling time (∼20 s).

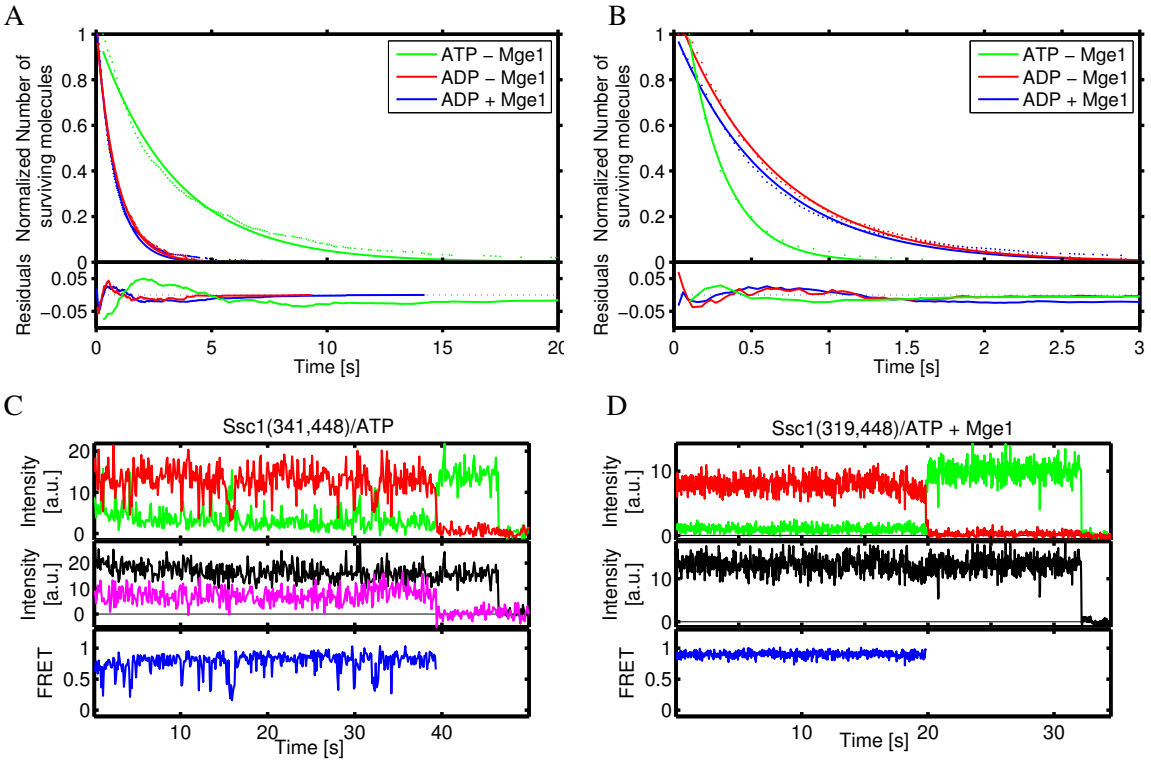


Figure 6.17: *Influence of Mdj1 on the inter-domain distance kinetics of Ssc1. (A,B) Cumulative dwell time histograms (dots) and single-exponential fits (lines) of the domain docking transitions (A) and the domain undocking transitions (B) calculated using the Viterbi paths from the HMM analysis for Ssc1(341,448) in the presence of 100 µM ADP (blue), in the presence of 100 µM ADP and 5 µM Mge1 (red), and in the presence of 1mM ATP (green). (C) Dynamic time traces of Ssc1(341,448) in the presence of 1 mM ATP without Mge1. (D) Static time traces of Ssc1(341,448) in the presence of 1 mM ATP and 5 µM Mge1. (C,D) The donor and acceptor intensity are shown in the upper panel, the total intensity in the middle panel and the FRET efficiency in the lower panel.*

Additionally, it is appealing to assume that Mge1 promotes the conformational change in Ssc1 on ATP binding, and thus increases the on rate.

6.3.9 Cycling of Ssc1 in the presence of ATP and Substrate

To understand the role of the putative nucleotide-free state of Ssc1 in the functional cycle of Ssc1, measurements of the domain sensor in the presence of ATP and substrate were performed. To this end, Ssc1(341,448) was encapsulated in 200 nm LUVs with 5 mM ATP and 0.5 mM P5. The inter-domain distance switches dynamically between a 90% FRET state and the 50% FRET state, similar to the measurement with Ssc1(341,448) with ATP only (Figure 6.18). However, additional transitions were observed leading to the peptide-bound (30% FRET) state. This state had a significantly longer dwell time than the 50% state, as expected for the substrate-bound state.

The molecules switch in all cases directly from the 90% FRET state to the 30% FRET state (Figure 6.18C). This could be interpreted by peptide binding only in the ATP-bound state but not in the nucleotide-free state. However, due to the short dwell time of the 50% FRET state, it is possible that population of the 50% FRET state on the way to the 30% FRET state is to short to be observed.

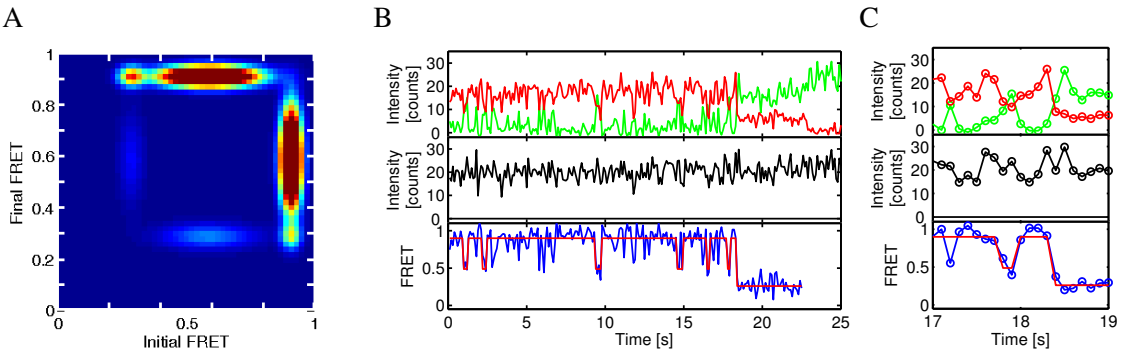


Figure 6.18: *Cycling of the inter-domain distance of Ssc1 in the presence of ATP and substrate. (A) Transition density plot for Ssc1(341,448) in the presence of ATP and substrate constructed from the transitions found by a global HMM analysis. (B) Typical trace of Ssc1(341,448) in the presence of ATP and substrate with donor (green) and acceptor (red) intensity (upper panel), total intensity (middle panel) and FRET efficiency (bottom panel). (C) Close-up of the transition to the 30% FRET state in (B).*

This hypothesis can be ruled out when the dwell time of the 50% FRET (∼250 ms) state is taken into account. The 90% FRET to 30% FRET transition was observed in 47 molecules with 100 ms time resolution. Assuming that the molecule needs to be in the 50% FRET state for more than 2 frames to observe the state, ∼45% of the transitions should show the step at 50% FRET. Thus, the data indicates with a high probability that Ssc1 does not go through the putative nucleotide-free state before substrate binding.

6.3.10 Interaction of Ssc1 with Tim44

Ssc1 is recruited to the translocation channel by Tim44, a peripheral membrane protein. This recruitment is essential for the function of Ssc1 as import motor of newly synthesized proteins into mitochondria. It is known that Ssc1 and Tim44 form nucleotide-dependent complexes in mitochondria (Rassow *et al.*, 1994; Schneider *et al.*, 1994; von Ahsen *et al.*, 1995; Voos *et al.*, 1996) which have been reconstituted from purified proteins (D'Silva *et al.*, 2004; Liu *et al.*, 2003; Slutsky-Leiderman *et al.*, 2007). SpFRET burst analysis experiments were performed to study the influence of Tim44 on the conformation of Ssc1. The spFRET efficiency histograms obtained using the domain sensor in the presence of ADP and Tim44 and in the presence of ADP, Tim44 and P5 are shown in Figure 6.19A-B.

In the presence of ADP only, a very heterogeneous distribution of inter-domain FRET efficiencies was observed (refer to Section 6.3.1). On the addition of Tim44, the FRET efficiency distribution is shifted to lower FRET efficiencies with a pronounced peak at approximately 20% FRET. The PDA fit gave an average distance of 7.8 nm ($\sigma = 1.0\,\mathrm{nm}$) for this peak. This inter-domain distance is similar to the distance observed in the substrate-bound state indicating that binding to Tim44 primes Ssc1 to a conformation suitable for substrate binding. However, this subpopulation accounted only for 49% of the molecules, and two additional distance distributions were necessary to describe the high FRET part of the histogram (Table 6.6). Additionally, the 2D histogram of the donor lifetime plotted against the FRET efficiency showed clear indications for the presence of dynamics between the low FRET subpopulation and the high-FRET subpopulations. Thus, although Tim44 binds stably to Ssc1 in the presence of ADP (Mapa, 2009), Ssc1 is cycling between a compact and two expanded conformations.

Peptide binding is known to displace Tim44 from Ssc1 (Mapa, 2009). Accordingly, addition of P5 to the Ssc1/Tim44 complexes in the presence of ADP resulted in a similar spFRET histogram as observed for substrate-bound Ssc1 in Section 6.3.2.

The results for the lid sensor are shown in Figure 6.19C-D. Tim44 binding results in an Ssc1 histogram consisting of one subpopulation similar to the substrate-bound state and species in a low-FRET ADP-like conformation. However, PDA analysis shows that in the Tim44-induced conformation (5.8 nm), the lid does not close as much as in the presence of substrate (5.2 nm). Addition of substrate results in displacement of Tim44, and the FRET efficiency observed in the lid sensor is shifted to approximately 70% similar to what was observed before for substrate-bound Ssc1 (Section 6.3.2). For the lid sensor, there were no dynamics detected in the donor lifetime – FRET efficiency histogram (Figure 6.20), indicating that interconversion between the two conformational states is much slower than the transition time of the molecules through the focus.

In (Mapa, 2009), it is shown that not only binding of substrate but also ATP can dissociate the Ssc1/Tim44/ADP complex. Here, spFRET experiments using non-hydrolyzable AMP-PNP were performed to find out if conformational changes are induced by AMP-PNP binding to the Ssc1/Tim44 complex. Figures 6.21A and C show the spFRET histograms for the domain and lid sensor in the presence of AMP-PNP and Tim44. The addition of Tim44 clearly changes the conformation of Ssc1 (compare histograms with AMP-PNP only in Figure 6.4), and the conformation also depends on the respective nucleotide present. However, visual inspection of the histograms in the presence of AMP-PNP (Figure 6.4C/D), in the presence of ADP and Tim44 (Figure 6.19A/C), and in the presence of AMP-PNP and Tim44 (Figure 6.21A-C) indicates that the conformation in the presence of AMP-PNP and Tim44 may consist of one fraction of Ssc1 in the conformation of the AMP-PNP-bound state and another fraction in the conformation measured in the presence of ADP and Tim44.

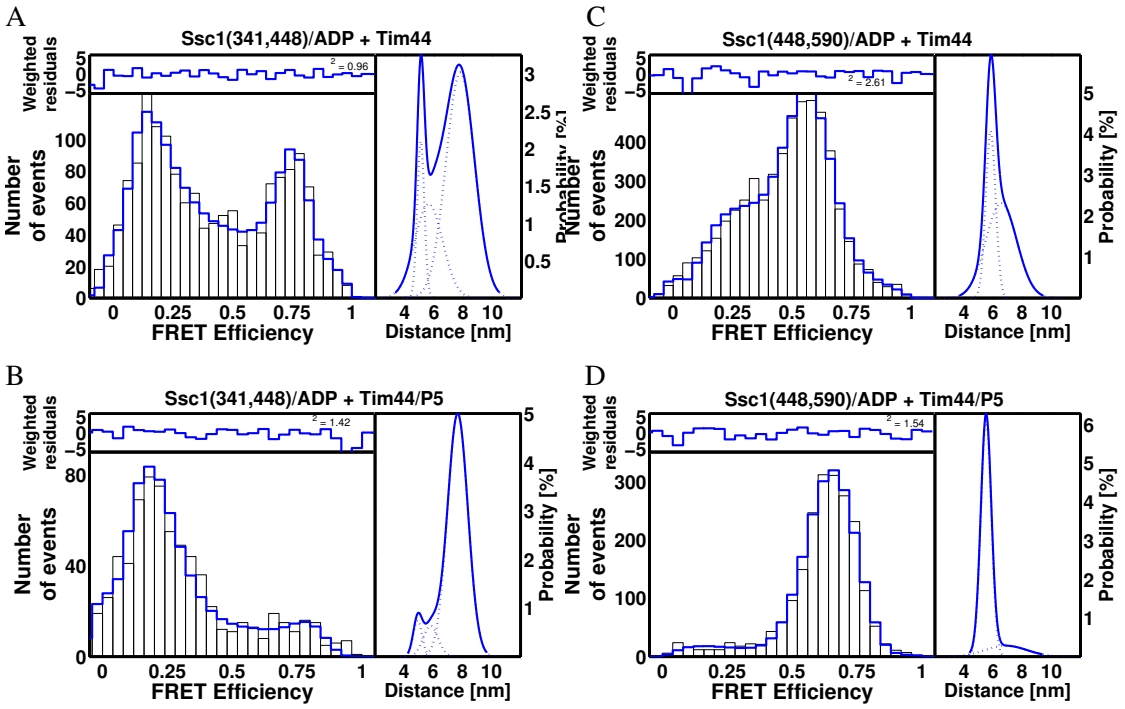
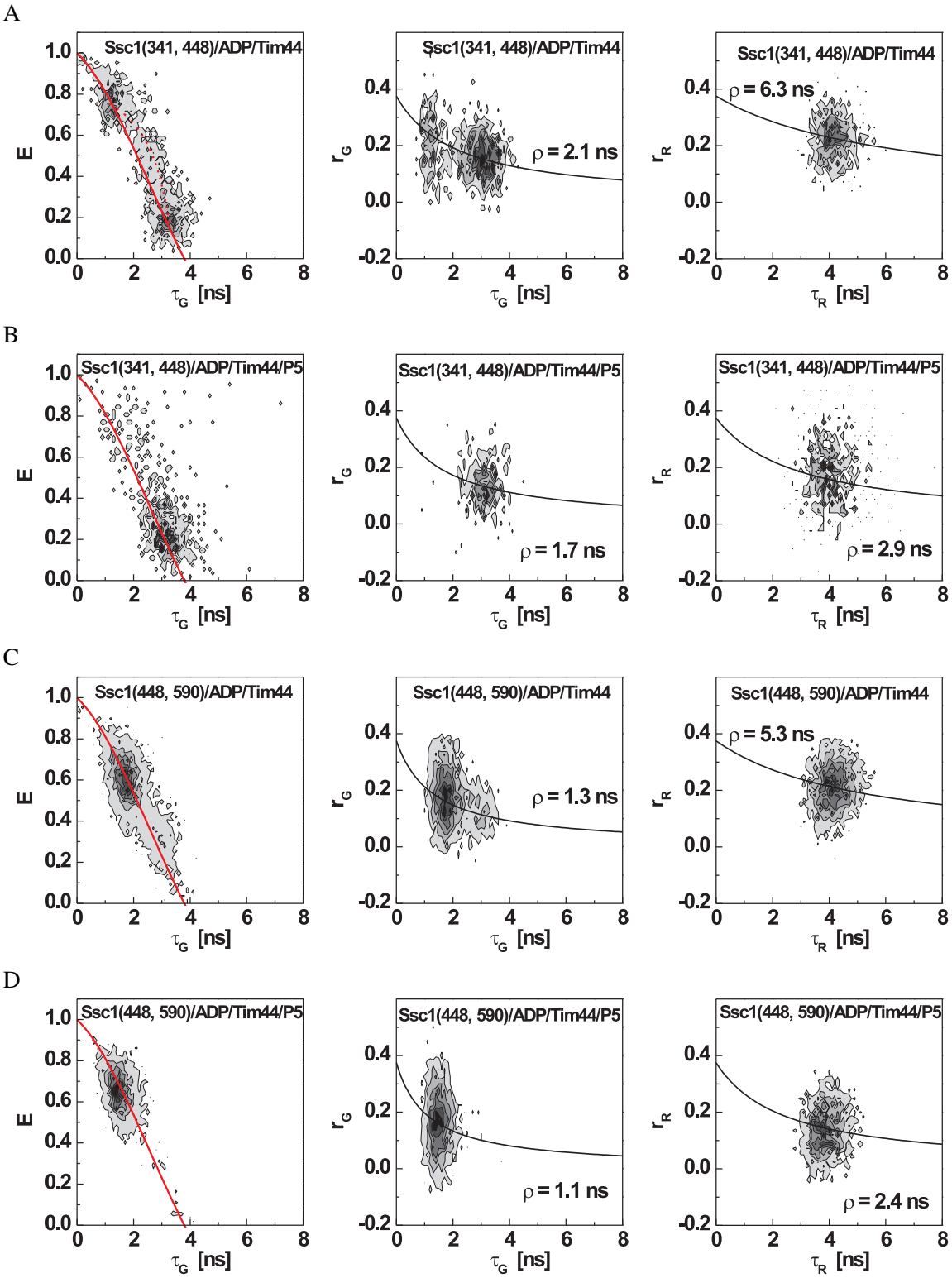


Figure 6.19: *SpFRET histograms with PDA fits (blue line) of Ssc1(341,448) in the presence of ADP and Tim44 (A, left panel) and ADP, Tim44, and P5(B, left panel), and of Ssc1(448,590) in the presence of ADP and Tim44 (C, left panel) and ADP, Tim44, and P5(D, left panel). Weighted residuals from the PDA fits are shown above the histograms, and the corresponding distance distributions determined using PDA are shown in the right panels.*

Figure 6.20: ***Following page:*** *Fluorescence lifetime and anisotropy histograms of Ssc1 in the presence of ADP, Tim44 and substrates. 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for Ssc1(341,448) in the presence of ADP and Tim44 (A) and ADP, Tim44, and P5 (B), and for Ssc1(448,590) in the presence of ADP and Tim44 (C) and ADP, Tim44, and P5 (D). (Left panels) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). For Ssc1(341,448)/ADP/Tim44, the data does not fit this relation but the relation for a dynamic sample (equation 6.1, dashed line). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*


A
Ssc1(341, 448)/ADP/Tim44
E
τ_G [ns]
Ssc1(341, 448)/ADP/Tim44
r_G
ρ = 2.1 ns
τ_G [ns]
Ssc1(341, 448)/ADP/Tim44
ρ = 6.3 ns
r_R
τ_R [ns]
B
Ssc1(341, 448)/ADP/Tim44/P5
E
τ_G [ns]
Ssc1(341, 448)/ADP/Tim44/P5
r_G
ρ = 1.7 ns
τ_G [ns]
Ssc1(341, 448)/ADP/Tim44/P5
ρ = 2.9 ns
r_R
τ_R [ns]
C
Ssc1(448, 590)/ADP/Tim44
E
τ_G [ns]
Ssc1(448, 590)/ADP/Tim44
r_G
ρ = 1.3 ns
τ_G [ns]
Ssc1(448, 590)/ADP/Tim44
ρ = 5.3 ns
r_R
τ_R [ns]
D
Ssc1(448, 590)/ADP/Tim44/P5
E
τ_G [ns]
Ssc1(448, 590)/ADP/Tim44/P5
r_G
ρ = 1.1 ns
τ_G [ns]
Ssc1(448, 590)/ADP/Tim44/P5
ρ = 2.4 ns
r_R
τ_R [ns]

