Investigation of a Folding Intermediate of RNase H Using Single Molecule FRET Spectroscopy

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des Karlsruher Instituts
für Technologie (KIT)

genehmigte

DISSERTATION

von

Robert Frank Rieger
aus Stuttgart – Bad Cannstatt

Tag der mündlichen Prüfung: 02.11.2012

Referent: Prof. Dr. G. U. Nienhaus

Korreferent: apl. Prof. Dr. W. Wenzel

Hiermit versichere ich, dass ich die vorliegende Doktorarbeit selbstständig verfasst und nach bestem Wissen und Gewissen keine als die angegebenen Quellen oder Hilfsmittel verwendet habe.

Robert Rieger
Freiburg i. Breisgau, 25.05.2012

Contents

1 Introduction **9**
- 1.1 Proteins and Protein Folding . . . 9
  - 1.1.1 A Dogma and a Paradox . . . 9
  - 1.1.2 Diffusion–Collision Models . . . 11
  - 1.1.3 The Folding Funnel . . . 12
  - 1.1.4 The Unfolded State of Proteins . . . 14
  - 1.1.5 Studies with Single-Molecule Resolution . . . 15
  - 1.1.6 Observing Protein Folding of Single Molecules . . . 16
- 1.2 RNase H . . . 16
- 1.3 Outline . . . 17

2 Theory **19**
- 2.1 The Phenomenon of Fluorescence . . . 19
  - 2.1.1 The Jablonski Diagram . . . 19
  - 2.1.2 Parameters of Fluorescence . . . 25
  - 2.1.3 Photobleaching . . . 26
  - 2.1.4 Dependence of the Fluorescence on the Environment . . . 27
  - 2.1.5 Trp as an Internal Fluorescent Probe . . . 27
- 2.2 FRET – Förster Resonance Energy Transfer . . . 28
  - 2.2.1 Introduction and Principle . . . 28
  - 2.2.2 Theoretical Calculation of the FRET efficiency . . . 30
  - 2.2.3 Experimental Determination of the FRET efficiency . . . 30
  - 2.2.4 Hidden Assumptions . . . 32
  - 2.2.5 The Solvent Dependence of the Förster Distance . . . 33
  - 2.2.6 Samples with Variable Inter-Dye Distance and Slow Dye Dynamics 37
- 2.3 Confocal Microscopy . . . 40
  - 2.3.1 The Confocal Principle . . . 40
  - 2.3.2 A Typical Confocal Microscopy Setup . . . 40
- 2.4 Single Molecule FRET measurements . . . 42
  - 2.4.1 Burst Detection . . . 42
  - 2.4.2 Selection of Bright Bursts . . . 45
  - 2.4.3 Burst Wise FRET Calculation . . . 47
  - 2.4.4 The Donor-Only Problem . . . 47

2.4.5 Donor-Only Corrected smFRET Histograms 51
2.4.6 Correction for Spectral Crosstalk and Direct Acceptor Excitation . 51
2.4.7 Determination of the γ Factor 54
2.5 Boltzmann Statistics 58
2.6 RNase H 59

3 Materials and Methods 63
3.1 Protein Labeling 63
3.2 Buffer Solutions 64
3.3 Cover Slip Coating 65
3.4 Sample Preparation for smFRET measurements 66
3.5 Microscope Systems 67
3.5.1 The M2 Setup 67
3.5.2 MicroTime 200 67
3.6 Absorption Measurements 68
3.7 Fluorescence Measurements 68
3.8 Quantum Yield and Total Detection Efficiency Measurements 68
3.9 Fitting the FRET histograms 69
3.9.1 Parameter Weighting 69
3.9.2 Variable Scaling 70
3.9.3 Error Determination 71

4 Experiments and Discussion 73
4.1 Determination of the γ factor 73
4.2 smFRET Measurements at pH 7.2 76
4.2.1 A Two-State Approach 77
4.2.2 Global Coupling of the Fractional Populations 79
4.2.3 Fractional Populations in the Two-State Model 80
4.2.4 Fixing the Folded State Distribution 81
4.2.5 Results of Fitting with the Two-State Model 82
4.2.6 Fixing the Width of the Unfolded State Distribution 82
4.2.7 Introduction of an Intermediate State 84
4.2.8 Fractional Populations in the Three-State Model 85
4.2.9 Results of Fitting with the Three-State Model 87
4.2.10 Comparison between the Two-State and the Three-State Model . . 87
4.3 smFRET Measurements at pH 3.0 90
4.3.1 Trp Fluorescence Measurements 92
4.3.2 Combined Fitting of Trp Fluorescence and smFRET Data 94
4.3.3 Breakdown of the Two-State Model 98
4.4 Discussion of the Free Energy Differences 99

5 Summary 105

Bibliography **107**

Appendix **126**
List of Publications 126
Curriculum Vitae 127
Danksagungen 128

Chapter 1

Introduction

1.1 Proteins and Protein Folding

1.1.1 A Dogma and a Paradox

It is no exaggeration to say that without proteins, life as we know it would not be possible. Proteins are, what Feynman meant when he talked about a "billion tiny factories" in his now famous talk at the annual meeting of the American Physical Society in 1959[1], and what has since simply been called nanomachines. As the name "nanomachine" implies, proteins are generally not much bigger than a few nanometers and they manipulate matter on an atomistic scale, just as Feynman envisioned. Nature employs them for nearly any task in living organisms: Proteins function as motors and rotors, provide scaffolding and structural support to cells, act as parts of signaling chains and pumps and catalyze all kinds of chemical reactions. They are involved in transport processes and regulation loops and they read, write and duplicate their own blueprints, the DNA. Proteins themselves are assembled by other proteins as well... and the list of protein functions goes on and on[2,3].

From a chemical perspective, proteins are linear biopolymers. The building blocks of proteins are a set of 20 naturally occurring amino acids, linked together covalently by peptide bonds between the carboxyl and the amino groups of adjacent residues. The characteristic sequence of amino acids that defines a protein is encoded in the DNA of an organism and is called its primary structure. Interactions of amino acids in close proximity along the backbone, especially through hydrogen bonds, lead to the formation of secondary structure. Segments of the polypeptide chain form spiral-shaped α-helices and flat β-sheets. These secondary structure elements position themselves in a well defined orientation to one another, an arrangement called the tertiary structure of a protein. The tertiary structure is mainly stabilized by the formation of a hydrophobic core, salt bridges, hydrogen bonds or disulfide bonds and van der Waals interactions. The tertiary structure, also called the fold or the native conformation of a protein, is necessary for its function. The grooves and cavities on the surface of a protein, for example, enable it to specifically recognize particular co-reactants.

The question of how a newly synthesized protein chain assumes this well defined native fold has been the topic of intense debate and research. Small globular proteins are able to fold independently and reversibly. According to the *thermodynamic hypothesis* this means that all the information about the conformation must be contained already in the amino acid sequence and that, under folding conditions, the native structure is a unique, stable and kinetically accessible minimum of the free energy. This hypothesis is also called *Anfinsen's dogma*, after Nobel laureate Christian B. Anfinsen[4].

Related to this thought is Levinthal's paradox. The number of conformations that an amino acid chain can assume grows exponentially with the number of its residues. A short chain of 100 amino acids and 9 conformations per amino acid (2 dihedral angles with 3 possible orientations each) has access to the astronomically high number of $9^{100} \approx 10^{95}$ possible conformations. Even with a chain reconfiguration that is fast enough to sample one conformation each picosecond, it would take $> 10^{83}\,\mathrm{s}$ to completely sample the whole conformational space, a time much longer than the estimated age of the universe ($\approx 10^{17}\,\mathrm{s}$). The "paradox" lies in the fact that, in reality, such a globular protein of 100 amino acids would most likely fold on a time scale of milliseconds or even microseconds[5,6]. Levinthal's paradox is considered more an educational thought experiment than a real scientific conundrum. It illustrates that a protein does not randomly search its conformational space until it stumbles upon the correct fold and it points out why brute force computational approaches for protein structure determination are doomed to fail.

As a solution to this paradox, Levinthal concluded that proteins fold through a succession of intermediate states[7,8]. The existence of such intermediate states has indeed been confirmed by experiment[9]. However, this idea only shifts the question to how the folding process reaches the first intermediate state. A clue to another possible solution to the paradox is hinted at by Dawkins in his book "The blind watchmaker"[10]. While discussing the similar problem of evolutionary progress by accumulation of small changes, he asks the following question: How many keystrokes would it take a monkey, hammering away randomly on a typewriter to produce a short, specific quote from Shakespeare's Hamlet. Similar to Levinthal's random search, this number is astronomically large ($\approx 10^{40}$). However, if the search is modified in such a way that the monkey cannot change letters that are already correctly in place, the phrase may be reached by a random search in only a few thousand keystrokes[6].

Levinthal proposed the idea of a specific folding pathway, a series of steps all protein molecules must follow to arrive at the folded state. After Levinthal's argument, several groups set out to explore the mechanism of protein folding experimentally[11,12]. A main goal of early experimental work was to find and observe possible intermediate states along the hypothetical folding pathways, in the hopes that these intermediates would hold the key to the secret of how proteins could fold so quickly. These early ensemble experiments used single or multi-exponential decays of some optical property to monitor the kinetics of protein folding. Single exponential folding kinetics were attributed to 'two-state' folding behavior. To account for multi exponential behavior, additional states were added to the folding models, called folding intermediates. As time went on, more sophisticated experimental techniques were added to the arsenal of protein researchers, such as high-resolution hydrogen exchange[13–18], and later pulsed hydrogen exchange[19–21], mass spectrometry[22–24] and mutational studies[25–28]. Laser-triggered methods even allowed the exploration of very early events in protein folding[29–31]. During the course of these investigations of protein folding, it became clear that there is a class of fast-folding two-state proteins that fold without ever accumulating significant amounts of intermediate populations[32–36]. New kinds of models were needed to account for these molecules, which folded without following a series of predetermined intermediate steps on their way.

1.1.2 Diffusion–Collision Models

One of the first attempts to develop a theory that explained the experimental results of intermediate-free two-state folding and at the same time 'solved' Levinthal's paradox were so-called *diffusion–collision models*, which introduced the concept of native contacts[37–40]. A native contact is formed by two amino acids which are not close to each other in the amino acid sequence, but are spatially close in the tertiary structure of the native state of the protein. If such native contacts are energetically favorable, they stabilize the chain and serve as "nucleation points in the folding process", limiting the accessible conformation space for subsequent folding. In this limited conformation space, additional native contacts are formed with even higher probability, thus accelerating the collapse of the protein chain into its native conformation[41]. The folding rate of a protein depends on the closeness of the contact-forming residues in the primary structure. Whereas fast folders usually have a high degree of mostly local structures such as helices and tight turns, slow folders have more nonlocal structure such as β-sheets[42,43], although there are exceptions[44]. This idea that the folding rate of a protein is determined by topological properties of its native structure was pioneered by Plaxco, Simons and Baker[43,45,46] and has since been confirmed. A number of topological parameters have been proposed that correlate with the speed of folding of a protein, such as the effective length of the chain of a protein (chain length minus the number of amino acids in helices)[47], secondary structure length[48], the number of sequence-distant contacts per residue[49], the fraction of sequence-distant native contacts[50] and the total contact distance[51].

These new kinds of models have replaced the single folding pathway of Levinthal with the concept of a multitude of parallel paths, any one of which leads an individual molecule from the unfolded to the folded state. Folding is regarded as a stochastic search, albeit not a random one, but a search in an energy landscape, in which the gradient of the potential energy guides the proteins towards the energetic minimum, the folded state. Uphill steps that increase the potential energy are possible, but less likely than downhill steps. Thus, the solution to Levinthal's paradox is 'funnels, not tunnels', as Ken Dill puts it[52].

1.1.3 The Folding Funnel

Typical protein energy landscapes are often illustrated in the form of said funnels (fig. 1.1)[53,54]. Plotted is the 'internal free energy' E_{micro} of an individual chain and surrounding solvent and each point of the energy landscape represents a possible microstate of the protein. E_{micro} includes energy contributions due to bond angles, torsion, stretches, van der Waals interactions, hydrogen bonds and the hydrophobic and solvation free energies gained by burying hydrophobic residues. The free energy level, $G_{\mathrm{macro}}(\xi)$ of a macrostate, ξ, of a protein can be expressed with the partition function, $Z(\xi)$, of that state:

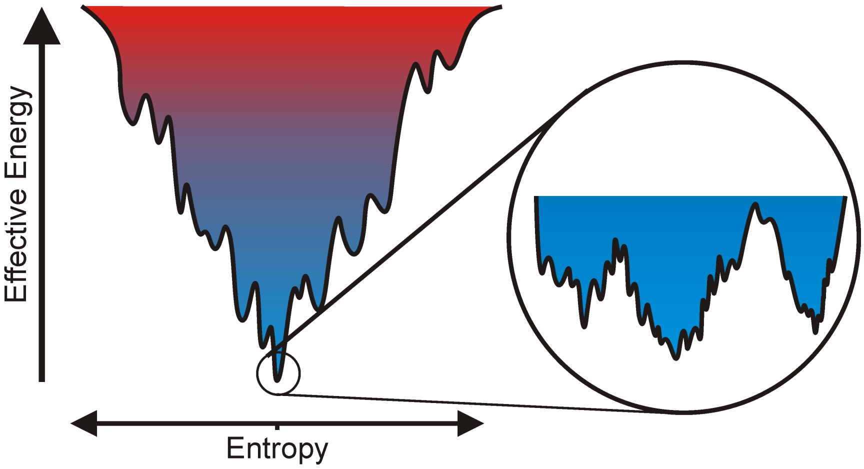


Figure 1.1: The funnel-shaped energy landscape of protein folding. The close-up illustrates the multitude of microstates that make up the native conformation macrostate.

$$G_{\mathrm{macro}}(\xi) = -kT\ \ln Z(\xi) \tag{1.1}$$

$$= -kT\ \ln \left(g(\xi) \exp\left[-\frac{E_{\mathrm{micro}}(\xi)}{kT} \right] \right), \tag{1.2}$$

where $g(\xi)$ is the number of microstates that contribute to the macrostate ξ and $E_{\mathrm{micro}}(\xi)$ the internal free energy of the microstates that make up ξ. k is the Boltzmann constant and T the absolute temperature of the system. Eqn. 1.2 simplifies to:

$$G_{\mathrm{macro}}(\xi) = E_{\mathrm{micro}}(\xi) - kT\ \ln\left(g(\xi)\right) \tag{1.3}$$

$$= E_{\mathrm{micro}}(\xi) - TS_{\mathrm{conform}}(\xi). \tag{1.4}$$

The conformational entropy of the protein chain, S_{conform}, is introduced in the last step using the statistical definition of entropy

$$S(\xi)_{\mathrm{conform}} = k \ln g(\xi), \tag{1.5}$$

which connects the entropy of macrostate ξ to the number of microstates, $g(\xi)$, that it represents.

The small ensemble of microstates at the bottom of the funnel constitutes the macrostate of the native, folded conformation, whereas the larger multitude of microstates at the edge of the funnel represent the unfolded macrostate. This means that folding comes at a huge loss of conformational entropy S_{conform}, which partially counteracts the microscopic free energy gain of the chain upon folding. Thus, the free energy gap that separates the folded from the unfolded macrostate is often remarkably small. For small globular proteins at room temperature the energy gain due to the stabilizing interactions, ΔE_{micro}, is only barely larger than the entropic cost due to the reduction of accessible microstates, $T\,\Delta S_{\text{conform}}$. Both these terms are on the order of MJ/mol, whereas their difference, the amount of free energy that stabilizes the folded state against the unfolded state, is only on the order of a few tens of kJ/mol, an amount comparable to the free energy content of a few hydrogen bonds.

In the funnel-shaped energy landscape, folding intermediates are represented by local minima in the energy landscape, 'gulleys' or 'ponds' that are common to many individual folding trajectories. Protein chains that find themselves in these intermediate states have to overcome an energy barrier, possibly by breaking already formed native contacts, before they can continue their folding progress [55–59]. Such local minima are also responsible for kinetic traps, such as in egg white lysozyme, in which a subpopulation undergoes fast overall folding, whereas a second subpopulation forms part of their secondary structure quickly followed by a much slower second folding step [21,24], a behavior called 'kinetic partitioning' [60]. However, not only local minima, but any feature of the energy landscape that slows folding can be the cause of an intermediate state. A flat plane in the internal free energy landscape may trap individual folding proteins as well, as they become lost in configurational space without a clear energy gradient. Only after some random diffusion do these molecules find the downhill slope again and can continue on their folding trajectory. This kind of diffusion-limited trapping can be visualized by an golf-course energy landscape, a flat plane with a hole in the middle. In this model, folding only progresses when the protein reaches a certain macrostate (the 'hole' of the golf-course), which, compared to the total number of microstates of the whole golf-course, consists of only a small number of microstates. Thus, reaching the hole correlates with a reduction of the conformational entropy, S_{conform}, of the protein chain, which, according to eqn. 1.4, creates a so-called entropic barrier in the macroscopic free energy G_{macro} [61,62].

Local minima of sufficient depth may trap subpopulations of a protein for significant amounts of time, leading to misfolded states. The folded state itself is composed of several substates, local minima in the rugged bottom of the folding funnel that are accessible at room temperature by thermal activation [63]. This flexibility of the native state is often of great importance for the function of a protein.The overall ruggedness of the energy landscape is known to influence the folding kinetics, in so far as fast folding is only possible if the overall slope towards the native state is large relative to the local ruggedness of the energy landscape[64,65].

So, while every molecule takes a distinct, individual path down the funnel, they all end up in the same global minimum, the folded conformation, thus satisfying Anfinsen's thermodynamic hypothesis. On the other hand, each molecule does so in a rapid and directed way, solving Levinthal's paradox.

1.1.4 The Unfolded State of Proteins

Characterizing the unfolded state of proteins proves challenging due to the structural heterogeneity of the unfolded ensemble. Traditional ensemble techniques such as small angle x-ray scattering (SAXS)[66] were used to measure properties of the unfolded conformation such as the radius of gyration, R_G, a measure widely used in polymer physics to describe the overall dimensions of polymer chains[67]. Indeed, at high denaturant concentrations, unfolded proteins were found to behave like random-coil polymers[68]. Comparing 28 crosslink-free chemically denatured polypeptides, it was revealed that the radius of gyration of proteins under strongly denaturing conditions only depends on the length of their polypeptide chain[69]. Another method to investigate the unfolded state is NMR spectroscopy, a technique that is able to characterize the unfolded ensemble in molecular detail[70–72]. Several groups have presented experimental evidence of residual structure even under strongly denaturing conditions[73–76]. This structure appears to take the form of hydrophobic clusters, forming predominantly around tryptophan (Trp) or histidine (His) residues[77–79]. Point mutations that replace Trp residues with glycin have been shown to disrupt the formation of such hydrophobic clusters in lysozyme[80]. This observation of residual structure might seem to be in conflict with the findings of random coil behavior. However, the presence of short, structured segments does not exclude random coil behavior, especially if the structures are only populated transiently[69,81–83]. Other ensemble techniques that have been employed to study the unfolded state of proteins include Raman optical activity[84], fluorescence correlation spectroscopy[85] and infrared absorption and vibrational circular dichroism spectroscopy[86].

To examine the unfolded state under weakly denaturing conditions, single molecule experiments are needed because in ensemble experiments, any signal under such conditions is dominated by the folded state of the protein[87,88]. As the environmental conditions become less denaturing, the unfolded protein coil collapses, as has been unequivocally proven by single molecule FRET studies[68,89,90]. Only the details of the collapse are open to discussion. While Schuler reports Gaussian chain behavior for the small cold shock protein Csp*TM*[91], measurements on chymotrypsin inhibitor 2 and acyl-CoA-binding protein show substantial deviations from Gaussian chain behavior for the collapsed denatured state[90].

1.1.5 Studies with Single-Molecule Resolution

Protein folding cannot be treated in the same way as simple chemical reactions, which are described as an ensemble of molecules that all perform the same sequence of steps to get from state A to state B. In contrast, the initial macrostate of protein folding, the unfolded conformation, comprises a vast collection of very distinct microstates. The individual molecules of the unfolded ensemble may be chemically identical, but they do posses a great variety of conformations [92]. Therefore, the folding trajectory of each protein of an ensemble is unique and the macroscopic folding process is a transition from disorder to order rather than a transition from one state to another [42].

The only way to gain meaningful results about such a heterogenous process as protein folding is to perform experiments on single molecules. Whereas ensemble measurements yield an averaged signal, weighted by population numbers and signals of the individual components, measuring every molecule on its own gives access to the complete distribution of observables [88,93,94]. It is the only way to distinguish between coexisting populations in equilibrium. In contrast to bulk experiments, studies on the single-molecule level allow us to directly resolve the heterogeneity of the folding process.

In the last fifteen years, methods to not only measure, but also to manipulate single molecules in aqueous environments have become a valuable part of the biophysical toolkit [93,95]. In single molecule force spectroscopy experiments, individual molecules are probed under the influence of external forces, which are applied, for example, with atomic force microscopes and optical or magnetic tweezers [95–101]. Fluorescence is a time honored method of observing biomolecules such as proteins, and has been employed in ensemble measurements for decades [102]. Especially with the advent of high-performance fluorescent dyes, it has turned into a suitable technique for single-molecule experiments [87,88,94,103–118]. Single dye molecules are attached to sites of interest at the biomolecule and then used in a variety of ways to report on the state and behavior of the biomolecule. Frequently one observes the quenching of the dye by nearby fluorescent quenchers, fluorescence anisotropy or the amount of resonant energy transfer to a second dye.

Single molecule detection requires tiny concentrations in the picomolar range. These concentrations result in a minuscule amount of sample consumption, which is an obvious advantage from a practical standpoint, considering how much work, time and money are necessary for the preparation of biological samples. The small concentrations also mean that one has to worry less about oligomerization or aggregation of the sample proteins, as typical experimental concentrations of about 10^{-9} M lie below the K_D values of all but the most tightly binding proteins [119,120].

1.1.6 Observing Protein Folding of Single Molecules

The characteristic time scales of protein folding are determined by two factors: The life times of the individual folded, unfolded or intermediate protein states and the transition times between these states, that is, the time it takes for a single protein to diffuse in the funnel-shaped energy landscape from one state to the other. The life times of the folded or the unfolded state can be directly observed in single molecule experiments with immobilized proteins[112]. These life times were found to be in the milliseconds to seconds range, depending on the environmental conditions. The transition times of small globular proteins, on the other hand, have only recently been experimentally determined to be smaller than $10\,\mu s$[121,122]. This time span is barely enough to detect a few photons from a single fluorescent molecule. However, to gain meaningful insights into the state of a protein, tens of photons have to be collected, requiring an observation time of hundreds of microseconds. Thus, single molecule fluorescence methods do not offer a sufficiently high time-resolution to observe single molecules on their way down the folding funnel. When examining protein folding, it is not the process of folding itself that is observed, but the populations of the multiple long-lived conformations situated in minima of the free energy landscape, such as folding intermediates. By addition of chemical denaturants, the energy landscape can be changed and the energy levels and populations of the different macrostates shift with regard to one another. The heterogeneity of states can then be studied in equilibrium at each denaturant concentration. From the examination and characterization of folding intermediates, one hopes to learn more about the folding process.

1.2 RNase H

Escherichia coli RNase H (protein data base accession code 2RN2) is a 155 amino acid-residue protein, with a secondary structure comprised of 5 α-helices and 5 β-sheets. It has a molecular weight of about 17.6 kDa[123]. Its crystal structure has been solved[124] and its tertiary structure is well characterized and of globular nature, as depicted in fig. 1.2a. It is a monomeric enzyme which catalyzes the cleavage of RNA in DNA-RNA hybrids[125]. After its initial isolation from calf thymus[126,127], it was subsequently found in a wide variety of organisms, from viruses and bacteria to eukaryotic multi-cell organisms such as humans[128]. Chemical and thermal denaturation of the protein was shown to be fully reversible *in vitro*[129].

RNase H has been the subject of intense investigation, especially by the group of Susan Marqusee. In the course of research on the folding of RNase H with equilibrium bulk experiments, an intermediate (I) conformation was detected using Circular Dichroism (CD) and Trp fluorescence spectroscopy[130]. Although already present in equilibrium with the folded (F) state under native conditions, the intermediate population can be enhanced by addition of chemical denaturants, such as guanidinium chloride (GdmCl) or urea. Because of its increased stability under acidic conditions ($\mathrm{pH} < 3$), this intermediate has also been termed the *acid state* of RNase H. It was further characterized by amide hydrogen exchange methods[131], yielding much of the structural knowledge we presently have about the intermediate state. The amide hydrogen exchange experiments revealed those secondary structure elements that most strongly resisted denaturation under environmental conditions which increasingly favor the unfolded state. This set of the most stable α-helices and β-sheets is situated in the center of the folded RNase H and is called the protein core. Specifically, the protein core is formed by the most stable α-helices A and D, and the slightly less stable α-helix B as well as β-strand 4, as depicted in fig. 1.2b. These regions were assumed to form the stable protein core, which is most resistant against unfolding.

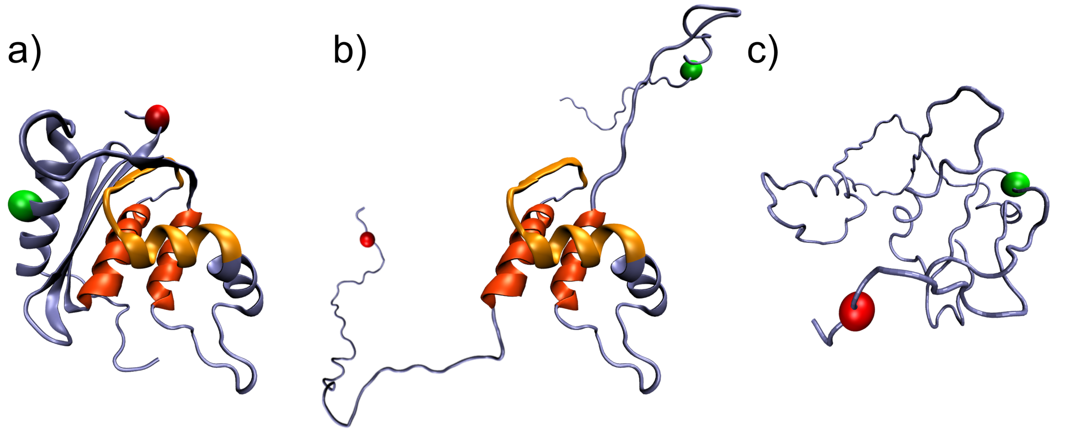


Figure 1.2: Schematic depiction of three conformations of RNase H. **a)** Folded structure (F) as obtained from the crystal structure. The protein core consists of the most stable α-helices A and D (dark orange) and the slightly less stable α-helix B as well as β-strand 4 (light orange). **b)** Folding intermediate (I), in which only the protein core retains its structure. **c)** Entirely unfolded state (U) at high denaturant concentration. The spheres at positions 3 (red) and 135 (green) mark the attachment sites of donor and acceptor dyes.

1.3 Outline

Detection and characterization of folding intermediates are currently the only windows into the folding process and, as such, key steps for the understanding of the folding pathway. While ensemble measurements can give strong support to the hypothesis of a folding intermediate of RNase H, due to the heterogenous nature of the folding process, only single molecule experiments can truly test and verify the hypothesis.

The goal of this thesis was to directly observe and characterize the folding intermediate of RNase H in freely diffusing single molecules via Förster resonance energy transfer and to unequivocally demonstrate the existence of a folding intermediate in RNase H. The performed experiments further allowed the characterization of RNase H in solution, as they provided the free energy relationships between the folded, the intermediate and the unfolded state under a wide range of denaturing conditions. A technique to interpret single molecule data with the help of additional information from Trp fluorescence measurements was developed in this thesis and may find application in future protein folding studies.

Chapter 2 gives a brief introduction into fluorescence and how it can be used to investigate internal protein dynamics. It is explained how the state of single molecules can be measured with the ”spectroscopic ruler” of Förster resonance energy transfer (FRET). Confocal microscopy is introduced, as well as the necessary algorithms to detect single diffusing molecules and calculate their individual FRET efficiencies. Trp fluorescence is introduced as another method to investigate protein conformations, exploiting the intrinsic protein fluorescence from Trp residues.

Chapter 3 discusses sample preparation. The fluorescence labeling of RNase H is detailed, as well as the employed buffers and the confocal microscope setups. The chapter also gives an overview of the fundamentals of the fitting routine developed for data evaluation.

Chapter 4 starts out with the calibration measurements necessary for data analysis. Single molecule FRET histograms of RNase H are presented and the procedures of data fitting are described. Certain constraints that had to be placed onto chosen fit parameters are introduced, justified and discussed. The measured data were modeled by both a two-state model and a three-state model, which are compared to prove the necessity of an intermediate state. The chapter further describes how the single molecule FRET measurements were extended to acidic conditions, which are known to stabilize the intermediate state. To circumvent disambiguities in the FRET data, Trp fluorescence was introduced as a supporting technique and smFRET and Trp fluorescence data were simultaneously fitted with one combined model. Finally, the free energy parameters that were gained by both measurements are discussed and compared to literature data.

Chapter 5 summarizes the experimental studies performed in this thesis and its results.

Chapter 2

Theory

2.1 The Phenomenon of Fluorescence

2.1.1 The Jablonski Diagram

The molecular processes that occur when molecules are electronically excited are often visualized with the aid of Jablonski diagrams (see fig. 2.1)[132]. The states of a molecule are depicted as horizontal lines. The diagram shows the electronic singlet ground state S_0, the first excited electronic singlet states S_1 and S_2 and the first excited electronic triplet state T_1. Each of these electronic energy levels is divided into several vibronic sub-states ($V_1 \ldots V_n$), describing the vibrational and rotational states of the molecule. For clarity, the whole vibrational spectrum of states is reduced to only five distinct states $V_1 \ldots V_5$.

Absorption

Electrons can be excited from low to higher energy levels by absorbing a photon with an energy that matches the energy difference between the two states, as indicated by the blue, upward arrows in the Jablonski diagram.

The available thermal energy, kT, at room temperature ($T \approx 300$ K) is about $4.1 \cdot 10^{-21}$ J which corresponds to a wavenumber of 200 cm^{-1}. While the vibrational modes of small molecules have wavenumbers of several thousand cm^{-1} (e.g. 4342 cm^{-1} for the H_2 stretching mode), larger molecules may contain vibrational modes with wavenumbers as low as only a few cm^{-1}. Thus, absorption occurs mostly from the vibrational ground state, V_0, and the low energy part of the vibronic spectrum of S_0. The transition to higher energy levels occurs in about 10^{-15} s, much faster than the time scale of nuclear motion. According to the *Franck-Condon principle*[133,134], the most probable transition is to a vibrationally excited level of the electronic excited state. From there, the excited electron decays back to the electronic ground state via several different pathways[135].

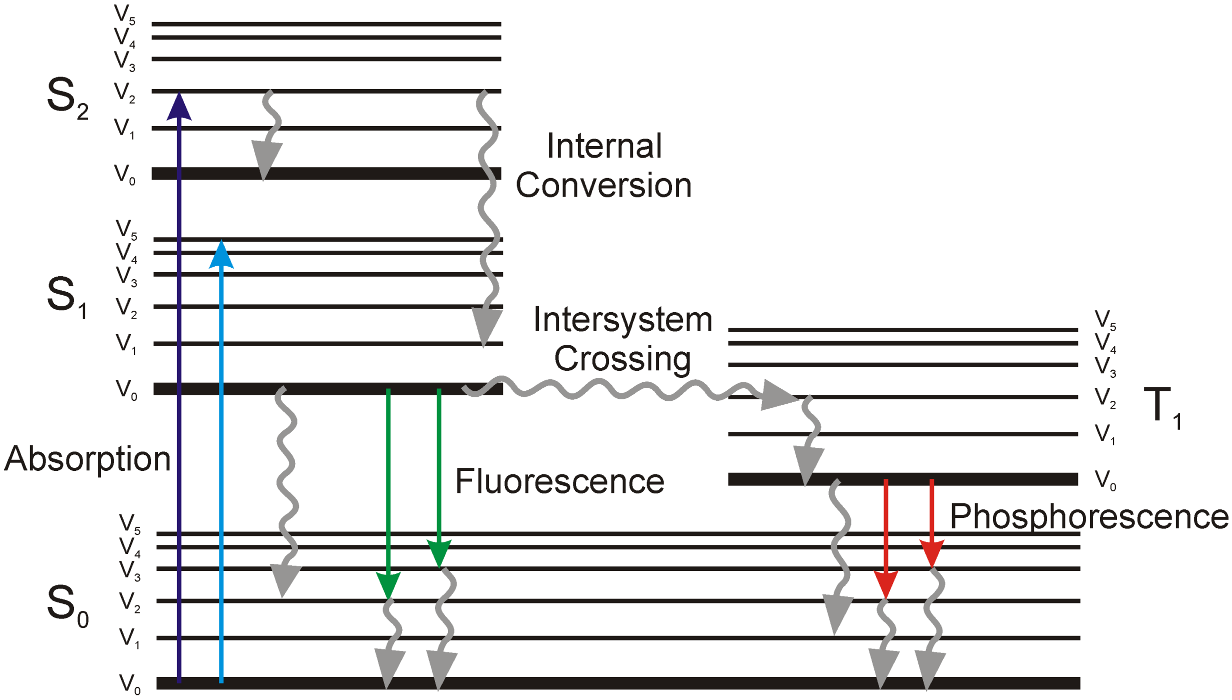

S$_2$
S$_1$
S$_0$
T$_1$
V$_5$
V$_4$
V$_3$
V$_2$
V$_1$
V$_0$
Internal Conversion
Intersystem Crossing
Absorption
Fluorescence
Phosphorescence

Figure 2.1: The first three singlet, and the first excited triplet state, each divided into vibrational energy sub-states. Radiative transitions such as absorption (violet and blue), fluorescence (green) and phosphorescence (red) are depicted as straight arrows, non-radiative processes such as vibrational relaxation, internal conversion and intersystem crossing as wiggly grey arrows.

The Franck-Condon principle

The Hamilton operator of a molecule consists of the kinetic energies of electrons and nuclei, $\boldsymbol{T}_{\mathrm{el}}$ and $\boldsymbol{T}_{\mathrm{nuc}}$, the electrostatic potentials describing the Coulombic interactions of the system including the repulsive interactions between nuclei, $\boldsymbol{V}_{\mathrm{nuc-nuc}}$, and electrons, $\boldsymbol{V}_{\mathrm{el-el}}$, as well as the attractive nucleus-electron interactions $\boldsymbol{V}_{\mathrm{nuc-el}}$:

$$\boldsymbol{H} = \boldsymbol{T}_{\mathrm{el}} + \boldsymbol{T}_{\mathrm{nuc}} + \boldsymbol{V}_{\mathrm{nuc-nuc}} + \boldsymbol{V}_{\mathrm{el-el}} + \boldsymbol{V}_{\mathrm{nuc-el}}. \tag{2.1}$$

The potentials for spin–spin coupling and spin–orbit coupling are omitted here due to their negligible contributions in comparison to the other potentials. Making an assumption called the *Born-Oppenheimer approximation*, the total wave function, Ψ_{total} of a molecule can be broken up into its electronic and its nuclear components, ψ_{el} and ψ_{nuc}, due to the great disparity in mass between the electrons and the nuclei of the molecule:

$$\Psi_{\mathrm{total}}(\boldsymbol{r}, \boldsymbol{R}) = \psi_{\mathrm{el}}(\boldsymbol{r}, \boldsymbol{R}) \cdot \psi_{\mathrm{nuc}}(\boldsymbol{R}). \tag{2.2}$$

While the nuclear part of the wave function depends only on the nuclear coordinates, $\boldsymbol{R}$, the electronic part depends on both the electron coordinates $\boldsymbol{r}$ and the nuclear coordinates, $\boldsymbol{R}$.

For the calculation of the electronic part of the wave function, ψ_{el}, the nuclei are considered stationary. They only provide a Coulombic potential $\boldsymbol{V}_{\mathrm{nuc-el}}$ for the Hamilton operator of the electrons. The electronic Hamilton operator further depends on the kinetic energies of the electrons $\boldsymbol{T}_{\mathrm{el}}$ and the Coulombic electron-electron interactions described by a potential $\boldsymbol{V}_{\mathrm{el-el}}$

$$\boldsymbol{H}_{\mathrm{el}} = \boldsymbol{T}_{\mathrm{el}} + \boldsymbol{V}_{\mathrm{el-el}} + \boldsymbol{V}_{\mathrm{nuc-el}}. \tag{2.3}$$

Solving the stationary Schrödinger equation,

$$\boldsymbol{H}_{\mathrm{el}}\, \psi_{\mathrm{el}}(\boldsymbol{r}, \boldsymbol{R}) = E_{\mathrm{el}}(\boldsymbol{R})\, \psi_{\mathrm{el}}(\boldsymbol{r}, \boldsymbol{R}), \tag{2.4}$$

then yields the electronic wave functions, ψ_{el}, and the associated electronic properties such as the probability densities of the electron positions as well as the energy of the electronic system, $E_{\mathrm{el}}(\boldsymbol{R})$, for a specific configuration of nuclear positions, $\boldsymbol{R}$.

The Hamilton operator for the nuclei consists of

$$\begin{aligned} \boldsymbol{H}_{\mathrm{nuc}} &= \boldsymbol{T}_{\mathrm{nuc}} + \underbrace{\boldsymbol{V}_{\mathrm{nuc-nuc}} + E_{\mathrm{el}}(\boldsymbol{R})}_{\boldsymbol{V}_{\mathrm{eff}}(\boldsymbol{R})} && (2.5) \\ &= \boldsymbol{T}_{\mathrm{nuc}} + \boldsymbol{V}_{\mathrm{eff}}(\boldsymbol{R}), && (2.6) \end{aligned}$$

with an effective potential $\boldsymbol{V}_{\mathrm{eff}}(\boldsymbol{R})$ that includes the nucleus – nucleus interaction as well as the energy of the electronic system obtained by solving the Schrödinger equation for the nuclear coordinates $\boldsymbol{R}$. Possible effective potentials for the electronic ground state S_0 and the first excited electronic state S_1 are sketched in fig. 2.2 for the simplest case of a diatomic molecule, in which case the effective potential $\boldsymbol{V}_{\text{eff}}$ only depends on the internuclear separation of the two nuclei. It determines the vibrational energy levels $\mathrm{V}_1 \dots \mathrm{V}_n$ of the molecule and the corresponding nuclear wave functions. For molecules with more than two atoms, $\boldsymbol{V}_{\text{eff}}$ can be described as a high-dimensional hypersurface. Any change in the electronic system of the molecule, such as excitation of the molecule to a higher electronic state, occurs on a time scale much shorter than the characteristic times of nuclear motion. These changes in the electronic system lead to a different effective potential $\boldsymbol{V}_{\text{eff}}(\boldsymbol{R})$, resulting in new energy eigenstates for the nuclear positions. The nuclear system has to switch from a nuclear state under the initial electronic conditions to a new state under the changed electronic conditions. The relative probability of such a transition compared to other possible transitions is proportional to the square of the overlap integral of initial and final nuclear wave functions. These relative transition probabilities are called *Franck-Condon factors.*

As depicted in fig. 2.2, the equilibrium position of the effective potential of the excited state is often shifted to a longer nuclear distance. Thus, the vibrationally excited states of S_1 generally exhibit a stronger overlap with the V_0 state of S_0 than the vibrational ground state of S_1. This means that, when the electronic system of a molecule is excited, the transition to an excited vibrational state of S_1 is more likely than a transition to the vibrational ground state of S_1. When the electronic system relaxes from S_1 to S_0, the nuclear system of the molecule ends up in a vibrationally excited state with a higher probability than in the vibrational ground state, for the same reasons.

