Advancing Fluorescence Fluctuation Microscopy in Living Cells:
From Non-stationary Signals to Ternary Protein Interactions

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Kwang Ho Hur

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Joachim D. Mueller, Advisor

October 2015

Acknowledgements

Dedication

This dissertation is dedicated to my parents Bong-Soo Hur and Si-Yeol Yang.

Abstract

Fluorescence fluctuation spectroscopy (FFS) is a powerful method for quantifying protein interactions. By exploiting the brightness of fluorescence intensity fluctuations we are able to measure the stoichiometry of protein complexes. FFS is particularly valuable because it allows real-time measurements within living cells, where protein complex formation plays a crucial role in the regulation of cellular processes. However, intensity fluctuations are frequently altered by the cell environment in subtle and unanticipated ways, which can lead to failure of the available FFS analysis methods. This thesis demonstrates that measuring in very small volumes, such as yeast and E. coli cells, can introduce a significant bias into the measured brightness as a result of cumulative sample loss, or photodepletion. This loss leads to a non-stationary signal, which is incompatible with the implicit assumption of a stationary process in conventional FFS theory. We addressed this issue by introducing a new analysis approach that serves as a foundation for extending FFS to non-stationary signals.

FFS measurements in cells are also currently limited to the study of binary interactions involving two different proteins. However, most cellular processes are mediated by protein complexes consisting of more than two different proteins. Observation of pairwise interactions is not sufficient to unequivocally determine the binding interactions involving three or more proteins. To address this issue, we extended FFS beyond binary interactions by developing tricolor heterospecies partition analysis to characterize ternary protein systems. The method is based on brightness analysis of fluorescence fluctuations from three fluorescent proteins that serve as protein labels. We verified tricolor heterospecies partition analysis by experiments on well-characterized protein systems and introduced a graphical representation to visualize interactions in ternary protein systems.

Table of Contents

List of Figures

List of Abbreviations

2DG	Two-dimensional Gaussian
3DG	Three-Dimensional Gaussian
ACF	Autocorrelation Function
APD	Avalanche Photodiode
BiFC	Bimolecular Fluorescence Complementation
cgf	cumulant generating function
CoA	Coactivator
CoR	Corepressor
DAQ	Data Acquisition
DBD	DNA Binding Domain
EGFP	Enhanced Green Fluorescence Protein
$EGFP_2$	dimeric EGFP
EYFP	Enhanced Yellow Fluorescence Protein
FCA	Fluorescence Cumulant Analysis
FCS	Fluorescence Correlation Spectroscopy
FFS	Fluorescence Fluctuation Spectroscopy
FP	Fluorescent Protein
FRET	Fluorescence Resonance Energy Transfer
GL	Gaussian-Lorentzian
HSP	Heterospecies Partition
kcps	kilocounts per second
LB	Lysogeny Broth

LBD   Ligand Binding Domain

mgf   moment generating function

mGL   modified Gaussian-Lorentzian

MSQ   Mean Segmented Q-value

NR   Nuclear Receptor

NCOA2   Nuclear Receptor Coactivator 2

NTF2   Nuclear Transport Factor 2

OD   Optical Density

PBS   phosphate-buffered saline

PCH   Photon Counting Histogram

pdf   probability distribution function

PSF   Point Spread Function

RAR   Retinoic Acid Receptor

RXR   Retinoid X Receptor

SBA   Segmented Brightness Analysis

SNR   Signal-to-Noise Ratio

TIFCA   Time-Integrated Fluorescence Cumulant Analysis

TIF2   Transcription Intermediate Factor 2

TPE   Two-Photon Excitation

1. Introduction

Biological cells were first observed by Hooke and van Leeuwenhoek with the early microscopes available in the 1660s and 1670s. A technical breakthrough in lens making by van Leeuwenhoek increased the resolution power by more than an order of magnitude. This development was crucial for the direct visualization of protozoa and bacteria by van Leeuwenhoek. He was the first to observe motility of single-cellular microorganisms, such as ciliate and sperm cells that propelled themselves through the water. He argued that since motility is a quality of life the objects he observed are living organisms. The cardinal importance of cells for all forms of living matter was not fully recognized until cell theory was established in 1855. The three tenets of cell theory are: 1) All organisms are composed of one or more cells; 2) the cell is the basic unit of structure and organization in organisms; 3) all cells come from preexisting cells.

The original cell theory emphasized structural ideas and morphological aspects. Over time it was realized that focusing on cellular processes instead of structural aspects offered a more promising avenue to advance our knowledge of cells. Since every cellular process involves a complex cascade of interacting biomolecules that carry out biochemical reactions, studying the function of DNA, RNA, and proteins became a central aspect of cell studies. Proteins are sometimes referred to as the "executive branch" of the cell, because they are responsible for carrying out most cellular function.

By the 1960s cells were viewed by many scientists as a vast collection of freely diffusing biomolecules that carry out reactions by two-body collisions between molecules [1]. In other words, the problem of predicting cellular behavior reduced to a complex set

of coupled second-order differential equations describing reactions of the form $A + B \rightleftarrows AB$. Today we know that this view of cellular processes was far too simplistic. Proteins are not simply colliding randomly with each other, but rather assemble into complexes containing several proteins. It has been indicated that over 80% of proteins do not operate alone but in complexes. In fact, every major cellular process involves protein complexes, which in turn interact with other complexes to carry out their biological function.

Complex formation of proteins provides a very flexible mechanism for regulating cellular processes. We mention two examples to illustrate this point: (1) Protein-protein interactions can modulate the activity of protein. For example, dimerization of the protein Bax starts a process that leads to apoptotic cell death [2]. (2) The association of proteins into a complex allows for substrate channeling, where the product of an enzyme is directly delivered to the next enzyme without its release into the bulk solution. This provides a more efficient and rapid processing of a substrate than would be possible for non-interacting enzymes. For example, tryptophan synthase guides its substrate through a protein tunnel that connects two active sites residing on different protein subunits [3].

Since all major cellular activities rely on the action of protein assemblies, any molecular-level model of cellular behavior requires information about the formation, interaction, and function of protein complexes. Thus, the identification of protein complexes and their characterization are pivotal research areas and much activity has been directed towards these aims. The vast majority of this research is conducted in vitro and ranges from studying a few targeted proteins to the collection of data from thousands of proteins simultaneously.

However, in vitro studies of protein complexes face a number of challenges and obstacles. For example, in vitro experiments require the isolation of proteins from cells or de novo synthesis of proteins. Extracting a complex of two or more associated proteins from a cell is a significant perturbation. There is no guarantee that the complex will stay intact, or that the change in environment during the extraction process leads to the attachment of new proteins not found in the original complex. Many proteins are covalently modified by enzymes after translation. These post-translational modifications add new functional groups that are important for the function and the interaction of the protein. For example, many enzymes and receptors are switched "on" or "off" by the addition of a phosphate group. This change can be reversed by cleavage of the phosphate group. Reversible phosphorylation and dephosphorylation is controlled by specific classes of enzymes whose activity is regulated by cellular signals. Post-translational modifications are a challenge for in vitro studies, because faithful recreation of the intracellular environment with its signaling pathways is far from trivial. Similarly, in vitro experiments of dilute solutions fail to create the conditions found in cells, where high concentration of biomolecules significantly reduces the volume available to free solvent molecules. This effect, termed molecular crowding, can exert significant influence on the thermodynamics and kinetics of protein interactions. It has been estimated that the reaction rates and binding equilibria of protein interactions determined in a test tube under uncrowded conditions may differ by orders of magnitude from the values observed under crowded conditions that are found inside cells.

Measuring inside the living cell probes protein assemblies in their natural environment and avoids the disadvantages associated with in vitro experiments.

However, quantitative experiments inside living cells are far from trivial. We still lack a diverse set of robust tools to identify and quantify cellular protein complexes. Two-hybrid screening is the most widely used molecular biology technique for identifying protein-protein interactions in cells. However, the specificity of the screen is not very high. In addition, it is not a real-time technique, because the read-out is typically achieved by initiating the expression of a reporter gene. Fluorescence-based identification of protein complexes has become increasingly more popular and offers advantages over existing molecular biology techniques. Fluorescence provides a very high level of detection sensitivity as demonstrated by various single-molecule experiments. It provides real-time information from time-resolved measurements of cellular fluorescence. Fluorescence experiments are also non-invasive, because the spectroscopic read-out leaves the cell intact.

While a variety of fluorescence-based methods exist, fluorescence fluctuation spectroscopy (FFS) is of special interest, because it has been developed to quantify the composition of protein complexes inside the living cell [4–6]. Experiments with model systems have verified that FFS can identify the oligomeric state of a cellular protein. Although FFS has been applied widely to study protein assemblies in the live cell, it is not yet a mature technique. This thesis focuses on overcoming some of the weaknesses of FFS that still exist.

For example, the original formulation of FFS assumes a stationary signal and an infinite sample reservoir. While these assumptions are approximately fulfilled for some FFS experiments in cells, they are clearly violated in others. Specifically, small sample compartments lead to a non-stationary signal because of depletion of fluorescent

4

molecules by the exciting laser. Part of this thesis is dedicated to expand FFS to such small sample compartments. The thesis addresses one further shortcoming of FFS. Ternary protein complexes, which are assemblies that contain three different proteins, are currently beyond the reach of the FFS technique. Yet, the formation of complexes involving more than two proteins is critical for many cellular processes, including signal transduction and transcriptional control. In addition, the interaction of three proteins is a minimal requirement for the presence of cooperative binding. Work presented in this thesis expands FFS to ternary protein complexes and characterizes qualitatively the binding model between two nuclear receptors and their coactivator [7,8].

The studies presented here significantly expand the reach of intracellular applications of FFS and lay the foundation for far more sophisticated experiments that probe the interplay of proteins across different cellular compartments, a topic that will be further discussed in the final chapter of the thesis.

1.1 Fluorescence detections of cellular protein complexes

Visualizing proteins in living cells had been a major challenge until the discovery of a protein that emits green fluorescence when excited with violet light. This protein, green fluorescent protein (GFP), was first expressed in E. coli cells in 1994 [9]. Joining the GFP gene with the gene of a specific protein generates a fusion gene, which if translated in a cell results in the expression of the protein tagged with GFP. Tagging renders the protein visible and is key to observe active processes within living cells. Mutations of GFP created improved versions and many new color variants of the protein [10,11]. In addition, fluorescent proteins (FPs) from other species have been discovered, resulting in further expansion of fluorescent tags. All in all, the palette of FPs available

has proven, and continues to prove, immensely useful for cellular applications. In particular, the detection of protein complexes in living cells relies strongly on the availability of FPs as protein labels. We briefly describe the methods that have been most widely used to identify protein assemblies by fluorescence in living cells.

The most widely-used method for detecting protein-protein interactions is fluorescence resonance energy transfer (FRET) from a donor FP to an acceptor FP [12,13]. Absorption of light by the donor leads to a transition of the donor molecule into an electronically excited state. This excited state decays via emission of a light. However, if an acceptor is within close contact with the donor, the donor's excited state energy can be non-radiatively transferred to the acceptor through a dipole-dipole interaction. The FRET efficiency is inversely proportional to the sixth power of the distance between donor and acceptor, and the typical range for FRET is between 1 and 10 nm. Since proteins are nanometer-sized objects, measurement of FRET provides an excellent marker to detect their association. The FRET efficiency is also affected by other factors such as the spectral overlap between the donor's emission spectrum and the acceptor's excitation spectrum, and the relative orientation between the donor emission dipole moment and the acceptor absorption dipole moment.

While a positive FRET signal is a tell-tale sign identifying protein-protein interactions, the absence of FRET cannot rule out complex formation. For example, if the size of a protein complex exceeds the range of FRET, the interaction cannot be detected. In addition, if the donor dipole is perpendicular to the acceptor dipole, FRET does not occur. Since the FRET efficiency depends on distance and orientation, which are not known a priory, calculating the stoichiometry of the protein complex is not feasible.

Thus, FRET is primarily a tool for detecting the presence of protein-protein interactions in the living cell.

Another detection method is bimolecular fluorescence complementation (BiFC), which is based on the spontaneous association of two protein fragments, which leads to fluorescence [14–16]. The two fragments are derived from an FP and are individually non-fluorescent. Attaching these fragments to proteins that interact brings the two components into close proximity, which initiates the reconstitution of the intact fluorescent protein. Thus, detection of a fluorescent signal indicates that the labeled proteins interact. However, the BiFC complex is irreversible, which prevents the dynamic dissociation and association of protein complexes. In addition, the association of the two fragments is not sufficient for fluorescence, which is delayed until the chromophore is formed, a process known as maturation. This delay is ~1 hour, which prohibits the real-time detection of protein interactions by BiFC.

Protein association can also be characterized by fluorescence fluctuation spectroscopy (FFS). This method relies on the fluorescent intensity fluctuations generated by the passage of proteins tagged with FP through the small optical observation volume of a microscope. The most widely used fluctuation technique is fluorescence correlation spectroscopy (FCS), which measures the autocorrelation function (ACF) of the intensity fluctuations [17–22]. The ACF characterizes the temporal decay of the fluctuations and identifies the diffusion coefficient of mobile proteins. The ACF detects protein association by a decrease in the diffusion coefficient, because the complex has a larger size than the individual protein. Unfortunately, the change in the ACF caused by the association of two similarly sized proteins is too small to be easily detected, even under

in vitro conditions [23]. Another challenge encountered in cells is the weak interaction of proteins with other cellular components, which influences its mobility. Thus, it is essentially impossible to predict the expected change in the diffusion coefficient for a protein complex.

FFS also analyzes the amplitude distribution of the intensity fluctuations [4,24–27]. The average amplitude characterizes the mean photon count rate per fluorescent particles, which we also refer to as brightness. The association of two labeled proteins into a complex doubles the fluorescence per particle. Thus, brightness is proportional to the oligomeric state of the labeled protein. Brightness has proven to be an excellent marker of protein association inside living cells, and is the main technique used in this work. A detailed description of FFS is found in the next chapter.

Not all FPs are suited for brightness experiments. FPs with a single and robust brightness state are optimal. Some FPs have photophysical properties, such as low photostability, that make them unsuitable for FFS. Others that exist in more than one brightness state can be used provided their photophysics can be accurately modeled [28]. Procedures for evaluating the potential of FPs in FFS experiments have been described elsewhere. The work presented in this thesis relies heavily on the enhanced GFP (EGFP), which is an amazingly robust fluorescent protein. The fluorescence-generating chromophore is embedded inside the beta-barrel shaped protein and largely insensitive to changes in the local environment, provided the pH remains unchanged. The brightness of EGFP doesn't change when it is tagged to another protein, nor when it is measured in the nucleus, cytoplasm or aqueous solution [4,29]. It is this property that makes EGFP such a valuable tool for FFS, and our first choice of fluorophore when pushing FFS experiments

into new directions.

1.2 Ternary Protein complexes

Because proteins often contain multiple interaction sites the formation of ternary or higher-order protein complexes are ubiquitous features of protein interactions and critical for many cellular processes, including signal transduction, transcriptional control, and cytoskeleton remodeling [30–32]. Ternary interactions provide a crucial element for regulation of cellular processes, because the stability of the complex depends on the binding energy of multiple protein contacts, which can be modulated by ligand binding to the proteins. Energetically, the interactions of higher-order protein complexes cannot always be described simply by the additive effects of the individual binary reactions that make up the overall complex, because of cooperative interactions between separate protein domains. In fact, cooperative binding of biomolecules is a widely used strategy for regulating cellular processes [33,34].

Because measuring the interactions between three protein species is technically very challenging, most experiments have been limited to look at interactions between two protein species in cells [5,6,35,36]. While immensely useful, knowledge of interactions between protein pairs is insufficient to answer basic questions of protein interactions of ternary complexes. For example, consider three proteins A, B, and C. Assume that A and B compete for the binding partner C, AC ↔ BC. Labeling two of the three proteins can detect interactions, but it cannot establish the presence of the competition. Only the simultaneous observation of both AC and BC provides a quantitative characterization of the process. Thus, in order to study ternary protein complexes inside cells, we need a technique that can distinguish between three different protein species.

1.3 Nuclear receptor and coactivator

Eukaryotic cells possess a very sophisticated system to regulate the expression of their genes. A very important group of regulatory proteins involved in this regulation is the nuclear receptor (NR) family, which are ligand-activated transcription factors. This group of proteins plays a key role in controlling the development, homeostasis, and metabolism of the organism. Many NRs sense the presence of specific ligands (such as steroids) and in response regulate the expression of specific genes. Furthermore, it has been recognized that NRs must recruit coregulatory proteins to modulate transcription of their target genes [37,38]. Coregulatory proteins are classified into two groups: coactivator (CoA) and corepressor (CoR). Coactivators enhance transcription activation, while corepressors suppress it [39,40].

Figure 1.1 presents a basic model of the essential interactions involved in transcription control by NRs with coregulatory proteins. The model stipulates that in the absence of ligand the NR dimer is bound to a corepressor (Figure 1.1A). Binding of an activating ligand, called agonist, to NR induces a conformational change that leads to the release of the corepressor followed by the recruitment of a coactivator (Figure 1.1B) [41–43]. Subsequent interactions with other cofactors lead to the initiation of gene expression.

Most representations found in the literature show the NRs bound to their response element at the DNA [38]. But recent studies have shown that NRs are very dynamic and only interact with chromosomal DNA for a short time [44–48]. Cellular experiments have also shown that activation with a ligand often transactivates other gene expression pathways that are thought to be independently regulated [49–51]. The origin for the cross-talk between different gene regulation paths is currently not clear. In some cases
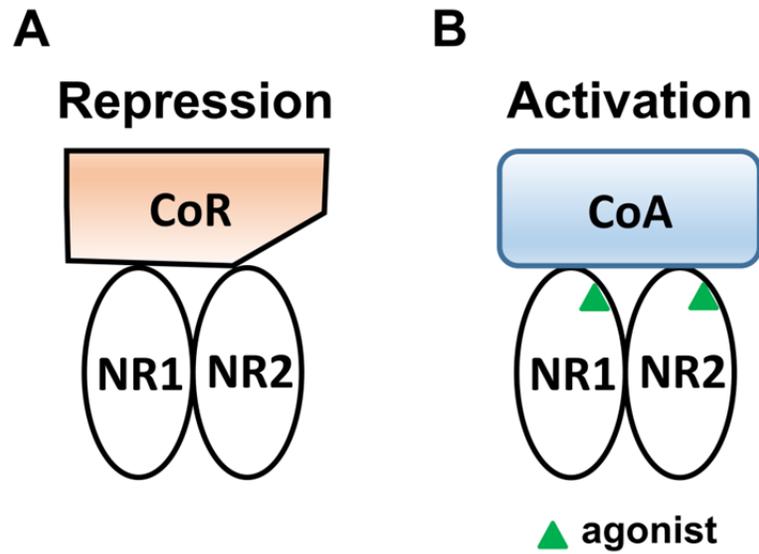
Figure 1.1 Gene regulation by NRs

(A) In the absence of agonist the NRs (NR1 & NR2) are believed to interact with a corepressor (CoR) and gene expression is suppressed. (B) Binding of agonist leads to dissociation of CoR and recruitment of a coactivator (CoA), which initiates gene activation.

evidence has been found that implies direct protein-protein interactions between transcription factor complexes, which could be responsible for cross-talk [52–54]. Another intriguing aspect of gene regulation is the presence of synergistic effects [55–57]. For example two ligands that activate the same nuclear receptor signaling pathway are sometimes much more potent together than individually. The molecular mechanism responsible for this effect is currently not established.

While it is relatively easy to assess the effect of gene regulation by monitoring the output (i.e. the expressed protein), it is very difficult to determine the protein interactions that initiate the whole process. Thus, while a number of models of gene regulation by NRs exist, quantitative knowledge of the protein interactions inside the cell is very limited. We address this challenge with tricolor brightness analysis in chapter 5 of this thesis by studying the binding of the NRs RAR (retinoic acid receptor) and RXR (retinoid X receptor) to a CoA.

NRs have a common structure that includes a ligand binding domain (LBD) and a DNA binding domain (DBD) (Figures 1.2A and B) [8]. The DBD has two zinc-finger motifs that mediate sequence-specific DNA recognition. The multi-functional LBD is the ligand binding site, dimerization site and co-regulator interaction surface. Binding of the agonist leads to a conformational change in the LBD domain, which facilitates CoA interaction. NRs function as dimers and interact with their response element at the DNA as a dimer. RAR requires RXR as a partner and forms a heterodimer [58,59]. RXR, on the other hand, can interact with itself and many other NRs [58–61].
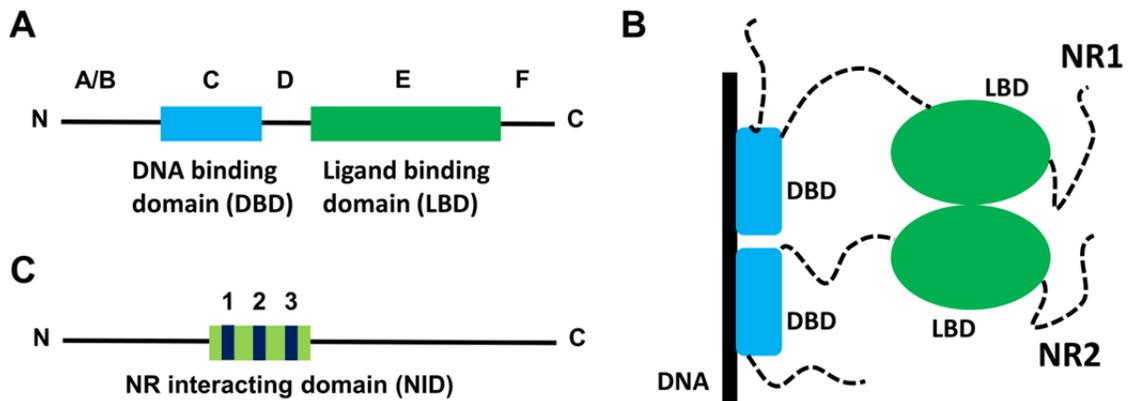
Figure 1.2 Schematic illustration of NR structure and binding of dimeric NRs to DNA and CoA

(A) Schematic NR construct. NR consists of five to six structural domains denoted by A to F. The DNA binding domain (DBD) is denoted by a blue box and the ligand binding domain (LBD) by a green box. (B) Two NRs interact through a dimerization surface in the LBD. The dimeric complex NR1-NR2 binds to DNA via the DBD domain. (C) CoA construct. The three NR boxes in the NR interacting domain (NID) are identified by 1, 2 and 3.

Nuclear receptor coactivator (NCOA2), also referred to as transcription intermediate factor 2 (TIF2), is a CoA that interacts with the nuclear receptor pair RAR-RXR [7,8]. The CoA contains three nuclear receptor binding domains (called NR boxes or LXXLL motifs) and has an intrinsic histone acetyltransferase activity (Figure 1.2C). Binding of NRs to the NR boxes brings the CoA in close proximity to the DNA. Subsequent local histone acetylation by CoA leads to chromatin decondensation, which provides the RNA polymerase access to the transcription site.

The ligand-dependent interactions between RAR-RXR and TIF2 are summarized in Figure 1.3. In the absence of ligand, the heterodimer RAR-RXR is bound to a corepressor (CoR) through an interaction between RAR and CoR (Figure 1.3A). An RAR-selective agonist induces an allosteric modification of RAR structure by which the heterodimer dissociates from the corepressor (Figure 1.3B). The dissociated heterodimer binds to CoA through interaction with RAR and one of the NR boxes of CoA (Figure 1.3C). The additional presence of an RXR-selective agonist leads to a conformational change of RXR LBD, which induces binding to a free NR box of CoA (Figure 1.3D). Note that the presence of only the RXR agonist is insufficient to induce dissociation of the heterodimer from the corepressor. In addition, there is no interaction between ligand-activated RXR and CoA, because the presence of CoR blocks access to the binding interface of RXR (Figure 1.3E). This effect is known as RXR subordination.

This system is an excellent candidate for tricolor FFS studies in cells. The three proteins RXR, RAR and TIF2 are soluble in the nucleus of mammalian cells, which avoids technical challenges caused by immobilization [62] and geometrical constraints encountered in the cytoplasm [63]. NR interaction with CoA is an important biological
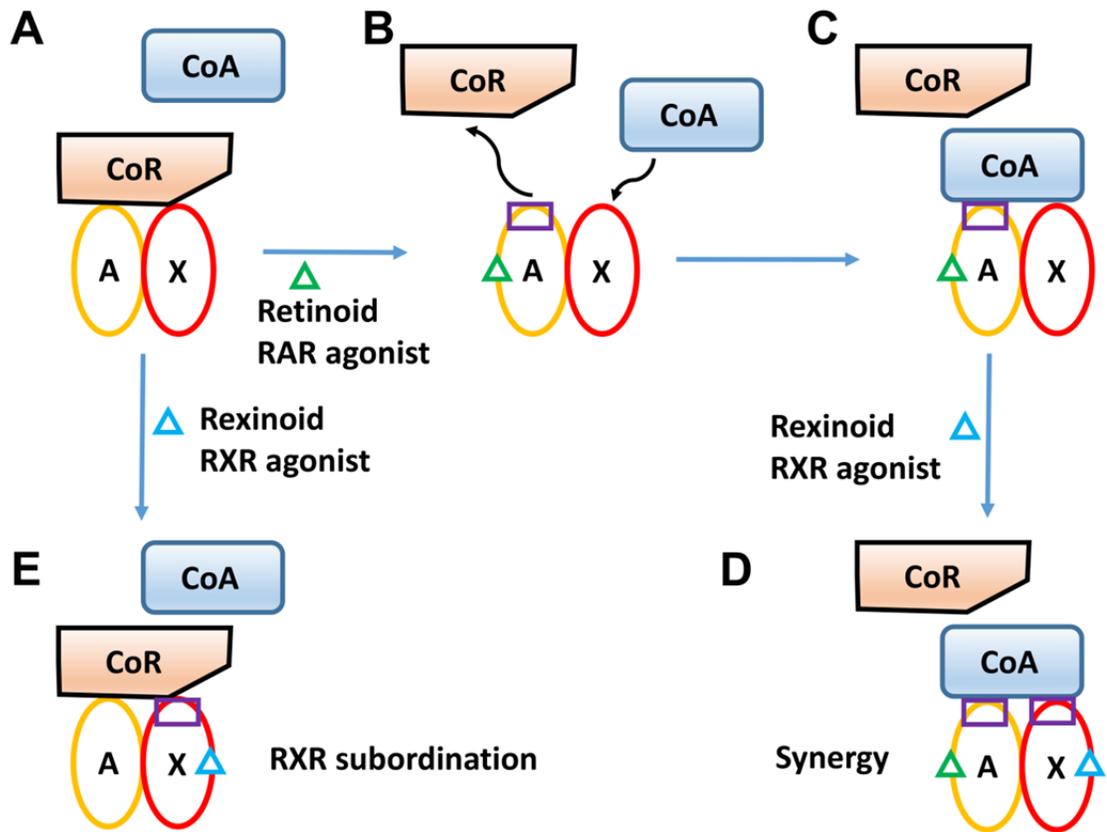
Figure 1.3 Schematic illustration of CoA and RAR-RXR interactions

The figure is adapted from [8]

system, because it represents the first step of gene activation of mammalian cells. Because of this fundamental role, NRs are linked to various diseases, including cancer, diabetes and obesity. While the behavior of NRs and CoA has been extensively studied, quantitative information about their physical interactions is rather limited. Tricolor FFS has the potential to fill in this gap and provide a starting point for detailed modeling of the interactions that regulate gene activation, which can be extended to include other NRs as well as their CoAs and CoRs.

1.4 Brief overview of thesis outline

The next chapter provides an introduction into FFS and describes the instrumentation, theory, and highlights a few selected analysis methods. Chapters 3 and 4 focus on the application of single-color FFS to yeast and bacterial cells. We choose the budding yeast *Saccharomyces cerevisiae* and *Escherichia coli* (E. coli) as model systems. The volume of these cells is significantly smaller than the cytoplasmic compartment of mammalian cells, which gives rise to new phenomena. We noticed during initial experiments in yeast an unusually large scatter in brightness values and traced it to the presence of photodepletion during the measurement process. We define photodepletion as the cumulative depletion of fluorescent molecules in a small reservoir due to photobleaching. This effect leads to a systematic decrease in the fluorescence intensity, which violates the assumption of a stationary signal as postulated by conventional FFS. We develop a model that takes the non-stationary signal into account and then apply it to demonstrate that quantitative and accurate FFS experiments in yeast cells are feasible.

Because the volume of E. coli cells is significantly smaller than for yeast, photodepletion is much more severe. We found that the analysis procedure used for yeast

cells fails under these conditions, and a new and more robust model had to be developed. In addition, the cylindrical shape of the E. coli cell needed to be accounted for in the FFS experiments to properly describe the data. This model was successfully applied to identify monomeric and dimeric proteins in E. coli cells.

In chapter 5, we turn towards the identification of ternary protein complexes in mammalian cells. Since the protein species are distinguished with fluorescent tags of different colors, we had to develop tricolor heterospecies partition (HSP) analysis and work out a procedure to identify relevant interactions from a mixture of protein complexes present within the cell. Experiments on model systems served to test tricolor HSP analysis. We next studied the interaction of the nuclear receptors RXR and RAR with their coactivator TIF2. We were specifically interested in demonstrating the presence of ternary protein complexes within the cell.

The work presented in chapter 3 and 4 has been published [64,65], while the material of Chapter 5 has been submitted for publication. The material presented in the appendices is unpublished.

2. Experimental Methods

This chapter presents background information on fluorescence fluctuation spectroscopy (FFS), the main experimental technique used in this work. We first introduce the concept of FFS and describe the instrumentation. The rest of the chapter focuses on the theoretical concepts behind FFS analysis of photon count data.

2.1 Introduction to FFS

Focusing a laser beam through a microscope objective creates a tight focal spot that excites fluorescent molecules along the beam path. In one-photon excitation, a pinhole before the detector is used to only pass fluorescence light generated in the near vicinity of the focal volume. In two-photon excitation, fluorescence is only generated in the focal region, because of the inherent z-sectioning effect of the non-linear excitation process [66]. Thus, the pinhole is only needed in one-photon excitation. Both excitation methods are routinely used to generate an excitation volume of the order of 0.2 fl. This volume is sufficiently small that the number of fluorescent molecules in the optical observation volume is small. For example, a solution with a concentration of 1 nM leads to average occupation number of ~0.1 molecules in the observation volume. Diffusion of the molecule in and out of the optical volume generates signal fluctuations in the fluorescence as schematically depicted in Figure 2.1A and B. We generally refer to the measurement and analysis of these fluorescence fluctuations as FFS.

Before discussing quantitative aspects of fluorescence intensity fluctuations, we want to start with a conceptual picture to illustrate some of the information embedded in the fluctuations.
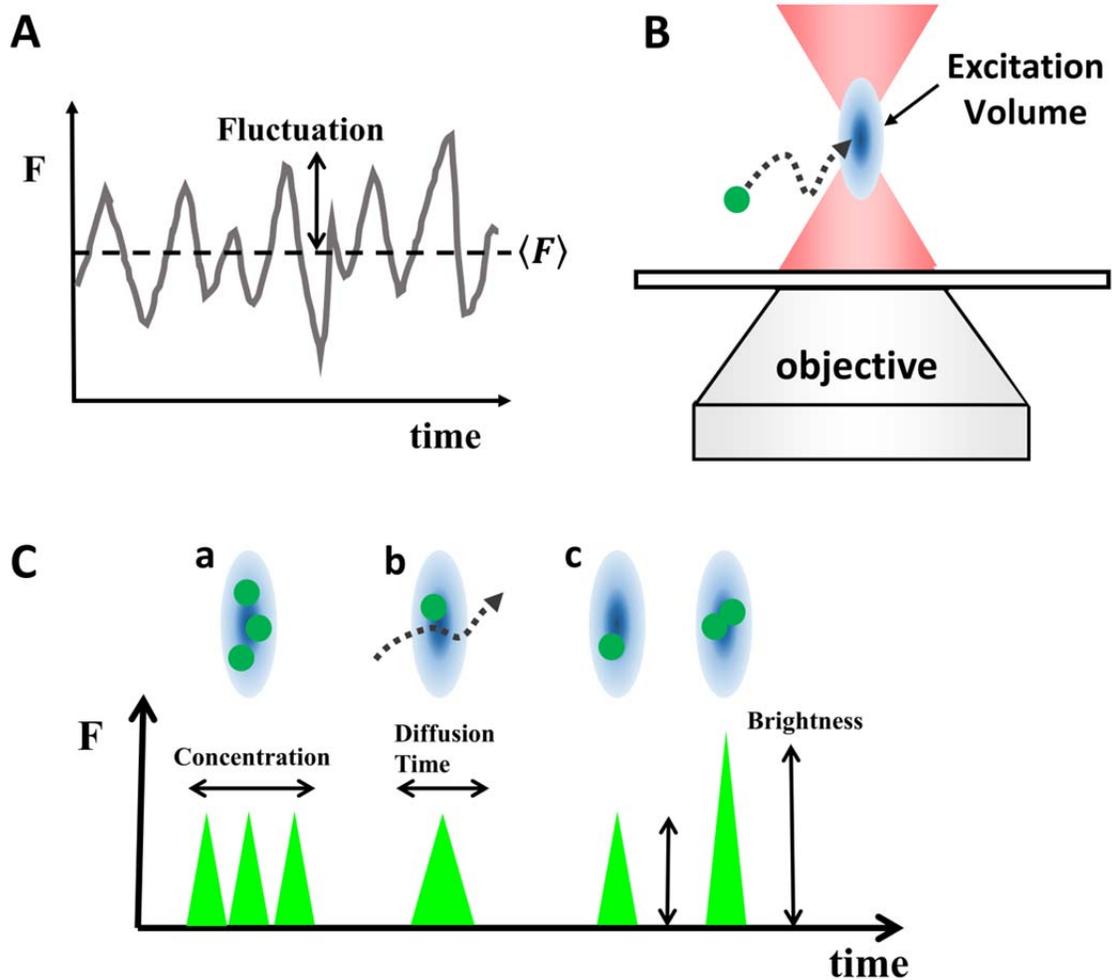
Figure 2.1 Schematic illustration of fluorescent fluctuations from diffusing fluorescent proteins.