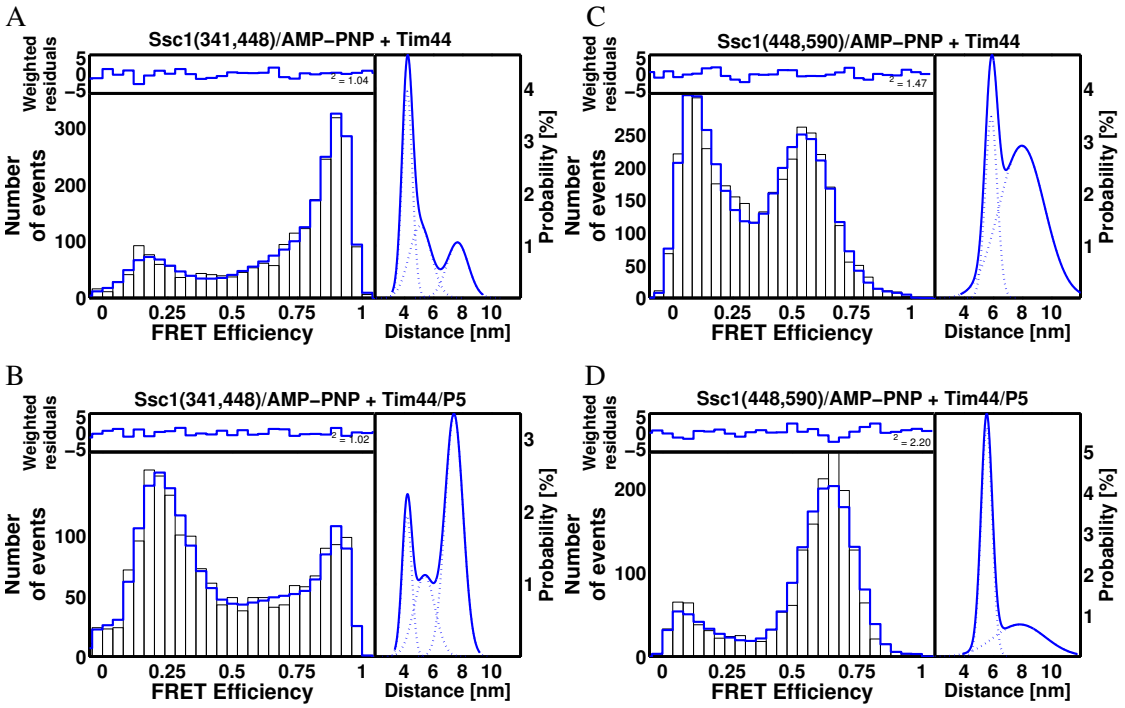
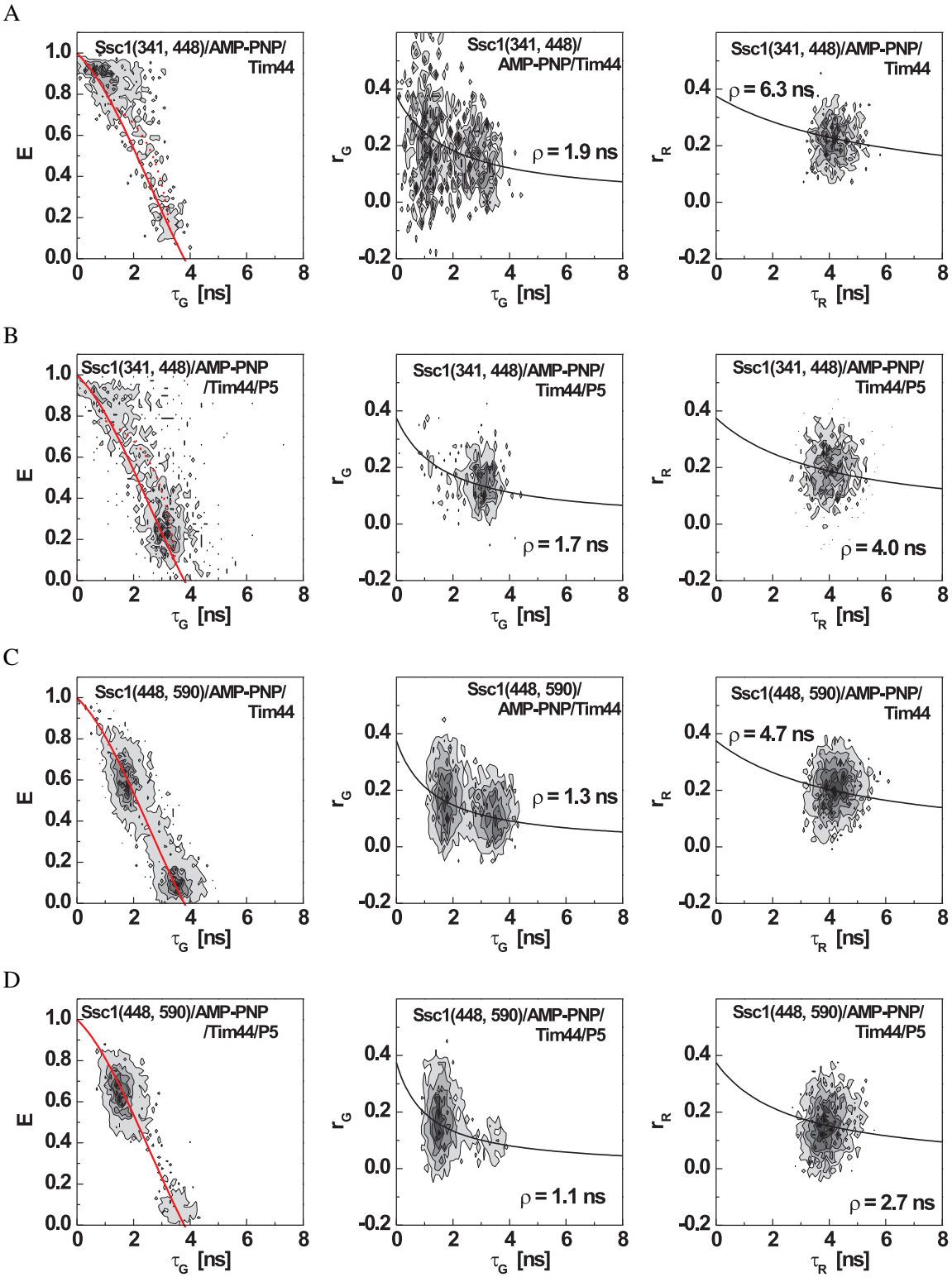


Figure 6.21: *SpFRET histograms with PDA fits (blue line) of Ssc1(341,448) in the presence of AMP-PNP and Tim44 (A, left panel) and AMP-PNP, Tim44, and P5(B, left panel), and of Ssc1(448,590) in the presence of AMP-PNP and Tim44 (C, left panel) and AMP-PNP, Tim44, and P5(D, left panel). Weighted residuals from the PDA fits are shown above the histograms, and the corresponding distance distributions determined using PDA are shown in the right panels.*

Figure 6.22: ***Following page:*** *Fluorescence lifetime and anisotropy histograms of Ssc1 in the presence of AMP-PNP, Tim44 and substrates. 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for Ssc1(341,448) in the presence of AMP-PNP and Tim44 (A) and AMP-PNP, Tim44, and P5 (B), and for Ssc1(448,590) in the presence of AMP-PNP and Tim44 (C) and AMP-PNP, Tim44, and P5 (D). (Left panels) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*


A
Ssc1(341, 448)/AMP-PNP/
Tim44
E
τG [ns]
Ssc1(341, 448)/
AMP-PNP/Tim44
ρ = 1.9 ns
rG
τG [ns]
Ssc1(341, 448)/AMP-PNP/
Tim44
ρ = 6.3 ns
rR
τR [ns]
B
Ssc1(341, 448)/AMP-PNP
/Tim44/P5
E
τG [ns]
Ssc1(341, 448)/AMP-PNP/
Tim44/P5
ρ = 1.7 ns
rG
τG [ns]
Ssc1(341, 448)/AMP-PNP/
Tim44/P5
ρ = 4.0 ns
rR
τR [ns]
C
Ssc1(448, 590)/AMP-PNP/
Tim44
E
τG [ns]
Ssc1(448, 590)/
AMP-PNP/Tim44
ρ = 1.3 ns
rG
τG [ns]
Ssc1(448, 590)/AMP-PNP/
Tim44
ρ = 4.7 ns
rR
τR [ns]
D
Ssc1(448, 590)/AMP-PNP
/Tim44/P5
E
τG [ns]
Ssc1(448, 590)/AMP-PNP/
Tim44/P5
ρ = 1.1 ns
rG
τG [ns]
Ssc1(448, 590)/AMP-PNP/
Tim44/P5
ρ = 2.7 ns
rR
τR [ns]

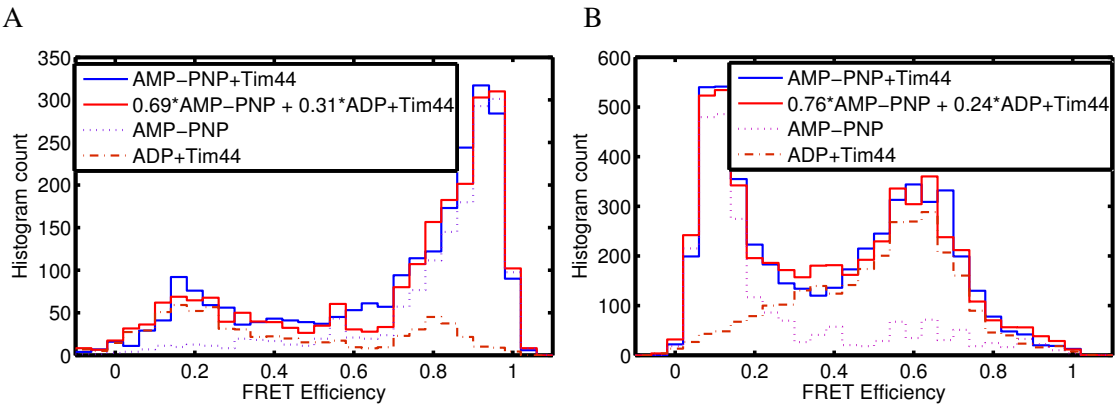


Figure 6.23: *The spFRET histogram of Ssc1(341,448) (A) and Ssc1(448,590) (B) in the presence of AMP-PNP and Tim44 (blue) is reproduced by a linear combination (solid red line) of the histograms in the presence of AMP-PNP and in the presence of ADP and Tim44 (dotted red lines).*

To clarify this, the histograms of the AMP-PNP measurements and the histograms of the ADP/Tim44 measurements were used to reconstruct the histograms of the measurements with AMP-PNP and Tim44 (Figure 6.23). For both sensors, the histograms in the presence of AMP-PNP and Tim44 can be reproduced accurately by a linear combination of the AMP-PNP and the ADP/Tim44 histograms. The relative amplitudes of the two components is similar for both sensors with the AMP-PNP histograms constituting approximately 75% of the AMP-PNP/Tim44 histograms. This is a good indication that only a fraction of 25% of the Ssc1 molecules have Tim44 bound at these conditions ($c_{Tim44} = 5\,\mu\mathrm{M}$ and $c_{Ssc1} = 25\,\mathrm{pM}$), in agreement with the lower binding affinity of Tim44 to Ssc1/AMP-PNP compared to Ssc1/ADP (Mapa, 2009). The fraction of Ssc1 that has Tim44 bound, however, has the same inter-domain and SBD conformation as the ADP/Tim44-bound Ssc1, at least to the extend measured by the spFRET sensors.

In the presence of AMP-PNP, not only the affinity of Tim44 to Ssc1 is lower than in the presence of ADP but also the affinity of substrate to Ssc1 (Mayer, 2010). This is indeed observed in the spFRET histograms presented in Figure 6.21 where P5 was added to the Ssc1/Tim44/AMP-PNP complexes. While the majority of the Ssc1 molecules now shows a FRET efficiency distribution typical for the substrate-bound state, there is an additional subpopulation for both the domain and lid sensor with the AMP-PNP conformation.

Table 6.6: *Distance distributions for Ssc1 in the presence of Tim44 and substrate*

	Subpopulation 1			Subpopulation 2			Subpopulation 3			D-only	
	%	d/nm	σ/nm	%	d/nm	σ/nm	%	d/nm	σ/nm	%	reduced χ^2
Ssc1(341,448)/ADP + Tim44	14.7	5.1	0.2	32.9	5.6	1.0	49	7.8	1.0	1.2	0.96
Ssc1(341,448)/ADP + Tim44/P5	7.3	4.9	0.3	12.2	5.8	0.5	72.7	7.7	0.7	7.7[1]	1.42
Ssc1(341,448)/AMP-PNP + Tim44	41.1	5.1	0.7	36.8	4.2	0.4	21.1	7.7	0.8	0.9	1.04
Ssc1(341,448)/AMP-PNP + Tim44/P5	30.3	5.4	0.7	13.9	4.2	0.3	52.9	7.4	0.6	2.9	1.02
Ssc1(448,590)/ADP + Tim44	57	6.6	1.1	41.1	5.8	0.3				1.9	2.61
Ssc1(448,590)/ADP + Tim44/P5	11.2	6.9	1.4	88.8	5.2	0.4				0	1.54
Ssc1(448,590)/AMP-PNP + Tim44	58.9	8.0	1.6	38.4	5.9	0.4				2.80	1.47
Ssc1(448,590)/AMP-PNP + Tim44/P5	24.8	7.8	1.8	75.4	5.6	0.4				0	2.20

[1] For most measurements, the amount of donor-only molecules in the PDA fit was 3% or less. The high amount of donor-only in the measurement with Ssc1(341,448)/ADP + Tim44/P5 is possibly due to a higher concentration than usual which results in inefficient filtering of donor-only molecules in the E-S-histogram.

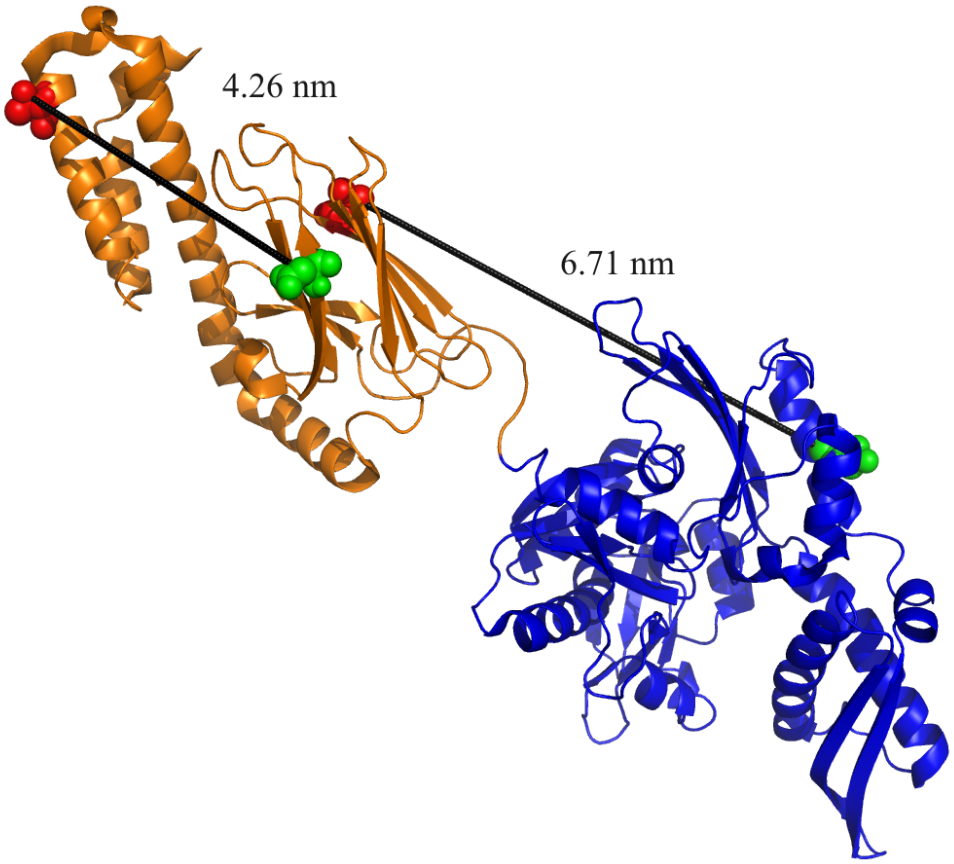


Figure 6.24: *The structure of DnaK with bound substrate (PDB 2KHO) visualized using PyMol in cartoon style. The residues where DnaK was labeled are shown as colored spheres together with the distances between each pair.*

6.3.11 Analysis of the conformational flexibility of DnaK

Although Ssc1 and DnaK, the major bacterial Hsp70 and the paradigm for Hsp70 research, are highly homologous, they cannot substitute for each other neither in *E. coli* nor in mitochondria. To understand the differences between these chaperones, spFRET experiments of the conformation of double-labeled DnaK in the presence of ATP and ADP were performed. The labeling positions of DnaK were chosen similar to the ones described above for Ssc1 (Figure 6.24). One sensor labeled at positions 318 and 425 was sensitive for the inter-domain distance, and a second sensor was used to follow the opening state of the substrate binding lid with labels at positions 458 and 563.

The ATP state of DnaK was similarly well defined as the ATP state of Ssc1. The spFRET histogram of the domain sensor was best fit with two gaussian distance distributions. In the major subpopulation (84% of the molecules) the two domains were closely docked with an average separation of 4.7 nm (width 0.3 nm). 15% of the molecules were in a more heterogenous conformation with more separated domains (6.8 nm, $\sigma = 1.6\,\mathrm{nm}$). The spFRET histogram of the lid sensor in the presence of ATP could be fit with a single gaussian distance distribution (8.7 nm, width 0.9 nm) (Figure 6.25 and Table 6.7).

In the presence of ADP, the SBD of DnaK adopted at least two distinct conformations (Figure 6.25D) with 7.9 nm and 5.9 nm donor-acceptor separation. However, in contrast to Ssc1 which showed a very

Table 6.7: *Distance distributions for DnaK*

	Major subpopulation			Minor subpopulation			D-only	
	%	d/nm	σ/nm	%	d/nm	σ/nm	%	reduced χ^2
DnaK(318,425)/ATP	84	4.7	0.3	15	6.8	1.6	1	1.51
DnaK(318,425)/ADP	89	8.4	0.9	8	5.3	0.5	3	1.15
DnaK(318,425)/ADP + P5	93	7.7	0.7	4.7	6.1	0.9	3.3	0.79
DnaK(458,563)/ATP	100	8.7	1.0	-	-	-	0	1.87
DnaK(458,563)/ADP	61	7.9	1.5	30	5.9	0.3	9	1.63
DnaK(458,563)/ADP + P5	86.1	5.4	0.3	11.1	6.6	1.2	2.7	1.55

heterogeneous FRET distribution and dynamic switching between different inter-domain separations, the domains of DnaK in the ADP state were more homogeneous ($\sigma = 0.9\,\mathrm{nm}$) and more separated (Figure 6.25C) with the majority of the molecules (89%) showing a donor-acceptor separation of 8.4 nm.

SpFRET measurements of substrate-bound Ssc1 showed an active influence of the substrate on the Ssc1 conformation. To compare this with DnaK, further spFRET experiments on DnaK in the presence of ADP and peptide were performed. Different than for Ssc1, substrate binding did not result in further undocking of the NBD and the SBD compared to the ADP state. On the contrary, upon substrate binding, the dominant subpopulation (93%) displayed a smaller donor-acceptor separation (7.7 nm, 0.7 nm width) than in the presence of ADP (8.4 nm, Table 6.7). This distance is roughly the same as predicted by the crystal structure (6.7 nm, Figure 6.24, Bertelsen *et al.*, 2009). For the DnaK(458,563) sensor, the spFRET histograms showed a dominant state similar in average distance and width to the minor subpopulation observed in the presence of ADP alone (5.8 nm, 0.3 nm width). This distance is significantly higher than expected from the crystal structure. According to the structure determined by Bertelsen *et al.* (2009) using solution NMR experiments, the residues 458 and 563 are 4.3 nm apart. This deviation is outside the error margins of the Förster radius (Table 6.1). A possible reason for the difference may be the different sequence of the substrate peptide.

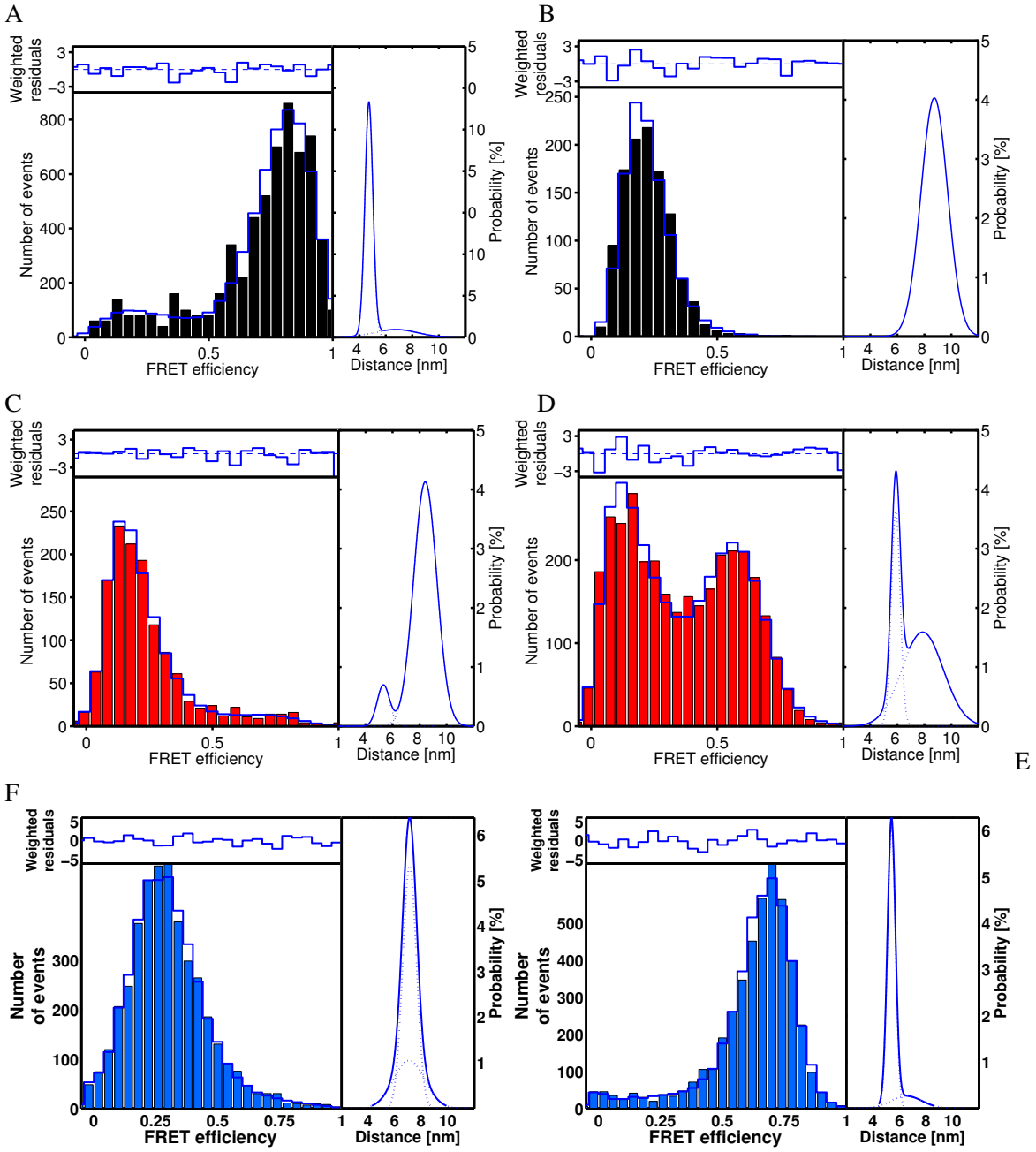


Figure 6.25: *SpFRET histograms of DnaK. (Left panels) SpFRET histograms with PDA fits (blue line) of DnaK (318,425) in the presence of ATP (black, A), ADP (red, C), and ADP and P5 (blue, E), and of DnaK (458,563) in the presence of ATP (black, B), ADP (red, D), and ADP and P5 (blue, F). Weighted residuals from the PDA fits are shown above the histograms, and the corresponding distance distributions determined using PDA are shown in the right panels.*