Vibrational Relaxation

Usually, the molecule relaxes quickly to V_0 of the excited electronic state by *vibrational relaxation.* In $\approx 10^{-14}$ – 10^{-11} s, the energy of the electronic system is deposited into vibrational modes of the molecule. An excited electron can also transit from one electronic level to a lower electronic level of the same spin, a process called *internal conversion.* Both processes are radiationless in the sense that they involve neither absorption nor emission of a photon and are depicted as wiggly arrows in the Jablonski diagram (fig. 2.1). The probability of internal conversion grows with the extent of overlap between the vibrational spectra of different electronic states. For most molecules in the condensed phase, this overlap is sufficiently pronounced at electronic levels above S_1. In consequence, these molecules relax to the S_1 level within less than 10^{-12} s. However, the large energy gap between S_0 and S_1 results in a lack of overlap between the vibrational manifolds of both states, resulting in a markedly slower vibrational relaxation from S_1 to S_0. In some molecules, the internal

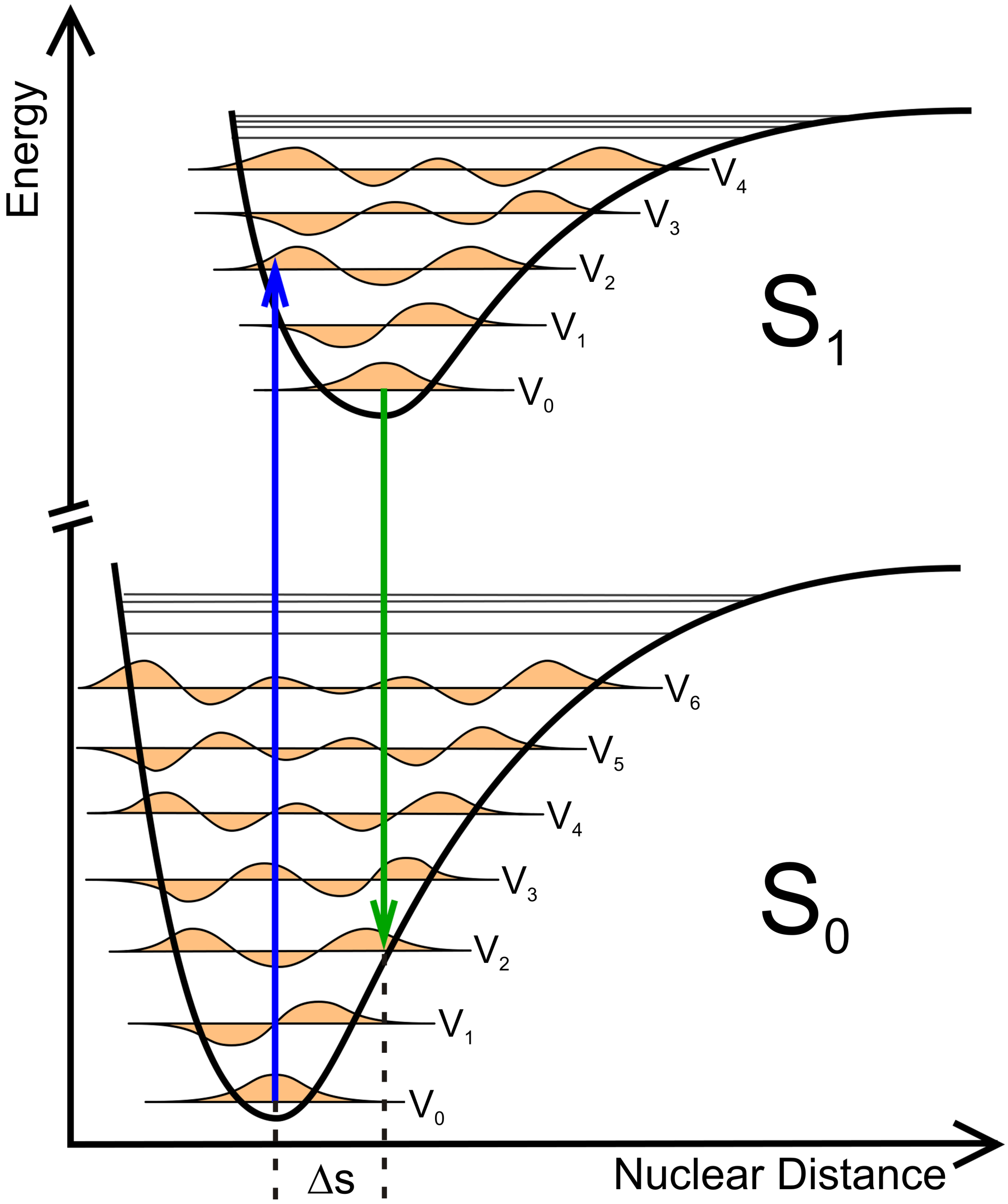


Figure 2.2: Illustration of the Franck-Condon principle for a molecule consisting of two atoms, based on a figure by Onno Gabriel[136]. The effective potentials $\boldsymbol{V}_{\mathrm{eff}}(\boldsymbol{R})$ of the electronic states S_0 and S_1 are plotted against the relative distance of the nuclei $|\boldsymbol{R}|$. The equilibrium position of the S_1 state is shifted by Δs with respect to the electronic ground state S_0. A possible absorption process from the vibrational ground state of S_0 to the second vibrational level of S_1 is depicted as the blue arrow, a relaxation process as the green arrow.

conversion to the electronic ground state is so slow that it competes with relaxation processes like fluorescence, which occur on time scales from 10^{-9} to 10^{-8} s. Such molecules are called fluorescent, whereas molecules that mainly relax via non-radiative pathways are called non-fluorescent.

Fluorescence and Phosphorescence

If relaxation of an excited electronic state to a lower electronic energy level of the same spin results in the emission of a photon, the decay is called *fluorescence*. It is depicted as a straight, downward arrow in the Jablonski diagram. As fluorescence is a slow process (10^{-9} – 10^{-8} s) compared to both vibrational relaxation and internal conversion to the lowest S_1 vibronic state, fluorescence emission typically occurs from the vibrational ground state of S_1, a fact referred to as *Kasha's rule*. As the Frank-Condon principle also applies to emission, fluorescence decay generally leaves the molecule in an excited vibrational state of S_0, followed by rapid vibrational relaxation to thermal equilibrium, which in most cases means relaxation to V_0.

In the singlet state S_0, all electron spins are paired and the total spin of the electronic system is zero. If the electronic system is excited to either S_1 or S_2, its total spin and, in consequence, the spin pairing are both conserved. If, however, a spin flip occurs upon excitation, the system reaches an excited triplet state T_1 in a process called *intersystem crossing* (ISC). This transition is "forbidden", according to strictly electronic selection rules, resulting in a relatively slow process on time scales that allow it to compete with fluorescence. Once in T_1, the same radiationless relaxation pathways as in the singlet states are available to the electron such as vibrational relaxation and internal conversion. Radiative relaxation from T_1 to the electronic singlet ground state S_0 is called *phosphorescence*. As it involves a change of spin, the time scale of phosphorescence is several orders of magnitude larger than for fluorescence. Due to this long time scale, phosphorescent decay has to compete with ISC back to S_0, which often strongly reduces the number of radiative phosphorescence events.

Stokes Shift

When a molecule is excited to higher electronic states, both its dipole moment and polarizability can change significantly from their values in the electronic ground state[137,138]. After excitation, this change in dipole moment prompts a reorientation and relaxation of surrounding, polar solvent molecules, lowering the energy of the excited state (depicted in fig. 2.3 as the transition from S_1 to S_1'). At the same time, this new solvent orientation is energetically less favorable for the ground state than the original orientation, destabilizing S_0 into a state with higher energy, S_0'. The reorientation of the solvent molecules occurs on a time scale of $\approx 10^{-10}$ s and, thus, generally precedes the emission of fluorescence. Both these changes in energy level lead to an energy difference between S_0' and S_1', that is smaller than the original energy difference between S_0 and S_1. Consequently, the emitted fluorescence light is red-shifted with respect to the light needed to excite the molecule.

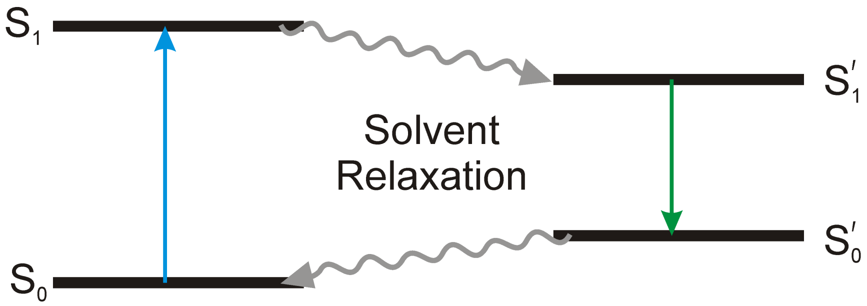


Figure 2.3: Jablonski diagram, illustrating the influence of solvent relaxation. Vibrational substates are omitted for clarity. After excitation, the electronic states S_0 and S_1 experience a shift in their respective energy levels to S'_0 and S'_1 due to solvent relaxation.

Nonradiative Decay

Several other nonradiative processes apart from intersystem crossing compete with fluorescence for the transition from exited states to the ground state. As already mentioned, internal conversion has a low but non-zero transition rate between S_1 and S_0. Another path is *collisional quenching*, the transfer of energy to other molecules through collisions. One of the most efficient collisional quenchers is molecular oxygen, which quenches nearly all known fluorophores[139,140]. Another form of quenching is *resonance energy transfer* (RET), in which the energy of the excited state is transferred to an acceptor molecule via nonradiative Coulombic interaction, lifting the acceptor molecule to an excited electronic state. If the acceptor is fluorescent itself, it may then emit a fluorescence photon. In contrast to collisional quenching, direct molecular contact is not required for RET, but can rather occur over distances of several tens of Ångströms.

2.1.2 Parameters of Fluorescence

The wavelength-dependent molar extinction coefficient, $\varepsilon(\lambda_{\text{ex}})$, describes how strongly a molecule absorbs light of a given excitation wavelength, λ_{ex}. By observing the frequency dependence of absorption, information about the structure of the energy levels of the sample can be gained.

Another important characteristic of a fluorophore is its quantum yield, ϕ, which is the probability of a fluorescence photon being emitted upon excitation of the fluorophore. It is defined by the ratio between the rate coefficient of radiative fluorescence decay, k_{fl}, and the total decay rate coefficient of a fluorophore

$$\phi = \frac{k_{\mathrm{fl}}}{k_{\mathrm{fl}} + \sum_i k_{i,\,\mathrm{nr}}}, \tag{2.7}$$

where $\sum_i k_{i,\,\mathrm{nr}}$ is the sum of the decay rate coefficients of all nonradiative processes. The quantum yield of a dye is independent of the excitation wavelength.

The brightness of a fluorophore excited with light of a particular wavelength, λ_{ex}, is the product of its extinction coefficient, $\varepsilon(\lambda_{\mathrm{ex}})$, and its quantum yield, ϕ:

$$\mathrm{BR}(\lambda_{\mathrm{ex}}) = \varepsilon(\lambda_{\mathrm{ex}})\,\phi. \tag{2.8}$$

2.1.3 Photobleaching

There are several pathways how a fluorophore can lose its ability to fluoresce, a process called *photobleaching*[141]. Due to its long lifetime, the excited triplet state is often considered to be the starting point of photochemical reactions that alter the fluorophore and possibly lead to photobleaching[142]. In aqueous, air-saturated solutions, molecular oxygen has been identified as an important source of photobleaching[143–145] and, as such, is a deciding factor for the stability of many organic dyes. Oxygen has been known as a potent quencher of excited states of organic dye molecules, singlet and triplet alike. The oxygen ground state is a triplet state. Upon quenching, it is converted into the higher energy oxygen singlet species, which is highly reactive[146,147]. Thus, quenching by oxygen increases the risk of oxidation of the fluorophore due to a reaction with the produced singlet oxygen[148]. For demanding applications, the removal of oxygen has been proposed as a measure to increase photostability[149]. Unfortunately the oxygen removal also removes one of the most important pathways for triplet state recovery. This leaves the dye trapped in the T_1 state for longer periods of time, decreasing the overall brightness and increasing blinking[147].

To retain the brightness of a dye in the absence of oxygen, the Tinnefeld group proposed the use of a *reducing and oxidizing system* (ROXS)[142,150] to quench the T_1 state and return the fluorescent system to the S_0 state as fast as possible via an electron-transfer reaction. Performing only either an oxidizing or reducing reaction would leave the system in a radical anionic or cationic state ground state (assuming a ground state that is normally neutral). Thus, the ROXS system involves an oxidizing as well as a reducing step, as depicted in fig. 2.4. The triplet state is rapidly depleted by electron-transfer reactions through either oxidation (forming a radical cation $\mathrm{F}^{\bullet+}$) or reduction (forming a radical anion $\mathrm{F}^{\bullet-}$). These reactions are carried out by the reducing and oxidizing agents that make up the ROXS. Both ionic states, $\mathrm{F}^{\bullet+}$ and $\mathrm{F}^{\bullet-}$, are then rapidly returned to the neutral singlet ground state S_0 by reduction or oxidation, respectively. The second step competes with side reactions that would take the reactive ionic species to photobleached products, P. The redox potentials of both the reducing and the oxidizing agent are chosen so that they do not react directly with each other. The energy of the excited triplet state is necessary to drive the electron-transfer processes.

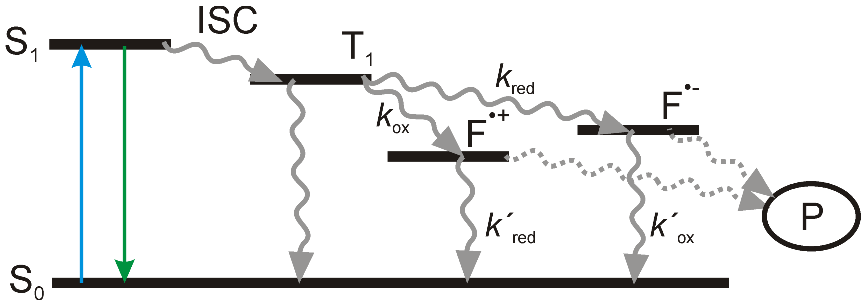


Figure 2.4: Jablonski diagram, depicting the working principle of a reducing and oxidizing system (ROXS). Vibrational states have been omitted for clarity. Triplet population is returned to the singlet ground state by oxidizing and reducing agents through the ionic states $F^{\bullet+}$ and $F^{\bullet-}$. The second oxidation or reduction step competes with pathways that would take the reactive ionic species to a photobleached state, P (light gray arrows).

2.1.4 Dependence of the Fluorescence on the Environment

The emission properties of a fluorescent molecule, such as the quantum yield and the peak position of the emission spectrum, are heavily dependent on its local environment. The most significant factor is the polarity of the solvent, but other solvent properties, such as the viscosity, the rate of solvent relaxation and the rigidity of the local environment may affect the fluorescence as well[102]. It can be difficult to determine to which degree each of those influences affects the fluorescence in a given experimental instance, and the multitude of effects makes it difficult to treat solvent-fluorophore interactions with a single, unified theory. However, even without detailed knowledge of the mechanisms, the strong dependence on the local environment makes fluorescence a potent and popular tool to assess the conformational state of proteins. One way to perform fluorescence measurements on proteins is to exploit the intrinsic protein fluorescence, which is due to the three aromatic amino acids phenylalanine (Phe), tyrosine (Tyr) and Trp. Another approach involves attaching fluorescent labels to points of interest in the protein.

2.1.5 Trp as an Internal Fluorescent Probe

As Gregorio Weber pointed out in an early review of fluorescence spectroscopy[151], the intrinsic fluorescence of proteins is mainly due to aromatic amino acids, which emit in the ultraviolet region of the spectrum. This fact was confirmed soon afterwards by the experimental observation of protein fluorescence and the fluorescence of aromatic amino acids[152–155]. Trp is a fluorescent, aromatic amino acid with a maximum absorption at $\approx 280\,\mathrm{nm}$. The emission spectrum of Trp is sensitive to the immediate surroundings of the indole ring of the Trp residue. The emission maximum λ_{max} of Trp dissolved in aqueous solution is at $\approx 355\,\mathrm{nm}$[156]. In proteins, the emission maximum varies from $\approx 308\,\mathrm{nm}$ (in azurin) to $\approx 355\,\mathrm{nm}$ (e.g. in glucagon), depending on the local environment and roughly correlates with the degree of solvent exposure of the chromophore[157]. Other strongly environment-dependent properties of Trp fluorescence are its quantum yield[158–160], anisotropy[161–163] and fluorescence lifetimes[164,165]. As such, Trp fluorescence is a widely used probe to assess the conformational state of proteins, substrate binding or denaturation[102,166–168]. About 300 papers per year are listed in "Biological Abstracts" alone in which Trp fluorescence in proteins is either exploited or studied[157].

2.2 FRET – Förster Resonance Energy Transfer

2.2.1 Introduction and Principle

If two fluorophores are brought into close proximity, they can exchange excitation energy via a nonradiative process called Förster resonance energy transfer (FRET), a process that competes with the fluorescence and other relaxation pathways of the donor. Upon excitation of the donor dye, the energy is transferred to the acceptor with probability E, the FRET efficiency. This efficiency depends on the inverse sixth power of the inter-dye distance R ($E \propto R^{-6}$). The range in which FRET is the dominant form of energy transfer lies between 10 and 100 Å, the size of many biological macromolecules. Below that range, it competes with direct energy transfer due to electron exchange (Dexter transfer) because of wavefunction overlap, which loses efficiency with an exponential dependence on the dye separation R[169,170]. At long ranges, the nature of energy transfer shifts to the classical radiative energy transfer between dipoles with a R^{-2} distance dependence. Combined with the fact that FRET can be measured in aqueous environments on commonly available optical microscopes, the strong distance dependence makes FRET a great tool to study intramolecular distances in biomolecules such as proteins. A typical experiment involves tagging the protein of interest with donor and acceptor dyes at well defined positions along the amino-acid chain, as pictured in fig. 2.5. Any conformational change that alters the distance between the dyes leads to a change in the FRET efficiency and can thus be detected and characterized by FRET spectroscopy.

To measure the FRET efficiency E, the donor dye is excited. The resulting fluorescence intensity is recorded and separated into light emitted by the donor, I_D, and acceptor, I_A, depending on the wavelength. E can be calculated from these intensities by determining which fraction of the totally emitted light, $I_\mathrm{D} + I_\mathrm{A}$, is emitted by the acceptor fluorophore:

$$E = \frac{I_\mathrm{A}}{\gamma I_\mathrm{D} + I_\mathrm{A}}, \tag{2.9}$$

where γ is a correction factor to account for different quantum yields and detection efficiencies of donor and acceptor dye. The exact calculation of E, necessary corrections and the determination of γ are described in detail in section 2.2.3.

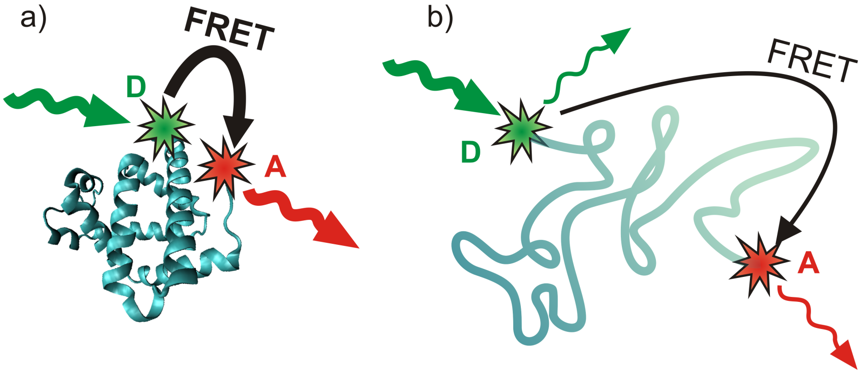


Figure 2.5: **a)** Sketch of a protein in its native conformation. Attachment sites for donor (D) and acceptor dye (A) are chosen to be in close proximity in the folded state, resulting in high FRET efficiency E. Upon excitation of the donor dye, most of the energy is transferred to the acceptor and emitted from there. **b)** In the unfolded conformation, the mean distance between D and A is increased, causing a decline in E. Compared to the folded conformation, fluorescence emission from the donor becomes more likely while, at the same time, the fluorescence intensity of the acceptor dye upon excitation of the donor dye decreases.

The mechanism of FRET, described by Förster in 1948[171], is based on the non-radiative dipole-dipole interaction between two fluorophores. A proper quantum electrodynamic treatment reveals this radiationless energy transfer to be the short-range limit of a more general theory of coulombic intermolecular interaction[170,172]. This unified theory also confirms the R^{-6} dependence of the transfer efficiency on distance, which was already predicted by Förster[173] and later confirmed experimentally[174]. At distances above ≈ 100 nm, retardation effects modify the R^{-6} distance dependence to R^{-2}, the classical result for radiative transfer,[170] which represents the long-range limit of the theory. As the interaction is electromagnetic in nature, the carrier particle of the interaction is a photon. FRET is called non-radiative in the sense that the energy is transferred by a *virtual* photon, a messenger particle which cannot be observed[172]. Due to its strong distance dependence, FRET is called a *spectroscopic ruler*[174].

2.2.2 Theoretical Calculation of the FRET efficiency

The efficiency E of the resonance energy transfer is a dimensionless number $0 \leq E \leq 1$. It is described as a function of the inter-dye distance R by the famous Förster formula[171],

$$E = \frac{1}{1 + (R/R_0)^6}. \tag{2.10}$$

Here, R_0 is the so-called Förster distance, which contains all parameters that factor into the efficiency of energy transfer except for the distance R. It contains the refractive index, n, of the surrounding medium, the quantum yield of the donor dye, ϕ_D, and the relative orientation of the transition dipoles of both dyes, $\boldsymbol{p}_\mathrm{D}$ and $\boldsymbol{p}_\mathrm{A}$, described by the orientation factor, κ^2. At last, the Förster distance also depends on the overlap integral J between the fluorescence spectrum $F_\mathrm{D}(\lambda)$ of the donor and the absorption spectrum $\varepsilon_\mathrm{A}(\lambda)$ of the acceptor dye. It is calculated as

$$R_0 = (8.79 \cdot 10^{23} n^{-4} \phi_\mathrm{D} J \kappa^2)^{1/6}, \tag{2.11}$$

with J and κ^2, in turn, being determined by

$$\begin{aligned} J &= 10^{-28} \frac{\int F_\mathrm{D}(\lambda) \varepsilon_\mathrm{A}(\lambda) \lambda^4 \, \mathrm{d}\lambda}{\int F_\mathrm{D}(\lambda) \, \mathrm{d}\lambda}, && (2.12) \\ \kappa^2 &= (\cos\theta_\mathrm{T} - 3\cos\theta_\mathrm{D} \cdot \cos\theta_\mathrm{A})^2. && (2.13) \end{aligned}$$

Here, θ_T represents the angle between the emission dipole of the donor molecule and the absorption dipole of the acceptor. θ_D and θ_A are the angles between the aforementioned dipoles and the vector $\boldsymbol{R}$ that connects donor and acceptor (see fig. 2.6)[102]. If the orientation of both dyes with regard to each other is static, κ^2 can assume any value between 0 and 4 with the extremes of parallel or antiparallel orientation ($\kappa^2 = 4$) on the one, and perpendicular orientation ($\kappa^2 = 0$) on the other hand. If the dyes are free to rotate independently of each other on a time scale much shorter than the time scale of the energy transfer process, an average $\langle\kappa^2\rangle$ must be used. In that case, integrating over all possible orientations of $\boldsymbol{p}_\mathrm{D}$ and $\boldsymbol{p}_\mathrm{A}$ yields $\langle\kappa^2\rangle = {}^2\!/_3$.

2.2.3 Experimental Determination of the FRET efficiency

The FRET efficiency E can be thought of as the quantum yield of the resonant transfer process, that is, the percentage of all donor excitations, N_total, that lead to energy transfer to the acceptor, N_A.

$$E = \frac{N_\mathrm{A}}{N_\mathrm{total}} = \frac{N_\mathrm{A}}{N_\mathrm{D} + N_\mathrm{A}}, \tag{2.14}$$

with N_D denoting the number of events in which the excitation is not transferred. Thus, E denotes the probability that the energy will migrate to the acceptor dye upon excitation of

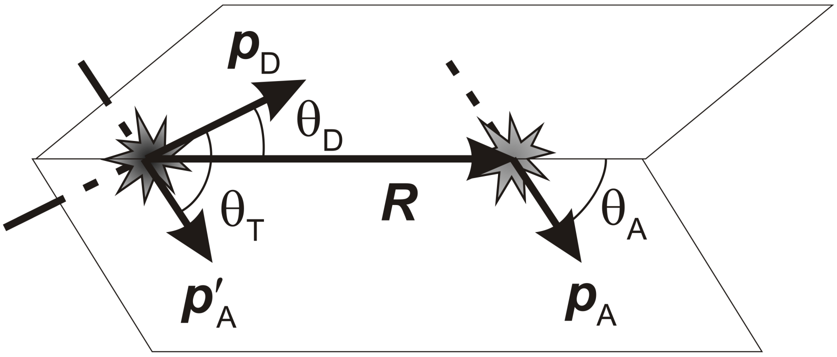

$\boldsymbol{p}_\mathrm{D}$
θ_D
θ_T
$\boldsymbol{R}$
θ_A
$\boldsymbol{p}'_\mathrm{A}$
$\boldsymbol{p}_\mathrm{A}$

Figure 2.6: Relative orientation of donor and acceptor transition dipoles $\boldsymbol{p}_\mathrm{D}$ and $\boldsymbol{p}_\mathrm{A}$. Vector $\boldsymbol{R}$ connects both dyes, vector $\boldsymbol{p}'_\mathrm{A}$ is parallel to $\boldsymbol{p}_\mathrm{A}$, shifted to intersect $\boldsymbol{p}_\mathrm{D}$. θ_D and θ_A describe the angles between $\boldsymbol{R}$ and the respective transition dipoles, whereas the angle between the dipoles themselves is called θ_T.

the donor. Unfortunately, with an optical microscope, one cannot directly observe which dye is excited. Instead, to detect excited states, we rely on the detection of fluorescence light that is only emitted with a certain probability given by the quantum yield ϕ of the dye. Furthermore, not every emitted photon is detected and the detection efficiency of the instrumental setup, η, is generally different for light of different wavelengths due to a variety of factors, such as the employed filtersets, dichroic mirrors and the wavelength dependent detector sensitivity. Therefore, one cannot assume that the existence of an exited state of either donor or acceptor dye leads to detection of a photon in the corresponding channel with equal probability. In consequence, E cannot be obtained via eqn. 2.14 by simply replacing the number of excitations of each dye, N_D and N_A, with the number of detected photons in the respective channel, represented by the observed photon counts F_D and F_A:

$$E = \frac{N_\mathrm{A}}{N_\mathrm{D} + N_\mathrm{A}} \neq \frac{F_\mathrm{A}}{F_\mathrm{D} + F_\mathrm{A}}. \tag{2.15}$$

The detected photon counts, F, can be described by the number of excitations of a fluorophore, N, multiplied by the probability of detecting a photon per excitation, $\Phi = \phi \cdot \eta$,

$$F_\mathrm{A} = N_\mathrm{A}\,\Phi_\mathrm{A} \qquad \text{and} \tag{2.16}$$

$$F_\mathrm{D} = N_\mathrm{D}\,\Phi_\mathrm{D}. \tag{2.17}$$

Insertion into eqn. 2.15 yields:

$$E = \frac{N_\mathrm{A}}{N_\mathrm{D} + N_\mathrm{A}} = \frac{\frac{F_\mathrm{A}}{\Phi_\mathrm{A}}}{\frac{F_\mathrm{D}}{\Phi_\mathrm{D}} + \frac{F_\mathrm{A}}{\Phi_\mathrm{A}}} = \frac{F_\mathrm{A}}{\gamma\, F_\mathrm{D} + F_\mathrm{A}}, \tag{2.18}$$

with the so called γ-factor:

$$\gamma = \frac{\Phi_\mathrm{A}}{\Phi_\mathrm{D}} = \frac{\phi_\mathrm{A}\eta_\mathrm{A}}{\phi_\mathrm{D}\eta_\mathrm{D}}. \tag{2.19}$$

The exact determination of the γ-factor is of great experimental interest. Possible methods for burst analysis experiments with non-immobilized molecules are explored in section 2.4.7.

2.2.4 Hidden Assumptions

The $\langle\kappa^2\rangle = {}^2\!/_3$ Assumption

The validity of the $\langle\kappa^2\rangle = {}^2\!/_3$ assumption and its implications on the method of distance determination via FRET has been the topic of intense discussion[175–178]. The averaging over all dipole orientations that leads to $\langle\kappa^2\rangle = {}^2\!/_3$ is only valid if the rotation happens on a time scale much smaller than that of the energy transfer process. In aqueous solutions, rotational diffusion of small molecules such as dye molecules happens on the picosecond time scale, whereas the FRET process occurs on time scales similar to the fluorescence lifetimes of organic dyes, which is on the order of a few nanoseconds[179]. It is, however, not clear if dye molecules tethered to a protein enjoy the same kind of rotational freedom as freely diffusing molecules. The often hydrophobic dye molecules may stick to hydrophobic areas on the surface of a protein, especially if the protein is already partially unfolded and exposes parts of its hydrophobic core. The dye exists in an equilibrium between a state in which it sticks to the protein and a state in which it can rotate freely. In addition, the motion of the dye may also be restricted by steric hindrance as described by motion-in-a-cone models[180,181] in which the orientation of the transition dipoles is confined to within a particular solid angle. An accurate determination of absolute distances with FRET is, therefore, not feasible without careful characterization of the mobility of both dyes. The validity of the $\langle\kappa^2\rangle = {}^2\!/_3$ assumption can be checked via anisotropy measurements of both dyes in the following way[182,183]:

If an ensemble of randomly oriented dyes is excited by polarized light, not all dye molecules are excited with equal probability. The smaller the angle between the transition dipole moment of a dye molecule and the axis of polarization of the incident light, the more likely the excitation of a dye molecule becomes. Dyes with a transition dipole moment perpendicular to the axis of polarization are not excited at all. Analogously, the polarization of the emitted light of any molecule also depends on the orientation of the transition dipole at the moment of emission. Thus, dyes with a fixed spatial orientation that are excited with linearly polarized light do not emit light of all polarizations with equal probability. Thus, the emission is anisotropic. If, however, the excited molecules have the ability to tumble around during the lifetime of the excited state, the directional dependence of the polarization of the emitted light on the polarization of the excitation light is lost over time. Due to the rotational diffusion of the molecules, the orientations of the excited molecules equilibrate until all directions of polarization are equally probable and the rate of decay of the anisotropy is a measure of rotational mobility of the fluorescent dye. Freely rotating dyes only exhibit one, fast anisotropy decay component. If, on the other hand, the dye population is (at least partially) attached to the protein or can only access a restricted number of orientations within a cone, a slow component describing protein motion appears in the anisotropy decay.

The Ideal Dipole Approximation (IDA)

In the derivation of the Förster formula, the transition densities of donor and acceptor are assumed to be ideal dipoles with no spatial extension. Higher terms of the multipole expansion are omitted. In reality, however, the electronic system of a dye molecule and thus its transition density extends over several Ångströms and cannot be described by only a dipole term. Studies show the ideal dipole approximation (IDA) to work reasonably well for distances $> 20\,\text{Å}$. However, this is only due to a cancellation of errors and requires full orientational averaging. In fact, for specific relative orientations, errors from the IDA are shown to be significant even beyond intermolecular distances of 50 Å[184,185]. Thus, the IDA and the $\langle\kappa^2\rangle = 2/3$ assumption both rely on the assumption of fast isotropic averaging of the relative orientation of the dyes.

2.2.5 The Solvent Dependence of the Förster Distance

In the way eqn. 2.10 is formulated, it may be tempting to interpret changes in FRET efficiency to be due only to changes in the inter-dye distance. One should be aware that the Förster distance R_0 is not necessarily a constant under variable experimental conditions. The protein denaturation experiments described in this thesis involve adding increasing concentrations of the chemical denaturant guanidinium chloride (GdmCl) to the sample solution. It is therefore important to check if R_0 is dependent on GdmCl concentration.

To compare FRET efficiencies obtained using different Förster distances, the following correction is employed. For a population of molecules, an initial FRET efficiency E_i is measured under initial conditions, which results in an initial Förster distance of $R_{0,\mathrm{i}}$. Solving eqn. 2.10 for R one obtains the inter-dye distance from E_i and $R_{0,\mathrm{i}}$. To get the FRET efficiency E_f that a molecule with inter-dye distance R would exhibit under different environmental conditions, all that is necessary is another application of eqn. 2.10, with the Förster distance $R_{0,\mathrm{f}}$ that corresponds to the new environment. The process is visualized in fig. 2.7 and calculated according to:

$$
\begin{aligned}
E_\mathrm{f} &= E_\mathrm{f}(R) = E_\mathrm{f}\left(R\left(E_\mathrm{i}\right)\right) && (2.20)\\
&= \left[1+\frac{R^6}{R_{0,\mathrm{f}}^6}\right]^{-1} && (2.21)\\
&= \left[1+\underbrace{\left(R_{0,\mathrm{i}}\sqrt[6]{\frac{1}{E_\mathrm{i}}-1}\right)}_{R}^{6}\Big/R_{0,\mathrm{f}}^6\right]^{-1} && (2.22)\\
&= \left[1+\left(\frac{R_{0,\mathrm{i}}}{R_{0,\mathrm{f}}}\right)^6\left(\frac{1}{E_\mathrm{i}}-1\right)\right]^{-1}. && (2.23)
\end{aligned}
$$

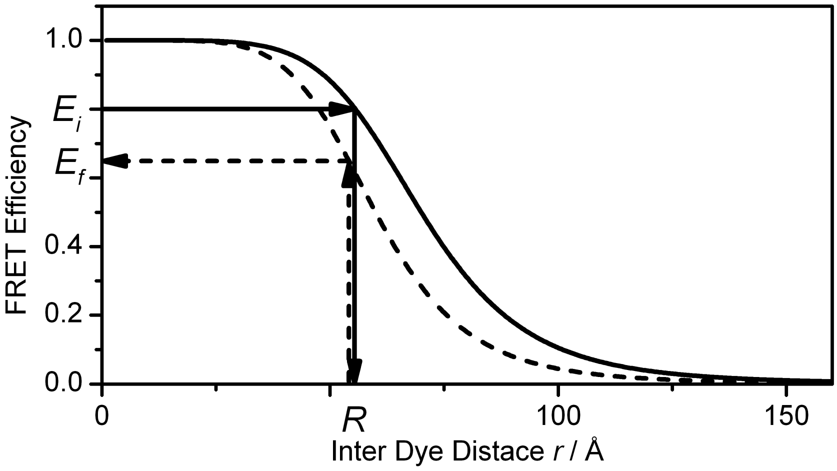


Figure 2.7: FRET efficiencies calculated according to eqn. 2.10 plotted for Förster radii of 60 Å (dashed line) and 70 Å (solid line). Starting from an initially measured FRET efficiency E_i, a distance R is calculated for a sample according to the solid curve. Changes in the Förster radius result in a different efficiency-distance relationship (dashed curve), leading to a different observed efficiency E_f for a sample of the same inter-dye distance. The size of the Förster radius shift of 10 Å chosen here is untypically large to make the effect better visible.

As is obvious from fig. 2.8, the magnitude of this correction depends on the initial FRET efficiency E_i and has its biggest effect around $E_\mathrm{i} = 0.5$.

Changes in Refractive Index n

The refractive index n of the protein solution increses with increasing GdmCl concentrations, influencing the Förster distance R_0, which is dependent on n according to $R_0 \propto n^{-4/6}$. Care has to be taken to assign the effective refractive index n_eff experienced by the dye

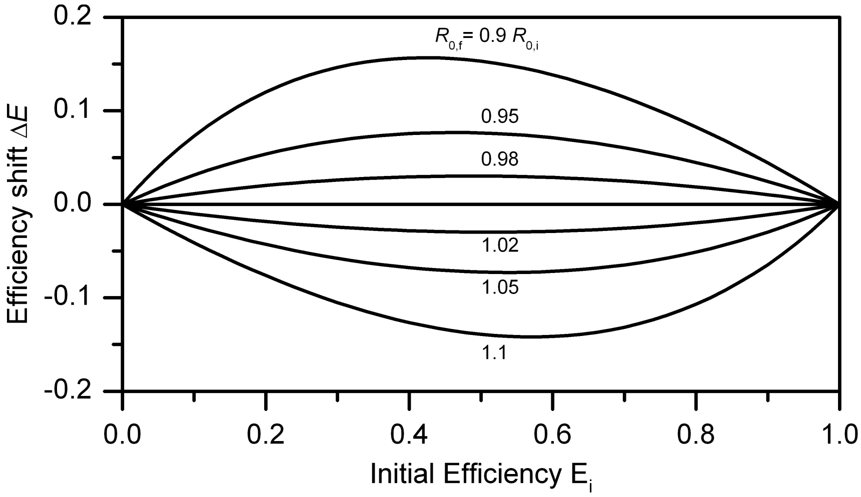


Figure 2.8: Shift in FRET efficiency, $\Delta E = E_\mathrm{f} - E_\mathrm{i}$, where E_f is calculated from E_i according to eqn. 2.23. The larger the difference between initial and final Förster distance ($R_\mathrm{0,i}$ and $R_\mathrm{0,f}$), the more pronounced the shift in FRET efficiency. The ratio $\frac{R_\mathrm{0,i}}{R_\mathrm{0,f}}$ is given next to each curve.

pair[186]. If the protein is folded, the volume between the dyes is mostly filled with protein matter, exhibiting a refractive index of $n_\mathrm{Prot} \approx 1.6$[187]. However, the dipole field, the carrier of the FRET interaction, extends into the surrounding medium as well, a volume that is made up by the buffer solution with a refractive index $n_\mathrm{Buff}(c)$ between 1.35 and 1.44 depending on the GdmCl concentration c. In denatured proteins, which are often imagined as extended random coils, the volume between both dyes is in fact filled with a mixture of both protein matter and buffer solution. The straightforward approach is to assume a weighted sum of the two components, with the buffer fraction p:

$$n_\mathrm{eff} = p\, n_\mathrm{Buff}(c) + (1-p)\, n_\mathrm{Prot}. \tag{2.24}$$

The $R_0 \propto n^{-4/6}$ relationship transforms this linear dependency of n_eff on c into a nonlinear one for the Förster distance. However, the effect is very subtle and $R_0(c)$ can be approximated as a linear relationship. For a reasonable assumption of $p = 0.5$, R_0 is only reduced by 2% at 6 M GdmCl when compared to the denaturant free solution.

Changes in the Quantum Yield ϕ_D of the Donor Dye

The dependence of the quantum yield ϕ_D of the donor on denaturant concentration c, is obviously dependent on the dye used as well as a host of other environmental parameters such as pH or buffer composition and, as such, has to be checked individually for each experiment. The topic of quantum yield determination is discussed in depth in section 2.4.7. As is evident from fig. 2.9, ϕ_D can exhibit a significant denaturant dependence and, in this special case, increases by a factor of 2.55. Although the sixth root dependence

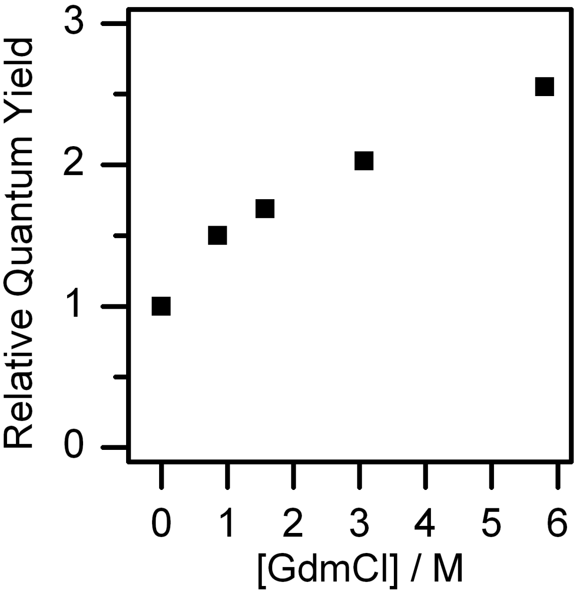


Figure 2.9: Relative quantum yield of Alexa Fluor 546 as a function of GdmCl concentration. Measurement in standard buffer at pH 3.0. Relative errors are approximately 2% for all points in the series. Error bars are much smaller than the symbols.

($R_0 \propto \phi_{\mathrm{D}}^{1/6}$) softens the impact on the Förster distance, this shift in ϕ_{D} still translates into a 17% increase of R_0 at the high end of GdmCl concentrations.

Changes in the Overlap Integral J

In the same vein as the quantum yield, other properties of fluorescent dyes such as the excitation and emission spectra may be sensitive to the GdmCl concentration. This may lead to changes in the overlap integral J. As GdmCl concentration increases, both the emission spectrum of Alexa Fluor 546 and the excitation spectrum of Alexa Fluor 647 exhibit a noticeable peak shift towards longer wavelengths. To illustrate the size of these shifts, the relevant spectra of the Alexa Fluor 546/647 dye pair are shown for 0 M and 6 M GdmCl concentrations in fig. 2.10. The overlap integral J of the Alexa Fluor 546/647 dye pair, calculated according to eqn. 2.12, decreases by up to 7% (at [GdmCl] = 6 M). However, because $R_0 \propto J^{1/6}$, the resulting effect on the Förster distance is reduced to $\approx 1\%$.