(A) The fluorescence intensity fluctuates around an average value as fluorescent proteins move in and out of the excitation volume (B) The microscope optics generates a submicron excitation volume which the fluorescent protein diffuses (C) Conceptual representation of the information embedded in the intensity fluctuations. a) The frequency of the intensity bursts indicates the concentration of the sample. b) The duration of a burst characterizes the diffusion time of the particle. c) The amplitude of the bursts is a measure of the brightness of the diffusing particle. A dimeric particle has, on average, twice the brightness as a monomeric particle.

Consider a solution of fluorescent proteins at a low concentration. Each passage of the fluorescent protein through the volume generates a small burst of light. The burst frequency is proportional to the protein concentration, the burst duration identifies how much time is needed to diffuse through the observation volume, and the burst amplitude identifies the fluorescence brightness of the protein (Figure 2.1C). The brightness is of special interest, because association of fluorescent proteins results in a larger and brighter complex that can be detected by FFS. This point will be discussed in more detail once we have introduced quantitative FFS models.

2.2 Instrumentation

Fluorescence fluctuation experiments were carried out on an AxioObserver microscope (Zeiss, Thornwood, NY) modified for two-photon excitation with a Ti:Sapphire laser (Tsunami or MaiTai, Spectra Physics, Mountain View, CA). The experimental setup of the instrument is shown in Figure 2.2. The laser produces an 80 MHz train of short laser pulses (~100 fs) that are reflected by a dichroic mirror (#740DCSPXR, Chroma, Rockingham, VT) and focused onto the sample by an objective. We used either a 63x C-Apochromat water immersion objective lens (NA=1.2, Zeiss) or a 63x Plan Apochromat oil immersion objective lens (NA=1.4, Zeiss) for the experiments described in this thesis. Two-photon excitation (TPE) of fluorophores near the focus of the beam results in fluorescence emission. Part of the emitted light was collected by the objective before entering the detection side of the instrument by passing through a dichroic and a clean-up filter (E700SP, Chroma Technology) to remove residual laser light from the beam.
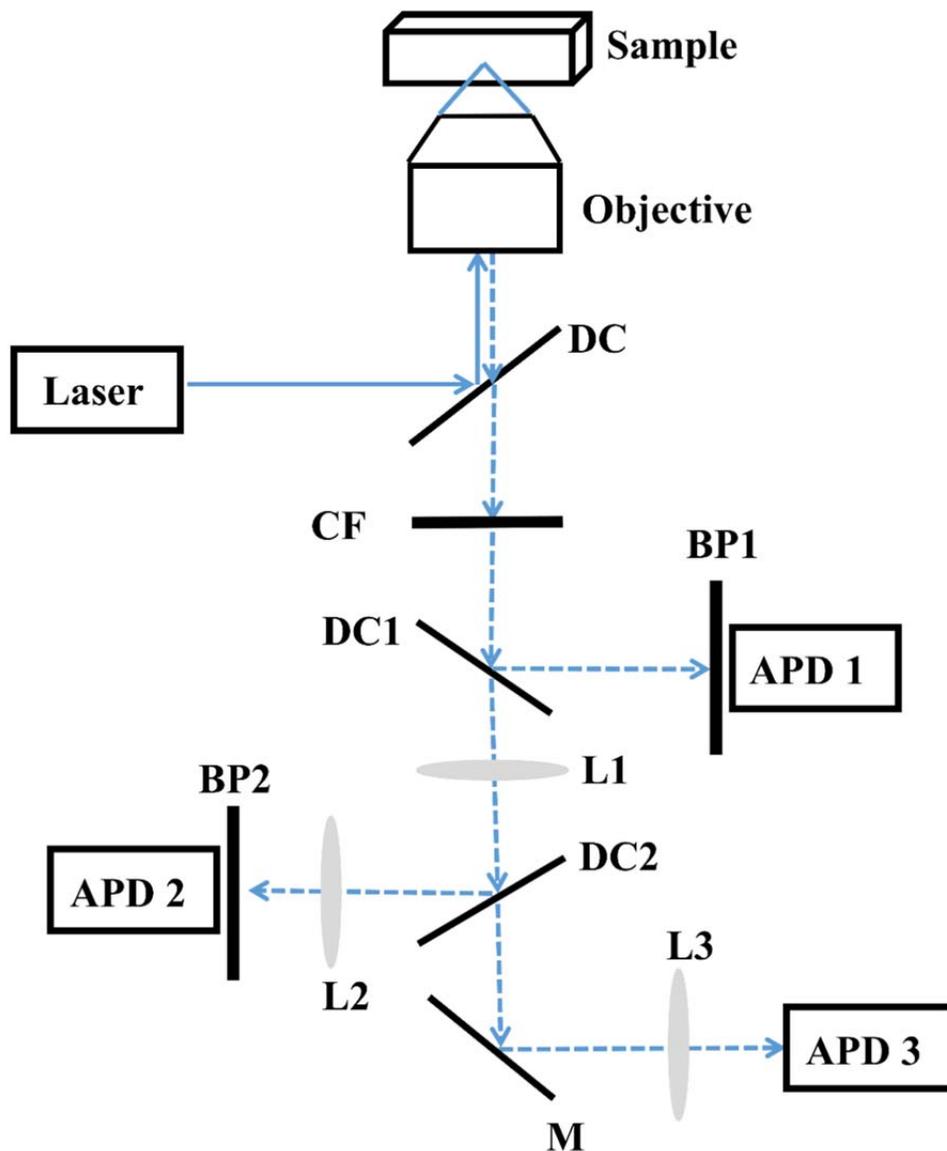
Figure 2.2 Schematic illustration of microscopy optics.

The laser beam is reflected by a dichroic mirror (DC) and focused through the objective into the sample. Fluorescence is collected by the objective and transmitted through the DC and a clean-up filter (CF) before entering the three-color detection module. The dichroic mirrors (DC1, DC2) and band-pass filters (BP1, BP2) determine the spectral range seen by each detector (APD: avalanche photodiode detector). The lenses (L1, L2, L3) focus the fluorescence onto the detectors. M is a mirror.

To perform three-color FFS experiments we had to build a three-channel detection set-up as shown in Figure 2.2. A dichroic mirror (DC1) reflects light within a certain wavelength region towards the first detection channel. The reflected light passes through an additional band pass filter (BP1) to completely remove unwanted wavelengths before detection in the first channel by an avalanche photodiode (APD 1). The light transmitted by DC1 encounters a second dichroic filter (DC2). Part of the light is reflected into the second channel and is detected by APD 2 after passing through BP2. The light transmitted through DC2 enters the third channel for detection by APD 3. The wavelength band detected by each channel is selected by proper configuration of the optical filters. Appendix A contains additional information about the detection optics.

Individual dichroic mirrors can be easily removed from the optical setup. Removal of DC1 and DC2 eliminates channel 1 and 2, respectively. This option is used to configure the instrument for the single and dual-color FFS experiments described in the next two chapters. The APDs (SPCM-AQ-14, Perkin-Elmer, Waltham, MA) operated in photon counting mode and were connected to a data acquisition card (RC1000-PP, Celoxica, UK or ISS, Champaign, IL) to preserve the photon count record for further analysis.

2.3 Point Spread Function Models

Focusing a laser beam through the objective leads to a spatial intensity distribution of the illumination light around the focus. The distribution defines the illumination point spread function (PSF), $\mathrm{PSF}_{\mathrm{ill}}(x, y, z)$. While this function $\mathrm{PSF}_{\mathrm{ill}}(x, y, z)$ can be calculated by using vectorial diffraction theory [67], there are no

analytical solutions for a typical FFS setup. As numerical solutions are inconvenient and do not provide insight, it is customary to approximate the PSF by a simple function for modeling purposes.

Before discussing model functions we need to introduce another type of PSF. FFS is not directly concerned with the illumination PSF, but with the fluorescence generated from a laser beam focused into a solution containing fluorophores. Thus, we introduce the excitation PSF ($\text{PSF}_{\text{exc}}(x, y, z)$), which describes the intensity distribution of excited fluorophores. In one photon excitation, the excitation rate of a fluorescent molecule is directly proportional to the illumination intensity as long as the illumination intensity is low enough to avoid saturation of the fluorescent molecule. Hence, in one photon excitation the illumination and excitation PSF are identical, $\text{PSF}_{\text{exc}}(x, y, z) = \text{PSF}_{\text{exc}}(x, y, z)$. In two-photon excitation, the excitation rate is proportional to the square of the illumination intensity, which implies $\text{PSF}_{\text{exc}}(x, y, z) = \text{PSF}_{\text{ill}}^2(x, y, z)$.

There are three model functions widely used to approximate the excitation PSF of an FFS experiment: the two-dimensional Gaussian (2DG) PSF model, the three-dimensional Gaussian (3DG) PSF model and the Gaussian-Lorentzian (GL) PSF model. We briefly discuss each of these models.

The 3DG PSF model is given by

$$PSF_{3DG}(x, y, z) = Exp\left(-2n\frac{x^2 + y^2}{\omega_0^2} - 2n\frac{z^2}{z_0^2}\right) \tag{2.1}$$

The radial beam waist ω_0 characterizes the effective size of the illumination beam in radial direction. The axial beam waist z_0 characterizes the extent of the PSF along the z-direction, which coincides with the optical axis of the beam. The excitation PSF depends on the order of the excitation process. One-photon excitation corresponds to $n=1$, while two-photon excitation is described by $n=2$. The 2DG PSF model is a special case of the 3DG PSF model, where the axial beam waist goes to infinity,

$$PSF_{2DG}(x, y, z) = Exp\left(-2n\frac{x^2 + y^2}{\omega_0^2}\right) \qquad (2.2)$$

The 2DG PSF is axially uniform, so it does not yield any intensity fluctuations along the z direction.

The GL PSF model,

$$PSF_{GL}(x, y, z) = \left(\omega_0/\omega(z)\right)^{2n} Exp\left(-2n\frac{x^2 + y^2}{\omega^2(z)}\right), \qquad (2.3)$$

is Gaussian in the radial direction, but has a Lorentzian shape along the z-direction with $\omega(z) = \omega_0\left(1 + z^2/z_0^2\right)^{1/2}$. This PSF model was widely used to describe two-photon FFS experiments [68]. For one-photon excitation, as used in confocal microscopy, a pinhole is added to axially restrict the optical observation volume. To take this restriction into account a collection profile function $\Omega(r, z)$ is combined with the PSF in order to define an effective PSF [19,69].

While these models are sufficient to describe the vast majority of FFS experiments, we have found that none of these PSF functions is suitable for z-scan FFS

experiments, where the excitation PSF is moved through the sample along the z-direction [63]. In order to accurately describe the experimental data of a z-scan FFS experiment, we had to introduce the modified Gaussian-Lorentzian (mGL) PSF model,

$$PSF_{mGL}(x,y,z) = (\omega_0/\omega(z))^{(1+\eta)n} Exp\left(-2n\frac{x^2+y^2}{\omega^2(z)}\right)$$ (2.4)

The mGL model introduces η as a new parameter, which modifies the shape of the Lorentzian function along the z axis. This allows for adjustments in the rate of the z-decay of the PSF, while preserving the Gaussian cross section of the PSF. For $\eta=1$, the mGL PSF model reduces to the GL PSF model 2.3. The xz cross section of the PSF models is shown in Figure 2.3 for η=1 and 1.9 with beam waists of $\omega_0=0.45$ and $z_0=0.95$. The mGL model is used in Chapter 4 to describe the z-scan experiments on E. coli cells.

2.4 Conventional FFS

Conventional FFS assumes that the excitation PSF is completely embedded in the sample. This condition is certainly fulfilled when the laser beam is focused into a thick solution. However, once the sample dimensions approach a few micrometers, one has to carefully check whether the assumption of a completely embedded PSF is still valid. FFS experiments in cells fall under this category. Some experiments in cells are accurately described by conventional FFS theory, while others require modifications. We focus in the following on conventional FFS theory. At the end of the chapter we discuss the influence of finite sample geometry on FFS experiments.
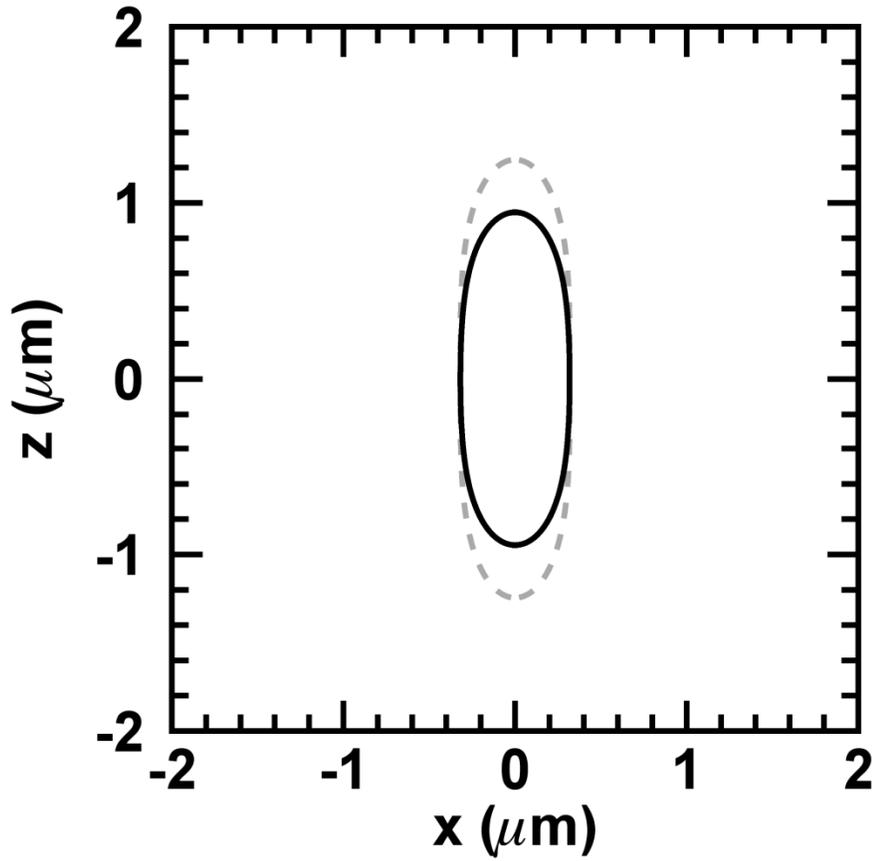
Figure 2.3 Contour plot of modified Gaussian-Lorentzian PSF model in x-z plane

The contour plot satisfies the condition $PSF_{mGL}(\vec{r}) = e^{-2}$. The dashed line corresponds to $\eta = 1$, while the solid line represents $\eta = 1.9$.

Conventional FFS implicitly assumes an infinite sample volume

$$V_{Sample} = \left\{ (x', y', z') \middle| x', y', z' \in (-\infty, \infty) \right\} \tag{2.5}$$

The PSF volume V_∞ is defined by integrating the excitation PSF over the sample volume

$$V_\infty = \int_{V_{Sample}} \text{PSF}(\mathbf{r}) d\mathbf{r} \tag{2.6}$$

We frequently refer to V_∞ as the excitation volume or the optical observation volume. For the mGL model, its value is described by

$$V_\infty = \frac{\pi^{3/2} z_0 \omega_0^2 \Gamma(\eta - 0.5)}{4\Gamma(\eta)}. \tag{2.7}$$

The subscript ∞ is used throughout this chapter to explicitly specify an infinite sample volume. The subscript is omitted in the literature and the next chapters, because an infinite volume is the default assumption of almost all FFS studies. We further define the m-th order PSF volume as the integral of the PSF raised to the m-th power over the sample space,

$$V_\infty^{(m)} = \int_{V_{Sample}} \text{PSF}^m(\mathbf{r}) d\mathbf{r}, \tag{2.8}$$

and define the m-th gamma- or shape-factors of the PSF by

$$\gamma_{m,\infty} = \frac{V_\infty^{(m)}}{V_\infty}. \tag{2.9}$$

The second gamma factor $\gamma_{2,\infty}$ is the most important for conventional FFS studies. Its value is 1/2, $\sqrt{2}/4$, and 3/16 for the 2DG, 3DG, and GL model, respectively. For the mGL model, its value is

$$\gamma_{2,\infty} = \frac{1}{2} \frac{\Gamma\left(2\eta + \frac{1}{2}\right)\Gamma(\eta)}{\Gamma(2\eta + 1)\Gamma\left(\eta - \frac{1}{2}\right)} \tag{2.10}$$

To illustrate the effect of η on the PSF, we calculated V_{∞} and $\gamma_{2,\infty}$ from Eqs. 2.7 and 2.10 with parameters $\omega_0 = 0.45\,\mu\text{m}$ and $z_0 = 0.95\,\mu\text{m}$ (Figure 2.4). As η increases, V_{∞} decreases, since the excitation volume is axially shrunk (Figure 2.3). The gamma factor $\gamma_{2,\infty}$ increases with η, which is caused by a sharper spatial decay of the PSF in axial direction. We obtain $V_{\infty} = 0.25\ \mu\text{m}^3$ and $\gamma_{2,\infty} = 0.27$ for $\eta = 1.9$.

The number of molecules within the observation volume is defined by

$$N(t) = \int_{V_{Sample}} \text{PSF}(\mathbf{r})c(\mathbf{r},t)\,d\mathbf{r}\,, \tag{2.11}$$

where $c(\mathbf{r},t)$ is the concentration of fluorophores at the location $\mathbf{r}$ at time t. The instantaneous fluorescence intensity is the product of the brightness λ with the number of molecules in the observation volume,

$$F(t) = \lambda \int_{V_{Sample}} \text{PSF}(\mathbf{r})c(\mathbf{r},t)\,d\mathbf{r} = \lambda N(t)\,. \tag{2.12}$$

For the majority of FFS experiments the concentration average $\langle c \rangle$ of a freely diffusing fluorophore is within good approximation independent of space and time. In this case the average number of molecules in the observation volume is

$$\langle N \rangle = cV_{\infty} \tag{2.13}$$
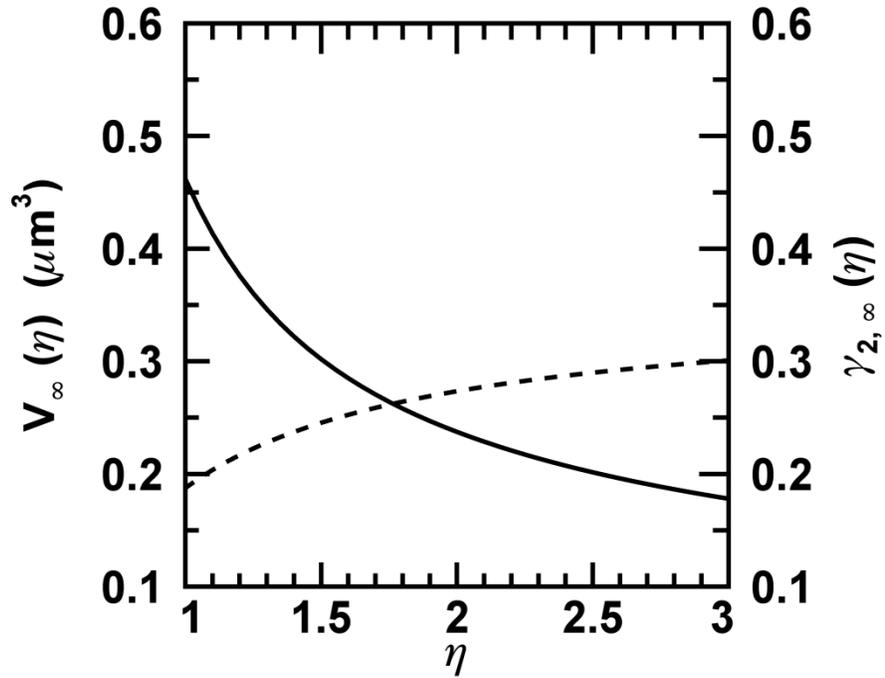
Figure 2.4 mGL PSF model's volume and gamma factor versus η

The volume V_∞ (solid line; left y-axis) and gamma factor $\gamma_{2,\infty}$ (dashed line, right y-axis) are shown for a beam waist of $\omega_0 = 0.45\,\mu\mathrm{m}$ and $z_0 = 0.95\,\mu\mathrm{m}$.

And the average fluorescence intensity is given by

$$\langle F \rangle = \lambda \langle N \rangle. \tag{2.14}$$

A fluorescence intensity fluctuation is defined by the deviation of the instantaneous signal from the mean,

$$\Delta F(t) = F(t) - \langle F \rangle. \tag{2.15}$$

We next consider the correlation $\langle \Delta F(t_1) \Delta F(t_2) \rangle$ of intensity fluctuations at two different times. Because conventional FFS assumes a stationary signal, the correlation only depends on the time difference $\tau = t_2 - t_1$ and can be written as $\langle \Delta F(0) \Delta F(\tau) \rangle$. Dividing this expression by the square of the average intensity defines the autocorrelation function (ACF),

$$G(\tau) = \frac{\langle \Delta F(0) \Delta F(\tau) \rangle}{\langle F \rangle^2}. \tag{2.16}$$

FFS had its origins in measurements of the ACF with dedicated hardware correlators and is known as fluorescence correlation spectroscopy (FCS). Fluctuation experiments that looked beyond the ACF were introduced only after FCS had been a well-established technique.

The pioneering work by Elson, Magde and Webb demonstrated the first FCS experiments measuring the diffusion coefficient and concentration of fluorescent molecules. Extension of FCS to measure flow, rotational diffusion, and the triplet state were introduced soon after. Here we focus on FCS of freely diffusing molecules, which is

the most relevant case for the studies described in this work. For the 3DG PSF model, the ACF of molecules with diffusion coefficient D is

$$G(\tau) = G(0) \frac{1}{\left(1 + \tau/\tau_D\right)\left(1 + \tau/\alpha\tau_D\right)^{1/2}} \tag{2.17}$$

where

$$\tau_D = \frac{\omega_0^2}{4nD} \tag{2.18}$$

is called the diffusion time or the residence time of the diffusing molecule. It represents a measure of the average time a diffusing particle spends within the observation volume. The squared ratio of the axial and radial beam waist is $\alpha = z_0^2/\omega_0^2$. The fluctuation amplitude

$$G(0) = \frac{\gamma_{2,\infty}}{\langle N \rangle} \tag{2.19}$$

is inversely proportional to the number of molecules in the PSF volume. The ACF has a non-exponential decay, which differs from the exponential decay of concentration fluctuations [70]. The non-exponential nature of the ACF is caused by integrating concentration correlations over the PSF volume [18].

An experimental ACF of the fluorescent protein EGFP is shown in Figure 2.5 together with a fit to Eq. 2.17. There are two fit parameters, the diffusion time τ_D and the fluctuation amplitude $G(0)$. The diffusion time τ_D is approximately the lag time that corresponds to a decay of the ACF amplitude to half its original value, $G(\tau_D) \approx G(0)/2$.
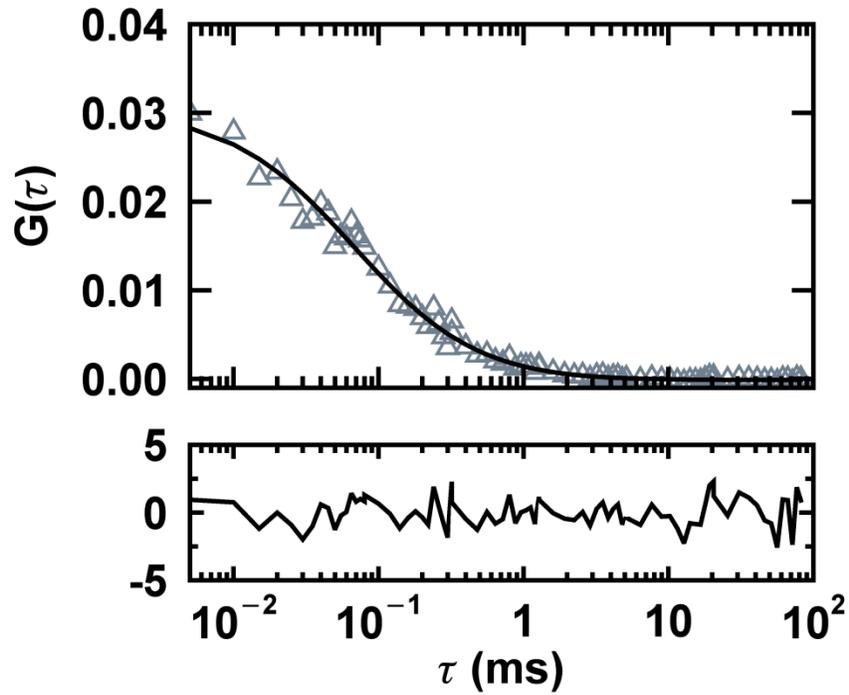
Figure 2.5 Autocorrelation function for EGFP in solution

The experimental autocorrelation function (triangles) was calculated from photon count data collected at 50 kHz. The fit (solid line) yields $G(0)$=0.030 and τ_D=0.27 ms with a reduced chi-square of χ^2=1.14.

The fitted value for τ_D can be converted into a diffusion coefficient by Eq. 2.18. For ω_0 =0.45 μm and two-photon excitation ($n = 2$) the diffusion coefficient is determined as D =0.093 $\mu m^2 / ms$. If we assume a spherical non-interacting particle in a solution of known viscosity η, the particle radius R can be estimated by the Stokes-Einstein relation, $D = k_B T / (6\pi R \eta)$, where $k_B T$ is the thermal energy. By using the viscosity of water (1.0 mPa·s at 20 °C) we obtained a hydrodynamic radius R=2.6 nm, which is consistent with previously reported values [71,72].

The number of molecules $\langle N \rangle$ in the observation volume can be calculated from the fluctuation amplitude $G(0)$. $\langle N \rangle$ is converted into a number concentration $c = \langle N \rangle / V_\infty$ or a molar concentration $\tilde{c} = \langle N \rangle / (V_\infty N_A)$ with the help of the PSF volume V_∞ and Avogadro's number N_A.

The ACF for the 2DG PSF and freely diffusing molecules is

$$G(\tau) = G(0) \frac{1}{(1 + \tau/\tau_D)}. \qquad (2.20)$$

It is a special case of the 3DG model for $\alpha \to \infty$ in Eq. 2.17. The 2DG PSF is axially uniform, so it does not yield any intensity fluctuations along the z direction. Because the PSF along the axial direction is much larger than in the radial direction, intensity fluctuations from molecules entering and leaving the PSF volume are mainly felt along the radial direction. Thus, the 2DG ACF typically provides a good approximation of ACF curves measured in solution. The 2DG model is also frequently used to describe

experiments involving two-dimensional sample geometries, such as measurements of protein diffusion along a planar membrane.

The ACF of the GL model has been discussed elsewhere [68]. Its functional shape is very well approximated by the 3DG ACF [73]. Thus, the shape of the ACF is fairly insensitive to the PSF model. Because of this fact, we use Eq. 2.17 to describe the ACF irrespective of the PSF model used. The amplitude $G(0)$ of the ACF, on the other hand, is sensitive to the ACF model, because of differences in the gamma factor $\gamma_{2,\infty}$. Thus, we fit ACF curves to Eq. 2.17, but interpret the fluctuation amplitude by using the proper PSF model for $\gamma_{2,\infty}$ in Eq. 2.19.

Moments of the fluorescence intensity distribution provide an alternative way to analyze fluctuations. We focus here on the first two moments, the mean $\langle F \rangle$ and the variance $\langle \Delta F^2 \rangle$. The variance can be expressed in terms of FFS parameters,

$$\left\langle \Delta F^2 \right\rangle = \gamma_{2,\infty} \lambda^2 \left\langle N \right\rangle . \tag{2.21}$$

Combining Eqs. 2.21 and 2.14 determines the brightness,

$$\frac{\left\langle \Delta F^2 \right\rangle}{\left\langle F \right\rangle} = \gamma_{2,\infty} \lambda , \tag{2.22}$$

and the fluctuation amplitude,

$$G(0) = \frac{\left\langle \Delta F^2 \right\rangle}{\left\langle F \right\rangle^2} = \frac{\gamma_{2,\infty}}{\left\langle N \right\rangle} . \tag{2.23}$$

However, unlike ACF analysis, moment analysis does not provide temporal information about the fluctuations.

Up to this point we assumed a single FFS species characterized by $\langle N \rangle$ and λ. Now we consider a mixture of species. The mean and variance of the intensity of the i-th species is $\langle F_i \rangle = \lambda_i \langle N_i \rangle$ and $\langle \Delta F_i^2 \rangle = \gamma_{2,\infty} \lambda_i^2 \langle N_i \rangle$, where $\langle N_i \rangle$ is the number of molecules and λ_i is the brightness of the i-th species. The total mean and variance of the intensity for the mixture is given by a sum over the mean $\langle F_i \rangle$ and variance $\langle \Delta F_i^2 \rangle$ of each species,

$$
\begin{aligned}
\langle F \rangle &= \sum_i \langle F_i \rangle = \sum_i \lambda_i \langle N_i \rangle \\
\langle \Delta F^2 \rangle &= \sum_i \langle \Delta F_i^2 \rangle = \sum_i \gamma_{2,\infty} \lambda_i^2 \langle N_i \rangle
\end{aligned}
\tag{2.24}
$$

Similarly, the ACF of the mixture is given by a sum over the ACF $G_i(\tau)$ of each species weighted by the square of the intensity fraction

$$
G(\tau) = \sum_i \frac{\langle F_i \rangle^2}{\langle F \rangle^2} G_i(\tau).
\tag{2.25}
$$

The theory presented in this section is based on intensities. While it provides an excellent tool for modeling, it does not accurately present FFS data, because our experiments record photon counts and not fluorescence intensities. Photon counting introduces shot noise, which is not accounted for in the above theory. Thus, we have to use photon count moments to properly analyze FFS data. Fortunately, intensities and photon counts are related by Mandel's formula, which allows us to connect each of the intensity-based equations to a description based on photon counts. Section 2.7 discusses this topic in more detail.

Conventional FFS has been widely applied to the study of *in vitro* systems. It has also been used for experiments in thick sections of live cells. For example, the nucleus of

many adherent mammalian cells has a height of ~5 μm, which is sufficient to approximately embed the entire PSF. We limited our discussion to a single detection channel, but extension of the concepts described in this section to two channels is fairly straightforward. We discuss dual-channel FFS in more detail later on in this chapter.

2.5 Single-Channel Brightness and Stoichiometry

The previous section introduced basic analysis tools that determine the diffusion time, fluctuation amplitude, and brightness λ of a sample. This section focuses on brightness, because it specifies the average rate of photon counts per fluorescent particle [74] and is of central importance in the next three chapters. While there are many techniques that detect protein interactions, brightness is unique in its ability to quantify interactions and protein stoichiometry both in vitro and in vivo.

We discuss brightness analysis using green fluorescent protein EGFP as an example. The brightness λ_{EGFP} of a sample containing EGFP is measured. Its value depends on the optical setup and laser power. However its value is a constant as long as the experimental setup remains unchanged. Next the brightness λ of a protein tagged with EGFP is measured. Comparing λ to the reference brightness λ_{EGFP} of the label identifies the stoichiometry of the labeled protein. The normalized brightness,

$$b = \frac{\lambda}{\lambda_{EGFP}}, \tag{2.26}$$

offers a convenient measure of the stoichiometry. For a monomeric protein the normalized brightness is $b = 1$ (Figure 2.6A). A dimeric protein complex has a normalized brightness of $b = 2$, because the complex contains two independently
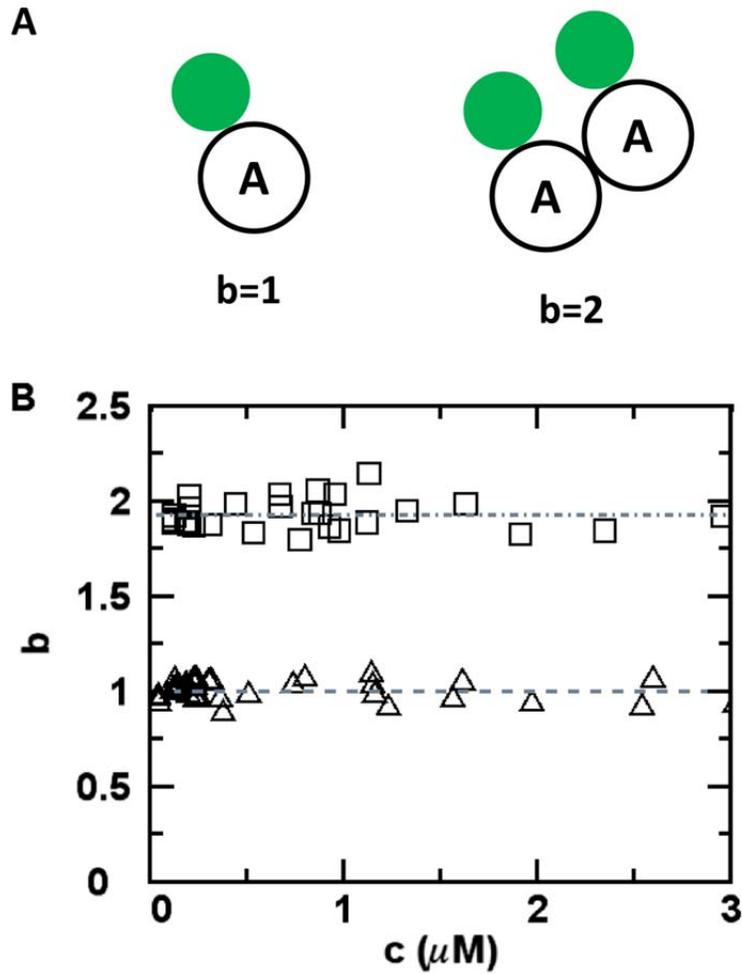
Figure 2.6 Normalized brightness of monomeric and dimeric proteins

(A) Schematic illustration of a monomeric protein tagged with EGFP and a homo-dimeric complex. The normalized brightness of the monomer is b = 1, while the dimeric complex has a normalized brightness of b = 2. (B) Normalized brightness of EGFP (triangles) and its tandem dimer EGFP$_2$ (squares) is graphed versus the molar concentration in units of the monomeric protein. The mean value for EGFP is, by definition, equal to b = 1 (dashed line). The experimentally determined mean value for EGFP$_2$ is b = 1.93 (dot-dashed line).

fluorescing EGFP labels (Figure 2.6A). In general, a normalized brightness $b = n$ identifies an n-meric protein. This concept of brightness analysis of protein stoichiometry was first demonstrated in living cells by comparing the brightness of EGFP with the brightness of a tandem-dimeric EGFP$_2$ complex, which served as an artificial dimer. The normalized brightness of both proteins is shown in Figure 2.6B versus protein concentration. The average brightness of EGFP is $b = 1$ and the brightness of EGFP$_2$ is $b = 1.93$. Brightness analysis is not limited to monomers and dimers, but has been applied to identify larger molecular complexes and even determine the protein copy number of viral particles.