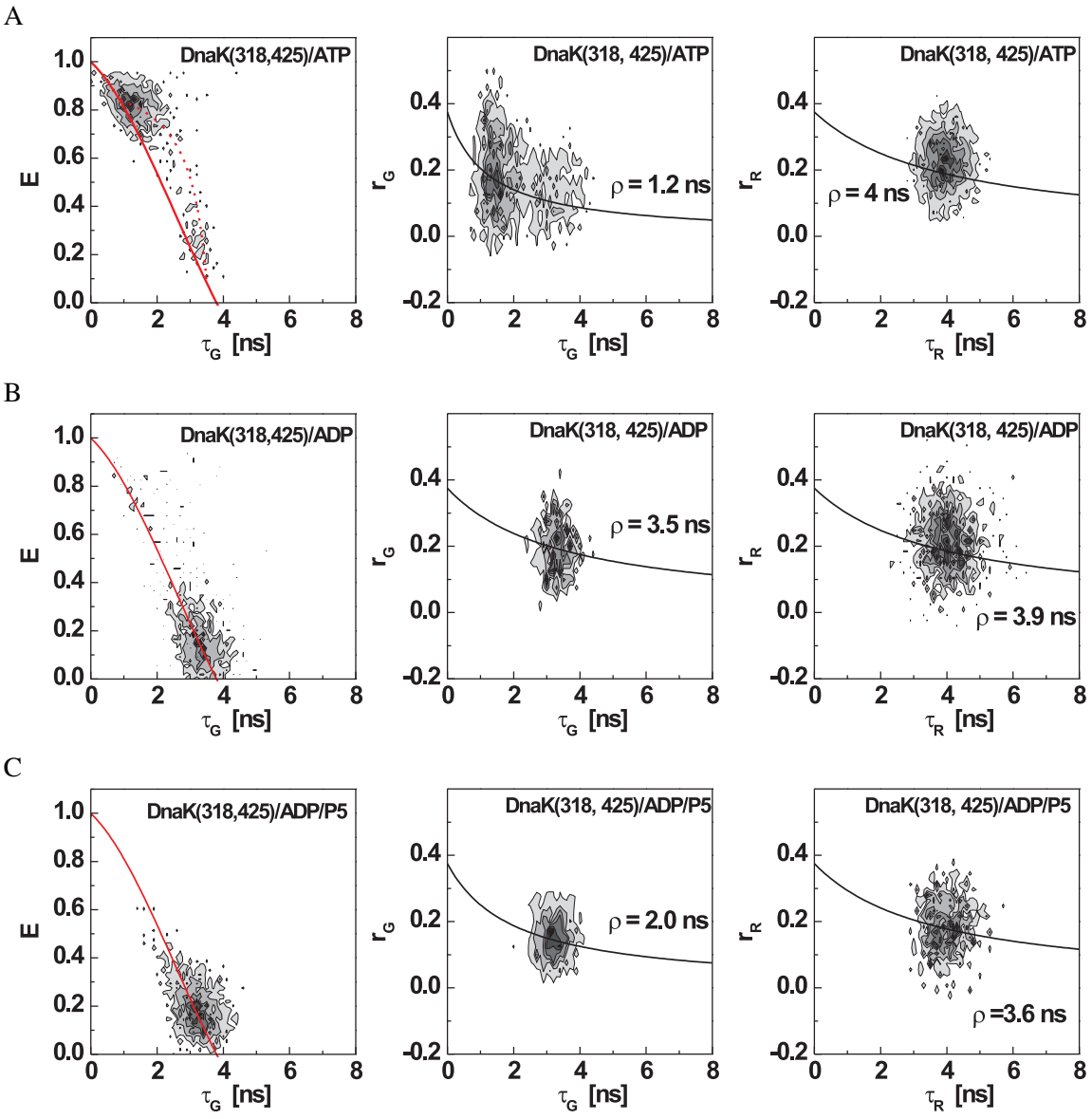


Figure 6.26: *Fluorescence lifetime and anisotropy histograms of DnaK(318,425). 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for DnaK(318,425) in the presence of ATP (A), ADP (B), and ADP and P5 (C). (left panel) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*

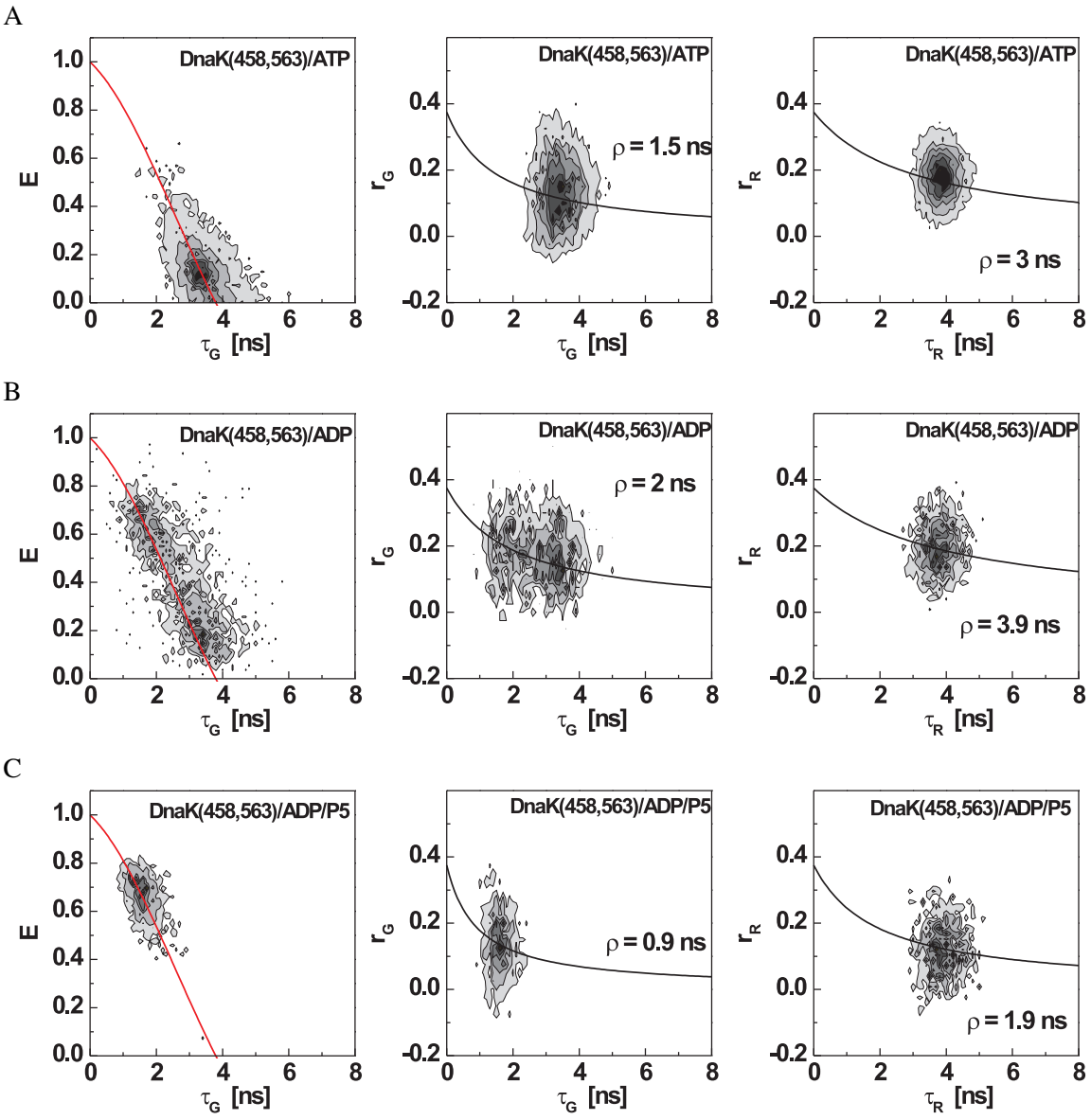


Figure 6.27: *Fluorescence lifetime and anisotropy histograms of DnaK(458,563). 2D histograms of the FRET efficiency E versus donor lifetime (left panels), donor anisotropy versus donor lifetime (middle panels), and acceptor anisotropy versus acceptor lifetime (right panels) for DnaK(458,563) in the presence of ATP (A), ADP (B), and ADP and P5 (C). (left panel) The solid line in the lifetime-E histograms describes the relation between donor lifetime and E expected for a static sample (equation 4.22). (Middle and right panels) The solid line is the Perrin equation, the theoretical relationship between the fluorescence lifetime and anisotropy, for a sample with one rotational mobility species. The rotational correlation time ρ is given in the graphs.*

6.4 Discussion

In this chapter, the conformation of the major mitochondrial Hsp70, Ssc1, was studied in different phases of its functional cycle using ensemble and spFRET. To cover timescales of dynamic changes between microseconds and seconds, MFD-PIE burst analysis experiments in solution and single-molecule TIRF experiments were combined. The results are summarized in Figure 6.28. The ATP-bound state of Ssc1 is well defined with an open lid and docked domains. These results on the ATP-bound state of Ssc1 are in agreement with the data obtained with DnaK in this thesis and with other Hsp70 chaperones Buchberger *et al.* (1995); Swain *et al.* (2007); Marcinowski *et al.* (2011). The tight docking of the NBD and SBD and the opening of the substrate binding lid in the presence of ATP thus seems to be a general feature of Hsp70 chaperones. The inter-domain distance derived from the MFD-PIE data using PDA analysis for Ssc1 (4.2 nm) is in good agreement with the distance between the respective residues in the crystal structure of Sse1 (4.1 nm). For the respective distance between residues 318 and 425 in DnaK, we estimated 4.7 nm from the MFD-PIE data. Interestingly, the calculated distance of 7.7 nm for the residues 448 and 590 in the base and the lid of SBD of Ssc1 is smaller than the distance in the crystal structure of Sse1 (8.5 nm) and in DnaK (8.7 nm). These differences reflect that different Hsp70 chaperones and the closely related Hsp110s share similarities among each other, but differ in other aspects depending on the specialization to different cellular demands.

In contrast to the clearly defined ATP state, in the presence of ADP, Ssc1 is heterogeneous. Observation on longer timescales allowed the discrimination of two distinct states populated in the presence of ADP. The opening state of the lid fluctuates between two conformations on a timescale of several seconds. In light of a recent study by Schlecht *et al.* (2011) who found three coexisting conformations of the SBD in the presence of ADP using EPR experiments, the non-exponential kinetics of the lid-closing transition are explained by the presence of a third conformation with a similar distance between residues 448 and 590 and thus indistinguishable using the lid sensor.

The inter-domain distance fluctuates between a state with similar close contact between NBD and SBD as in the presence of ATP, and a more separated conformation. The timescale of these fluctuations is approximately 5 times faster than the fluctuations observed in the SBD. This is a good indication that although, in the ATP and substrate-bound states. strong allosteric coupling exists between the SBD and the NBD, the conformations of NBD and SBD are uncoupled in the presence of ADP. An additional indication for conformational uncoupling is the ADP dependence of the association rate of the domains but no ADP dependence of the lid dynamics. The ADP dependence of the association rate of the domains suggests the interpretation of the state with more separated domains as a nucleotide-free state. This notion is supported by the observation that the putative nucleotide-free state is detected both in the presence of ATP and ADP and that the presence of the nucleotide exchange factor Mge1 prohibits its detection in the presence of ATP by increasing the rate of ATP binding.

Diverging descriptions of the ADP states of different Hsp70 chaperones exist in the literature. For mammalian Hsc70, extensive inter-domain interactions were suggested (Jiang *et al.*, 2007, 2005). However, in this study and in earlier reports, the NBD and SBD of DnaK were found to behave independently in its ADP-bound form (Bertelsen *et al.*, 2009; Swain *et al.*, 2007). Thus, the ADP-bound state of Ssc1 appears more similar to Hsp70s of higher eukaryotes than to DnaK.

Taking into account the similarities of the SBD conformation in different Hsp70s as well as Hsp110s in the presence of both ADP and ATP, and the variations of the domain interactions observed at least in DnaK and Ssc1 here, it appears that the tuning of Hsp70s for specific functions has affected

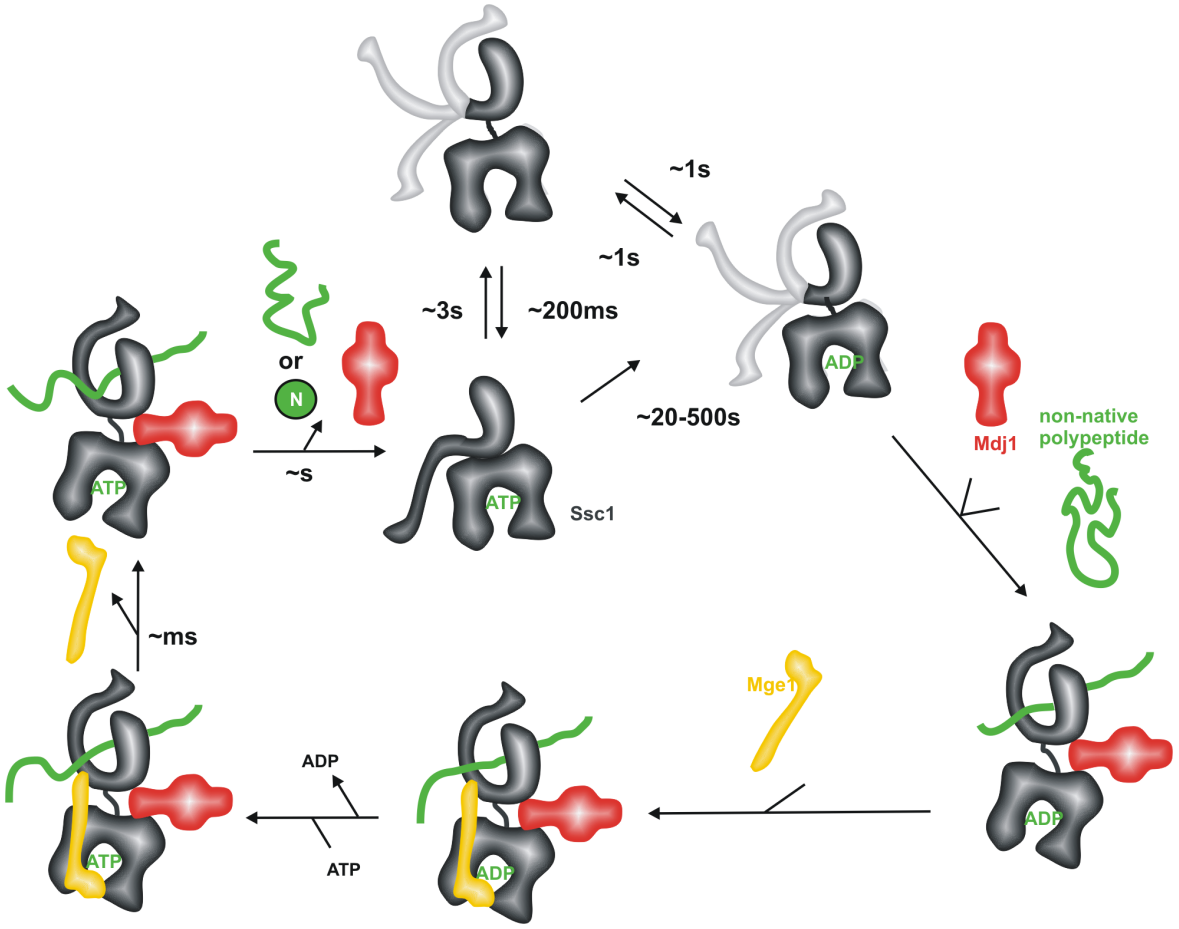


Figure 6.28: *Model of the folding cycle of Ssc1. Ssc1 is a compact molecule with an open SBD and the two domains docked when ATP is bound. Spontaneous hydrolysis slowly converts the ATP to ADP. At the same time, ATP and ADP can be released and rebound with a significantly higher rate. The ADP state has a slightly higher inter-domain distance than the ATP state and a flexible lid. In the nucleotide-free state, the lid is similarly flexible as in the ADP state, but the domains are more separated. The J protein Mdj1 stimulates binding of a substrate to the ATP state of Ssc1 by increasing the hydrolysis rate of ATP. The ADP state of Ssc1 with the bound substrate is characterized by the domains being completely undocked and the SBD lid closed. Binding of Mge1 increases the rate of ATP rebinding. In the presence of ATP, Mge1 will be released from the complex within milliseconds, followed by conformational changes of Ssc1, which induce release of Mdj1 and the substrate from the chaperone, allowing a new cycle to start.*

mainly the inter-domain communication but not the SBD. Substrate binding is a general feature of all Hsp70s, thus conservation of the SBD structure may be highly desirable. On the other hand, triggering of substrate release and/or release of the Hsp70 from co-chaperones may depend strongly on the specific function. Thus, the inter-domain communication may have been modified to meet each demand. One possible explanation for the observed differences between Ssc1 and DnaK is the involvement of Ssc1 in the translocation process in addition to the chaperone activity. During translocation, binding of the substrate to Ssc1 needs to be communicated to the translocation channel in order to ensure the release of Ssc1 only after substrate has bound. Substrate binding and release need to be coordinated to allow the generation of an effective inward motion of the substrate. For the chaperone activity, other timing may be required and thus another trigger mechanism may be used.

Due to the difficulty in obtaining Hsp70s in ADP-bound but substrate-free form, the role of substrates in the conformational cycle of Hsp70s has remained unclear (Swain and Gierasch, 2006). The data presented here demonstrate an active role of substrates in determining the conformation of both Ssc1 and DnaK. For Ssc1, the conformation of the ADP-bound state was found to be very different from the substrate-bound state. In contrast, according to our data with DnaK, ADP alone is sufficient to undock the domains and substrate binding results in a slightly more compact conformation. The SBD of both Ssc1 and DnaK was flexible in the presence of ADP and binding of the substrate is needed for its closure.

Additionally, the conformational changes of Ssc1 upon interaction with several co-chaperones were studied. The J protein partner of Ssc1 in the folding cycle, Mdj1, bound to the ATP form of Ssc1 and stimulated ATP hydrolysis as expected. However, study of the interaction kinetics using FRET between labeled Mdj1 and Ssc1 revealed a rather unexpected behavior. Mdj1 remained bound to Ssc1 for approximately 5 min, significantly longer than the folding cycle (17 s based on the rate of hydrolysis in the presence of Mdj1 and P5). This agrees well with the observation that the Mdj1-Ssc1 complex was recovered from mitochondria solubilized in the presence of ATP (Horst *et al.*, 1997). In contrast, other J proteins, such as DnaJ and Sec63, were proposed to bind more transiently to their respective Hsp70 partners (Bukau, 1998; Jiang *et al.*, 2007). Stable binding was only observed when the J proteins seemed to act as substrates for the Hsp70s (Misselwitz *et al.*, 1999). As shown by the spFRET experiments here, Mdj1 does not bind to Ssc1 in a substrate-like manner.

Prolonged binding of the J protein to Ssc1 is advantageous in the context of protein folding in the mitochondrial matrix. Mdj1 is released from Ssc1 within a second in the presence of ATP and Mge1, allowing re-initiation of the folding cycle. However, during translocation of proteins into mitochondria by the TIMM23 complex, prolonged interactions would slow down the import process significantly, and quick release of the substrate by Mge1 and ATP would prevent effective movement of the substrate into the mitochondrial matrix. This discrepancy may have been solved in mitochondria by the involvement of at least two additional co-chaperones in the translocation process, Tim44 and Tim16. Study of the conformation of Ssc1 in the presence of Tim44 revealed that Tim44 not only recruits Ssc1 to the translocation pore but also actively remodels the conformation of Ssc1 to prime it for substrate binding. However, several questions regarding the translocation process are still open, especially the role of Tim16, and the tuning of the interaction time between Ssc1 and its J protein partner Tim14. Ultimately, it would be desirable to elucidate the precise mechanism of preprotein import into mitochondria, i.e. which model, the Brownian ratchet model, the power stroke model, or the entropic pulling model describes the process best. Clearly, single-molecule studies are able to contribute significantly to the study of this system, and will help to finally get a complete picture not only of the folding cycle presented here but also of the translocation cycle.

7 Rescue of a kinetically trapped folding intermediate by GroEL

The results of this chapter have been published previously in (Chakraborty *et al.*, 2010).

7.1 Introduction

7.1.1 Overview of the function of GroEL

GroEL is a bacterial chaperone that assists in folding of a large number of substrates in an ATP-dependent manner. It was detected first in *e. coli* mutants specifically blocking assembly of bacteriophage λ heads (Georgopoulos *et al.*, 1973). Hemmingsen and Ellis (1986) found a protein of similar tetradecameric ring structure to interact with Rubisco during Rubisco synthesis but not with the final protein. The association of GroEL with protein folding was made a few years later (Laminet *et al.*, 1990; Martin *et al.*, 1991).

GroEL is a barrel-shaped $\sim$800 kDa tetradecameric protein complex with 57 kDa subunits (Figure 7.1A, Xu *et al.*, 1997). Each subunit consists of an apical domain (A), an intermediate domain (I1, I2), and an equatorial domain (E1,E2). The domains are ordered in the sequence as E1-I1-A-I2-E2. Seven subunits form a ring structure and two of these rings are stacked back to back with their equatorial domains. In the absence of ATP and GroES, the apical domains expose hydrophobic surfaces at the opening of the cavity formed inside the ring (Figure 7.1B and 7.2). These hydrophobic regions catch an unfolded protein in one ring (*cis* ring). ATP binding to the equatorial domains of the *cis* ring triggers an allosteric mechanism that results in dramatic changes in the conformation of the GroEL subunits (Figure 7.1C) increasing the affinity of the apical domain to the co-chaperone GroES. Upon GroES binding, the substrate is encapsulated inside the cavity of the *cis* ring. The inside surface of the cavity becomes hydrophilic due to the conformational change in the GroEL subunits, thereby promoting folding of the substrate (Weissman *et al.*, 1996; Mayhew *et al.*, 1996). The substrate is enclosed inside the GroEL cavity during the time required for the hydrolysis of the seven ATP molecules in the *cis* ring. GroES and ADP release from the *cis* ring is triggered by binding of seven ATP molecules to the opposite (*trans*) ring. The substrate is now free to leave the GroEL cavity. At this point, the two rings exchange their roles with the former *cis* ring now becoming the *trans* ring and vice versa, and the cycle begins again. Successfully folded protein will not be recaptured by the hydrophobic patches of the nucleotide free *trans* ring, but proteins still unfolded can be recaptured quickly for another folding cycle. The functional cycle of GroEL is sketched in Figure 7.2.

GroEL assists folding of approximately 10 % of the proteins in *E. coli*. Proteins requiring GroEL for proper folding are usually between 20 and 60 kDa [1] in size and have complex α/β or $\alpha + \beta$

[1] The cavity can hold proteins up to 60 kDa, however, there is speculation that GroEL assists folding of bigger substrates by encapsulation of individual domains (Chaudhuri *et al.*, 2001).