Changes in the Orientation Factor κ^2

The least accessible and controllable parameter that enters into the calculation of R_0 is κ^2. The interaction between the dyes and the labeled protein is most likely solvent dependent. The $\langle\kappa^2\rangle = 2/3$ assumption may or may not be valid, depending on the solvent composition of the experiment. In consequence, if and how much κ^2 deviates from its ⅔ average value may dependent on the solvent. There is no simple estimation of the effect that solvent changes can have on κ^2. Instead, κ^2 has to be controlled independently at all solvent environments via anisotropy measurements.

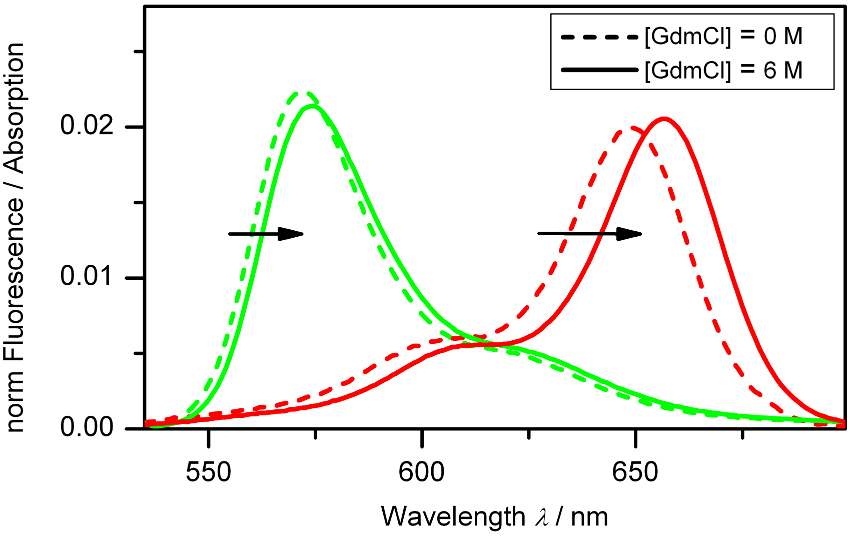


Figure 2.10: Fluorescence and absorption spectra for the Alexa Fluor 546/647 dye pair at pH 3.0 in standard buffer. The fluorescence spectrum of Alexa Fluor 546 (green) and the absorption spectrum of Alexa Fluor 647 (red) are shown for 0 M GdmCl (dashed lines) and 6 M GdmCl (solid lines), all normalized to an area of one.

2.2.6 Samples with Variable Inter-Dye Distance and Slow Dye Dynamics

In the simple $E(R)$ relationship given by the Förster equation (eqn. 2.10), the distance R is considered to be constant. While this may be a reasonable assumption for folded proteins, it breaks down as soon as at least parts of the protein unfold. An unfolded protein lacks the well defined structure of its folded counterpart and the distance R between donor and acceptor varies with time. The inter-dye distance has to be described as a distribution of distances with a probability density function $P(R)$.

In unfolded molecules, the inter-dye distance can be replaced by an average value if the unfolded chain samples its conformation space on time scales faster than the inverse rate of energy transfer τ_{FRET}, a quantity on the order of the magnitude of fluorescence lifetimes (10^{-9} – 10^{-8} s). At the same time, the treatment of κ^2 as a constant with the value of $\kappa^2 = {}^2\!/_3$, is only applicable if the rotational dynamics of both dyes are faster than τ_{FRET}.

If either assumption breaks down, eqn. 2.10 is no longer applicable in its pure form and has to be corrected. Kuzmenkina *et al.* derived the corrections to eqn. 2.10 for the case that both dye reorientation and inter-dye distance variations happen on a time scale that is slower than the fluorescence lifetime of the dyes but faster than the time resolution of the experiment, in this case the bin-time of the fluorescence trace[113,188]. In a straightforward way, the averaging process over the conformational space in R and κ^2 is expressed as

$$\langle E\rangle = \int_0^{l_c}\int_0^4 \frac{1}{1+\frac{2/3}{\kappa^2}\left(\frac{R}{R_0}\right)^6}\,P(R)\,p(\kappa^2)\,d\kappa^2\,dR, \tag{2.25}$$

where $P(R)$ and $p(\kappa^2)$ are probability density functions for inter-dye distance and orientation factor, respectively. The upper limit l_c of the integral over R is the contour length of the protein chain between the dyes, i.e., the maximally possible inter-dye distance in the case of a completely stretched chain. The Förster distance R_0 is calculated according to eqn. 2.11 with a value of $\kappa^2 = {}^2\!/_3$.

First, let us assume that R remains fixed (represented by $P(R)$ collapsed to a delta function for the purpose of evaluating eqn. 2.25), and only variations in κ^2 are considered. If all spatial orientations of both dyes are equally probable, the isotropic probability density for κ^2 is given by[94]:

$$p(\kappa^2) = \begin{cases} \frac{1}{2\sqrt{3\kappa^2}}\ln(2+\sqrt{3}) & \text{for} \quad 0 \le \kappa^2 \le 1 \\ \frac{1}{2\sqrt{3\kappa^2}}\ln\left(\frac{2+\sqrt{3}}{\sqrt{\kappa^2}+\sqrt{\kappa^2-1}}\right) & \text{for} \quad 1 < \kappa^2 \le 4 \end{cases} \tag{2.26}$$

The resulting $E(R)$ dependency can be obtained by numerically integrating eqn. 2.25 and is depicted as the dotted line in fig. 2.11. It is less steep than the original Förster relation and reduces the FRET efficiency for all distances.

Another regime arises when the inter-dye distance is free to vary. The unfolded state of most proteins can, at least in first approximation and at high denaturant concentrations, be described with a Gaussian chain model[69], for which the probability density function of the end-to-end distance is given by

$$P(R) = \frac{4\pi R^2}{(2\pi R_a^2)^{3/2}}\exp\left[-\frac{3R^2}{2R_a^2}\right]. \tag{2.27}$$

Here R_a is $\langle R^2\rangle^{1/2}$, the root-mean-square average of R. It has been shown that this model can still be feasible for protein states with significant amounts of residual structure[81], making it applicable to unfolded states at lower denaturant concentrations. The protein covers this conformational space on a time scale between dye photophysics and the bin-time of a burst analysis experiment. Numerical integration yields $E(R_a)$, the dashed line in fig. 2.11. This solution can be approximated by a function of the same form as eqn. 2.10[113]:

$$E = \frac{1}{1+0.975\left(\frac{R_a}{R_0}\right)^{2.650}} \qquad \text{for} \quad \frac{R_a}{R_0} < 3. \tag{2.28}$$

This analytical function offers an easy conversion of FRET efficiencies into mean inter-dye distances R_a, e.g., for unfolded proteins.

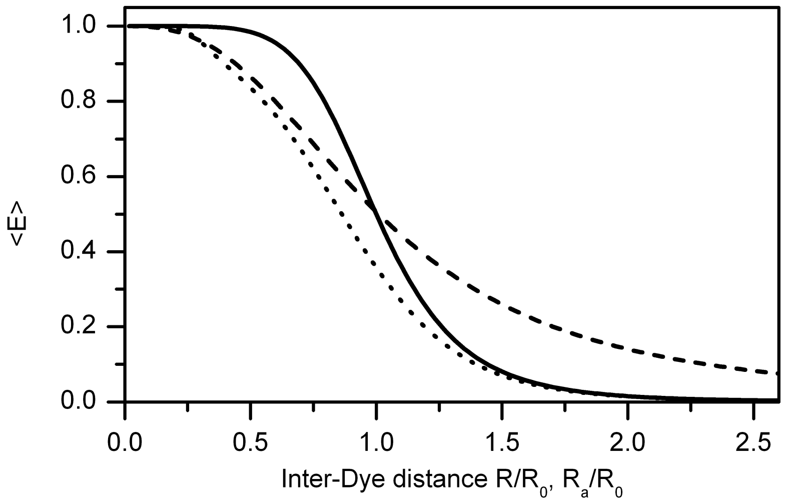


Figure 2.11: The distance-FRET efficiency relationship for different experimental conditions. The classical Förster formula (solid line) represents the regime in which dye distance and κ^2 are constant, or average faster than the fluorescence lifetime of the dyes. The dotted line represents the case where this still holds for R, but κ^2 averages on a time scale in between the dye lifetime and the experimental time resolution. Allowing the inter-dye distance to vary as well on the same time scale (dashed line), leads to the dependence given by eqn. 2.28.[113]

2.3 Confocal Microscopy

2.3.1 The Confocal Principle

The foundations of confocal microscopy were laid in the 1950s when Marvin Minsky introduced the basic confocal principle (and patented it in 1957)[189]. Confocal microscopy is performed in an epi-fluorescence setup, which means that excitation and collection of fluorescence light are done with the same objective. As a consequence, not the whole sample is illuminated but the excitation beam is focused onto a small, diffraction limited spot in the sample. However, regions above and below the focal spot still get excited, although the intensity of illumination decreases with increasing distance from the focal plane. To restrict the observation volume, the confocal pinhole is introduced to block emission light from non-focal planes from entering the detector, as illustrated in fig. 2.12. Only light from the focal plane is imaged onto the plane of the confocal pinhole and can pass the pinhole relatively unhindered. The majority of light from other planes, however, is rejected. A confocal microscope only delivers information about one specific point of the sample. To image larger areas, either the beam or the sample have to be scanned, and the picture reconstructed by a computer.

2.3.2 A Typical Confocal Microscopy Setup

The setup used for this work is schematically depicted in fig. 2.13. Laser 1 excites the donor dye, whereas Laser 2 is used to directly excite the acceptor dye. Both beams are combined using a dichroic mirror and fed into a single mode optical fiber. For advanced single molecule techniques, such as *alternating laser excitation*[190,191] or *pulsed interleaved excitation*[192] (ALEX and PIE, explained in sections 2.4.4 and 2.4.4), the excitation is required to alternate rapidly between the two excitation wavelengths. One way to achieve this is to use pulsed laser sources, synchronized so as to pulse in an alternating fashion. Alternatively, an acousto-optical tunable filter (AOTF) can be used to generate pulsed output from continuous wave lasers. The excitation dichroic mirror, designed to exclusively reflect light in a small window around the excitation wavelengths, reflects the excitation light towards a microscope objective with high numerical aperture, which focuses the excitation light into a diffraction limited spot inside the sample solution. Fluorescent light is collected with the same objective and can, because it is Stokes shifted towards longer wavelengths, pass through the excitation dichroic.

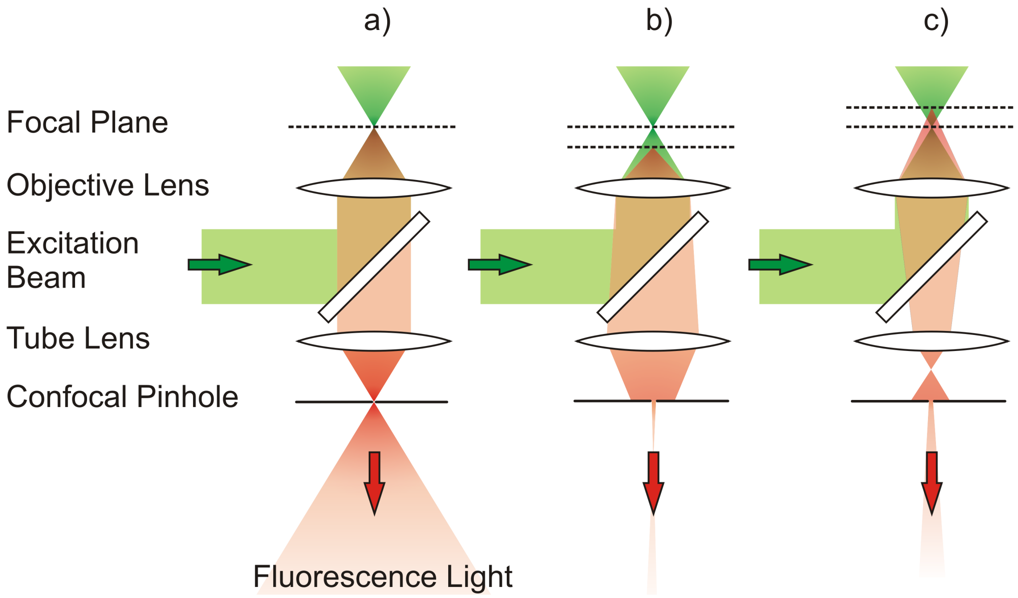

a)
b)
c)
Focal Plane
Objective Lens
Excitation Beam
Tube Lens
Confocal Pinhole
Fluorescence Light

Figure 2.12: The confocal principle. Excitation light is focused by the objective lens onto the focal plane. Fluorescence light is collected with the same objective and separated from the excitation light via a dichroic mirror. The optical system consisting of the objective and the tube lens images the focal plane onto the plane in which the confocal pinhole is installed. **a)** A point source on the optical axis in the focal plane is thus refocused exactly on the pinhole and most of its fluorescence light passes through to the detectors. **b)** and **c)** Light emitted from planes above or below the focal plane, however, is focused above and below the confocal pinhole. A large amount of fluorescence from these planes is consequently blocked.

It is then refocused onto the confocal pinhole, filtering light from non-focal planes according to the confocal principle explained above. The emission dichroic separates fluorescence from donor and acceptor, based on the emission spectra of both dyes. Additional bandpass filters in both paths limit spectral crosstalk and block scattered excitation light. Finally, the light is collected by single photon-sensitive detectors, connected to a computer, where the intensity is recorded as a function of time. These time traces are subsequently subjected to burst detection and FRET analysis.

2.4 Single Molecule FRET measurements

To measure the FRET efficiency of single molecules (smFRET), a method called *burst analysis* is used. The labeled sample is diluted to concentrations in the picomolar range so that the probability of two proteins being present in the focal volume at the same time becomes negligibly small. Single molecules will statistically diffuse through the focal volume of the microscope of roughly 1 fl and emit a burst of fluorescence light as they pass. From this burst of photons, the FRET efficiency of the single molecule can be calculated. The transit times of small biomolecules such as RNase H in aqueous solutions at room temperature are typically several 100 µs, during which some tens of photons can be detected. An example of the raw data collected during a burst analysis experiment is shown in fig. 2.14.

2.4.1 Burst Detection

Burst detection is the process of identifying those bins in a time trace that contain photon counts from sample molecules passing through the focus and of separating them from the vast majority of bins that contain only background noise. To decide which bins constitute a burst, a somewhat arbitrary threshold n_{min} has to be set. All bins containing more photons than this threshold are considered to contain meaningful data, whereas bins with less photons are discarded. Obviously, the threshold should be chosen dependent on the overall noise level of the trace. To make a sensible choice for n_{min}, it is useful to consider how the noise is distributed.

The background noise is mainly due to *dark counts* of the detectors, false counting events that are generated by the detectors themselves, even in the absence of any light. Additional background may be due to scattered excitation light. A small fraction will pass the detection filters and contribute to the signal background. In addition, stray light may enter the setup and be detected. In such an experiment, the molecules of interest are diluted to such a degree that bursts are well separated in time, so the number of bins that contain only background photons is much larger than that of bins containing light due to actual burst events. A photon counting histogram of a time trace is therefore dominated by background noise. Because background noise consists of independent stochastic events with a constant

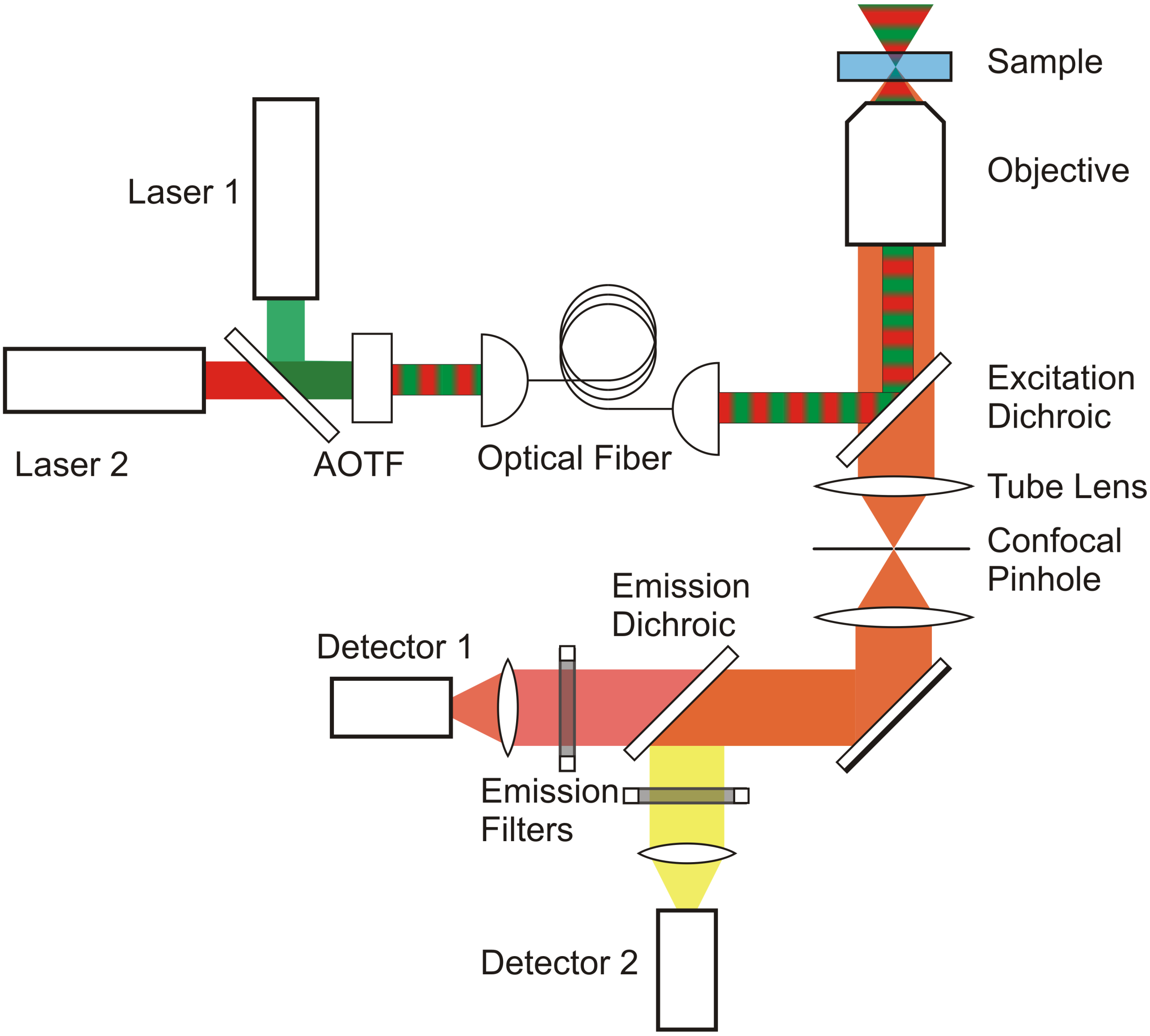

Sample
Objective
Laser 1
Excitation
Dichroic
Laser 2
AOTF
Optical Fiber
Tube Lens
Confocal
Pinhole
Emission
Dichroic
Detector 1
Emission
Filters
Detector 2

Figure 2.13: Schematic setup of a confocal microscope. Excitation light is combined, modulated, led to the microscope via an optical fiber and focused into the sample by an objective lens with a high numerical aperture. Fluorescence from the focal region is collected by the same objective, passed through the focal pinhole, separated into donor and acceptor emission and subsequently detected.

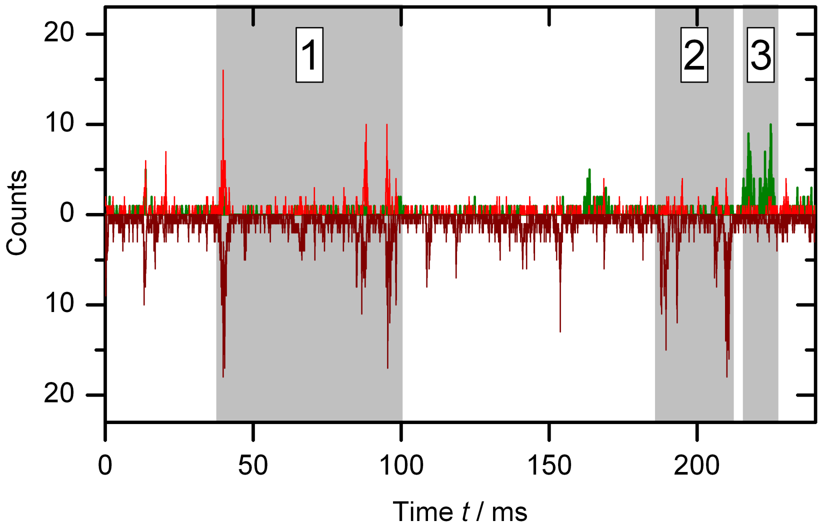


Figure 2.14: Fluorescence time trace of labeled RNase H at pH 7.2 and 0 M GdmCl with a bin size of 100 µs. Signals in the donor (green) and acceptor (light red) channels are plotted in the upper half of the panel, the signal of the acceptor channel upon direct excitation (dark red) is plotted with negative amplitude in the lower half. Fluorescence bursts of properly labeled molecules in a high-FRET state are shown in interval 1. Upon donor excitation, the acceptor emission dominates, presence of an acceptor dye is evidenced by the signal upon acceptor excitation. Acceptor-only bursts only show up upon acceptor excitation. They are caused by molecules labeled with only acceptor dyes (2). Molecules labeled only with a donor dye give rise to donor-only peaks. They show no acceptor component upon donor excitation and no signal at all upon acceptor excitation (3).

average rate, the number of photons per time bin, i.e., the sum over these events within a certain amount of time, is described by a Poisson distribution (fig. 2.15). The mean μ of the Poissonian describes the average background of the trace and is calculated for each channel separately. The Poissonian also delivers a meaningful criterion for choosing n_{min}, depending on how high a risk of mistaking noise for real burst events one is willing to take. We decided to include bins for which the Poissionian distribution of background noise has dropped below 0.1% (marked as dashed lines in fig. 2.15). If several adjacent bins all contain more photons than n_{min}, they are considered to be part of the same molecule transit event and their photon count values in each channel are added. In the end, a burst is characterized by three numbers: The number of detected photons in each channel upon donor excitation and the number of detected photons in the acceptor channel upon direct excitation of the acceptor.

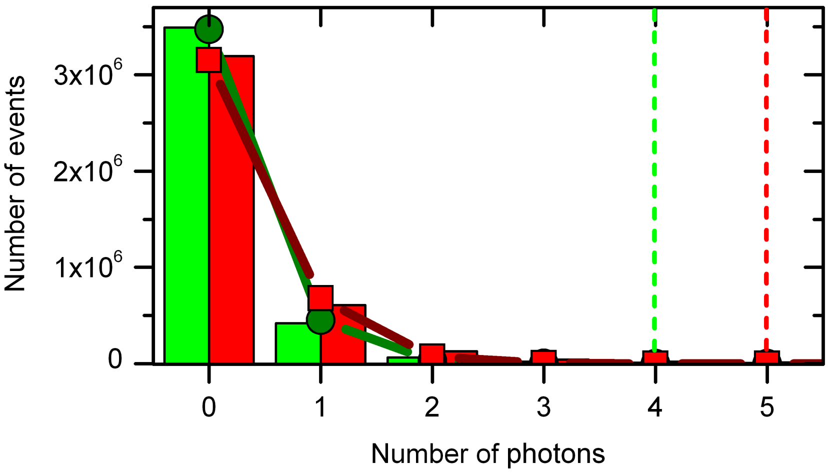


Figure 2.15: Photon counting histograms of donor and acceptor channels of a 400 second time trace (green and red columns). Connected symbols in the respective colors are least squares fits of Poissonian distributions to these data sets. Means of the Poissonians are $\mu_{\mathrm{D}} = 0.13$ for the donor and $\mu_{\mathrm{A}} = 0.22$ for the acceptor channel. The noise threshold, marked as dashed lines for the two channels, is set to the value where the Poissonian distributions have dropped below 0.1%. Bins above that threshold are considered to be part of a burst.

2.4.2 Selection of Bright Bursts

As diffusion is a stochastic process, some molecules will visit only the edges of the detection volume, while others may stay in the center for longer times. In consequence, there is no typical number of photons per burst, but instead, the "size" of the bursts is distributed, as visualized in a burst-size histogram (BSH) in fig. 2.16. An arbitrary threshold, n_{min}, is chosen that represents the minimal number of photons for a burst to be considered for further analysis. For each burst that is accepted, a FRET value E is calculated separately. All FRET values are then collected in a histogram (for an example, see fig. 2.18). When choosing n_{min}, a balance has to be found between the following effects:

The width of any peak in a smFRET histogram will be due to a variety of reasons, a detailed list of which is given in section 4.2.4. One of the contributing factors is the uncertainty of E due to *shot noise*. Whenever a signal is formed as a sum of independent, stochastic events, the average signal carries an inherent statistical uncertainty, the so-called shot noise. When a burst of fluorescence light is detected, the photon counts of both the donor and the acceptor channel constitute such signals. Consequently, the FRET efficiency, E, calculated from these photon counts, is also subject to statistical fluctuations. Shot noise is unavoidable due to the nature of light emission and detection in the form of distinct and independent photons.

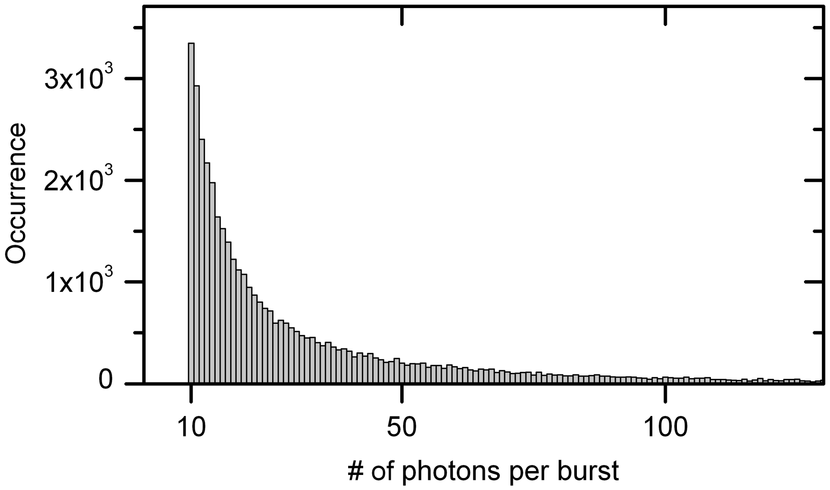


Figure 2.16: Burst-size histogram (BSH) of a 30 min measurement of RNase H at pH 7.2 and a GdmCl concentration of ≈ 1.5 M. Only bursts containing a minimum of 10 detected photons are considered for this histogram. The mean number of photons per burst in this case is 36.

It is, thus, the source of the minimal standard deviation, σ_{SN}, of any distribution in a smFRET histogram. The standard deviation due to shot noise can be calculated as[121]

$$\sigma_{\mathrm{SN}} = \left[\frac{E(1-E)}{n_{\mathrm{tot}}}\right]^{1/2}, \tag{2.29}$$

which is dependent on the total number of photons n_{tot} per burst. The behavior of σ_{SN} for several values of n_{tot} is depicted in fig. 2.17. So, on the one hand, a higher n_{min} results in a higher average number of photons per burst, which in turn leads to lower shot noise and narrower distributions. Increasing n_{min}, on the other hand, reduces the overall number of events in the final smFRET histogram, which leads to poor statistics and ill-defined fits. In this work, a burst had to contain at least 50 photons to be considered for further analysis.

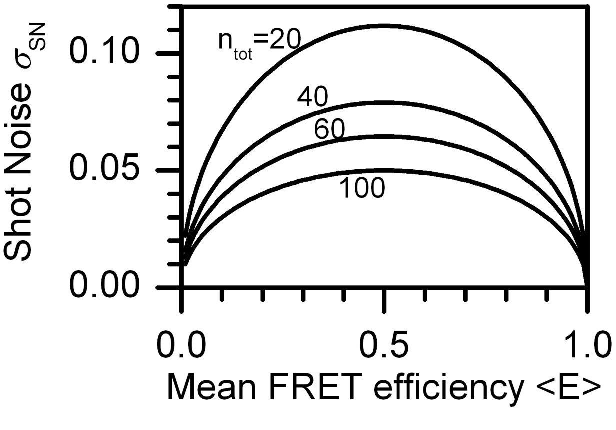


Figure 2.17: Shot noise contribution to the width of smFRET distributions, depending on the mean FRET efficiency of the distribution. The number of photons per burst, n_{tot}, is given next to each curve.

2.4.3 Burst Wise FRET Calculation

Using eqn. 2.18, the FRET efficiency value of each burst was calculated separately from the photon counts in donor and acceptor channel, F_{D} and F_{A}. The FRET efficiency values of all bursts in a measurement were collected in a smFRET histogram, such as the one depicted in fig. 2.18.

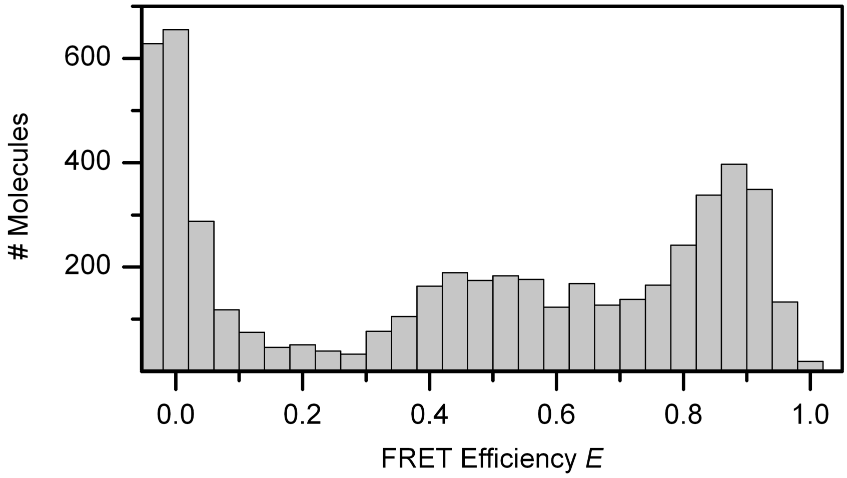


Figure 2.18: Example of a smFRET histogram (RNase H at pH 7.2, [GdmCl] = 1.73 M). The histogram reveals at least three distinct distributions. The population around $E = 0$ is most likely due to donor-only molecules.

2.4.4 The Donor-Only Problem

One problem in the burst analysis is the presence of so-called donor-only molecules. These particles only show up in the donor channel under donor excitation. A few of these particles may be fluorescent contaminants. Most are, however, improperly labeled proteins that lack an acceptor dye and are labeled with one or two donor fluorophores. These donor-only molecules give rise to a peak at zero FRET efficiency in FRET histograms, possibly obscuring data from properly labeled molecules at low FRET efficiencies (see fig. 2.18).

Alternating Laser Excitation

To overcome the donor-only problem, Kapanidis and coworkers introduced a technique called *Alternating Laser Excitation* (ALEX)[190,191]. On a time scale shorter than the transit time of a molecule through the detection volume, the excitation wavelength is switched so as to excite donor and acceptor dye in an alternating fashion. By synchronizing the detectors with the excitation it is possible to discern which fraction of the detected light is emitted upon excitation of the donor dye, and which fraction is due to the direct excitation of the acceptor. During donor excitation, donor and acceptor photon counts (F_{D}, F_{A}) are recorded as previously described and used to calculate the FRET efficiency according to eqn. 2.18. Upon direct acceptor excitation, only the photon count in the acceptor channel $F_{\mathrm{A,dir}}$ is collected. From these data, the stoichiometry S of each burst is calculated as

$$S = \frac{\gamma F_{\mathrm{D}} + F_{\mathrm{A}}}{\gamma F_{\mathrm{D}} + F_{\mathrm{A}} + F_{\mathrm{A,dir}}} = \frac{F_{\mathrm{tot}}}{F_{\mathrm{tot}} + F_{\mathrm{A,dir}}}. \tag{2.30}$$

All observed bursts are collected in a two-dimensional histogram called an E-S map with FRET efficiency E and stoichiometry S on the two axes (fig. 2.19). As donor-only molecules lack an acceptor dye, they exhibit no FRET ($E = 0$). In addition, donor-only molecules show no fluorescence emission when illuminated with acceptor excitation light ($F_{\mathrm{A,dir}} = 0$) and therefore have a stoichiometry of $S = 1$. Thus, donor-only molecules are situated in the upper left corner of the histogram at $E = 0$ and $S = 1$ (marked (1) in fig. 2.19). Molecules that possess an acceptor, on the other hand, will exhibit acceptor intensity under direct excitation and, therefore, have a reduced stoichiometry value ($F_{\mathrm{A,dir}} > 0$ and $S < 1$). Such events are marked by (2) in fig. 2.19. This representation allows separation of donor-only bursts from meaningful bursts generated by molecules that possess both donor and acceptor. In fig. 2.19, a possible threshold has been marked by a white line at $S = 0.6$. Only bursts below that line would be considered for further analysis.

$F_{\mathrm{A,dir}}$ is directly dependent upon the power of the acceptor excitation. The S-position of meaningful bursts and thus their separation in stoichiometry from donor-only bursts can therefore be adjusted via the power of the laser used for acceptor excitation. One of the requirements of the ALEX detection scheme is that the excitation can be switched on a time scale faster than the observation time. This is possible with acousto-optical modulators which are able to switch between illumination beams with a dead time of $\approx 7\,\mu\mathrm{s}$. However, the switching rate cannot be increased arbitrarily. At a period of $100\,\mu\mathrm{s}$ for an illumination cycle (as chosen in this work), the dead time of switching the illumination twice, already amounts to a loss of observation time and therefore a loss of signal of 14%. This reduction in signal gets worse as the switching frequency and with it the relative dead time increases.

The observation time of the burst analysis method is limited by the time that the freely diffusing molecules stay in the confocal observation volume. Transit times of globular proteins such as RNase H with molecular weights in the range of some 10 kDa through the focus of a confocal microscope are typically in the range of a few hundred microseconds. As the burst detection algorithm selects bursts that contain a high number of photons, the bursts that are considered for analysis possibly have a somewhat higher residence time in the focus. Still, if ALEX is implemented with acousto-optical modulators, it only barely satisfies the requirement of fast switching. By measuring F_{tot} and $F_{\mathrm{A,dir}}$ in sequence, the observation conditions change in between the two measurements, as the molecule diffuses from one region of the observation volume to another. This change in position means that the measurements of F_{tot} and $F_{\mathrm{A,dir}}$ occur at different excitation intensities and detection efficiencies. Thus, the ratio between F_{tot} and $F_{\mathrm{A,dir}}$ may vary strongly from one burst to the other, resulting in a broadening of the FRET distribution in the E-S map along the

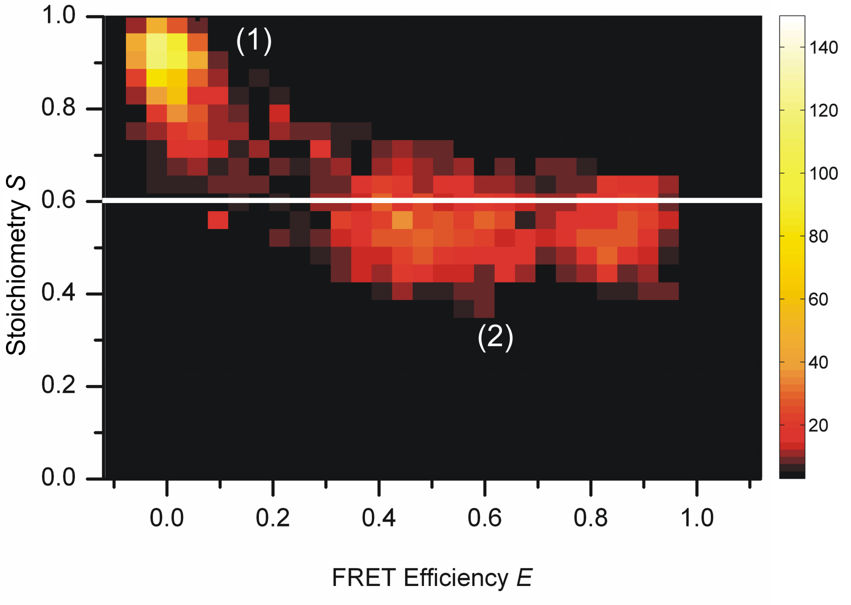


Figure 2.19: E-S map of an ALEX burst analysis experiment with RNase H at pH 7.2 and 1.89 M GdmCl. The donor-only peak (1) is caused mainly by mislabeled protein molecules, which lack an acceptor dye. The distribution due to properly labeled molecules with both donor and acceptor dyes (2) is situated at intermediate stoichiometries. The white line at $S = 0.6$ is a possible threshold to separate the two populations.

stoichiometry axis. The longer the individual observation time slots are, the more distinct the observation conditions between the different illumination schemes and the stronger the broadening becomes.

Pulsed Interleaved Excitation

The ALEX detection scheme has been refined to the *Pulsed Interleaved Excitation* (PIE) method by the group of Don Lamb[192]. Here, switching between two excitation wavelengths happens on a much shorter time scale due to the use of pulsed laser sources. As illustrated in fig. 2.20, the lasers emit pulses with well defined time delays after a signal from the reference clock (blue lines), synchronized in an alternating fashion (red and green lines). The reference clock also triggers the detection system, so that it is possible to measure the arrival time Δt of each photon with respect to the synchronization pulse. Via these arrival times, each detected photon can be associated with the laser pulse that caused it. The repetition rate of the lasers has to be chosen low enough for the fluorescence of the excited molecules to decay before the next laser pulse is triggered. Otherwise, a photon could be erroneously attributed to the subsequent excitation pulse, a process called *temporal crosstalk*. The time between two laser pulses (50 ns) is 10,000 fold shorter than the molecular transit time in burst analysis. In contrast to ALEX, the requirement of fast switching is satisfied by PIE.

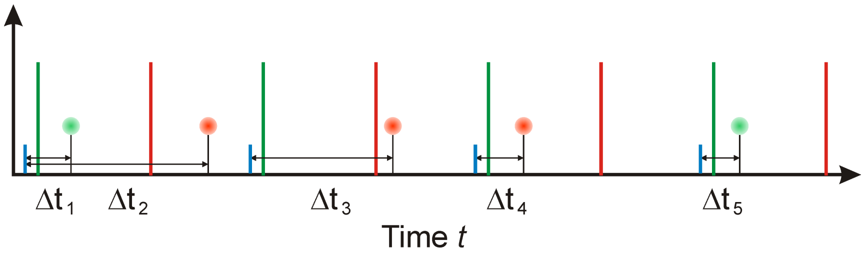


Figure 2.20: Sketch of the PIE excitation system. Pulsed lasers fire (green and red lines) at well defined time delays after a signal from the reference clock (blue lines).

As donor excitation happens immediately after the synchronization pulse, photons that arrive in the first half of the synchronization cycle are counted towards F_{D} or F_{A}, depending on which detection channel they are registered in. The acceptor excitation pulse occurs shortly after the half time of the synchronization cycle. All photons that arrive in the acceptor channel in the second half of the synchronization cycle are counted towards $F_{\mathrm{A,dir}}$ (fluorescence due to direct acceptor excitation) (see fig. 2.21).

The single photons are subsequently binned into a fluorescence time trace and subjected to burst detection. FRET efficiency E and stoichiometry S are calculated from individual bursts as previously described.