2.5.1 Apparent Brightness

Most proteins inside the cell exist in multiple oligomeric states with a composition that depends on protein concentration and other environmental factors. Thus, FFS experiments on EGFP-labeled proteins in cells typically involve a mixture of brightness states. Thus, the brightness calculated from the first two intensity moments does not present a single state, but rather a weighted average over all brightness states of the sample. We refer to this quantity as the apparent brightness of the mixture,

$$b_{app} = \frac{\sum_i b_i^2 \langle N_i \rangle}{\sum_i b_i \langle N_i \rangle} = \sum_i f_i b_i \,, \tag{2.27}$$

where b_i and $\langle N_i \rangle$ are the brightness and number of molecules of the i-th species in the excitation volume. The expression can be derived from Eqs. 2.22 and 2.24. The fractional intensity $f_i = \langle F_i \rangle / \langle F \rangle$ is the ratio of fluorescence intensity of species i and the total

fluorescence of the sample. We continue to report the apparent brightness of a protein throughout the rest of this thesis, but it will simply be noted as b rather than b_{app}.

The degree of protein association depends on the total protein concentration. According to the law of mass action, high concentrations favor the associated state and low concentrations favor the dissociated state. Thus, changes in the concentration of a labeled protein determine the oligomeric composition of the sample and thereby its apparent brightness. In other words, the changes in the apparent brightness as a function of protein concentration provide a direct measure of the protein binding curve. We explore this idea by considering a monomer-dimer equilibrium of a labeled protein to provide a concrete example (Figure 2.7). At sufficiently low concentration, the protein is completely dissociated, which corresponds to $b = 1$. At high concentrations, the proteins fully associate into dimeric complexes. This situation corresponds to a single brightness state with $b = 2$. At intermediate concentrations, the sample consists of a mixture of monomers and dimers. The brightness lies between 1 and 2, reflecting the presence of a monomeric and dimeric brightness state. The exact value of the apparent brightness of the mixture is determined by Eq. 2.27.

We refer to the brightness plot of Figure 2.7 as a brightness titration. The brightness titration curve provides information about the average degree of association of a protein sample and the midpoint of the curve indicates the strength of interaction. Most importantly, brightness titration provides direct access to the protein binding curve inside the living cell. We perform such experiments by transient transfection of cells with the labeled protein. Since transiently transfected cells vary widely in the level of protein expressed, we take advantage of this behavior by selecting cells from low to high

39
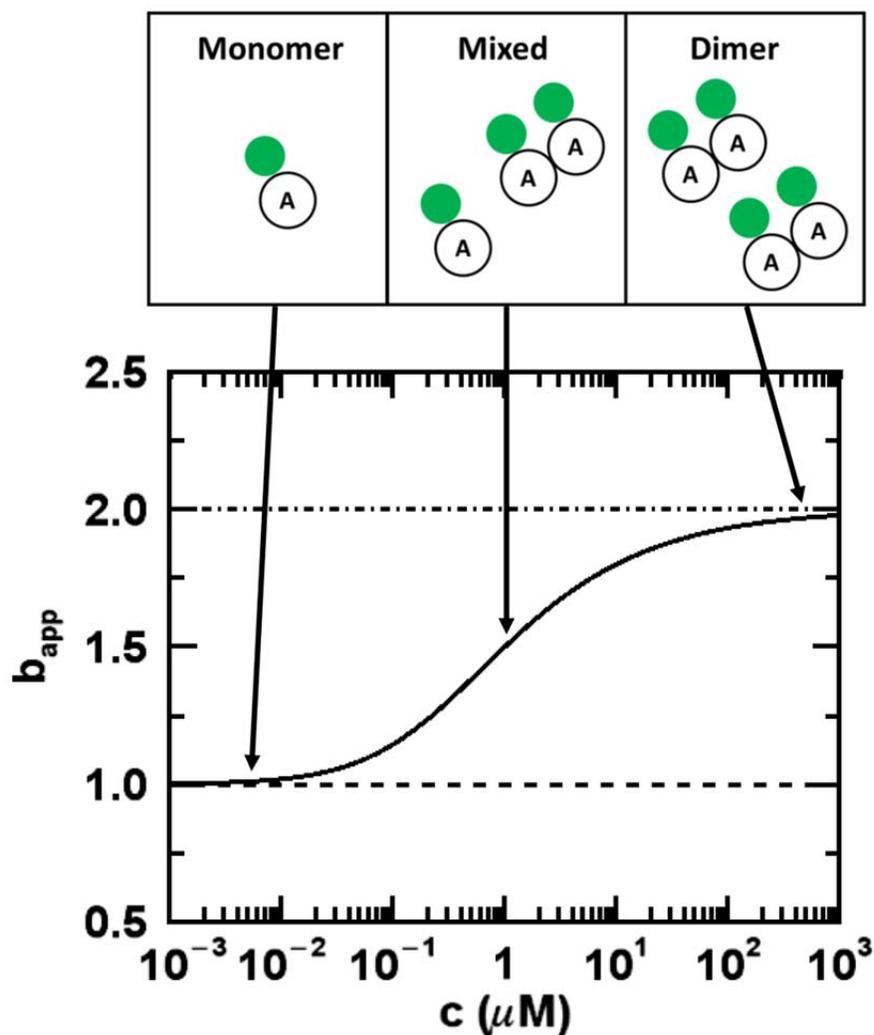
Figure 2.7 Brightness titration curve of b_{app}

Conceptual illustration of the change in the apparent brightness b_{app} of a homo-dimerization reaction, $2A \leftrightarrow A_2$, with a dissociation constant $K_D = 1\,\mu M$. The brightness curve was generated by Eq. 2.27 with b_{app} of 1 at low concentrations, which corresponds to the monomer A. The brightness increases with concentration and reaches a limiting value of 2 at high concentrations, which represents the homo-dimeric complex A_2. The brightness lies between 1 and 2 at intermediate concentrations and indicates the presence of a mixture of monomers and dimers.

expression levels. The experimental brightness titration curve is constructed by graphing the brightness of each cell versus the labeled protein concentration.

2.6 Time–integrated Intensities and Cumulants

Section 2.4 discussed the instantaneous intensity F. However, any measurement process samples data with a finite sampling time T. Thus, it is more accurate to speak of a time-integrated intensity

$$W(t) = \int_t^{t+T} F(t') dt',$$

(2.28)

instead of the instantaneous intensity. Clearly, if the time-scale of intensity fluctuations is much slower than the sampling time T, then the intensity stays approximately constant during the sampling process and Eq. 2.28 reduces to

$$W(t) = F(t)T,$$

(2.29)

which we refer to as the oversampling regime. Undersampling, on the other hand, implies that $F(t)$ changes during the sampling time T, which necessitates the use of Eq. 2.28. Thus, the theory described in section 2.4 is only valid in the oversampling regime, where $W(t)$ and $F(t)$ are proportional to each other.

While it is generally possible to choose a sampling time short enough to ensure oversampling conditions, it is not always a sound experimental strategy, because reducing the sampling time corresponds to less detected light, which inflates the experimental uncertainty of the measured quantity. It has been shown that choosing undersampling conditions frequently results in an increase of the signal-to-noise ratio of brightness experiments [75], which has been repeatedly exploited in cellular studies.

However, in order to employ undersampling conditions we need a theory based on Eq. 2.28, which is valid for arbitrary sampling times.

Instead of calculating moments of $W(t)$ it is advantageous to use cumulants of $W(t)$. The r-th order cumulant κ_r of a single species with brightness λ, number of molecules $\langle N \rangle$ and diffusion time τ_D is

$$\kappa_r = \gamma_{r,\infty} \lambda^r \langle N \rangle B_r(T,\tau_D). \tag{2.30}$$

The r-th order binning function is defined by

$$B_r(T,\tau_d) = r! \int \cdots \int_{0 \leq t_1 \leq t_2 \leq \cdots \leq t_r \leq T} G(t_2 - t_1, t_3 - t_2, \cdots, t_r - t_{r-1}; \tau_d) dt_1 dt_2 \cdots dt_r, \tag{2.31}$$

with

$$G_r(\tau_2, \cdots \tau_r; \tau_d, \alpha) = \frac{\left\{ \begin{array}{l} \int_V \cdots \int_V \mathrm{PSF}(\bar{r}_1)\mathrm{PSF}(\bar{r}_2)\cdots\mathrm{PSF}(\bar{r}_r) \dfrac{1}{(4\pi D\tau_2)^{\frac{3}{2}}} \exp\left(-\dfrac{\|\bar{r}_2 - \bar{r}_1\|^2}{4D\tau_2}\right) \\[3mm] \cdots \dfrac{1}{(4\pi D\tau_r)^{\frac{3}{2}}} \exp\left(-\dfrac{\|\bar{r}_r - \bar{r}_{r-1}\|^2}{4D\tau_r}\right) d\bar{r}_1 d\bar{r}_2 \cdots d\bar{r}_r \end{array} \right\}}{\int_V \mathrm{PSF}(\bar{r})^r d\bar{r}}. \tag{2.32}$$

The values of the binning function for $r > 2$ are determined numerically [75]. The time-integrated fluorescence cumulants κ_r can be expressed as ordinary moments of W. The cumulants to fourth-order are explicitly listed below for reference,

$$
\begin{aligned}
\kappa_1 &= \langle W \rangle & &= \lambda N T \\
\kappa_2 &= \langle \Delta W^2 \rangle & &= \gamma_{2,\infty} \lambda^2 N B_2 \left(T, \tau_D \right) \\
\kappa_3 &= \langle \Delta W^3 \rangle & &= \gamma_{3,\infty} \lambda^3 N B_3 \left(T, \tau_D \right) \\
\kappa_4 &= \langle \Delta W^4 \rangle - 3 \langle \Delta W^2 \rangle^2 &&= \gamma_{4,\infty} \lambda^4 N B_4 \left(T, \tau_D \right)
\end{aligned}
\tag{2.33}
$$

The binning function simplifies in the oversampling regime to

$$
B_r(T, \tau_d) = T^r . \tag{2.34}
$$

2.7 From Analogue to Digital: Photon Counting Process

FFS data are almost always collected in photon counting mode. The data acquisition records the number of photons k detected during the sampling time T. Thus, the data of single detector are represented by a discrete time-ordered sequence of non-negative photon counts $\{k_i\}_{i=1}^{N}$, where i represents the i-th bin at time iT. Unlike intensities, the photon counting process is affected by shot noise (Poisson noise) [76,77]. Mandel's formula

$$
p(k) = \int_0^\infty \mathrm{Poi}(k, \eta W)\, p(W)\, dW , \tag{2.35}
$$

relates the probability distribution function (pdf) of the time-integrated intensity, $p(W)$, with the pdf $p(k)$ of the photon counts [24,78]. $\mathrm{Poi}(k, \langle k \rangle)$ is the Poisson function with mean $\langle k \rangle$ and the parameter η describes the sensitivity of the detector. It is customary in FFS experiments to set $\eta = 1$ for simplicity. With this definition intensities $F(t)$ are reported in counts per second and time-integrated-intensities $W(t)$ are expressed in counts.

With the help of Mandel's formula it can be shown that the moment generating function (mgf) of the integrated intensities is equal to the factorial mgf of the photon counts [79]. The logarithm of the mgf defines the corresponding cumulant generating function (cgf). Thus, the cgf of the integrated intensity is identical to the factorial cgf of the photon counts. This result proves that the r-th integrated intensity cumulants κ_r is equal to the r-th factorial cumulant $\kappa_{[r]}$ of the photon counts,

$$\kappa_{[r]} = \kappa_r .$$ (2.36)

The first four factorial cumulants expressed in terms of photon counts are

$$\begin{aligned}
\kappa_{[1]} &= \langle k \rangle \\
\kappa_{[2]} &= \langle \Delta k^2 \rangle - \langle k \rangle \\
\kappa_{[3]} &= \langle \Delta k^3 \rangle - 3\langle \Delta k^2 \rangle + 2\langle k \rangle \\
\kappa_{[4]} &= \langle \Delta k^4 \rangle - 6\langle \Delta k^3 \rangle - 3\langle \Delta k^2 \rangle^2 + 11\langle \Delta k^2 \rangle - 6\langle k \rangle
\end{aligned}$$ (2.37)

Experimentally, we determine factorial cumulants from the calculated moments of the photon count data. We next use Eqs. 2.37 and 2.30 to analyze data. For example, the ratio of the second and first factorial cumulant is used to determine the brightness λ,

$$\frac{\kappa_{[2]}}{\kappa_{[1]}} = \frac{\langle \Delta k^2 \rangle - \langle k \rangle}{\langle k \rangle} = \frac{\kappa_2}{\kappa_1} = \gamma_{2,\infty} \lambda \frac{B_2(T, \tau_D)}{T} .$$ (2.38)

This equation simplifies to

$$\frac{\kappa_{[2]}}{\kappa_{[1]}} = \gamma_{2,\infty} \lambda T ,$$ (2.39)

in the case of oversampling. This equation is very similar to Eq. 2.22, except for the factor T, which appears to account for the time-integration, $\left\langle \Delta W^2 \right\rangle / \left\langle W \right\rangle = \left(\left\langle \Delta F^2 \right\rangle T^2 \right) / \left(\left\langle F \right\rangle T \right)$.

Cumulants of the sum of statistically independent variables are given by the sum of the cumulants of the individual variables. This property makes it straightforward to expand the theory to a mixture of brightness species. Thus, the fluorescence intensity cumulants scale with the number of molecules in the observation volume, and the corresponding cumulant for a mixture of species is simply given by the sum of the cumulants of each species,

$$\kappa_r = \sum_i \kappa_{r,i} = \sum_i \gamma_{r,\infty} \lambda_i^r \left\langle N_i \right\rangle B_r \left(T, \tau_{D,i} \right), \tag{2.40}$$

where $\kappa_{r,i}$ is the r-th cumulant of the integrated fluorescence intensity of the i-th species.

Time-integrated fluorescence cumulant analysis (TIFCA) analyzes FFS data based on the model described in the last two sections [75]. The experimental photon counts taken with a specific sampling frequency are repeatedly rebinned in software to generate data for a range of sampling times T. For each sampling time, the first few factorial cumulants $\kappa_{[r]}(T)$ are calculated and then fit to a model based on Eq. 2.40. TIFCA has been used to identify the brightness, concentration, and diffusion time of a binary mixture of species and has been applied to cell experiments. TIFCA has also been expanded to two detection channels to describe dual-color FFS experiments [80]. In chapter 5 we consider TIFCA applied to three detection channels.

2.8 Sample Geometry

Conventional brightness experiments assume that fluorescent molecules access all areas of the PSF with equal probability as discussed in section 2.4. While this assumption is usually fulfilled for FFS experiments in solution and within thick regions of a cell, it breaks down at thin sections of the cell because the illumination volume extends beyond the thickness of the sample. As a consequence, a fluorescent molecule within the thin section cannot access all parts of the illumination PSF. Thus, excitation is confined to the overlap of the illumination PSF with the sample space. We make the finite sample size explicit by rewriting the expression of Eq. 2.6, which calculates the PSF volume,

$$V_{Sample} = \int_{V_{Sample}} PSF(\mathbf{r}) d\mathbf{r} = \int PSF(\mathbf{r}) S(\mathbf{r}) d\mathbf{r}. \tag{2.41}$$

The function $S_{Sample}(\mathbf{r})$ defines the sample geometry by

$$S_{Sample}(\mathbf{r}) = \begin{cases} 0 & \mathbf{r} \notin V_{Sample} \\ 1 & \mathbf{r} \in V_{Sample} \end{cases}. \tag{2.42}$$

Similarly, we define the r-th order PSF volume as

$$V_{Sample}^{(r)} = \int PSF^r(\mathbf{r}) S_{Sample}(\mathbf{r}) d\mathbf{r}, \tag{2.43}$$

and the gamma factors by

$$\gamma_{r, Sample} = V_{Sample}^{(r)} \big/ V_{Sample}. \tag{2.44}$$

The above symbols use the subscript *Sample* instead of the subscript ∞ used in section 2.4 to emphasize that the sample space is finite. All models discussed above for the

infinite sample space are still valid for the finite case provided $V_\infty^{(r)}$ and $\gamma_{r,\infty}$ are replaced by the equivalent expressions for the finite case.

This theory was first applied to a thin fluorescent layer starting at height z_1 and ending at height z_2. We refer to this geometry as a slab-layer. Applying

$$S_{slab}(z) = \begin{cases} 1 & z_1 \geq z \geq z_2 \\ 0 & else \end{cases},$$ (2.45)

to Eq. 2.43 yields an *r*-th order PSF volume

$$V_{slab}^{(r)}(z) = \int PSF^r(x', y', z') S_{slab}(z' - z) d\mathbf{r},$$ (2.46)

and a gamma factor

$$\gamma_{2,slab}(z) = \frac{V_{slab}^2(z)}{V_{slab}(z)}.$$ (2.47)

The variable z defines the position of the center of the excitation PSF. The average and variance of the fluorescence intensity depend on the location of the PSF with respect to the slab layer,

$$\begin{aligned} \langle F(z) \rangle &= \lambda c V_{slab}(z) \\ \langle \Delta F^2(z) \rangle &= \gamma_{2,slab}(z) \lambda^2 c V_{slab}(z) \end{aligned}.$$ (2.48)

FFS experiments in a slab are performed by first performing a scan of the laser beam along the z-axis through the slab layer. The z-scan produces a fluorescence intensity profile $\langle F(z) \rangle$, which is fit to the model described in Eq. 2.48 to determine the locations z_1 and z_2 of the slab (Figure 2.8). Once the location of the layer is known the
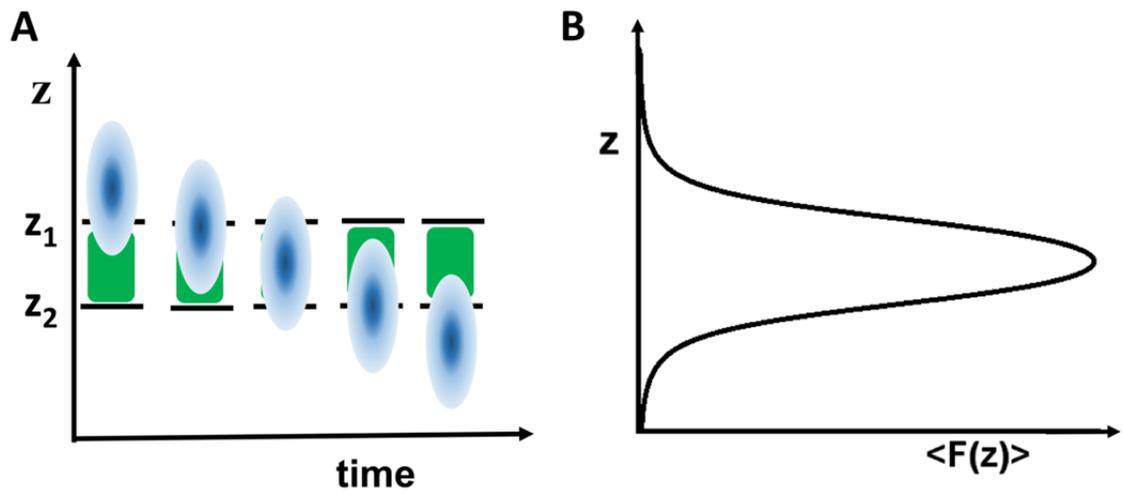
Figure 2.8 Schematic illustration of z-scan FFS through a slab layer

(A) Snapshots of the PSF location during a z-scan through a fluorescent sample layer starting at location z_1 and ending at z_2. (B) The z-scan results in an intensity profile that reflects the convolution of the PSF with the sample profile.

gamma factor $\gamma_{2,slab}(z)$ can be calculated for any position z with respect to the center of the PSF. Next, the laser is focused at mid height z_M of the slab to collect FFS data. The brightness λ is then calculated by

$$\frac{\left\langle \Delta F^2 \left(z_M\right) \right\rangle}{\left\langle F\left(z_M\right) \right\rangle} = \gamma_{2,slab}\left(z_M\right)\lambda \,. \tag{2.49}$$

This experimental approach is referred to as z-scan FFS. It was shown that z-scan FFS is crucial to accurately determine the brightness in thin sections of cells. Conventional FFS leads to large bias under these circumstances. Z-scan FFS has recently been expanded to multiple fluorescent layers, such as encountered for proteins that reside both at the plasma membrane and in the cytoplasm. Chapter 4 describes the extension of z-scan FFS to a cylindrical geometry. Appendix B derives expressions for z-scan FFS in a spherical geometry.

2.9 Dual-color FFS

Since single-color FFS is limited to the observation of homo-protein interactions, dual-color FFS was developed to study hetero-protein interactions between two protein species [6,80,81]. Dual-color FFS distinguishes the two proteins by employing differently colored FPs, such as EGFP and mCherry, as tags [6]. We will use G to denote a green-labeled protein, while R identifies a red-labeled protein. For simplicity, we assume that the emission of the green label is confined to detector channel 1, while the emission of the red label is only seen by detector channel 2. In this case, hetero-protein complexes are identified by concomitant intensity fluctuations in both detection channels (Figure 2.9).
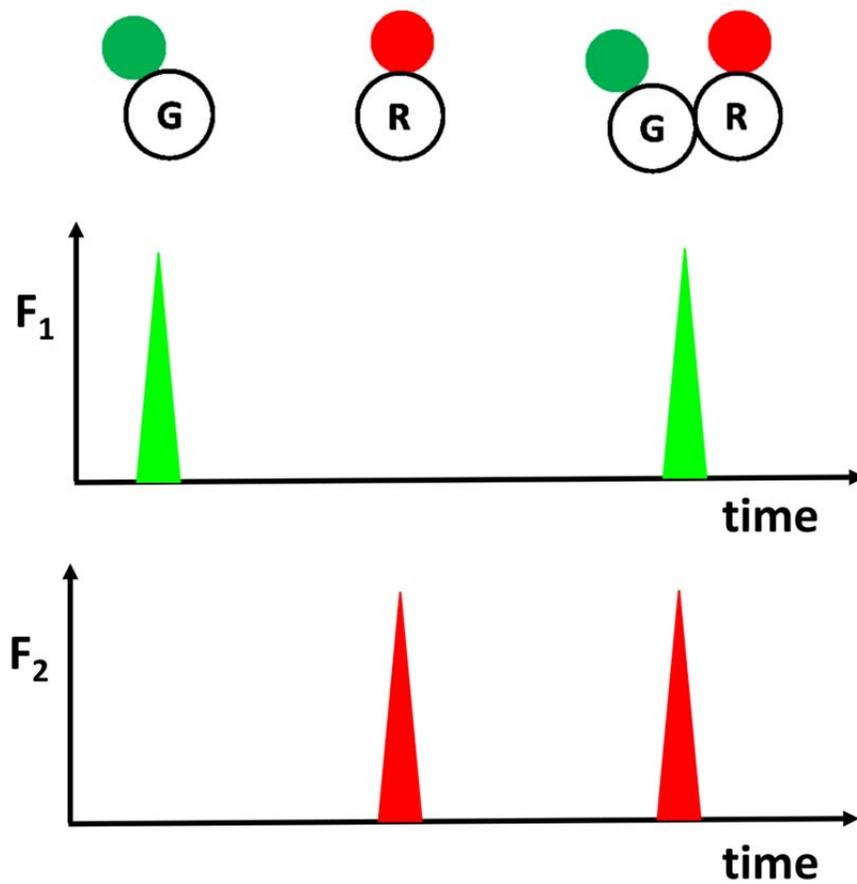
Figure 2.9 Schematic illustration of hetero-protein complex identification by dual-color FFS

Protein *G* labeled with a green FP induces a fluorescent signal in the 1st detection channel, while protein *R* labeled with a red FP leads to a fluorescent signal in the 2nd detection channel. The hetero-dimer *GR* results in the simultaneous appearance of a fluorescent signal in both detection channels.

Cross-correlation between the signals from each channel, and $F_1(t)$ and $F_2(t)$, would provide a simple measure to identify hetero-protein interactions, because the absence of a cross-correlation signal indicates the absence of hetero-interactions, while a positive result demonstrates the presence of hetero-protein complexes.

Unfortunately, the picture presented in Figure 2.9 is not achieved in actual experiments. The emission spectrum of fluorescent protein is wide with a long tail towards the red part of the emission spectrum. This leads to spectral overlap between the green and the red emitting FPs. This issue is illustrated in Figure 2.10A using EGFP and mCherry as example. While it is possible to restrict the emission of the red FP with proper choice of the dichroic mirror to a single detection channel (Figure 2.10), the fluorescence of the green FP exhibits cross talk into the second detection channel. The cross-talk introduces a positive cross-correlation signal even in the absence of hetero-protein complexes, which significantly complicates the analysis of dual-FFS data.

A powerful way of accounting for cross-talk is dual-color brightness analysis, where each fluorescent species is represented by a two component brightness vector $\vec{\lambda} = (\lambda_1, \lambda_2)$, where λ_i represents the single-color brightness of the species in the i-th detection channel. The brightness vector $\vec{\lambda}$ is a generalization of single-color FFS to two detection channels and provides a unique representation of each FP. We denote the brightness vector of a protein tagged with EGFP by $\vec{\lambda}_G = (\lambda_{1,G}, \lambda_{2,G})$ and a protein tagged with mCherry by $\vec{\lambda}_R = (0, \lambda_{2,R})$. The hetero-dimer GR is represented by $\vec{\lambda}_{GR} = \vec{\lambda}_G + \vec{\lambda}_R$ in the absence of FRET [28]. These three brightness vectors are described
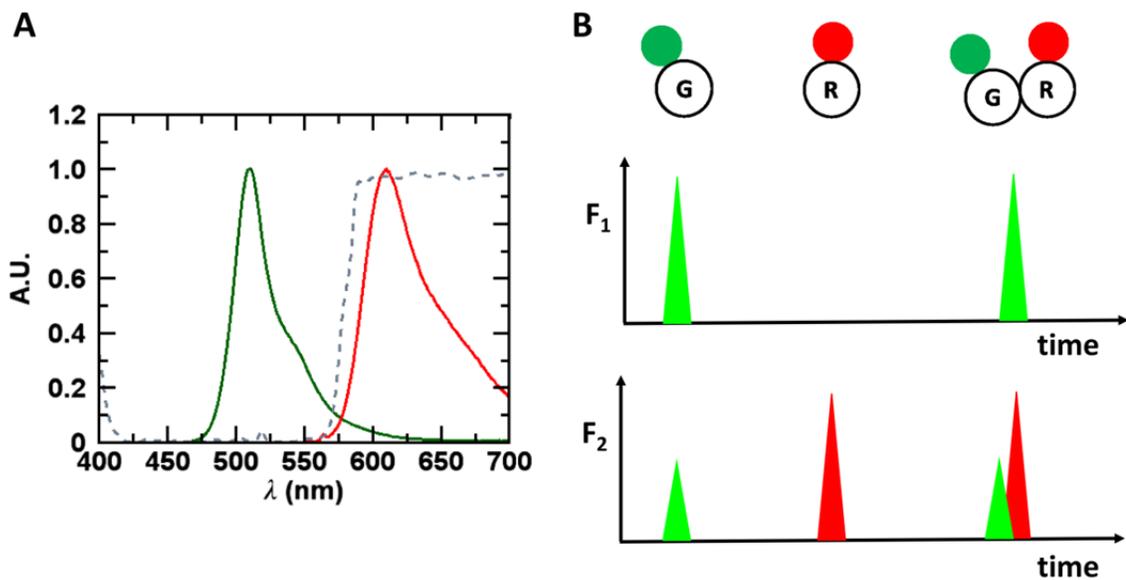
Figure 2.10 Spectral cross-talk

(A) The emission spectrum of EGFP (green line) and mCherry (red line) are plotted together with the transmission spectrum of the dichroic filter (dashed line) that separates the emission into two detection channels. (B) EGFP is detected in both channels, because part of the emission leaks into the second detection channel. The mCherry signal is only registered in the second channel. The signal of the hetero-dimer is the superposition of the mCherry and EGFP contributions.
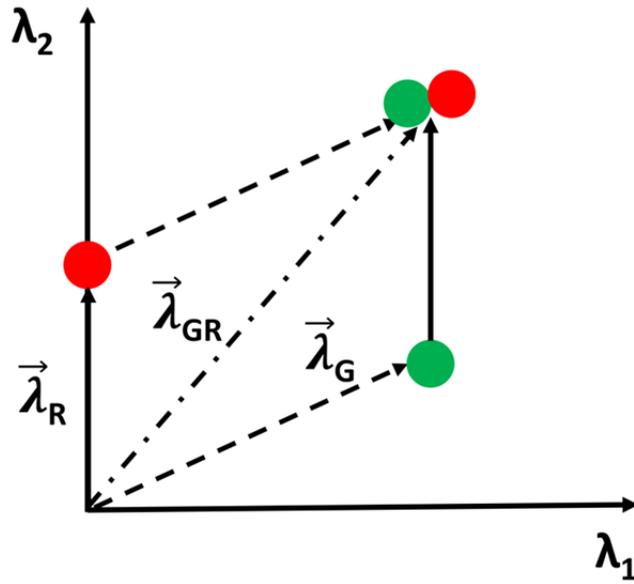
Figure 2.11 Brightness vectors for *G*, *R* and *GR* states in $\left(\lambda_1, \lambda_2\right)$ plane

The *G* state (green circle) with brightness vector $\vec{\lambda}_G = \left(\lambda_{1,G}, \lambda_{2,G}\right)$ corresponds to a monomer labeled with EGFP. The *R* state (red circle) with $\vec{\lambda}_R = \left(0, \lambda_{2,R}\right)$ represents a monomer labeled with mCherry. The *GR* state (green and red circle) with brightness vector $\vec{\lambda}_{GR} = \vec{\lambda}_G + \vec{\lambda}_R$ identifies the heterodimer in the absence of FRET.

in Figure 2.11. A hetero-dimer interaction, $G + R \leftrightarrow GR$, is characterized by three brightness states G, R and GR. Analysis of FFS data in principle identify these brightness states and resolve their individual concentrations.

2.10 Dual-color TIFCA

Identifying the brightness vector from dual-color FFS is achieved by generalizing single-color FFS analysis methods. For example, dual-color photon counting histogram (PCH) employs the joint probability of detecting photon count in both detection channels [81]. TIFCA improves the signal-to-noise ratio (SNR) over PCH by rebinning of photon count data, while accounting for undersampling as shown in section 2.6. This advantage becomes even more important in dual-channel detection, because the fluorescent signal is split into two channels, which leads to a reduced SNR in each channel compared to the SNR of single-color FFS. We briefly describe the relevant theory of dual-color TIFCA [80].

In dual-color FFS, each FP is characterized by a brightness vector $\vec{\lambda} = (\lambda_1, \lambda_2)$ and a diffusion time τ_D. The fluorescent intensity $F_i(t)$ of the i-th channel is

$$F_i(t) = \lambda_i \int_{V_{Sample}} \text{PSF}(\mathbf{r}) c(\mathbf{r}, t) d\mathbf{r}, \tag{2.50}$$

which corresponds to Eq. 2.12 in single-color FFS with concentration $c(\mathbf{r}, t)$.

The time-integrated intensity W_i is given by

$$W_i(t) = \int_t^{t+T} F_i(t') dt'. \tag{2.51}$$

The bivariate cumulants $\kappa_{i,j}$ of the time-integrated intensities W_1 and W_2 are derived as [80],

$$\kappa_{i,j}(T) = \gamma_{i+j,\infty} \lambda_1^i \lambda_2^j N B_{i+j}(T, \tau_D). \tag{2.52}$$

where i and j represent the cumulant order of the first and second channel, respectively, and N is the average occupation number in the observation volume.

A mixture of brightness species is described by summing the cumulants of each species s,

$$\kappa_{i,j}(T) = \sum_s \kappa_{i,j}(T, \tau_{D,s}), \tag{2.53}$$

Just as in single-color TIFCA, the bivariate cumulants $\kappa_{i,j}$ of the integrated intensity are equal to the bivariate factorial cumulants $\kappa_{[i,j]}$ of the photon counts, $\kappa_{i,j} = \kappa_{[i,j]}$. In chapter 5, we extend dual-color TIFCA to tri-color TIFCA by generalizing the theory described here from two to three detection channels.

2.11 Dual-color heterospecies partition (HSP) analysis

In principle, dual-color TIFCA would identify the brightness species by a fit of the experimental cumulants to Eq. 2.53. However, the achievable SNR of cell experiments is not high enough to reliably identify more than two species. This performance is not even sufficient to identify the three brightness states of protein hetero-dimerization $G + R \leftrightarrow GR$. Dual-color heterospecies partition (HSP) analysis was developed to overcome this challenge [6]. Instead of resolving all species present in the system, dual-color HSP analysis identifies a composite species, referred to as

heterospecies brightness, which provides a measure of the average stoichiometry of the hetero-interacting proteins.