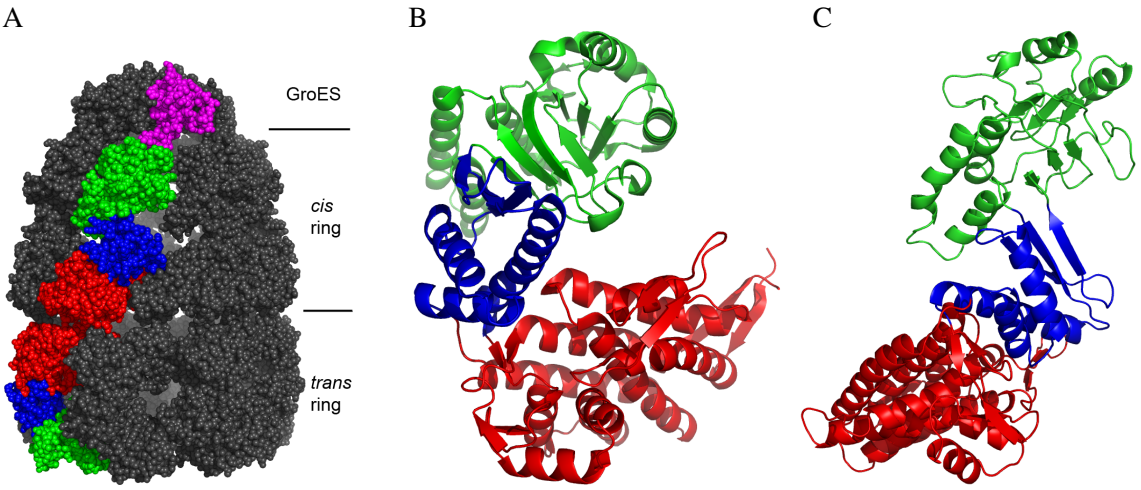


Figure 7.1: *Structure of the GroEL-GroES complex. (A) Model from the crystal structure determined by Xu et al. (1997) (Protein Data Bank (PDB): 1AON). GroEL is composed of two cylindrical rings each composed of seven subunits. The ring capped by GroES is called the* cis *ring, the opposite ring the* trans *ring. One subunit in each ring is shown in color with the apical domain in green, the intermediate domain in blue and the equatorial domain in red. One subunit of GroES is shown in magenta. (B) Cartoon diagram of a nucleotide free subunit in the* trans *ring using the same color coding for the domains. (C) Cartoon diagram of subunit with bound ATPs in the* cis *ring using the same color coding for the domains.*

multi-domain topologies (Ewalt *et al.*, 1997; Houry *et al.*, 1999). These proteins have large stretches of hydrophobic amino acids that are buried in the core of the folded protein. Exhibition of these hydrophobic patches in unfolded states makes these proteins highly prone to aggregation. Bacterial RuBisCo (ribulose-1,5-bisphosphate carboxylase-oxygenase) is a typical stringent substrate of GroEL sharing this domain topology. RuBisCo is known to fold efficiently only within the GroEL-GroES cage (Brinker *et al.*, 2001).

7.1.2 Previous single-molecule studies of GroEL

GroEL is the prototype chaperone system and has received much attention over the last two decades. The development of single-molecule fluorescence methods happened during the same period of time. Consequently, emerging single-molecule techniques were often used to study the function of GroEL.

The first single-molecule fluorescence observation of GroEL interactions with a substrate was already in 1999 (Yamasaki *et al.*, 1999). There, GroEL was labeled specifically by the introduction of a Cysteine residue at position 133, and the interaction with labeled substrate was followed by single-molecule TIRF microscopy. This basic concept remained the same for many of the following single-molecule studies of GroEL, although technological advances made much more detailed studies possible. In 2001, Taguchi *et al.* used a similar approach to study the interaction of immobilized labeled GroEL with labeled GroES. Most of their findings were in agreement with previous studies using ensemble methods. However, the additional information only accessible in single-molecule experiments allowed them the observation of single-exponential binding kinetics but double-exponential release kinetics in the presence of various unfolded substrates (Taguchi *et al.*, 2001). They attributed these kinetics to the existence of a kinetic intermediate of the GroEL/GroES/substrate complex before dissociation. Similar double-exponential release kinetics were already observed by Viani *et al.*

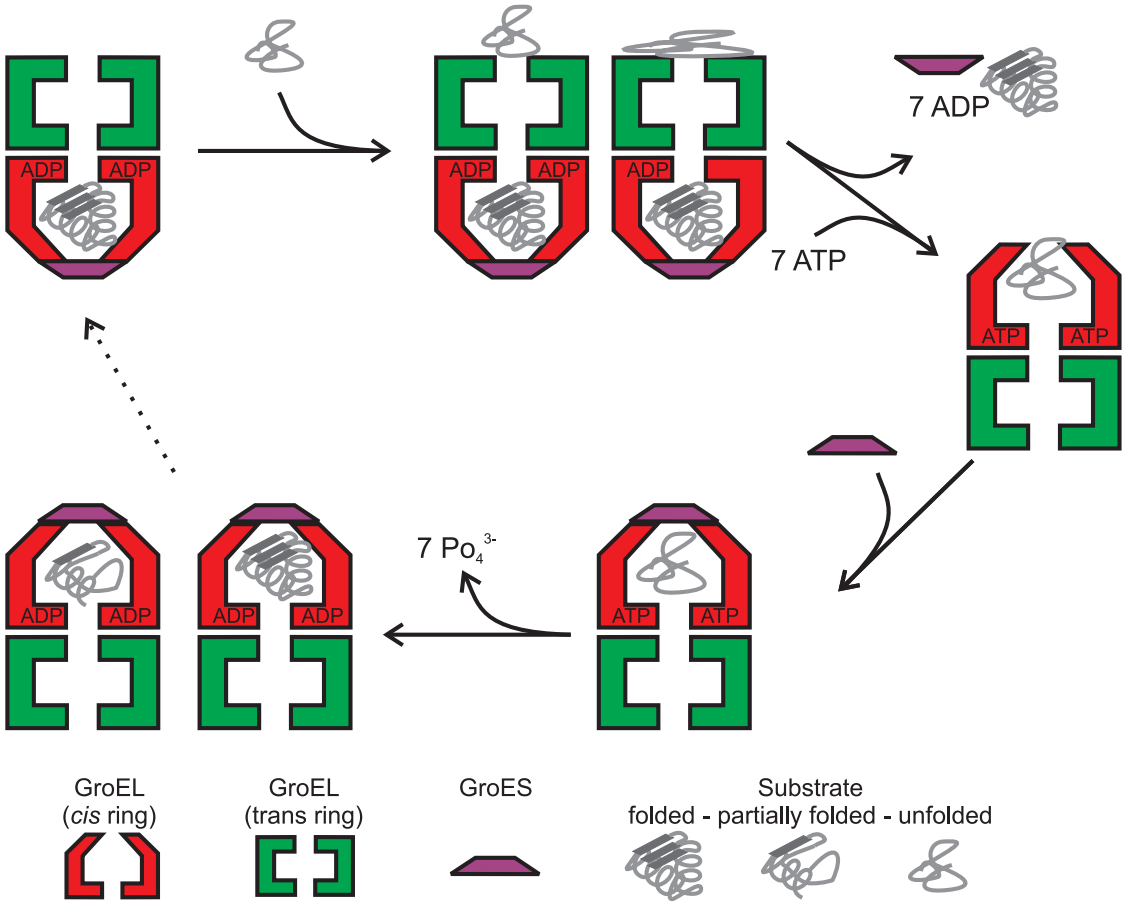


Figure 7.2: *The functional cycle of GroEL. In the upper left corner, the upper ring of the GroEL molecule is currently in the* trans *state (green). The lower ring is currently bound by GroES, encapsulating an already folded substrate (*cis *state, red). An unfolded substrate binds to the* trans *ring and assumes either a compact or an expanded conformation. Upon ATP binding to the* trans *ring, GroES, ADP and substrate are released from the* cis *ring. Binding of GroES completes the change of roles, making the upper ring the* cis *ring and the lower ring the* trans *ring. The substrate is free to fold during the ATP hydrolysis, resulting in a folded (right) or partially folded (left) substrate. Now, the cycle starts again either with a new substrate or by re-capture of the partially folded substrate.*

(2000), who used atomic force microscopy to measure the height of single GroEL/GroES complexes on a mica surface, allowing to detect the presence or absence of a GroES molecule on top of GroEL.

This two-timer mechanism of GroEL mediated folding was again observed in a later study where GFP was used as a substrate and the folding of GFP was monitored. Ueno *et al.* (2004) followed the folding of GFP assisted by immobilized GroEL and were able to quantify the dwell times of the two successive steps of GroEL-assisted folding. The first step with a characteristic time of $\sim$3 s comprises caging of substrate into the GroEL cavity followed by folding during a second step of $\sim$5 s. The duration of the caging step was later found to depend on the presence of the flexible C-terminal region of GroEL (Suzuki *et al.*, 2008). To exclude that the observed two-step behavior is an artifact of the low concentrations necessary for TIRF microscopy, Miyake *et al.* (2008) used zero-mode waveguides to decrease the detection volume compared to TIRF. This allowed the observation of the interaction between labeled GroEL and GroES at micromolar concentrations confirming the existence of two-step kinetics at intracellular concentrations.

Frank *et al.* approached a different aspect of the cycling of GroEL. Saturating ATP concentrations should in theory fully shift the equilibrium toward the conformation without exposed hydrophobic surfaces, which would prevent substrate binding. They designed an optical switch to follow the conformation of GroEL (Frank *et al.*, 2008) and used it to monitor conformational cycling using fluorescence correlation spectroscopy (FCS) and singular value decomposition (Frank *et al.*, 2010). GroEL is cycling even in the presence of saturating amounts of ATP with $\sim$50% of the molecules in the different conformational states of the subunits. This is explained by slow release of ADP from GroEL on the same timescale as hydrolysis, $\sim$10 s.

The formation of symmetric complexes of one GroEL bound to two GroES molecules (football-shaped complex) under conditions of high ADP concentration and high unfolded substrate concentration in addition to the regularly observed asymmetric complex with a 1:1 stoichiometry (bullet-shaped complex) was proposed by various authors (compare references 6-14 in Sameshima *et al.*, 2010). Recently, Sameshima *et al.* (2010) observed the football-shaped complex and its decay using single-molecule fluorescence in zero-mode waveguides showing that the order of GroES release from the two sides of GroEL is not determined by the binding order but rather is a probabilistic process purely based on stochastic probabilities of release from the two sides.

Instead of labeling GroEL and GroES or substrate to monitor interactions, our group focused on spFRET measurements of the conformation of different substrates during GroEL-assisted folding. Sharma *et al.* (2008) used a double-mutant of maltose binding protein (DM-MBP) known to fold efficiently but more slowly in the absence of GroEL. In the presence of GroEL and GroES, folding of DM-MBP speeds up by a factor of $\sim$7. By choosing several different labeling positions on the substrate, various distances were monitored. The starting point of the chaperone-assisted folding pathway is the unfolded state showing large separation between donor and acceptor dye for all observed distances. After incubation with GroEL, however, two distinct conformational states of the substrate were observed, an expanded, unfolded-like state and a more condensed state. Addition of ATP results in transient stretching of the substrate and disappearing of the unfolded-like conformation. Finally, addition of GroES allows successful folding resulting in the native state with high FRET efficiencies between all monitored distances.

Hillger *et al.* (2008) did similar experiments using labeled rhodanese as a substrate. In contrast to Sharma *et al.* (2008), they did not observe an expanded subpopulation of GroEL-bound rhodanese possibly implying a substrate-dependent process of GroEL action. Analysis of the time course of the FRET efficiency between different labeling positions in rhodanese in the initial 5000 s of spontaneous and GroEL-assisted folding by singular value decomposition allowed the extraction of the folding

rates of individual subdomains (Hofmann *et al.*, 2010). Spontaneous folding of the C-terminal domain is approximately six times faster than folding of the N-terminal domain and contact formation between the domains. This order is maintained in chaperonin-assisted folding. However, the folding rate of the C-terminal domain is decelerated depending on temperature by a factor of two to eight inside the GroEL cavity. This decrease can be explained by confinement effects and by interactions of the substrate with the cavity walls that become rate-limiting for the fast-folding C-terminal domain but not for the slower-folding N-terminal domain

Kim *et al.* (2010) used another substrate, the Von Hippel-Lindau tumor suppressor protein VHL, to follow the conformation of substrate during GroEL-assisted folding over time with TIRF microscopy. The hydrophobic core of this substrate is on average compacted by GroES binding while the hydrophilic N- and C-terminal ends are expanded. Time-resolved TIRF measurements showed that this compaction/expansion is of probabilistic nature since only 84% of the observed transitions in the hydrophobic core after ATP/GroES addition are actually compactions and only 69% of the transitions in the hydrophilic regions were expansions.

Single-molecule experiments have contributed significantly to the understanding of the mechanism of GroEL-assisted protein folding. However, there are still numerous open questions. In particular, it is not yet clear if and to what extend GroEL plays an active role beyond providing an isolated environment for protein folding, and which features make a specific protein dependent on GroEL for proper folding.

7.1.3 Mechanisms of GroEL-assisted protein folding

The bacterial chaperonin GroEL is essential for proper folding of a large subset of proteins (Section 7.1.1). It provides a compartment for single protein molecules to fold undisturbed by outer influences (Mayhew *et al.*, 1996; Weissman *et al.*, 1996; Brinker *et al.*, 2001). The exact mechanism by which GroEL promotes effective folding, however, is still subject of discussion. Several recent publications have suggested that GroEL facilitates folding purely by the prevention of aggregation without an additional active mechanism (Apetri and Horwich, 2008; Tyagi *et al.*, 2011).

Previous single-molecule studies showed conformational expansion of DM-MBP, a spontaneously slow-folding double-mutant form of Maltose Binding Protein, on encapsulation in the GroEL cavity. Although this expansion is not critical for accelerated folding (Brinker *et al.*, 2001; Tang *et al.*, 2006; Sharma *et al.*, 2008; Tang *et al.*, 2008), it supports an active role of GroEL in the folding process.

DM-MBP is an ideal model substrate to investigate whether GroEL/ES improves the folding properties of a protein by aggregation prevention or by an additional active mechanism. It folds spontaneously with high yield but a slow rate ($t_{1/2} \approx 20\,\mathrm{min}$ at $25\ ^\circ\mathrm{C}$). Its folding rate is increased 10 fold in the presence of GroEL/ES (Tang *et al.*, 2006). This could reflect an active role of GroEL in catalyzing DM-MBP folding. Alternatively, GroEL may accelerate folding in a passive manner by preventing the formation of reversible aggregates that would reduce the folding rate but not the yield (Apetri and Horwich, 2008).

In this chapter, I used FCS and FCCS to show that the increased folding rate of DM-MBP is not due to the prevention of aggregation. Unfolding and refolding traces measured by ensemble Tryptophan fluorescence and circular dichroism revealed the presence of a folding intermediate significantly populated at 0.5 M GuHCl (Chakraborty *et al.*, 2010). I used spFRET and PDA to characterize the structure of this folding intermediate. The results have been published together with H/D exchange experiments and measurements of DM-MBP folding with additional constrains induced by disulfide bridges in (Chakraborty *et al.*, 2010).

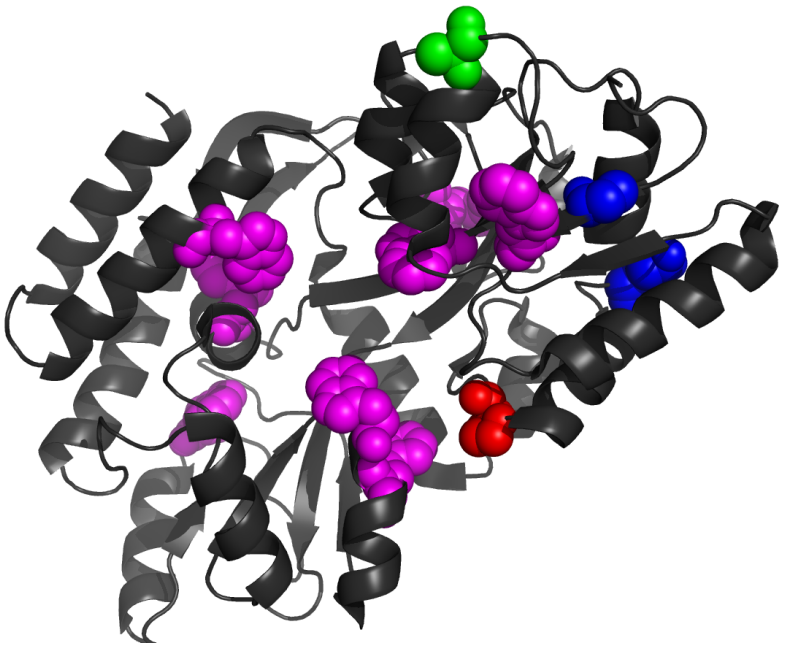


Figure 7.3: *The structure of Maltose binding protein (PDB ID 1OMP) visualized by PyMol. The labeling sites are indicated as green and red spheres. Green: Position 52 labeled with Atto 532. Red: Position 298 labeled with Atto 647N. The mutations V8G and Y283D are indicated by blue spheres. The positions of the intrinsic Tryptophan residues are indicated as purple spheres.*

7.2 Experimental procedures

7.2.1 Protein expression, purification, and labeling

DM-MBP (Chun *et al.*, 1993) was created in the lab of Ulrich Hartl by mutation of Valine to Glycine at position 8 (V8G) and Tyrosine to Aspartic acid at position 283 (Y283D). Additional Cysteines were introduced at positions 52, substituting Alanine (A52C), and 298, substituting Proline (P298C), for specific labeling of DM-MBP with fluorophores using maleimide chemistry. This construct, GroEL and GroES were expressed and purified by our collaborators in the Hartl Lab as described in (Brinker *et al.*, 2001; Tang *et al.*, 2006; Sharma *et al.*, 2008).

DM-MBP labeling was also performed in the Hartl lab and only a short description of the process is given here. The Cysteine at position 298 is shielded by Maltotriose binding to DM-MBP. In the first labeling step, DM-MBP is incubated for 2 h at room temperature with a 1.1-fold molar excess of Atto 532-Maleimide and 500 mM Maltotriose resulting in preferential labeling of position 15. Maltotriose and unbound fluorophore is removed using a 10 kDa Amicon spin column (Millipore) followed by incubation with a 2.5-fold molar excess of Atto 647N-Maleimide. After 2 h incubation, excess dye is again removed by a 10 kDa spin column. The labeling efficiency was measured by the absorption of MBP ($\epsilon_{280nm} = 69,000\,\mathrm{M^{-1}cm^{-1}}$), Atto 532 ($\epsilon_{532nm} = 115,000\,\mathrm{M^{-1}cm^{-1}}$) and Atto 647N ($\epsilon_{647nm} = 150,000\,\mathrm{M^{-1}cm^{-1}}$) and was higher than 90%. The structure of MBP with the mutations V8G and Y283D (Sharff *et al.*, 1992), the labeling positions and the positions of the intrinsic Tryptophan residues are shown in Figure 7.3.

7.2.2 FCS, FCCS, and single-pair FRET experiments

Fluorescence correlation spectroscopy (FCS), fluorescence cross-correlation spectroscopy (FCCS) and single-pair FRET (spFRET) experiments were performed using the setup described in Chapter 4 or on a Microtime 200 instrument (Picoquant). FCS was used to detect aggregation during spontaneous refolding by monitoring the diffusion time of a single-labeled sample. FCCS was employed to increase the sensitivity to detect also smaller aggregates down to dimers. There, single-labeled DM-MBP species with Atto 532 and with Atto 647N, respectively, were mixed and the amplitude of the CCF was measured.

The conformation of DM-MBP in different environments was measured by spFRET, and PDA was used to extract the center and width of the distance distributions from the FRET efficiency histograms as described in Section 4.1.5.7.

7.3 Results

7.3.1 Aggregation is not responsible for the reduced folding rate in DM-MBP

FCS was employed to detect higher-order aggregates of fluorescently labeled DM-MBP during refolding by a change of the diffusion time through the sample volume. FCS measurements were performed with DM-MBP single labeled with Atto 532 at position 52. The diffusion time of refolding DM-MBP (10 nM, Figure 7.4[2]) was 392 µs and did not change significantly when 1 or 2 µM of unlabeled DM-MBP was added to the refolding reaction. This indicates the absence of large aggregates. GroEL-bound DM-MBP (∼800 kDa) was used as a control and showed a diffusion time of 1.1 ms (Figure 7.4). The increase in diffusion time of a factor of 2.8 is in reasonable agreement with the expected change (2.5×) resulting from the difference in mass of a GroEL-bound MBP molecule to a free DM-MBP molecule.

For satisfactory fitting of the data, a stretched exponential term was added to the model for a single diffusing species (Equation 4.35) to account for the strong afterpulsing of the detectors used for these experiments (MPD, Picoquant).

$$G(\tau) = G_D(\tau) + A * exp\left(\left(\frac{\tau}{\tau_A}\right)^{\beta}\right) \tag{7.1}$$

Here, G_D is the correlation function of a single diffusing species (equation 4.33) and A is the afterpulsing amplitude. The average decay time of the stretched exponential, $\tau_A/\beta \cdot \Gamma(1/\beta)$, was for all fits on the order of 5-10 µs with typically $\beta = 0.3$. The additional stretched exponential term did not interfere with the fit for the diffusion term.

To test whether smaller aggregates, down to dimers, formed during refolding, additional FCCS experiments were performed. Single-labeled DM-MBP with either Atto 532 or Atto 647N at position 52 was unfolded as a 1:1 mixture and refolding initiated at a final concentration of ∼5 nM for both single-labeled species. FCCS measurements were performed within the first 250 s of refolding. According to simulations performed in (Chakraborty *et al.*, 2010) that assumed aggregation-limited folding kinetics, aggregates are expected to be populated on average to ∼75% during the first 250 s

[2]The amplitude of the weighted residuals should be distributed with a normal distribution with $\sigma = 1$. Clearly, the distribution in Figures 7.4 and 7.5 is narrower. This is probably caused by enlarged error bars due to concentration changes by adhesion of protein to the walls of the sample chamber.