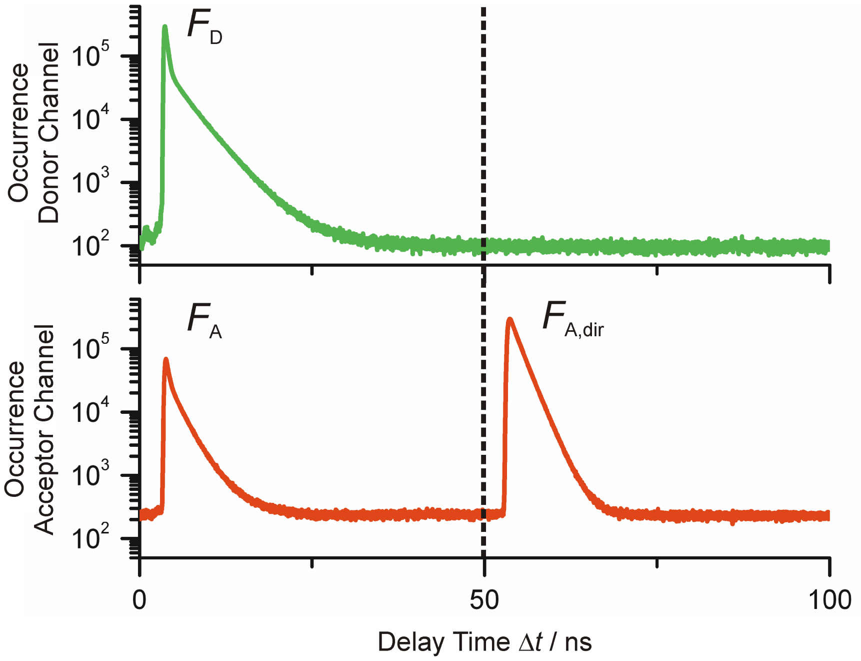


Figure 2.21: Collecting photons on a PIE setup over many pulse cycles allows the creation of arrival time histograms, shown here for a cycle length of 100 ns. Depending on their arrival time, detected photons count towards F_D, F_A ($0\,\mathrm{ns} < \Delta t \leq 50\,\mathrm{ns}$) or $F_\mathrm{A,dir}$ ($50\,\mathrm{ns} < \Delta t \leq 100\,\mathrm{ns}$). The distinction between F_D and F_A is made according to the detector channel in which a photon is detected.

2.4.5 Donor-Only Corrected smFRET Histograms

After excluding donor-only bursts via the E-S map, FRET efficiency values of all remaining bursts are collected in one donor-only corrected histogram, as depicted in fig. 2.22.

2.4.6 Correction for Spectral Crosstalk and Direct Acceptor Excitation

The excitation and emission spectra of the fluorescent dyes used in the FRET experiments are depicted in fig. 2.23. Due to the long wavelength tail of the emission spectrum of the donor, a fraction of the donor fluorescence bleeds through into the acceptor channel, as visualized in fig. 2.23, causing a systematic overestimation of the acceptor fluorescence due to crosstalk, F_CT. The absorption spectrum of the acceptor often exhibits a long tail towards higher energies. Therefore, the acceptor dye may be directly excited by the laser for donor excitation, which increases the acceptor signal by a certain amount, F_DE and, thus, causes a systematic overestimation of the acceptor fluorescence. The observed

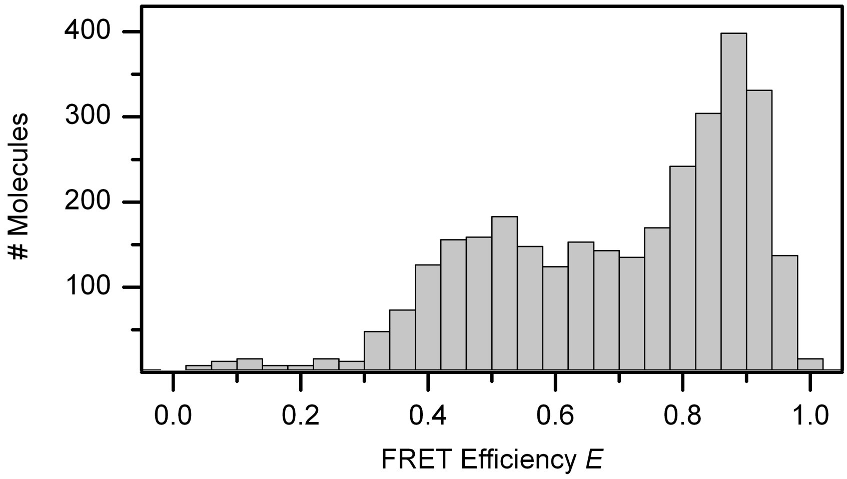


Figure 2.22: Example of a smFRET histogram (RNase H at pH 7.2, [GdmCl] = 1.73 M) after removal of the donor-only peak. The histogram reveals at least two distinct distributions.

acceptor photon count, $F_{\mathrm{A,obs}}$, has to be corrected for these two effects to obtain the value of acceptor emission due to FRET transfer from the donor, F_{A}, which is used for further analysis:

$$F_{\mathrm{A}} = F_{\mathrm{A,obs}} - F_{\mathrm{CT}} - F_{\mathrm{DE}}. \tag{2.31}$$

Determining the Spectral Crosstalk

The ratio α of donor emission that bleeds into the acceptor channel and donor emission detected in the donor channel is constant ($F_{\mathrm{CT}}/F_{\mathrm{D}} = \alpha$). It is easily accessible by performing a calibration measurement with donor-only molecules, that is, molecules which are only labeled with the donor dye. In an ALEX or PIE setup, α can be determined during donor-only bursts, rendering a separate calibration measurement unnecessary. If α is known, the cross talk fraction of the observed acceptor signal can be calculated as

$$F_{\mathrm{CT}} = \alpha F_{\mathrm{D}}. \tag{2.32}$$

Determining the Direct Acceptor Excitation

Under donor illumination, a certain amount of counts in the acceptor channel, F_{DE}, is due to direct excitation of the acceptor dye by the donor excitation laser. To determine F_{DE}, a second light source at a higher wavelength is needed which exclusively excites the acceptor dye. This means that the determination of F_{DE} is only possible in ALEX or PIE setups. The ratio between F_{DE} (the acceptor photon count due to direct excitation by the donor excitation laser) and $F_{\mathrm{A,dir}}$ (the acceptor photon count under excitation by the

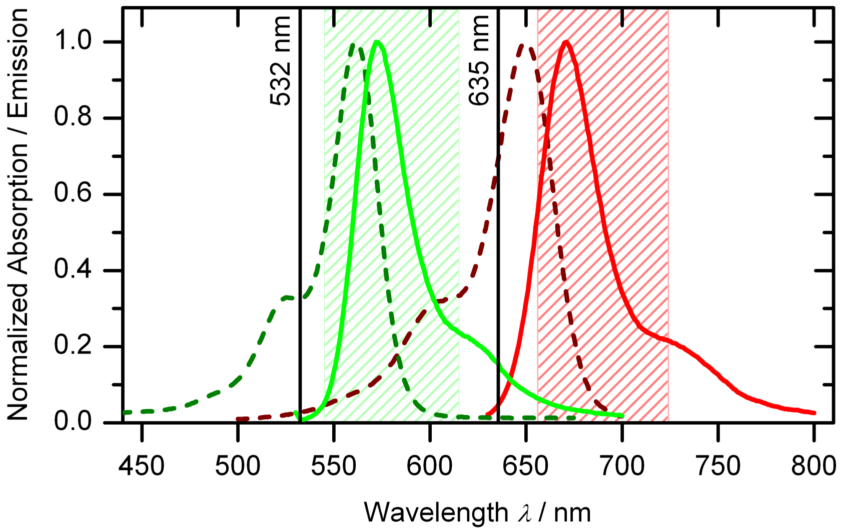


Figure 2.23: Absorption and emission spectra of the organic dyes Alexa Fluor 546 (green) and Alexa Fluor 647 (red) as provided by the manufacturer (Molecular Probes, Eugene, OR). Excitation spectra are marked as dashed, emission spectra as solid lines. The shaded areas represent the detection windows of the donor (light green) and acceptor (light red) channels, for the filterset used in PIE measurements on the Microtime 200 microscope. The excitation wavelengths of the Microtime 200 are marked by vertical black lines.

long wavelength acceptor excitation laser) is a constant, and depends on the extinction coefficients of the acceptor dye at the wavelengths of both the donor and the acceptor excitation lasers (ε_D and ε_A) and the excitation intensities of both lasers (I_D and I_A):

$$\frac{F_\mathrm{DE}}{F_\mathrm{A,dir}} = \underbrace{\frac{I_\mathrm{D}\varepsilon_\mathrm{A}(\lambda_\mathrm{D})}{I_\mathrm{A}\varepsilon_\mathrm{A}(\lambda_\mathrm{A})}}_{\beta} = \beta. \tag{2.33}$$

In a smFRET experiment, β is determined with acceptor-only labeled molecules, by calculating the ratio of their acceptor fluorescence upon excitation with both the donor and the acceptor excitation laser, $\beta = F_\mathrm{A}/F_\mathrm{A,dir}$. For the correct determination of β, the illumination powers of both donor and acceptor excitation laser, I_D and I_A, have to be the same as in the actual burst analysis experiment. With known β, the direct excitation error can be calculated as

$$F_\mathrm{DE} = \beta F_\mathrm{A,dir}. \tag{2.34}$$

Note that, despite the fact that $F_\mathrm{A,dir}$ is used to determine the amount of direct excitation, F_DE is independent of the acceptor excitation intensity I_A. The dependence of $F_\mathrm{A,dir}$ on the intensity of the direct acceptor excitation, $F_\mathrm{A,dir} \propto I_\mathrm{A}$, is counteracted by the inverse proportionality of $\beta \propto 1/I_\mathrm{A}$ (see eqn. 2.33).

2.4.7 Determination of the γ Factor

The γ factor is a correction factor used for the calculation of FRET efficiencies (see. section 2.2.3). It corrects for different quantum yields ϕ and instrumental detection efficiencies η of the donor and acceptor dyes and is calculated as:

$$\gamma = \frac{\phi_{\mathrm{A}}\eta_{\mathrm{A}}}{\phi_{\mathrm{D}}\eta_{\mathrm{D}}}. \tag{2.35}$$

Quantum yields have been published for a wide variety of commonly used organic dyes. However, the quantum yield of a dye depends on the solvent environment. Changes in pH, additives to the solvent or the fact that the dye molecule is chemically bound to the protein may influence this value. The wavelength dependent instrumental detection efficiency η of a confocal microscope is also difficult to calculate *ab initio*. While it is straightforward to obtain transmission spectra of components such as dichroic mirrors and fluorescence filters, it is more challenging to accurately assess the frequency dependent detector sensitivity or transmission rate through the confocal pinhole which may be wavelength dependent due to chromatic aberration in the optical system of the microscope.

The combination of all these difficulties makes it nearly impossible to calculate γ by using literature data for the values of ϕ and η. Directly measuring the quantum yields and the instrumental detection efficiencies separately is extremely challenging and, at the very least, time-consuming. An alternative experimental method of determining γ is presented below. For the following derivation it is useful to combine the quantum yield ϕ and the instrumental detection efficiency η into one quantity, the total detection efficiency $\Phi = \phi\eta$. Whereas the quantum yield ϕ represents the probability of a fluorescence photon being *emitted* by a fluorophore upon excitation, the total detection efficiency Φ gives the probability of the fluorescence photon being *detected* in the microscope. While ϕ is only a property of the dye molecule and its environment, Φ also depends on instrumental parameters, such as the alignment of the microscope, characteristics of the employed detectors and the fluorescence filters that are in place. The γ factor can be reexpressed with the total detection efficiencies as:

$$\gamma = \frac{\phi_{\mathrm{A}}\eta_{\mathrm{A}}}{\phi_{\mathrm{D}}\eta_{\mathrm{D}}} = \frac{\Phi_{\mathrm{A}}}{\Phi_{\mathrm{D}}}. \tag{2.36}$$

Calculation of γ via the Stoichiometry Map

In an E-S histogram, events generally are spread out along the E-axis as shown in fig. 2.19. Although the donor of all labeled molecules is on average excited the same number of times, the total photon count in both detection channels upon donor excitation, $F_{\mathrm{tot}} = F_{\mathrm{D}} + F_{\mathrm{A}}$, is not necessarily the same for all molecules. If the detection efficiencies of donor and acceptor dye are different, the total photon count, F_{tot}, of a specific molecule depends on how the excitation is distributed between the two dyes, which in turn is determined by the FRET efficiency of that molecule. If γ is chosen correctly, the corrected brightness of the donor is the same as the brightness of the acceptor dye and the corrected total photon count of both dyes, $F_{\mathrm{tot}} = \gamma\, F_{\mathrm{D}} + F_{\mathrm{A}}$, is independent of FRET efficiency. In consequence, the stoichiometry $S = F_{\mathrm{tot}}/F_{\mathrm{tot}} + F_{\mathrm{A,dir}}$ is independent on E as well. Thus, γ can be found via an iterative procedure: It is varied and the E-S map is recalculated for each new value of γ until the mean stoichiometry is the same for all FRET efficiencies. It is instructive to have a look at the proper derivation:

The stoichiometry S of a burst is calculated according to eqn. 2.30 from the observed intensities F_{D} and F_{A} in the donor and acceptor channels. These intensities are directly proportional to the number of excitations, N, that the donor and acceptor dyes experience during the burst. Proportionality factors are the respective total detection efficiencies, Φ, of both fluorophores:

$$F = \Phi N. \tag{2.37}$$

This allows expressing the stoichiometry in terms of the number of times each fluorophore is excited:

$$S = \frac{\gamma\, \Phi_{\mathrm{D}} N_{\mathrm{D}} + \Phi_{\mathrm{A}} N_{\mathrm{A}}}{\gamma\, \Phi_{\mathrm{D}} N_{\mathrm{D}} + \Phi_{\mathrm{A}} N_{\mathrm{A}} + \Phi_{\mathrm{A}} N_{\mathrm{A,dir}}} \tag{2.38}$$

$$= \frac{\left(\gamma\, \dfrac{\Phi_{\mathrm{D}}}{\Phi_{\mathrm{A}}}\right) N_{\mathrm{D}} + N_{\mathrm{A}}}{\underbrace{\left(\gamma\, \dfrac{\Phi_{\mathrm{D}}}{\Phi_{\mathrm{A}}}\right)}_{\xi} N_{\mathrm{D}} + N_{\mathrm{A}} + N_{\mathrm{A,dir}}} \tag{2.39}$$

$$= \frac{\xi\, N_{\mathrm{D}} + N_{\mathrm{A}}}{\xi\, N_{\mathrm{D}} + N_{\mathrm{A}} + N_{\mathrm{A,dir}}}, \tag{2.40}$$

with the correction factor

$$\xi = \gamma \cdot \frac{\Phi_{\mathrm{D}}}{\Phi_{\mathrm{A}}}. \tag{2.41}$$

Upon donor excitation, the two-fluorophore system is excited depending on the power of the donor excitation laser, I_{D}, and the extinction coefficient of the donor fluorophore, ε_{D}. Because the two dyes form a FRET pair, the energy from the excitation event is distributed between the two dyes depending on the FRET efficiency, E. The energy will be transferred to the acceptor with a probability of E, while it is retained at the donor with a probability of $(1 - E)$:

$$N_{\mathrm{D}} = (1 - E)\ \varepsilon_{\mathrm{D}} I_{\mathrm{D}}, \tag{2.42}$$

$$N_{\mathrm{A}} = E\ \varepsilon_{\mathrm{D}} I_{\mathrm{D}}. \tag{2.43}$$

The number of direct acceptor excitations is calculated as

$$N_{\mathrm{A,dir}} = \varepsilon_{\mathrm{A}} I_{\mathrm{A}}, \tag{2.44}$$

where I_{A} denotes the laser power of the laser used for direct acceptor excitation, and ε_{A} is extinction coefficient of the acceptor dye at this wavelength. Inserting eqns. 2.42–2.44 into 2.40 yields

$$S = \frac{\xi\ (1-E)\,\varepsilon_{\mathrm{D}} I_{\mathrm{D}}\ +\ E\varepsilon_{\mathrm{D}} I_{\mathrm{D}}}{\xi\ (1-E)\,\varepsilon_{\mathrm{D}} I_{\mathrm{D}}\ +\ E\varepsilon_{\mathrm{D}} I_{\mathrm{D}}\ +\ \varepsilon_{\mathrm{A}} I_{\mathrm{A}}} \tag{2.45}$$

$$= \frac{[\xi + (1-\xi)E]\ \varepsilon_{\mathrm{D}} I_{\mathrm{D}}}{[\xi + (1-\xi)E]\ \varepsilon_{\mathrm{D}} I_{\mathrm{D}} + \varepsilon_{\mathrm{A}} I_{\mathrm{A}}}. \tag{2.46}$$

Eqn. 2.46 shows that S does generally depend on E. Only in the special case where $\xi = 1$, E cancels out and eqn. 2.46 simplifies to

$$S = \frac{\varepsilon_{\mathrm{D}} I_{\mathrm{D}}}{\varepsilon_{\mathrm{D}} I_{\mathrm{D}} + \varepsilon_{\mathrm{A}} I_{\mathrm{A}}}. \tag{2.47}$$

In the case of $\xi = 1$, the stoichiometry is thus independent of FRET efficiency. All distributions in an E-S map, with the exception of the donor-only peak, should therefore lie at the same 'height', that is, the same S value. For $\xi = 1$ to be true, however, relation 2.36 has to be satisfied. Note that the individual values of the quantum yields, ϕ, and instrumental detection efficiencies, η, do not have to be known to satisfy relation 2.36. It is sufficient that the ratio of the total detection efficiencies Φ_{A} and Φ_{D} equals the value of γ.

Thus, the correct γ factor can be found by plotting the E-S map for varying values of γ until the mean stoichiometry is the same at all FRET efficiencies or, in other words, until a linear regression through all events of the E-S map (with exception of the donor-only peak) yields a slope of zero. Note that this method is only feasible if the distribution in the E-S map is sufficiently spread out along the E-axis. In this work, it is therefore well applicable at medium GdmCl concentrations, where the middle of the E-S map is populated by possibly multiple distributions, spanning large intervals of the E-axis. At low GdmCl concentrations, on the other hand, where only the narrow distribution of the folded state is present at the upper edge of the E-axis, γ cannot be determined with high precision.

Unfortunately, it is not possible to simply determine γ under conditions that are well suited for the task and apply it to all concentrations because the quantum yields of the fluorophores and the instrumental detection efficiencies are most likely dependent on the GdmCl concentration. Instead of a single value of γ that is valid for all denaturant concentrations, the correction factor becomes a function of denaturant concentration, $\gamma(c)$ and the problem of finding the right value of γ turns into the problem of finding the right function $\gamma(c)$. To determine $\gamma(c)$, it is necessary to measure the denaturant dependence of the total detection efficiencies $\Phi(c)$ of both dyes, a procedure explained in section 4.1. $\gamma(c)$ can then be expressed as $\gamma(c) = \beta(c)\,\gamma(0\,\mathrm{M\ GdmCl})$, where $\beta(c)$ is obtained by measuring the concentration dependence of the $\Phi(c)$ values of both dyes, and the γ-factor at denaturant free conditions remains the only free parameter. The γ-factors obtained under favorable conditions could thus be used to calculate the γ-factors for E-S maps where the calculation is troublesome.

Normally, quantum yields are measured on sample solutions in glass cuvettes that fit into absorption and fluorescence spectrometers. An absorption spectrum is taken, although the only point of interest is the absorbance at the wavelength at which the sample is excited, when the fluorescence spectrum is recorded in a second step. The fluorescence intensity I that is used is actually the integrated fluorescence over the whole emission spectrum. The instrumental detection efficiency η of a confocal microscope most likely also depends on GdmCl concentration. Especially the increase of the refractive index that accompanies higher GdmCl concentrations influences the optical system that projects the detection volume onto the confocal pinhole. Instead of trying to determine the concentration dependence of $\eta(c)$ directly, it is more useful to combine the measurements of $\phi(c)$ and $\eta(c)$ to directly obtain the total detection efficiency $\Phi(c) = \phi(c)\eta(c)$. As Φ depends on instrumental parameters, the fluorescence has to be collected in exactly the same way as in the burst analysis measurement. The sample is placed on the confocal microscope and fluorescence is collected through a coverslip. At the same time, the sample holder has to be suitable for absorption measurements in a spectrometer. That means the beam path through the sample needs to be well defined and reproducible, to make consecutive measurements comparable. All of these requirements were met by a home made setup based on a modified chamber slide, presented in section 3.8.

As with quantum yield determination, this method does not yield absolute values of Φ, but rather values at all GdmCl concentrations relative to each other. These values can, without loss of generality, be given in relation to the denaturant-free value as:

$$\Phi(c) = \alpha(c)\;\Phi(0\,\mathrm{M}). \tag{2.48}$$

If the factors $\alpha(c)$ for both donor and acceptor dye are known at a particular GdmCl concentration, γ can also be expressed in relation to its value at zero molar GdmCl according to:

$$\gamma(c) = \underbrace{\frac{\alpha_{\mathrm{A}}(c)}{\alpha_{\mathrm{D}}(c)}}_{\beta(c)}\;\gamma(0\,\mathrm{M}) = \beta(c)\;\gamma(0\,\mathrm{M}). \tag{2.49}$$

$\beta(c)$ describes the behavior of the γ-factor due to changes in total detection efficiencies over the whole concentration range of GdmCl. By choosing a value for $\gamma(0\,\mathrm{M})$, one curve $\gamma(c)$ is picked out of the set of possible denaturant dependencies.

However, instead of determining γ at one concentration and calculating the γ-factors at all the other concentrations, one can gain some information by looking at the GdmCl series as a whole by varying $\gamma(0\,\mathrm{M GdmCl})$ until the mean stoichiometry is the same in the E-S maps at all concentrations, instead of only one chosen E-S map.

However, if the extinction coefficients ε depend on denaturant concentration c, it is evident from eqn. 2.47, that the expected stoichiometry becomes denaturant dependent as well:

$$S(c) = \frac{\varepsilon_{\mathrm{D}}(c)\, I_{\mathrm{D}}}{\varepsilon_{\mathrm{D}}(c)\, I_{\mathrm{D}} + \varepsilon_{\mathrm{A}}(c)\, I_{\mathrm{A}}}. \tag{2.50}$$

In that case it is not enough to find a $\gamma(c)$ that causes the *same* average stoichiometry in every E-S map at all GdmCl concentrations. Instead, $\gamma(c)$ has to be found that reproduces the expected stoichiometry according to eqn. 2.50 in every E-S map. The task of measuring the dependence of the extinction coefficients of the used dyes on GdmCl concentration is explained in section 4.1.

In summary, if multiple E-S maps are recorded, according to eqn. 2.50, the expected S-value will depend upon the denaturant concentration at which the map was obtained. The correct function $\gamma(c)$ has to be chosen in such a way that the measured mean stoichiometries $\langle S \rangle$ of all distributions match $S(c)$ in each E-S map. At the same time, in each E-S map, the mean stoichiometry has to be independent of the FRET efficiency E.

2.5 Boltzmann Statistics

For the global analysis of series of smFRET experiments, it is important to understand the connection between the energy level of a molecular state and its relative population, a connection provided by the the laws of Boltzmann statistics. Developed in the 1870s by Ludwig Boltzmann and Josiah Willard Gibbs, Boltzmann statistics, in its most general form, describes the probability, p_i, of a system to be in a particular state, i, given that the system is in thermodynamic equilibrium and coupled to an external heat bath so that its temperature remains constant[193]. The probability p_i depends on the energy level, E_i, of state i according to:

$$p_i = \frac{1}{Z(T)} \exp\left[-\frac{E_i}{k_{\mathrm{B}}T}\right], \tag{2.51}$$

where T is the temperature and k_{B} denotes the Boltzmann constant. The normalization parameter $Z(T)$ is called the canonical partition function. It represents a sum over all states of the system:

$$Z(T) = \sum_{\text{all states}} \exp\left[-\frac{E_j}{k_{\mathrm{B}}T}\right]. \tag{2.52}$$

In a system of N_{tot} non-interacting particles, the expected number of particles in a particular state i is given by

$$N_i = \frac{N_{\mathrm{tot}}}{Z(T)} \exp\left[-\frac{E_i}{k_{\mathrm{B}}T}\right]. \tag{2.53}$$

Relative populations of different states, A and B, are thus related to each other by the energy difference between the states, $\Delta E = E_\mathrm{B} - E_\mathrm{A}$:

$$\frac{N_\mathrm{A}}{N_\mathrm{B}} = \exp\left[-\frac{\Delta E_\mathrm{AB}}{k_\mathrm{B}T}\right]. \tag{2.54}$$

The proper thermodynamic potential for the description of chemical reactions and internal molecular processes is the Gibbs free energy, G, which describes the capacity of a system to do non-mechanical work at constant pressure and constant temperature. Thus, when applying Boltzmann statistics on a set of non-degenerate intramolecular states such as folding conformations of a protein, the relative populations of these states are determined by the differences in Gibbs free energy, ΔG, between them. For the treatment of molecular states under constant pressure and temperature, eqn. 2.54 is consequently expressed using differences in Gibbs free energy:

$$\frac{N_\mathrm{A}}{N_\mathrm{B}} = \exp\left[-\frac{\Delta G_\mathrm{AB}}{k_\mathrm{B}T}\right]. \tag{2.55}$$

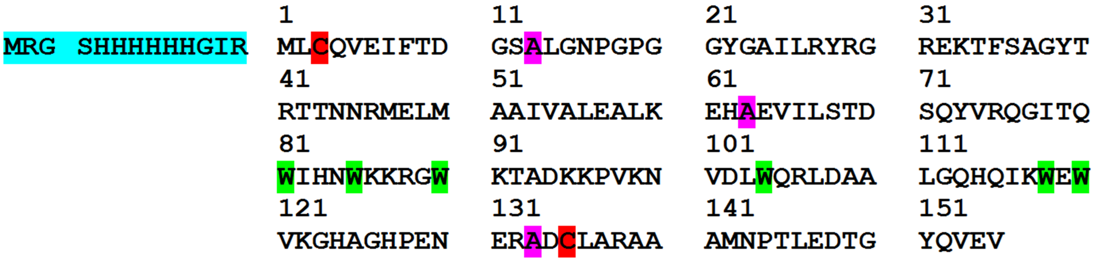
2.6 RNase H

Experimental characterization of RNase H and its folding behavior started out with ensemble measurements. During first examinations, stopped-flow CD spectroscopy allowed direct observation of an intermediate state during folding, which was found to strongly resemble the acid state. The formation of the intermediate was observed in the so-called burst phase, a change in signal during the dead time of the experiment that was markedly stronger than expected by changing the solution environment. This limits the time scale of the formation of the intermediate to below a few microseconds. The intermediate subsequently develops into the folded form on a longer time scale[194]. These experiments were later confirmed by protein engineering studies[195]. Single molecule optical tweezers experiments even showed that the I state is obligatory and resides on the folding path from the unfolded conformation (U) to F[98,196]. Mutating selected residues of RNase H can have drastic effects on the stability of the intermediate state. It can be stabilized to be present in significant amounts even under near-native conditions or destabilized to an extent that the three-state folding behavior vanishes and the protein becomes a two-state folder[197,198]. Taken as a whole, this body of research implies a hierarchical folding model for RNase H, in which the I state plays a crucial role as the central folding core[195,199].

It might be interesting to note that a number of proteins apparently fold through rapidly formed folding intermediates, such as SNase[200,201], myoglobin[202,203] or cytochrome c[20,204]. These intermediates are compact and exhibit some secondary structure, but lack a well defined tertiary structure. States that satisfy these conditions are called *molten globules* and believed to play an important role in the mechanism of protein folding[205,206].

Kuzmenkina *et al.* first studied the chemical denaturation of RNase H with single molecule resolution on surface-immobilized molecules[112,113]. The model used to explain the data did not contain an intermediate state, but rather employed a distribution of unfolded states. The first single molecule force spectroscopy experiments with RNase H were the optical tweezers experiments by Cecconi *et al.*[196]: Induced by the external force of an AFM, RNase H unfolds in a two-state manner and refolds through an intermediate state that resembles the molten globule-like intermediate found in bulk studies. As complete refolding originated from the I state in the majority of cases, the intermediate is considered to be obligatory for refolding. However, it is not clear if mechanically unravelling a protein represents the same event, and follows the same pathways as chemically denaturing it[207,208]. Unfortunately, due to the high experimental complexity of the AFM studies, the number of observed folding events was relatively small (23 times).

For fluorescence labeling of RNase H, the three naturally occurring cysteines were replaced by alanines (C13A, C63A, C133A) and two cysteine residues were subsequently reintroduced at the desired labelling positions near the C- and N-termini (K3C and E135C). The plasmid pJAL135C, containing the gene for a single-cysteine mutant of RNase H with the C13A, C63A, C133A and E135C mutations and ampicillin resistance was a gift by Dr. S. Kanaya (Osaka University, Japan). The K3C mutation was introduced by Dr. E. Kuzmenkina via site-directed mutagenesis. A His-tag was added to the protein at the N-terminus by Stefanie Bacher and Uwe Theilen (fig. 2.24). The resulting mutant is referred to as RNase H 3-135.



```
             1          11         21         31
MRG SHHHHHHGIR  MLCQVEIFTD GSALGNPGPG GYGAILRYRG REKTFSAGYT
             41         51         61         71
             RTTNNRMELM AAIVALEALK EHAEVILSTD SQYVRQGITQ
             81         91         101        111
             WIHNWKKRGW KTADKKPVKN VDLWQRLDAA LGQHQIKWEW
             121        131        141        151
             VKGHAGHPEN ERADCLARAA AMNPTLEDTG YQVEV
```

Figure 2.24: Sequence of the *Escherichia coli* RNase H mutant protein. A His-tag was introduced at the N-terminus for affinity purification (marked in blue). Three naturally occurring cysteines were replaced by alanine (marked in purple) and two cysteines were re-introduced for fluorescence labeling (marked in red). The protein contains six Trp residues (marked in green).

RNase H contains six Trp residues at positions 81, 85, 90, 104, 118 and 120, depicted in figs. 2.24 and 2.25. With the notable exception of Trp81, all Trp residues are hidden in the native conformation. Residues Trp85 on helix C, Trp90 in the loop between helices C and D as well as Trp118 and Trp120 on β-strand 5 are all situated outside of the core. In the intermediate state, the secondary structure around those residues vanishes, increasing their exposure to the solvent. Although they might still transiently interact with the folded core, they experience the solvent to an extent comparable to that of the unfolded state. Trp81 is situated at the edge of the core region and, thus, its solvent exposure is the same in the I and the F state. Trp 104 is buried inside the protein core and thus only exposed to solvent in the completely unfolded U state.

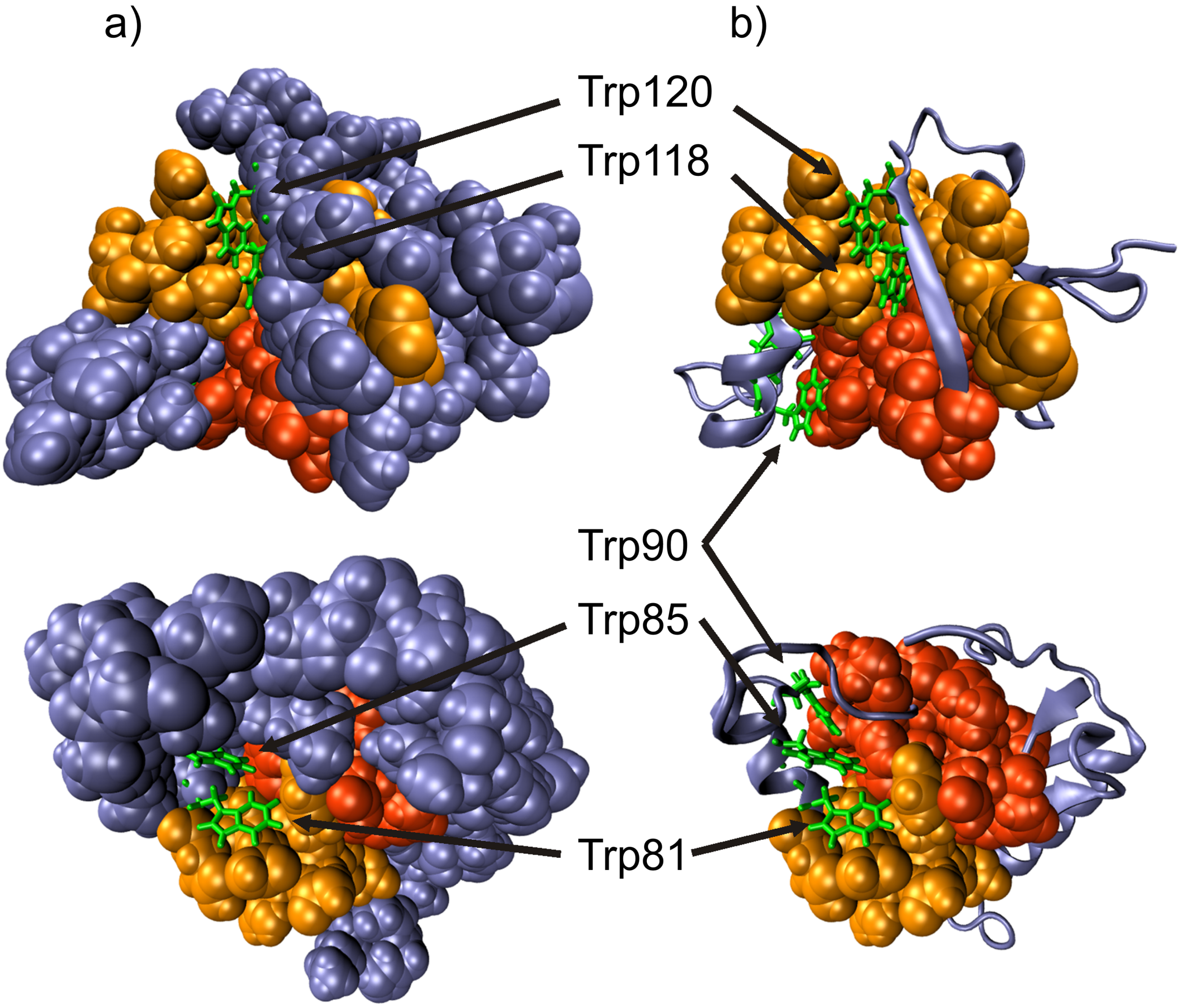


Figure 2.25: Structural models of RNase H in its native conformation (protein data bank accession code 2RN2). The stable core that forms the intermediate state is colored in orange in the same way as in fig. 1.2, whereas those parts of the protein that unfold in the F → I transition are depicted in blue. Trp residues are shown as green stick models. **a)** The van-der-Waals sphere representation highlights how the six Trp residues are shielded from solvent interactions in the folded state. **b)** Here, the regions that are unfolded in the I-state are shown in cartoon representation to enable a better view of the Trp residues.

Chapter 3

Materials and Methods

3.1 Protein Labeling

Introduction of cysteine residues at the chosen labeling sites, residues 3 and 135, enabled site specific attachment of the maleimide derivatives of organic fluorophores. The double bond in the maleimide ring structure reacts with the cysteine thiol group, thereby forming a stable carbon-sulfur bond. RNase H was initially labelled with the acceptor dye Alexa Fluor 647 (Invitrogen, Carlsbad, CA) and singly labeled proteins were separated from unlabeled or doubly labeled fractions via Fast Protein Liquid Chromatography (FPLC). Subsequently, RNase H was labeled with the donor dye Alexa Fluor 546 (Invitrogen).

150 µM RNase H dissolved in 50 mM potassium-phosphate buffer (pH 7.0) was mixed with a tenfold molar excess of tris(2-carboxyethyl)phosphine (TCEP) to reduce the cysteine side chains (20 min at room temperature). A 1.5-fold molar excess of Alexa Fluor 647 was added while carefully stirring the protein solution. The solution was incubated in the dark for 2 h at room temperature. Singly labeled RNase H proteins were separated from doubly and unlabeled RNase H proteins via FPLC (Äkta 900, GE Healthcare Europe, Freiburg, Germany).

For this separation, RNase H in 50 mM potassium phosphate buffer (pH 7.0) was loaded onto a Resource S cation exchange column (GE Healthcare Europe) with a column volume

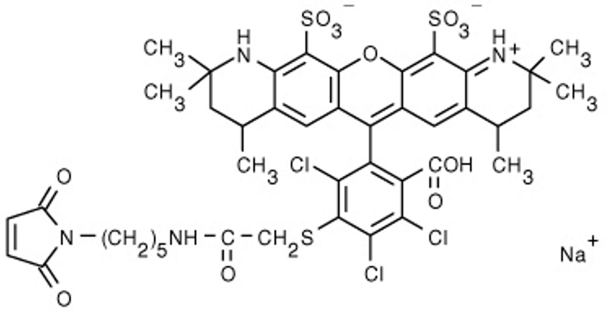


Figure 3.1: Structure of the maleimide derivative of Alexa Fluor 546. The maleimide group is coupled to the fluorophore by a 5-carbon linker. Picture by Invitrogen (Carlsbad, CA).

(CV) of 1 ml and washed for 10 min at a flow rate of 1 ml/min with 50 mM potassium phosphate buffer (pH 7.0). Elution was then started at a flow rate of 1 ml/min (flow speed 180 cm/h on this particular column), using a 25 – 40% gradient of 1 M NaCl (in 50 mM potassium phosphate buffer, pH 7.0) with a slope of 1.5% NaCl/CV. The absorption of the eluate was monitored at 280 nm (absorption of the protein) and 650 nm (absorption of the Alexa Fluor 647 dye) to identify the relevant fractions. These fractions were pooled and reconcentrated by membrane ultrafiltration via Vivaspin centrifugation columns at 7,000 rpm and 4°C (Molecular weight cut-off of 10 kDa, Sartorius Stedim Biotech GmbH, Göttingen, Germany).

The same labeling and purification process was subsequently repeated with Alexa Fluor 546 with a different elution gradient, which ranged from 20% to 35% of 1 M NaCl. In addition, the absorption of the eluate was also measured at 554 nm to detect the Alexa 546 dye.

A fraction of proteins intended for Trp fluorescence experiments was set aside and remained unlabeled.

3.2 Buffer Solutions

All measurements were performed in 50 mM potassium phosphate buffer. The first protonation equilibrium of phosphate at $\mathrm{p}K_{\mathrm{a1}} = 2.15$ was used to buffer the solution at pH 3.0. For pH 7.2 the second protonation equilibrium at $\mathrm{p}K_{\mathrm{a2}} = 7.20$ was used.

- **pH 7.2 standard buffer**
  15.4 mM KH_2PO_4, 34.5 mM K_2HPO_4, 50 mM KCl, 10 mM $MgCl_2$, 1 mM ascorbic acid, 1 mM methyl viologen, 8.1 µM Tween 20

- **pH 7.2 GdmCl stock solution**
  15.4 mM KH_2PO_4, 34.5 mM K_2HPO_4, 50 mM KCl, 10 mM $MgCl_2$, 1 mM ascorbic acid, 1 mM methyl viologen, 8.1 µM Tween 20, 6 M GdmCl

- **pH 3.0 standard buffer**
  5.1 mM H_3PO_4, 44.8 mM KH_2PO_4, 50 mM KCl, 10 mM $MgCl_2$, 1 mM ascorbic acid, 1 mM methyl viologen, 8.1 µM Tween 20

- **pH 3.0 GdmCl stock solution**
  5.1 mM H_3PO_4, 44.8 mM KH_2PO_4, 50 mM KCl, 10 mM $MgCl_2$, 1 mM ascorbic acid, 1 mM methyl viologen, 8.1 µM Tween 20, 6 M GdmCl

Water (Mol Bio grade) was purchased from 5 Prime (Hamburg, Germany), GdmCl from Merck (Darmstadt, Germany). After preparation, all solutions were adjusted precisely to the desired pH by addition of concentrated HCl or NaOH as necessary. The pH was measured with a glass electrode (Titan Benchtop Meter, Sentron Integrated Sensor Technology, Roden, The Netherlands). It should be noted that, for varying counterion concentrations, there exists a subtle problem with pH determination[209]. For GdmCl concentrations below 2 M the pH value measured with glass electrodes deviates less than 0.1 pH units from its true value. For higher GdmCl concentrations, though, the difference between perceived and real pH grows with increasing denaturant concentration and can become as large as about 0.7 pH units at 6 M GdmCl. As the main point of interest of the present work, the exchange between folded, intermediate and unfolded states of RNase H, mainly occurs at GdmCl concentrations below 2, this error in pH determination was considered small and therefore neglected. All buffers were prepared in a laminar flow box. After preparation, the buffers were screened for fluorescent contaminations on a confocal microscope by performing a burst analysis under the same conditions as a smFRET experiment (see section 2.4). The buffer was considered clean if the number of detected fluorescence bursts did not exceed about one per ten seconds over a measurement time of a few minutes.

To prepare a buffer with a specific GdmCl concentration, standard buffer and 6 M GdmCl stock solution were mixed in the appropriate ratio. The exact GdmCl concentration c was determined by measuring the refractive index[210].