Consider a binary system of protein G having s binding sites for protein R. The hetero-protein interaction creates species G, GR, GR_2, $\cdots$, GR_s. Homo-protein interaction is allowed only for R, which leads to species R, R_2, R_3 and so on. Homo-complexes of G, such as G_2, are not considered to keep the discussion as simple as possible [6]. Such a system will generally exist as a mixture of many brightness states which is illustrated in Figure 2.12A for the states G, R, GR, R_2, GR_2. We partition these states into two groups. The first group is referred as the heterospecies $H = \{G, GR_1, GR_2, \cdots\}$ and consists of states that contain at least one G. The second group is the free species $F = \{R, R_2, \cdots\}$, which is comprised of states without G (Figure 2.12B). These two composite states are described by five parameters. The state H is represented by $\vec{\lambda}_H = (\lambda_{1,H}, \lambda_{2,H})$ and N_H, and the state F by $\vec{\lambda}_F = (0, \lambda_{2,F})$ and N_F. These five parameters can be determined by dual-color brightness analyses [80,81] and related to the average stoichiometry of G and R of the mixture of brightness states [6].

All brightness species of the composite species H carry at least a single G protein (G, GR, GR_2, $\cdots$, GR_s), and as a consequence contribute intensity fluctuations to the first channel. The F species, on the other hand, only consists of brightness complexes of R, which cannot contribute intensity fluctuations to the first channel, because R has a zero brightness to the first channel. This feature is exploited by HSP to partition the fluctuations into two different classes. The brightness of the heterospecies H identifies
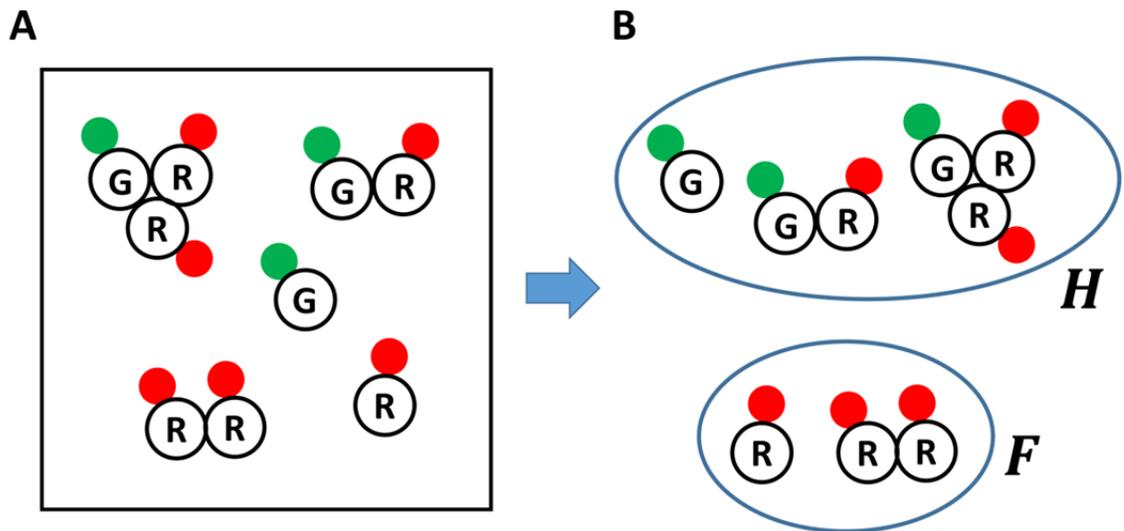
Figure 2.12 Heterospecies representation of dual-color FFS

(A) A hypothetical sample with protein complexes in the brightness states G, R, GR_1, R_2, GR_2. (B) The HSP separates this mixture into two different brightness states. The heterospecies H, which consists of protein complexes with at least one G present. The free species F consists of protein complexes that are free of G.

the average stoichiometry of the G and R interaction as discussed below.

For the sake of simplicity, we describe HSP analysis with cumulants assuming oversampling conditions $T \ll \tau_D$ which leads the $B_r(T, \tau_D) = T^r$ (Eq. 2.34). HSP analysis fits the experimental cumulants $\kappa_{[i,j]}$ of the photon counts up to the second order $(1 \leq i + j \leq 2)$ to the two species H and F of the HSP model,

$$\kappa_{i,j}^H(T) = \gamma_{i+j} \lambda_{1,H}^i \lambda_{2,F}^j N_H T^{i+j} + \gamma_{i+j} \delta_{i,0} \lambda_{2,F}^j N_F T^{i+j} . \tag{2.54}$$

where $\delta_{i,0}$ is the Kronecker delta. This fit uniquely determines the five FFS parameters $\{\lambda_{1,H}, \lambda_{2,H}, N_H, \lambda_{2,F}, N_F\}$ from the cumulants up to the second order $\left(\kappa_{1,0}^H, \kappa_{0,1}^H, \kappa_{1,1}^H, \kappa_{2,0}^H, \kappa_{0,2}^H\right)$. We now focus on the interpretation of the parameters describing the heterospecies $\lambda_{1,H}$, $\lambda_{2,H}$, and N_H.

The physical system consists of a mixture of brightness states. The cumulants of this mixture are according to Eq. 2.53 given by

$$\kappa_{i,j}(T) = \sum_{n=0}^{s} \gamma_{i+j} \lambda_{1,GR_n}^i \lambda_{2,GR_n}^j N_{GR_n} T^{i+j} + \sum_{n=1}^{s} \gamma_j \delta_{i,0} \lambda_{2,R_n}^j N_{R_n} T^j . \tag{2.55}$$

Species $G_m R_n$ is characterized by a brightness vector $\vec{\lambda}_{G_m R_n} = \delta_{m,1} \vec{\lambda}_G + n \vec{\lambda}_R$ in the absence of FRET and an average occupation number $N_{G_m R_n}$. The first term of Eq. 2.55 specifies the contributions from protein complexes that belong to the heterospecies H, and the second term defines the contributions associated with the free species F.

By matching the cumulants $\kappa_{1,0}^H = \kappa_{1,0}$, $\kappa_{2,0}^H = \kappa_{2,0}$ and $\kappa_{1,1}^H = \kappa_{1,1}$ in Eqs 2.54 and 2.55 we derive the following expressions for the heterospecies brightness,

$$\lambda_{1,H} = \lambda_{1,G}$$

$$\lambda_{2,H} = \lambda_{2,G} + \lambda_{2,R} \sum_{n=0}^{s} n f_n = \lambda_{2,G} + \lambda_{2,R} y, \qquad (2.56)$$

and obtain $N_H = \sum_{n=0}^{s} N_{GR_n}$ for the average occupation number of the heterospecies, which

corresponds to the total occupation number $N_{G,tot}$ of all G proteins expressed. The

parameter $f_n = N_{GR_n} / N_{G,tot}$ is the number fraction of species GR_n. We also defined

$y = \sum_{n=0}^{s} n f_n$, which is the degree of hetero-binding. It describes the mean number of R

proteins bound to protein G. For a HSP value $\hat{\lambda}_{2,H}$, the mean hetero-binding

stoichiometry is calculated as

$$y = \left(\hat{\lambda}_{2,H} - \lambda_{2,G} \right) / \lambda_{2,R}. \qquad (2.57)$$

As an example, consider the heterodimer model, $G + R \leftrightarrow GR$, with a

dissociation constant K_D. Proteins G, R and GR are represented by brightness vectors

$\vec{\lambda}_G$, $\vec{\lambda}_R$ and $\vec{\lambda}_{GR}$, respectively. Their occupation numbers are given by N_G, N_{GR} and

N_R. Applying HSP analysis results in a brightness vector of H given by

$$\vec{\lambda}_H = \left(\lambda_{1,G}, \ \lambda_{2,G} + y \lambda_{2,R} \right), \ y = N_{GR} / \left(N_G + N_{GR} \right). \qquad (2.58)$$

with an average occupation number $N_H = N_G + N_{GR}$. The parameter y varies from 0 to 1,

as shown in the brightness titration curve (Figure 2.13).
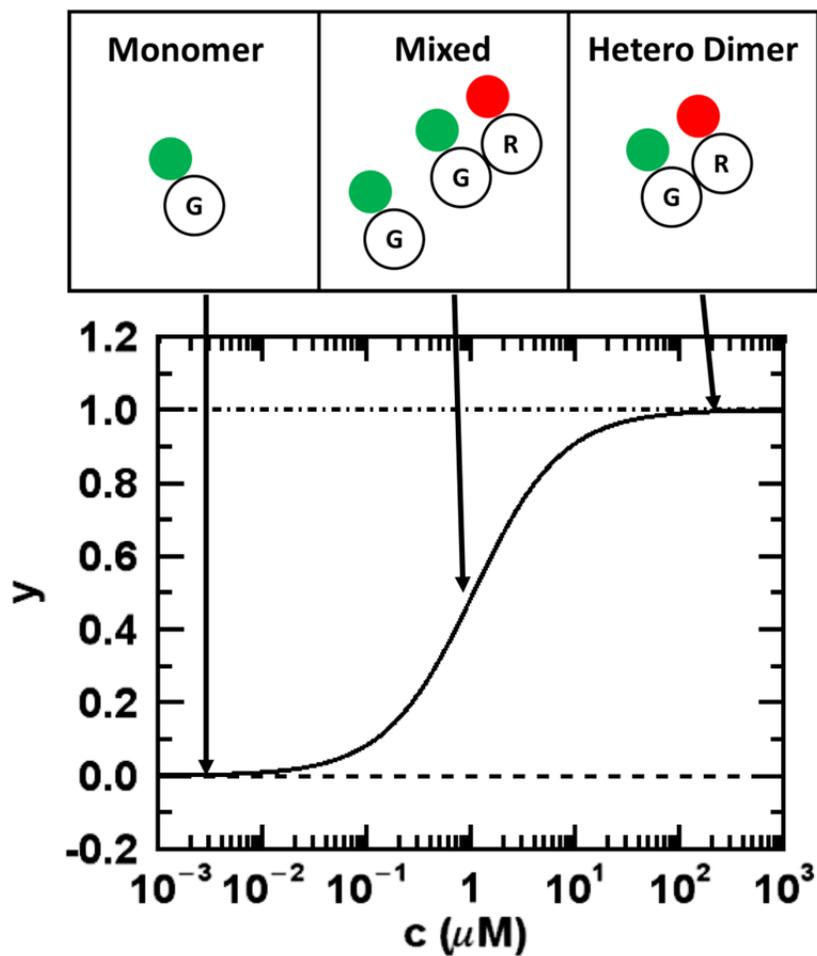
Figure 2.13 HSP brightness titration curve

The degree of hetero-binding y for the reaction $G + R \leftrightarrow GR$ is determined by HSP analysis and graphed versus the molar concentration c of the total expressed R protein for a dissociation constant $K_D = 1\,\mu M$. The value of y is ~0 for low concentrations, which corresponds to the monomeric state *G*. The *y*-value reaches 1 at concentrations much larger than K_D and represents the heterodimeric state *GR*. In the middle of concentration range, *y* is between 0 and 1, which indicates the mixture of *G* and *GR*.

3. Quantitative measurement of brightness from living cells in the presence of photodepletion

The brightness of fluorescently labeled proteins provides an excellent marker for identifying protein interactions in living cells. Quantitative interpretation of brightness, however, hinges on a detailed understanding of the processes that affect the signal fluctuation of the fluorescent label. This chapter focuses on the cumulative influence of photobleaching on brightness measurements in cells. Photobleaching within the finite volume of the cell leads to a depletion of the population of fluorescently labeled proteins with time. The process of photodepletion reduces the fluorescence signal which biases the analysis of brightness data. We characterized the bias and introduced an analysis method that accounts for photodepletion as verified by experiments with mammalian and yeast cells. In addition, photodepletion experiments with the fluorescent protein EGFP revealed the presence of a photoconversion process, which leads to a marked decrease in the brightness of the EGFP protein.

3.1 Introduction

A powerful application of FFS lies in the characterization of protein-protein interactions in living cells by brightness analysis [4,24–26,82] of homo-protein and hetero-protein complexes [4–6]. However, analysis of FFS data requires caution as signal fluctuation can be affected in subtle but significant ways by the sample environment [63]. The vast majority of cellular FFS experiments has been performed in mammalian cells. However, when we carried out FFS measurements in yeast cells, we observed a surprisingly large scatter of brightness values that was absent in mammalian cells

measured under otherwise identical conditions. The cause of this scatter is the cumulative but subtle reduction of the fluorophore concentration by photobleaching, which we will refer to as the photodepletion henceforth.

Less than 10% of photodepletion can lead to a bias in brightness of over 100%. This bias is problematic, because it obscures the correct interpretation of protein interaction from brightness data. We found that the impact of photodepletion on brightness depends strongly on the concentration of the fluorescently labeled protein. A simple model explains the brightness bias and identifies experimental conditions where photodepletion is of concern. We further describe segmented brightness analysis to effectively eliminate the influence of photodepletion on brightness data and verify it experimentally using a monomeric fluorescent protein (EGFP) and its tandem dimeric construct (EGFP$_2$).

Closer inspection of the data over a wide range of photodepletion fractions for EGFP revealed the presence of photoconversion in addition to photobleaching. The photoconversion leads to a significantly reduced brightness state of the EGFP protein. Although the presence of photoconversion complicates quantitative fluorescence experiments, we demonstrate that under most conditions its effect on brightness experiments is negligible. Thus, segmented brightness analysis offers a robust method to investigate protein interactions in the presence of photodepletion.

Our results reveal that brightness experiments in small sample compartments are vulnerable to photodepletion artifacts. The decay in the fluorescence intensity violates the implicit assumption of a stationary signal that forms the basis of conventional FFS theory.

We broaden FFS theory by rigorously including the non-stationary photodepletion process. This enhanced formulation of FFS provides a framework for brightness experiments not only in yeast cells, but also in other small compartments, such as cellular organelles or bacterial cells, and extends the reach of brightness experiments significantly.

3.2 Materials and Methods

3.2.1 Experimental setup.

The instrumentation is described in section 2.1. Single-color FFS measurements on EGFP samples were conducted at either 1000 nm or 905 nm with an excitation power of ~1 mW as measured at the objective. Alexa-488 solution was measured with the same power at a wavelength of 900 nm. Z-scan intensity profile measurements used an arbitrary waveform generator (33522A, Agilent Technologies, Santa Clara, CA) running a linear ramp signal with a frequency of 200 mHz and a peak-to-peak amplitude of 1.0 V. This voltage corresponds to an axial travel distance of ~10.0 μm. One-photon photobleaching was conducted with a FluoArc mercury lamp (Zeiss) run between 80% and 100% power with light filtered by a (450 – 490 nm) optical bandpass filter (Chroma Technology). The spectrum of EGFP was measured with an Acton SP-2150i spectrograph (Princeton Instruments, Acton, MA) connected to an iXon 897 camera (Andor Technology, Belfast, UK).

3.2.2 Microdroplets.

A volume of 100 μL of Alexa-488 (Molecular Probes, Eugene, OR) dissolved in water was combined with 900 μL of silicon oil (Fisher Scientific, Fairlawn, NJ), pipetted for 5 seconds and then vortexed for 20 seconds. The emulsion was allowed to stand for

three minutes while the larger droplets settle before removing a few µL from the top and transferring it onto a glass slide. A coverslip was pressed down on top and affixed at the corners with nail polish. FFS data were collected in the presence of photodepletion after focusing the two-photon focal spot at the center of the droplet.

3.2.3 Yeast expression vector, cell line, sample preparation, and experimental protocol.

Yeast strain of the EGFP vector, derived from the base S288C, was grown in a synthetic medium containing 2% raffinose overnight at ~23 C°. For the expression of EGFP, galactose was added to the yeast culture (~2% final concentration) when the optical density (OD) was about 0.4~0.5 at 600 nm. When the OD reached 0.7~0.8, the yeast culture was spun down (3000G, 30s) and resuspended with fresh synthetic medium. After repeating this step twice the concentration of yeast cells was concentrated 5X through resuspension in a reduced volume of medium. The concentrated yeast medium was mixed with low-temperature agar (1% final concentration) at ~30 C°, and 2 µl of the mixture was pipetted on a microscope slide containing 5-µm microspheres that act as a spacer. The microscope slide was covered with a cover slip, and the slide's borders were sealed with nail polish. Sample preparation and culturing of yeast strain 3165 (described in [83]) expressing the dimeric construct $EGFP_2$ was identical to the procedure above except that cells were grown in synthetic medium containing 2% glucose. Yeast cells were identified in bright field microscopy. We carefully selected a measurement position that avoided the nucleus and vacuoles, and took a z-scan measurement. Following that, FFS data were collected with a stationary beam focused into the cell.

3.2.4 Mammalian expression vectors, cell Lines, sample preparation, and experimental protocol.

U2OS, COS-1, MRC-5 and CV-1 cells (American Type Culture Collection, Manassas, VA) were transfected with either an EGFP-C1 plasmid or a tandem dimeric EGFP (EGFP$_2$) plasmid as described previously [4]. These mammalian cells were maintained in a mixture of DMEM medium and 10% fetal bovine serum (Hyclone Laboratories, Logan, UT). U2OS, CV-1, COS-1 and MRC-5 cells were transfected using TransFectin reagent (Bio-Rad, Hercules, CA) according to the manufacturer's instructions 24 hours before measurement. All cells were subcultured into eight-well coverglass chamber slides (Nalge Nunc International, Rochester, NY) with the media replaced by Leibovitz L15 medium (Gibco, Auckland, NZ) immediately before measurement. FFS measurements on cells were performed as previously described [29]. For photodepletion experiments, cells were exposed repeatedly for short time intervals to epifluorescence light. After each exposure the instrument performed a short two-photon FFS measurement to record the brightness and the photodepletion fraction.

3.2.5 Data analysis.

Photon count data collected at a frequency of 20 kHz were analyzed with code written in IDL 8.0 (Research Systems, Boulder, CO). The brightness $\bar{\lambda}$ of the sample was determined by photon counting histogram (PCH) analysis and photon count moment analysis, in which deadtime and afterpulsing effects were corrected as previously described [24,74,76,80,84]. The brightness was further corrected for the finite thickness of the cell by analysis of the z-scan intensity profile [63]. We measured the brightness $\bar{\lambda}_{EGFP}$ of EGFP either in water or in the nucleus of U2OS cells to establish a reference

brightness for the fluorescent label. We use a bar over the symbol to stress that brightness is calculated from a time-average and not an ensemble average. The standard deviation (SD) of the reference brightness was less than 10%. The normalized brightness of a yeast cell measurement is $\bar{b} = \bar{\lambda}/\bar{\lambda}_{\text{EGFP}}$. A dimeric protein carrying two EGFPs is represented by a normalized brightness $\bar{b} = 2$, while a monomeric protein results in $\bar{b} = 1$. The photodepletion rate coefficient k_D was determined from a fit of the intensity trace $F(t)$ to a decaying exponential function, $F_0 \exp(-k_D t)$, where F_0 is the initial intensity. The photodepletion fraction f_D was calculated from the fluorescence intensity trace by $f_D = (F_0 - F(t))/F_0$. In segmented data analysis the photon count data was sliced into segments with a time interval T_s. Brightness b_S was calculated independently for each segment. We noticed the presence of undulations in some intensity traces from yeast cells. Such data was discarded, because it likely reflects the motion of vacuoles into and out of the excitation volume or the presence of focus drift during the measurement.

3.3 Results

We performed single-color FFS experiments on yeast cells expressing EGFP by focusing the laser beam into the cytoplasm and collecting the fluorescence signal. After completing the FFS measurement an intensity z-scan was carried out to identify the thickness of the cytoplasmic layer at the measurement position as previously described [63]. The brightness $\bar{\lambda}$ of the sample was determined by PCH analysis corrected for the axial thickness at the measured location [24,84]. We converted it to the normalized brightness $\bar{b} = \bar{\lambda}/\bar{\lambda}_{\text{EGFP}}$ with the help of the reference brightness $\bar{\lambda}_{\text{EGFP}}$. For convenience,

we hereafter refer to normalized brightness simply as brightness. The analysis also identified the number of EGFP molecules in the optical observation volume, which was converted into a molar concentration. Since the amount of expressed EGFP differed between cells, repeating the experiment on many different yeast cells established the brightness over a wide concentration range. The result of this experiment, shown in Figure 3.1A, revealed an unexpected finding. While the brightness $\bar{b}$ equaled one at low concentrations, as expected for a monomeric EGFP protein, the brightness at higher concentrations scattered between one and three. Brightness values larger than one indicate association between EGFP proteins [4]. However, EGFP is known to be monomeric at or below micromolar concentrations [4,63,85], as illustrated in Figure 1B, which depicts monomeric brightness values for EGFP measured in mammalian cells under the same experimental conditions as the yeast experiment. The brightness data shown in Figure 3.1A and B were determined by PCH analysis. As an additional check we reevaluated these data with an alternative analysis method based on photon count moments [74,76,80], which returned brightness values that are within a few percent identical to the PCH generated values (Figure 3.1A).

We suspected that an unknown experimental artifact was responsible for the difference in results between yeast and U2OS cells. Close inspection of the data revealed a small decrease ($\leq 10\%$) in the intensity over the 30-second measurement period, which did not occur in the mammalian cell measurements. Because we anticipated that photobleaching plays a functional role in distorting brightness, we performed another set of yeast experiments with much longer data acquisition times. The new data show a
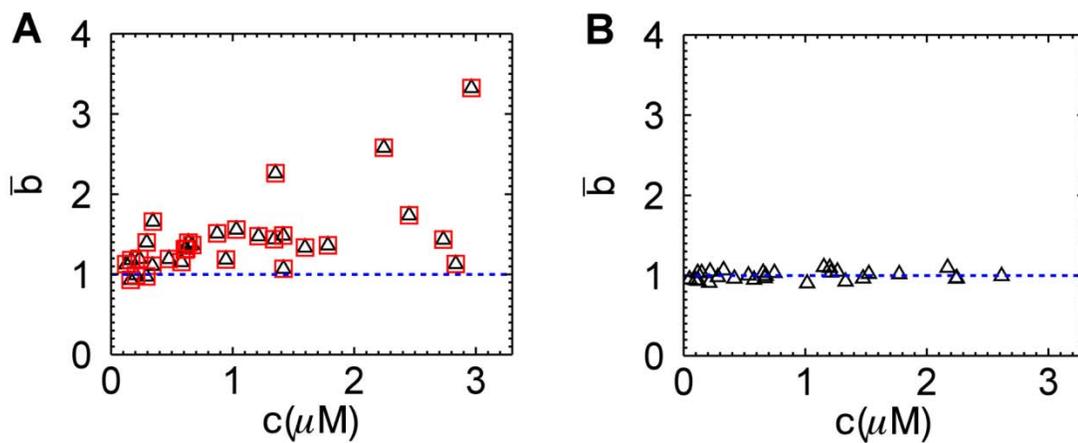
Figure 3.1 Normalized brightness of EGFP versus concentration.

(A) EGFP in yeast cells results in brightness values that scatter from 1 to 3. Brightness is analyzed by PCH analysis (triangles) and by moment analysis (squares). (B) EGFP in U2OS cells exhibits a brightness close to 1 at all concentrations as expected for a monomeric protein. The blue dashed line represents the brightness value expected for monomeric EGFP.

pronounced intensity decay with time (inset of Figure 3.2A). Fitting an exponential decay

of the form $F(t)/F_0 = \exp(-k_D t)$ to the intensity traces determined the depletion rate

coefficient k_D, where F_0 is the fluorescence intensity at the start of the experiment. The

photodepletion rate of yeast cells varied (inset of Figure 3.2A). Larger yeast cells had a

lower depletion rate than small cells, because it takes longer to deplete fluorophores in a

large reservoir than in a small one. Normalizing the fluorescence intensity trace to an

amplitude of one and a rescaled time with respect to the depletion rate coefficient k_D

mapped all intensity traces to the same functional shape (Figure 3.2A). The relative

decrease in fluorescence intensity is characterized by the photodepletion fraction

$f_D = (F_0 - F(t))/F_0$. We calculated brightness $\bar{b}(f_D)$ from fluorescence data as a

function of the photodepletion fraction f_D by truncating the fluorescence data at the

point where the relative fluorescence decrease equaled the desired photodepletion

fraction f_D. The brightness $\bar{b}(0.2)$ calculated from yeast data truncated at a

photodepletion fraction of 0.2 depended strongly on the initial fluorescence intensity F_0

(Figure 3.2B, squares). If we instead calculated the brightness $\bar{b}(0.1)$ from a shorter

segment of the data, so that the photodepletion fraction is only 0.1, we still observed a

strong dependence of brightness on intensity (Figure 3.2B, triangles), but it was less

pronounced than for the case $f_D = 0.2$.

The above observation demonstrated a link between the observed brightness bias

and photodepletion. Larger photodepletion led to a stronger bias as seen in Figure 3.2B.
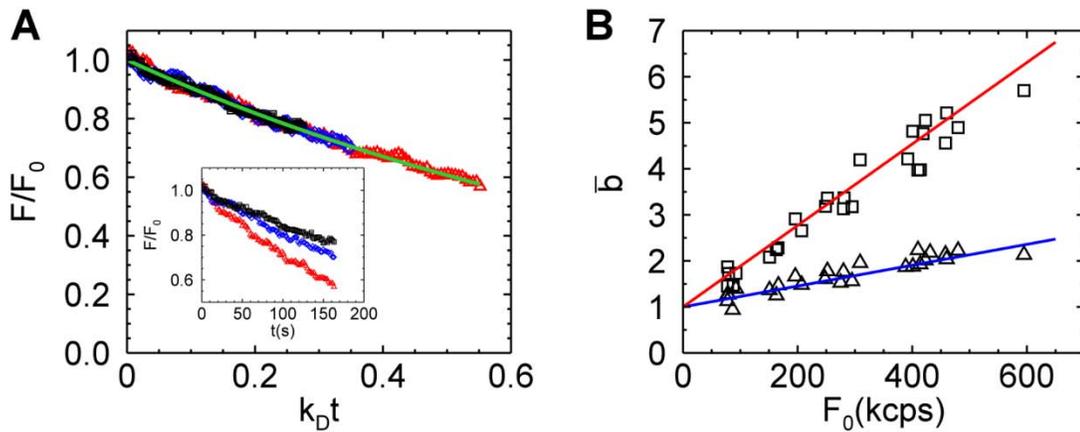
Figure 3.2 Fluorescent intensity decays in yeast cells and its effect on brightness values.

(A) The fluorescent intensity decay from three different cell experiments (black, blue, and red symbols) is graphed versus the scaled time $k_D t$. An exponential decay function $\exp(-k_D t)$ (green line) describes the experimental fluorescent intensity curves. The decay rate coefficient k_D of the three cells differ (inset). (B) Brightness versus the initial intensity for data with a photodepletion fraction $f_D = 0.1$ (triangles) and $f_D = 0.2$ (squares). Modeling by Eq. 3.4 with photodepletion fractions of 0.2 and 0.1 is shown as the red and blue solid line, respectively.

However, even if the photodepletion fraction was kept the same, the bias was not constant, but depended on intensity. At low intensities the bias was almost negligible, while substantial at high intensities even for depletion fractions as low as 0.1. Thus, identifying the presence of photodepletion bias in brightness experiments seems important in order to avoid misinterpretation of data. Conventional FFS theory cannot predict the magnitude of this artifact, because it assumes a stationary fluorescence signal, which is violated in the presence of photodepletion. To account for the non-stationary signal in brightness calculations we consider a single photobleaching step converting the fluorescent protein from a fluorescent state F to a non-fluorescent dark state D with a rate coefficient that depends on the fluorophore and excitation light. Such a process leads to an exponential decay of the fluorescence intensity with time. At each time point t' along the intensity trace a well-defined ensemble-averaged moment of the fluorescence intensity exists. The first and second ensemble-averaged moments are $\langle F(t') \rangle = F_0 \exp(-k_D t')$ and $\langle \Delta F^2(t') \rangle = \gamma_2 \lambda \langle F(t') \rangle$, which utilizes the relationship between brightness and the first two intensity moments, $\lambda = \langle \Delta F^2 \rangle / (\gamma_2 \langle F \rangle)$ [24,86]. The FFS experiment determines time-averaged moments with the first time-averaged moment $\bar{F}(t)$ given by

$$\bar{F}(t) = \frac{1}{t} \int_0^t \langle F(t') \rangle \, dt' = F_0 \frac{1 - \exp(-k_D t)}{k_D t}, \qquad (3.1)$$

where t represents the measurement time. All properties based on time-averaged moments will be denoted by a bar over the symbol. Applying the same procedure to the second central moment leads to

$$\overline{\Delta F^2}(t) = \gamma_2 \lambda \overline{F}(t) + F_0 \cdot \overline{F}(t) \left(\frac{1 + \exp(-k_D t)}{2} - \frac{1 - \exp(-k_D t)}{k_D t} \right). \tag{3.2}$$

A detailed derivation of Eqs. 3.2 to 3.4 is found in section 3.5.1. It is convenient to rewrite the above equations in terms of the photodepletion fraction f_D,

$$\overline{F}(f_D) = -F_0 f_D / \ln(1 - f_D)$$
$$\overline{\Delta F^2}(f_D) = \left[\gamma_2 \lambda + F_0 \left(\frac{2 - f_D}{2} + \frac{f_D}{\ln(1 - f_D)} \right) \right] \overline{F}(f_D). \tag{3.3}$$

The time-averaged brightness $\bar{\lambda}_D(f_D)$ in the presence of photodepletion is determined by $\overline{\Delta F^2} / (\gamma_2 \overline{F})$ as

$$\bar{\lambda}_D(f_D) = \lambda + \frac{F_0}{\gamma_2} \left(\frac{2 - f_D}{2} + \frac{f_D}{\ln(1 - f_D)} \right). \tag{3.4}$$

The above equation demonstrates that the time-averaged brightness $\bar{\lambda}_D$ is larger than the ensemble-averaged brightness λ, if photodepletion is present.

Eq. 3.4 was tested using microdroplets containing Alexa488 solution embedded in silicon oil. Individual droplets were measured for a long enough time period to achieve photodepletion fractions in excess of 80%. Data were analyzed by systematically truncating the data at different lengths to vary the photodepletion fraction f_D continuously. The brightness $\bar{\lambda}_D(f_D)$ was divided by the reference brightness $\lambda_{Alexa488}$ obtained from a measurement of a dye solution to get the normalized brightness $\bar{b}_D(f_D)$. Figure 3.3A shows the brightness $\bar{b}_D(f_D)$ from two droplets, one containing a high

concentration of dye and the other containing a very low concentration of dye, as a function of f_D, together with their respective fits to Eq. 3.4. The agreement between data and fits validated the simple model in an aqueous solution environment even for brightness biases as large as several hundred percent. We also tested the photodepletion model on cellular data by reexamining the brightness data from yeast cells shown in Figure 3.2B. Dividing Eq. 3.4 by the reference brightness λ_{EGFP} determined the time-averaged normalized brightness $\overline{b}_D(f_D)$. We plot $\overline{b}_D(f_D)$ using photodepletion fractions of 0.2 and 0.1 as solid lines in Figure 3.2B and achieved excellent agreement with the experimental data.

After establishing the validity of the simple photodepletion model we investigated the influence of concentration on the brightness bias. The data in Figure 3.3A demonstrate that the sample with $N_0 = 16$ dye molecules in the observation volume was very susceptible to photodepletion, while at the single molecule level ($N_0 = 0.69$), the bias was only noticeable at very high photodepletion fractions. N_0 describes the initial number of fluorescence molecules in the optical observation volume before photodepletion occurred. This dependence on concentration is also predicted by Eq. 3.4, because the initial intensity is proportional to the number of molecules, $F_0 = \lambda N_0$. The relative error or bias can be written as,

$$e(f_D) = \frac{\overline{\lambda}_D(f_D) - \lambda}{\lambda} = \frac{1}{\gamma_2}\left(\frac{2 - f_D}{2} + \frac{f_D}{\ln(1 - f_D)}\right) N_0. \tag{3.5}$$
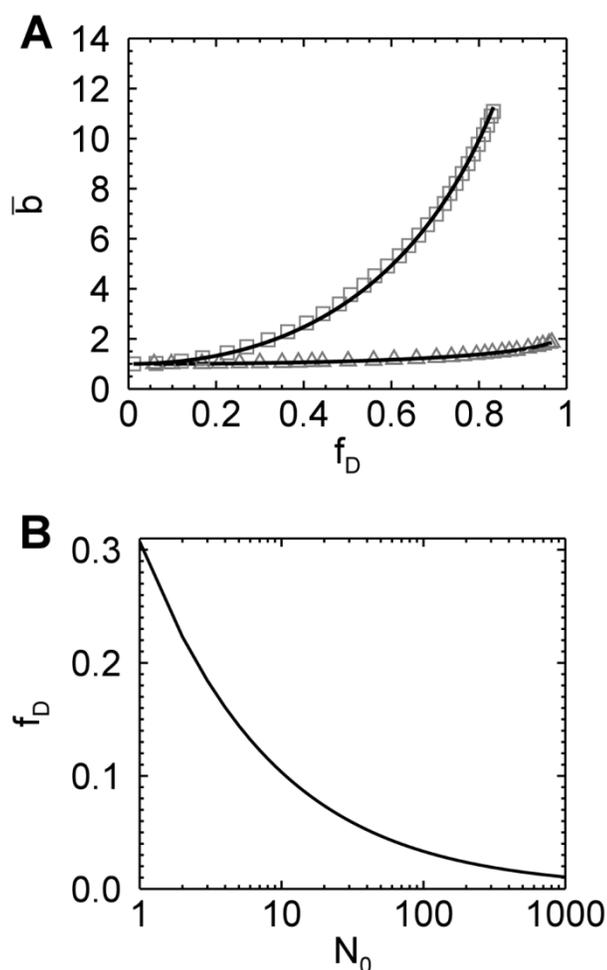
Figure 3.3 Time-averaged brightness bias.

(A) The time-averaged brightness of Alexa488 as a function of the photodepletion fraction f_D as measured in a microdroplet at a high (squares) and a low (triangles) concentration. The increase in brightness with f_D is an artifact caused by photodepletion. The solid lines represent the fit of the data to Eq. 3.4, which resulted in N_0 of 16 and 0.69 for the high and low concentration data, respectively. (B) The relationship between photodepletion fraction f_D and the initial number of fluorescent molecules N_0 in the optical observation volume that result in a brightness error of 5%.