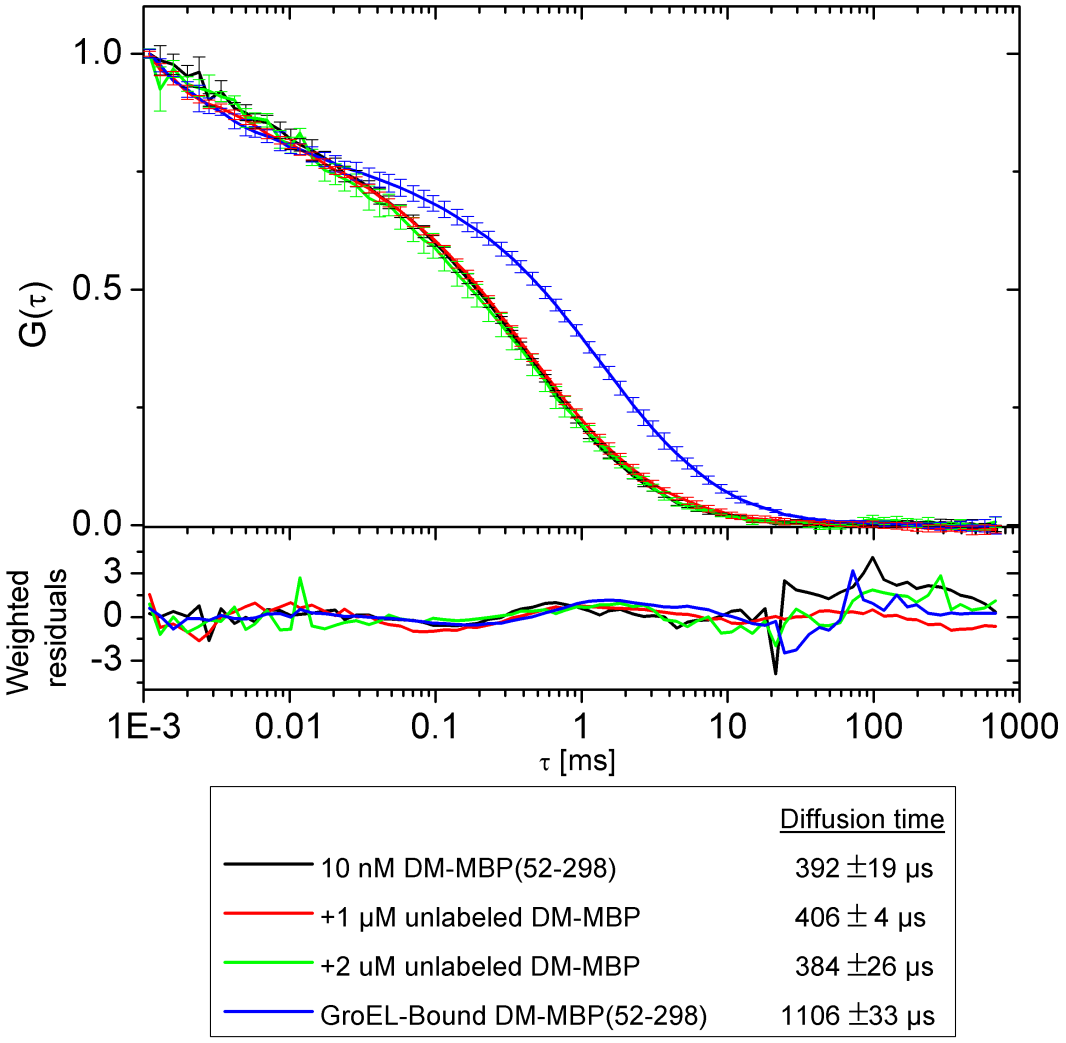


Figure 7.4: *FCS measurements of refolding DM-MBP to show the absence of large aggregates. Normalized autocorrelation curves for 10 nM Atto 532 labeled DM-MBP alone (black) and with the addition of 1 µM (red) and 2 µM (green) unlabeled DM-MBP. GroEL-bound DM-MBP was measured as a positive control (blue). For the calculation of the error bars, each measurement was divided into 8 parts and the average value and the standard error of the mean of the correlation curves from the individual parts were calculated. The diffusion time was determined from the data by a nonlinear least squares fit to equation 7.1 and the error of the diffusion time was calculated from fits to three independent measurements. Weighted residuals for the fits are shown in the lower panel.*

of refolding, corresponding to ∼35% of particles containing both labels in case of exclusive dimer formation. This would result in significant cross-correlation amplitudes as indicated in Figure 7.5. However, no cross-correlation between the Atto 532 and the Atto 647N signal were detected indicating the absence of dimeric or multimeric species (Figure 7.5).

Figure 7.5A shows autocorrelation and cross-correlation functions (ACFs and CCF) of a mixture of free dyes Atto 532 and Atto 647N as a negative control. The ACFs show that both dyes are present in solution at similar concentration (∼5 nM). No cross-correlation was observed as expected for the non-interacting dyes using PIE (Müller *et al.*, 2005). An HPLC- purified DNA containing both labels was measured as a positive control for cross-correlation. The theoretical CCF for an ideally double-labeled sample would have the average amplitude of the two ACFs. Imperfect overlap of the observation volumes, photobleaching, and imperfect labeling limit the CCF to a lower amplitude as shown in Figure 7.5B. This measurement yields a positive control for the maximum cross-correlation to be expected for fully interacting species.

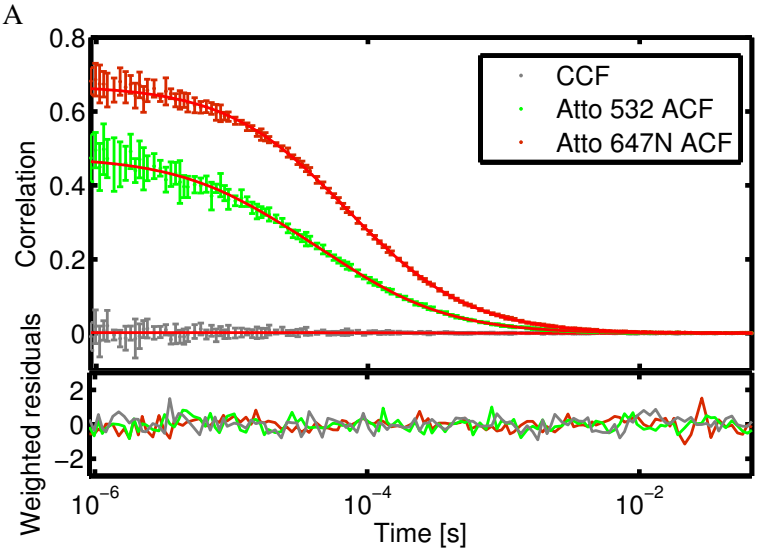
The combined FCS and FCCS measurements presented in this section exclude both small (FCCS) and large (FCS) aggregates as the cause of slow spontaneous DM-MBP refolding.

7.3.2 SpFRET characterization of a folding Intermediate

In the absence of transient aggregation, another possibility for the observed slow spontaneous refolding of DM-MBP compared to GroEL-assisted folding is the formation of a monomeric, kinetically trapped folding intermediate. To define this species, our collaborators first analyzed the GuHCl-dependent unfolding and refolding profiles of DM-MBP by Tryptophan (Trp) fluorescence and circular dichroism (CD) measurements (Chakraborty *et al.*, 2010). They detected a prominent hysteresis effect in the denaturation and renaturation curves. This is an indication for the presence of a folding intermediate. According to the hysteresis curves, this intermediate is populated at 0.5-0.8 M GuHCl during refolding. In contrast, the native state is stable up to approximately 0.8 M GuHCl. According to the CD spectra, the trapped intermediate of DM-MBP at 0.5 M GuHCl contained only approximately 22% α-helical structure compared to ∼50% in the native state (Sharff *et al.*, 1992). The low intensity of Trp fluorescence compareable to the unfolded state suggests solvent exhibition of the Trp residues (the positions of the Trp residues are indicated in Figure 7.3) indicating the lack of tertiary structure. Spectroscopic analysis immediately upon removal from denaturant (burst phase amplitude) indicated that the trapped intermediate is also populated under refolding conditions (Chakraborty *et al.*, 2010).

To analyze the compactness of the intermediate at 0.5 M GuHCl compared to the native and unfolded states, I performed single-pair FRET (spFRET) burst analysis measurements in solution and probability distribution analysis (PDA, Section 4.1.5.7). DM-MBP was double-labeled with Atto 532 (position 52) as the fluorescence donor and Atto 647N (position 298) as the acceptor as described in Section 7.2.1. According to the crystal structure (Sharff *et al.*, 1992), these two residues are 33 Å apart in the native structure (Figure 7.3). This value corresponds well to the distance of 36 Å determined by PDA (Figure 7.6A and Table 7.1). The width of the distance distribution in the native state is 4 Å. This value suggests a very rigid conformation, with the residual distance distribution originating mainly from acceptor photophysics. The unfolded protein in 3 M GuHCl shows a distance distribution that is much wider than in the native state ($\sigma = 12.3$ Å) with an average distance of 81.5 Å between the fluorophores (Figure 7.6B). This reflects the unfolding of the native structure to a more flexible gaussian chain. The higher value for the χ^2 of the fit for the unfolded protein (Table 7.1) may be due to the fact that the distribution of distances in a gaussian chain is not gaussian as assumed in the fitting model.

A

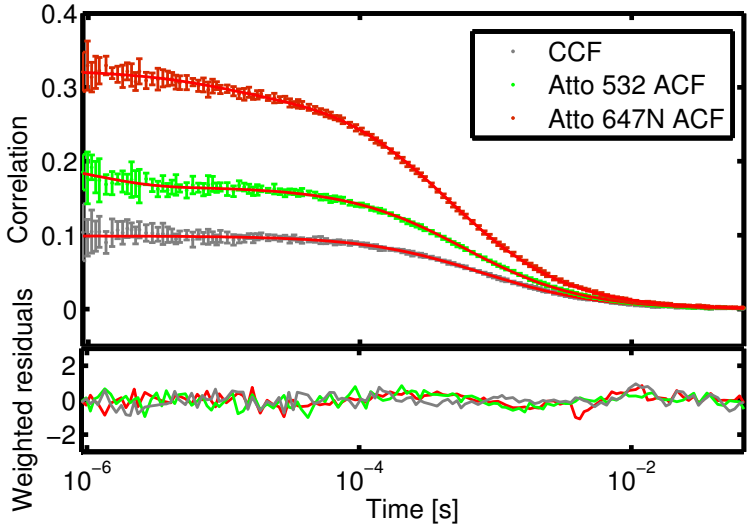


CCF
N = (2.6 ± 3.3) 10^{-3}

Atto 532 ACF
N = 0.87 ± 0.03
$_{diff}$ = (58 ± 3) µs
T = 0.15 ± 0.03
$_{trip}$ = (11 ± 4) µs

Atto 647N ACF
N = 0.56 ± 0.02
$_{diff}$ = (79 ± 3) µs
T = 0.06 ± 0.02
$_{trip}$ = (25 ± 11) µs

B

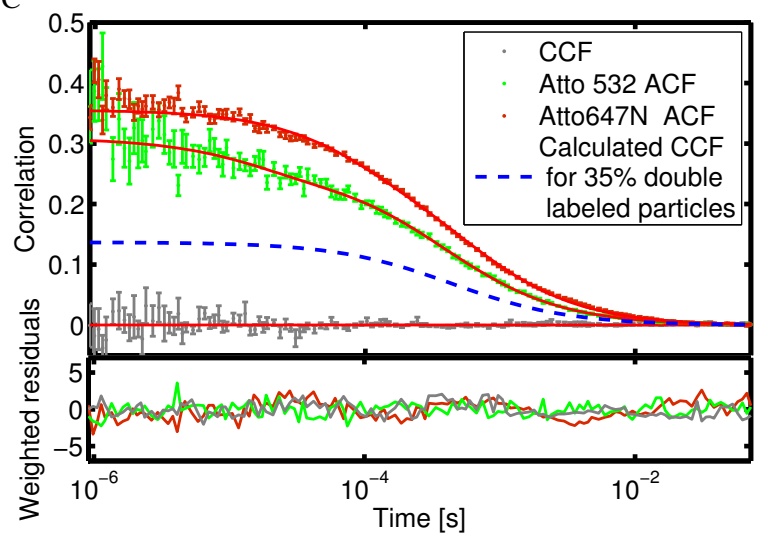


CCF
N = 0.713 ± 0.001
$_{diff}$ = (793 ± 12) µs

Atto 532 ACF
N = 2.138 ± 0.009
$_{diff}$ = (587 ± 8) µs
T = 0.2 ± 0.1
$_{trip}$ = (1.1 ± 0.6) µs

Atto 647N ACF
N = 1.190 ± 0.008
$_{diff}$ = (444 ± 5) µs
T = 0.081 ± 0.009
$_{trip}$ = (8.1 ± 1.7) µs

C



CCF
N = (1.1 ± 1.8) 10^{-5}

Atto 532 ACF
N = 1.39 ± 0.04
$_{diff}$ = (383 ± 18) µs
T = 0.18 ± 0.05
$_{trip}$ = (15 ± 8) µs

Atto 647N ACF
N = 1.19 ± 0.04
$_{diff}$ = (500 ± 31) µs
T = 0.17 ± 0.02
$_{trip}$ = (74 ± 16) µs

c(GuHCl) [M]	Mean distance [Å]	Distribution Width [Å]	χ^2
0	36.0	4.0	1.00
0.5	40.0	9.4	1.54
3	81.5	12.3	2.92

Table 7.1: *Results of the PDA analsis of DM-MBP.*

The intermediate state at 0.5 M GuHCl displayed a broader distribution ($\sigma = 9.4$ Å) than the native state (Figure 7.6C). The average distance, however, increased only sightly to 40 Å. This suggests that the species trapped in the refolding phase has a similar compaction than the native protein but has a larger variability in structure, in agreement with the lower amount of tertiary structure shown by Trp fluorescence. The broadening observed in the PDA analysis of the unfolded and intermediate states suggests that the typical timescale of the conformational changes is on the millisecond timescale or slower. Another possible reason for the broadening may be aditional photophysics of the acceptor, e.g. due to interactions with Trp residues that are more exposed than in the native state. The distance distributions are shown in Figure 7.7 and the peak and width values of the distance distributions are given together with the χ^2 of the PDA fit in Table 7.1.

7.4 Discussion

It was discussed that accelerated folding of substrate proteins like RuBisCo or DM-MBP by GroEL may be explained by a purely passive effect of the encapsulation by the prevention of reversible or irreversible aggregation of folding intermediates (Apetri and Horwich, 2008; Tyagi *et al.*, 2011). Here, we have employed several techniques, including FCS and FCCS experiments, to exclude transient aggregation as the cause for the accelerated folding of DM-MBP. We find that the actual cause for the slow folding of DM-MBP without the presence of GroEL is the formation of a kinetically trapped intermediate state. GroEL employs an active mechanism to rescue DM-MBP from this kinetic trap thus increasing the folding rate. For other substrates, GroEL does indeed provide a passive enclosed space to prevent aggregation. Both mechanisms work together to increase the folding speed and yield of different substrates.

GroEL substrates are typically proteins with complex topologies and are often stabilized not only by local interactions and hydrophobic forces but also by long range contacts of residues far apart in the sequence. The formation of these is entropically unfavorable (Plaxco *et al.*, 1998) and results in

Figure 7.5: ***Previous page:*** *FCCS measurements of refolding DM-MBP to show the absence of reversible dimer formation. (A) ACFs and CCF of a mixture of Atto 532 and Atto 647N as a negative control. (B) ACFs and CCF of dsDNA labeled with Atto 532 and Atto 647N as positive control. (C) ACFs and CCF of a mixture of DM-MBP single-labeled with Atto 532 or with Atto 647N and calculated CCF for 35% double-labeled particles. (A-C) The ACF of the signal in the detection channel for Atto 532 is shown in green, the ACF of the signal in the detection channel for Atto 647N is shown in red, and the CCF of the signals of both dyes is shown in gray. The red lines overlaying the data are fits using equation 4.36. The lower panel shows weighted residuals of the fits. Fit results and error estimates are shown to the right of the figures, where N is the average number of molecules bearing a green and red label in the green and red ACFs and the average number of molecules bearing both labels in the CCFs.*

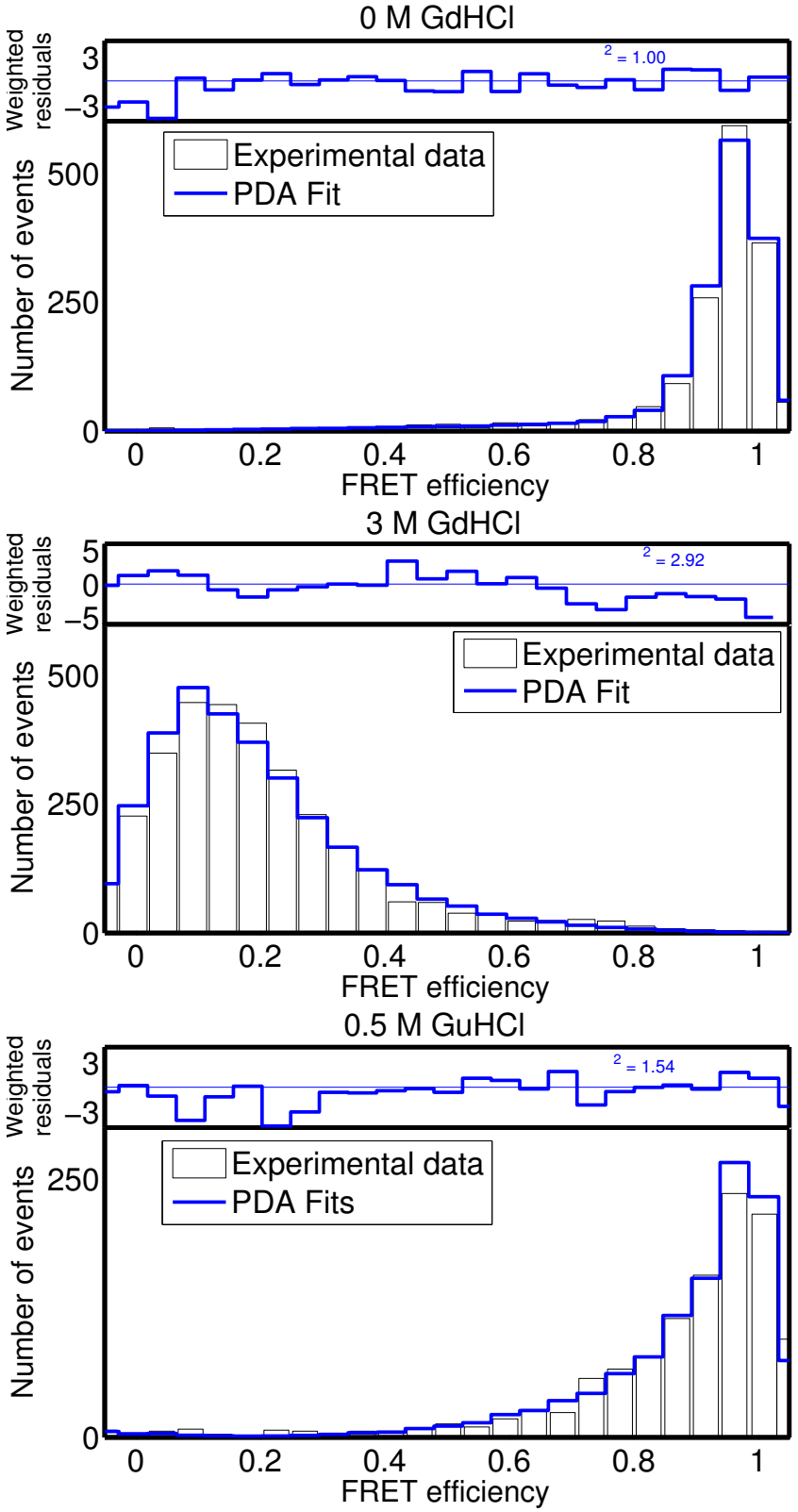


Figure 7.6: *SpFRET histograms of double-labeled DM-MBP in 0 M (upper panel), 3 M (middle panel), and 0.5 M GuHCL (lower panel). The blue line represents the PDA fit to the data using a single gaussian distance distribution. The weighted residuals of the fit are shown in the upper panels.*

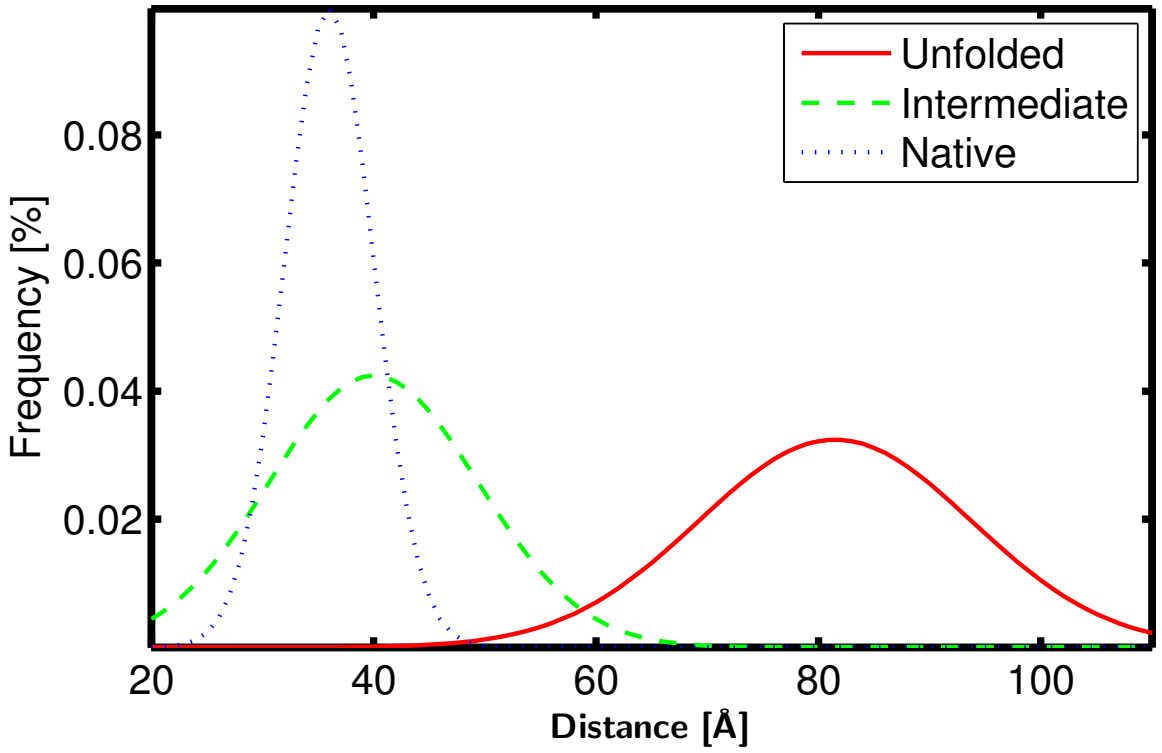


Figure 7.7: *Distance distributions calculated by PDA from the spFRET histograms in Figure 7.6 for double-labeled DM-MBP in 0 M GuHCL (dotted blue line), 3 M (solid red line), and 0.5 M (dashed green line).*

slow folding rates. For DM-MBP, a kinetically trapped intermediate is significantly populated during folding as shown by the pronounced hysteresis in unfolding/refolding CD and Trp fluorescence traces (Chakraborty *et al.*, 2010). The intermediate is stable upon refolding in the presence of 0.5 M GuHCL and was studied under these conditions using single-pair FRET burst analysis and PDA. It is collapsed and similar in compactness as the native state. However, lacking tertiary structure results in more structural flexibility suggesting that the intermediate state is not stabilized by non-native contacts between specific amino acids but by kinetic hindrance to reach the native state.