Ascorbic acid (Sigma Aldrich, St. Lois, MO) and methyl viologen (Sigma Aldrich) together form a reducing and oxidizing system (ROXS), which, in conjunction with oxygen depletion of the sample, increases brightness and photostability of the fluorescent dyes[142]. 100 mM stock solutions of ascorbic acid and methyl viologen were prepared freshly every week and stored at -20°C. Together with Tween 20 (Thermo Fisher Scientific Inc., Waltham, MA), they were added to the buffers only directly before the start of a measurement.

3.3 Cover Slip Coating

The glass surface onto which a drop of sample solution was placed for the smFRET measurements was passivated against protein adsorption by coating it with an extensively cross-linked star-shaped polyethylene glycol (PEG) layer[211,212]. All the following steps were performed in a laminar flow box. Glass cover slips ($32 \times 24\,\mathrm{mm}^2$, Menzel, Braunschweig, Germany) were amino-silanized with the commercial aminosilane Vectabond (Biozol Diagnostica, Eching, Germany) according to the manufacturer's instructions: After cleaning the glass slides in an air plasma (PDC002 Plasma Cleaner, Harrick Plasma, Ithaca, NY) for 10 min, they were immersed in acetone for 5 min and then placed in Vectabond solution for another 5 min. The reaction was stopped and slides were cleaned by repeated immersion in water (Ultrapure Milli-Q water, $> 18\,\mathrm{M\Omega\,cm}$ at 25°C). The hydrophobic slides retained only little water when pulled out of it. The remaining water was dried in the airflow of the flow box.

Star shaped PEG molecules, functionalized with isocyanate groups, were kindly provided by the group of Prof. Martin Möller (RWTH Aachen, Germany)[211,212]. 20 mg of these

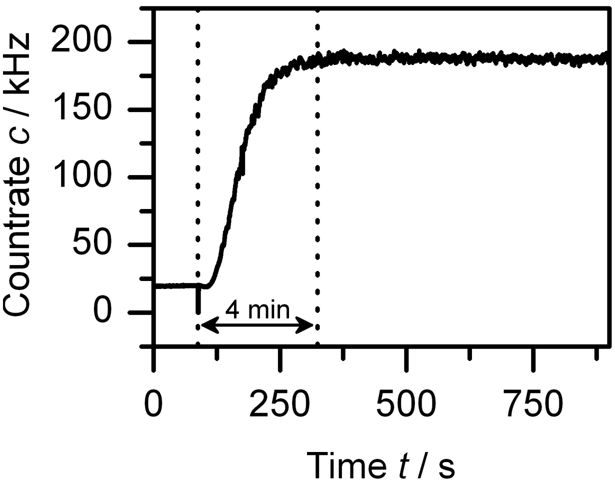


Figure 3.2: Fluorescence trace of the oxygen sensitive dye $Ru(dpp)_3(PF_6)_2$ in ethanol. A droplet of ≈ 70 µl was placed on a coverslip, imitating the setup of the smFRET measurements. The box which covers the sample-holder was flooded with nitrogen (intentionally marked with the dip in fluorescence). Four minutes later, the fluorescence reached a stable, higher level, indicating that the oxygen concentration inside the box reached a minimum.

dried star-PEG molecules were dissolved in water free tetrahydrofuran (Sigma Aldrich). The solution was then diluted 10-fold with water and kept at room temperature for 5 min. This solution was subsequently passed through a 10 µm Anatop syringe filter (Whatman, Maidstone, UK), followed by a 0.02 µm Anatop syringe filter and deposited drop-wise on the Vectabond treated cover slips. A thin layer was formed by spin coating (2, 500 rpm for 40 s). The isocyanate groups react with the amine groups on the amino-silanized cover slip and with each other, thus attaching the PEG layer to the underlying surface and cross linking it. Individual slides were then sealed in petri dishes with parafilm. After incubating the cover slips at room temperature over night, they were stored at -20°C until used.

3.4 Sample Preparation for smFRET measurements

A drop of ≈ 70 µl sample solution containing the fluorescently labeled RNase H was placed on a star-PEG coated coverslip and mounted on the confocal microscope. Due to the hydrophobic nature of the surface coating, the aqueous sample formed stable droplets when placed on top of the cover slips. To remove oxygen from the buffer solution, samples were kept in a nitrogen atmosphere during measurement. For this purpose, the sample stage was covered with a box which was subsequently purged with nitrogen. To minimize evaporation under this constant flow of gas, the nitrogen gas was saturated with water by passing it through a gas washing bottle filled with distilled water. Time needed for oxygen depletion was characterized using the oxygen sensitive ruthenium complex *Tris(4,7-diphenyl-1,10-phenanthroline) ruthenium(II) bis (hexafluorophosphate)* ($Ru(dpp)_3(PF_6)_2$, Sigma-Aldrich, St. Louis, MO). It is dynamically quenched by molecular oxygen, making it a useful probe for the amount of dissolved oxygen in solution[213]. As shown in fig. 3.2, 4 min after the onset of the nitrogen flow, the oxygen concentration in the sample solution has reached a minimum.

3.5 Microscope Systems

The smFRET data in this work were measured with two different experimental setups. While data at pH 7.2 were obtained on a custom-built confocal microscope called *M2*, later experiments at pH 3.0 were conducted with a commercial *MicroTime 200* from PicoQuant (PicoQuant GmbH, Berlin, Germany). Both setups are described below.

3.5.1 The M2 Setup

The core of the home-built M2 setup was a Zeiss Axiovert 135 TV frame. A 532-nm diode-pumped solid state laser (Excelsior 532-100-SLM-CDRH, Spectra Physics, Santa Clara, CA) at a power of 43 kW cm^{-2} was used to excite the donor fluorophore. Another solid state laser (DL 638-100, CrystaLaser, Reno, NV) with an excitation power of 110 kW cm^{-2} at 641 nm served for direct excitation of the acceptor dye. To obtain the alternating excitation necessary for ALEX, both beams were passed through and modulated by an acousto-optical tunable filter (AOTF) (A-A Optoelectronic, Orsay, France) at a frequency of 10 kHz. 70% and 30% of each cycle were allotted to donor and acceptor excitation, respectively. The time necessary for switching between illumination regimes was experimentally determined as 7 µs. Data recorded in the first 7 µs after switching were thus discarded to avoid temporal crosstalk. The modulated excitation beam was focused into the sample by the objective (C-Apochromat 63x/1.2w, Zeiss), which was also used to collect the fluorescence emission. The microscope tube lens refocused the fluorescence onto the confocal pinhole with a diameter of 100 µm. Light passing through the pinhole was split into two channels by a 640 nm dichroic mirror (HQ640DCXR, AHF, Tübingen, Germany). To avoid spectral crosstalk, the light in each channel was passed through an appropriate bandpass fluorescence filter (550 – 610 nm for the donor and 650 – 750 nm for the acceptor channel). Both signals were detected by avalanche photo diodes (APDs)(SPCM-CD 3017, PerkinElmer, Boston, MA) and counted with a data acquisition card (NI PCI-6229, National Instruments, München, Germany).

3.5.2 MicroTime 200

The MicroTime 200 is based on an IX71 microscope body (Olympus, Hamburg, Germany). Pulsed Interleaved Emission (PIE) was used instead of ALEX to alternate between donor and acceptor excitation, using two pulsed diode lasers synchronized by a PDL 828-L Sepia II laser driver (PicoQuant). A LDH-P-FA-530 laser with an excitation power of 60 kW cm^{-2} at 530 nm was used for donor excitation and a LDH-C-640B laser with 21 kW cm^{-2} at 640 nm for excitation of the acceptor (both PicoQuant). Both lasers emitted pulses at a frequency of 10 MHz with a phase shift of π between them, generating excitation pulses of alternating color every 50 ns. Thus, the delay between pulses was much longer than the fluorescence lifetimes of the dyes, thus avoiding temporal crosstalk. Excitation light was focused and fluorescence light collected with an UPLSAPO objective (60XW, 60x/1.2 W, Olympus). After being focused onto a 100 µm diameter confocal pinhole, the emission light was split into two channels by a dichroic mirror (620DRXXR, Chroma, AHF, Tübingen, Germany) and passed through appropriate filters (HQ580/70 for the donor and HQ690/70 M for the acceptor channels, both Chroma, AHF). Detection was accomplished by avalanche photodiodes (SPCM-CD 3017, PerkinElmer, Boston, MA).

3.6 Absorption Measurements

When light passes through a sample, its initial intensity, I_0 decays exponentially with the pathlength. The intensity, I, of light of wavelength λ after passing through a sample of thickness l is thus

$$I = I_0\, 10^{-\varepsilon(\lambda) l c}, \tag{3.1}$$

where c is the concentration of the absorbing substance and $\varepsilon(\lambda)$ its wavelength-dependent extinction coefficient. The absorbance A of such a sample is given by the *Beer-Lambert Law*:

$$A(\lambda) = -\log_{10}\left(\frac{I}{I_0}\right) = \varepsilon(\lambda) l c. \tag{3.2}$$

The measurement of absolute extinction coefficients is challenging, as precise control of the concentration is required, which is difficult to achieve. However, relative changes in ε can be measured more easily. Absorption spectra were recorded on a Cary 100 spectrophotometer (Varian, Darmstadt, Germany) in cuvettes with a path length of 1 cm.

3.7 Fluorescence Measurements

Fluorescence spectra were taken with a Fluorolog-3 fluorimeter (Horiba Jobin Yvon Inc., Edison, NY). For the measurements, the sample was held in quartz cuvettes.

3.8 Quantum Yield and Total Detection Efficiency Measurements

A well established technique to determine quantum yields is the comparative method of Williams *et al.*[214]. The absorbance, A, of a dye solution is directly proportional to the number of times fluorophores get excited, N_x. The observed intensity of fluorescence emission, I, from a sample is directly proportional to the number of emitted photons, N_e:

$$\phi = \frac{N_\mathrm{e}}{N_\mathrm{x}} = \frac{c_\mathrm{e} I}{c_\mathrm{x} A}. \tag{3.3}$$

Generally, the dye in question is diluted in a series of steps. At each concentration, a fluorescence spectrum is measured as well as the absorbance of the dye at the excitation wavelength used to obtain the fluorescence spectrum. The integrated fluorescence intensities from these spectra, I, are then plotted against the absorbances at the excitation wavelength, A. This data set should form a line through the origin, as is clear if we solve eqn. 3.3 for I:

$$I = \phi \frac{c_\mathrm{x}}{c_\mathrm{e}} \cdot A. \tag{3.4}$$

The slope $m = \phi\,(c_\mathrm{x}/c_\mathrm{e})$ of such a plot represents the quantum yield multiplied by the unknown factor $c_\mathrm{e}/c_\mathrm{x}$. In most cases, this proportionality factor is difficult to determine. However, assuming that c_e and c_x are constants of the experimental setup and identical for all samples, the unknown quantum yield of a test sample ϕ_x can be determined relative to the known ϕ_r of a reference sample according to

$$\phi_\mathrm{x} = \phi_\mathrm{r}\, \frac{m_\mathrm{x}}{m_\mathrm{r}}. \tag{3.5}$$

Thus, absolute quantum yields can be simply obtained by comparison with reference samples of known quantum yield.

If the two samples are measured in solvents with different refractive indices, n, a correction term has to be added to eqn. 3.5:

$$\phi_\mathrm{x} = \phi_\mathrm{r}\, \frac{m_\mathrm{x}}{m_\mathrm{r}}\, \frac{n_\mathrm{x}^2}{n_\mathrm{r}^2}. \tag{3.6}$$

For the measurement of total detection efficiencies, a Lab-Tek chamber slide (Nalge Nunc International, Rochester, NY), was modified as depicted in fig. 3.3. A home made mounting bracket for the spectrometer ensured repeatable mounting positions in the Cary 100 spectrophotometer for absorption measurements. Fluorescence was measured on the MicroTime 200 under the same conditions as single molecule experiments.

3.9 Fitting the FRET histograms

The FRET data were fitted with self-written MATLAB scripts, on versions ranging from MATLAB 7.5 to MATLAB 7.10. The optimization routine used the `lsqcurvefit` command, an implementation of the Levenberg Marquardt algorithm[215,216].

3.9.1 Parameter Weighting

In a least-squares fitting routine, the residual sum of squares, χ^2, is a measure of how well the model function f approximates the given data set. It is calculated as a sum of all residuals, weighted by the inverse square of the statistical uncertainty:

$$\chi^2 = \sum_{i=1\ldots n} \frac{1}{\sigma_i^2}\,\left(y_i - f(x_i, \boldsymbol{\Theta})\right)^2. \tag{3.7}$$

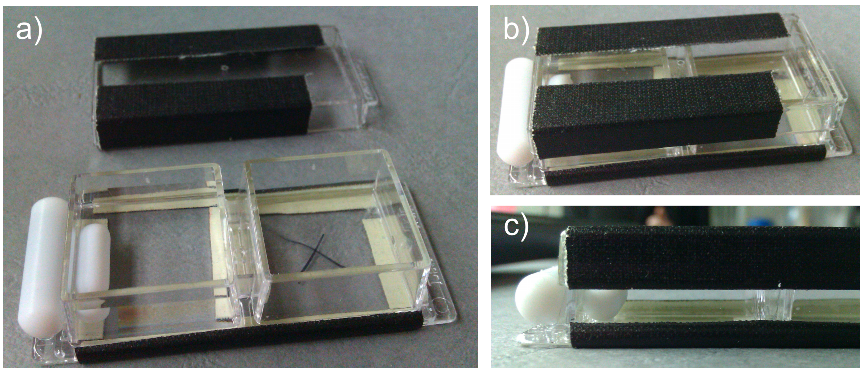


Figure 3.3: Sample chamber for measuring total detection efficiencies. For fluorescence measurements, it is placed on a confocal microscope. The sample is excited and the fluorescence light is collected through the coverslip that forms the bottom of the chamber. For absorption measurements, light is passed through the chamber through the transparent side walls. **a)** Sample chamber with the lid taken off. The magnetic stir bar inside the chamber allows swift mixing after solvent exchanges. During the absorption and fluorescence measurements, it is held at the side of the sample chamber, as shown. **b)** and **c)** Closed sample chamber. The height of the beam path for absorption measurements is restricted by the lower edge and the lid to ensure a defined beam path regardless of the liquid level in the chamber.

The sum extends to n, the total number of data-points (x_i, y_i). $\boldsymbol{\Theta} = (\Theta_1 \dots \Theta_p)$ is a vector of length p, containing all fit parameters.

When fitting smFRET histograms, the weighting factors σ_i^{-2} are directly accessible. The histograms are formed by collecting and counting the occurrence of independent events: Each bin in a histogram counts the number of times a molecule with a FRET efficiency within a certain interval is detected. The value of such a bin represents a signal that is formed as the sum of independent events, and, as such, obeys Poissonian statistics. Using the property of Poissonian distributions that mean equals variance of the distribution ($\mu = \sigma^2$), the weighting factor of each data point y_i is calculated as

$$\sigma_i^{-2} = \mu_i^{-1} = y_i^{-1}. \tag{3.8}$$

In practical applications, to avoid division by zero errors, one normally calculates the weighting factors according to

$$\sigma_i^{-2} = \max(1, y_i)^{-1}. \tag{3.9}$$

3.9.2 Variable Scaling

The set of fit parameters contains a wide variety of variables, describing all kinds of physical quantities. Their numerical values differ by up to five orders of magnitude if straightforwardly expressed in their native units. Whereas the widths of Gaussian distributions in smFRET histograms lie around 0.1, the peak positions of Trp fluorescence, another parameter of the global fitting routine, has a numerical value in the tens of thousands when expressed in wavenumbers. Such a set of fit parameters is called *poorly scaled* and leads to numerical instabilities if left unresolved[217]. A first measure is to implement upper and lower boundaries for all fit parameters to limit the range in which each parameter is allowed to vary. However, the best solution to overcome this problem is to scale all variables (and, of course, their boundaries as well) to match the order of magnitude of their values. The set of fit parameters $\boldsymbol{\Theta} = (\Theta_1 \dots \Theta_p)$ is thus prescaled to $\boldsymbol{\Theta}_{\text{scaled}} = (s_1\Theta_1 \dots s_p\Theta_p)$ and handed over to the fit routine of MATLAB. In the model function itself, the scaling is reversed before calculation by multiplying $\boldsymbol{\Theta}_{\text{scaled}}$ element wise with the inverse scaling factors $(s_1^{-1} \dots s_p^{-1})$. The fit algorithm of MATLAB now operates on the scaled parameter set $\boldsymbol{\Theta}_{\text{scaled}}$. When the fit routine modifies two elements of $\boldsymbol{\Theta}_{\text{scaled}}$ by the same *absolute* amount, this translates into changes of roughly the same *relative* amount of these elements in the descaled $\boldsymbol{\Theta}$.

3.9.3 Error Determination

The standard error of any fit parameter Θ is determined according to the following procedure[218]. Starting point is the sum of all residual χ^2 as defined in eqn. 3.7. It is divided by the number of residuals, n, minus the number of independent fit parameters, p, to yield the mean residual variance:

$$s^2 = \frac{\chi^2}{n-p}. \tag{3.10}$$

The fit routine of MATLAB delivers the set of best fit parameters $\boldsymbol{\Theta}'$ that minimizes $\chi^2(\boldsymbol{x}, \boldsymbol{\Theta}')$ and the value of χ^2 at that minimum. In addition, it also yields the Jacobi matrix of dimension $n \times p$:

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial}{\partial\Theta_1} f(x_1, \boldsymbol{\Theta'}) & \cdots & \frac{\partial}{\partial\Theta_p} f(x_1, \boldsymbol{\Theta'}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial\Theta_1} f(x_n, \boldsymbol{\Theta'}) & \cdots & \frac{\partial}{\partial\Theta_p} f(x_n, \boldsymbol{\Theta'}) \end{pmatrix}, \tag{3.11}$$

which contains the derivatives of the fit function with regard to the fit parameters, evaluated at all data points. With the Jacobi matrix, one can calculate the so-called Hessian matrix of dimension $p \times p$,

$$\boldsymbol{H} = \boldsymbol{J}' \cdot \boldsymbol{J} = \begin{pmatrix} \frac{\partial^2}{\partial \Theta_1^2} f & \frac{\partial^2}{\partial \Theta_1 \, \partial \Theta_2} f & \cdots & \frac{\partial^2}{\partial \Theta_1 \, \partial \Theta_p} f \\ \frac{\partial^2}{\partial \Theta_2 \, \partial \Theta_1} f & \frac{\partial^2}{\partial \Theta_2^2} f & \cdots & \frac{\partial^2}{\partial \Theta_2 \, \partial \Theta_p} f \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \Theta_p \, \partial \Theta_1} f & \frac{\partial^2}{\partial \Theta_p \, \partial \Theta_2} f & \cdots & \frac{\partial^2}{\partial \Theta_p^2} f \end{pmatrix}, \tag{3.12}$$

which, in turn, is needed for the calculation of the variance-covariance matrix:

$$\boldsymbol{C} = (\boldsymbol{J}' \cdot \boldsymbol{J})^{-1} \, s^2 = \boldsymbol{H}^{-1} \, s^2. \tag{3.13}$$

The error of any fit parameter Θ_j is obtained from the diagonal of the variance-covariance matrix, (where the variances are situated)

$$\Delta\Theta_j = \sqrt{\boldsymbol{C}_{jj}}. \tag{3.14}$$

Combined in one equation, the error can be evaluated at the previously determined value $\boldsymbol{\Theta}'$ which minimizes $\chi^2(\boldsymbol{x}, \boldsymbol{\Theta})$:

$$\Delta\Theta_j = \sqrt{\left[\frac{\partial^2}{\partial \Theta_j^2} \, f(\boldsymbol{\Theta}')\right]^{-1} \frac{\chi^2}{n-p}}. \tag{3.15}$$

Thus, the standard error of any fit parameter is dependent on the inverse second derivative of the fit function $f(\boldsymbol{x}, \boldsymbol{\Theta}')$ with regard to that parameter and the normalized sum of residuals.

Chapter 4

Experiments and Discussion

4.1 Determination of the γ factor

Measuring Total Detection Efficiencies

For the calculation of total detection efficiencies, dilution series of Alexa Fluor dyes 546 and 647 in standard buffer were taken at five GdmCl concentrations. The fluorescence count rate was plotted against the absorbance at the excitation wavelength. The results are shown in fig. 4.1. The relative total detection efficiencies, calculated from the slopes of these plots, and the resulting behavior of the γ-factor are presented in fig. 4.2.
The relative γ-factors in fig. 4.2b were interpolated with a cubic B-spline to estimate γ in between the measured values. It is evident that the most significant change of the γ-factor happens at denaturant concentrations between 0 M and 1 M, whereas γ remains relatively constant at GdmCl concentrations above 1 M.

Measuring the GdmCl Dependence of the Extinction Coefficients

One milliliter dye solution with a maximal absorbance of ≈ 1 was prepared in the same buffer used for the single molecule experiment. In a series of measurements, 100 µl of 5.7 M GdmCl solution were added ten times, and an absorbance spectrum was taken each time. The amount of added solvent was controlled by weighing the cuvette during each addition with a precision of 1 mg, yielding a relative error $< 1\%$. With the measured densities of both solutions ($\rho_{0\mathrm{M}} = (1,007.6 \pm 0.9)$ g/l and $\rho_{5.7\mathrm{M}} = (1,147.8 \pm 1.5)$ g/l), the precise volume and, in consequence, relative dye concentration, could be calculated for every dilution step. The resulting spectra were corrected for dilution and are shown in fig. 4.3. A possible problem is the adsorption of dye molecules to the cuvette walls or to pipet tips when adding new solution, which changes the concentration in an uncontrolled way. To avoid contact of the sample solution with the pipet tips, the GdmCl solution was dripped into the cuvette. To properly mix the solution after solvent addition, the cuvette was closed with a cap and shaken. As a negative control experiment, the same dilution

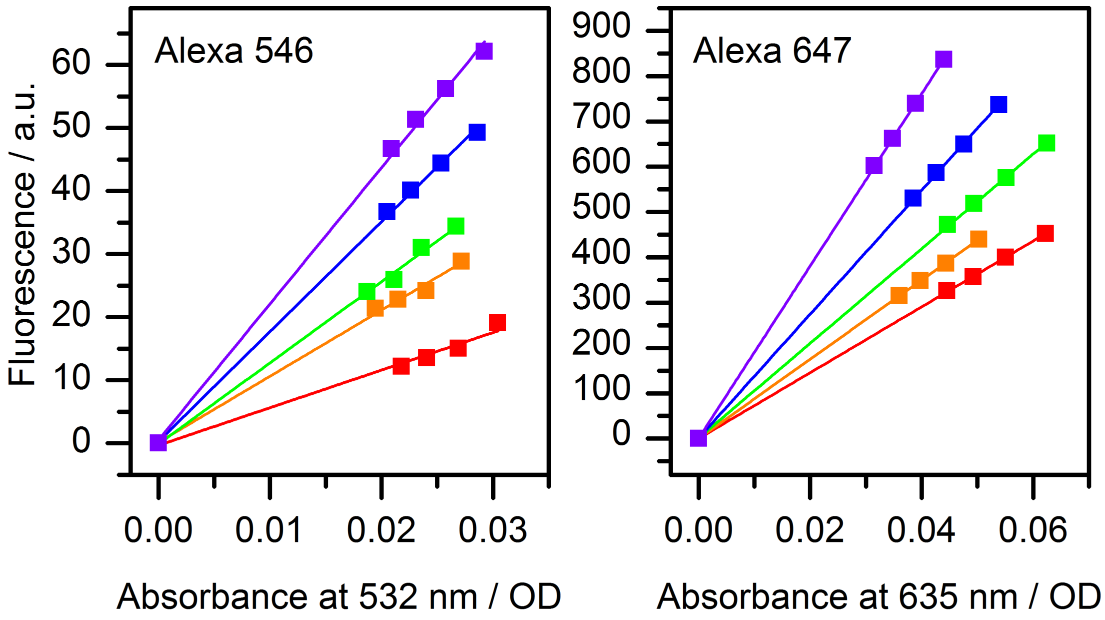


Figure 4.1: Fluorescence intensities of Alexa Fluor 546 and Alexa Fluor 647 dye solutions in standard buffer measured with the MicroTime 200 and plotted against the absorbance at the excitation wavelength. The experiment was performed at GdmCl concentrations of 0 M (red), 0.84 M (orange), 1.54 M (green), 2.99 M (blue) and 5.88 M (violet). The relative total detection efficiencies that were calculated from the slopes of the plots are depicted in fig. 4.2.

series was performed with buffer instead of 5.7 M GdmCl solution. The dilution-corrected absorbance values all were within 2% of each other.

The integrated absorbances of the spectra of both dyes shown in fig. 4.3 as well as the absorbance at the single molecule excitation wavelengths are plotted in fig. 4.4. Although the integrated absorbance of Alexa 647 does not change much, a significant peak shift leads to a decrease of the extinction coefficient $\varepsilon_{\mathrm{Al647}}(635\text{ nm})$ by almost 25% at the excitation wavelength. The absorbance integral of Alexa 546 only exhibits changes of a few percent, while the absorbance at the excitation wavelength $\varepsilon_{\mathrm{Al546}}(532\text{ nm})$ shows a sudden jump by almost 10% upon addition of 0.5 M GdmCl. At higher GdmCl concentrations, it stays relatively stable. Cubic B-splines have been fitted to obtain a closed curve that describes the behavior. Splines are piecewise-polynomial functions that consist of a number of ordered points called knots, connected by polynomial functions. The order, n, of the connecting polynomials, is called the order of the spline. The polynomial functions have to be chosen in such a way that the joint spline is continuously differentiable $n-1$ times at all but the two outermost knots. For interpolation, the most commonly used type of spline is cubic,

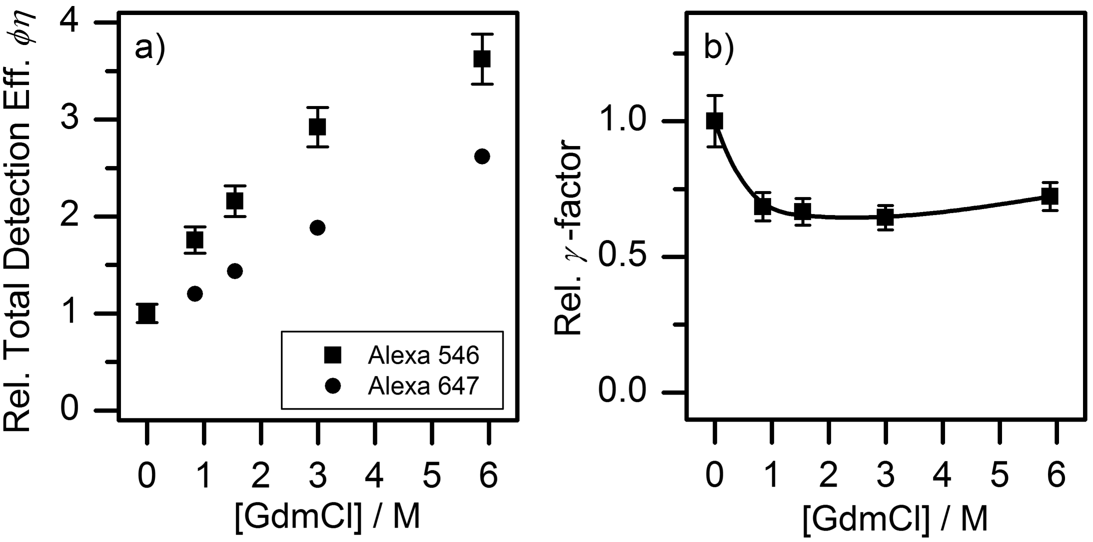


Figure 4.2: **a)** Total detection efficiencies $\Phi = \phi\eta$ for Alexa Fluor 546 (squares) and Alexa Fluor 647 (circles), relative to the respective values at [GdmCl] = 0 M. Error bars for Alexa 647 are smaller than 1% for all points and, thus, smaller than the symbols. **b)** Expected γ factor response to increasing GdmCl concentration, also relative to the value in denaturant free solution, interpolated by a cubic B-spline (solid line).

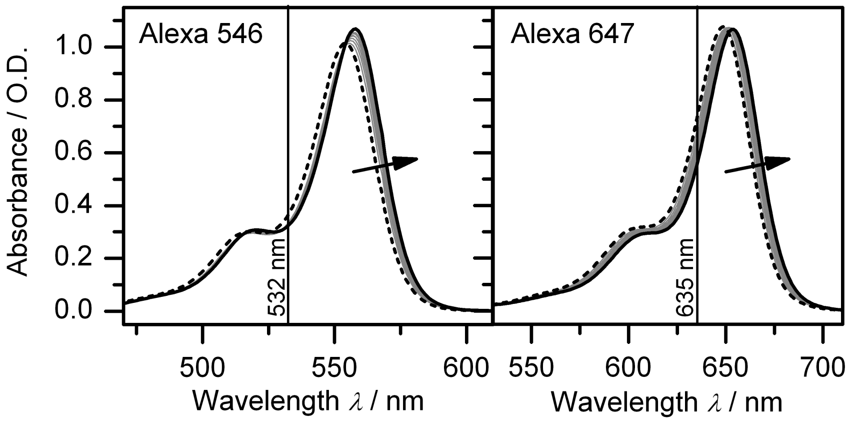


Figure 4.3: Absorption spectra of Alexa Fluor 546 and Alexa Fluor 647 at pH 3 at increasing GdmCl concentrations from 0 M (dashed line) to 3.3 M (bold solid line). The wavelengths of 532 nm and 635 nm, which will be used for donor and acceptor excitation in single molecule experiments, are marked by vertical lines.

i.e. splines of order $n = 3$, which guarantees curves that are smooth and differentiable twice over the whole range of the function. For interpolation, splines are often favored over polynomials as they offer more degrees of freedom and avoid *Runge's phenomenon*[219], the fact that higher order polynomials often introduce undesired oscillations when used for interpolation. The name "B-spline" refers to the way a spline function is represented: One possible base of the vector space of piecewise-polynomial functions is a set called B-spline base functions and splines that are represented as a superposition of B-spline base functions are often simply referred to as B-splines[220].

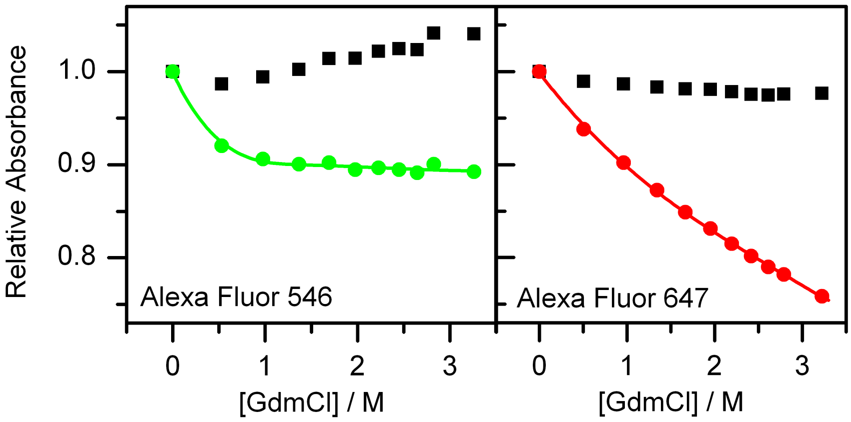


Figure 4.4: Relative absorbances of Alexa Fluor dyes as a function of GdmCl concentration. Black squares depict the integrated absorbance. Absorbances at 532 nm for Alexa Fluor 546 ($\varepsilon_{\text{Al546}}$(532 nm), green circles) and 635 nm for Alexa Fluor 647 ($\varepsilon_{\text{Al647}}$(635 nm), red circles) are shown as well. The lines depict cubic B-splines, fitted to the respective data sets.

With eqn. 2.50, it is straightforward to give the expected mean stoichiometry $\langle S \rangle$ at all GdmCl concentrations. A set of curves calculated according to eqn. 2.50 is shown in fig. 4.5. The size of the shift in $\langle S \rangle$ depends on the relation of the donor and the acceptor excitation power. At low direct acceptor excitation intensity, which results in high overall stoichiometry values, the shift in expected stoichiometry amounts to 4.6%, whereas at high direct acceptor excitation intensity and low overall stoichiometries, the shift is twice as large and amounts to 9.2%.

4.2 smFRET Measurements at pH 7.2

FRET-pair labeled RNase H 3-135 molecules were observed at pH 7.2 on the M2 setup. The FRET values of freely diffusing single RNase H molecules were measured as they passed through the confocal volume of the microscope and collected into a histogram. A series

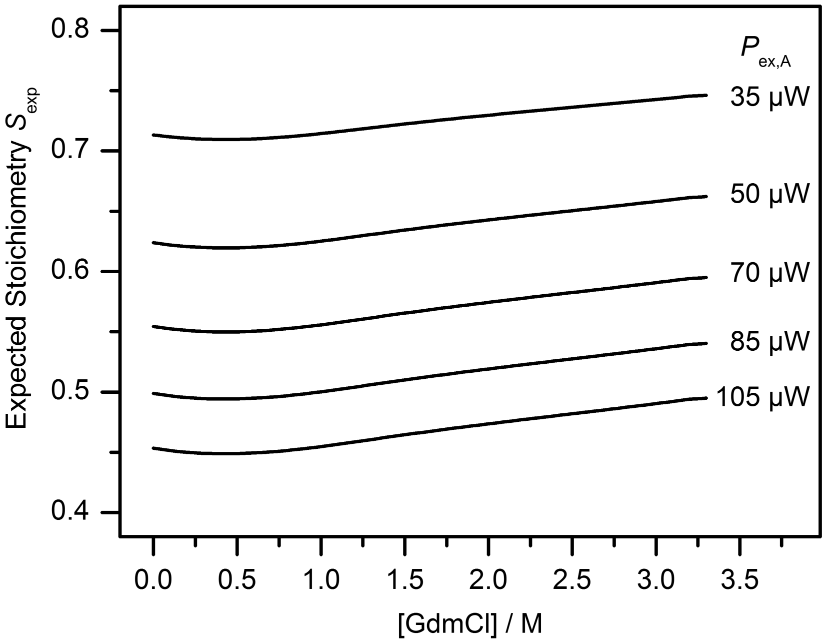
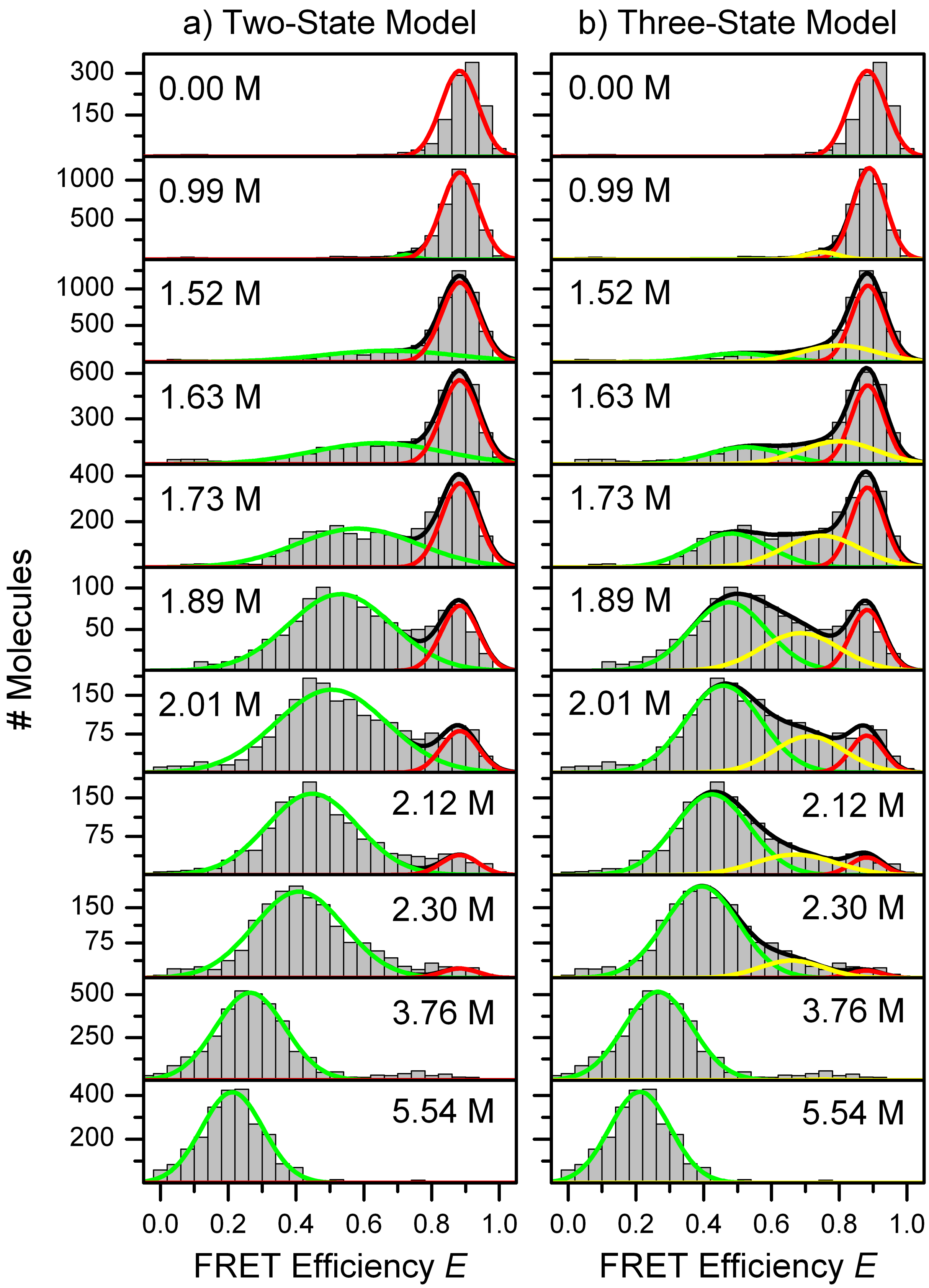


Figure 4.5: Expected stoichiometries for an Alexa 546/647 dye system, using a donor excitation power of $P_{\mathrm{ex,D}} = 200$ µW and various acceptor excitation powers, $P_{\mathrm{ex,A}}$, given in the figure.

of 22 such burst-analysis measurements was made over a range of GdmCl concentrations from 0 to ≈ 6 M. Measurement time was 30 min per concentration. A selection of the resulting histograms is shown in fig. 4.6 (grey bars). At low GdmCl concentrations, only one distribution with high FRET efficiency is observed ($0.8 < E < 1$), which is assigned to the folded state (F). In the range of intermediate denaturant concentrations, i.e., between ≈ 1.2 and 2.3 M, however, a second state appears at FRET efficiencies < 0.8, which was previously interpreted by our group as the unfolded conformation of the protein (U) because the resulting histograms have a similar appearance as previously obtained results on surface-immobilized proteins from our group[112,113,188,211].

4.2.1 A Two-State Approach

While the F-state dominates at low GdmCl concentrations, the F state exchanges population with U in the region of $1.2 < c < 2.3$ M, until, at concentrations above that range, F vanishes completely and only U remains. A straightforward approach to fit the data is to employ a two-state model with normalized Gaussian distributions $G(E; \mu_i, \sigma_i)$, one for each conformation. Each distribution was scaled with a factor A_i to account for its fractional population in the mixture.


a) Two-State Model
b) Three-State Model
# Molecules
0.00 M
0.99 M
1.52 M
1.63 M
1.73 M
1.89 M
2.01 M
2.12 M
2.30 M
3.76 M
5.54 M
300
150
1000
500
1000
500
600
300
400
200
100
50
150
75
150
75
150
75
500
250
400
200
0.0
0.2
0.4
0.6
0.8
1.0
FRET Efficiency E

The sum of fractional populations was normalized to a total area of one ($\sum_i A_i = 1$) and then multiplied with the total number of events contained in the histogram, N_{tot}, to yield the model function:

$$g(E) = N_{\text{tot}} \sum_{i=\text{F,U}} A_i \, G(E; \mu_i, \sigma_i). \tag{4.1}$$

One such function is needed for every denaturant concentration at which a histogram is obtained, each with its own set of variables for positions, μ_i, widths, σ_i and fractional populations, A_i.