This formula is very convenient for estimating the potential bias of a brightness experiment. The initial number of molecules and the photodepletion fraction are the only factors needed to estimate the bias. Since the experimental uncertainty of brightness experiments is ~10%, it is reasonable to set a bias limit that is half of the experimental uncertainty to ensure the absence of noticeable artifacts in brightness data. Eq. 3.5 was solved numerically for $e = 5\%$ to determine the limiting photodepletion fraction f_D as a function of the initial number concentrations N_0 (Figure 3.3B). The photodepletion fraction that guarantees a bias of $\leq 5\%$ decreases with increasing concentration N_0. Because fluorescence fluctuation experiments at concentrations higher that $N_0 = 1000$ are rarely feasible, a photodepletion fraction of $\leq 1\%$ guarantees that brightness experiments in cells are free of the photodepletion artifact (Figure 3.3B). For reference, the highest concentration measured in this study is $N_0 \sim 600$. Since the fastest photodepletion rate coefficient obtained from the yeast cells is ~0.006 s^{-1}, a data segment length of ~1.6 s guarantees a photodepletion fraction of $\leq 1\%$.

Thus, it seems that dividing the data into sufficiently short segments provides a simple remedy to avoid artifacts due to photodepletion. However, photodepletion not only affects brightness through the introduction of a non-stationary signal, but also alters the brightness of oligomeric protein complexes. This is readily demonstrated by taking a closer look at the photobleaching process (Figure 3.4A). We assumed a simple model wherein a fluorescent protein F with normalized brightness $b = 1$ is irreversibly converted into a non-fluorescent state D with brightness zero. Consider first the case of a monomeric protein F. Photobleaching leads to two populations, F and D. Only state F
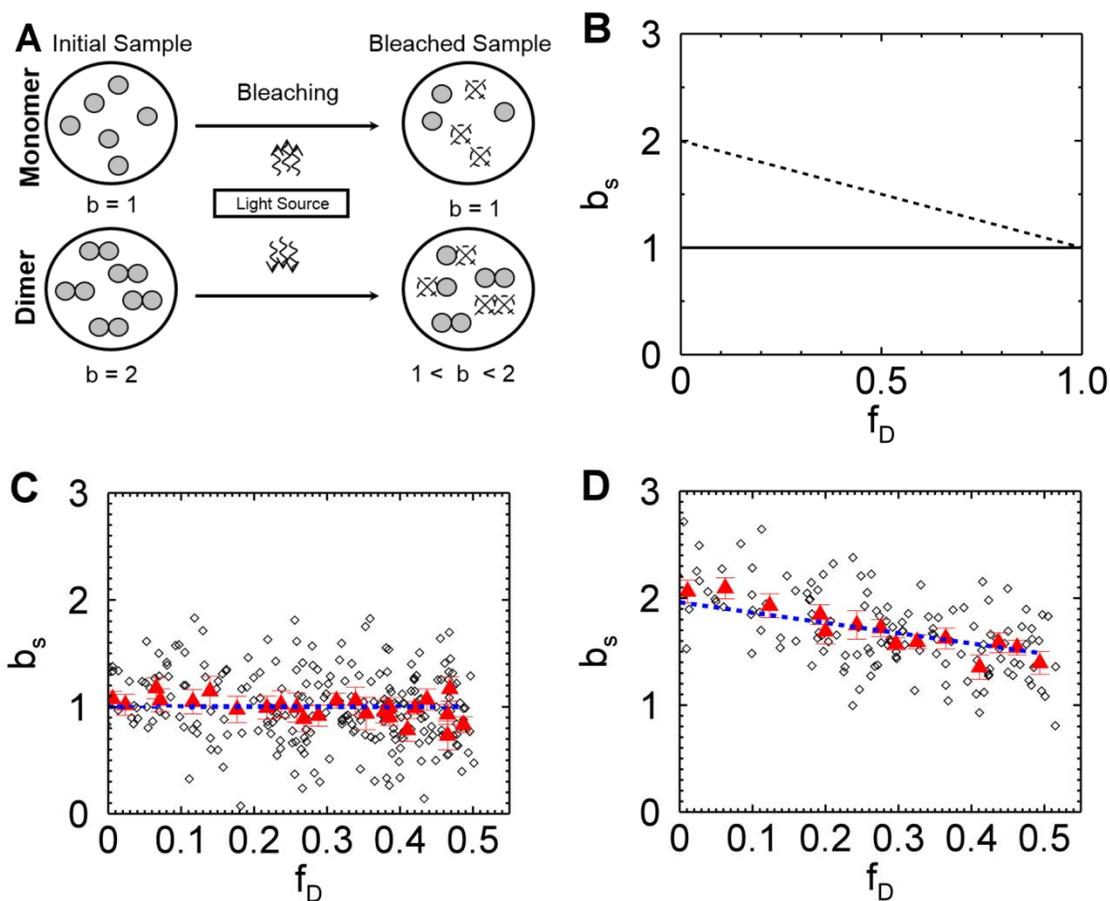
Figure 3.4 Segmented brightness analysis of monomers and dimers.

(A) Illustration of photodepletion for monomers and dimers. Fluorescent molecules are depicted as filled circles and photobleached molecules are pictured as broken circles. The normalized brightness of monomers ($b = 1$) remains unchanged by photodepletion, In contrast, photobleaching of a dimeric sample with initial brightness of 2 leads to a reduction of brightness as explained in the text. (B) Theoretical brightness based on segmentation analysis of a monomeric (solid line) and dimeric (dashed line) sample as a function of photodepletion. (C) Brightness of EGFP from a yeast cell by segmentation analysis versus photodepletion fraction. Brightness values (diamonds) for a data segment size of 1.6 s. Ten consecutive brightness values are averaged (red triangles) to better visualize the trend of the data. The dashed blue line represents a fit of the brightness values to Eq. 3.6 with a fitted value of $b_0 = 1.00 \pm 0.03$. (D) Brightness of EGFP$_2$ from a yeast cell by segmentation analysis versus photodepletion fraction. Symbols are described under (C). The blue dashed line represents a fit of the brightness values to Eq. 3.6 with a fitted value of $b_0 = 1.96 \pm 0.03$.

contributes to the fluorescence signal. Because each protein in state F has the same brightness, photobleaching has no effect on the brightness of the sample. A population of dimers F_2, on the other hand, initially has a normalized brightness of $b_0 = 2$. Photobleaching introduces three distinct species that differ in their brightness (Figure 3.4A). If both fluorophores of the dimer are photobleached (state D_2), then the complex is dark with a brightness of zero. If both fluorphores survive (state F_2), the brightness of the complex remains that of a dimer. If one of the two fluorophores survives (state FD), the complex has a brightness of 1. This mixture of brightness states leads to an apparent brightness between 1 and 2. Since the population of states FD and D_2 increases with time, the brightness of the dimer decreases in the presence of photodepletion [4].

Let us explicitly treat the case of a population of n-mers F_n. Bleaching of s chromophores leads to the species $F_{n-s}D_s$ with normalized brightness $b_{n-s} = n - s$. Since photobleaching of individual chromophores is statistically independent, the probability p of a single chromophore to be bleached equals the photodepletion fraction, $p = f_D$. Thus, the probability of an n-mer to be in state $F_{n-s}D_s$ is given by $p_{n-s} = \binom{n}{s} p^s (1-p)^{n-s}$. The sample consists of a mixture of populations $F_{n-s}D_s$, which leads to an average or apparent normalized brightness of [4]

$$b_S\left(f_D\right) = \frac{\sum_{s=0}^{n} b_{n-s}^2 p_{n-s}}{\sum_{s=0}^{n} b_{n-s} p_{n-s}} = b_0 - \left(b_0 - 1\right) f_D \qquad (3.6)$$

where b_0 represents the initial normalized brightness of the sample in the absence of photodepletion. Note that Eq. 3.6 specifies the brightness b_S from a short data segment of a sample with a photodepletion fraction f_D. For an n-mer the initial normalized brightness is $b_0 = n$. The behavior of Eq. 3.6 is illustrated in Figure 3.4B for a dimer $(b_0 = 2)$ and monomer $(b_0 = 1)$ sample. The normalized brightness starts at a value of b_0 in the absence of photodepletion $(f_D = 0)$ and decreases linearly to a value of one as the photodepletion fraction approaches one. This result reflects that the last surviving fluorescent population of an n-mer is F_1D_{n-1} with exactly one remaining fluorophore, which implies a normalized brightness of one. Of course, Eq. 3.6 also predicts that the brightness of a monomeric protein sample stays constant as discussed earlier. While we derived Eq. 3.6 for a homogenous sample of n-mers, it is straightforward to show that the equation remains correct for a mixture of oligomeric states with b_0 representing the apparent brightness of the mixture. Note that we removed the bar over the brightness symbol to emphasize that the segmented brightness equals the ensemble-averaged brightness.

We performed segmented brightness analysis on data taken from a yeast cell expressing EGFP. The brightness of each segment is graphed as a function of the photodepletion fraction (Figure 3.4C). The brightness values showed significant scatter reflecting the poor statistics due to the short segment size of 1.6 s. We also graph the brightness averaged over 10 segments, which reduces the scatter and aids in visualizing data trends. The initial brightness of the cell was determined by a fit of the segmented

brightness values with Eq. 3.6. The fit (dashed line, Figure 3.4C) resulted in a brightness

of $b_0 = 1.00 \pm 0.03$ (reduced Chi-squared = 1.0), as expected for monomer EGFP. Next,

we examined a dimeric fluorescent protein by expressing the tandem construct $EGFP_2$ in

yeast cells. The data was subjected to the same analysis as described above. The

segmented brightness appeared to diminish with depletion fraction (Figure 3.4D), which

is a trend predicted by the model (Figure 3.4A & B). The dashed line describes the fit of

the data to Eq. 3.6 with an initial brightness $b_0 = 1.96 \pm 0.03$ (reduced Chi-squared = 1.1),

which is consistent with dimeric EGFP.

While fitting of the segmented brightness values by Eq. 3.6 is feasible, there is a

simpler alternative. The average of all segmented brightness values of the experiment,

$\langle b_S \rangle = f_D^{-1} \int_{x=0}^{f_D} b_S(x)dx$ is, according to Eq. 3.6, related to the initial brightness,

$$ b_0 = \frac{\langle b_S \rangle}{1 - f_D/2} - \frac{f_D/2}{1 - f_D/2}. \tag{3.7} $$

We used Eq. 3.7 to compute the initial brightness b_0 for the data shown in Figure 3.4C &

D, which yielded 0.98 for EGFP and 1.93 for $EGFP_2$. These values agree with the results

from the earlier analysis based on Eq. 3.6. However, Eq. 3.7 is more convenient, because

no fitting is required.

We performed segmented brightness analysis on the yeast data previously shown

in Figure 3.1A using a segment length of 1.6 s. The initial brightness b_0 was calculated

with Eq. 3.7 to determine the initial brightness (Figure 3.5). We see that the new analysis

successfully removed the earlier brightness scatter (Figure 3.1A) and produced a
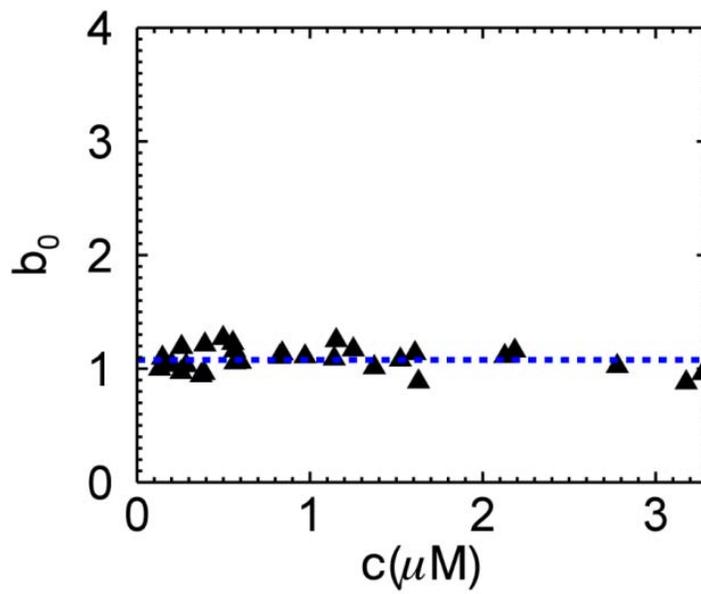
Figure 3.5 Normalized brightness of EGFP versus concentration in yeast cells.

The same data shown in Figure 3.1A are reanalyzed with segmented brightness analysis, which removes the bias (mean and SD of brightness data: .1.08 ± 0.10).

brightness b_0 close to one (mean and SD: 1.08 ± 0.10), which were in good agreement with the result obtained for mammalian cells in Figure 3.1B (mean and SD: 1.01 ± 0.06).

While the above results demonstrate that quantitative brightness analysis in the presence of photodepeletion is feasible, we have not yet examined the range of photodepletion fractions covered by our model. This is especially important for EGFP, since the photophysics of GFP-like proteins is remarkably complex [87,88], while our model is based on a single photobleaching step. To address this question we performed extended photodepletion experiments both in mammalian and yeast cells as described in the Materials and Methods section to achieve photodepletion fractions in excess of 80%. The segmented brightness of several mammalian cells expressing EGFP is graphed as a function of the photodepletion fraction (Figure 3.6A). The segmented brightness initially remained at one, as expected for a monomer (Eq. 3.6). However, once the photobleaching fractions exceeded 60% a decrease in the segmented brightness is noted. This result indicates that our bleaching model is too simplistic. Analogous photodepletion experiments were also performed on mammalian cells expressing the tandem construct EGFP$_2$. The segmented brightness values closely followed the curve (blue line) expected for a dimer for $f_D \leq 60\%$ (Figure 3.6B), but was falling off faster than predicted by theory for photodepletion fractions exceeding 60%.

We further conducted extended photodepletion experiments on yeast cells to identify whether the unexpected brightness behavior at high photodepletion fractions of mammalian cells was also found in yeast cells. Figure 3.6C shows the averaged segmented brightness (red circles) from several yeast cells expressing EGFP together
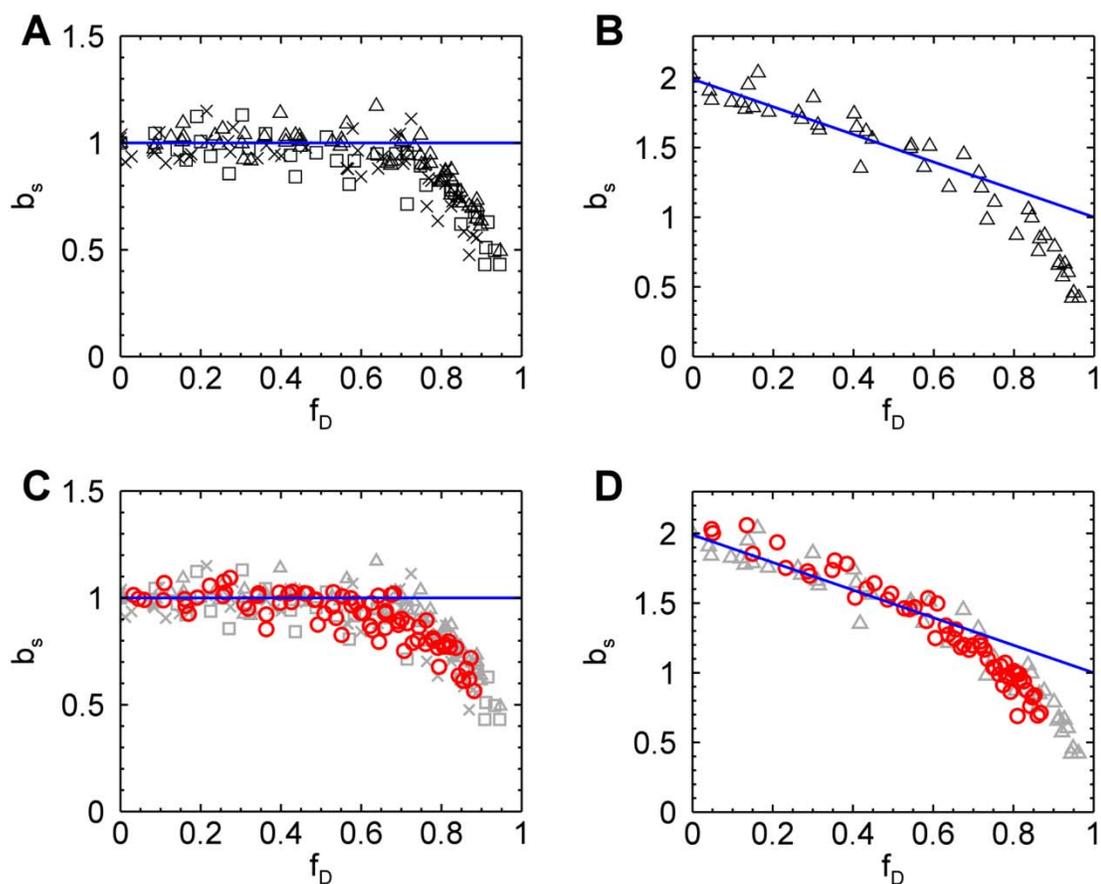
Figure 3.6 Segmented brightness of EGFP and EGFP$_2$ in mammalian and yeast cells versus photodepletion fraction.

Photobleaching of mammalian cells was accomplished by one-photon excitation, while yeast cells were photobleached by two-photon excitation. (A) The normalized brightness of EGFP in mammalian cells (five COS cells (triangles), five CV-1 cells (squares), six MRC5 cells (crosses)) is close to the theoretical value (blue line) until the photodepletion fraction exceeds 0.6, at which point it begins to drop. (B) The normalized brightness of EGFP$_2$ in mammalian cells agrees with theory (blue line) until the photodepletion fraction reaches 0.6. (C) The normalized brightness of EGFP from three yeast cells (red circles) shows the same behavior as seen in mammalian cells (grayed symbols, same as shown in panel A). (D) The normalized brightness of EGFP$_2$ from two yeast cells (red circles) close follows the brightness observed in mammalian cells (grayed symbols, same as shown in panel B).

with the earlier results obtained from mammalian cells. The corresponding data (red circles) from yeast cells expressing EGFP$_2$ are graphed together with the results from mammalian cells in Figure 3.6D. We observed an identical response of segmented brightness with photodepletion fraction for yeast and mammalian cells.

Our model is based on a photobleaching reaction from a single bright to a non-fluorescent state, $F \rightarrow D$. The data show that the brightness behavior of EGFP was well approximated by this simple model provided the photodepletion fraction was less than 60%. However, the decrease in brightness of monomeric EGFP for $f_D > 0.6$, necessitates the appearance of a second brightness state, which we denote as F^*. This new state is likely populated by a photoconversion process, as explained later, and has to be less bright than the original state F to explain the drop in brightness observed in the data.

Photoconversion of EGFP from a green to a red fluorescent state by an electron transfer process has been recently reported [89,90]. We measured the fluorescence emission spectrum of a U2OS cell before and after photobleaching to identify whether the appearance of state F^* is associated with a strong shift in the emission spectrum towards the red. The emission spectrum after strong photodepletion ($f_D = 0.87$) was virtually identical to the emission spectrum of the unbleached sample (Figure 3.7). Because state F and F^* are both green fluorescent states, the observed photoconversion process is distinct from the reddening of EGFP. We also measured the fluorescence lifetime in the absence ($f_D = 0$) and presence ($f_D = 0.77$) of photodepletion in U2OS cells (see section 3.5.3). The time-resolved fluorescence intensity decay is close to a mono-
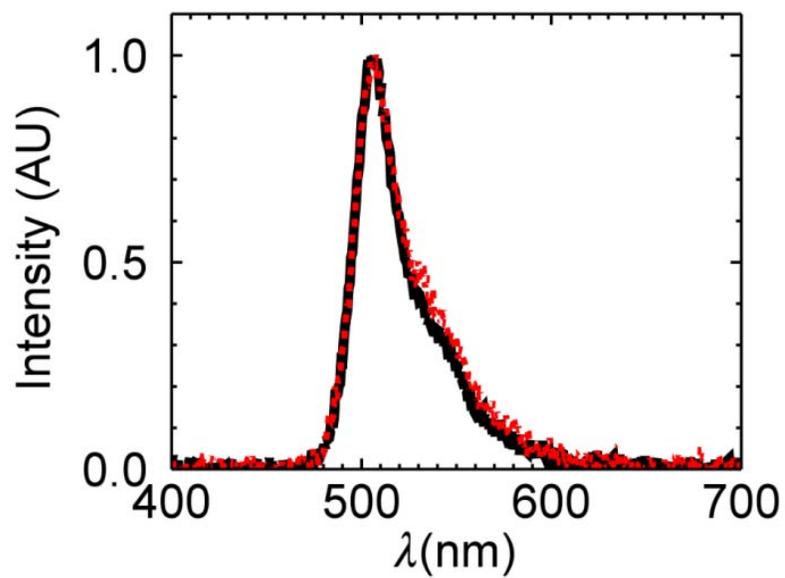
Figure 3.7 Fluorescence emission spectrum of EGFP before and after photodepletion.

The initial spectrum (black line) before photodepletion is virtually identical to the spectrum (red line) taken at a photodepletion fraction of 0.87. Both spectra have been scaled to a maximum amplitude of one to facilitate visual comparison.
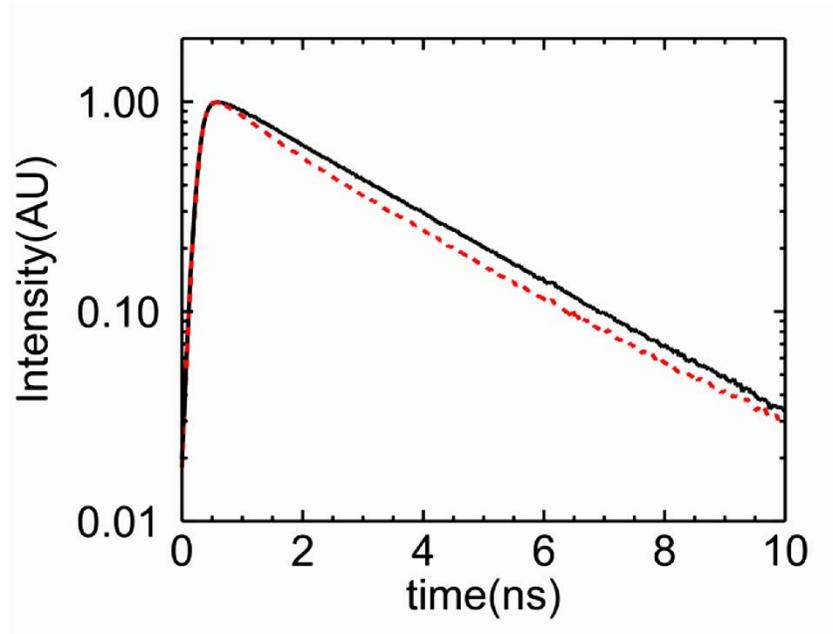
Figure 3.8 Time-resolved fluorescence decay curve of EGFP in the presence and absence of photodepletion.

The time-resolved fluorescence decay curves of EGFP in the absence of photodepletion ($f_D = 0$, black line) and in the presence of photodepletion ($f_D = 0.77$, red line) are normalized to an amplitude of one to facilitate visual comparison. The decay in the absence of photodepletion is approximately described by a single exponential decay with a lifetime of ~2.6 ns. The fluorescence decay in the presence of photodepletion (red line) exhibits an initial fast decay, which reveals the presence of a short lifetime component. The slope of the decay at later times is roughly equal to the slope of the original decay curve (black line), indicting the presence of a second lifetime component of ~2.6 ns.

exponential for $f_D = 0$, while photodepletion leads to the appearance of a second, shorter lifetime component, which is responsible for the initial faster decay of the intensity trace (Figure 3.8). The change in the time-resolved fluorescence decay with photodepletion supports the existence of a photoconverted EGFP state as suggested by the brightness experiments.

3.4 Discussion

Because correcting photobleaching effects is not straightforward, most FFS experiments use conditions where the probability of a fluorophore being photobleached is sufficiently small to not directly distort the statistics of the fluctuations. The occasional photobleaching event still reduces the number of fluorophores in the sample. This reduction has a negligible effect on concentration for sufficiently large sample reservoirs. However, this is not the case for small sample volumes, where the cumulative effect of photobleaching leads to a measurable reduction in the concentration of fluorophores over the measurement period. The budding yeast *Saccharomyces cerevisiae* is spherical-shaped with a diameter of ~5 μm. Its volume of ~60 fL is significantly smaller than that of a mammalian cell with a volume of a few pL. The data in Figure 3.1 demonstrate that photodepletion which is negligible for mammalian cells, cannot be ignored for budding yeast cells under identical experimental conditions. Photodepletion effects are also variable within a given cell population. While the median cell size of budding yeast depends on species and strain, there is significant variability in cell size in any given population of yeast, which gives rise to the differences in the observed photodepletion rates (inset, Figure 3.2A).

The analysis of FCS and related techniques is based on a stationary fluorescence signal, which is violated in the presence of photodepletion. We developed a model that explicitly takes the non-stationary signal due to the depletion of fluorophores into account. This model explains the observed brightness bias of conventional analysis (Figure 3.2B and Figure 3.3A). It further predicts the linear relation between bias and initial sample concentration (Eq. 3.5), which explains the absence of significant brightness error at low concentration. For example, a photodepletion fraction of 10% leads to a bias of less than 20% for concentration < 200 nM with a focal volume of 0.2 fL. These conditions were met in an earlier study reporting the first brightness measurements in *S. cerevisiae* [83]. However, the analysis bias is not negligible at higher concentrations, and artificially increased brightness values would lead to an erroneous conclusion about protein complex formation.

What factors are responsible for the observed increase in brightness when photodepletion occurs? Conventional theory states that for a stationary process the brightness is proportional to the ratio of variance to mean of the fluorescence, $\lambda \propto \langle \Delta F^2 \rangle / \langle F \rangle$ [24,86]. Photodepletion increases $\langle \Delta F^2 \rangle$, because the resulting intensity decrease constitutes an extra variation of the signal that is added to the intensity variations caused by fluorescent proteins diffusing in and out of the observation volume. By the same token, photodepletion decreases the mean fluorescence $\langle F \rangle$. Both factors increase the ratio $\langle \Delta F^2 \rangle / \langle F \rangle$, which explains the observed inflation of the brightness value.

Segmentation of the photon count data provided an effective strategy to eliminate biases due to photodepletion. The appropriate segment size is calculated using the photodepletion rate and Eq. 3.5. Because the segment size is short, the scatter in brightness is large (Figure 3.4C and D), and further data processing is necessary to identify the initial brightness of the sample. The segmented brightness b_S decreases linearly as a function of the photodepletion fraction f_D with a slope that depends on the initial brightness (Eq. 3.6). While fitting of the slope provides the unbiased brightness, we prefer to directly compute the initial brightness from Eq. 3.7. Applying this method to the measurements of EGFP in *S. cerevisiae* eliminates the scatter seen in Figure 3.1A and yields the expected brightness for monomeric EGFP at all concentrations (Figure 3.5). The standard deviation of segmented brightness analysis is ~10% and represents a significant improvement over the uncertainty of previously reported brightness data in yeast [83,91]. This value is close to the standard deviation achieved in mammalian cells.

The method introduced in this paper determines the maximum segment length that guarantees a relative brightness bias of e or less (for a brief summary of the protocol see section 3.5.2). There also is a minimum length requirement, because enough fluctuations need to be sampled during a single segment to ensure a meaningful calculation of brightness. Based on our experience 100 independent fluctuations are sufficient to provide enough sampling for determining brightness. The diffusion time is a measure of the duration of a fluctuation. Since the number concentration N of experiments in cells is larger than one, a segment time of 100 diffusion times ensures the sampling of 100 independent fluctuations. Because the diffusion time of soluble proteins in cells is typically a few milliseconds, we estimate a minimum segment time of a few

hundred milliseconds. Note that our analysis method is not applicable if the minimum segment time exceeds the maximum segment time. Such a situation may arise for slowly diffusing proteins, such as membrane proteins, and needs to be checked before applying segmented brightness analysis. In our case there is no concern, because a diffusion time of ~1ms for EGFP in yeast leads to a minimum segment length (~100 ms), which is much shorter than the maximum segment length of 1.6 s.

We assumed a simple photobleaching process that converts a fluorescent state F with normalized brightness 1 to a non-fluorescent state D. EGFP deviates from this simple model, because we observed a drop in the segmented brightness for $f_D > 0.6$ (Figure 3.5A). This behavior provides conclusive evidence that the state F is not the only brightness state of EGFP. For simplicity we consider just one additional state F^*. As mentioned earlier this state must have a lower brightness than state F. It has been shown that the presence of a mixture of brightness states within a fluorescent protein leads to a dimeric brightness that is less than double [28,92]. Thus, EGFP is initially well described by a single brightness state F, because we observed (Figure 3.6D) within experimental uncertainty brightness doubling for the dimeric EGFP construct [4,93]. Because the lower brightness state F^* contributes less to the overall brightness of the sample compared to state F, the drop in brightness is not observed until a significant population of the fluorescent proteins is in state F^*. Thus, the most likely explanation for the presence of a large population of F^* at large photodepletion fractions is the presence of a photoconversion process that populates state F^* in addition to the photobleaching process. This photoconversion process, however, is not associated with a change in the emission spectrum (Figure 3.7).

Because the photophysics of EGFP is complex [87,88], identifying the exact nature of the state F^* will require additional studies. However, the presence of more than one state of EGFP is supported by additional experiments. One- and two-photon photobleaching studies of EGFP have reported non-exponential decay characteristics [94,95], which support the existence of more than one state. Conversely, it has been argued that the non-exponential photobleaching kinetics might be explained by Lévy statistics [96]. However, single molecule studies have reported that occasionally two photobleaching steps are observed for single EGFP molecules [87,97]. This observation is consistent with the presence of a second brightness state of EGFP.

We would like to stress that despite the appearance of a second brightness state, our analysis with the simple bleaching model is successful as long as the photodepletion fraction is less than ~60%. Since this condition is met for virtually all FFS experiments, the segmentation method described in this paper provides a robust analysis method. Segmentation has been originally suggested to lessen the influence of cytoplasmic intensity drifts on brightness measurements [29] The same approach has been used to correct distortions in the autocorrelation functions caused by photobleaching [98]. However, the importance of data segmentation in brightness analysis has gone largely unnoticed [99]. For example, it is common to apply PCH analysis to the entire data set [4]. This paper provides the first quantitative formulation of segmented brightness analysis and a framework for future investigation of non-stationary processes by brightness. A novel aspect of this technique is the identification of a potential photoconversion process of EGFP by relying on brightness instead of emission color. The existence of a photoconversion process is further corroborated by changes in the fluorescence lifetime

of EGFP. Thus, segmented brightness analysis could prove useful for providing insights into the behavior of EGFP and other fluorescent proteins that are difficult to obtain by other methods. The properties of EGFP are of particular interest, because EGFP serves as the fluorescent tag of a vast number of cellular studies. We expect that characterization of brightness conversion processes should prove important for fluorescence-based cellular studies. For example, stepwise photobleaching experiments count the number of fluorescently-labeled subunits in a protein complex [97,100]. Photoconversion of the fluorophore into a different brightness state compromises the count statistics of the experiment.

While EGFP is relatively photostable, some other fluorescent proteins are much more photolabile. For example, photodepletion of a red fluorescent protein has been observed in two-photon FFS measurements in mammalian cells [92]. In addition, while photobleaching by two-photon excitation is strictly confined to the focal volume, photobleaching by one-photon excitation occurs also outside the focal volume, which potentially accelerates the appearance of photodepletion and its artifacts. Thus, photodepletion effects are potentially relevant not only for yeast experiments, but also for measurements in larger volumes, such as in mammalian cells.

The ability of FFS to perform brightness titrations is a powerful tool [4], but only if brightness can be correctly related to protein stoichiometry and concentration. This paper introduces a general theory for incorporating a non-stationary process into the analysis of fluorescence fluctuations. This expanded formulation of FFS was essential for the correct identification of brightness and concentration in the presence of photodepletion as demonstrated for the budding yeast *S. cerevisiae*. The new analysis

approach should also prove useful for brightness experiments in other small compartments, such as cellular organelles or bacterial cells. We expect that the modified FFS theory provides a useful framework for future investigation of protein interactions of non-stationary processes in living matter by brightness techniques.

3.5 Supplemental Materials

3.5.1 Derivation of time-averaged variance of the fluorescence intensity

The time-averaged fluorescence $\overline{F}(t)$ for a measurement starting at time 0 and ending at time t is given by Eq. 3.1. We define the fluctuation as the deviation of the instantaneous fluorescence $F(t')$ at time t' from the time-averaged fluorescence $\overline{F}(t)$, $\delta F(t') = F(t') - \overline{F}(t)$. The ensemble variance is determined from the fluctuations by $\langle \Delta F^2 \rangle = \langle \delta F(t')^2 \rangle$, which for a non-stationary process is time-dependent. The FFS experiment determines the time-averaged variance by

$$\overline{\Delta F^2}(t) = \frac{1}{t}\int_0^t \langle \delta F(t')^2 \rangle dt' = \frac{1}{t}\int_0^t \langle (F(t') - \overline{F}(t))^2 \rangle dt'. \tag{3.8}$$

To evaluate the integral we first rewrite the integrand $\langle (F(t') - \overline{F}(t))^2 \rangle$ in the form

$\langle F^2(t') \rangle - 2\langle F(t') \rangle \overline{F}(t) + \overline{F}(t)^2$ and apply the identity $\langle F(t')^2 \rangle = \langle \Delta F(t')^2 \rangle + \langle F(t') \rangle^2$.