Our collaborators tested this hypothesis by the introduction of disulfide bonds between residues close in the native state but far apart along the sequence (Chakraborty *et al.*, 2010). These disulfide bonds produce configurational constraints that have a similar effect as the confinement in the GroEL cage, depending on the location of the disulfide bond. A disulfide bond in the N-terminal domain between residues 18 and 296 increased the folding rate by a factor of 5 and removed the hysteresis in the unfolding-refolding profiles. This bond may have a similar stabilizing effect as a hydrogen bond between residues 296 (aspartic acid) and the backbone amine of residue 19. The disulfide bridge accelerated folding when it was formed before protein collapse as well as when formed after protein collapse, indicating that the residues are close also in the intermediate state. Constraining already close residues accelerates folding only if the protein has significant conformational flexibility in the intermediate state, in agreement with the broadened distance distribution observed in the PDA analysis.

About 10% of all proteins require GroEL for folding *in vivo*, and deletion of GroEL is lethal for the cell under all conditions. Several redundant chaperone systems exist for the prevention of aggregation (Section 3.3). Rescue of trapped intermediates is a unique property of GroEL and thus explains why it is essential for cell viability and cannot be substituted by other chaperones. Additionally, GroEL may be able to buffer mutations that do not change the thermodynamic stability of the native state but introduce kinetic hindrances for productive folding as in DM-MBP. This may facilitate protein evolution by increasing the number of foldable proteins (Tokuriki and Tawfik, 2009; Chakraborty *et al.*, 2010).

7.5 Outlook

7.5.1 Resolving fast fluctuations in DM-MBP using FCS

Although the experiments presented here shed light on the mechanism of GroEL-assisted folding, they also raise several new questions. The intermediate state of DM-MBP observed at 0.5 M GuHCL shows a significantly broader distance distribution as the native state, implying more conformational flexibility. The timescale of these fluctuations, however, is not accessible by PDA [3].

FCS is theoretically able to resolve fluctuations on timescales from ps to s (Felekyan *et al.*, 2005; Nettels *et al.*, 2008). Conformation-dependent signal fluctuations for the study of the conformational dynamics of DM-MBP can be induced by several different labeling schemes. Specific labeling on one position and the observation of quenching by aromatic residues can yield insights into the overall dynamics of a protein (Chen *et al.*, 2007). Specific labeling on two positions and the observation of fluctuations in the donor or acceptor signal due to FRET allows the discrimination of fluctuations in

[3] Dynamics can be resolved by PDA only in a limited range of timescales. The range is limited at the lower end by the number of photons necessary to calculate a sensible FRET efficiency to approximately 10 times the photon rate ($\sim$200 µs), and on the upper end by the duration of the bursts ($\sim$5 ms).

specific parts of the protein. Differences in fluctuation times in different parts of the protein, their dependence on the GuHCL content and changes in the presence of GROEL/ES may give interesting insights into the folding process of DM-MBP and the principles of GroEL assistance.

7.5.2 spFRET experiments in a microfluidic mixer

Many of the folding reactions in GroEL occur on short timescales and are not accessible in equilibrium experiments. Time-resolved ensemble methods have often been employed to resolve these fast processes, e.g. the transient compaction of DM-MBP in the GroEL cage after ATP addition (Sharma *et al.*, 2008). These methods have the necessary time resolution to resolve these processes. Stopped flow mixing has a time resolution of approximately 1 ms, and the use of continuous flow hydrodynamic focusing techniques has even reduced the time resolution to less than 10 µs (Knight *et al.*, 1998). Several techniques have been employed in combination with these mixing techniques including circular dichroism, FRET, fluorescence anisotropy, and small angle x-ray scattering (Wu *et al.*, 2008). However, all of these techniques suffer from ensemble averaging and are not able to resolve the heterogeneities of the processes involved in GroEL-assisted folding (Section 2.4).

The observation of single molecules is limited by the photon detection rate. For a single molecule with saturating excitation and usual detection efficiencies of ∼5%, the photon count rate is 100-1000 kHz. Even under optimum conditions and with 30 photons on average per bin for the calculation of FRET efficiencies, the maximum possible time resolution is limited to ∼35 µs. Under realistic conditions, single-molecule mixing has been limited to a time resolution of ∼1 ms. Recently, Gambin *et al.* (2011) have developed a novel microfluidic mixer that enables single-molecule measurements with mixing times down to 200 µs by fast mixing and subsequent flow deceleration.

Several groups have used hydrodynamic focusing together with spFRET burst analysis either in microfabricated devices (Lipman *et al.*, 2003; Pfeil *et al.*, 2009; Hofmann *et al.*, 2010) or using glass capillaries pulled to opening diameters in the µm range (Hamadani and Weiss, 2008). Hofmann *et al.* (2010) have used their microfluidic mixer to study spontaneous and GroEL-assisted folding of Rhodanese, a classic GroEL substrate, but did not observe conformational changes during the first seconds of folding. This is in contrast to the transient stretching observed in DM-MBP in (Sharma *et al.*, 2008). One possible reason may be the use of a single-ring GroEL mutant in (Hofmann *et al.*, 2010). Similar experiments using DM-MBP and wild-type GroEL might be able to elucidate the cause for the transient stretching of DM-MBP during GroEL-assisted folding.

It is clear that single-molecule fluorescence experiments have the ability to produce important new insights into the function of chaperones like GroEL and will contribute new results in the future. The combination of single-molecule fluorescence methods with other techniques and the application of new labeling schemes may even increase the possibilities of GroEL research.

8 Conclusions and Summary

Protein folding is an essential part of protein biosynthesis and misfolding can result in uncontrolled protein aggregation which is characteristic for several severe diseases. Although several models have been proposed to describe protein folding, a complete microscopic theory to describe folding of all classes of proteins is still lacking. In this thesis, I have employed single-molecule fluorescence spectroscopy to study protein folding using model systems on different levels of complexity.

Myoglobin, a simple globular protein, is used as a model system to study a protein folding intermediate. The acidic compact state of myoglobin is stabilized at acidic pH conditions and has several similarities to the molten globule state, a common protein folding intermediate and is studied by a combination of molecular dynamics (MD) simulations and several fluorescence spectroscopic techniques. Using two focus FCS, it is shown that the acidic compact state of myoglobin is less compact than the native state, but not as expanded as the fully unfolded state. The agreement between the predictions of the MD simulations and the experimental results validates that the MD simulations well describe the acidic compact state of Myoglobin. The analysis of exposed hydrophobic regions in the acidic structures generated by the MD simulations reveals potential candidates involved in the aggregation and fibril formation processes of myoglobin in the acidic compact state. These results may contribute to the understanding of disease-related fibril formation. These results contribute to better understand disease-related fibril formation which will hopefully lead ultimately to better treatments for diseases like Alzheimer's disease.

While myoglobin is a typical example of a small globular protein that is able to fold in the test tube without assistance by other proteins, the situation is much more complicated for larger protein complexes or proteins with more complex topology. An additional obstacle for productive folding is the highly crowded environment of the living cell. A huge machinery of specialized proteins, the molecular chaperones, has evolved to assist protein folding in the cell. I have studied several members of this machinery. The conformation of the mitochondrial heat shock protein 70 (Hsp70), Ssc1, was probed in different stages of its functional cycle using single-pair FRET experiments on the single-molecule level. Ssc1 has a very well defined conformation in the ATP state with closely docked substrate binding domain and nucleotide binding domain but shows significantly more heterogeneity in the presence of ADP. The dependence of the rates between the different FRET states on the ADP concentration suggests that this heterogeneity is presumably due to binding and release of ADP. The nucleotide-free state, which has not been experimentally accessible until now due to the high tendency of Ssc1 to aggregate in a nucleotide-free environment, has less inter-domain contacts than the ATP or ADP-bound states. However, the addition of a substrate peptide induces closing of the substrate binding lid along with a further increase of the distance between the domains. Thus, substrate binding plays an active role in the structural remodeling of Ssc1. This behavior is strikingly different than in DnaK, the major bacterial Hsp70. In DnaK, domain undocking in the presence of ADP was observed in all molecules, followed by a slight re-compaction upon substrate binding. These differences may reflect tuning of Ssc1 to meet specific functions, i.e. protein import into mitochondria, in addition to protein folding.

Ssc1 requires the assistance of several cofactors modulating its function depending on the specific task at hand. In the protein folding cycle, Mdj1 delivers unfolded substrates to Ssc1. Using spFRET

and ensemble kinetic measurements, it is shown here that the interaction between Mdj1 and Ssc1 is transient but with a surprisingly long interaction time. This may reflect specific requirements when binding unfolded protein chains to prevent their aggregation and promote their folding in the mitochondrial matrix. In the protein import context, Ssc1 is recruited to the import channel by the peripheral membrane protein Tim44. SpFRET experiments show that Ssc1 adopts a similar conformation when bound to Tim44 independent of the nucleotide present. Tim44 primes Ssc1 to a conformation similar to the substrate-bound form. However, the spFRET MFD-PIE data suggests significant dynamical changes in Tim44-bound Ssc1 in the presence of ADP between the substrate-bound conformation and a more ADP-like conformation.

While Hsp70s perform a number of different tasks in conjunction with several different cofactors, the GroEL/ES complex is a highly specialized molecular machine that is essential for folding of a large subset of proteins. The properties of proteins typically requiring the assistance of GroEL are not completely understood yet. For some proteins, GroEL assists folding by providing an isolated environment where the protein can fold unaffected by aggregation. However, it is shown here that GroEL plays an active role in the folding of DM-MBP. DM-MBP populates a kinetically trapped intermediate state when folding spontaneously, and GroEL rescues DM-MBP by the introduction of entropic constraints, and thereby increases its folding rate. These findings suggest that proteins with a tendency to populate kinetically trapped intermediates during folding may require GroEL assistance. The capacity of GroEL to rescue proteins from such folding traps may be part of the explanation of the unique role of GroEL within the cellular chaperone machinery.

Clearly, the combination of a variety of single-molecule fluorescence techniques employed in this thesis can yield significant new insights into protein folding and the role of molecular chaperones. The "ideal" experiment with the ability to give a complete description of protein folding may be to monitor the conformation of a single molecule over the complete process of spontaneous or chaperone-assisted folding with sufficient time resolution to resolve the individual steps of the process. Although this "ideal" experiment is still out of reach, the technological advances of the past years allowed experiments that would have been unthinkable before, and today, researchers throughout the world contribute pieces to the puzzle that will ultimately give a complete picture of protein folding.

Bibliography

Alcala, J. R., Gratton, E., and Prendergast, F. G., 1987. Interpretation of fluorescence decays in proteins using continuous lifetime distributions. *Biophysical journal*, 51(6):925–936.

Andrec, M., Levy, R., and Talaga, D., 2003. 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.

Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C., 2006. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *The Journal of Physical Chemistry B*, 110(13):6970–6978.

Apetri, A. C. and Horwich, A. L., 2008. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proceedings of the National Academy of Sciences*, 105 (45):17351–17355.

Baldwin, R., 1989. How does protein folding get started? *Trends in biochemical sciences*, 14(7): 291–294.

Baldwin, R. L., 1994. Matching speed and stability. *Nature*, 369(6477):183–184.

Barraclough, R. and Ellis, R., 1980. Protein synthesis in chloroplasts ix. assembly of newly-synthesized large subunits into ribulose bishopshate carboxylase in isolated intact pea chloroplasts. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, 608(1):19–31.

Barrick, D. and Baldwin, R. L., 1993. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Science*, 2(6):869–876.

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

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

Baumann, F., Milisav, I., Neupert, W., and Herrmann, J., 2000. Ecm10, a novel hsp70 homolog in the mitochondrial matrix of the yeast saccharomyces cerevisiae. *Febs Letters*, 487(2):307–312.

Becker, D., Krayl, M., Strub, A., Li, Y., Mayer, M. P., and Voos, W., 2009. Impaired interdomain communication in mitochondrial hsp70 results in the loss of inward-directed translocation force. *Journal of Biological Chemistry*, 284(5):2934–2946.

Becker, W. *The bh TCSPC handbook*. Becker & Hickl GmbH, 2005a.

Becker, W. *Advanced time-correlated single photon counting techniques*. Springer Verlag, 2005b.

Bertelsen, E., Chang, L., Gestwicki, J., and Zuiderweg, E., 2009. Solution conformation of wild-type e. coli hsp70 (dnak) chaperone complexed with adp and substrate. *Proceedings of the National Academy of Sciences*, 106(21):8471.

Bilmes, J., 1998. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4:126.

Bismuto, E. and Irace, G., 1994. Unfolding pathway of apomyoglobin : Simultaneous characterization of acidic conformational states by frequency domain fluorometry. *Journal of molecular biology*, 241(1):103–109.

Bismuto, E., Irace, G., Servillo, L., Giovane, A., and Colonna, G., 1984. Conformational stability and basal metabolic rate: Reexamination of the case of myoglobin. *Cellular and Molecular Life Sciences*, 40(12):1400–1401.

Bismuto, E., Colonna, G., Savy, F., and Irace, G., 1985. Myoglobin structure and regulation of solvent accessibility of heme pocket. *International Journal of Peptide and Protein Research*, 26 (2):195–207.

Bismuto, E., Gratton, E., and Irace, G., 1988. Effect of unfolding on the tryptophanyl fluorescence lifetime distribution in apomyoglobin. *Biochemistry*, 27(6):2132–2136.

Bismuto, E., Di Maggio, E., Pleus, S., Sikor, M., Röcker, C., Nienhaus, G., and Lamb, D., 2009. Molecular dynamics simulation of the acidic compact state of apomyoglobin from yellowfin tuna. *Proteins: Structure, Function, and Bioinformatics*, 74(2):273–290.

Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., and Pfanner, N., 2008. Multiple pathways for sorting mitochondrial precursor proteins. *EMBO reports*, 9(1):42–49.

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

Brinker, A., Pfeifer, G., Kerner, M., Naylor, D., Hartl, F., and Hayer-Hartl, M., 2001. Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*, 107(2):223–233.

Brodsky, J., 1996. Post-translational protein translocation: not all hsc70s are created equal. *Trends in Biochemical Sciences*, 21(4):122–126.

Bronson, J., Fei, J., Hofman, J., Gonzalez Jr, R., and Wiggins, C., 2009. Learning rates and states from biophysical time series: a bayesian approach to model selection and single-molecule fret data. *Biophysical journal*, 97(12):3196–3205.

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

Buchberger, A., Theyssen, H., Schröder, H., McCarty, J., Virgallita, G., Milkereit, P., Reinstein, J., and Bukau, B., 1995. Nucleotide-induced conformational changes in the atpase and substrate binding domains of the dnak chaperone provide evidence for interdomain communication. *Journal of Biological Chemistry*, 270(28):16903.

Bukau, B., 1998. The hsp70 and hsp60 chaperone machines. *Cell*, 92:351–366.

Bukau, B., Weissman, J., and Horwich, A., 2006. Molecular chaperones and protein quality control. *Cell*, 125(3):443–451.

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

Chaudhuri, T., Farr, G., Fenton, W., Rospert, S., and Horwich, A., 2001. Groel/groes-mediated folding of a protein too large to be encapsulated. *Cell*, 107(2):235–246.

Chen, H., Rhoades, E., Butler, J., Loh, S., and Webb, W., 2007. Dynamics of equilibrium structural fluctuations of apomyoglobin measured by fluorescence correlation spectroscopy. *Proceedings of the National Academy of Sciences*, 104(25):10459.

Cheng, M., Hartl, F., Martin, J., Pollock, R., Kalousek, F., Neupert, W., Hallberg, E., Hallberg, R., and Horwich, A., 1989. Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature*, 337(6208):620.

Chiti, F. and Dobson, C. M., 2006. Protein misfolding, functional amyloid, and human disease. *Annual review of biochemistry*, 75(1):333–366.

Chun, S., Strobel, S., Bassford, P., and Randall, L., 1993. Folding of maltose-binding protein. evidence for the identity of the rate-determining step in vivo and in vitro. *Journal of Biological Chemistry*, 268(28):20855.

Cocco, M. J. and Lecomte, J. T. J., 1990. Characterization of hydrophobic cores in apomyoglobin: a proton nmr spectroscopy study. *Biochemistry*, 29(50):11067–11072.

Craig, E. and Marszalek, J., 2002. A specialized mitochondrial molecular chaperone system:Â¶ a role in formation of fe/s centers. *Cellular and Molecular Life Sciences*, 59(10):1658–1665.

Craig, E., Kramer, J., and Kosic-Smithers, J., 1987. Ssc1, a member of the 70-kda heat shock protein multigene family of saccharomyces cerevisiae, is essential for growth. *Proceedings of the National Academy of Sciences of the United States of America*, 84(12):4156.

Craig, E., Huang, P., Aron, R., and Andrew, A., 2006. The diverse roles of j-proteins, the obligate hsp70 co-chaperone. *Reviews of physiology, biochemistry and pharmacology*, pages 1–21.

Daggett, V. and Fersht, A., 2003. Is there a unifying mechanism for protein folding? *Trends in biochemical sciences*, 28(1):18–25.

Dahan, M., Deniz, A., Ha, T., Chemla, D., Schultz, P., and Weiss, S., 1999. Ratiometric measurement and identification of single diffusing molecules. *Chemical Physics*, 247(1):85–106.

Dale, R., Eisinger, J., and Blumberg, W., 1979. The orientational freedom of molecular probes. the orientation factor in intramolecular energy transfer. *Biophysical journal*, 26(2):161–193.

De Los Rios, P., Ben-Zvi, A., Slutsky, O., Azem, A., and Goloubinoff, P., 2006. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proceedings of the National Academy of Sciences of the United States of America*, 103(16):6166–6171.

Deloche, O., Kelley, W., and Georgopoulos, C., 1997. Structure-function analyses of the ssc1p, mdj1p, and mge1p saccharomyces cerevisiae mitochondrial proteins in escherichia coli. *Journal of Bacteriology*, 179(19):6066.

Dertinger, T., Pacheco, V., von der Hocht, I., Hartmann, R., Gregor, I., and Enderlein, J., 2007. Two focus fluorescence correlation spectroscopy: A new tool for accurate and absolute diffusion measurements. *ChemPhysChem*, 8(3):433–443.

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M., 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences*, 25 (7):331–339.

D'Silva, P., Liu, Q., Walter, W., and Craig, E., 2004. Regulated interactions of mthsp70 with tim44 at the translocon in the mitochondrial inner membrane. *Nature structural & molecular biology*, 11 (11):1084.

Durbin, R. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.

Eftink, M. R. and Ghiron, C. A., 1981. Fluorescence quenching studies with proteins. *Analytical Biochemistry*, 114(2):199–227.

Eggeling, C., Berger, S., Brand, L., Fries, J., Schaffer, J., Volkmer, A., and Seidel, C., 2001. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *Journal of biotechnology*, 86(3):163–180.

Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E., 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Mol Biol*, 18(2):148–155.

Elson, E. L. and Magde, D., 1974. Fluorescence correlation spectroscopy. i. conceptual basis and theory. *Biopolymers*, 13(1):1–27.

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

Enderlein, J., Gregor, I., Patra, D., and Fitter, J., 2004. Art and artefacts of fluorescence correlation spectroscopy. *Current Pharmaceutical Biotechnology*, 5(2):155–161.

Enderlein, J., Gregor, I., Patra, D., Dertinger, T., and Kaupp, U., 2005. Performance of fluorescence correlation spectroscopy for measuring diffusion and concentration. *ChemPhysChem*, 6(11):2324–2336.

Ewalt, K., Hendrick, J., Houry, W., and Hartl, F., 1997. In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell*, 90(3):491–500.

Felekyan, S., KÃ¼hnemuth, R., Kudryavtsev, V., Sandhagen, C., Becker, W., and Seidel, C., 2005. Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Review of Scientific Instruments*, 76:083104.

Fink, A. L., 1995. Compact intermediate states in protein folding. *Annual review of biophysics and biomolecular structure*, 24(1):495–522.

Fändrich, M., Fletcher, M., and Dobson, C., 2001. Amyloid fibrils from muscle myoglobin. *Nature*, 410(6825):165–166.

Frank, G. A., Kipnis, Y., Smolensky, E., Daube, S. S., Horovitz, A., and Haran, G., 2008. Design of an optical switch for studying conformational dynamics in individual molecules of groel. *Bioconjugate Chemistry*, 19(7):1339–1341.

Frank, G. A., Goomanovsky, M., Davidi, A., Ziv, G., Horovitz, A., and Haran, G., 2010. Out-of-equilibrium conformational cycling of groel under saturating atp concentrations. *Proceedings of the National Academy of Sciences*, 107(14):6270–6274.

Frauenfelder, H., Nienhaus, G. U., and Johnson, J. B., 1991. Rate-processes in proteins. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics*, 95(3):272–278.

Frauenfelder, H., Austin, R., Schultz, C., Nienhaus, U., and Young, R. *Physics of proteins: An introduction to biological physics and molecular biophysics (Biological & medical physics, biomedical engineering)*. Springer, 2010.

Fries, J., Brand, L., Eggeling, C., Köllner, M., and Seidel, C., 1998. Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *The Journal of Physical Chemistry A*, 102(33):6601–6613.

Frydman, J., 2001. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annual review of biochemistry*, 70(1):603–647.

Gambin, Y., VanDelinder, V., Ferreon, A., Lemke, E., Groisman, A., and Deniz, A., 2011. Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing. *Nature Methods*, 8(3): 239–241.

Gansen, A., Valeri, A., Hauger, F., Felekyan, S., Kalinin, S., Tóth, K., Langowski, J., and Seidel, C., 2009. Nucleosome disassembly intermediates characterized by single-molecule fret. *Proceedings of the National Academy of Sciences*, 106(36):15308.

Gell, C., Brockwell, D. J., Beddard, G. S., Radford, S. E., Kalverda, A. P., and Smith, D. A., 2001. Accurate use of single molecule fluorescence correlation spectroscopy to determine molecular diffusion times. *Single Molecules*, 2(3):177–181.

Genevaux, P., Georgopoulos, C., and Kelley, W. L., 2007. The hsp70 chaperone machines of escherichia coli: a paradigm for the repartition of chaperone functions. *Molecular Microbiology*, 66 (4):840–857.

Georgopoulos, C., Hendrix, R., Casjens, S., and Kaiser, A., 1973. Host participation in bacteriophage lambda head assembly* 1. *Journal of molecular biology*, 76(1):45–50.

Goloubinoff, P. and De Los Rios, P., 2007. The mechanism of hsp70 chaperones: (entropic) pulling the models together. *Trends in Biochemical Sciences*, 32(8):372–380.