Instead of fitting all histograms independently ("locally"), the fractional populations, A_i, were globally coupled by a linear free energy relation between the two states, in which the free energy difference between F and U follows the GdmCl activity, D, linearly. The necessity for this global coupling is not evident from the data depicted in fig. 4.6, which apparently only represents two well separated populations. Global coupling is, however, necessary if populations overlap more strongly, such as in measurements at lower pH or models that approximate the dataset of fig. 4.6 with more than two populations, both of which are introduced in the following sections. The means and widths of both F and U population that result from a globally coupled model deviate only slightly from the results of a completely local model, as seen in fig. 4.7. The only noticeable differences between the global and the local method of fitting concern the F population at GdmCl concentrations around 2 M.

4.2.2 Global Coupling of the Fractional Populations

As the areas A_i represent fractional populations, they must conform to the laws of thermodynamics and statistics. The areas were coupled over the whole range of denaturant

Figure 4.6: Page 78: smFRET histograms of RNase H at pH 7.2 (grey bars), obtained by burst analysis experiments on the M2 setup with a γ-factor of 1.0. Each histogram was fitted with a sum of Gaussian distributions (black line), with the areas globally coupled via Boltzmann statistics and a linear free energy model. The measurements were performed in the order of increasing GdmCl concentrations, as indicated in the respective panels. **a)** Two-state analysis, where the fit function contains only two Gaussian components, representing folded (red line) and unfolded (green line) conformations. **b)** Three-state analysis, in which an additional intermediate population (yellow line) was introduced. While means and widths of each component were adjusted independently for each histogram in the case of I, some of these parameters were subject to global coupling for the F and U species. For details about the fitting procedures, see section 4.2.2.

concentrations using Boltzmann statistics, which governs the relative populations of states in thermodynamic equilibrium. For states F and U

$$\frac{A_{\mathrm{F}}}{A_{\mathrm{U}}} = \exp\left(-\frac{\Delta G_{\mathrm{FU}}}{RT}\right), \tag{4.2}$$

with temperature T and universal gas constant R. ΔG_{FU} represents the difference in Gibbs free energy between the F and U states per mole. For proteins, these differences in free energy are due to a complex interplay between the aqueous solvent, the denaturant molecules and the solvent accessible surface area (ASA) of the protein. The solvation free energy of amino acid side chains, buried inside of the protein in its native conformation, increases with GdmCl concentration. The hydrophobic effect that drives the internal residues out of the solvent becomes less pronounced and the unfolded state more stable, as denaturant is added.[113,221–223]. The effect is markedly nonlinear for GdmCl. The relationship between denaturant concentration and Gibbs free energy can be linearized by introducing the GdmCl activity, D, which is connected with the GdmCl concentration, c, according to the empirical relationship

$$D = \frac{C_{0.5} \cdot c}{C_{0.5} + c}, \tag{4.3}$$

with the coefficient $C_{0.5} = 7.5$ M for GdmCl in water[223,224].
The change in Gibbs free energy of any state can be expressed as a function of GdmCl activity with the so-called cooperativity parameter m:

$$G(D) = G(0) + mD. \tag{4.4}$$

Consequently, the free energy difference between the two conformations F and U is also linearly dependent on D,

$$\Delta G_{\mathrm{FU}}(D) = \Delta G_{\mathrm{FU}}(0) + \Delta m_{\mathrm{FU}} D. \tag{4.5}$$

The Δm values show a strong correlation with the amount of surface area that is exposed to the solvent upon unfolding[225,226],

$$\Delta m \propto \Delta \mathrm{ASA}. \tag{4.6}$$

4.2.3 Fractional Populations in the Two-State Model

Eqns. 4.2 and 4.5 describe the ratio between the subpopulations F and U over the whole range of denaturant concentrations, only requiring two global parameters: The free energy difference between the two states under denaturant free conditions, $\Delta G_{\mathrm{FU}}(0)$, and the difference of the cooperativity parameters of both states, Δm_{FU}. The actual fractional populations follow directly from the free energy changes with GdmCl activity.

As $A_{\mathrm{F}}(D)$ and $A_{\mathrm{U}}(D)$ represent fractional populations, they are normalized according to

$$A_{\mathrm{U}}(D) + A_{\mathrm{F}}(D) = 1 \tag{4.7}$$

at all activities D independently. Inserting eqn. 4.2 into eqn. 4.7 and solving for $A_{\mathrm{F}}(D)$ yields:

$$A_{\mathrm{F}}(D) = \frac{1}{1 + \exp\left(\frac{\Delta G_{\mathrm{FU}}(D)}{RT}\right)}, \tag{4.8}$$

with $\Delta G_{\mathrm{FU}}(D)$ given by eqn. 4.5. Analogously, $A_{\mathrm{U}}(D)$ is

$$A_{\mathrm{U}}(D) = \frac{1}{1 + \exp\left(-\frac{\Delta G_{\mathrm{FU}}(D)}{RT}\right)}. \tag{4.9}$$

4.2.4 Fixing the Folded State Distribution

The native protein state is characterized by a well defined structure. Therefore, it should be represented by a Gaussian distribution with identical means, μ_{F}, and widths, σ_{F}, in all histograms, independent of denaturant concentration. However, the mean of the F state, μ_{F}, exhibits a small but noticeable shift to lower values as the GdmCl concentration, c, increases. One possible reason for that effect is discussed in section 2.2.5: The Förster distance depends on the denaturant concentration through changes in refractive index of the solution, changes in the overlap integral due to spectral shifts of the donor and the acceptor dye, changes in the quantum yield of the donor, as well as possibly changes in the κ^2-factor. Other possible causes might be solvent dependent photophysical properties of the fluorophores or the way in which the protein–dye–linker system behaves under increasing denaturant concentrations. The width σ_{F}, however, indeed appears to be constant within the experimental precision. Thus, a linear behavior was introduced for the means and widths of F:

$$\begin{aligned} \mu_{\mathrm{F}}(D) &= \mu_{\mathrm{F}}(0) - \alpha_{\mu_{\mathrm{F}}} \cdot c, && (4.10) \\ \sigma_{\mathrm{F}}(D) &= \sigma_{\mathrm{F,global}}. && (4.11) \end{aligned}$$

Instead of one local variable $\mu_{\mathrm{F,D}}$ at each concentration, two global variables $\mu_{\mathrm{F}}(0)$ (the mean of the F state at $c = 0\,\mathrm{M}$ GdmCl) and $\alpha_{\mu_{\mathrm{F}}}$ describe the behavior of $\mu_{\mathrm{F}}(D)$ at all concentrations. Moreover, only one global value of $\sigma_{\mathrm{F,global}}$ was used for all histograms, instead of a multitude of concentration dependent local values. The linear constraints for F are indicated in fig. 4.7 as solid black lines.

Significant deviations from this linear behavior of μ_F and σ_F appear only at the high-concentration edge of the transition region, where the population of F is small and μ_F and σ_F are only weakly determined in consequence. Considering that F adheres to the model at lower denaturant concentrations, where the fit is more meaningful, these deviations appear to be mere fitting artifacts.

4.2.5 Results of Fitting with the Two-State Model

The means and widths that the fit of the two-state model yields for the folded and the unfolded distributions are depicted in fig. 4.7a with solid circles and open squares. It only changes the original two-state fit without a fixed folded state slightly, but extrapolates the behavior of F from the low denaturant regime where F is well defined to regions where a completely free fit cannot yield meaningful results for F. This model with the globally fixed linear behavior of F and the free energy coupling between the fractional populations is called the *Two-State Model* and the fit according to the model is depicted in figs. 4.6a (red and green lines) and 4.7a (solid circles and open squares). The best fit parameter for the global width of F is $\sigma_\mathrm{F,global} = (0.048 \pm 0.001)$, slightly larger than the lower limit set by shot noise, $\sigma_\mathrm{SN} = 0.035$ for the folded peak (calculated according to eqn. 2.29 with $E = 0.9, n_\mathrm{tot} = 75$). As is apparent from figs. 4.6 and 4.7, the value of μ_F at $c = 0\,\mathrm{M}$ GdmCl does not fit that linear relationship. Otherwise, the model describes the data well.

4.2.6 Fixing the Width of the Unfolded State Distribution

The U state shows a strong peak shift towards low FRET efficiencies as the denaturant concentration increases. U first appears at $\mu_\mathrm{U}(1.5\,\mathrm{M\ GdmCl}) \approx 0.73$ and ends up at $\mu_\mathrm{U}(5.5\,\mathrm{M\ GdmCl}) = (0.212 \pm 0.002)$. This effect indicates a strong swelling of the unfolded coil structure, as solvent–coil contacts become more and more favorable with increasing denaturant concentration[113]. It was first observed for the cold shock protein from *Thermotoga maritima* (Csp*Tm*)[88], has since been found in a multitude of other proteins by single molecule methods[68,89,90,181,227,228] and, as such, represents no surprising behavior. More remarkable about the U distribution is the strong denaturant dependence of its width. Ranging from $\sigma_\mathrm{U}(1.5\,\mathrm{M\ GdmCl}) = (0.198 \pm 0.017)$ to $\sigma_\mathrm{U}(5.5\,\mathrm{M\ GdmCl}) = (0.090 \pm 0.002)$, the width of the U-state distribution more than doubles when GdmCl concentrations are lowered to intermediate concentrations from the high end values of ≈ 6 M. Overall, the U distribution is 2–4 times broader than the distribution representing F. The physical reasons for this behavior can be manifold:

- The shot noise, which is due to the statistical fluctuations of the number of photons that are detected per burst in each channel is calculated according to eqn. 2.29 and depends on the position of a distribution on the FRET efficiency axis, E, and the average number of photons per burst, n_tot. Calculating σ_SN at $c = 0.2\,\mathrm{M}$ reveals

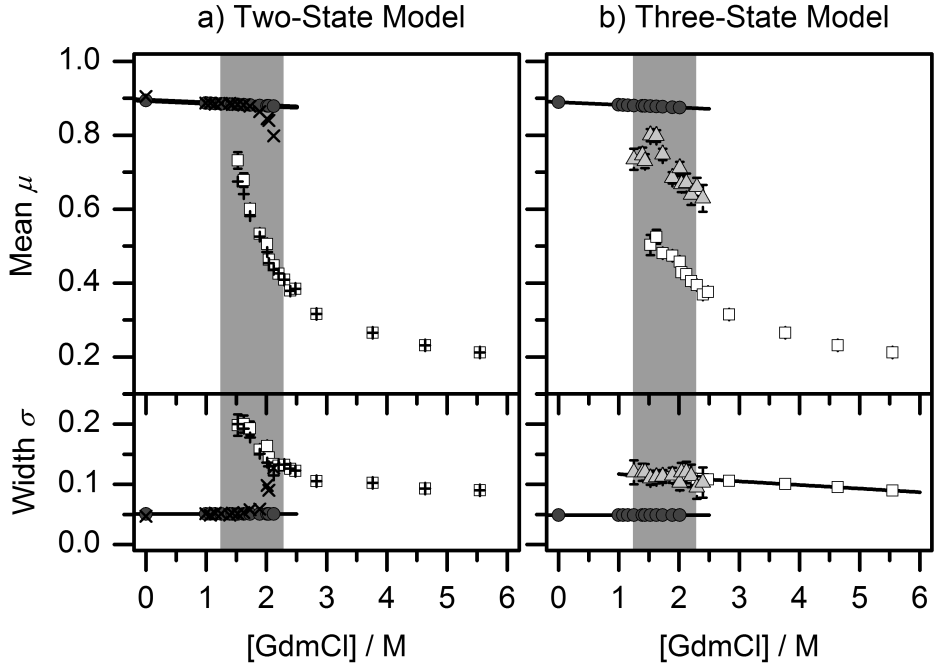


Figure 4.7: Means, μ, and widths, σ, of the Gaussian distributions used to fit the smFRET histograms in fig. 4.6. The transition region, where several configurations coexist in solution, is shaded in grey. **a)** Two-state analysis, where the fit function consisted of two Gaussian distributions representing the F ($\times$) and U ($+$) conformations. Means and widths of both Gaussians were allowed to vary freely. Forcing means and widths of F to behave linearly with GdmCl concentration leads to the results shown as closed circles for F and open squares for U. **b)** For the three-state analysis, a third, intermediate population was added (grey triangles) and, in addition to the linear constraints on F, the widths of U were forced to obey a linear dependence.

that, at low GdmCl concentrations, shot noise is only responsible for about one fourth of the observed variance[1] of the U distribution, σ^2_{obs} ($\sigma^2_{\text{shot}}(E = 0.2) = 0.0020$ and $\sigma^2_{\text{obs}}(E = 0.2) = 0.0081$). As the U distribution moves towards the middle of the FRET axis, at intermediate GdmCl concentrations, the observed width, σ_{obs}, increases by a factor of 2.2, while the shot noise, σ_{shot}, increases by barely 10%. Thus, shot noise alone is neither sufficient to fully explain the width of U nor the amount of increase in width that U exhibits towards intermediate FRET efficiencies.

---

[1] In cases where several non-correlated sources contribute to the total error of a value the different distributions can be convoluted with each other to obtain the total error distribution. If all individual errors are normally distributed or at least obey sufficiently similar distributions, the variances of the separate error distributions can simply be added to get the variance of the total error distribution ($\sigma^2_{\text{tot}} = \sigma^2_1 + \sigma^2_2 + \ldots + \sigma^2_n$) because the convolution of two Gaussians is again a Gaussian with a variance that is the sum of the original variances.

- Conformational fluctuations of the protein chain, which, in its unfolded state, is often modeled as a random coil or worm-like-chain[69,81], may affect the width of the U distribution. However, these fluctuations would have to occur on comparable or slower time scales than the experimental observation time of ≈ 1 ms in order to not be averaged out[116,229–234]. A donor-acceptor cross-correlation analysis performed on immobilized RNase H molecules by Kuzmenkina *et al.*[112] showed no evidence of any fluctuations between 20 µs and 1 ms. Other groups have also observed fast chain reconfigurations of the unfolded state and determined their time scales to be on the sub-microsecond level[121,235–237]. Therefore, the chain reconfigurations presumably do not influence the width of the U distribution.
  However, Kuzmenkina *et al.*[112] also reported observations on the dynamics of U on a time scale slow enough to account for the observed heterogeneity. It has been proposed that RNase H may retain some residual structure in the unfolded state, due to the formation of hydrophobic clusters, formed by aromatic or aliphatic residues[238,239]. Interconversion between these states may be slow enough to have an effect on σ_{U}.

- Photophysical artifacts of the fluorescent dye molecules, as proposed by Eaton and coworkers, may have to be considered[121,240]. Quenching of the dye molecules by Trp and Tyr residues has profound effects on the FRET efficiency[241,242]. While this effect may be small in the folded state, where aromatic amino acids are hidden in the core of the protein, the increased flexibility of the unfolded state may facilitate these dye-quencher interactions.

Of all these possible effects, only shot-noise predicts a broadening of U towards low GdmCl concentrations. As the mean FRET efficiency of the collapsing chain increases towards $E = 0.5$, so does the shot noise (see fig. 2.17). However, once the FRET efficiency grows beyond 0.5, shot noise is reduced again, whereas the observed width of U reaches its maximum at $\mu_{\mathrm{U}} = 0.7$. The other possible explanations predict a behavior for U that is in conflict with our observations. With the compaction of the polymer chain one would expect a decrease in heterogeneity, not an increase.

It is noteworthy that the model function describing U broadens in the transition region to almost twice its width at high denaturant concentrations. This fact lacks a plausible explanation. This counterintuitive behavior had already been observed in previous work on immobilized proteins. The smFRET histograms in earlier studies[112,113] also revealed some population in between F and U, which at the time was not captured well by the fit.

4.2.7 Introduction of an Intermediate State

Although the aforementioned two-state model approximates the data reasonably well, it raises a question about the existence of a folding intermediate, a state that has been proposed from optical observations in bulk as well as single-molecule experiments with optical tweezers[131,194–196,198,243]. These observations, combined with the conundrum of the broadening U-state, motivates the introduction of a third population to the model, representing an intermediate state (I). This intermediate distribution is not visible as a distinct peak in the histogram, likely because of a strong overlap with the F and U peaks and its small fractional weight. To fit this *three-state model*, an additional Gaussian distribution representing the I state is added to the fit function

$$g(E) = N_{\mathrm{tot}} \sum_{i=\mathrm{F,I,U}} A_i \, G(E; \mu_i, \sigma_i). \tag{4.12}$$

The folded state is subject to the same restrictions as in the two-state model, which means the width of F is kept constant, while the mean of F shifts linearly with GdmCl concentration. The width of U is forced to follow a weak linear dependence as well (solid black lines in 4.7b):

$$\sigma_{\mathrm{U}}(D) = \sigma_{\mathrm{U}}(0) - \alpha_{\sigma_{\mathrm{U}}} \cdot c. \tag{4.13}$$

This means that the parameters μ_F, σ_{F} and σ_{U}, characterizing the mean of F and the widths of both the F and the U states, are determined at conditions where F and U are present as the sole conformation, respectively. The linear dependencies extrapolate the behavior of both states into the transition region where both distributions overlap with the I distribution. These restrictions on F and U are imposed to give the I distribution as much local freedom as possible. The mean and the width of the I distribution are local parameters, independently adjusted in each histogram, with the sole restriction that μ_{I} has to lie in between the values of F and U ($\mu_{\mathrm{U}} < \mu_{\mathrm{I}} < \mu_{\mathrm{F}}$). As the unfolded protein chain is expected to collapse towards low GdmCl concentrations, a shift in μ_{U} is expected, and μ_{U} is kept as a free local parameter for each concentration.

The populations of all three conformations were connected, using the same thermodynamic free energy coupling as in the case of the two-state fit (see section 4.2.2). For a three-state system, two more thermodynamic parameters have to be added, e.g. $\Delta G_{\mathrm{FI}}(0)$ and Δm_{FI}. As is evident from fig. 4.8, the third energy difference can be easily calculated from the other two. Thus, the total number of fit parameters needed to describe the fractional populations of the F, U and I states at all GdmCl concentrations reduces from two local variables per histogram to a total of four global variables.

4.2.8 Fractional Populations in the Three-State Model

The derivation for the fractional populations of three components follows basically the same path as for the two-state case, with one added component $A_{\mathrm{I}}(D)$. For three components U, I and F, three initial equations are written as:

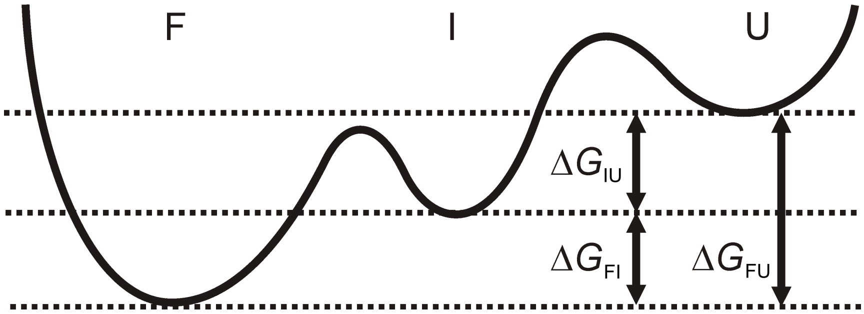


Figure 4.8: Free energy landscape of a three-state system.

$$\frac{A_{\mathrm{F}}(D)}{A_{\mathrm{U}}(D)} = \exp\left(-\frac{\Delta G_{\mathrm{FU}}(D)}{RT}\right), \tag{4.14}$$

$$\frac{A_{\mathrm{F}}(D)}{A_{\mathrm{I}}(D)} = \exp\left(-\frac{\Delta G_{\mathrm{FI}}(D)}{RT}\right), \tag{4.15}$$

$$A_{\mathrm{F}}(D) + A_{\mathrm{I}}(D) + A_{\mathrm{U}}(D) = 1. \tag{4.16}$$

Inserting eqns. 4.14 and 4.15 into eqn. 4.16 eliminates A_{U} and A_{I}:

$$A_{\mathrm{F}}(D)\left[1 + \exp\left(\frac{\Delta G_{\mathrm{FU}}(D)}{RT}\right) + \exp\left(\frac{\Delta G_{\mathrm{FI}}(D)}{RT}\right)\right] = 1. \tag{4.17}$$

Solving for $A_{\mathrm{F}}(D)$ yields an expression that describes $A_{\mathrm{F}}(D)$ over the whole range of denaturant concentrations:

$$A_{\mathrm{F}}(D) = \frac{1}{1 + \exp\left(\frac{\Delta G_{\mathrm{FU}}(D)}{RT}\right) + \exp\left(\frac{\Delta G_{\mathrm{FI}}(D)}{RT}\right)}. \tag{4.18}$$

In an analogous fashion, expressions for $A_{\mathrm{U}}(D)$ and $A_{\mathrm{I}}(D)$ are derived as

$$A_{\mathrm{U}}(D) = \frac{1}{1 + \exp\left(-\frac{\Delta G_{\mathrm{FI}}(D)}{RT}\right) + \exp\left(\frac{\Delta G_{\mathrm{FU}}(D) - \Delta G_{\mathrm{FI}}(D)}{RT}\right)}, \tag{4.19}$$

$$A_{\mathrm{I}}(D) = \frac{1}{1 + \exp\left(-\frac{\Delta G_{\mathrm{FU}}(D)}{RT}\right) + \exp\left(\frac{\Delta G_{\mathrm{FI}}(D) - \Delta G_{\mathrm{FU}}(D)}{RT}\right)}. \tag{4.20}$$

4.2.9 Results of Fitting with the Three-State Model

The fit curves that were obtained by modeling the measured data with the three-state model function are shown in fig. 4.6b. The values of μ_{I} and σ_{I} are depicted in fig. 4.7, where it is also clear from the errors that the parameters describing I, especially σ_{I}, are less well defined than those of the other populations, despite all the restrictions placed on F and U. All global best-fit parameters are compiled in table 4.1. The free energy relations and the resulting fractional populations are depicted in fig. 4.9 as functions of the GdmCl activity.

The swelling effect of U is preserved in the three-state model, albeit in a less pronounced form, from $\mu_{\mathrm{U}}(1.5\,\mathrm{M\ GdmCl}) \approx 0.5$ to $\mu_{\mathrm{U}}(5.5\,\mathrm{M\ GdmCl}) = (0.212 \pm 0.002)$. Thus, part of the initially observed extremely strong shift of μ_{U} appears to be a mere fitting artifact of the two-state model. In the absence of an intermediate distribution, the two-state fit has to use the U distribution to capture the population around $E \approx 0.7$ that, in reality, belongs to the intermediate state. The three-component fit places the I state at mean FRET efficiencies between 0.6 and 0.8, right where an excess in population was noticed in earlier work on immobilized samples[113]. The intermediate I exhibits the same kind of swelling as U, which is not surprising given its partially coil-like nature.

4.2.10 Comparison between the Two-State and the Three-State Model

It is hardly surprising that the three-state fit better approximates the data set than its two-state counterpart. This seems clear even from simple visual comparison of the fit curves with the histograms (fig. 4.6), and is confirmed by a closer look at the residuals of both fits (two examples are shown in fig. 4.10). Whereas the residuals of the three-state fit scatter statistically around zero, the residuals of the two-state model reveal systematic oscillations, indicating that the addition of a third, intermediate distribution is indeed necessary for a proper description of the observed data.

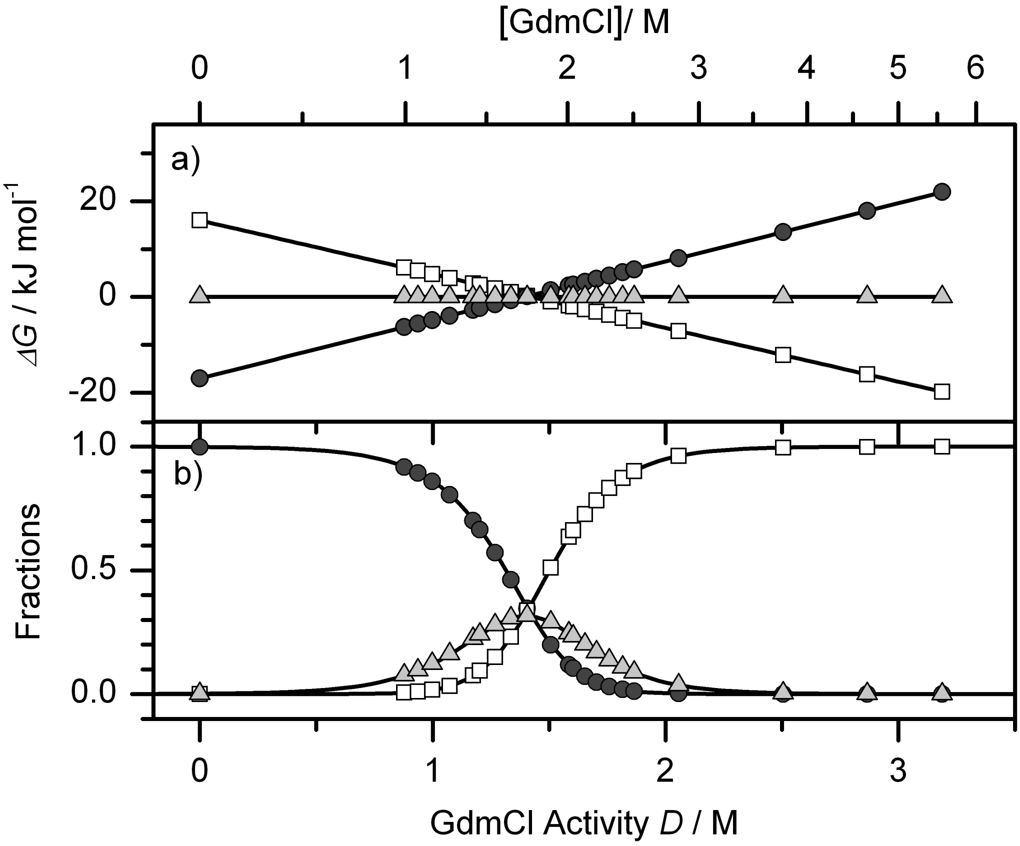


Figure 4.9: a) Free energy changes of the F (closed circles) and U (open squares) conformations, relative to I (grey triangles), plotted against GdmCl activity. The free energy differences are obtained from the three-state model, in which $\Delta G_{\mathrm{UI}}(0)$ and $\Delta G_{\mathrm{UF}}(0)$ as well as the associated Δm parameters are global fit parameters. **b)** Resulting fractional populations, calculated from the free energy differences above, according to Boltzmann statistics (eqns. 4.18, 4.19 and 4.20). Conformations are marked with the same symbols as in a).

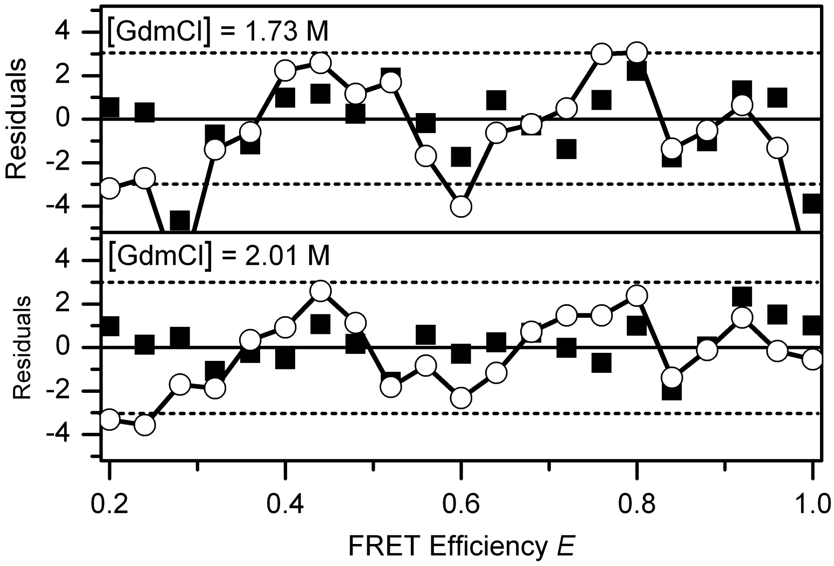


Figure 4.10: Normalized residuals of a two-state (open circles) and a three-state model (closed squares) fit of the FRET efficiency histograms taken at 1.73 M and 2.01 M GdmCl (shown in fig. 4.6). Dotted lines mark the 3σ interval.

Table 4.1: Compilation of the best fit parameters of the three-state model, pH 7.2 data set, a series of 22 histograms. While the global variables are unique, each local variable exists as a separate quantity in each histogram.

Variable Type	Number	Describes	Variable Name	Value
Global	9	F-State	$\mu_{\mathrm{F}}(0)$	(0.896 ± 0.001)
			$\alpha_{\mu_{\mathrm{F}}}$	$(0.007 \pm 0.001)\,\mathrm{M}^{-1}$
			$\sigma_{\mathrm{F,global}}$	(0.048 ± 0.001)
		U-State	$\sigma_{\mathrm{U}}(0)$	(0.155 ± 0.006)
			$\alpha_{\sigma_{\mathrm{U}}}$	$(0.012 \pm 0.002)\,\mathrm{M}^{-1}$
		Fractional Populations	$\Delta G_{\mathrm{UF}}(0)$	$(33 \pm 5)\,\mathrm{kJ\,mol}^{-1}$
			Δm_{UF}	$(23 \pm 1)\,\mathrm{kJ\,mol}^{-1}\,\mathrm{M}^{-1}$
			$\Delta G_{\mathrm{UI}}(0)$	$(17 \pm 2)\,\mathrm{kJ\,mol}^{-1}$
			Δm_{UI}	$(12 \pm 1)\,\mathrm{kJ\,mol}^{-1}\,\mathrm{M}^{-1}$
Local	88	Total Area	N_{tot}	
		U-State	μ_{U}	see fig. 4.7b
		I-State	μ_{I}	see fig. 4.7b
			σ_{I}	see fig. 4.7b

4.3 smFRET Measurements at pH 3.0

Expanding upon the investigations of RNase H at pH 7.2, further experiments were conducted at pH 3.0, a condition expected to stabilize and thus enhance the population of the intermediate conformation. It was earlier observed that, at very low pH, the I state of RNase H appears even without the addition of denaturant in a cooperative transition described by a midpoint at pH 1.9[130]. pH 3.0 marks the beginning of this transition, where the I state is just not yet populated in a denaturant-free environment.

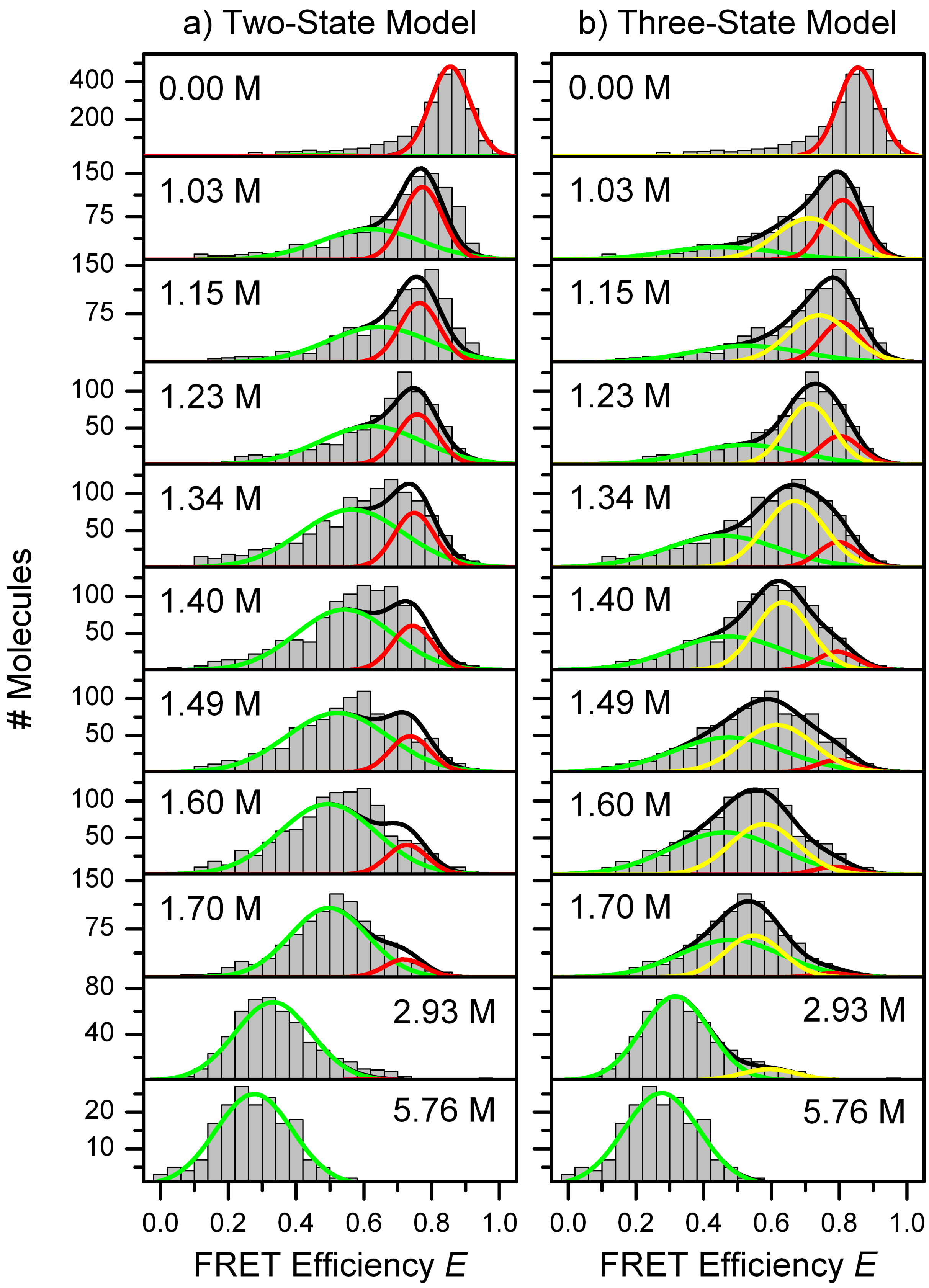
Freely diffusing RNase H 3-135 molecules were probed by burst analysis measurements in the same way as in the experiment at pH 7.2 (see section 4.2) on the MicroTime 200 setup with a γ-factor of 0.6. The results were compiled into a series of 16 smFRET histograms, a selection of which is shown as grey bars in fig. 4.11. At low GdmCl concentrations ($c < 0.5\,\mathrm{M}$), the histogram consists of practically only one distribution, which is assigned to the folded state. Upon increasing denaturant concentration, this distribution develops a shoulder towards lower FRET efficiencies. At the same time it widens, and its center of gravity shifts towards the middle of the FRET efficiency axis.

In the denaturant regime between 0.5 M and 2 M, folded, intermediate and unfolded conformations all coexist, leading to a featureless, broad distribution. At GdmCl concentrations above 3 M the distribution represents only one component again, the unfolded state of RNase H. As in the experiment at pH 7.2, the histograms were modeled with a two-state model and a three-state model, both sums of Gaussian distributions, according to eqn. 4.1:

$$g(E) = N_{\mathrm{tot}} \sum_{i=\mathrm{F},(\mathrm{I}),\mathrm{U}} A_i \, G(E; \mu_i, \sigma_i). \tag{4.21}$$

However, even with the global restraints on F and U that were developed for the analysis of the data at pH 7.2, it was not possible to obtain meaningful results in the case of the

Figure 4.11: Page 91: smFRET histograms of RNase H at pH 3.0 (grey bars), obtained by burst analysis experiments. Each histogram was fitted with a sum of Gaussian distributions (black line), the areas of which were globally coupled via Boltzmann statistics and a linear free energy model. The model also simultaneously fits Trp fluorescence data (see fig. 4.13), coupled to the smFRET measurement via the fractional populations. For details about the fitting procedures, see section 4.2.2. **a)** Two-state analysis, where the fit function only contains two Gaussian components, representing folded (red line) and unfolded (green line) conformations. **b)** Three-state analysis, in which an additional intermediate population (yellow line) is added. While means and widths were adjusted independently for each histogram in the case of I, some of these parameters were subject to global coupling for the F and U species.


a) Two-State Model
b) Three-State Model
# Molecules
FRET Efficiency E
0.00 M
1.03 M
1.15 M
1.23 M
1.34 M
1.40 M
1.49 M
1.60 M
1.70 M
2.93 M
5.76 M
400
200
150
75
100
50
80
40
20
10
0.0
0.2
0.4
0.6
0.8
1.0

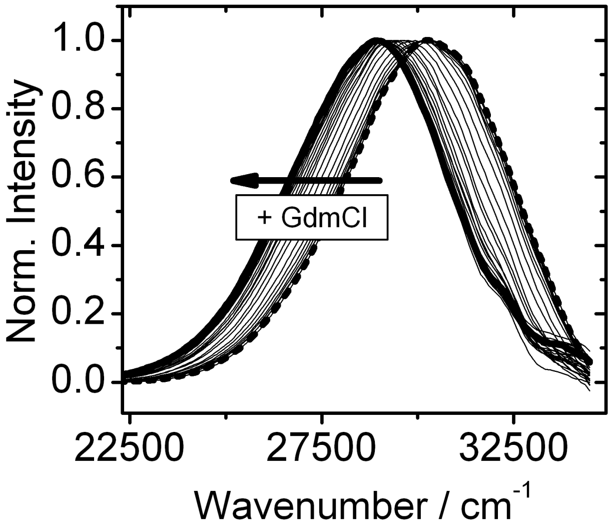


Figure 4.12: Shift of the Trp fluorescence peak of RNase H at increasing GdmCl concentrations. Spectra were taken at pH 3.0 and T = 23°C. After linear background correction, all spectra were normalized to the same peak height.

three-state model. The three Gaussian distributions overlap to such a degree that they appear as only one, smeared out distribution without apparent underlying structure. Fit parameters, such as widths and means of the individual Gaussians, are not independent from each other, so that changes in one parameter can be counteracted by changes in other parameters, ultimately leaving all parameters ambiguous and ill defined. The straightforward fit with the three-state model became unstable to a point at which the starting conditions of the fit parameters determined the outcome of the fitting procedure. In a case like this, where many different local minima exist for the fit, additional data from a different experimental technique can help to determine the right description of the data set. One such data source is the Trp fluorescence of RNase H, which strongly depends on the folding state of the protein. While a Trp fluorescence experiment obviously cannot help with assigning FRET efficiencies and widths of FRET distributions, it may, however, provide additional information that can be used to determine the fractional populations of the three conformations.

4.3.1 Trp Fluorescence Measurements

Trp fluorescence measurements were performed at 33 concentrations evenly distributed between 0 and 5.6 M GdmCl in standard buffer. RNase H samples were prepared at a concentration of 22 µM in standard buffer and emission spectra were recorded in the range from 290 – 450 nm with a resolution of 1 nm and an excitation wavelength of 280 nm. The spectra are shown in fig. 4.12. After linear background correction, the peak positions, ν_{obs}, of all spectra were determined via Gaussian fits with Origin 8.1. They are plotted in fig. 4.13 as open circles. RNase H contains six Trp residues at positions 81, 85, 90, 104, 118 and 120, as depicted in fig. 2.25. In the F-state, these Trp residues are buried inside the folded globule and are inaccessible to the solvent, with the exception of Trp81, which has one face of its indole side chain exposed to the solvent. The emission peak for this conformation was determined to be $\nu_{\mathrm{F}} = (30,328 \pm 6)\,\mathrm{cm}^{-1}$, or $\lambda_{\mathrm{F}} = (329.7 \pm 0.1)\,\mathrm{nm}$.

In the denatured U-state, all Trp residues are solvent accessible to the highest possible degree, resulting in a significantly red-shifted emission band at $\nu_U = (28,814 \pm 7)\,\mathrm{cm}^{-1}$, or $\lambda_U = (347.1 \pm 0.1)\,\mathrm{nm}$. Although not as directly accessible as the peak positions of F and U, the emission maximum of I can be determined with a global fitting procedure (see section 4.3.2), yielding $\nu_I = (29,280 \pm 30)\,\mathrm{cm}^{-1}$, or $\lambda_I = (341.5 \pm 0.3)\,\mathrm{nm}$. Thus, the peak shift from the F → I transition is roughly twice as large as the peak shift resulting from the transition from I to U. This difference corresponds with the fact that during the F → I transition four buried Trp residues (85, 90, 118 and 120) get exposed to the solvent, whereas the I → U transition increases the solvent exposure of only two Trp residues (81 and 104).