Inserting the reformulated integrand into Eq. 3.8 yields

$$\overline{\Delta F^2}(t) = \frac{1}{t}\int_0^t \left(\langle \Delta F^2(t') \rangle + \langle F(t') \rangle^2 - 2\langle F(t') \rangle \overline{F}(t) + \overline{F}(t)^2 \right) dt'. \tag{3.9}$$

By applying the sum rule and taking all constant factors out of the integral we get

$$\overline{\Delta F^2}(t) = \frac{1}{t}\int_0^t \left\langle \Delta F^2(t') \right\rangle dt' + \frac{1}{t}\int_0^t \left\langle F(t') \right\rangle^2 dt' - \overline{F}(t)\frac{1}{t}\int_0^t \left\langle F(t') \right\rangle dt'. \tag{3.10}$$

Evaluating the integrals requires the ensemble-averaged intensity $\left\langle F(t') \right\rangle = F_0 \exp(-k_D t')$

and its variance $\left\langle \Delta F^2(t') \right\rangle = \gamma_2 \lambda \left\langle F(t') \right\rangle$ as introduced in the manuscript. The three terms

of Eq. 3.10 evaluate to

$$\frac{1}{t}\int_0^t \left\langle \Delta F^2(t') \right\rangle dt' = \frac{1}{t}\int_0^t \gamma_2 \lambda \left\langle F(t') \right\rangle dt' = \gamma_2 \lambda \overline{F}(t)$$

$$\frac{1}{t}\int_0^t \left\langle F(t') \right\rangle^2 dt' = \overline{F}(t) F_0 \frac{1+\exp(-k_D t)}{2} \tag{3.11}$$

$$\overline{F}(t)\frac{1}{t}\int_0^t \left\langle F(t') \right\rangle dt' = \overline{F}(t)^2$$

The sum of the three terms simplifies to

$$\overline{\Delta F^2}(t) = \gamma_2 \lambda \overline{F}(t) + F_0 \cdot \overline{F}(t)\left(\frac{1+\exp(-k_D t)}{2} - \frac{1-\exp(-k_D t)}{k_D t} \right), \tag{3.12}$$

which is Eq. 3.2.

It is advantageous to express $\overline{F}(t)$ and $\overline{\Delta F^2}(t)$ as a function of the

photodepletion fraction f_D by using the relation $f_D = (F_0 - F(t))/F_0 = 1 - \exp(-k_D t)$. By

inserting $k_D t = -\ln(1-f_D)$ into Eqs. 3.1 and 3.12, we recover Eq. 3.3,

$$\overline{F}(f_D) = -F_0 f_D / \ln(1-f_D)$$

$$\overline{\Delta F^2}(f_D) = \left[\gamma_2 \lambda + F_0 \left(\frac{2-f_D}{2} + \frac{f_D}{\ln(1-f_D)} \right) \right] \overline{F}(f_D). \tag{3.13}$$

The time-averaged brightness $\overline{\lambda}_D(f_D)$ is determined from the above equation by

$$\bar{\lambda}_D(f_D) = \frac{1}{\gamma_2} \frac{\overline{\Delta F^2}(f_D)}{\overline{F}(f_D)} = \lambda + \frac{F_0}{\gamma_2}\left(\frac{2-f_D}{2} + \frac{f_D}{\ln(1-f_D)}\right), \tag{3.14}$$

which corresponds to Eq. 3.4. As expected, the time-averaged brightness $\bar{\lambda}_D(f_D)$ reduces to the true brightness λ in the absence of photodepletion ($f_D = 0$) as easily verified by applying l'Hopital's rule to Eq. 3.14.

3.5.2 Protocol for Brightness Analysis in Small Sample Compartments

A prudent first step when analyzing data from a small sample compartment is testing whether photodepletion is present. The intensity trace of the experiment is fit to an exponential decay of the form $F(t) = F(0)\exp(-k_D t)$ to determine the depletion rate coefficient k_D and the photodepletion fraction $f_D = (F(0) - F(T))/F(0)$, where $F(0)$ and $F(T)$ represent the initial and final fluorescence intensity of the experiment. Next, choose the relative brightness bias e that is experimentally acceptable. We suggest a value of $e = 0.05$ as explained in the manuscript. The highest concentration that guarantees an error of e or less is determined by Eq. 3.5. Because concentration and initial fluorescence intensity are related, $F(0) = \lambda_{EGFP} N_0$, we rewrite Eq. 3.5 to determine the highest initial intensity $F_{max}(0)$ that ensures a brightness error of e or less,

$$F_{max}(0) = \gamma_2 \frac{e}{\dfrac{2-f_D}{2} + \dfrac{f_D}{\ln(1-f_D)}} \bar{\lambda}_{EGFP}, \tag{3.15}$$

where $\bar{\lambda}_{EGFP}$ is the calibrated brightness of the fluorescent protein measured in a large sample reservoir. If the initial intensity of the experiments is less than $F_{max}(0)$, then photodepletion can be safely neglected.

However, if the initial intensity $F(0)$ of the experiments exceeds $F_{max}(0)$, the data need to be segmented to reduce the photodepletion fraction from f_D to f_D^*. The value of f_D^* is determined by graphically or numerically solving the equation

$$F(0) = \gamma_2 \frac{e}{\frac{2-f_D^*}{2} + \frac{f_D^*}{\ln(1-f_D^*)}} \bar{\lambda}_{EGFP}.$$
(3.16)

The segment length is determined by the time T^* that results in a photodepletion fraction f_D^*,

$$T^* = -\frac{\ln(1-f_D^*)}{k_D}.$$
(3.17)

Segmenting the data into time intervals of length T^* followed by segmented brightness analysis as described in the manuscript provides brightness values that are immune to photodepletion.

3.5.3 Fluorescence Lifetime Measurement

Fluorescence lifetime measurements were performed on a two-photon microscope. Data were collected using the same experimental conditions as employed in FFS measurements except that the emission light passed through a polarizer set to magic angle conditions. The emission light was detected by a hybrid PMT (HPM-100-40, Becker & Hickl, Berlin, Germany) connected to time-correlated single photon counting module (TimeHarp 200, PicoQuant, Berlin, Germany). The reference timing signal of the TimeHarp module was obtained from the 80MHz clock output of the laser (Mai-Tai, Spectra Physics, Mountain View, CA) and subsequently filtered and amplified (91018

pulse conditioner, ISS Inc., Champaign, IL). The instrument response function was determined by recording second harmonic generation light obtained from urea crystals (ICN Biomedical, Aurora, OH). Experiments were performed on U2OS cells expressing EGFP. After measuring a fluorescence decay histogram in the absence of photodepletion ($f_D = 0$) the cell was photobleached using the same protocol used for brightness experiments. A second measurement of the fluorescence decay histogram was performed on the photobleached cell. The fluorescence decay histograms were analyzed using routines written in IDL 8.0 (Research Systems, Boulder, CO).

4. Quantitative Brightness Analysis of Fluorescence Intensity Fluctuations in E. Coli

The brightness measured by fluorescence fluctuation spectroscopy specifies the average stoichiometry of a labeled protein in a sample. Here we extended brightness analysis, which has been mainly applied in eukaryotic cells, to prokaryotic cells with E. coli serving as a model system. The small size of the E. coli cell introduces unique challenges for applying brightness analysis that are addressed in this work.

4.1 Introduction

This chapter extends brightness analysis of single color FFS to prokaryotes, using the bacterium E. coli as a model system. The volume of a typical bacterial cell is on the order of a femtoliter, while a typical mammalian cell has a volume of a few picoliters. This reduction in volume introduces significant challenges. First, the size of the bacterium is smaller than the point spread function (PSF) of the optical microscope. The incomplete overlap between PSF and sample changes the amplitude distribution of the fluorescence intensity fluctuations and therefore distorts the brightness [63,101]. Second, because the excitation beam illuminates a large part of the bacterial volume, photobleaching results in a noticeable decrease in the number of active fluorophores. We refer to this cumulative decrease in the population of fluorescent molecules as photodepletion, which has been discussed in more detail in the previous chapter [64]. Photodepletion is not accounted for by traditional FFS analysis and can cause spurious results [64].

Z-scan FFS and segmented brightness analysis (SBA) were independently developed to account for incomplete PSF overlap and photodepletion in eukaryotic cells, respectively [63,64]. However, we found that these methods were not adequate for experiments on prokaryotes. In response, we developed in this work mean segmented Q-value (MSQ) analysis and combined it with a modified z-scan FFS theory to evaluate the brightness of labeled proteins in E.coli. We present a rigorous derivation of MSQ analysis. While the derivation is lengthy and proceeds through a number of intermediate steps, the final result is simple and elegant. We demonstrate that MSQ analysis coupled with z-scan FFS recovers not only brightness but also the concentration and diffusion time. We first applied the technique to recover the brightness of enhanced green fluorescent protein (EGFP) in mammalian, yeast, and E. coli cells. Next, EGFP-labeled nuclear transport factor 2 (NTF2), which has been shown to be dimeric in U2OS cells [63,102], was measured in E. coli, to evaluate the potential of our new algorithm to determine the oligomeric state of a soluble protein in a bacterial cell. We found that the performance of MSQ analysis in E. coli cells is comparable to established brightness analysis methods in mammalian cells.

4.2 Materials and Methods

4.2.1 Instrumentation

FFS measurements were performed with an excitation wavelength of 1000 nm and a power of ~1 mW. The photon count data was recorded with a frequency of 20 kHz. The stage was driven by a voltage signal from an arbitrary waveform generator (33250A, Agilent Technologies, Santa Clara, CA). The signal waveform was a linear ramp

function with a frequency of 0.1 Hz and a peak-to-peak amplitude of 0.8 V, which corresponds to 8.04 μm of axial travel. The z-scan intensity profile was sampled at 20 kHz.

4.2.2 Expression vectors

EGFP was amplified from the pEGFP-C1 plasmid (Clontech, Mountain View, CA) with a 5' primer that encodes a BamHI restriction site and a 3' primer that encodes an XhoI site. The product was cloned into the pRSET-B vector (Invitrogen, Carlsbad, CA), which is referred to as pB-G and serves as the E. coli expression vector. NTF2 was amplified from human NTF2 expression vector (Genbank accession number: BC002348) with a 5' primer that encodes an XhoI restriction site and a 3' primer that encodes an EcoRI site. The result was cloned into the EcoRI/HindIII site of pB-G. Expression vectors for yeast and U2OS cells have been described in the previous chapter [64].

4.2.3 Sample preparations

The competent E. coli strain BL21(DE3)pLysS (Promega, Madison, WI) was used in this study. E. coli cells carrying either the EGFP or NTF2-EGFP vector were cultured overnight at ~30 C° in Lysogeny broth (LB) medium with 1 mM ampicillin. The medium was diluted to $0.2 - 0.3$ OD_{600nm} with fresh LB medium the next morning. After growing to $0.6 - 0.8$ OD_{600nm} the medium was centrifuged at 6000 g for ~10 s. After removing the old medium the cells were resuspended with fresh LB medium and mixed with low-melting point agarose dissolved in PBS medium at ~32 C°. A volume of 0.5 μl of the 1% agarose/medium mixture was transferred to a microscope slide and covered by a coverslip, which was gently pressed to achieve a layer thickness of ~1 μm between the

microscope slide and the coverslip. This process resulted in an orientation of E. coli cells parallel to the glass interface. The slide borders were sealed with nail polish. The preparation of yeast and U2OS samples has been described in the previous chapter [64]. In addition, EGFP was purified as reported [94] and dissolved in Biacore's HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) for solution measurements.

4.2.4 Measurement Protocol and Analysis

The FFS experiments in U2OS and yeast cells were performed as described in the previous chapter [4,29,64]. For experiments on E. coli the bacteria were first identified in bright-field illumination using a CCD camera. The focal point of the two-photon beam was aligned with the geometric center of the imaged E. coli cell, followed by a z-scan at a reduced power of ~0.3 mW, which ensured that photodepletion was negligible during the scan. Before performing the FFS measurement, the beam position was moved axially until the fluorescent intensity was maximized, which corresponds to a focus at the midpoint of the E. coli cell. The beam power was then switched to ~1 mW to collect photon counts for the FFS experiment. The analysis of the FFS experiments and the z-scan intensity profiles is described in the Results section (see section 4.3). Artifacts due to undersampling of fluctuations are negligible, since data were sampled faster than the residence time of the labeled protein [75,81]. We also performed solution measurements of EGFP to provide a reference brightness λ_{EGFP} or reference Q-factor Q_{EGFP} for the cell experiments. These measurements were taken with the focus ~10 μm into the solution to

ensure that the PSF is completely embedded in the solution. All data were analyzed with programs written in IDL 8.3 (Research Systems, Boulder, CO).

4.2.5 Z-scan Calibration of PSF

A modified squared Gaussian-Lorentzian (mGL) model,

$$PSF(x, y, z) = \left(\frac{z_0^2}{z_0^2 + z^2} \right)^{(1+\eta)} \exp\left(-\frac{4z_0^2}{w_0^2} \frac{\left(x^2 + y^2 \right)}{z_0^2 + z^2} \right), \tag{4.1}$$

provides a good approximation of the PSF of our two-photon microscope [63,101]. A z-scan calibration procedure was performed as described previously [63] to determine the free parameters of our model. The calibration resulted in $z_0 = 0.86 \pm 0.08$ μm, $\eta = 2.20 \pm 0.22$, and $\omega_0 = 0.43 \pm 0.05$ μm, where ω_0 and z_0 describe the radial and axial beam waist, while η characterizes the axial decay shape of the PSF. The mGL PSF volume is determined by [63]

$$V_\infty = \left(\frac{1}{4} \pi w_0^2 z_0 \right) \frac{\sqrt{\pi}\ \Gamma\left(\eta - \frac{1}{2} \right)}{\Gamma(\eta)}, \tag{4.2}$$

which yields 0.18 fl for the calibrated parameters.

4.3 Results

Before measuring in E. coli we performed a control FFS experiment in the nucleus of U2OS cells expressing EGFP by recording the sequence of photon counts k_i with a sampling time T of 50 μs. Each photon count k_i is related to the photon count rate by $F(iT) = k_i/T$, which is traditionally also referred to as the fluorescence intensity as
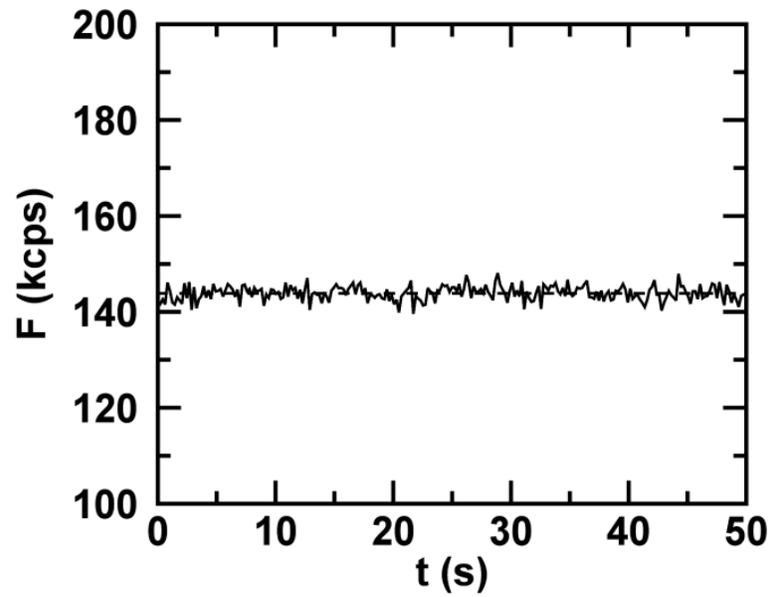
Figure 4.1 Fluorescence intensity trace *F*(t) of EGFP in U2OS cell.

The fluorescence intensity of EGFP measured inside the nucleus is stationary with an average value of 144 kcps.

discussed in more detail elsewhere [74]. Because the intensity is constant (Figure 4.1), conventional FFS theory, which assumes a stationary fluorescence signal, applies. We used Mandel's Q-parameter to determine the brightness λ of the sample from the photon counts [103],

$$Q = \frac{\kappa_{[2]}}{\kappa_{[1]}} = \frac{\langle \Delta k^2 \rangle - \langle k \rangle}{\langle k \rangle} = \frac{\langle \Delta F^2 \rangle}{\langle F \rangle} T = \gamma_2 \lambda T \ , \tag{4.3}$$

where γ_2 is the PSF gamma factor [74,75]. This equation summarizes important relations of Q that hold in conventional FFS theory. The population mean $\langle k \rangle$ and variance $\langle \Delta k^2 \rangle$ of the recorded photon counts k_i are linked to the first $\kappa_{[1]} = \langle k \rangle$ and second $\kappa_{[2]} = \langle \Delta k^2 \rangle - \langle k \rangle$ factorial cumulant of k_i [74]. The mean and variance of the fluorescence intensity are given by $\langle F \rangle = \langle k \rangle / T$ and $\langle \Delta F^2 \rangle = \left(\langle \Delta k^2 \rangle - \langle k \rangle \right) / T^2$. By using Eq. 4.3 we determined $Q = 0.018$ for EGFP in the U2OS cell, which corresponds to a brightness of $\lambda = 1.28$ kcps. We typically convert λ or Q into a normalized value

$$b = \frac{Q}{Q_{EGFP}} = \frac{\lambda}{\lambda_{EGFP}} \tag{4.4}$$

by taking the ratio with the reference brightness λ_{EGFP} or Q-value Q_{EGFP} of the label EGFP, which were determined from an independent solution measurements of EGFP. The normalized brightness reflects the average stoichiometry of the fluorescently labeled protein. In other words, a monomeric protein corresponds to $b = 1$, while a dimeric protein results in $b = 2$. We measured a Q-value Q_{EGFP} of 0.019 and determined a normalized brightness of $b = 0.95$, which is consistent with a monomeric EGFP in

cells,since the typical uncertainty of b measured in mammalian cells is ~10% [104,105].

Eq. 4.4 assumes that both Q-values are measured with the same PSF and comparable

overlap between PSF and sample as will be discussed later.

When we performed an FFS experiment in E. coli expressing EGFP, the

fluorescence intensity $F(t)$ was not stationary as in the U2OS cell, but decayed

exponentially (Figure 4.2A) from an initial intensity F_0,

$$F(t) = F_0 \exp\left(-k_D t\right), \quad (4.5)$$

because photobleaching within the very small volume of the bacterium leads to

photodepletion with a rate coefficient k_D. Because the decaying signal is non-stationary,

applying Eq. 4.3, which is based on conventional FFS theory, can result in strongly

biased brightness values [64]. SBA theory was introduced to circumvent this bias by

dividing the intensity trace into segments (Figure 4.3A) short enough that the intensity

decay per segment is negligible [64]. This process leads to quasistationary data within a

segment provided that the segment time T_S does not exceed a limit $T_{S,\text{limit}}$, which is

determined by SBA theory from the intensity decay curve. SBA calculates the unbiased

brightness λ from the segmented FFS data as previously demonstrated [64]. Applying

SBA analysis to the E. coli data of Figure 4.2A determined a very short limit ($T_{S,\text{limit}} =$

0.2 s), which reflects the relatively fast intensity decay within the bacterium. To test the

SBA model for such short data sections, we calculated the brightness for segment times

of 0.2 s, 0.05 s, and 0.025 s and recovered 1.78, 1.49, and 1.21 kcps, respectively. Instead

of recovering the same value as expected from SBA theory, we observed a decrease in
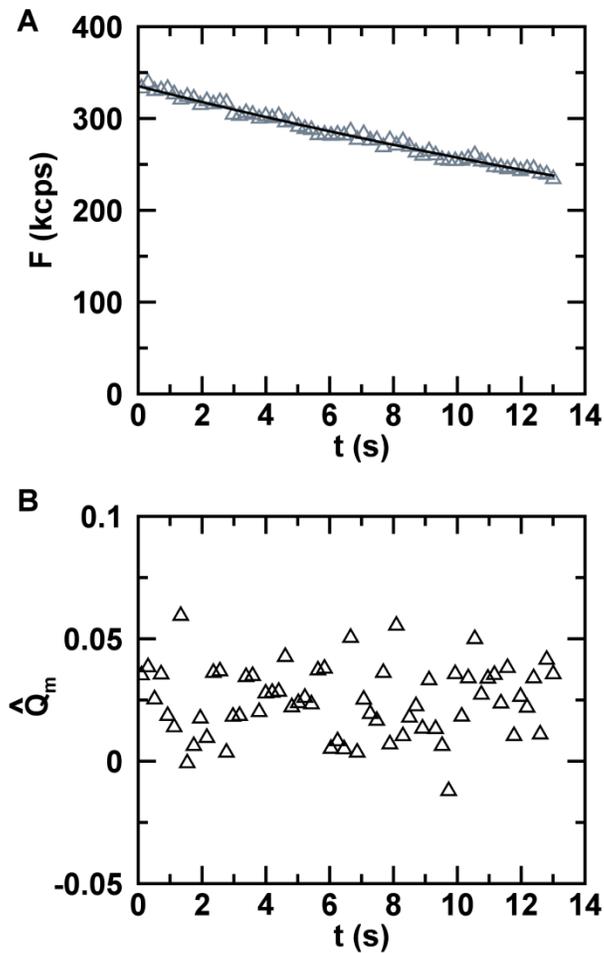
Figure 4.2 Fluorescence from EGFP in E. coli cell.

(A) Fluorescence intensity (triangles) decays with time as a result of photodepletion. The fit (solid line) to an exponential decay function recovered an initial intensity $F_0 = 336$ kcps and a depletion rate coefficient $k_D = 0.026 \text{ s}^{-1}$. (B) Segmented Q-values $\widehat{Q_m}(T_S)$ for a segment time of $T_S = 0.2$ s with an average value of 0.025.

brightness at shorter segment times. This result demonstrated that SBA analysis is not suitable for E. coli samples.

Thus, we set out to develop a robust analysis method that is equally applicable in large eukaryotic and small prokaryotic cells. We define a few quantities used throughout the paper. The experimental data are divided into $M = T_{\mathrm{DAQ}}/T_S$ segments with T_{DAQ} being the total data acquisition time and T_S representing the segment time (Figure 4.3A). The m-th segment defined by the time interval $\left[(m-1)T_S, mT_S\right]$ contains $N = T_S/T$ sampled photon count events $k_{m,i}$ sampled with a time interval T. The unbiased estimator of the first two photon count moments, $\widehat{k_m} \equiv N^{-1}\sum_{i=1}^{N} k_{m,i}$ and $\widehat{k_m^2} \equiv N^{-1}\sum_{i=1}^{N} k_{m,i}^2$, were used to construct an estimator of Q for the m-th segment based on Eq. 4.3

$$\widehat{Q_m}(T_S) = \frac{\widehat{\Delta k_m^2} - \widehat{k_m}}{\widehat{k_m}} = \frac{\widehat{k_m^2}}{\widehat{k_m}} - \widehat{k_m} - 1, \tag{4.6}$$

with an estimator of the variance defined by $\widehat{\Delta k_m^2} \equiv N^{-1}\sum_{i=1}^{N}\left(k_{m,i} - \widehat{k_m}\right)^2$. Applying Eq. 4.6 determines the Q-value for each segment as illustrated in Figure 4.3B. The experimental segmented Q-values for the E.coli data depicted in Figure 4.2A are shown in Figure 4.2B for $T_S = 0.2$ s. We also define the average of the Q-estimator over all segments,

$$\mathrm{MSQ}(T_S) = M^{-1}\sum_{m=1}^{M} \widehat{Q_m}(T_S), \tag{4.7}$$
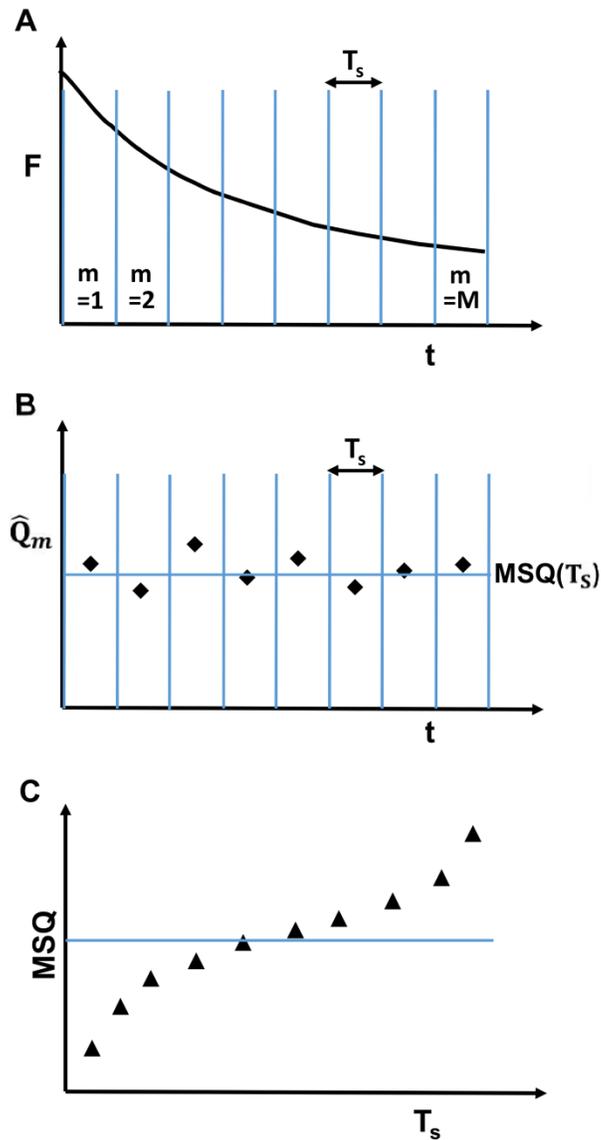
Figure 4.3 Schematic representation of MSQ analysis procedure.

(A) The decaying fluorescence intensity trace is divided into M segments. Each segment has a length of T_S. (B) The Q-value $\widehat{Q}_m(T_S)$ is calculated from the photon count data of each segment, followed by the calculation of the mean of the segmented Q-values $\mathrm{MSQ}(T_S)$. (C) The above steps are repeated for different segment lengths to calculate MSQ as a function of T_S. Conventional FFS theory predicts that MSQ is independent of the segment length (solid line). The presence of photodepletion and estimator bias introduces curvature into the MSQ-curve (triangles).

which we refer to as the mean of the segmented Q-values (MSQ). The MSQ-curve is constructed by calculating Eq. 4.7 for many different segment times and will be the centerpiece of the new analysis method introduced here (Figure 4.3C). Conventional FFS as described by Eq. 4.3 predicts a MSQ-curve that is independent of the segment length T_S (solid line, Figure 4.3C). Any observed changes of MSQ with T_S reflect the presence of an artifact that needs to be accounted for. In the following, we will present experimental MSQ-curves and develop the theory to model the data.

We calculated the MSQ-curve for FFS data of EGFP measured in U2OS, yeast and E. coli (Figure 4.4A, 4.4B and 4.4C) and observed a clear dependence of MSQ on T_S. Similarly, repeating the process on data from E. coli expressing NTF2-EGFP resulted in a pronounced dependence of MSQ on T_S (Figure 4.4D). The MSQ-curve from the U2OS cell expressing EGFP (Figure 4.4A) comes closest to the ideal behavior. The MSQ-factor stays essentially constant for $T_S > 1\,\text{s}$ and only appreciably drops for T_S less than ~0.4s. Performing the same experiment in yeast cells resulted in a MSQ-curve (Figure 4.4B) with a similar decline at short segment times as seen with the U2OS cells. However, unlike the U2OS cells, the MSQ-curve rises at long segment times, indicating an apparent increase in brightness. We previously demonstrated that photobleaching, which leads to a depletion of the fluorophores within the small volume of the yeast cell, introduces artificially inflated brightness values [64]. Because the volume of a U2OS cell vastly exceeds that of yeast, the same photobleaching process results in an entirely negligible depletion of the fluorophore population in the larger cell [64]. The MSQ-curves

108
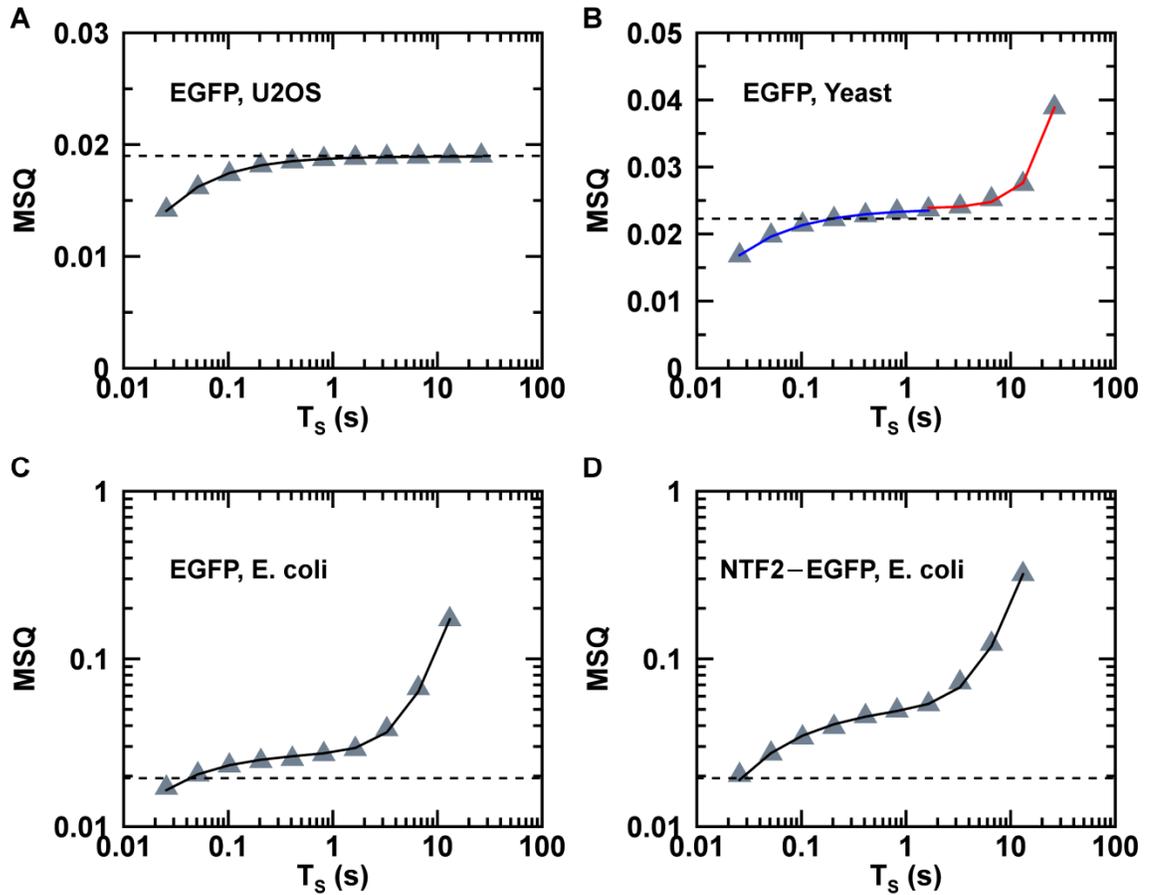
Figure 4.4 MSQ curves.

(A) MSQ-curve (triangles) for EGFP in U2OS cell and fit (solid line) to MSQ model with $Q = 0.0193$ and a diffusion time of 0.8 ms. (B) MSQ-curve (triangles) of EGFP in yeast cell. Fit (blue line) of data with $T_S < 1.6$ s to Eq. 4.8 yielded $Q = 0.0237$ and $\tau_D = 1.2$ ms. Fit (red line) of data with $T_S > 1.6$ s to Eq. 4.12 determined $Q = 0.0238$ and $k_D = 4.46 \times 10^{-3}\,\text{s}^{-1}$. (C) MSQ-curve (triangles) of EGFP in E.coli cell and fit (solid line) to Eq. 4.14 with $Q = 0.028$, $\tau_D = 2.7$ ms and $k_D = 2.7 \times 10^{-2}\,\text{s}^{-1}$. (D) MSQ-curve (triangles) of NTF2-EGFP in E.coli cell and fit (solid line) to Eq. 4.14 with $n = 2.1$, $\tau_D = 10$ ms and $k_D = 5.3 \times 10^{-2}\,\text{s}^{-1}$. The dashed line in each panel represents the reference Q-value of EGFP in solution, which was measured at the same power as the corresponding MSQ data.

for E. coli (Figure 4.4C and 4.4D) are graphed with a logarithmic y-axis and display the same general behavior as observed for yeast, only more pronounced.

A common feature of all experimental MSQ-curves is the observed decrease at short segment times. We suspected that estimator bias is responsible for this effect, because it also affects the autocorrelation function [106]. The Q-estimator of Eq. 4.6 involves the ratio of two unbiased estimators ($\widehat{k_m^2}$ and $\widehat{k_m}$) and therefore is only asymptotically unbiased [107]. We started with the definition of Eq. 4.6 and derived the expectation value of the MSQ function accounting for the estimator bias (see section 4.5.1),

$$\mathrm{MSQ}_{EB}\left(T_S\right) = Q - \frac{1}{N} - Q\frac{B_2\left(NT, \tau_D\right)}{\left(NT\right)^2}, \tag{4.8}$$

where $B_2\left(T, \tau_D\right)$ represents the second-order binning function [74,75] with τ_D as the diffusion time. Fitting the MSQ-curve from the U2OS cell to Eq. 4.8 leads to a good representation of the experimental data with $\tau_D = 0.80$ ms and $Q = 0.019$ (Figure 4.4A). The fitted diffusion time is in good agreement with the diffusion time of 0.72 ms determined by a fit of the autocorrelation function of the U2OS data. The fitted Q-value matches the calibration value ($Q_{EGFP} = 0.019$) for EGFP in solution.

The relative bias between the MSQ-value and the true Q-value depends on the second and third term of Eq. 4.8. The influence of the second term on the MSQ-value is mostly negligible, since $N \geq 500$ for all data shown in Figure 4.4, which results in a
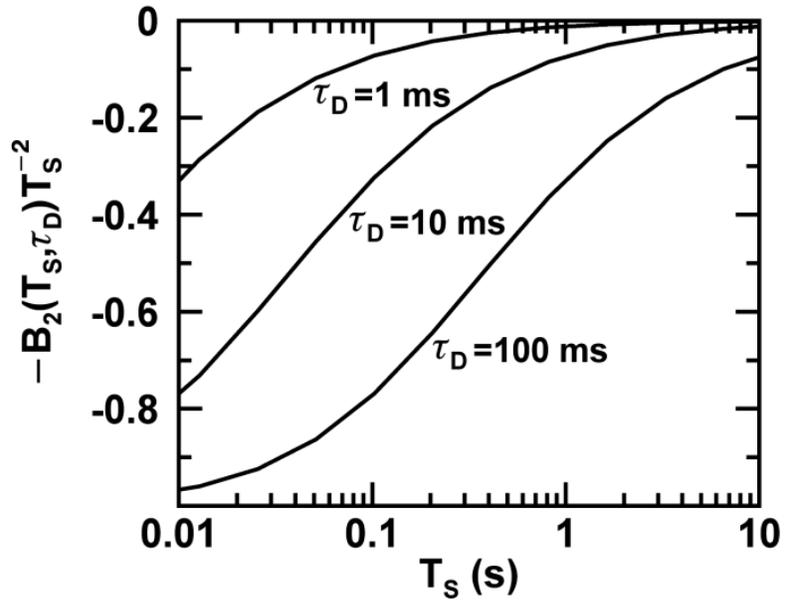
Figure 4.5 Relative bias in the Q-value introduced by diffusion.