Goloubinoff, P., Christeller, J., Gatenby, A., and Lorimer, G., 1989a. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and mg-atp. *Nature*, 342(6252):884–889.

Goloubinoff, P., Gatenby, A., and Lorimer, G., 1989b. Groe heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in escherichia coli. *Nature*, 337(6202):44.

Goto, Y. and Fink, A. L., 1990. Phase diagram for acidic conformational states of apomyoglobin. *Journal of molecular biology*, 214(4):803–805.

Gratton, E. and Limkeman, M., 1983. A continuously variable frequency cross-correlation phase fluorometer with picosecond resolution. *Biophysical journal*, 44(3):315–324.

Ha, T., Enderle, T., Ogletree, D., Chemla, D., Selvin, P., and Weiss, S., 1996. 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.

Hamadani, K. and Weiss, S., 2008. Nonequilibrium single molecule protein folding in a coaxial mixer. *Biophysical journal*, 95(1):352–365.

Hemmingsen, S. and Ellis, R., 1986. Purification and properties of ribulosebisphosphate carboxylase large subunit binding protein. *Plant physiology*, 80(1):269.

Hennessy, F., Nicoll, W., Zimmermann, R., Cheetham, M., and Blatch, G., 2005. Not all j domains are created equal: Implications for the specificity of hsp40-hsp70 interactions. *Protein Science*, 14 (7):1697–1709.

Herschel, J., 1845. 'a o no. i. on a case of superficial colour presented by a homogeneous liquid internally colourless. *Philosophical Transactions of the Royal Society of London*, pages 143–145.

Hertzog, D., Michalet, X., Jäger, M., Kong, X., Santiago, J., Weiss, S., and Bakajin, O., 2004. Femtomole mixer for microsecond kinetic studies of protein folding. *Analytical chemistry*, 76 (24):7169–7178.

Hillger, F., Hänni, D., Nettels, D., Geister, S., Grandin, M., Textor, M., and Schuler, B., 2008. Probing protein-chaperone interactions with single molecule fluorescence spectroscopy. *Angewandte Chemie International Edition*, 47(33):6184–6188.

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

Hofmann, H., Hillger, F., Pfeil, S., Hoffmann, A., Streich, D., Haenni, D., Nettels, D., Lipman, E., and Schuler, B., 2010. Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proceedings of the National Academy of Sciences*, 107(26):11793.

Horst, M., Oppliger, W., Rospert, S., Schönfeld, H., Schatz, G., and Azem, A., 1997. Sequential action of two hsp70 complexes during protein import into mitochondria. *The EMBO Journal*, 16 (8):1842–1849.

Houry, W., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F., 1999. Identification of in vivo substrates of the chaperonin groel. *Nature*, 402(6758):147–154.

Huang, B., 2010. Super-resolution optical microscopy: multiple choices. *Current opinion in chemical biology*, 14(1):10–14.

Hughson, F., Wright, P., and Baldwin, R., 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science*, 249(4976):1544–1548.

Irace, G., Bismuto, E., Savy, F., and Colonna, G., 1986. Unfolding pathway of myoglobin: Molecular properties of intermediate forms. *Archives of Biochemistry and Biophysics*, 244(2):459–469.

Itzhaki, L., Otzen, D., and Fersht, A., 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *Journal of molecular biology*, 254(2):260–288.

Ivanov, V., Li, M., and Mizuuchi, K., 2009. Impact of emission anisotropy on fluorescence spectroscopy and fret distance measurements. *Biophysical journal*, 97(3):922–929.

Jennings, P. and Wright, P., 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, 262(5135):892–896.

Jiang, J., Prasad, K., Lafer, E., and Sousa, R., 2005. Structural basis of interdomain communication in the hsc70 chaperone. *Molecular cell*, 20(4):513–524.

Jiang, J., Maes, E., Taylor, A., Wang, L., Hinck, A., Lafer, E., and Sousa, R., 2007. Structural basis of j cochaperone binding and regulation of hsp70. *Molecular cell*, 28(3):422–433.

Kalinin, S., Felekyan, S., Antonik, M., and Seidel, C., 2007. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *The Journal of Physical Chemistry B*, 111(34):10253–10262.

Kalinin, S., Felekyan, S., Valeri, A., and Seidel, C., 2008. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *The Journal of Physical Chemistry B*, 112(28):8361–8374.

Kang, P., Ostermann, J., Shilling, J., Neupert, W., Craig, E., and Pfanner, N., 1990. Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. *Nature*, 348 (6297):137.

Kapanidis, A., Lee, N., Laurence, T., Doose, S., Margeat, E., and Weiss, S., 2004. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (24):8936.

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

Karplus, M. and Weaver, D., 1994. Protein folding dynamics: The diffusion collision model and experimental data. *Protein Science*, 3(4):650–668.

Kawahara, K. and Tanford, C., 1966. Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *Journal of Biological Chemistry*, 241(13):3228–3232.

Kendrew, J., Bodo, G., Dintzis, H., Parrish, R., Wyckoff, H., and Phillips, D., 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610): 662–666.

Kim, S., Miller, E., Frydman, J., and Moerner, W., 2010. Action of the chaperonin groel/es on a non-native substrate observed with single-molecule fret. *Journal of molecular biology*, 401(4): 553–563.

Knight, J., Vishwanath, A., Brody, J., and Austin, R., 1998. Hydrodynamic focusing on a silicon chip: mixing nanoliters in microseconds. *Physical Review Letters*, 80(17):3863–3866.

Koshioka, M., Sasaki, K., and Masuhara, H., 1995. Time-dependent fluorescence depolarization analysis in three-dimensional microspectroscopy. *Applied spectroscopy*, 49(2):224–228.

Kubo, Y., Tsunehiro, T., Nishikawa, S., Nakai, M., Ikeda, E., Toh-e, A., Morishima, N., Shibata, T., and Endo, T., 1999. Two distinct mechanisms operate in the reactivation of heat-denatured proteins by the mitochondrial hsp70/mdj1p/yge1p chaperone system1. *Journal of molecular biology*, 286 (2):447–464.

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

Kudryavtsev, V., Sikor, M., Kalinin, S., Seidel, C., and Lamb, D. Combining mfd and pie for accurate fret measurements. *submitted*.

Lakowicz, J. and Masters, B. *Principles of fluorescence spectroscopy, second edition*. Springer, 1999.

Lamb, D., Schenk, A., Röcker, C., Scalfi-Happ, C., and Nienhaus, G., 2000. Sensitivity enhancement in fluorescence correlation spectroscopy of multiple species using time-gated detection. *Biophysical journal*, 79(2):1129–1138.

Laminet, A., Ziegelhoffer, T., Georgopoulos, C., and Plückthun, A., 1990. The escherichia coli heat shock proteins groel and groes modulate the folding of the beta-lactamase precursor. *The EMBO Journal*, 9(7):2315.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M., and Hartl, F., 1992. Successive action of dnak, dnaj and groel along the pathway of chaperone-mediated protein folding. *Nature*, 356 (6371):683–689.

Levinthal, C., 1969. How to fold graciously. *Mossbauer spectroscopy in biological systems*, pages 22–24.

Lipman, E., Schuler, B., Bakajin, O., and Eaton, W., 2003. Single-molecule measurement of protein folding kinetics. *Science*, 301(5637):1233.

Lisse, T. and Schwarz, E., 2000. Functional specificity of the mitochondrial dnaj protein, mdj1p, in saccharomyces cerevisiae. *Molecular and General Genetics MGG*, 263(3):527–534.

Liu, Q., D'Silva, P., Walter, W., Marszalek, J., and Craig, E., 2003. Regulated cycling of mitochondrial hsp70 at the protein import channel. *Science*, 300(5616):139.

Magde, D., Elson, E., and Webb, W. W., 1972. Thermodynamic fluctuations in a reacting system: Measurement by fluorescence correlation spectroscopy. *Physical Review Letters*, 29(11):705.

Magde, D., Elson, E. L., and Webb, W. W., 1974. Fluorescence correlation spectroscopy. ii. an experimental realization. *Biopolymers*, 13(1):29–61.

Mapa, K. *Conformational Dynamics of the Mitochondrial TIM23 Preprotein Translocase*. PhD thesis, LMU München, 2009.

Mapa, K., Sikor, M., Kudryavtsev, V., Waegemann, K., Kalinin, S., Seidel, C. A. M., Neupert, W., Lamb, D. C., and Mokranjac, D., 2010. The conformational dynamics of the mitochondrial hsp70 chaperone. *Molecular cell*, 38(1):89–100.

Marcinowski, M., Höller, M., Feige, M., Baerend, D., Lamb, D., and Buchner, J., 2011. Substrate discrimination of the chaperone bip by autonomous and cochaperone-regulated conformational transitions. *Nature Structural & Molecular Biology*.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A., and Hartl, F., 1991. Chaperonin-mediated protein folding at the surface of groel through a'molten globule'-like intermediate. *Nature*, 352(6330):36.

Mayer, M. and Bukau, B., 2005. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62(6):670–684.

Mayer, M. P., 2010. Gymnastics of molecular chaperones. *Molecular cell*, 39(3):321–331.

Mayhew, M., da Silva, A., Martin, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F., 1996. Protein folding in the central cavity of the groel-groes chaperonin complex. *Nature*, 379(6564): 420.

Mayor, U., Johnson, C. M., Daggett, V., and Fersht, A. R., 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proceedings of the National Academy of Sciences*, 97(25):13518–13522.

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

Meacham, G., Patterson, C., Zhang, W., Younger, J., and Cyr, D., 2001. The hsc70 co-chaperone chip targets immature cftr for proteasomal degradation. *Nature cell biology*, 3(1):100–104.

Mickler, M., Hessling, M., Ratzke, C., Buchner, J., and Hugel, T., 2009. The large conformational changes of hsp90 are only weakly coupled to atp hydrolysis. *Nature Structural & Molecular Biology*, 16(3):281–286.

Minsky, M. Microscopy apparatus, 1961.

Misselwitz, B., Staeck, O., Matlack, K., and Rapoport, T., 1999. Interaction of bip with the j-domain of the sec63p component of the endoplasmic reticulum protein translocation complex. *Journal of Biological Chemistry*, 274(29):20110.

Miyake, T., Tanii, T., Sonobe, H., Akahori, R., Shimamoto, N., Ueno, T., Funatsu, T., and Ohdomari, I., 2008. Real-time imaging of single-molecule fluorescence with a zero-mode waveguide for the analysis of protein- protein interaction. *Analytical chemistry*, 80(15):6018–6022.

Mokranjac, D. and Neupert, W., 2009. Thirty years of protein translocation into mitochondria: unexpectedly complex and still puzzling. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1793(1):33–41.

Mokranjac, D., Bourenkov, G., Hell, K., Neupert, W., and Groll, M., 2006. Structure and function of tim14 and tim16, the j and j-like components of the mitochondrial protein import motor. *The EMBO Journal*, 25(19):4675–4685.

Moro, F., Okamoto, K., Donzeau, M., Neupert, W., and Brunner, M., 2002. Mitochondrial protein import: molecular basis of the atp-dependent interaction of mthsp70 with tim44. *Journal of Biological Chemistry*, 277(9):6874.

Morozova-Roche, L. and Malisauskas, M., 2007. A false paradise - mixed blessings in the protein universe: The amyloid as a new challenge in drug development. *Current Medicinal Chemistry*, 14 (11):1221–1230.

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

Müller, C., Loman, A., Pacheco, V., Koberling, F., Willbold, D., Richtering, W., and Enderlein, J., 2008. Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL (Europhysics Letters)*, 83:46001.

Muñoz, V., 2007. Conformational dynamics and ensembles in protein folding. *Annual review of biophysics and biomolecular structure*, 36(1):395–412.

Murphy, K. Hidden markov model (hmm) toolbox for matlab, 1998.

Nettels, D., Hoffmann, A., and Schuler, B., 2008. Unfolded protein and peptide dynamics investigated with single-molecule fret and correlation spectroscopy from picoseconds to seconds. *The Journal of Physical Chemistry B*, 112(19):6137–6146.

Neupert, W. and Brunner, M., 2002. The protein import motor of mitochondria. *Nature Reviews Molecular Cell Biology*, 3(8):555–565.

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

Okumus, B., Wilson, T., Lilley, D., and Ha, T., 2004. Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophysical journal*, 87(4):2798–2806.

Ostermann, J., Horwich, A., Neupert, W., and Hartl, F., 1989. Protein folding in mitochondria requires complex formation with hsp60 and atp hydrolysis. *Nature*, 341(6238):125.

Parak, F. G. and Nienhaus, G. U., 2002. Myoglobin, a paradigm in the study of protein dynamics. *ChemPhysChem*, 3(3):249–254.

Perrin, F., 1934. Mouvement brownien d'un ellipsoide-i. dispersion diélectrique pour des molécules ellipsoidales. *Journal de Physique et le Radium*, 5(10):497–511.

Perrin, F., 1936. Mouvement brownien d'un ellipsoide (ii). rotation libre et dépolarisation des fluorescences. translation et diffusion de molécules ellipsoidales. *Journal de Physique et le Radium*, 7 (1):1–11.

Pfeil, S., Wickersham, C., Hoffmann, A., and Lipman, E., 2009. A microfluidic mixing system for single-molecule measurements. *Review of Scientific Instruments*, 80:055105.

Plaxco, K., Simons, K., and Baker, D., 1998. Contact order, transition state placement and the refolding rates of single domain proteins1. *Journal of molecular biology*, 277(4):985–994.

Plotkin, S. S. and Onuchic, J. N., 2002a. Understanding protein folding with energy landscape theory - part i: Basic concepts. *Quarterly reviews of biophysics*, 35(2):111–167.

Plotkin, S. S. and Onuchic, J. N., 2002b. Understanding protein folding with energy landscape theory - part ii: Quantitative aspects. *Quarterly reviews of biophysics*, 35(3):205–286.

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

Rabiner, L. and Juang, B., 1986. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4–16.

Rassow, J., Maarse, A., Krainer, E., Kübrich, M., Müller, H., Meijer, M., Craig, E., and Pfanner, N., 1994. Mitochondrial protein import: biochemical and genetic evidence for interaction of matrix hsp70 and the inner membrane protein mim44. *The Journal of cell biology*, 127(6):1547.

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

Rowley, N., Prip-Buus, C., Westermann, B., Brown, C., Schwarz, E., Barrell, B., and Neupert, W., 1994. Mdj1p, a novel chaperone of the dnaj family, is involved in mitochondrial biogenesis and protein folding. *Cell*, 77(2):249–259.

Saibil, H., 2008. Chaperone machines in action. *Current opinion in structural biology*, 18(1):35–42.

Sameshima, T., Iizuka, R., Ueno, T., and Funatsu, T., 2010. Denatured proteins facilitate the formation of the football-shaped groel-(groes) 2 complex. *The Biochemical journal*, 427(2):247.

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

Schenk, A., Ivanchenko, S., Röcker, C., Wiedenmann, J., and Nienhaus, G. U., 2004. Photodynamics of red fluorescent proteins studied by fluorescence correlation spectroscopy. *Biophysical journal*, 86(1):384–394.

Schlecht, R., Erbse, A., Bukau, B., and Mayer, M., 2011. Mechanics of hsp70 chaperones enables differential interaction with client proteins. *Nature Structural & Molecular Biology*, 18(3):345–351.

Schlüsche, P. *Untersuchung der Funktion und Dynamik von DNA-Transkriptionsfaktoren mittels Einzelmolekül-Fluoreszenzmikroskopie*. PhD thesis, LMU München, 2007.

Schneider, H., Berthold, J., Bauer, M., Dietmeier, K., Guiard, B., Brunner, M., and Neupert, W., 1994. Mitochondrial hsp70/mim44 complex facilitates protein import. *Nature*, 371(6500):768.

Schröder, G. and Grubmüller, H., 2003. Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments. *The Journal of chemical physics*, 119:9920.

Sela, M., White Jr, F., and Anfinsen, C., 1957. Reductive cleavage of disulfide bridges in ribonuclease. *Science*, 125(3250):691–692.

Selkoe, D., 2004. Cell biology of protein misfolding: the examples of alzheimer's and parkinson's diseases. *Nature cell biology*, 6(11):1054–1061.

Sharff, A., Rodseth, L., Spurlino, J., and Quiocho, F., 1992. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry*, 31(44):10657–10663.

Sharma, S., Chakraborty, K., Muller, B., Astola, N., Tang, Y., Lamb, D., Hayer-Hartl, M., and Hartl, F., 2008. Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell*, 133(1):142–153.

Sichting, M., Mokranjac, D., Azem, A., Neupert, W., and Hell, K., 2005. Maintenance of structure and function of mitochondrial hsp70 chaperones requires the chaperone hep1. *The EMBO Journal*, 24(5):1046–1056.

Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Müller, S., and Seidel, C., 2011. Accurate distance determination of nucleic acids via förster resonance energy transfer: Implications of dye linker length and rigidity. *Journal of the American Chemical Society*.

Sirangelo, I., Malmo, C., Iannuzzi, C., Mezzogiorno, A., Bianco, M. R., Papa, M., and Irace, G., 2004. Fibrillogenesis and cytotoxic activity of the amyloid-forming apomyoglobin mutant w7fw14f. *Journal of Biological Chemistry*, 279(13):13183–13189.

Slutsky-Leiderman, O., Marom, M., Iosefson, O., Levy, R., Maoz, S., and Azem, A., 2007. The interplay between components of the mitochondrial protein translocation motor studied using purified components. *Journal of Biological Chemistry*, 282(47):33935.

Strickler, S. and Berg, R., 1962. Relationship between absorption intensity and fluorescence lifetime of molecules. *The Journal of chemical physics*, 37(4):814.

Suzuki, M., Ueno, T., Iizuka, R., Miura, T., Zako, T., Akahori, R., Miyake, T., Shimamoto, N., Aoki, M., Tanii, T., Ohdomari, I., and Funatsu, T., 2008. Effect of the c-terminal truncation on the functional cycle of chaperonin groel. *Journal of Biological Chemistry*, 283(35):23931–23939.

Swain, J. F. and Gierasch, L. M., 2006. The changing landscape of protein allostery. *Current opinion in structural biology*, 16(1):102–108.

Swain, J. F., Dinler, G., Sivendran, R., Montgomery, D. L., Stotz, M., and Gierasch, L. M., 2007. Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Molecular cell*, 26(1):27–39.

Taguchi, H., Ueno, T., Tadakuma, H., Yoshida, M., and Funatsu, T., 2001. Single-molecule observation of protein-protein interactions in the chaperonin system. *Nature biotechnology*, 19(9): 861–865.

Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M., Hartl, F., and Hayer-Hartl, M., 2006. Structural features of the groel-groes nano-cage required for rapid folding of encapsulated protein. *Cell*, 125(5):903–914.

Tang, Y., Chang, H., Chakraborty, K., Hartl, F., and Hayer-Hartl, M., 2008. Essential role of the chaperonin folding compartment in vivo. *The EMBO Journal*, 27(10):1458–1468.

Teale, F. W. J., 1959. Cleavage of the haem-protein link by acid methylethylketone. *Biochimica et Biophysica Acta*, 35:543–543.

Tissieres, A., Mitchell, H., and Tracy, U., 1974. Protein synthesis in salivary glands of drosophila melanogaster: Relation to chromosome puffs* 1. *Journal of molecular biology*, 84(3):389–392.

Tokuriki, N. and Tawfik, D., 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459(7247):668–673.

Tyagi, N. K., Fenton, W. A., Deniz, A. A., and Horwich, A. L., 2011. Double mutant mbp refolds at same rate in free solution as inside the groel/groes chaperonin chamber when aggregation in free solution is prevented. *FEBS Letters*, 585(12):1969–1972.

Ueno, T., Taguchi, H., Tadakuma, H., Yoshida, M., and Funatsu, T., 2004. Groel mediates protein folding with a two successive timer mechanism. *Molecular cell*, 14(4):423–434.

Uversky, V. and Fink, A., 2004. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics*, 1698(2):131–153.

Viani, M., Pietrasanta, L., Thompson, J., Chand, A., Gebeshuber, I., Kindt, J., Richter, M., Hansma, H., and Hansma, P., 2000. Probing protein-protein interactions in real time. *Nature Structural & Molecular Biology*, 7(8):644–647.

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

von Ahsen, O., Voos, W., Henninger, H., and Pfanner, N., 1995. The mitochondrial protein import machinery. *Journal of Biological Chemistry*, 270(50):29848.

Voos, W. and Röttgers, K., 2002. Molecular chaperones as essential mediators of mitochondrial biogenesis. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1592(1):51–62.

Voos, W., von Ahsen, O., Müller, H., Guiard, B., Rassow, J., and Pfanner, N., 1996. Differential requirement for the mitochondrial hsp70-tim44 complex in unfolding and translocation of preproteins. *The EMBO Journal*, 15(11):2668.

Walsh, P., Bursac, D., Law, Y. C., Cyr, D., and Lithgow, T., 2004. The j-protein family: modulating protein assembly, disassembly and translocation. *EMBO reports*, 5(6):567–571.

Watson, J. and Crick, F., 1953. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738.

Wegele, H., Müller, L., and Buchner, J., 2004. Hsp70 and hsp90 - a relay team for protein folding. *Ergebnisse der Physiologie, biologischen Chemie und experimentellen Pharmakologie*, 151(1):1–44.

Weiss, C., Niv, A., and Azem, A., 2002. Two-step purification of mitochondrial hsp70, ssc1p, using mge1 (his) 6 immobilized on ni-agarose. *Protein Expression and Purification*, 24(2):268–273.

Weissman, J., Rye, H., Fenton, W., Beechem, J., and Horwich, A., 1996. Characterization of the active intermediate of a groel-groes-mediated protein folding reaction. *Cell*, 84(3):481–490.

Westermann, B., Gaume, B., Herrmann, J., Neupert, W., and Schwarz, E., 1996. Role of the mitochondrial dnaj homolog mdj1p as a chaperone for mitochondrially synthesized and imported proteins. *Molecular and Cellular Biology*, 16(12):7063.

Wetlaufer, D., 1962. Ultraviolet spectra of proteins and amino acids. *Advances in Protein Chemistry*, 17:303–390.

Widengren, J., Rigler, R., and Mets, Ü., 1994. Triplet-state monitoring by fluorescence correlation spectroscopy. *Journal of Fluorescence*, 4(3):255–258.

Widengren, J., Mets, Ü., and Rigler, R., 1995. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. *The Journal of Physical Chemistry*, 99(36): 13368–13379.

Widengren, J., Mets, Ü., and Rigler, R., 1999. Photodynamic properties of green fluorescent proteins investigated by fluorescence correlation spectroscopy. *Chemical Physics*, 250(2):171–186.

Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Röcker, C., Salih, A., Spindler, K.-D., and Nienhaus, G. U., 2004. Eosfp, a fluorescent marker protein with uv-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45):15905–15910.

Wu, Y., Kondrashkina, E., Kayatekin, C., Matthews, C., and Bilsel, O., 2008. Microsecond acquisition of heterogeneous structure in the folding of a tim barrel protein. *Proceedings of the National Academy of Sciences*, 105(36):13367.

Xu, Z., Horwich, A. L., and Sigler, P. B., 1997. The crystal structure of the asymmetric groel-groes-(adp)7 chaperonin complex. *Nature*, 388(6644):741–750.

Yamasaki, R., Hoshino, M., Wazawa, T., Ishii, Y., Yanagida, T., Kawata, Y., Higurashi, T., Sakai, K., Nagai, J., and Goto, Y., 1999. Single molecular observation of the interaction of groel with substrate proteins. *Journal of molecular biology*, 292(5):965–972.

Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T., Rech, I., Cova, S., Xun, L., and Xie, X., 2003. Protein conformational dynamics probed by single-molecule electron transfer. *Science*, 302 (5643):262.

Yoneda, T., Benedetti, C., Urano, F., Clark, S., Harding, H., and Ron, D., 2004. Compartment-specific perturbation of protein handling activates genes encoding mitochondrial chaperones. *Journal of Cell Science*, 117(18):4055.

Young, J., Agashe, V., Siegers, K., and Hartl, F., 2004. Pathways of chaperone-mediated protein folding in the cytosol. *Nature Reviews Molecular Cell Biology*, 5(10):781–791.

Zelent, B., Kusba, J., Gryczynski, I., Johnson, M. L., and Lakowicz, J. R., 1998. Time-resolved and steady-state fluorescence quenching of n-acetyl-tryptophanamide by acrylamide and iodide. *Biophysical Chemistry*, 73(1-2):53–75.

Zhu, X., Zhao, X., Burkholder, W., Gragerov, A., Ogata, C., Gottesman, M., and Hendrickson, W., 1996. Structural analysis of substrate binding by the molecular chaperone dnak. *Science*, 272 (5268):1606.

A Technical details

A.1 TCSPC parameters

For the TCSPC cards used in this work, a number of parameters has to be set to ensure correct detection of the photon arrival times. The parameter settings depend strongly on the type of detector used, and optimal parameters may even vary for different detectors of the same type. Here, for all detectors a common set of parameters could be used. The settings are given in Table A.1.

The APDs generate a voltage pulse with a varying amplitude due to the stochastic mechanism of signal amplification. Using a simple thresholding mechanism, it would not be possible to assign an arrival time with sub-nanosecond time resolution to each photon due to the different pulse heights. Thus, a *constant fraction discriminator* (CFD) is used to determine the arrival time. The CFD triggers at a given constant fraction of the overall signal amplitude. It determines the arrival time from the zero-crossing point of the sum of the signal with the inverted and shifted signal (Becker *et al.*, 2009). The time of the zero-crossing point is independent of the pulse amplitude and thus, this timing method minimizes the timing jitter caused by the amplitude jitter of the detector pulses.

The threshold for the zero-crossing detection is given by the *CFD ZC Level* parameter. The *CFD Limit Low* parameter sets an additional threshold to discard small background pulses and noise. The same procedure is applied to detect the arrival time of the Sync pulse. An additional parameter allows division of the Sync frequency by a constant factor (*Sync frequency divider*).

The *Time to Amplitude Converter* (TAC) is used to determine the time between the photon detection and the Sync pulse detection. A voltage ramp is started by the detection of the photon signal, and

Parameter	Setting
Operation Mode	FIFO
Trigger	Falling edge
CFD Limit Low	-100 mV
CFD ZC Level	-12.09 mV
Dither range	1/16
TAC Range	80 ns
TAC Gain	2
TAC Offset	3.14%
TAC Limit Low	0%
TAC Limit High	100 %
Sync ZC Level	-12.85 mV
Sync Frequency devider	1
Sync Threshold	-101.96 mV
Settings	Common

Table A.1: *List of TCSPC parameter values typically used for data collection*

Parameter	Setting
Acquisition mode	Kinetic
Triggering	Internal
Exposure time	30-500 ms
Kinetic series length	1000-2000
Number of prescans	0
Frame transfer	on
Vertical shift speed	0.6 µs
Vertical clock voltage amplitude	normal
Readout rate	10 MHz at 14 bit
Pre-amplifier gain	2.4×
Electron multiplier gain	300
Image size	512 × 512

Table A.2: *EMCCD settings*

after stopping by the Sync signal, the voltage amplitude is registered. The duration of the voltage ramp is given by the product of the *TAC range* and the *TAC gain*. The *TAC offset* adds a constant offset to the TAC signal, thus shifting the detected TCSPC pattern in time. Similar to the CFD limit low parameter, the parameters *TAC limit low* and *TAC limit high* reject pulses outside a given window of the TAC range.

To ensure the synchronized start of the measurement in all four detection cards, an external trigger is used to start the measurement. The trigger signal is generated by a National Instruments USB-6002 device (*falling edge*).

A.2 EMCCD parameters

For the detection of single molecules, a high sensitivity of the camera is required. Regular CCD cameras are limited by the readout noise of the CCD chip which is on the order of 10 electrons. In EMCCD cameras, the charge per pixel is amplified in a series of amplifying registers, thus increasing the number of electrons per pixel above the level of the readout noise even for light levels of few or one photons per pixel. The images in this work were acquired using frame transfer. There, after exposure of an image in the illuminated area of the CCD chip, the image is shifted vertically to a covered area of the CCD chip. This vertical shift is fast and allows quick re-exposure of another image while the charge of the last image is still read out.

A.3 Spectra

Careful selection of the emission filters is necessary to maximize the detection efficiency of the setup for the selected dyes, while at the same time minimizing the amount of contamination by spectral crosstalk and scattered laser light. Figures A.1 and A.2 show the normalized absorption and emission spectra of the dye pairs used for spFRET experiments throughout this thesis, Atto 532 and Atto 647N together with the emission filters, dichroic mirrors, and laser lines used in the MFD-PIE setup (Fig. A.1) and in the TIRF setup (Fig. A.2).

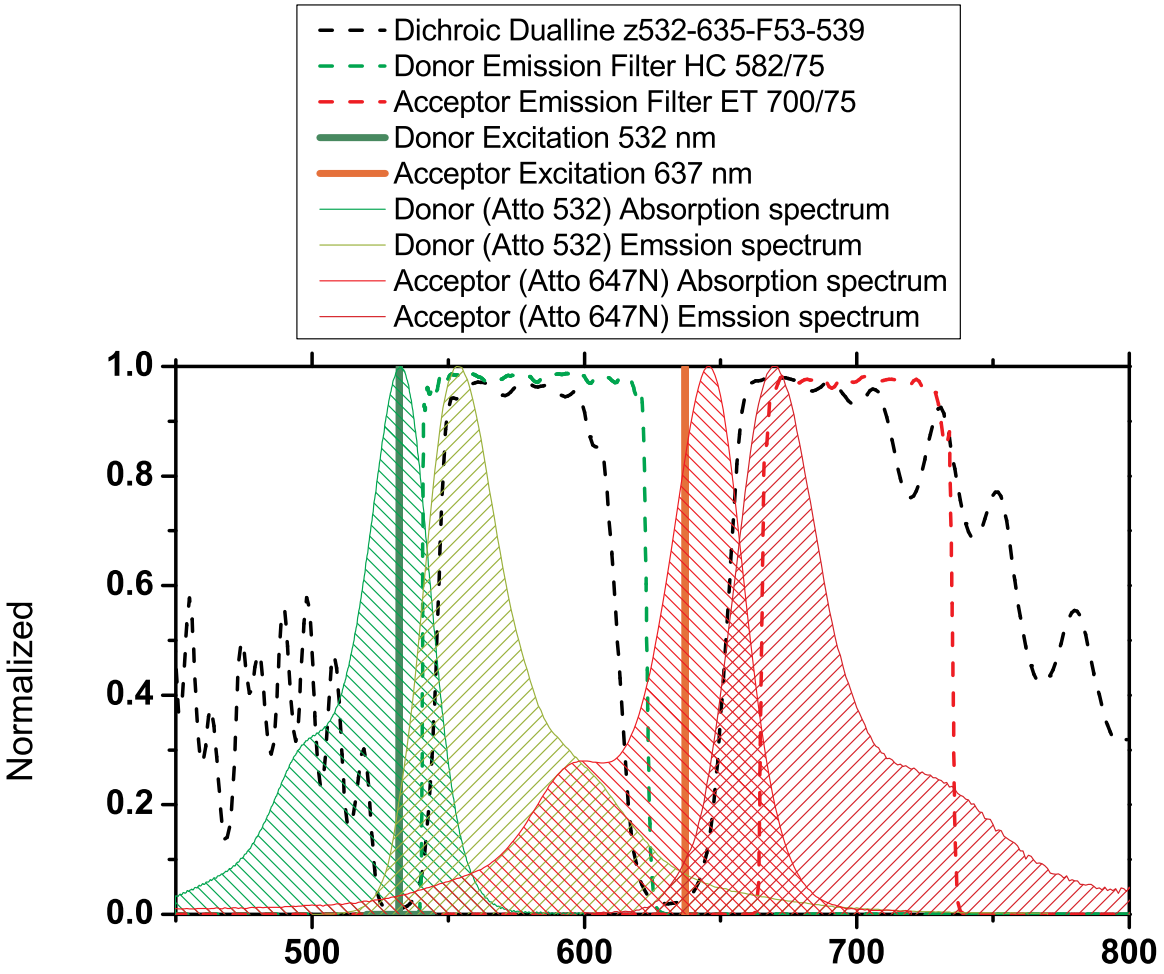


Figure A.1: *Spectra for the MFD-PIE setup. The normalized absorption and emission spectra of Atto 532 and Atto 647N are shown in green and red together with the transmission spectra of the dichroic mirror Dualline z532-635-F53-539 (black) and the emission filters for Atto 532 (HQ 582/75, dashed green line) and Atto 647N (ET 700/75, dashed red line). The laser lines used for excitation are indicated by vertical lines in green ($\lambda = 532\,\mathrm{nm}$) and red ($\lambda = 637\,\mathrm{nm}$).*

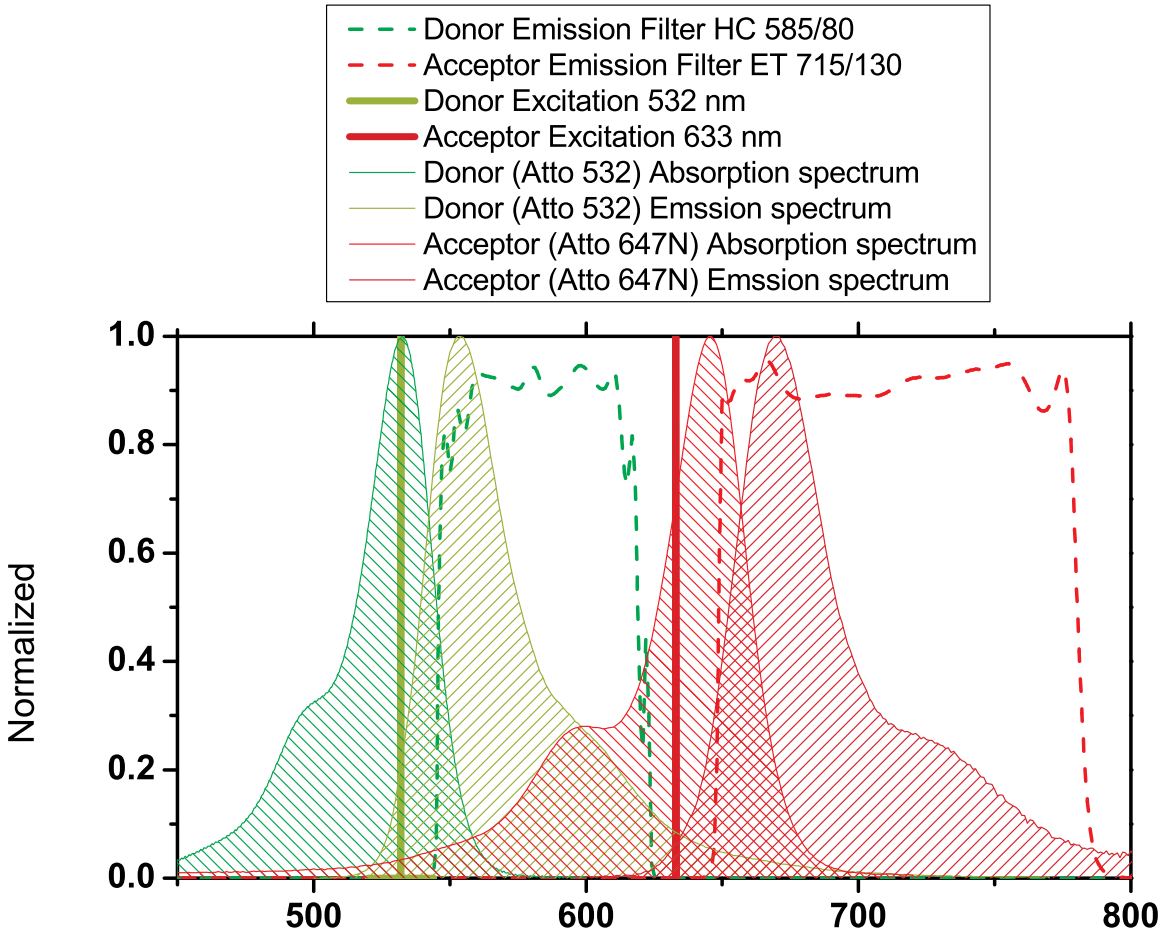


Figure A.2: *Spectra for the TIRF setup. The normalized absorption and emission spectra of Atto 532 and Atto 647N are shown in green and red together with the transmission spectra of the emission filters for Atto 532 (HC585/80, dashed green line) and Atto 647N (ET 715/130, dashed red line). The laser lines used for excitation are indicated by vertical lines in green ($\lambda = 532\,\mathrm{nm}$) and red ($\lambda = 633\,\mathrm{nm}$).*

List of Abbreviations

2cFCCS	Two-color fluorescence cross-correlation spectroscopy, page 39
2fFCS	Two focus Fluorescence Correlation Spectroscopy, page 36
ACF	Autocorrelation function, page 38
ADP	Adenosine 5'-diphosphate, page 68
ADR	Acceptor-to-donor ratio, page 67
ALEX	Alternating laser excitation, page 17
AMP-PNP	Adenosine 5'-(β,γ-imido)triphosphate, page 68
AOTF	Acousto-optical tunable filter, page 40
APBS	All photon burst search, page 21
APD	Avalanche photo diode, page 16
ATP	Adenosine 5'-triphosphate, page 68
BiP	Immunoglobulin Binding Protein, page 61
CCF	Cross-correlation function, page 38
CD	Circular dichroism, page 113
CFD	Constant fraction discriminator, page 19
CT	Crosstalk, page 22
DE	Direct excitation, page 22
DM	Dichroic mirror, page 18
DM-MBP	Double-mutant maltose binding protein, page 108
DNA	Deoxyribonucleic acid, page 1
DOPC	1,2-Dioleoyl-sn-glycero-3-phosphocholine, page 42
EF	Emission filter, page 18
EMCCD	electron multiplying charged coupled device, page 41
ER	endoplasmic reticulum, page 14
FCCS	Fluorescence cross-correlation spectroscopy, page 39
FCS	Fluorescence correlation spectroscopy, page 33
FRET	Förster resonance energy transfer, page 5
GFP	Green fluorescent protein, page 23
GuHCL	Guanidine hydrochloride, page 54
HEPES	2-(4-(2-Hydroxyethyl)- 1-piperazinyl)-ethansulfonsäure, page 54
HMM	Hidden markov modeling, page 45
HPLC	High pressure liquid chromatography, page 113
IRF	Instrument response function, page 23
J protein partner of Ssc1 for protein folding in the mitochondrial matrix a, page 63	
LUVs	Large unilamellar vesicles, page 44

MBP	Maltose binding protein, page 108
MD	Molecular dynamics, page 51
MFD	Multi-parameter Fluorescence Detection, page 15
Mge1	Nucleotide exchange factor for Ssc1, page 63
MLE	Maximum likelihood estimator, page 22
MLVs	Multilamellar vesicles, page 42
NA	Numerical aperture, page 16
NBD	Nucleotide binding domain, page 61
NEF	Nucleotide exchange factor, page 61
NHS	N-Hydroxysuccinimid, page 54
NMR	Nuclear magnetic resonance, page 1
P5	Model substrate peptide for Hsp70s, sequence CALLLSAPRR, page 73
PBS	Polarizing beam splitter, page 18
PDA	Probability distribution analysis, page 32
PDB	Protein data bank, page 12
PEG	Polyethylene glycol, page 40
PH	Pinhole, page 37
PIE	Pulsed interleaved excitation, page 15
PMT	Photomultiplier tube, page 31
PSF	Point spread function, page 16
RCMLA	Reduced carboxymethylated α-lactalbumin, page 73
RNA	Ribonucleic acid, page 13
SBD	Substrate binding domain, page 61
spFRET	Single-pair Förster resonance energy transfer, page 2
Ssc1	The major mitochondrial heat shock protein 70, page 62
TAC	Time-to-amplitude converter, page 19
TCSPC	time-correlated single photon couting, page 18
TDP	Transition density plot, page 48
TIMM23	Translocase of the inner mitochondrial membrane 23, page 63
TIRF	Total internal reflection, page 40
TL	Tube lens, page 37
Trp	Tryptophan, page 113
TTL	Transistor-transistor logic, page 18
TTTR,T3R	Time-tagged time resolved, page 19

Publications

Kudryavtsev, V., **Sikor M.**, Kalinin, S., Seidel, C.A.M., and Lamb D.C., *Combining Multiparameter Fluorescence Detection and Pulsed Interleaved Excitation for accurate single-pair FRET measurements*, submitted.

Chakraborty K., Chatila M., Sinha J., Shi Q., Poschner B.C., **Sikor M.**, Jiang G., Lamb D.C., Hartl F.U. and Hayer-Hartl M., *Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding* (2010), **Cell**, 142(1), 112-122.

Sikor M., Sabin J., Keyvanloo A., Schneider M.F., Thewalt J.L., Bailey A.E. and Frisken B.J., *Interaction of a Charged Polymer with Zwitterionic Lipid Vesicles* (2010), **Langmuir**, 26(6), 4095-4102.

Mapa K., **Sikor M.**, Kudryavtsev V., Waegemann K., Kalinin S., Seidel C.A.M., Neupert N., Lamb D.C. and Mokranjac D., *The Conformational Dynamics of the Mitochondrial Hsp70 Chaperone* (2010), **Molecular Cell**, 38(1), 89-100.

Griepernau, B., Leis, S., Schneider, M.F., **Sikor, M.**, Steppich, D., Böckmann, R., *1-Alkanols and membranes - a story of attraction*, **Biochim. Biophys. Acta (2007)**, doi:10.1016/j.bbamem.2007.08.002.

Bismuto E., Di Maggio E., Pleus S., **Sikor M.**, Röcker C., Nienhaus G.U. and Lamb D.C., *Molecular dynamics simulation of the acidic compact state of apomyoglobin from yellowfin tuna* (2008), **Proteins: Structure, Function, and Bioinformatics**,74(2), 273-290.

Scientific meetings and presentations

03/2011	Talk at the Annual Meeting of the Biophysical Society: *Conformational Dynamics of Mitochondrial Hsp70*
09/2010	Talk at the workshop on "Single Molecule Spectroscopy and Ultra Sensitive Analysis in the Life Sciences": *The Conformational Dynamics of the Mitochondrial Hsp70 Chaperone*
03/2010	Poster at the Annual Meeting of the Biophysical Society: *The Conformational Cycle of Mitochondrial Hsp70*
01/2007	Talk at biophysics seminar of the Simon Fraser University, Vancouver, Canada: *From blood flow to the thermodynamics of membranes - from Augsburg to SFU*

Acknowledgements

Looking back over the last years, I am very grateful for all the support I have received from many people. This thesis would not have been possible without their support, and I would like to express my deep gratitude to everyone who contributed to the successful completion of this work.

The person who had most influence on this thesis and on my scientific development in these years is my advisor, Prof. Don C. Lamb. I am impressed by his ability to see the smallest details without losing track of the bigger picture, and his joy of doing science has always been an example to me.

Prof. Christoph Bräuchle has provided much of the great infrastructure that allowed me to do the research presented here. I am also grateful for his support in the IDK-NBT as my second advisor.

The research presented in this thesis was the result of great collaborations. The work with Koyeli Mapa and Dejana Mokranjac was a excellent example for a perfect collaboration; they allowed me to focus on my part of the work by taking perfect care of the biological part of the mtHsp70 project. I am grateful to Stanislav Kalinin, Suren Felekyan and Claus Seidel for giving us access to their analysis methods and for the advice on how to use them. The GroEL project would not have been possible without the work of Kausik Chakraborty, Jyoti Sinha, Manal Chatila, Manajit Hayer-Hartl and Ulrich Hartl. I am also grateful to Ettore Bismuto and Emiddio di Maggio for giving me the opportunity to be part of the Myoglobin project, and to Jörg Enderlein for his advice regarding the two-focus FCS experiments.

I have also been so lucky to be part of a great group that made bad times looking better, and good times looking fablabrific. I am thankful to everyone in the lab for the great times we had on various trips and coffee breaks, for the wonderful atmosphere, for the fact that when I had a question there was always someone able to answer it, and for always taking the time to do so. I enjoyed it so much to work closely with Matthias from the first day of my PhD on, discussing technical details, pair programming, procrastinating, and having identical thoughts. I am grateful to Gregor for his advice on TIRF and programming, and for the many discussions we had on topics as different as hidden markov models, diets, energy storage, and anything else that matters in life. He and my other office mates, Aurelie, Jens, and Iko, were always there to help me when I had a question about science, or to join me in the kitchen when I needed a break. Without Niko and his efforts to keep things running, from new setups to computers and, most important, out-of-lab activities, the group wouldn't be the same. I thank Doro for sharing humor, laughing and lots of coffee with me, Vova for his terrific programming skills that made my life so much easier sometimes, my students Martina and Mangai for their help with various experiments, Moritz for his skilled help with administrative paperwork and presentation style, and Axel Gersdorf, Ralf Hiermaier and the other people in the electronics and fine-mechanical workshops for their quick and professional help with all the small technical problems.

I am very grateful to my parents for their lifelong support, and I thank my brother for setting an example of an amazing scientist to me. Finally, I thank Olli for always being there for me as the best part of my life.