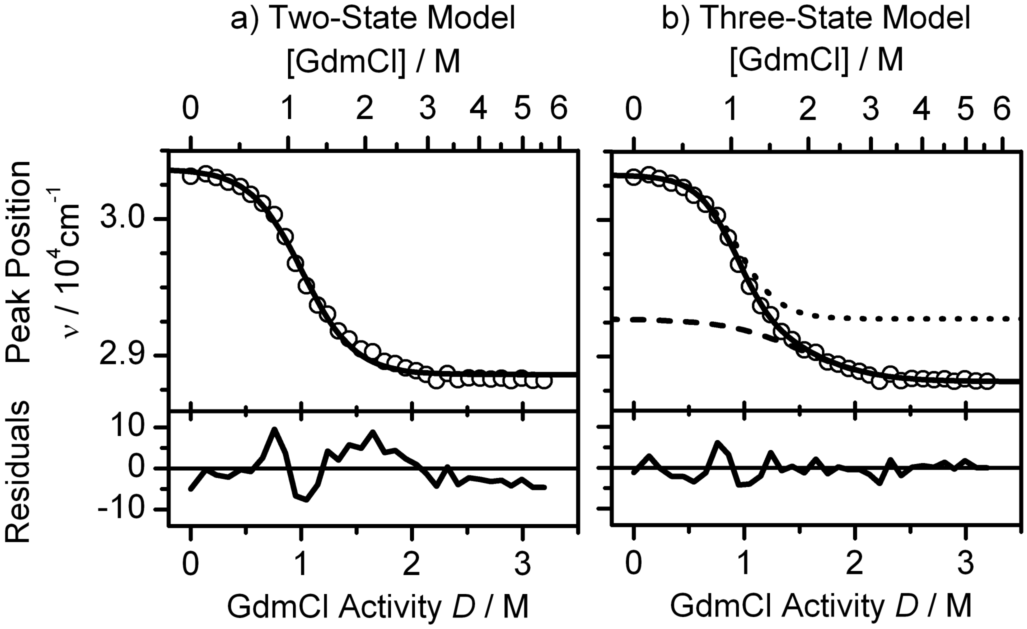


Figure 4.13: Trp emission peak position of RNase H at pH 3.0, plotted as a function of GdmCl activity D (open circles), and fitted with two-state and three-state models, respectively (solid lines). Relative errors are smaller than 0.2% for all measurements and the error bars are smaller than the symbols. This fit was performed in conjunction with the smFRET data (see fig. 4.11), using one global model to fit both data sets simultaneously. For the three-state model, the transitions between F and I (dotted line) and between I and U (dashed line) are shown as well. The residuals indicate the systematic deviations of the two-state model, whereas the residuals of the three-state model only show the expected statistical scatter around zero.

4.3.2 Combined Fitting of Trp Fluorescence and smFRET Data

Analogously to the smFRET histograms, Trp emission spectra were fitted with a function that models each Trp fluorescence spectrum as a linear superposition of Gaussian bands, each representing one of the three protein conformations F, I and U. As is obvious from the Trp emission spectra (fig. 4.12), these bands are not well resolved. Therefore, only one Gaussian was fitted to each spectrum, and the observed peak wavenumber, ν_{obs}, was assumed to be a population-weighted sum:

$$\nu_{\mathrm{obs}} = \frac{\sum_{i=\mathrm{F,I,U}} A_i(D) B_i \nu_i}{\sum_{i=\mathrm{F,I,U}} A_i(D) B_i}. \tag{4.22}$$

The ν_i represent the denaturant-independent peak positions of the spectra representing the different states (F, I and U). The additional weighting factors B_i account for the different brightnesses of the conformations and are assumed to be denaturant independent as well. The treatment of replacing three Gaussian distributions by only one is justified by the following consideration: The first moment of a normalized Gaussian $G(x, \mu, \sigma)$ is μ by definition and thus:

$$\int_{-\infty}^{\infty} x \, G(x, \mu, \sigma) \, \mathrm{dx} = \mu. \tag{4.23}$$

The total mean μ_{tot} of a sum of normalized Gaussians with weighting factors A_i is thus calculated according to

$$\begin{aligned} \mu_{\mathrm{tot}} &= \int_{-\infty}^{\infty} x \sum_i A_i \, G(x, \mu, \sigma) \, \mathrm{dx} && (4.24) \\ &= \sum_i A_i \underbrace{\int_{-\infty}^{\infty} x \, G(x, \mu, \sigma) \, \mathrm{dx}}_{\mu_i} && (4.25) \\ &= \sum_i A_i \, \mu_i. && (4.26) \end{aligned}$$

This relation is true in general, whether the Gaussians overlap or not. It is specifically also true in the case of Gaussians that are close enough together to appear unimodal. According to Behboodian[244], a mixture of two normal distributions is unimodal if their respective means μ_1 and μ_2 are closer than

$$|\mu_2 - \mu_1| \leq 2 \min(\sigma_1, \sigma_2). \tag{4.27}$$

Any sum of Gaussians for which eqn. 4.27 is true, is again Gaussian in shape. The mean μ_{tot} of a Gaussian distribution resulting from such a sum can, therefore, be experimentally obtained from the data by applying a single Gaussian fit, without prior knowledge about the underlying components. For the Trp emission spectra presented here, the largest difference in peak positions was between the pure F spectrum in GdmCl-free conditions and the pure U spectrum at high GdmCl concentrations, $|\mu_\mathrm{F} - \mu_\mathrm{U}| \approx 1{,}500\,\mathrm{cm}^{-1}$. The widths of the spectra of F and U were found to be $\sigma_\mathrm{F} \approx 2{,}000\,\mathrm{cm}^{-1}$ and $\sigma_\mathrm{U} \approx 1{,}700\,\mathrm{cm}^{-1}$. As such, eqn. 4.27 holds, at least for the spectra of F and U, and any spectrum formed as a sum of these two individual spectra can, in fact, be fitted with only one Gaussian distribution. While the width of I could not be directly determined, the spectrum of I would have to be more than twice narrower than the spectra of F and U to violate inequality 4.27.

The connection between smFRET and Trp fluorescence measurements is achieved via the fractional populations, $A_i(D)$, which play crucial roles in both model functions (eqn. 4.21 for smFRET histograms and eqn. 4.22 for Trp fluorescence). The behavior of the fractional populations is again described globally by a linear free energy model that determines the fractional distributions of population among the different conformations at all GdmCl concentrations (see section 4.2.2): Given the thermodynamic parameters $\Delta G(0)$ and Δm, a set of fractional populations A_i can be obtained for any denaturant activity D, allowing the combination of smFRET and Trp fluorescence measurements, which are not necessarily taken at the same denaturant concentrations. The $\Delta G(0)$ and Δm parameters have to be chosen to simultaneously describe FRET histograms and Trp peak positions in one simultaneous global fit.

The B_i parameters only describe the relative brightnesses of different conformations with respect to each other. Because eqn. 4.22 is normalized, the B_I values do not necessarily need to be normalized, but they may as well be without loss of generality: $\sum_i B_i = 1$. Therefore, knowledge of all but one B_i immediately defines the last B parameter as well. Therefore, using all three B_i as fit parameters would leave the system overparameterized and the brightness ratios $\alpha_\mathrm{UF} = B_\mathrm{U}/B_\mathrm{F}$ and $\alpha_\mathrm{IF} = B_\mathrm{I}/B_\mathrm{F}$ are used instead. α_UF is directly accessible by comparing the intensities of the Trp fluorescence at the highest and the lowest GdmCl concentrations and is determined as $\alpha_\mathrm{UF} = I_\mathrm{U}/I_\mathrm{F} = (0.59 \pm 0.01)$. The parameter α_IF was obtained by the global fit, which yielded $\alpha_\mathrm{IF} = (0.73 \pm 0.08)$. As I is more folded than U and, thus, more similar to F, α_IF being closer to unity than α_UF is expected.

The results of the Trp fluorescence part of the fit are shown as black lines in fig. 4.13. Depicted as well is the exchange between F and I (dotted line), which exhibits an approximately twofold larger peak shift than the one of the I-to-U transition (dashed line). When transitioning from F to I, four Trp residues (Trp85, Trp90, Trp118 and Trp120) experience changes in their solvent environment, while the I-to-U transition results in changes for only two Trp groups (Trp 81 and Trp104). The restrictions on the F and the U distribution were the same as for pH 7.2: The mean, μ_F, of the F state as well as the width, σ_U, of U were again globally constrained to behave linearly, σ_F was kept constant over the whole denaturant range. The U-state position μ_U was again fitted locally, and shows the expected collapse of the unfolded state towards low denaturant concentrations. Only the I-state was free of any global restrictions concerning its mean and width: μ_I and σ_I were both local parameters and, thus, fitted independently for each histogram. All best-fit parameters of the three-state model are depicted in fig. 4.14 and table 4.2. The free energy

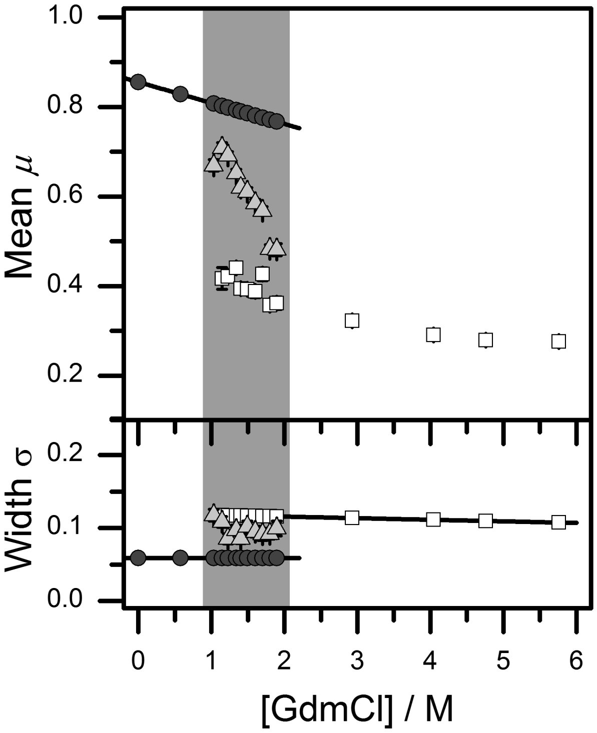


Figure 4.14: Means μ and widths σ of the three Gaussian distributions that represent folded (closed circles), intermediate (grey triangles) and unfolded (open squares) populations in the model function fitted to the smFRET histograms (fig. 4.11). To facilitate comparisons, the scale has been chosen to be the same as in the corresponding plot for the pH 7.2 data (fig. 4.7). As indicated by the black lines, the mean of F as well as the widths of F and U were globally constrained to exhibit linear behavior.

relationships between the three conformations and the resulting fractional populations are depicted in fig. 4.15. Contrary to pH 7.2, the I state is the most stable conformation between $\approx 1\,\mathrm{M} < D < 1.5\,\mathrm{M}$ GdmCl. In that region, it has the lowest free energy of the three states and is, consequently, most strongly populated.

In hindsight, it becomes clear why the smFRET data alone would not have been sufficient to obtain meaningful results for I, due to the strong overlap of the I distribution with the F and U distributions in the smFRET histogram. If the mixture of any two normal distributions is unimodal (according to the criterium given before in eqn. 4.27) it can be fitted with only one Gaussian distribution with three independent fit parameters A, μ and σ describing its area, mean and width. Using twice that number of parameters to model two Gaussian distributions obviously leaves the system over-parameterized: Changes in one parameter can be compensated by adjustments in other parameters, ultimately leaving all parameters ambiguous and of little significance.

To see if the distributions representing F, I and U in the smFRET histogram fits at pH 3.0 overlap to an extent that satisfies eqn. 4.27, it is sufficient to take a look at the widths and positions obtained by the three-state fit. The width of the folded population is directly accessible at low GdmCl concentrations, where only F is present ($\sigma_{\mathrm{F}} = 0.059$). The states U and I have widths $\sigma_{\mathrm{I/U}}$ of roughly 0.1 and above. Considering that, at GdmCl concentra-

Table 4.2: List of all parameters required to fit the pH 3.0 data set, a series of 16 sm-FRET histograms combined with Trp fluorescence data. While the global variables are unique, each local variable exists as a separate instance in each histogram.

Variable Type	Number	Describes	Variable Name	Value
Global	14	F-State	$\mu_{\mathrm{F}}(0)$	(0.85 ± 0.02)
			$\alpha_{\mu_{\mathrm{F}}}$	$(0.047 \pm 0.003)\,\mathrm{M}^{-1}$
			σ_{F}	(0.0590 ± 0.001)
		U-State	$\sigma_{\mathrm{U}}(0)$	(0.120 ± 0.007)
			$\alpha_{\sigma_{\mathrm{U}}}$	$(0.002 \pm 0.002)\,\mathrm{M}^{-1}$
		Fractional Populations	$\Delta G_{\mathrm{IF}}(0)$	$(11.6 \pm 0.5)\,\mathrm{kJ\,mol}^{-1}$
			Δm_{IF}	$(13.2 \pm 0.5)\,\mathrm{kJ\,mol}^{-1}\,\mathrm{M}^{-1}$
			$\Delta G_{\mathrm{UI}}(0)$	$(10.6 \pm 0.6)\,\mathrm{kJ\,mol}^{-1}$
			Δm_{UI}	$(7.3 \pm 0.5)\,\mathrm{kJ\,mol}^{-1}\,\mathrm{M}^{-1}$
		Relative Brightness	α_{UF}	(0.59 ± 0.01)
			α_{IF}	(0.73 ± 0.08)
		Trp Emission Maxima	ν_{F}	$(30,328 \pm 6)\,\mathrm{cm}^{-1}$
			ν_{I}	$(29,280 \pm 30)\,\mathrm{cm}^{-1}$
			ν_{U}	$(28,814 \pm 7)\,\mathrm{cm}^{-1}$
Local	64	Total Area	N_{tot}	
		U-State	μ_{U}	see fig. 4.14
		I-State	μ_{I}	see fig. 4.14
			σ_{I}	see fig. 4.14

tions of $c < 1.2\,\mathrm{M}$, the difference between μ_{F} and μ_{I} is less than 0.1, the two distributions cannot be distinguished in a free fit even in theory, let alone in noisy histogram data. The same holds true for the distinction between the I and U distributions at denaturant concentrations of $c > 1.5\,\mathrm{M}$. (See fig. 4.14).

The problem of having to fit two Gaussian distributions to a data set in which the two populations appear unimodal is overcome by reducing the number of free parameters of the fit. For the F state, μ_{F} and σ_{F} are globally determined, meaning that the width and the position of F are basically determined at low GdmCl concentrations where F is the only population. The global coupling fixes σ_{F} at this value for all histograms and allows μ_{F} only a slightly deviate from its initial value. Thus, both parameters are essentially fixed in the transition region, where the F and I populations coexist. This reduces the number of free, local parameters by two. The ratio between the two fractional populations is a parameter which is not removed from the fit in a strict sense. However, the fit is heavily biased towards a relation that also describes the Trp fluorescence data, reducing the number of free parameters to three, thereby removing the over-parametrization in the region where the F and I distributions overlap. The unfolded distribution is not quite as heavily restricted as F, and some ambiguity remains above $c = 1.5\,\mathrm{M}$, where the exchange from I to U happens.

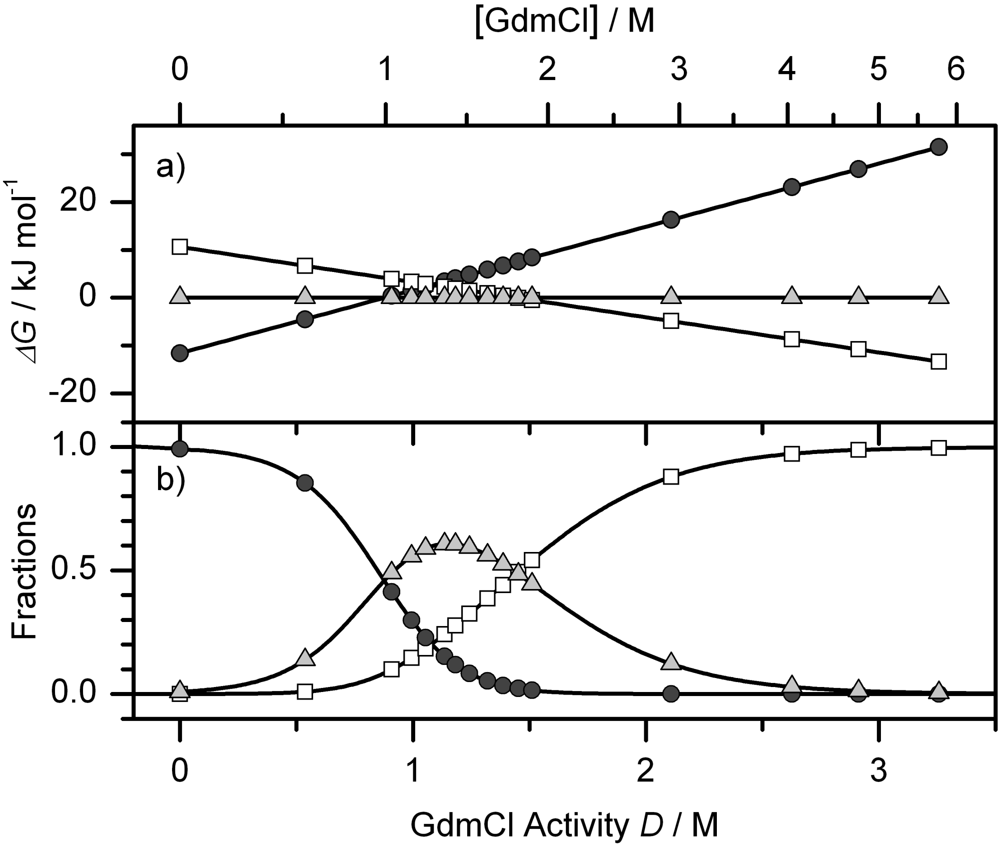


Figure 4.15: **a)** Free energies of F (closed circles) and U (open squares) conformations, relative to I (grey triangles), plotted against the GdmCl activity. The free energy differences are gained from the three-state fit, in which $\Delta G_{\mathrm{UI}}(0)$ and $\Delta G_{\mathrm{UF}}(0)$ as well as the associated Δm are global fit parameters. **b)** Resulting fractional populations, calculated from the free energy differences above, according to Boltzmann statistics (eqns. 4.18, 4.19 and 4.20). Conformations are marked with the same symbols as in a).

4.3.3 Breakdown of the Two-State Model

In addition to the three-state fit, a two-state model without an intermediate state was fitted to the data set (see fig. 4.11a). The F-state distribution was linearly constrained in mean, μ_{F}, and width, σ_{F}, whereas the mean, μ_{U}, and the width, σ_{F}, of the U state distribution were treated as local parameters. The populations of the two states F and U were globally coupled via linear free energy relations. Like its three-state counterpart, the two-state model had to simultaneously describe the Trp fluorescence data (see. fig. 4.13a).

Table 4.3: Reported values for free energy differences in the absence of denaturant $\Delta G(0)$ and Δm values, describing the free energy relation between folded, intermediate and unfolded conformation of RNase H at various pH values.

pH	$\Delta G(0)/\mathrm{kJ\,mol^{-1}}$			$\Delta m/\mathrm{kJ\,mol^{-1}\,M^{-1}}$			Ref.
	UF	IF	UI	UF	IF	UI	
3.0	$(22.2 \pm 0.8)^{a}$	11.6 ± 0.5	10.6 ± 0.6	(20.5 ± 0.7)	13.2 ± 0.5	7.3 ± 0.5	[245]
5.5	41.3	$(26.5)^{a}$	14.8	8.8 ± 0.4^{b}	1.9^{b}	5.2^{b}	[195]
5.5	39.6 ± 1.5	$(19.6 \pm 2.7)^{a}$	20.0 ± 2.3	21.9 ± 0.8	$(6.9 \pm 1.7)^{a}$	15.0 ± 1.5	[246]
7.2	33 ± 5	$(16 \pm 5)^{a}$	17 ± 2	23 ± 1	$(11 \pm 1)^{a}$	12 ± 1	[247]
7.4	34 ± 3^{c}	–	–	20 ± 2^{c}	–	–	[211]
7.4	–	17 ± 2^{d}	–	–	12 ± 1^{d}	–	[112]

[a] Values in parentheses were calculated from the two other reported parameters for comparison.
[b] These Δm values are smaller because of the use of urea as denaturant instead of GdmCl.
[c] Obtained with a two-state model
[d] Parameters of the most compact unfolded state from a model assuming a continuum of unfolded states.

Comparing the residuals of the two-state and the three-state model fits in fig. 4.13 leaves no doubt which model describes the Trp fluorescence data better. The two-state model exhibits systematic deviations from the data because one sigmoidal exchange is not sufficient to describe the observed behavior. The three-state model, on the other hand, lacks these systematic deviations. Its residuals scatter statistically around zero.

The two-state fit to the smFRET histograms is shown in fig. 4.11a together with the experimental data. Even without plotting the residuals, the insufficiency of the two-state model is obvious, especially in the concentration range between 1.15 M and 1.60 M GdmCl, the region where the intermediate appears most prominently. In these histograms, molecules are present at FRET efficiencies between those of F and U that the two-state fit is incapable of capturing. The constraints placed on the fractional population by the inclusion of Trp data into the joint fit leave the two-state model incapable of properly approximating the smFRET data. By contrast, the three-state model gives an excellent, simultaneous description of both data sets.

4.4 Discussion of the Free Energy Differences

To put the results of this work into context, table 4.3 is a compilation of the best fit parameters of the free energy differences, ΔG, and the corresponding Δm values obtained at pH 7.2 and pH 3.0 as well as literature data of those parameters measured by other groups.

The fit at pH 7.2 yielded $\Delta G_{\mathrm{UF}}(0) = (33 \pm 5)\mathrm{kJ\,mol^{-1}}$ and $\Delta m_{\mathrm{UF}} = (23 \pm 1)\mathrm{kJ\,mol^{-1}\,M^{-1}}$ for the F–U relationship. These results are the same within the error as the values of $\Delta G_{\mathrm{UF}}(0) = (34 \pm 3)\,\mathrm{kJ\,mol^{-1}}$ and $\Delta m_{\mathrm{UF}} = (20 \pm 2)\,\mathrm{kJ\,mol^{-1}\,M^{-1}}$, obtained from a two

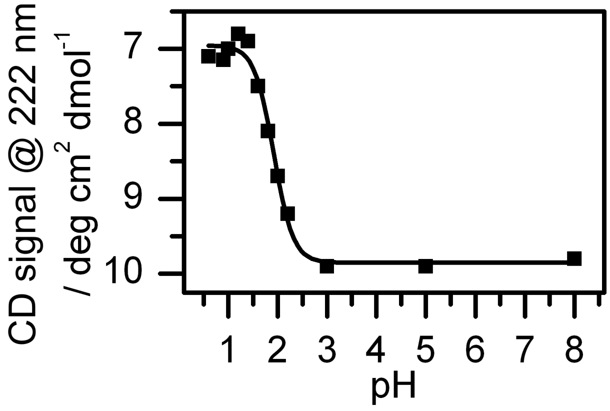


Figure 4.16: Fit of circular dichroism data from a pH denaturation experiment by Dabora *et al.*[130] with a Henderson-Hasselbalch equation.

state analysis of smFRET data measured previously in our lab at pH 7.4[211]. In another previous smFRET experiment at pH 7.4[112], in an effort to describe the swelling of U, the U state was described as a continuum of unfolded sub-states which exchange population as the GdmCl concentration increases. Although a direct comparison is not feasible, it should be remarked that the free energy difference between the F state and the most compact non-folded state U′ is $\Delta G_{\mathrm{U'F}}(0) = (17 \pm 2)\,\mathrm{kJ\,mol^{-1}}$ and exhibits a GdmCl dependency of $\Delta m_{\mathrm{U'F}} = (12 \pm 1)\,\mathrm{kJ\,mol^{-1}\,M^{-1}}$. This compact unfolded state is represented by the intermediate state in this work, which behaves identically with respect to the folded state: ($\Delta G_{\mathrm{IF}}(0) = (16 \pm 5)\,\mathrm{kJ\,mol^{-1}}$ and $\Delta m_{\mathrm{IF}} = (11 \pm 1)\,\mathrm{kJ\,mol^{-1}\,M^{-1}}$).

Comparable studies in the groups of Marqusee[195] and Kanaya[246] were performed at pH 5.5, where RNase H appears to be ≈ 20% more stable than at pH 7.2 in denaturant free conditions, with F being stabilized by a free energy difference of $\Delta G_{\mathrm{UF}}(0) \approx 40\,\mathrm{kJ\,mol^{-1}}$ with respect to U (see table 4.3). The I state lies halfway in between F and U on the free energy scale at all pH values ($\Delta G_{\mathrm{UI}}(0) \approx \Delta G_{\mathrm{IF}}(0)$).

At pH 3.0, the stability of the native conformation in the absence of denaturant is markedly reduced and $\Delta G_{\mathrm{UF}}(0)$ as well as $\Delta G_{\mathrm{IF}}(0)$ drop to $(22.2 \pm 0.8)\ \mathrm{kJ\,mol^{-1}}$ and $(11.6 \pm 0.5)\,\mathrm{kJ\,mol^{-1}}$ respectively. $\Delta G_{\mathrm{IF}}(0)$ can be compared to pH denaturation data measured by Dabora and Marqusee[130] in an experiment, in which RNase H was converted from the folded into the intermediate form as a function of pH. The data were extracted from the published figure by eye and fitted with a Henderson-Hasselbalch equation, a variant of the law of mass action that describes protonation equilibria in dependence of pH[248]:

$$A_+ = 1 - A_0 = \frac{1}{1 + 10^{n\,(\mathrm{pK}-\mathrm{pH})}}, \tag{4.28}$$

where A_+ and A_0 are the fractional populations of the protonated and unprotonated species, respectively. pK describes the transition midpoint and n is a cooperativity parameter, which is unity for single protonation equilibria and $n > 1$ for cooperative transitions involving multiple protonations. The fit yields a transition midpoint of $\mathrm{pK} = (1.92 \pm 0.03)$ and a cooperativity parameter of $n = (2.2 \pm 0.4)$ (see fig. 4.16). The free energy differences between F and I in the transition region can be extrapolated to arbitrary pH values using:

$$\Delta G(\mathrm{pH}) = \frac{RT}{\log(e)}\, n\, (\mathrm{pK} - \mathrm{pH}). \tag{4.29}$$

At pH 3.0, eqn. 4.29 yields a free energy difference $\Delta G_{\mathrm{IF}} = (13.5 \pm 2.5)\,\mathrm{kJ\,mol^{-1}}$ between F and I, a value that is equal within the error to $\Delta G_{\mathrm{IF}}(0) = (11.6 \pm 0.5)\,\mathrm{kJ\,mol^{-1}}$, the value obtained via the smFRET experiments in this work.

The Δm_{UF} values of all measurements with GdmCl as denaturant is $\approx 22\,\mathrm{kJ\,mol^{-1}\,M^{-1}}$ regardless of pH, which indicates that the total change in solvent accessible surface area does not depend on pH. These similar Δm values, combined with the differences in free energy at denaturant free conditions, indicate that the transition from F to U occurs at different GdmCl concentration under different pH conditions (see. fig. 4.17). At pH 5.5, RNase H is most stable: The transition out of the F state happens at a higher GdmCl activity ($D_{\mathrm{t}} = 1.81\,\mathrm{M}$) than at pH 3.0 and pH 7.2 and the I state is never populated more strongly than 4% (fig. 4.17b). A different picture emerges at higher and lower pH values. At pH 7.2, the I state is well developed and the transition point ($D_{\mathrm{t}} = 1.43\,\mathrm{M}$) from F to U is lower than at pH 5.5 (fig. 4.17c). At pH 3.0, the $\mathrm{F} \rightarrow \mathrm{U}$ transition is situated at an even lower value of $D_{\mathrm{t}} \approx 1.1\,\mathrm{M}$ and, in contrast to the data at the other pH values, the I state becomes the dominant species in the interval of $0.87\,\mathrm{M} < D < 1.46\,\mathrm{M}$ (fig. 4.17a).

With regard to the ASA, the I state seems to be more similar to U at pH 3.0 than at higher pH, as evidenced by the fact that Δm_{UI} at pH 3.0 is smaller than Δm_{UI} at pH 5.5 and 7.2 by almost a factor of two. As the experiment at pH 5.5 by Marqusee was performed with the denaturant urea, the Δm values of that experiment cannot be directly compared with the experiment using GdmCl.

The presented results should also be compared to preceding single-molecule experiments on RNase H, namely those of Cecconi *et al.*[196] and Kuzmenkina *et al.*[112,113]. Cecconi and coworkers observed the force-induced unfolding and refolding of single RNase H molecules immobilized between functionalized beads, one of them held by optical tweezers. Under the influence of an external force, a single protein molecule was forced to unfold and the subsequent refolding was then observed, while the force exerted on the molecule was kept constant. However, force-induced folding experiments raise the question of biological relevance[207,208], as the two attachment sites pose additional boundary conditions for the folding process, which may not be present for freely diffusing molecules in solution. Cecconi *et al.* estimated the Gibbs free energy difference between U and I with three different methods as $\Delta G_{\mathrm{UI-1}}(0) = (17 \pm 13)\,\mathrm{kJ\,mol^{-1}}$, $\Delta G_{\mathrm{UI-2}}(0) = (17 \pm 8)\,\mathrm{kJ\,mol^{-1}}$ and $\Delta G_{\mathrm{UI-3}}(0) = (16 \pm 3)\,\mathrm{kJ\,mol^{-1}}$ at a pH value of 7.0, which lie within the error of the result of this work of $\Delta G_{\mathrm{UI}}(0) = (17.2 \pm 2)\,\mathrm{kJ\,mol^{-1}}$, obtained at pH 7.2. Thus, this work shows that at least the energy differences between the conformational states found by force-induced unfolding correspond to those obtained by chemical denaturation experiments. While our work did not allow the observation of individual molecules *during* the folding process as

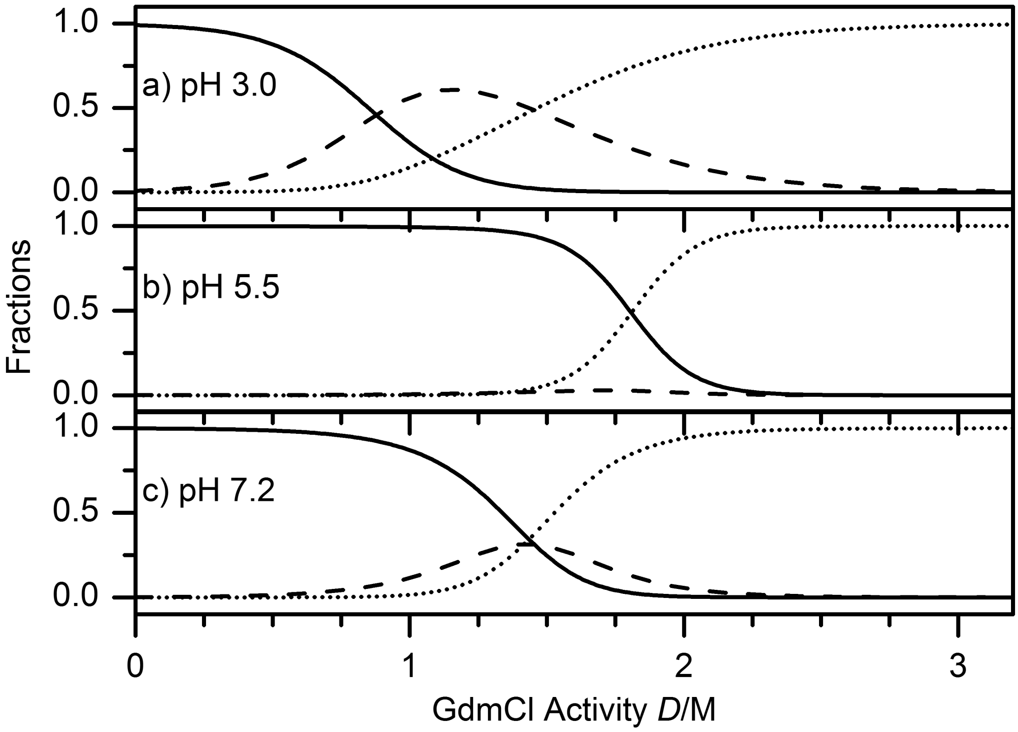


Figure 4.17: Fractional populations of F (solid lines), I (dashed lines) and U (dotted lines) populations as a function of GdmCl activity. All graphs were calculated from the $\Delta G(0)$ and Δm values in table 4.3 with eqns. 4.18 to 4.20 **a)** Fractions from the pH 3.0 measurement of this work. **b)** Fractions from the pH 5.5 measurement in the group of Kanaya[246]. **c)** Fractions from the pH 7.2 measurement of this work.

in force-induced unfolding, the burst analysis method employed here yields much better statistics, as it quickly probes thousands instead of tens of molecules.

Other notable single-molecule measurements on RNase H were done by Kuzmenkina *et al.*[112,113] When fitted with a two-state model, the U state exhibited a strong, counterintuitive broadening at intermediate GdmCl concentrations. They proposed a model with a distribution of several distinct U substates. In this work, an additional state is added to the binary picture of only folded and unfolded state. However, instead of dividing the single unfolded state into a distribution of multiple unfolded states, only one intermediate state is included. Much like Kuzmenkina's U-state continuum, the intermediate of this work also shifts to lower FRET efficiencies as the denaturant concentration increases. Both models offer an explanation for the seeming broadening of the U state in two-state models. Thus, the data presented here offer an alternative model that also fits the earlier data of Kuzmenkina, but is more in line with the three-state description of RNase H favored by other groups.

Future worthwhile studies on RNase H may include mutational studies, in which the folding intermediate of RNase H is stabilized or destabilized by specific point mutations similar to the ensemble experiments by Spudich and Connell[197,198]. In contrast to the indirect detection of the intermediate in these ensemble experiments, a single-molecule setup would be able to directly observe changes of the population of the intermediate state in the sm-FRET histograms. Furthermore, RNase H mutants in which the two attachment sites for fluorescent dyes are moved to different locations would be of great interest.

A particular mutant with dye attachment sites at positions at the beginning and the end of the protein core comes to mind directly. The FRET efficiency of a dye pair of such a mutant would mainly be sensitive to the folding and unfolding of the inner core part of the protein, that is the part of the protein which, according to current knowledge, constitutes the intermediate state. A mutant of this kind, thus, could allow the direct observation of the U–I transition and could confirm the assignment of the intermediate state to the protein core.

Chapter 5

Summary

Single molecule FRET has been established as a mature experimental technique in the biophysical toolkit. It might be surprising to find that most studies with this technique are performed on two-state folders, where the only distinction is between folded and unfolded conformations. As yet, smFRET experiments have not generally dealt with more than two states (the notable exception being the work of Slaughter *et al.*[249]). This may be due to the overlap of the broad FRET distributions, for which, despite intense efforts[116,121,229–242], no conclusive explanation has been found. This overlap obscures different populations and the finer structure of the histogram in general.

FRET studies were performed on freely diffusing RNase H molecules at pH 7.2. The application of ALEX and PIE illumination and detection schemes allowed the accumulation of donor-only free smFRET data with excellent statistics. However, even with good statistics, a straightforward three-component fit only yielded ambiguous results, a situation that could only be remedied by imposing physically sensible global restrictions on the means and widths of the F and the U state, as well as coupling the fractional populations of all states by linear free energy relationships. With these carefully chosen measures in place, however, not only is the evidence for I clear, the intermediate can also be observed and characterized. The approach of coupling the fractional populations in one global fit yielded free energy differences between the populations F, I and U over the whole spectrum of GdmCl concentrations, which fit well into the picture established by previous experiments[112,113,195,211,246]. The direct observation of an intermediate with a single molecule technique such as smFRET validated the preceding bulk studies that hinted at the existence of an intermediate state. It provides direct proof of the existence of the folding intermediate of RNase H on a fundamental level.

To make an even stronger case for the existence of the intermediate state, further smFRET experiments were performed at pH 3, where the intermediate state is expected to be more stable than at pH 7.2. However, the pH 3.0 histograms revealed heavy overlap between the distributions of all three conformations, making a straightforward fit impossible. Trp fluorescence was introduced as an additional reporter of the protein conformation. The distinct data sets of smFRET data and Trp fluorescence were joined via the global coupling of the fractional populations and the fit was done globally to both data sets simultaneously. While a two-state model did accommodate the smFRET data or the Trp fluorescence data reasonably well individually, the attempt to fit the combined data sets simultaneously utterly failed. This failure further underlines the necessity of an intermediate state for the interpretation of folding experiments of RNase H.

This, to my knowledge, novel technique of using ensemble Trp fluorescence data to aid in the interpretation of a single molecule experiment may be widely applicable for smFRET studies. The principle of coupling ensemble and single-molecule data in one global analysis may well be extended beyond just Trp fluorescence to include other ensemble techniques such as circular dichroism spectroscopy, IR spectroscopy or NMR methods. With the ever increasing number of single-molecule experiments, bolstering the ability to distinguish between coexistent conformations in these kinds of experiments will hopefully find applications in future studies.

Bibliography

[1] Feynman, R.P. (1960) There's Plenty of Room at the Bottom. *Eng. Sci.*, **23**, 22 – 36.

[2] Alberts, B. (1998) The cell as a collection overview of protein machines: Preparing the next generation of molecular biologists. *Cell*, **92**, 291–294.

[3] Lodish, H., Berk, A., Matsudaira, P., Kaiser, C.A., Krieger, M., Scott, M.P., Zipurksy, S.L., and Darnell, J. (2004) *Molecular Cell Biology (5th ed.).* New York, New York: WH Freeman and Company.

[4] Anfinsen, C.B. et al. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223–230.

[5] Levinthal, C. (1968) Are there pathways of folding. *J. Chim. Phys.*, **65**, 44–45.

[6] Zwanzig, R., Szabo, A., and Bagchi, B. (1992) Levinthal's paradox. *Proc. Natl. Acad. Sci. USA*, **89**, 20–22.

[7] Levinthal, C. (1969) How to fold graciously. *Mössbauer Spectrosc. Biol. Syst.*, **67**, 22–24.

[8] Rooman, M., Dehouck, Y., Kwasigroch, J.M., Biot, C., and Gilis, D. (2002) What is paradoxical about Levinthal paradox? *J. Biomol. Struct. Dyn.*, **20**, 327–329.

[9] Kim, P.S. and Baldwin, R.L. (1990) Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.*, **59**, 631–660.

[10] Dawkins, R. (1987) *The Blind Watchmaker*. Norton, New York.

[11] Ikai, A. and Tanford, C. (1971) Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature*, **230**, 100–102.

[12] Tsong, T.Y., Baldwin, R.L., and Elson, E.L. (1971) The sequential unfolding of ribonuclease A: Detection of a fast initial phase in the kinetics of unfolding. *Proc. Natl. Acad. Sci. USA*, **68**, 2712–2715.

[13] Englander, S.W. and Poulsen, A. (1969) Hydrogen-tritium exchange of the random chain polypeptide. *Biopolymers*, **7**, 379–393.

[14] Woodward, C.K. and Rosenberg, A. (1971) Studies of hydrogen exchange in proteins. *J. Biol. Chem.*, **246**, 4105–4121.

[15] Englander, S.W., Downer, N.W., and Teitelbaum, H. (1972) Hydrogen exchange. *Annu. Rev. Biochem.*, **41**, 903–924.

[16] Woodward, C.K. and Hilton, B.D. (1979) Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu. Rev. Biophys. Bioeng.*, **8**, 99–127.

[17] Bai, Y., Sosnick, T.R., Mayne, L., and Englander, S.W. (1995) Protein folding intermediates: Native-state hydrogen exchange. *Science*, **269**, 192–197.

[18] Bai, Y. and Englander, S.W. (1996) Future directions in folding: The multi-state nature of protein structure. *Proteins: Struct., Funct., Genet.*, **24**, 145–151.

[19] Udgaonkar, J.B. and Baldwin, R.L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature*, **335**, 694–699.

[20] Roder, H., Elöve, G.A., and Englander, S.W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature*, **335**, 700–704.

[21] Radford, S.E., Dobson, C.M., and Evans, P.A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, **358**, 302–307.

[22] Miranker, A., Robinson, C.V., Radford, S.E., Aplin, R.T., and Dobson, C.M. (1993) Detection of transient protein folding populations by mass spectrometry. *Science*, **262**, 896–900.

[23] Miranker, A., Robinson, C.V., Radford, S.E., and Dobson, C.M. (1996) Investigation of protein folding by mass spectrometry. *FASEB J.*, **10**, 93–101.