Correlations in the photon counts introduced by the diffusion time τ_D give rise to a relative bias in the Q-estimator that depends on the segment period T_S. The relative bias for diffusion times of 1 ms, 10 ms, and 100 ms is shown.

maximum relative bias of ~10% at the shortest segment length. The third term, which arises from the correlation in the photon counts, becomes more important as the ratio T_S/τ_D decreases. The relative bias in MSQ due to the third term is $-B_2\left(T_S,\tau_D\right)/T_S^2$ and exceeds 10% once $T_S < 50\tau_D$ (Figure 4.5). This result demonstrates that slow diffusing species are more prone to estimator bias than fast diffusing species. The bias in the MSQ decreases with increasing T_S and disappears in the limit $T_S \to \infty$, which demonstrates that Eq. 4.8 describes an asymptotically unbiased estimator.

Next we analyzed photodepletion following a previously discussed approach [64] to model the increase in the MSQ-curve at long segment times (Figure 4.4B and 4.4C). Consider an unbiased estimator $\widehat{g_m}$ of the form $N^{-1}\sum_{i=1}^{N} g_{m,i}$. A non-stationary signal introduces a time-dependent population mean $\left\langle g_m(t)\right\rangle$. Since $\widehat{g_m}$ involves a summation over the m-th segment of duration T_S, the expectation value $E\widehat{g_m}\equiv\left\langle g_m\right\rangle$ represents the time-average of $\left\langle g_m(t)\right\rangle$ over the segment,

$$\overline{\left\langle g_m\right\rangle}\equiv N^{-1}\sum_{i=1}^{N}\left\langle g_m\left(t_i\right)\right\rangle\approx\frac{1}{T_S}\int_{(m-1)T_S}^{mT_S}\left\langle g(t)\right\rangle dt\,,\qquad(4.9)$$

where we approximated the summation by a time integral and $t_i = \left(m-1\right)T_S + iT$. We are using the convention that a bar over a variable specifies the time average of that variable over one segment. By applying Eq. 4.9 to the estimators $\widehat{k_m}$ and $\widehat{\Delta k_m^2}$ we derived an

112

expression for the expectation value of the Q-estimator $\widehat{EQ_m}$ using time-averaged cumulants (see section 4.5.2),

$$\widehat{EQ_m} = \frac{\overline{\kappa_{[2]m}} + \overline{\kappa_{[1]m}^2} - \left(\overline{\kappa_{[1]m}}\right)^2}{\overline{\kappa_{[1]m}}}.$$

(4.10)

Eq. 4.10 can only be applied to long segment times, because estimator bias has been ignored for now.

We evaluate Eq. 4.10 for the special case of a monomeric protein, such as EGFP, in the presence of photodepletion with rate coefficient k_D and a photobleaching reaction, $F \rightarrow D$, that converts the fluorescent state F into a non-fluorescent state D. The first two time-dependent cumulants of this model have been derived in previous work [64],

$$\begin{aligned}
\kappa_{[1]}(t) &= \lambda T N_0 e^{-k_D t}, \\
\kappa_{[2]}(t) &= \gamma_2 \lambda T \kappa_{[1]}(t),
\end{aligned}$$

(4.11)

where N_0 is the initial average number of fluorescence proteins in the observation volume, which is related to the initial fluorescence intensity by $F_0 = \lambda N_0$. We calculated the time-averaged expressions for $\overline{\kappa_{[2]m}}$, $\overline{\kappa_{[1]m}^2}$ and $\overline{\kappa_{[1]m}}$ based on Eq. 4.11, which were inserted into Eq. 4.10. Next, we determined a model for the MSQ curve in the presence of photodepletion by averaging over all segments, $\mathrm{MSQ}_{PD}(T_S) = M^{-1}\sum_{m=1}^{M} \widehat{EQ_m}(T_S)$,

$$\mathrm{MSQ}_{PD}(T_S) = \mathrm{A}(Q_1, n) + F_0 T \frac{1 - (1 - \Delta f_D)^M}{M \Delta f_D} \left(\frac{2 - \Delta f_D}{2} + \frac{\Delta f_D}{\ln(1 - \Delta f_D)} \right).$$

(4.12)

113

A detailed derivation of this equation is found in section 4.5.3. The equation depends on the number of segments $M = T_{DAQ}/T_S$, the initial fluorescence intensity F_0, and the depletion fraction,

$$\Delta f_D(T_S) = 1 - \exp(-k_D T_S). \qquad (4.13)$$

The depletion fraction Δf_D describes the fractional decrease of the fluorescence intensity from the beginning to the end of a segment, $\Delta f_D = 1 - F(mT_S)/F((m-1)T_S)$, which for an exponential decay with the depletion rate constant k_D reduces to Eq. 4.13. The function $A(Q_1,n)$, which depends on the Q-factor Q_1 for a monomer and the stoichiometry n of the protein sample, will be discussed in more detail later. From here on we have to distinguish carefully between the Q-factor Q_1 of a monomer and the Q-factor $Q = nQ_1$ of an n-mer. For the special case of a monomer ($n = 1$) considered for now, the function reduces to $A(Q_1,1) = Q_1$.

We applied the above theory to the MSQ-curve of EGFP in yeast (Figure 4.4B). SBA analysis of the yeast data indicated that photodepletion effects are only significant for $T_S > 1.6$ s. Thus, we fit the MSQ-curve for $T_S > 1.6$ s to Eq. 4.12 with $A(Q_1,1) = Q_1$. The only fit parameters were Q_1 and the depletion rate coefficient k_D of Eq. 4.13. The number of segments M for each T_S is known and the initial intensity F_0 was determined from the intensity decay curve of the yeast data. The experimental MSQ-curve for $T_S > 1.6$ s is reproduced by the fit (red line, Figure 4.4B) to the MSQ_{PD} model with $Q_1 = $

0.0238. The fitted depletion rate coefficient of 0.0045 s^{-1} is in good agreement with the value of 0.0044 s^{-1} recovered by an independent fit of the intensity decay (Figure 4.6). Next, we modeled the experimental MSQ-curve for $T_S < 1.6$ s. Because we expected that this part of the curve is only influenced by estimator bias, we modeled it using Eq. 4.8. The fit (blue line, Figure 4.4B) to this equation faithfully describes the MSQ-curve at small segment times with a fitted Q-value of 0.0237. Both fitted Q-values are in close agreement. Comparing both Q-values with the in vitro calibrated monomeric Q-value ($Q_{EGFP,\infty} = 0.0223$, Figure 4.4B) results in a normalized brightness of 1.06, which is consistent with a monomeric protein.

The yeast data demonstrated that photodepletion bias and estimator bias can affect different domains of the MSQ-curve. Short segment lengths suffer from estimator bias, while long segment lengths are affected by photodepletion bias. The plateau in the MSQ curve around $T_S \approx 1$ s separates these two domains. Segment times that correspond to the plateau region are essentially free from either artifact, which validates our previous SBA analysis of yeast FFS data [64]. The MSQ-curve for E. coli (Figure 4.4C), on the other hand, lacks the plateau region found for yeast, which implies that no region of the MSQ-curve is free of bias. This observation further implies that a range of segment lengths exists where both biases are present simultaneously. Thus, modeling of the MSQ-curve requires the following formula that combines both photodepletion and estimator bias (see section 4.5.4),
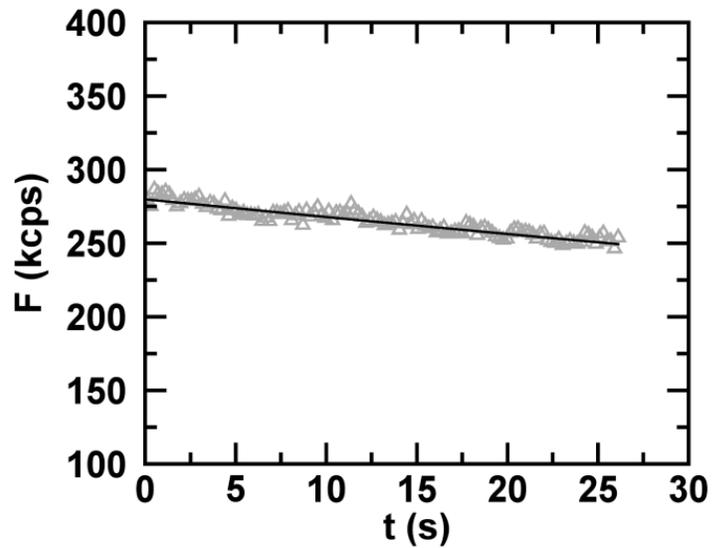
Figure 4.6 Fluorescence intensity trace $F(t)$ of EGFP in yeast cell.

The fluorescence intensity of EGFP measured in the cytoplasm of yeast decreases. The data are described by an exponential decay (solid line) with $F_0 = 280$ kcps and a rate coefficient $k_D = 0.0044$ s^{-1}.

$$\mathrm{MSQ}(T_S) = \mathrm{A}(Q_1, n) + F_0 T \frac{1 - (1 - \Delta f_D)^M}{M \Delta f_D} \left(\frac{2 - \Delta f_D}{2} + \frac{\Delta f_D}{\ln(1 - \Delta f_D)} \right)$$
$$- \frac{1}{N} - \mathrm{A}(Q_1, n) \frac{B_2(T_S, \tau_D)}{T_S^2}. \quad (4.14)$$

The first two terms are identical to Eq. 4.12 and capture the influence of photodepletion. The next two terms describe the estimator bias and are similar to Eq. 4.8, the only difference being the last term where Q has been replaced by $\mathrm{A}(Q_1, n)$. We fit the entire MSQ-curve for E. coli expressing EGFP to Eq. 4.14 with $\mathrm{A}(Q_1, 1) = Q_1$, since EGFP is a monomeric protein. The only free fit parameters were k_D, τ_D, and Q_1, since $\Delta f_D = 1 - \exp(-k_D T_S)$, $M = T_{DAQ}/T_S$, and $N = T_S/T$. The initial intensity F_0 was determined by a fit of the intensity decay curve (Figure 4.2A). Applying Eq. 4.14 to the MSQ data of E. coli results in a fit (solid line, Figure 4.4C) that closely matches the experimental curve. However, the recovered monomeric Q-value ($Q_1 = 0.028$) was significantly higher than the calibration Q-value of EGFP ($Q_{EGFP} = 0.019$) measured in solution. This result implies a normalized brightness of 1.47, which is inconsistent with a monomeric sample.

This apparent contradiction is caused by differences in the overlap between the sample and the PSF volume [63,101]. The solution measurements of EGFP were taken with the focus sufficiently deep in the solution, so that the entire PSF volume is embedded in the sample. This situation mimics an infinite sample reservoir and is the standard condition assumed in traditional FFS analysis. These differences in overlap prompted us to distinguish from now on between different sample geometries. The

117

subscript ∞ is used to mark properties that are measured or calculated for the infinite sample geometry. Thus, EGFP with brightness λ_{EGFP} measured in solution leads to a Q-factor $Q_{EGFP,\infty} = \gamma_{2,\infty}\lambda_{EGFP}T$ with $\gamma_{2,\infty}$ describing the gamma factor of the entire PSF. The geometry of E. coli can be approximated by a cylinder with diameter of ~1 μm, which is too small to enclose the entire PSF volume. This incomplete overlap between bacterium and the PSF influences the measured Q-parameter [63]. FFS parameters that are measured or calculated for the cylindrical geometry of E. coli are identified by the subscript *cyl*. EGFP measured in E.coli leads to a Q-factor $Q_{EGFP,cyl} = \gamma_{2,cyl}\lambda_{EGFP}T$ with $\gamma_{2,cyl}$ describing the gamma factor of the PSF that overlaps with the sample. Thus samples with the same brightness λ_{EGFP} measured in two different sample geometries result in different Q-factors. Calculating the normalized brightness by Eq. 4.4 implicitly assumes that the Q-factors are taken with the same overlap between sample and PSF. Thus, we need to convert the solution Q-factor $Q_{EGFP,\infty}$ of EGFP into the equivalent Q-factor $Q_{EGFP,cyl}$ for E.coli before employing Eq. 4.4. The equations $Q_{EGFP,\infty} = \gamma_{2,\infty}\lambda_{EGFP}T$ and $Q_{EGFP,cyl} = \gamma_{2,cyl}\lambda_{EGFP}T$ relate both Q-factors by $Q_{EGFP,cyl} = Q_{EGFP,\infty}\gamma_{2,cyl}/\gamma_{2,\infty}$. However, this procedure requires knowledge of the gamma factor ratio, which can be determined by z-scan FFS [63].

We originally developed z-scan FFS based on earlier work by the Hof group [108] to correctly determine the brightness of thin layers, such as a thin cytoplasmic slab [63]. The fluorescence intensity profile of a z-scan through the sample determines the sample geometry, which is then used to identify the correct gamma factor for FFS experiments

[63]. We followed the same approach and performed an axial scan of the PSF through the E. coli bacterium with the scan trajectory perpendicular to the rotation axis of the cylinder. The scan passes through the geometric center (Figure 4.7A) and generates a z-scan intensity profile (Figure 4.7B). Previous z-scan analysis of the intensity profile accounted for the finite sample size only along the z-direction, which for E. coli is no longer sufficient, because the finite width of the bacterium is comparable to the size of the PSF in the y direction. The length of the bacterium is sufficiently large that its finite size is not a concern. Thus, we modeled the bacterium as a cylinder of radius R assuming for simplicity an infinite length along the x-direction. The cylinder is defined by the set

$$V_{cyl}(R) = \left\{ (x', y', z') \mid x' \in (-\infty, \infty), y' \in (-\sigma, \sigma), z' \in (-R, R) \right\} \tag{4.15}$$

with $\sigma = \sqrt{R^2 - z'^2}$. Using a coordinate system with the origin placed at the geometric center (Figure 4.7A), we define the effective PSF volume of order r by

$$V_{eff}^{(r)}(R, z) = \int_{V_{cyl}(R)} \mathrm{PSF}^r (x', y', z' - z) d\mathbf{r}'. \tag{4.16}$$

Eq. 4.16 describes the volume overlap of PSF raised to the r-th power with the cylindrical sample volume, where the PSF is located at position z with respect to the geometric center along the z-axis. Evaluating Eq. 4.16 for the modified-Gaussian Lorentzian PSF (Eq. 4.1) results in

$$V_{eff}^{(r)}(R, z) = \frac{\pi \omega_0^2 z_0}{4r} \int_{-\hat{R}}^{\hat{R}} \left(1 + (\hat{z}' - \hat{z})^2 \right)^{1 - r(1+\eta)} \mathrm{Erf}\left(\phi_r(\hat{z}) \right) d\hat{z}' \tag{4.17}$$
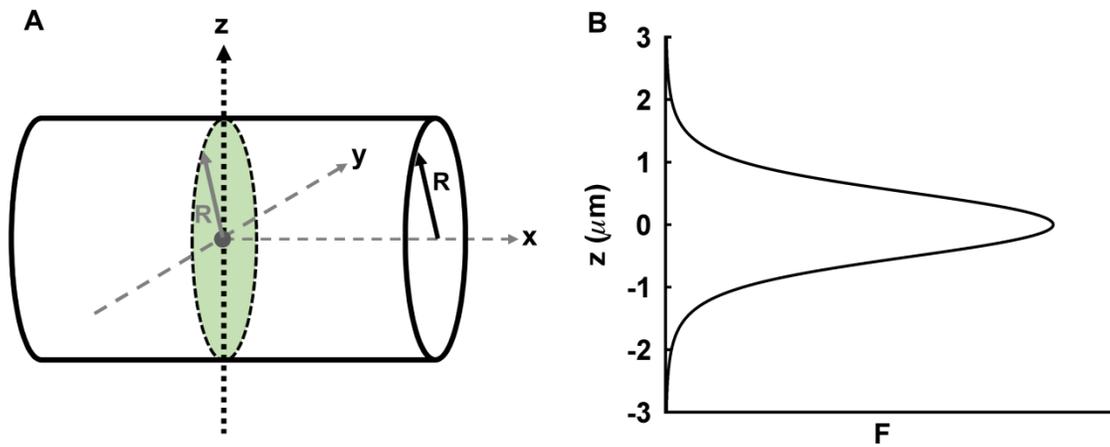
Figure 4.7 Schematic illustration of z-scan for E. coli experiments.

(A) Approximation of E. coli geometry by a cylinder with radius R. The scan axis z passes through the geometric center of the cylinder. (B) Z-scan fluorescence intensity profile of cylinder is modeled by Eq. 4.18.

with $\hat{z} = z/z_0$, $\hat{R} = R/z_0$, $\phi_r(\hat{z}) = 2\sqrt{r}\alpha\sqrt{\dfrac{\hat{R}^2 - \hat{z}^2}{1+(\hat{z}' - \hat{z}^2)}}$, $\alpha = z_0/\omega_0$, and the error function

Erf. Following the theory of z-scan FFS [63], the intensity profile of the scan is given by

$\langle F(z)\rangle = F_\infty V_{\text{eff}}^{(1)}(R,z)/V_\infty$, where F_∞ represents the intensity of an infinite sample and

V_∞ is the volume of the entire PSF. Evaluating $\langle F(z)\rangle$ for the modified Gaussian-

Lorentzian PSF using Eq. 4.17 describes the shape of the intensity profile,

$$\langle F(R,z)\rangle = F_\infty \int_{-\hat{R}}^{\hat{R}} \left(1 + (\hat{z}' - \hat{z})^2\right)^{-\eta} \text{Erf}\left(\phi_1(\hat{z})\right) d\hat{z}' . \qquad (4.18)$$

We performed eight consecutive z-scans through the geometric center of an E. coli cell

with reduced laser power to ensure the absence of photobleaching during the scans. The

intensity profiles of the consecutive scans are shown in Figure 4.8. Each profile was fit

by Eq. 4.18 to determine F_∞ and R, which recovered the averaged fit parameters $F_\infty =$

72±2 kcps and $R = 0.45\pm0.01$ μm. Inserting the averaged fit parameters into Eq. 4.18

resulted in a modeled intensity profile (red solid line, Figure 4.8), which is in good

agreement with the experimental data. We repeated this experiment on several E. coli

cells ($n = 14$). The peak intensity differed for each cell, reflecting the variations in the

EGFP concentration from cell to cell. However, the radius was essentially identical for all

cells. The averaged radius was 0.45±0.026 μm.

Since the gamma factor is defined by $\gamma_2 = V_{\text{eff}}^{(2)}/V_{\text{eff}}^{(1)}$ [63], its value $\gamma_2(z,R)$

depends on the z position of the PSF and the radius of the E. coli bacterium. The FFS

experiment was performed with the focus positioned at the geometric center of the
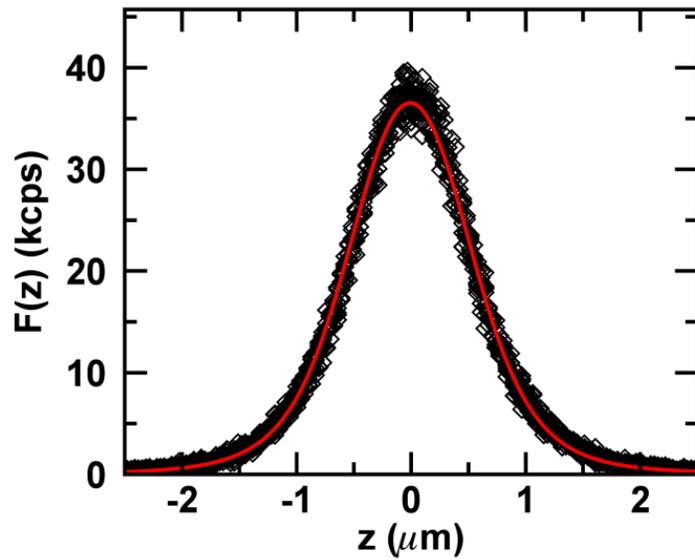
Figure 4.8 Experimental z-scan intensity profiles of EGFP from E. coli cell.

Experimental z-scan intensity data (diamonds) from eight consecutive z-scans together with model function (red curve) for $F_{\infty} = 72$ kcps and $R = 0.45$ μm.

E. coli cell (Figure 4.7A). Because this condition correspond to $z = 0$, the correct gamma factor that accounts for the overlap of the modified Gaussian-Lorentzian PSF with the sample is given by

$$\gamma_{2,cyl}(R,0) = \frac{V_{eff}^{(2)}(R,0)}{V_{eff}^{(1)}(R,0)} = \frac{1}{2} \frac{\int_{-\hat{R}}^{\hat{R}} \left(1 + \hat{z}'^2\right)^{-1-2\eta} \text{Erf}\left(\phi_2(0)\right) d\hat{z}'}{\int_{-\hat{R}}^{\hat{R}} \left(1 + \hat{z}'^2\right)^{-\eta} \text{Erf}\left(\phi_1(0)\right) d\hat{z}'}.$$

(4.19)

As mentioned earlier, a sample with brightness λ results in a Q-factor of $Q_\infty = \gamma_{2,\infty} \lambda T$ for an infinite sample and a Q-factor of $Q_{cyl}(R,0) = \gamma_{2,cyl}(R,0) \lambda T$ when measured at the geometric center of an E. coli cell with radius R. The Q-factors of both geometries are related by

$$Q_{cyl}(R,0) = Q_\infty \frac{\gamma_{2,cyl}(R,0)}{\gamma_{2,\infty}}.$$

(4.20)

The ratio $\gamma_{2,cyl}(R,0)/\gamma_{2,\infty}$ calculated from Eq. 4.19 is shown in Figure 4.9 as a function of the radius R. Because the radius of the E. coli bacteria was constant at 0.45 μm, the gamma ratio for E. coli is $\gamma_{2,cyl}(0.45 \text{ μm}, 0)/\gamma_{2,\infty} = 1.51$.

We applied Eq. 4.20 to the calibration Q-value of EGFP ($Q_{EGFP,\infty}$ = 0.019) measured in solution to get the reference Q-factor appropriate for E. coli measurements, $Q_{EGFP,cyl}$ = 0.0287. Next, we converted the Q-value (Q_1=0.028) recovered from the fit to the MSQ-curve (Figure 4.4C) with Eq. 4.4 into a normalized brightness using $Q_{EGFP,cyl}$ as the reference, which resulted in $b = 0.98$. Thus, accounting for the cell geometry in MSQ
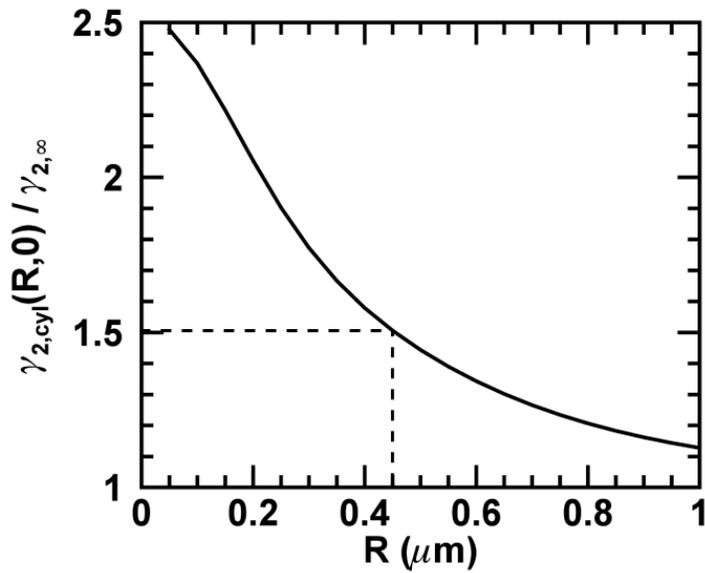
Figure 4.9 Ratio of gamma factors.

The ratio $\gamma_{2,cyl}(R,0)\big/\gamma_{2,\infty}$ as determined from Eq. 4.19 is shown as a function of the radius R. The dashed line indicates the values for the measured E. coli cells.

analysis recovered a monomeric brightness.

Additional E. coli cells expressing EGFP were measured to test our analysis procedure. The FFS data taken at the geometric center were fit to Eq. 4.14 and the normalized brightness was calculated from the recovered Q_1 with the help of Eqs. 4.4 and 4.20. The radius of the E. coli cell was either determined from the z-scan intensity profile or taken as 0.45 μm. We plotted the normalized brightness b versus the initial fluorescence intensity F_0 (Figure 4.10). The values of b are close to 1 with a mean of 0.98 and a standard deviation of 0.09. This result correctly identifies the bacterially expressed EGFP as a monomeric protein. The right axis shows the biased normalized brightness $b^* = Q_1/Q_{EGFP,\infty}$ that results if the incomplete PSF overlap is not accounted for. A value of b^* close to 1.51 would lead to the misleading conclusion that the sample is a mixture of monomers and dimers. Thus, accounting for photodepletion and geometry of the bacterium is crucial to avoid misinterpretation of FFS brightness experiments inside E.coli.

The theory developed up to this point is still incomplete. So far, we described an unbiased procedure to determine the brightness from a bacterial cell for the special case of a monomeric protein. Our model does not yet work for samples containing protein complexes, because of an additional effect of photodepletion on brightness [64]. To illustrate this issue consider a dimeric protein that contains two fluorescent labels. Photobleaching of one of the two labels results in a dimer with a reduced brightness, since only one label remains fluorescent. This process creates different brightness populations of the dimeric protein with population levels that depend on the amount of
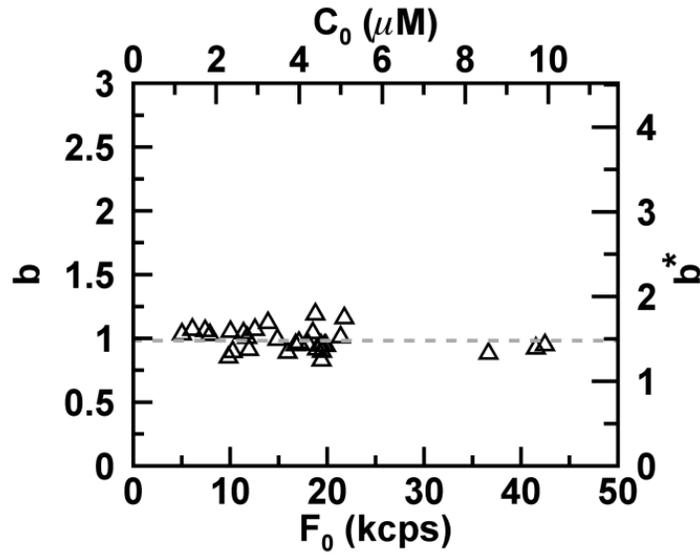
Figure 4.10 Normalized brightness of EGFP from E. coli cells.

MSQ-curves were fit to Eq. 4.14 with $A(Q_1,1)=Q_1$ and converted into a normalized brightness by $b=Q_1/Q_{cyl,\infty}$. The normalized brightness is independent of the initial fluorescence intensity F_0. The average brightness (dashed line) is 0.98 ± 0.09. The top axis represents the initial protein concentration, while the right axis displays the biased normalized brightness $b^* = Q_1/Q_{EGFP,\infty}$, when the finite size of the bacterium is ignored.

126

photodepletion. We recently examined this process for an n-meric protein with brightness $n\lambda$ and showed that photodepletion leads to a time-dependence of the first two cumulants [64],

$$\kappa_{[1]}(t) = \sum_{s=0}^{n} \lambda (n-s) T p_{n-s} N_0,$$
$$\kappa_{[2]}(t) = \sum_{s=0}^{n} \gamma_2 \lambda^2 (n-s)^2 T^2 p_{n-s} N_0,$$

(4.21)

with λ representing the brightness of a single label, N_0 being the initial number of n-mers in the observation volume and $p_{n-s} = \binom{n}{s} p^s (1-p)^{n-s}$, where $p = 1 - e^{-k_D t}$ is the probability that a fluorescent label is photobleached at time t. Following the same steps applied to the cumulants of a monomeric protein (Eq. 4.11) to the cumulants of an n-mer (Eq. 4.21) produced an MSQ function $\text{MSQ}_{PD}(T_S)$ accounting for photodepletion that is identical to Eq. 4.12 (see section 4.5.3). Next, we included the effect of estimator bias (see section 4.5.4), which produced an MSQ function that is identical to Eq. 4.14. The only difference to the monomeric case is found in the function $A(Q_1, n)$, which for a protein with stoichiometry n is given by,

$$A(Q_1, n) = Q_1 \left[1 + (n-1) \frac{2 - \Delta f_D}{2} \times \frac{1 - (1 - \Delta f_D)^M}{M \Delta f_D} \right].$$

(4.22)

Now that we have a complete theory, we decided on the following strategy to analyze the MSQ-curve from an E. coli sample with unknown stoichiometry n. The experimental MSQ-curve is fit to Eqs. 4.14 and 4.22 with n, k_D, and τ_D as the only fit parameters. F_0 is determined from a fit of the intensity decay curve, while $N = T_S / T$,

$M = T_{DAQ}/T_S$, and $\Delta f_D = 1 - \exp\left(-k_D T_S\right)$ are functions of T_S . The monomeric Q-factor

Q_1 of the function A is needed as a calibration factor and set equal to $Q_{EGFP,cyl}$ to account

for the geometry of the bacterium. Because the normalized brightness b and the

stoichiometry n are numerically identical, $b = n$, we use both parameters interchangeably

and at times refer to n as the normalized brightness. As a first test of this procedure we

reanalyzed the FFS data from E. coli expressing EGFP with the new fit strategy to

recover the stoichiometry of the sample. The analysis returned a normalized brightness n

of ~1 for all samples (mean of 0.98 ± 0.10) as expected for a monomeric protein (Figure

4.11). The fit parameter k_D varied slightly from cell to cell (mean 0.022 s^{-1} and standard

deviation 0.0073 s^{-1}), because of volume variations caused by different lengths of the E.

coli cells. The diffusion time τ_D was approximately the same with a mean of 2.5 ±

0.9 ms.

We turned to the MSQ-curve taken in E. coli expressing NTF2-EGFP (Figure

4.4D). A fit (solid line, Figure 4.4D) to Eqs. 4.14 and 4.22 with $Q_1 = Q_{EGFP,cyl}$ resulted in a

normalized brightness of n = 2.1, a diffusion time of 10 ms, and a depletion rate

coefficient of 0.053 s^{-1}. The initial intensity F_0 was determined from a fit of the intensity

decay curve (Figure 4.12). The normalized brightness indicates a dimeric protein

complex, which is consistent with the observation of dimeric NTF2-EGFP in U2OS cells

and in solution [63,93]. We applied the same analysis to additional FFS data from E. coli

cells expressing mammalian NTF2-EGFP. The fitted normalized brightness was ~2 in all

cases (Figure 4.11), indicating that NTF2-EGFP exists as a dimer in the E. coli cytosol.
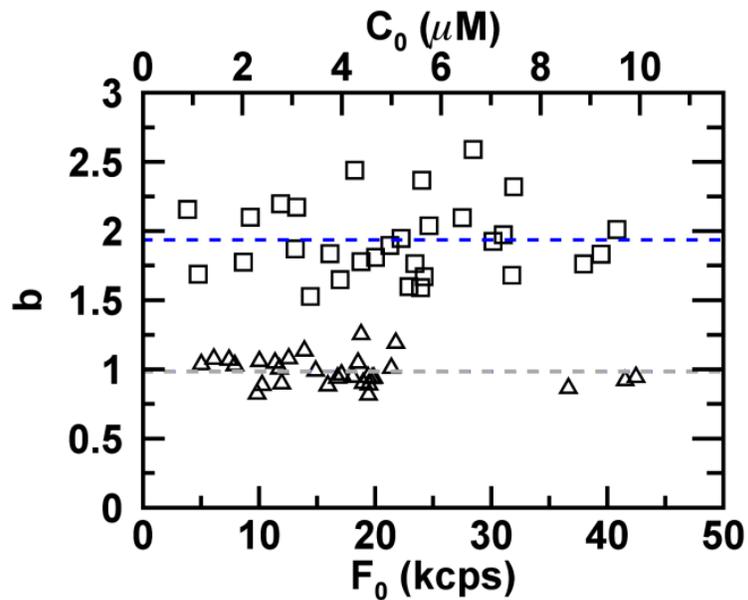
Figure 4.11 Measured stoichiometry of proteins in E. coli cells.

MSQ-curves were fit to Eqs. 4.14 and 4.22 to determine the average stoichiometry n for EGFP (triangles) and NTF2-EGFP (squares) as a function of the initial fluorescence intensity F_0. The average stoichiometry of EGFP (gray dashed line) is 0.98 ± 0.10. The average stoichiometry of NTF2-EGFP (blue dashed line) is 1.94 ± 0.27. The top axis displays the initial protein concentration.

The mean of the normalized brightness was 1.94 ± 0.27 and the fitted diffusion time τ_D had a mean of 9.5 ± 3.4 ms.

Finally, the theory developed here also allowed us to convert the initial fluorescence intensity F_0 into a concentration. The value of F_0 is given by $F_0 = \lambda_{EGFP} N_0$ with N_0 representing the initial number of the EGFP-labeled proteins in the PSF volume. The molar concentration is determined by dividing N_0 by Avogadro's number N_A and the effective PSF volume $V_{\text{eff}}^{(1)}(R,0)$ of the mGL-PSF focused at the geometric center of an E. coli of radius R, $c = N_0 \big/ \left(V_{\text{eff}}^{(1)}(R,0) \cdot N_A \right)$. The top axis of Figure 4.10 and Figure 4.11 depicts the concentration of EGFP and NTF2-EGFP of the E. coli measurements, respectively.