[24] Radford, S.E. and Dobson, C.M. (1995) Insights into protein folding using physical techniques: studies of lysozyme and α-lactalbumin. *Phil. Trans. R. Soc. Lond.*, **348**, 17–25.

[25] Chen, B.L., Baase, W.A., and Schellman, J.A. (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry*, **28**, 691–699.

[26] Fersht, A.R. (1995) Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.*, **5**, 79–84.

[27] Matouschek, A., Kellis, J.T., Serrano, L., Bycroft, M., and Fersht, A.R. (1990) Transient folding intermediates characterized by protein engineering. *Nature*, **346**, 440–445.

[28] Neira, J.L., Davis, B., Ladurner, A.G., Buckle, A.M., de Prat Gay, G., and Fersht, A.R. (1996) Towards the complete structural characterization of a protein folding pathway: The structures of the denatured, transition and native states for the association/folding of two complementary fragments of cleaved chymotrypsin inhibitor 2. Direct evidence for a nucleation-condensation mechanism. *Folding Des.*, **1**, 189–208.

[29] Jones, C.M., Henry, E.R., Hu, Y., Chan, C.K., Luck, S.D., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W.A. (1993) Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. USA*, **90**, 11860–11864.

[30] Pascher, T., Chesick, J.P., Winkler, J.R., and Gray, H.B. (1996) Protein folding triggered by electron transfer. *Science*, **271**, 1558–1560.

[31] Williams, S., Causgrove, T.P., Gilmanshin, R., Fang, K.S., Callender, R.H., Woodruff, W.H., and Dyer, R.B. (1996) Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry*, **35**, 691–697.

[32] Jackson, S.E. and Fersht, A.R. (1991) Folding of chymotrypsin inhibitor 2. *Biochemistry*, **30**, 10428–10435.

[33] Sosnick, T.R., Mayne, L., Hiller, R., and Englander, S.W. (1994) The barriers in protein folding. *Nat. Struct. Mol. Biol.*, **1**, 149–156.

[34] Huang, G.S. and Oas, T.G. (1995) Submillisecond folding of monomeric lambda repressor. *Proc. Natl. Acad. Sci. USA*, **92**, 6878.

[35] Schindler, T., Herrler, M., Marahiel, M.A., and Schmid, F.X. (1995) Extremely rapid protein folding in the absence of intermediates. *Nat. Struct. Mol. Biol.*, **2**, 663–673.

[36] Sosnick, T.R., Shtilerman, M.D., Mayne, L., and Englander, S.W. (1997) Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. USA*, **94**, 8545–8550.

[37] Karplus, M. and Weaver, D.L. (1979) Diffusion–collision model for protein folding. *Biopolymers*, **18**, 1421–1437.

[38] Weaver, D.L. (1984) Alternative pathways in diffusion–collision controlled protein folding. *Biopolymers*, **23**, 675–694.

[39] Karplus, M. and Weaver, D.L. (1994) Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.*, **3**, 650–668.

[40] Gruebele, M. (1999) The fast protein folding problem. *Annu. Rev. Phys. Chem.*, **50**, 485–516.

[41] Nienhaus, G.U. (2004) Physik der Proteine. *Phys. J.*, **3**, 37–44.

[42] Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. (2008) The protein folding problem. *Annu. Rev. Biophys.*, **37**, 289–316.

[43] Plaxco, K.W., Simons, K.T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.*, **277**, 985–994.

[44] Xu, Y., Purkayastha, P., and Gai, F. (2006) Nanosecond folding dynamics of a three-stranded β-sheet. *J. Am. Chem. Soc.*, **128**, 15836–15842.

[45] Plaxco, K.W., Kim, T., Ruczinski, I., and Baker, D. (2000) Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*, **39**, 11177–11183.

[46] Baker, D. (2000) A surprising simplicity to protein folding. *Nature*, **405**, 39–42.

[47] Ivankov, D.N. and Finkelstein, A.V. (2004) Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. *Proc. Natl. Acad. Sci. USA*, **101**, 8942–8944.

[48] Huang, J.T., Cheng, J.P., and Chen, H. (2007) Secondary structure length as a determinant of folding rate of proteins with two-and three-state kinetics. *Proteins: Struct., Funct., Bioinf.*, **67**, 12–17.

[49] Gromiha, M.M. and Selvaraj, S. (2001) Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J. Mol. Biol.*, **310**, 27–32.

[50] Mirny, L. and Shakhnovich, E. (2001) Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 361–396.

[51] Zhou, H. and Zhou, Y. (2002) Folding rate prediction using total contact distance. *Biophys. J.*, **82**, 458–463.

[52] Dill, K.A. (1987) The stabilities of globular proteins. *Protein Eng.*, pp. 187–192.

[53] Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nature Struct. Biol.*, **4**, 10–19.

[54] Wolynes, P.G. and Eaton, W.A. (1999) The physics of protein folding. *Phys. World*, **12**, 39–44.

[55] Camacho, C.J. and Thirumalai, D. (1993) Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA*, **90**, 6369–6372.

[56] Abkevich, V.I., Gutin, A.M., and Shakhnovich, E.I. (1994) Free energy landscape for protein folding kinetics: intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.*, **101**, 6052–6062.

[57] Chan, H.S. and Dill, K.A. (1994) Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.*, **100**, 9238–9257.

[58] Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Genet.*, **21**, 167–195.

[59] Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995) Navigating the folding routes. *Science*, **267**, 1619–1620.

[60] Thirumalai, D. (1995) From minimal models to real proteins: time scales for protein folding kinetics. *J. Phys. I*, **5**, 1457–1467.

[61] Hill, T.L. (1976) Diffusion frequency factors in some simple examples of transition-state rate theory. *Proc. Natl. Acad. Sci. USA*, **73**, 679–683.

[62] Zhou, H.X. and Zwanzig, R. (1991) A rate process with an entropy barrier. *J. Chem. Phys.*, **94**, 6147–6152.

[63] Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991) The energy landscapes and motions of proteins. *Science*, **254**, 1598–1603.

[64] Go, N. (1983) Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.*, **12**, 183–210.

[65] Brooks, C.L., Gruebele, M., Onuchic, J.N., and Wolynes, P.G. (1998) Chemical physics of protein folding. *Proc. Natl. Acad. Sci. USA*, **95**, 11037–11038.

[66] Millett, I.S., Doniach, S., and Plaxco, K.W. (2002) Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv. Protein Chem.*, **62**, 241–262.

[67] Flory, P.J. (1953) *Principles of polymer chemistry*. Cornell Univ. Pr.

[68] Merchant, K.A., Best, R.B., Louis, J.M., Gopich, I.V., and Eaton, W.A. (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA*, **104**, 1528–1533.

[69] Kohn, J.E., et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA*, **101**, 12491–12496.

[70] 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.*, **5**, 148–155.

[71] Dyson, H.J. and Wright, P.E. (2004) Unfolded proteins and protein folding studied by NMR. *Chem. Rev.*, **104**, 3607–3622.

[72] Religa, T.L., Markson, J.S., Mayor, U., Freund, S.M.V., and Fersht, A.R. (2005) Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.

[73] Shortle, D. and Abeygunawardana, C. (1993) NMR analysis of the residual structure in the denatured state of an unusual mutant of staphylococcal nuclease. *Structure*, **1**, 121–134.

[74] Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J., and Dobson, C.M. (1996) Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem.*, **100**, 2661–2666.

[75] Blanco, F.J., Serrano, L., and Forman-Kay, J.D. (1998) High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J. Mol. Biol.*, **284**, 1153–1164.

[76] Mok, Y.K., Kay, C.M., Kay, L.E., and Forman-Kay, J. (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.*, **289**, 619–638.

[77] Neri, D., Billeter, M., Wider, G., and Wuthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science*, **257**, 1559–1563.

[78] Ropson, I.J. and Frieden, C. (1992) Dynamic NMR spectral analysis and protein folding: identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by 19F NMR. *Proc. Natl. Acad. Sci. USA*, **89**, 7222–7226.

[79] Saab-Rincon, G., Gualfetti, P.J., and Matthews, C.R. (1996) Mutagenic and thermodynamic analyses of residual structure in the α subunit of tryptophan synthase. *Biochemistry*, **35**, 1988–1994.

[80] Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science*, **295**, 1719–1722.

[81] Fitzkee, N.C. and Rose, G.D. (2004) Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA*, **101**, 12497–12502.

[82] Jha, A.K., Colubri, A., Freed, K.F., and Sosnick, T.R. (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA*, **102**, 13099–13104.

[83] Ding, F., Jha, R.K., and Dokholyan, N.V. (2005) Scaling behavior and structure of denatured proteins. *Structure*, **13**, 1047–1054.

[84] Barron, L.D., Blanch, E.W., and Hecht, L. (2002) Unfolded proteins studied by Raman optical activity. *Adv. Protein Chem.*, **62**, 51–90.

[85] Frieden, C., Chattopadhyay, K., and Elson, E.L. (2002) What fluorescence correlation spectroscopy can tell us about unfolded proteins. *Adv. Protein Chem.*, **62**, 91–109.

[86] Keiderling, T.A. and Xu, Q. (2002) Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. *Adv. Protein Chem.*, **62**, 111–161.

[87] Schuler, B. and Eaton, W.A. (2008) Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.*, **18**, 16–26.

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

[89] Sherman, E. and Haran, G. (2006) Coil–globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA*, **103**, 11539–11543.

[90] Laurence, T.A., Kong, X., Jäger, M., and Weiss, S. (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. USA*, **102**, 17348–17353.

[91] Hoffmann, A., et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA*, **104**, 105–110.

[92] Kelley, A.M., Michalet, X., and Weiss, S. (2001) Single-molecule spectroscopy comes of age. *Science*, **292**, 1671–1672.

[93] Zhuang, X. and Rief, M. (2003) Single-molecule folding. *Curr. Opin. Struct. Biol.*, **13**, 88–97.

[94] Schuler, B., Lipman, E.A., Steinbach, P.J., Kumke, M., and Eaton, W.A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA*, **102**, 2754–2759.

[95] Zlatanova, J. and van Holde, K. (2006) Single-molecule biology: what is it and how does it work? *Mol. Cell*, **24**, 317–329.

[96] Mehta, A.D., Rief, M., Spudich, J.A., Smith, D.A., and Simmons, R.M. (1999) Single-molecule biomechanics with optical methods. *Science*, **283**, 1689–1695.

[97] Leckband, D. and Prakasam, A. (2006) Mechanism and dynamics of cadherin adhesion. *Annu. Rev. Biomed. Eng.*, **8**, 259–287.

[98] Dill, J., Bustamante, C., and Marqusee, S. (2010) The Unfolding Behavior of RNase H Under Force. *Biophys. J.*, **98**, 617–617.

[99] Sotomayor, M. and Schulten, K. (2007) Single-molecule experiments in vitro and in silico. *Science*, **316**, 1144–1148.

[100] Strick, T., Allemand, J.F., Croquette, V., and Bensimon, D. (2000) Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol.*, **74**, 115–140.

[101] Zlatanova, J. and Leuba, S.H. (2003) Magnetic tweezers: a sensitive tool to study DNA and chromatin at the single-molecule level. *Biochem. Cell Biol.*, **81**, 151–159.

[102] Lakowicz, J.R. (2006) *Principles of fluorescence spectroscopy, Third Edition*. Springer Science+Business Media, New York.

[103] Weiss, S. (1999) Fluorescence spectroscopy of single biomolecules. *Science*, **283**, 1676–1683.

[104] Weiss, S. (1999) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Biophys. J.*, **76**, 409–413.

[105] Deniz, A.A., Laurence, T.A., Dahan, M., Chemla, D.S., Schultz, P.G., and Weiss, S. (2001) Ratiometric single-molecule studies of freely diffusing biomolecules. *Annu. Rev. Phys. Chem.*, **52**, 233–253.

[106] Ha, T. (2001) Single-molecule fluorescence resonance energy transfer. *Methods*, **25**, 78–86.

[107] Haustein, E. and Schwille, P. (2003) Ultrasensitive investigations of biological systems by fluorescence correlation spectroscopy. *Methods*, **29**, 153–166.

[108] Ha, T. (2004) Structural dynamics and processing of nucleic acids revealed by single-molecule spectroscopy. *Biochemistry*, **43**, 4055–4063.

[109] Haustein, E. and Schwille, P. (2004) Single-molecule spectroscopic methods. *Curr. Opin. Struct. Biol.*, **14**, 531–540.

[110] Peterman, E.J.G., Sosa, H., and Moerner, WE (2004) Single-molecule fluorescence spectroscopy and microscopy of biomolecular motors. *Annu. Rev. Phys. Chem.*, **55**, 79–96.

[111] Neuweiler, H. and Sauer, M. (2004) Using photoinduced charge transfer reactions to study conformational dynamics of biopolymers at the single-molecule level. *Curr. Pharm. Biotechnol.*, **5**, 285–298.

[112] Kuzmenkina, E.V., Heyes, C.D., and Nienhaus, G.U. (2005) Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA*, **102**, 15471–15476.

[113] Kuzmenkina, E.V., Heyes, C.D., and Nienhaus, G.U. (2006) Single-molecule FRET study of denaturant induced unfolding of RNase H. *J. Mol. Biol.*, **357**, 313–324.

[114] Nienhaus, G.U. (2006) Exploring Protein Structure and Dynamics under Denaturing Conditions by Single-Molecule FRET Analysis. *Macromol. Biosci.*, **6**, 907–922.

[115] Kobitski, A.Y., Hengesbach, M., Helm, M., and Nienhaus, G.U. (2008) Sculpting an RNA Conformational Energy Landscape by a Methyl Group Modification—A Single-Molecule FRET Study. *Angew. Chem., Int. Ed.*, **47**, 4326–4330.

[116] Gopich, I.V., Nettels, D., Schuler, B., and Szabo, A. (2009) Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.*, **131**, 095102–(1–5).

[117] Choi, U.B., Strop, P., Vrljic, M., Chu, S., Brunger, A.T., and Weninger, K.R. (2010) Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.*, **17**, 318–324.

[118] Santoso, Y., Joyce, C.M., Potapova, O., Le Reste, L., Hohlbein, J., Torella, J.P., Grindley, N.D.F., and Kapanidis, A.N. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. USA*, **107**, 715–720.

[119] Kleanthous, C. (2000) *Protein-protein recognition*, vol. 31. Oxford University Press, USA.

[120] Archakov, A.I., Govorun, V.M., Dubanov, A.V., Ivanov, Y.D., Veselovsky, A.V., Lewi, P., and Janssen, P. (2003) Protein-protein interactions as a target for drugs in proteomics. *Proteomics*, **3**, 380–391.

[121] Chung, H.S., Louis, J.M., and Eaton, W.A. (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA*, **106**, 11837–11844.

[122] Chung, H.S., Gopich, I.V., McHale, K., Louis, J.M., and Eaton, W.A. (2011) Photon-By-Photon Analysis of Single Molecule Fluorescence Trajectories Determines an Upper Bound for the Transition Path Time in Protein Folding. *Biophys. J.*, **100**, 349a.

[123] Kanaya, S. and Crouch, R.J. (1983) DNA sequence of the gene coding for Escherichia coli ribonuclease H. *J. Biol. Chem.*, **258**, 1276–1281.

[124] Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1992) Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution. *J. Mol. Biol.*, **223**, 1029–1052.

[125] Berkower, I., Leis, J., and Hurwitz, J. (1973) Isolation and characterization of an endonuclease from Escherichia coli specific for ribonucleic acid in ribonucleic acid·deoxyribonucleic acid hybrid structures. *J. Biol. Chem.*, **248**, 5914–5921.

[126] Stein, H. and Hausen, P. (1969) Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. *Science*, **166**, 393–395.

[127] Hausen, P. and Stein, H. (1970) Ribonuclease H. *Eur. J. Biochem.*, **14**, 278–283.

[128] Kanaya, S. and Ikehara, M. (1995) Functions and structures of ribonuclease H enzymes. *Sub-Cell. Biochem.*, **24**, 377–422.

[129] Kanaya, S., Katsuda, C., Kimura, S., Nakai, T., Kitakuni, E., Nakamura, H., Katayanagi, K., Morikawa, K., and Ikehara, M. (1991) Stabilization of Escherichia coli ribonuclease H by introduction of an artificial disulfide bond. *J. Biol. Chem.*, **266**, 6038–6044.

[130] Dabora, J.M. and Marqusee, S. (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: Characterization of a partially folded state. *Protein Sci.*, **3**, 1401–1408.

[131] Chamberlain, A.K., Handel, T.M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Mol. Biol.*, **3**, 782–787.

[132] Jablonski, A. (1935) Über den Mechanismus der Photolumineszenz von Farbstoffphosphoren. *Z. Phys.*, **94**, 38–46.

[133] Franck, J. and Dymond, E.G. (1926) Elementary processes of photochemical reactions. *Trans. Faraday Soc.*, **21**, 536–542.

[134] Condon, E. (1926) A Theory of Intensity Distribution in Band Systems. *Phys. Rev.*, **28**, 1182–1201.

[135] Jaffe, H.H. and Miller, A.L. (1966) The fates of electronic excitation energy. *J. Chem. Educ.*, **43**, 469–473.

[136] Wikipedia (2012), Franck-Condon-Prinzip – Wikipedia, Die freie Enzyklopädie. [Online; Stand 26. Mai 2012], `http://de.wikipedia.org...` `.../w/index.php?title=Franck-Condon-Prinzip&oldid=101520430`.

[137] Bagchi, B., Oxtoby, D.W., and Fleming, G.R. (1984) Theory of the time development of the Stokes shift in polar media. *Chem. Phys.*, **86**, 257–267.

[138] Liptay, W. (1974) Dipole moments and polarizabilities of molecules in excited electronic states. *Excited states*, **1**, 129–229.

[139] Kautsky, H. (1939) Quenching of luminescence by oxygen. *Trans. Faraday Soc.*, **35**, 216–219.

[140] Potashnik, R., Goldschmidt, C.R., and Ottolenghi, M. (1971) Triplet state formation in the quenching of fluorescence by molecular oxygen. *Chem. Phys. Lett.*, **9**, 424–425.

[141] Eggeling, C., Widengren, J., Rigler, R., and Seidel, C.A.M. (1998) Photobleaching of fluorescent dyes under conditions used for single-molecule detection: evidence of two-step photolysis. *Anal. Chem.*, **70**, 2651–2659.

[142] Vogelsang, J., Kasper, R., Steinhauer, C., Person, B., Heilemann, M., Sauer, M., and Tinnefeld, P. (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed.*, **47**, 5465–5469.

[143] Panzer, O., Göhde, W., Fischer, U.C., Fuchs, H., and Müllen, K. (1998) Influence of oxygen on single molecule blinking. *Adv. Mater.*, **10**, 1469–1472.

[144] Zondervan, R., Kulzer, F., Kol'chenko, M.A., and Orrit, M. (2004) Photobleaching of rhodamine 6G in poly(vinyl alcohol) at the ensemble and single-molecule levels. *J. Phys. Chem. A*, **108**, 1657–1665.

[145] Lill, Y. and Hecht, B. (2004) Single dye molecules in an oxygen-depleted environment as photostable organic triggered single-photon sources. *Appl. Phys. Lett.*, **84**, 1665–1667.

[146] Hübner, C.G., Renn, A., Renge, I., and Wild, U.P. (2001) Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen. *J. Chem. Phys.*, **115**, 9619–9622.

[147] Renn, A., Seelig, J., and Sandoghdar, V. (2006) Oxygen-dependent photochemistry of fluorescent dyes studied at the single molecule level. *Mol. Phys.*, **104**, 409–414.

[148] Wilkinson, F., McGarvey, D.J., and Olea, A.F. (1994) Excited Triplet State Interactions with Molecular Oxygen: Influence of Charge Transfer on the Bimolecular Quenching Rate Constants and the Yields of Singlet Oxygen ($O_2^*, {}^1\delta_g$) for Substituted Naphthalenes in Various Solvents. *J. Phys. Chem.*, **98**, 3762–3769.

[149] Aitken, C.E., Marshall, R.A., and Puglisi, J.D. (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.*, **94**, 1826–1835.

[150] Cordes, T., Vogelsang, J., and Tinnefeld, P. (2009) On the mechanism of trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.*, **131**, 5018–5019.

[151] Weber, G. (1953) Rotational Brownian motion and polarization of the fluorescence of solutions. *Adv. Protein Chem.*, **8**, 415–459.

[152] Bowman, R.L., Caulfield, P.A., and Udenfriend, S. (1955) Spectrophotofluorometric assay in the visible and ultraviolet. *Science*, **122**, 32–33.

[153] Duggan, D.E. and Udenfriend, S. (1956) The spectrophotofluorometric determination of tryptophan in plasma and of tryptophan and tyrosine in protein hydrolysates. *J. Biol. Chem.*, **223**, 313–319.

[154] Shore, V.G. and Pardee, A.B. (1956) Fluorescence of some proteins, nucleic acids and related compounds. *Arch. Biochem. Biophys.*, **60**, 100–107.

[155] Vladimirov, I.A. and Konev, S.V. (1959) On possible mechanisms of energy migration in the protein molecule. *Biofizika*, **4**, 533–540.

[156] Du, H., Fuh, R.C.A., Li, J., Corkan, L.A., and Lindsey, J.S. (1998) PhotochemCAD: A Computer-Aided Design and Research Tool in Photochemistry. *Photochem. Photobiol.*, **68**, 141–142.

[157] Vivian, J.T. and Callis, P.R. (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.*, **80**, 2093–2109.

[158] Van Duuren, B.L. (1961) Solvent effects in the fluorescence of indole and substituted indoles. *J. Org. Chem.*, **26**, 2954–2960.

[159] Cowgill, R.W. (1967) Fluorescence and protein structure: Reappraisal of solvent and structural effects. *Biochim. Biophys. Acta*, **133**, 6–18.

[160] Kirby, E.P. and Steiner, R.F. (1970) Influence of solvent and temperature upon the fluorescence of indole derivatives. *J. Phys. Chem.*, **74**, 4480–4490.

[161] Chuang, T.J. and Eisenthal, K.B. (1972) Theory of fluorescence depolarization by anisotropic rotational diffusion. *J. Chem. Phys.*, **57**, 5094–5097.

[162] Szabo, A. (1984) Theory of fluorescence depolarization in macromolecules and membranes. *J. Chem. Phys.*, **81**, 150–167.

[163] Munro, I., Pecht, I., and Stryer, L. (1979) Subnanosecond motions of tryptophan residues in proteins. *Proc. Natl. Acad. Sci. USA*, **76**, 56–60.

[164] Gudgin, E., Lopez-Delgado, R., and Ware, W.R. (1981) The tryptophan fluorescence lifetime puzzle. A study of decay times in aqueous solution as a function of pH and buffer composition. *Can. J. Chem*, **59**, 1037–1044.

[165] Gudgin, E., Lopez-Delgado, R., and Ware, W.R. (1983) Photophysics of tryptophan in water, deuterium oxide and in nonaqueous solvents. *J. Phys. Chem.*, **87**, 1559–1565.

[166] Andrews, L.J. and Forster, L.S. (1972) Protein difference spectra. Effect of solvent and charge on tryptophan. *Biochemistry*, **11**, 1875–1879.

[167] Beechem, J.M. and Brand, L. (1985) Time-resolved fluorescence of proteins. *Annu. Rev. Biochem.*, **54**, 43–71.

[168] Eftink, M.R. (1991) *Fluorescence techniques for studying protein structure*. Wiley Online Library.

[169] Dexter, D.L. (1953) A theory of sensitized luminescence in solids. *J. Chem. Phys.*, **21**, 836–850.

[170] Andrews, D.L. (1989) A unified theory of radiative and radiationless molecular energy transfer. *Chem. Phys.*, **135**, 195–201.

[171] Förster, T. (1948) Intermolecular energy migration and fluorescence. *Ann. Phys.*, **2**, 55–75.

[172] Andrews, D.L. and Bradshaw, D.S. (2004) Virtual photons, dipole fields and energy transfer: a quantum electrodynamical approach. *Eur. J. Phys.*, **25**, 845–858.

[173] Förster, T. (1949) Experimentelle und theoretische Untersuchung des zwischenmolekularen Übergangs von Elektronenanregungsenergie. *Z. Naturforsch. A*, **4**, 321.

[174] Stryer, L. and Haugland, R. P. (1967) Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. USA*, **58**, 719–726.

[175] dos Remedios, C.G. and Moens, P.D.J. (1995) Fluorescence Resonance Energy Transfer Spectroscopy Is a Reliable "Ruler" for Measuring Structural Changes in Proteins Dispelling the Problem of the Unknown Orientation Factor. *J. Struct. Biol*, **115**, 175–185.

[176] Van der Meer, B.W. (2002) Kappa-squared: from nuisance to new sense. *Rev. Mol. Biotechnol.*, **82**, 181–196.

[177] VanBeek, D.B., Zwier, M.C., Shorb, J.M., and Krueger, B.P. (2007) Fretting about FRET: correlation between κ and R. *Biophys. J.*, **92**, 4168–4178.

[178] Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Müller, S., and Seidel, C.A.M. (2011) Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.*, **133**, 2463–2480.

[179] Marmé, N., Knemeyer, J.P., Sauer, M., and Wolfrum, J. (2003) Inter-and intramolecular fluorescence quenching of organic dyes by tryptophan. *Bioconjugate Chem.*, **14**, 1133–1139.

[180] Dale, R.E., Eisinger, J., and Blumberg, W.E. (1979) The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys. J.*, **26**, 161–193.

[181] Deniz, A.A., Laurence, T.A., Beligere, G.S., Dahan, M., Martin, A.B., Chemla, D.S., Dawson, P.E., Schultz, P.G., and Weiss, S. (2000) Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA*, **97**, 5179–5184.

[182] Rothwell, P.J., Berger, S., Kensch, O., Felekyan, S., Antonik, M., Wöhrl, B.M., Restle, T., Goody, R.S., and Seidel, C.A.M. (2003) Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. *Proc. Natl. Acad. Sci. USA*, **100**, 1655–1660.

[183] Johnson, C., Slaughter, B., Unruh, J., and Price, E. (2006) Fluorescence Probes of Protein Dynamics and Conformations in Freely Diffusing Molecules. *Rev. Fluoresc.*, **2006**, 239–259.

[184] Khan, Y.R., Dykstra, T.E., and Scholes, G.D. (2008) Exploring the Förster limit in a small FRET pair. *Chem. Phys. Lett.*, **461**, 305–309.

[185] Muñoz-Losa, A., Curutchet, C., Krueger, B.P., Hartsell, L.R., and Mennucci, B. (2009) Fretting about FRET: failure of the ideal dipole approximation. *Biophys. J.*, **96**, 4779–4788.

[186] Knox, R.S. and van Amerongen, H. (2002) Refractive index dependence of the Förster resonance excitation transfer rate. *J. Phys. Chem. B*, **106**, 5289–5293.

[187] Vörös, J. (2004) The density and refractive index of adsorbing protein layers. *Biophys. J.*, **87**, 553–561.

[188] Kuzmenkina, E. (2005) *Protein Folding Dynamics – Single-Molecule Studies of Ribonuclease HI on Biocompatible Surfaces*. Ph.D. thesis, Universität Ulm.

[189] Minsky, M. (1961) Microscopy Apparatus. *US Patent Nr.: US3013467 (A)*.

[190] Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E., and Weiss, S. (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. USA*, **101**, 8936–8941.

[191] Lee, N.K., Kapanidis, A.N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.*, **88**, 2939–2953.

[192] Müller, B.K., Zaychikov, E., Bräuchle, C., and Lamb, D.C. (2005) Pulsed interleaved excitation. *Biophys. J.*, **89**, 3508–3522.

[193] Uffink, J. (2008) Boltzmann's Work in Statistical Physics. Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy*, winter 2008 edn.

[194] Raschke, T.M. and Marqusee, S. (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat. Struct. Mol. Biol.*, **4**, 298–304.

[195] Raschke, T.M., Kho, J., and Marqusee, S. (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nature Struct. Biol.*, **6**, 825–830.

[196] Cecconi, C., Shank, E.A., Bustamante, C., and Marqusee, S. (2005) Direct observation of the three-state folding of a single protein molecule. *Science*, **309**, 2057–2060.

[197] Spudich, G.M., Miller, E.J., and Marqusee, S. (2004) Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J. Mol. Biol.*, **335**, 609–618.

[198] Connell, K.B., Horner, G.A., and Marqusee, S. (2009) A Single Mutation at Residue 25 Populates the Folding Intermediate of E. coli RNase H and Reveals a Highly Dynamic Partially Folded Ensemble. *J. Mol. Biol.*, **391**, 461–470.

[199] Robic, S., Berger, J.M., and Marqusee, S. (2002) Contributions of folding cores to the thermostabilities of two ribonucleases H. *Protein Sci.*, **11**, 381–389.

[200] Nishimura, C., Riley, R., Eastman, P., and Fink, A.L. (2000) Fluorescence energy transfer indicates similar transient and equilibrium intermediates in staphylococcal nuclease folding. *J. Mol. Biol.*, **299**, 1133–1146.

[201] Su, Z.D., Arooz, M.T., Chen, H.M., Gross, C.J., and Tsong, T.Y. (1996) Least activation path for protein folding: investigation of staphylococcal nuclease folding by stopped-flow circular dichroism. *Proc. Natl. Acad. Sci. USA*, **93**, 2539–2544.

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

[203] Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I., and Fujisawa, T. (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. *Proc. Natl. Acad. Sci. USA*, **101**, 1171–1176.

[204] Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y., and Fujisawa, T. (2002) Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA*, **99**, 1329–1334.

[205] Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins*, **6**, 87–103.

[206] Ptitsyn, O. (1996) How molten is the molten globule? *Nat. Struct. Mol. Biol.*, **3**, 488–490.

[207] Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., and Gaub, H.E. (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, **276**, 1109–1112.

[208] Oberhauser, A.F., Marszalek, P.E., Erickson, H.P., and Fernandez, J.M. (1998) The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, **393**, 181–185.

[209] Garcia-Mira, M.M. and Sanchez-Ruiz, J.M. (2001) pH corrections and protein ionization in water/guanidinium chloride. *Biophys. J.*, **81**, 3489–3502.

[210] Nozaki, Y. (1972) The preparation of guanidine hydrochloride. *Methods Enzymol.*, **26**, 43–50.

[211] Amirgoulova, E.V., Groll, J., Heyes, C.D., Ameringer, T., Röcker, C., Möller, M., and Nienhaus, G.U. (2004) Biofunctionalized polymer surfaces exhibiting minimal interaction towards immobilized proteins. *ChemPhysChem*, **5**, 552–555.

[212] Heyes, C.D., Groll, J., Möller, M., and Nienhaus, G.U. (2007) Synthesis, patterning and applications of star-shaped poly (ethylene glycol) biofunctionalized surfaces. *Mol. BioSyst.*, **3**, 419–430.

[213] Mills, A. (1997) Optical oxygen sensors. *Platinum Met. Rev.*, **41**, 115–126.

[214] Williams, A.T.R., Winfield, S.A., and Miller, J.N. (1983) Relative fluorescence quantum yields using a computer-controlled luminescence spectrometer. *Analyst*, **108**, 1067–1071.

[215] Levenberg, K. (1944) A method for the solution of certain nonlinear problems in least squares. *Quart. Appl. Math*, **2**, 164–168.

[216] Marquardt, D.W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, **11**, 431–441.

[217] Conn, A.R., Gould, N.I.M., and Toint, P.L. (2000) *Trust-Region Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

[218] Bates, D.M. and Watts, D.G. (2007) *Nonlinear regression analysis and its applications*. Wiley series in probability and statistics. Probability and statistics section, John Wiley & Sons.

[219] Runge, C. (1901) Über empirische Funktionen und die Interpolation zwischen äquidistanten 0rdinaten. *Z. Math. Phys.*, **46**, 20.

[220] De Boor, C. (2001) *A practical guide to splines*, vol. 27. Springer Verlag.

[221] Tanford, C. (1970) Protein Denaturation Part C: Theoretical Models for The Mechanism of Denaturation. *Adv. Protein Chem.*, **24**, 1–95.

[222] Staniforth, R.A., Burston, S.G., Smith, C.J., Jackson, G.S., Badcoe, I.G., Atkinson, T., Holbrook, J.J., and Clarke, A.R. (1993) The energetics and cooperativity of protein folding: A simple experimental analysis based upon the solvation of internal residues. *Biochemistry*, **32**, 3842–3851.

[223] Parker, M.J., Spencer, J., and Clarke, A.R. (1995) An Integrated Kinetic Analysis of Intermediates and Transition States in Protein Folding Reactions. *J. Mol. Biol.*, **253**, 771–786.

[224] Parker, M.J. and Clarke, A.R. (1997) Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry*, **36**, 5786–5794.

[225] Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, **4**, 2138–2148.

[226] Courtenay, E.S., Capp, M.W., Saecker, R.M., and Record Jr, M.T. (2000) Thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding: Interpretation of urea and guanidinium chloride m-values and their correlation with changes in accessible surface area (ASA) using preferential interaction coefficients and the local-bulk domain model. *Proteins Struct. Funct. Bioinf.*, **41**, 72–85.

[227] Tezuka-Kawakami, T., Gell, C., Brockwell, D.J., Radford, S.E., and Smith, D.A. (2006) Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys. J.*, **91**, L42–44.

[228] Mukhopadhyay, S., Krishnan, R., Lemke, E.A., Lindquist, S., and Deniz, A.A. (2007) A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc. Natl. Acad. Sci. USA*, **104**, 2649–2654.

[229] Gopich, I.V. and Szabo, A. (2005) Theory of photon statistics in single-molecule Förster resonance energy transfer. *J. Chem. Phys.*, **122**, 014707–(1–18).

[230] Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C.A.M. (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B*, **110**, 6970–6978.

[231] Nir, E., Michalet, X., Hamadani, K.M., Laurence, T.A., Neuhauser, D., Kovchegov, Y., and Weiss, S. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B*, **110**, 22103–22124.

[232] Gopich, I.V. and Szabo, A. (2007) Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B*, **111**, 12925–12932.

[233] Gopich, I.V. and Szabo, A. (2009) Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B*, **113**, 10965–10973.

[234] Kalinin, S., Valeri, A., Antonik, M., Felekyan, S., and Seidel, C.A.M. (2010) Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B*, **114**, 7983–7995.

[235] Nettels, D., Gopich, I.V., Hoffmann, A., and Schuler, B. (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA*, **104**, 2655–2660.

[236] Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W., and Kiefhaber, T. (2007) Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci. USA*, **104**, 2163–2168.

[237] Fierz, B. and Kiefhaber, T. (2007) End-to-end vs interior loop formation kinetics in unfolded polypeptide chains. *J. Am. Chem. Soc.*, **129**, 672–679.

[238] Wu, Y., Vadrevu, R., Kathuria, S., Yang, X., and Matthews, C.R. (2007) A Tightly Packed Hydrophobic Cluster Directs the Formation of an Off-pathway Sub-millisecond Folding Intermediate in the α Subunit of Tryptophan Synthase, a TIM Barrel Protein. *J. Mol. Biol.*, **366**, 1624–1638.

[239] Ratcliff, K. and Marqusee, S. (2010) Identification of Residual Structure in the Unfolded State of Ribonuclease H1 from the Moderately Thermophilic Chlorobium tepidum: Comparison with Thermophilic and Mesophilic Homologues. *Biochemistry*, **49**, 5167–5175.

[240] Chung, H.S., Louis, J.M., and Eaton, W.A. (2010) Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. *Biophys. J.*, **98**, 696–706.

[241] Chen, H., Rhoades, E., Butler, J.S., Loh, S.N., and Webb, W.W. (2007) Dynamics of equilibrium structural fluctuations of apomyoglobin measured by fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA*, **104**, 10459–10464.

[242] Doose, S., Neuweiler, H., and Sauer, M. (2009) Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *ChemPhysChem*, **10**, 1389–1398.

[243] Connell, K.B., Miller, E.J., and Marqusee, S. (2009) The Folding Trajectory of RNase H Is Dominated by Its Topology and Not Local Stability: A Protein Engineering Study of Variants that Fold via Two-State and Three-State Mechanisms. *J. Mol. Biol.*, **391**, 450–460.

[244] Behboodian, J. (1970) On the modes of a mixture of two normal distributions. *Technometrics*, **12**, 131–139.

[245] Rieger, R. and Nienhaus, G.U. (2012) A combined single-molecule FRET and tryptophan fuorescence study of RNase H folding under acidic conditions. *Chem. Phys.*, **396**, 3–9.

[246] Yamasaki, K., Saito, M., Oobatake, M., and Kanaya, S. (1995) Characterization of the internal motions of Escherichia coli ribonuclease HI by a combination of 15N-NMR relaxation analysis and molecular dynamics simulation: Examination of dynamic models. *Biochemistry*, **34**, 6587–6601.

[247] Rieger, R., Kobitski, A., Sielaff, H., and Nienhaus, G.U. (2011) Evidence of a Folding Intermediate in RNase H from Single-Molecule FRET Experiments. *ChemPhysChem*, **12**, 627–633.

[248] Nienhaus, K., Kriegl, J.M., and Nienhaus, G.U. (2004) Structural dynamics in the active site of murine neuroglobin and its effects on ligand binding. *J. Biol. Chem.*, **279**, 22944–22952.

[249] Slaughter, B.D., Allen, M.W., Unruh, J.R., Urbauer, R.J.B., and Johnson, C.K. (2004) Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution. *J. Phys. Chem. B*, **108**, 10388–10397.

List of Publications

Rieger R., Röcker C. and Nienhaus G.U. (2005) Fluctuation correlation spectroscopy for the advanced physics laboratory. *Am. J. Phys.*, **73**, 1129-1134

Rieger R., Kobitski A., Sielaff H. and Nienhaus G.U. (2011) Evidence of a Folding Intermediate in RNase H from Single-Molecule FRET Experiments. *ChemPhysChem*, **12**, 627-633

Rieger R. and Nienhaus G.U. (2011) A Combined Single-molecule FRET and Tryptophan Fluorescence Study of RNase H Folding under Acidic Conditions. *Chem. Phys.*, **396**, 3-9

Curriculum Vitae

Persönliche Daten

Name	Robert Frank Rieger
Geburtsdaten	13.10.1979 in Stuttgart – Bad Cannstatt
Nationalität	Deutsch

Ausbildung

06/2006 – 05/2011	Dissertation in Biophysik an der Universität Ulm und ab 10/2009 am Karlsruher Institut für Technologie (KIT)
10/1999 – 05/2006	Studium der Physik an der Universität Ulm, Diplomarbeit: "Schneller Flüssigkeitsmischer für kinetische Experimente an Biomolekülen mit hoher Zeitauflösung"
1990 – 1999	Gymnasium Donauwörth, mathematisch-naturwissenschaftlicher Zweig
1986 – 1990	Sebastian-Franck-Grundschule in Donauwörth

Danksagungen

An dieser Stelle möchte ich meinen Dank all denen gegenüber ausdrücken, die mir bei der Erstellung meiner Doktorarbeit hilfreich zur Seite standen.

An erster Stelle danke ich Herrn Prof. Dr. Nienhaus, der mir in seiner Arbeitsgruppe die Erstellung dieser Doktorarbeit erst ermöglichte. Mein physikalisches Verständnis machte in den Diskussionen mit ihm mehr als einmal einen großen Sprung vorwärts.

Danke auch meinem Zweitgutachter, Herrn apl. Prof. Dr. Wolfgang Wenzel, dass er sich die Zeit zur Begutachtung der Arbeit genommen hat.

Ich danke den Herren Dr. Jürgen Groll und Prof. Martin Möller, die mir das Star-PEG Material für meine Oberflächen-Beschichtungen zur Verfügung gestellt haben.

Beim Aufbau meiner Experimente standen mir Florian Junginger und Alexander Hepting stets mit Rat, und mehr als nur einmal auch mit Tat, zur Seite. Herzlichen Dank dafür.

Danke auch an Steffi Bacher, Uwe Theilen, Hendrik Sielaff, Susanne Matschulo und Michael Herrmann für technische Unterstützung in biochemischen Gefilden. Ich habe viel gelernt.

Speziell in den letzten Monaten der Arbeit war mir Karin Nienhaus beim Ordnen der Arbeit und dem Finden der richtigen Formulierungen eine große Hilfe, wofür ich mich herzlich bedanken will.

Die Diskussionen mit Charlie Röcker, Jochen Fuchs, Andrei Kobitski und Susi Gayda waren nicht nur fachlich sehr fruchtbar, sie gehören auch zu meinen liebsten Erinnerungen an meine Zeit in der Abteilung. Herzlichen Dank.

Das größte "Danke" geht ganz klar an meine Eltern und meine Partnerin Simone, die mich durch die Jahre nicht nur moralisch unterstützt haben. Ohne Euch hätte ich es nicht geschafft.