4.4 Discussion

This work provides a detailed account of the development of quantitative brightness analysis of FFS data in prokaryotic cells with E. coli chosen as our model system. The function $\text{MSQ}(T_S)$ is central to our analysis strategy, and is experimentally determined from FFS data by Eqs. 4.6 and 4.8. The final formulation of MSQ theory (Eqs. 4.22 and 4.14) takes photobleaching, which depletes the fluorophore population and reduces the brightness of protein complexes, and estimator bias into account. The MSQ method is a significant advancement over SBA analysis. In essence, SBA analysis can only be applied, if the MSQ-curve has a plateau. Thus, simple inspection of the MSQ-curves of Figure 4.4B and Figure 4.4C demonstrates that the SBA model is sufficient for
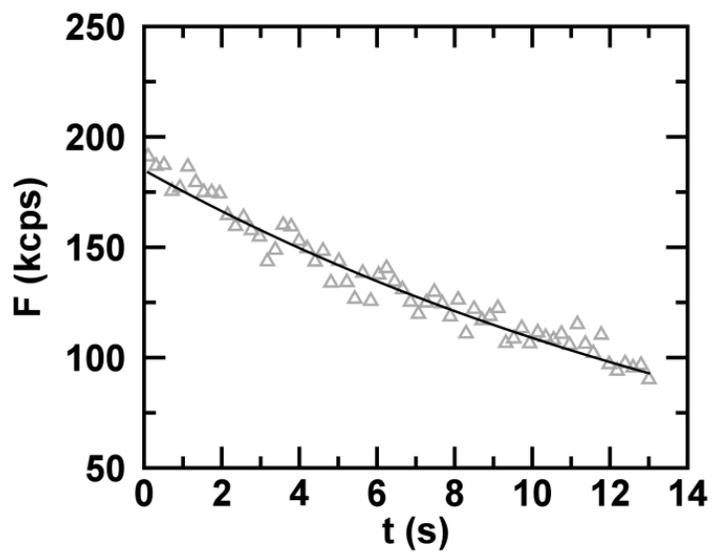
Figure 4.12 Fluorescence intensity trace F(t) of NTF2-EGFP in E. coli cell.

The experimental intensity data are described by an exponential decay (solid line) with $F_0 = 185$ kcps and a rate coefficient $k_D = 0.053$ s^{-1}.

the yeast measurement, but fails in case of the E. coli data. In contrast, the MSQ model was successfully applied to FFS data obtained from mammalian, yeast, and E. coli cells.

MSQ is directly determined from the experimental photon count data without the need for any model. Thus, connecting MSQ with the brightness, which depends on the PSF and its overlap with the sample, requires additional information. Two external factors F_0 and $Q_{EGFP,cyl}$ are needed for fitting the MSQ curve. F_0 represents the initial fluorescence intensity and is determined from a fit of the intensity decay curve. $Q_{EGFP,cyl}$ is the reference Q-factor of the monomer that accounts for the geometry of the E. coli cell. Its calculation by Eq. 4.20 requires $\gamma_{2,cyl}(R,0)/\gamma_{2,\infty}$, which accounts for the overlap between the geometry of the sample and the PSF. We extended the theory of z-scan FFS to E. coli, determined the radius R from the z-scan intensity profile and calculated $\gamma_{2,cyl}(R,0)/\gamma_{2,\infty}$. The ratio is ~1.5, which reflects a significant correction of brightness due to the geometry.

Effects associated with the geometric size of prokaryotic cells have not been accounted for in previous fluorescence fluctuation studies [109,110], except for one study where a Monte-Carlo simulation was used to estimate a correction factor accounting for the bacterial geometry assuming a 3D-Gaussian PSF [111]. However, commonly used model functions, such as the 3D-Gaussian model, do not accurately reflect the actual PSF shape [63]. Thus, quantitative studies should use an experimentally characterized PSF instead of a general model function, because the numerical values of the effective volume and gamma factor depend on the form of the PSF. This study used the modified

Gaussian-Lorentzian function with calibrated parameters (Eq. 4.1), which we have found to be a sufficiently precise model for two-photon microscopy [63,101].

In the derivation of the general MSQ model we assumed an irreversible photobleaching reaction of the label from a single bright to a non-fluorescent state. We earlier demonstrated with experiments both in mammalian and in yeast cells that the brightness behavior of EGFP was well approximated by this simple model provided the intensity loss by photodepletion was less than 60% [64]. Because the intensity reduction of our E. coli experiments never exceeded ~50%, the simple photobleaching model is sufficient. However, for experiments using a different fluorescent protein as label it is important to perform control experiments to establish the validity of the photobleaching model employed here.

The analysis of MSQ data of E. coli cells by Eqs. 4.14 and 4.22 included three fitting parameters, n, τ_D, and k_D. Because the depletion rate coefficient k_D can also be determined from the fit of the intensity decay curve $F(t)$, k_D can be changed to a fixed instead of a free parameter. We chose to keep it as a free parameter in the MSQ analysis and found that the recovered values of k_D by MSQ and from the intensity decay agreed within a few percent.

The main goal of this study was the determination of the normalized brightness or stoichiometry of soluble proteins in E. coli. The analysis accurately identified EGFP as a monomer and NTF2-EGFP as a dimer (Figure 4.11). The mean of the normalized brightness values deviated less than 5% from the ideal monomer and dimer brightness.

Remarkably, the uncertainty in the normalized brightness of a single measurement was ~10% for EGFP and ~14% for EGFP-NTF2, which is similar to the brightness uncertainty achieved in U2OS cells [104,105]. Thus, MSQ analysis of fluctuation data from a bacterium achieved an accuracy that matched standard FFS experiments in mammalian cells. This result speaks to the robustness of the MSQ approach. We choose NTF2 for this study, because it forms a very tight dimer and is not found in prokaryotes. The absence of unlabeled endogenous NTF2 ensures that all dimers formed contain two labels, as confirmed by our analysis, which represents the first quantitative brightness analysis of protein interactions in E. coli cells.

We were able to calculate the absolute concentration of a labeled protein in E. coli, which is a byproduct of MSQ analysis. Measuring concentrations inside bacterial cells is of sufficient interest that many studies have been devoted to this topic. By comparing the fluorescence intensity of bacteria with a reference sample of known concentration and applying correction factors absolute concentrations can be estimated (reviewed in [112]). A powerful method for measuring absolute concentrations in bacterial cells is based on single molecule counting [113], but this approach is technically challenging and works best at low protein concentrations. A different approach to measure concentrations uses intensity fluctuations between daughter cells after bacterial cell division [114,115]. Concentrations determined from intensity fluctuation of imaged bacteria have also been reported [111,116]. Because the amplitude of fluctuations is directly related to concentration, fluctuation methods are a powerful approach to measure absolute concentrations without relying on a concentration reference. However, these methods are often intricate and care needs to be exercised to minimize systematic errors [117]. The

134

MSQ method has been carefully characterized and accounts for sample geometry, PSF shape, photobleaching, and statistical bias, which should eliminate many sources of systematic error. Thus, MSQ may offer an attractive approach for measuring bacterial protein concentrations.

MSQ analysis also recovers the diffusion time τ_D through the dependence of the MSQ on the binning function $B_2(t,\tau_D)$. The standard method of measuring the diffusion time in FFS is the autocorrelation function. Because the binning function B_2 is mathematically related to the autocorrelation function [74,75], it also can recover the diffusion time. We found in U2OS cells, where photodepletion is negligible and only estimator bias contributes to the shape of the MSQ curve (Figure 4.4A), that MSQ and autocorrelation analysis recover diffusion times that agree within ~10%. Extending autocorrelation analysis from eukaryotic cells to bacterial samples faces challenges. Photodepletion, brightness reduction of protein complexes due to photobleaching, and the presence of potential estimator bias can affect the fluctuation amplitude of each diffusing species. The derivation of MSQ theory provides a rigorous framework for incorporating these effects into autocorrelation analysis. Of course, the partial overlap of the sample with the PSF affects the amplitude as well. An additional complication is the effect of diffusion in a confined space on the shape of the autocorrelation function [118]. Because this effect appears to be sufficiently small in E. coli cells, determination of diffusion times should generally be possible [109,110]. Thus, we expect that the finite size of E. coli has no significant effect on the diffusion-time dependent shape of the MSQ curve. In fact, we can justify this statement with a simple estimate. Figure 4.5 shows the influence

of diffusion on the MSQ amplitude. Given the experimental uncertainty, it seems reasonable to disregard effects that introduce less than ~20% bias. This implies that diffusion significantly influences the MSQ curve only for segment times $T_S \leq 20\tau_D$. The maximum time $T_S \sim 20\tau_D$ leads to a root-mean square displacement Δx_{rms} of $\sqrt{2DT_S}$ = $\frac{1}{2}w_0\sqrt{T_S/\tau_D} \sim 1$ μm, which is approximately the shortest dimension of the E. coli cell. Thus, diffusion for $T_S \leq 20\tau_D$ is essentially not confined, which justifies our diffusion analysis of the MSQ curve.

MSQ analysis of EGFP in E. coli recovered an average diffusion time of 2.5 ms with a standard deviation of ~35%. Converting the diffusion time into a diffusion coefficient, $\tau_D = \omega_0^2/8D$, results in a value of $9.2^{+5.2}_{-2.5}$ μm^2/s. The diffusion coefficient of GFP in E. coli has been measured in many studies with techniques that include FRAP, FCS, and single molecule tracking [119–121]. The majority of values reported by these studies range from 6 to 14 μm^2/s [121], which is consistent with our result. Finally, the diffusion time is not affected by photodepletion, because the timescales are vastly different. Diffusion times occur on the millisecond timescales, while the characteristic photodepletion time is on the order of tens of seconds. Thus, the probability of a given fluorophore to undergo photobleaching while passing through the PSF is vanishingly low. This point has been discussed in more detail recently [64].

The MSQ method offers a fairly straightforward and simple analysis approach. The algorithm for calculating the MSQ-curve is easy to implement and consists of data rebinning followed by the calculation of MSQ values from the average and variance of

the rebinned data. In addition, the FFS data taken inside the small sample compartment contain all relevant information, except for the influence of geometry on FFS parameters, which is established by taking a separate z-scan fluorescence intensity profile. The initial intensity F_0 and the MSQ-curve determined from the FFS data identify the concentration, brightness, diffusion time, and depletion rate coefficient with the help of Eq. 4.14. We successfully demonstrated combined MSQ and z-scan analysis on proteins that are soluble and uniformly distributed inside E. coli. Extending MSQ analysis to a non-uniform protein distribution, such as generated by a protein bound to the cell membrane of E. coli, would require additional development work and is beyond the scope of this study. However, a recent paper discusses brightness experiments of non-uniformly distributed proteins [101], which might serve as a suitable starting point for the development of a generalized MSQ theory.

Combining laser scanning with fluorescence fluctuation measurements is a very popular approach [116,122–124]. One of the reported advantages of scanning is the reduction of photobleaching effects on fluctuation experiments [125]. Because the probability of photobleaching increases with exposure time, scanning the beam causes a reduction in the amount of photobleaching at any one particular volume. However, this advantage becomes marginal when the sample size is similar to the size of the laser beam as is the case for bacterial cells. Scanning also does not reduce the depletion of fluorophores compared to stationary measurements. The amount of fluorophores depleted from a sample depends directly on the integrated laser power the sample was exposed to irrespective whether scanning occurred or not. However, a significant advantage of

scanning over stationary measurements is the possibility to probe all bacteria in the field of view simultaneously, while our approach requires manual repositioning and aligning of the cell with respect to the beam, which is a time-consuming process. A recent study used this advantage of scanning to determine protein concentration in Bacillus subtilis cells with N&B analysis [111,116]. Because the N&B algorithm is similar to calculating the Q-factor, it should be possible to adapt MSQ theory to scanning applications that account for photodepletion and estimator bias, and thereby reducing systematic errors in the data analysis.

In summary, the MSQ method enables quantitative brightness analysis of soluble proteins in samples ranging from mammalian to bacterial cells. Geometric overlap between the bacterial cell and the PSF, which was characterized by z-scan FFS, has to be considered to correctly connect the Q-value with the stoichiometry of a protein complex. MSQ analysis was used to successfully identify monomers, as well as dimers, in E. coli cells. In addition, MSQ provides the diffusion times of the labeled protein. The results of this work demonstrate that quantitative FFS analysis of protein complexes and their concentrations in femtoliter-sized compartments is feasible. We expect that the MSQ approach will prove useful as a robust analysis method for FFS studies of bacterial samples. The concepts of MSQ theory might also provide a useful starting point for future FFS studies of small organelles in mammalian cells, such as the endoplasmic reticulum or the nucleolus.

4.5 Supplemental Materials: Derivation of MSQ theory

4.5.1 Q-factor estimator for stationary signals

We rewrite the estimators of the first two photon count moments of the m-th

segment, $\widehat{k_m} = N^{-1} \sum_{i=1}^{N} k_{m,i}$ and $\widehat{k_m^2} = N^{-1} \sum_{i=1}^{N} k_{m,i}^2$ by substituting $k_{m,i} = \langle k_m \rangle + \delta k_{m,i}$ with

$\delta k_{m,i}$ being the instantaneous fluctuation around the mean $\langle k_m \rangle$ and inserting it into the

definition of the Q-estimator of the m-th segment given by Eq. 4.6. Because the

estimators $\widehat{k_m}$ and $\widehat{k_m^2}$ are unbiased by the mean ergodic theorem [126], their expectation

values are equal to the population moments of the photon counts, $E\widehat{k_m} = \langle k_m \rangle$ and

$E\widehat{k_m^2} = \langle k_m^2 \rangle$. The estimator $\widehat{k_m}$ in the denominator can be expressed as $\widehat{k_m} = \langle k_m \rangle + \Delta$

with $\Delta = N^{-1} \sum_{i=1}^{N} \delta k_{m,i}$. As N increases, Δ must vanish to satisfy the mean ergodicity

theorem, which implies $\Delta / \langle k_m \rangle \ll 1$. Taking the Taylor expansion of the denominator up

to the second order of $\Delta / \langle k_m \rangle$ leads to

$$\widehat{Q_m}(T_S) = \frac{N^{-1} \sum_{i=1}^{N} \delta k_{m,i}^2 - \langle k_m \rangle}{\langle k_m \rangle} - \frac{N^{-2} \sum_{i,j=1}^{N} \delta k_{m,i} \delta k_{m,j}}{\langle k_m \rangle} + O\left(\left(\frac{\Delta}{\langle k_m \rangle}\right)^3\right). \quad (4.23)$$

The expectation value of this estimator is its ensemble average or population mean given

by

$$E\widehat{Q_m}(T_S) = \langle \widehat{Q_m}(T_S) \rangle = Q_m - \frac{N^{-2} \sum_{i,j=1}^{N} \langle \delta k_{m,i} \delta k_{m,j} \rangle}{\langle k_m \rangle} \quad (4.24)$$

where we used $\langle \delta k_{m,i}^2 \rangle = \langle \Delta k_m^2 \rangle$ and introduced the Q-factor of the m-th segment by applying the same definition as used in traditional FFS theory (see Eq. 4.3),

$$Q_m = \frac{\kappa_{[2],m}}{\kappa_{[1],m}} = \frac{\langle \Delta k_m^2 \rangle - \langle k_m \rangle}{\langle k_m \rangle} = Q = \gamma_2 \lambda T . \tag{4.25}$$

Because the signal is stationary, Q_m has to have the same value Q for all segments.

We next rewrite the summation in Eq. 4.24 as a sum of variances and covariances

$$\sum_{i,j=1}^{N} \langle \delta k_{m,i} \delta k_{m,j} \rangle = \sum_{i=1}^{N} \langle \delta k_{m,i}^2 \rangle + 2 \sum_{i=1<j}^{N} \langle \delta k_{m,i} \delta k_{m,j} \rangle . \tag{4.26}$$

Each of the two sums is expressed as factorial cumulants of the photon counts as detailed in [75],

$$\begin{aligned} \sum_{i=1}^{N} \langle \delta k_{m,i}^2 \rangle &= N \kappa_{[2],m}(T) + N \kappa_{[1],m}(T) \\ 2 \sum_{i=1<j}^{N} \langle \delta k_{m,i} \delta k_{m,j} \rangle &= \kappa_{[2],m}(T_S) - N \kappa_{[2],m}(T) \end{aligned} . \tag{4.27}$$

The first two factorial cumulant of the m-th segments are given by $\kappa_{[1],m}(T) = \lambda T N_0$ and $\kappa_{[2],m}(t) = \gamma_2 \lambda^2 N_0 B_2(t, \tau_D)$. B_2 is called the second-order binning function [74,75]. The first moment $\langle k_m \rangle$ is equal to the first factorial cumulant $\kappa_{[1],m}$ [74,75]. We insert the above relations into Eq, 4.24 to express the expectation value of the Q-estimator

$$E\widehat{Q_m}(T_S) = \langle \widehat{Q_m}(T_S) \rangle = Q_m - \frac{1}{N} - Q_m \frac{B_2(T_S, \tau_D)}{T_S^2} , \tag{4.28}$$

where we used Eq. 4.25 and the absence of undersampling as assumed throughout the manuscript,

$$Q_m = \frac{\kappa_{[2],m}(T)}{\kappa_{[1],m}(T)} = \gamma_2 \lambda \frac{B_2(T,\tau_D)}{T} = \gamma_2 \lambda T .$$ (4.29)

The binning function reduces to $B_2(T,\tau_D) = T^2$ in the absence of undersampling [75].

Finally, we take the average of Eq. 4.28 over all segments to derive the expectation value of the MSQ-function due to estimator bias $\mathrm{MSQ}_{EB}(T_S) = \langle \mathrm{MSQ}(T_S) \rangle = M^{-1}\sum_{m=1}^{M} \langle \widehat{Q_m}(T_S) \rangle$. Since Q_m has the same value Q for all segments for a stationary signal, Eq. 4.28 is independent of the segment number, which results in

$$\mathrm{MSQ}_{EB}(T_S) = Q - \frac{1}{N} - Q \frac{B_2(T_S,\tau_D)}{T_S^2} .$$ (4.30)

4.5.2 Time-averaged Q-estimator for a non-stationary signal

We consider the data segment defined by the time interval $\left[(m-1)T_S, mT_S\right]$. For simplicity, we assume a long enough segment so that estimator bias is negligible. Estimators for the mean and variance of the photon counts over a segment are defined by

$$\begin{aligned} \widehat{k_m} &\equiv \frac{1}{N}\sum_{i=1}^{N} k_{m,i} = \frac{1}{T_S}\int_{(m-1)T_S}^{mT_S} k_m(t')\,dt' \\ \widehat{\Delta k_m^2} &\equiv \frac{1}{N}\sum_{i=1}^{N} \left(k_{m,i} - \overline{k_m}\right)^2 = \frac{1}{T_S}\int_{(m-1)T_S}^{mT_S} \left(k_m(t') - \overline{k_m}\right)^2 dt' \end{aligned}$$ (4.31)

The sum is converted into an integration since the sampling time T is much smaller than the segment time T_S. A bar over a variable defines the time-average over the segment period as defined in Eq. 4.9. Thus, $\overline{k_m}$ denotes the time-average of the photon counts (see

Eq. 4.32). The variance is estimated by subtracting the time-averaged mean from the instantaneous photon count $k_m(t)$. The expectation values of the above estimators are

$$
\begin{aligned}
E\hat{k}_m &= \frac{1}{T_S}\int_{(m-1)T_S}^{mT_S}\langle k_m(t')\rangle dt' = \overline{k_m} \\
E\widehat{\Delta k^2}_m &= \frac{1}{T_S}\int_{(m-1)T_S}^{mT_S}\left\langle\left(k_m(t')-\overline{k_m}\right)^2\right\rangle dt' = \overline{k_m^2} - \overline{k_m}^2
\end{aligned}
\tag{4.32}
$$

Next, we express the time-average of the first two photon count moments in terms of time-averaged factorial cumulants,

$$
\begin{aligned}
\overline{k_m} &= \overline{\kappa_{[1],m}} \\
\overline{k_m^2} &= \overline{\kappa_{[2],m}} + \overline{\kappa_{[1],m}^2} + \overline{\kappa_{[1],m}} + \overline{\kappa_{[1],m}}^2
\end{aligned}
\tag{4.33}
$$

The above relations are based on known relations between raw moments and factorial cumulants [74]. Specifically, the mean of the photon counts equals the first factorial cumulant of the photon counts, $\langle k(t')\rangle = \kappa_{[1]}(t')$, while the second moment is given by $\langle k^2\rangle = \kappa_{[2]} + \kappa_{[1]}^2 + \kappa_{[1]}$. Applying Eq. 4.9 to these relations results in Eq. 4.33. Finally, evaluating the expectation value of the Q-estimator (Eq. 4.6) with the help of Eqs. 4.32 and 4.33 leads to

$$
E\widehat{Q_m} = \frac{E\left(\widehat{\Delta k_m^2} - \hat{k}_m\right)}{E\hat{k}_m} = \frac{\overline{\kappa_{[2],m}} + \overline{\kappa_{[1],m}^2} - \overline{\kappa_{[1],m}}^2}{\overline{\kappa_{[1],m}}}.
\tag{4.34}
$$

4.5.3 MSQ function in the presence of photodepletion

Before deriving the MSQ function we must evaluate Eq. 4.34 for a monomeric or n-meric protein sample. Let us assume a fluorescently-labeled protein F that associates to form an n-mer F_n with a brightness λn, where λ is the brightness of the monomer. We

postulate that a chromophore in the fluorescent state F converts irreversibly and independently to a non-fluorescent dark state D as a result of photobleaching. Thus, photobleaching of exactly one chromophore leads to the state $F_{n-1}D_1$ with brightness $\lambda(n-1)$. The probability for a fluorophore to be photobelached is given by $p = 1 - \exp(-k_D t)$. The n-mer's brightness state F_n changes into the state $F_{n-s} D_s$ of brightness $\lambda(n-s)$ with the probability $p_{n-s} = \binom{n}{s} p^s (1-p)^{n-s}$. The initial state at $t=0$ is the n-mer F_n with the number of molecules equal to N_0. The number of molecules of each state at time t is given by the number $N_0 p_{n-s}$. The factorial cumulants for an n-meric protein in the presence of photodepletion (Eq. 4.21) were previously derived assuming the absence of undersampling [64]. These two cumulants simplify to

$$
\begin{aligned}
\kappa_{[1]}(t) &= n\lambda T N_0 e^{-k_D t}, \\
\kappa_{[2]}(t) &= \gamma_2 \lambda^2 T^2 N_0 \left\{ (n^2 - n) e^{-2k_D t} + n e^{-k_D t} \right\}
\end{aligned}
\tag{4.35}
$$

Next we calculated time-integrated cumulant values of Eq. 4.35 for the m-th segment as defined by Eq. 4.9,

$$
\begin{aligned}
\overline{\kappa_{[1]m}} &= n\lambda T N_0 e^{-k_D (m-1) T_S} \left(1 - e^{-k_D T_S} \right) \big/ k_D T_S \\
\overline{\kappa_{[1]m}^2} &= \left(n\lambda T N_0 \right)^2 e^{-2k_D (m-1) T_S} \left(1 - e^{-2k_D T_S} \right) \big/ 2k_D T_S \\
\overline{\kappa_{[2]m}} &= \gamma_2 \lambda^2 T^2 N_0 \left\{ \begin{array}{l} n e^{-k_D (m-1) T_S} \left(1 - e^{-k_D T_S} \right) \big/ k_D T_S \\ + (n^2 - n) e^{-2k_D (m-1) T_S} \left(1 - e^{-2k_D T_S} \right) \big/ 2k_D T_S \end{array} \right\}
\end{aligned}
\tag{4.36}
$$

where $\overline{\kappa_{[1]m}}$ and $\overline{\kappa_{[1]m}^2}$ are affected by the overall intensity drop due to photodepletion, while $\overline{\kappa_{[2]m}}$ is sensitive to the variation in brightness states caused by photodepletion. We inserted the expressions of Eq. 4.36 into Eq. 4.10 to evaluate the expectation value of the Q-estimator of the m-th segment,

$$E\widehat{Q_m} = Q_1\left\{1+(n-1)\frac{2-\Delta f_D}{2}e^{-k_D(m-1)T_S}\right\} + F_0Te^{-k_D(m-1)T_S}\left(\frac{2-\Delta f_D}{2}+\frac{\Delta f_D}{\ln(1-\Delta f_D)}\right), \qquad (4.37)$$

where we used $Q_1 = \gamma_2\lambda T$ for the monomeric Q-factor and the photodepletion fraction $\Delta f_D = 1-e^{-k_D T_S}$. For the special case $n = 1$ and $m = 1$, Eq. 4.37 reduces to a previously derived equation [64]. From here on we have to distinguish carefully between the Q-factor Q_1 of a monomer and the Q-factor $Q = nQ_1$ of an n-mer.

We determine a function describing the MSQ-curve in the presence of photodepletion by averaging Eq. 4.37 over all segments, $\text{MSQ}_{PD}(T_S) = M^{-1}\sum_{m=1}^{M}E\widehat{Q_m}(T_S)$. By using the geometric sum $\sum_{m=1}^{M}e^{-k_D(m-1)T_S} = \left(1-(1-\Delta f_D)^M\right)\Delta f_D^{-1}$, we derive

$$\text{MSQ}_{PD}(T_S) = A(Q_1,n) + F_0T\frac{1-(1-\Delta f_D)^M}{M\,\Delta f_D}\left(\frac{2-\Delta f_D}{2}+\frac{\Delta f_D}{\ln(1-\Delta f_D)}\right), \qquad (4.38)$$

which corresponds to Eq. 4.12. The term $A(Q_1,n)$ is defined by

$$A(Q_1,n) = Q_1\left\{1+(n-1)\frac{2-\Delta f_D}{2}\times\frac{1-(1-\Delta f_D)^M}{M\,\Delta f_D}\right\}, \qquad (4.39)$$

which reduces to Q_1 for the case of a monomeric protein ($n = 1$).

4.5.4 MSQ function in the presence of photodepletion and estimator bias

To combine estimator and photodepletion bias we start with the Q-estimator for the m-th segment given by Eq. 4.6 as $\widehat{Q}_m = \left(\widehat{\Delta k_m^2} - \widehat{k_m}\right)\Big/\widehat{k_m}$ together with $\widehat{k_m}$ and $\widehat{\Delta k_m^2}$ defined by Eq. 4.31. The estimator $\widehat{k_m}$ is rewritten as $\widehat{k_m} = \overline{k_m} + \sum_{i=1}^{N} \Delta k_{m,i}$ with $\overline{k_m}$ representing the expectation value of the estimator and $\Delta k_{m,i} = k_{m,i} - \overline{k_m}$. By following the steps outlined in section 4.5.1 an equation for the expectation value of Q-estimator is found,

$$E\widehat{Q}_m = \frac{\overline{\Delta k_m^2} - \overline{k_m}}{\overline{k_m}} - \frac{N^{-2}\sum_{i,j=1}^{N}\left\langle \Delta k_{m,i}\Delta k_{m,j}\right\rangle}{\overline{k_m}}, \tag{4.40}$$

where the second term represents the estimator bias. This equation is equivalent to Eq. 4.24, except that it also includes photodepletion. The sum of the second term can be expressed as

$$\sum_{i,j=1}^{N}\left\langle \Delta k_{m,i}\Delta k_{m,j}\right\rangle = \sum_{i=1}^{N}\left\langle \Delta k_{m,i}^{2}\right\rangle + 2\sum_{i=1<j}^{N}\left\langle \Delta k_{m,i}\Delta k_{m,j}\right\rangle. \tag{4.41}$$

Following the procedure used for the derivation of Eq. 4.27, we can rewrite both sums as

$$\begin{aligned} \sum_{i=1}^{N}\left\langle \Delta k_{m,i}^{2}\right\rangle &= N\overline{\kappa_{[2],m}} - N\overline{\kappa_{[1],m}} \\ 2\sum_{i=1<j}^{N}\left\langle \Delta k_{m,i}\Delta k_{m,j}\right\rangle &\approx \overline{\kappa_{[2],m}}B_2\left(T_S,\tau_D\right)T^{-2} - N\overline{\kappa_{[2],m}} \end{aligned}. \tag{4.42}$$

Inserting Eqs. 4.33, 4.41, and 4.42 into Eq. 4.40 results in an expectation value of the Q-estimator given by,

$$E\widehat{Q}_m = \frac{\overline{\kappa_{[2],m}} + \overline{\kappa^2_{[1],m}} - \overline{\kappa_{[1],m}}^2}{\overline{\kappa_{[1],m}}} - \frac{1}{N} - \frac{\overline{\kappa_{[2],m}B_2\left(T_S,\tau_D\right)}}{\overline{\kappa_{[1],m}}T_S^2}.$$ (4.43)

We calculate the mean segmented Q-value from this equation by

$$\mathrm{MSQ}\left(T_S\right) = M^{-1}\sum_{m=1}^{M}\frac{\overline{\kappa_{[2],m}} + \overline{\kappa^2_{[1],m}} - \overline{\kappa_{[1],m}}^2}{\overline{\kappa_{[1],m}}} - M^{-1}\sum_{m=1}^{M}\frac{1}{N} - M^{-1}\sum_{m=1}^{M}\frac{\overline{\kappa_{[2],m}}}{\overline{\kappa_{[1],m}}}\frac{B_2\left(T_S,\tau_D\right)}{T_S^2}.$$ (4.44)

Because the first term equals the MSQ function for photodepletion only (Eq. 4.38), we rewrite this equation in its final form,

$$\mathrm{MSQ}(T_S) = \mathrm{MSQ}_{PD}\left(T_S\right) - \frac{1}{N} - \mathrm{A}\left(Q_1,n\right)\frac{B_2\left(T_S,\tau_D\right)}{T_S^2},$$ (4.45)

with the function $\mathrm{A}\left(Q_1,n\right)$ is given by $M^{-1}\sum_{m=1}^{M}\left(\overline{\kappa_{[2],m}}\big/\overline{\kappa_{[1],m}}\right)$. This equation is equivalent to Eq. 4.14.

5. Characterization of ternary protein systems in living cells with tricolor heterospecies partition analysis

Tools and assays that characterize protein-protein interactions are of fundamental importance to biology, because protein assemblies play a critical role in the control and regulation of nearly every cellular process. The availability of fluorescent proteins has facilitated the direct and real-time observation of protein-protein interactions inside living cells, but existing methods are mostly limited to binary interactions between two proteins. Because of the scarcity of techniques capable of identifying ternary interactions, we developed tricolor heterospecies partition analysis. The technique is based on brightness analysis of fluorescence fluctuations from three fluorescent proteins that serve as protein labels. We identified three fluorescent proteins suitable for tricolor brightness experiments. In addition, we developed the theory of identifying interactions in a ternary protein system using tricolor heterospecies partition analysis. The theory was verified by experiments on well-characterized protein systems. A graphical representation of the heterospecies partition data was introduced to visualize interactions in ternary protein systems. Lastly, we performed fluorescence fluctuation experiments on cells expressing a coactivator and two nuclear receptors and applied heterospecies partition analysis to explore the interactions of this ternary protein system.

5.1 Introduction

The execution of most cellular processes relies on protein complexes [1,127,128]. Identifying their composition, binding energies, and the processes that regulate their assembly and disassembly provides insights into the principles of cellular organization

and function. The advent of green fluorescent protein (GFP) and its many variants has provided a powerful tool set for studying protein complexes directly inside the living cell [10,129]. Fluorescence resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS), and bimolecular fluorescence complementation (BiFC) have proven effective methods for detecting the interactions between two proteins in the live cell [16,36,83,130–132].

Virtually every cellular process involves protein complexes of more than two proteins [1]. Probing pairwise interactions between proteins, while powerful, is not sufficient to predict the behavior of protein complexes with multiple binding sites. For example, cooperative binding interactions or competition of proteins for a specific binding site is not captured by the study of binary interactions. Because fluorescence-based methods that characterize ternary interactions of proteins are very limited [133,134], this work introduces a method for characterizing ternary protein interactions involving three different protein species. The technique is based on fluorescence fluctuation spectroscopy (FFS) combined with brightness analysis [4–6,83,123] to determine the average stoichiometry of protein complexes within cells.

We first identified three fluorescent proteins suitable for FFS studies with three detection channels. Next, we introduced tricolor heterospecies partition (HSP) analysis to extract the average composition of the heterointeracting brightness species. We evaluated the feasibility of the HSP assay using protein constructs with well-characterized protein-protein interactions. Data interpretation was based on a graphical representation of the HSP analysis, which provides a convenient way to identify interactions between the proteins. Lastly, we applied tricolor HSP to investigate the interaction between a

coactivator (CoA) and two nuclear receptors (NRs) [7,8]. FFS experiments probed the binding of the CoA transcription intermediate factor 2 (TIF2) and the NRs retinoic acid receptor (RAR) and retinoid X receptor (RXR) in order to identify the presence of ternary protein complexes.

5.2 Materials and Methods

5.2.1 Instrumentation

A home-built two-photon fluorescence instrument based on an Axiovert 200 microscope (Zeiss, Thornwood, NY) with a mode-locked Ti-Sapphire Laser (Mai-Tai, Spectra Physics, Mountain View, CA) was used for FFS measurements in living cells using a 990 nm excitation beam with a power of ~0.5 mW after the 63x C-Apochromat water immersion objective (NA = 1.2, Zeiss). A dichroic mirror (Chroma Technology, Rockingham, VT) separated excitation and emission light. Two long-pass dichroic mirrors (FF506-Di03, Semrock, Rochester, NY; 580dcxr, Chroma Technology) and two band pass filters (FF02-475/50 and FF01-531/40, Semrock) were used to spectrally separate the emitted light into three detection channels (Figure 5.1). The light in each channel was detected by a single photon counting module (SPCM-AQ-14, Perkin-Elmer, Waltham, MA) and recorded by a data acquisition card (PP1000, Celoxica, UK) with a sampling frequency of 20 kHz.

5.2.2 Expression vectors and sample preparation

The vectors pEGFP-C1, pEYFP-C1, and pmCherry-C1 were used for expressing EGFP, EYFP, and mCherry in mammalian cells [28]. For simplicity, we refer to EGFP as G, to EYFP as Y and to mCherry as R when naming protein constructs. The ligand binding domains of the NRs RARα and RXRα are denoted by AL and XL, while the full
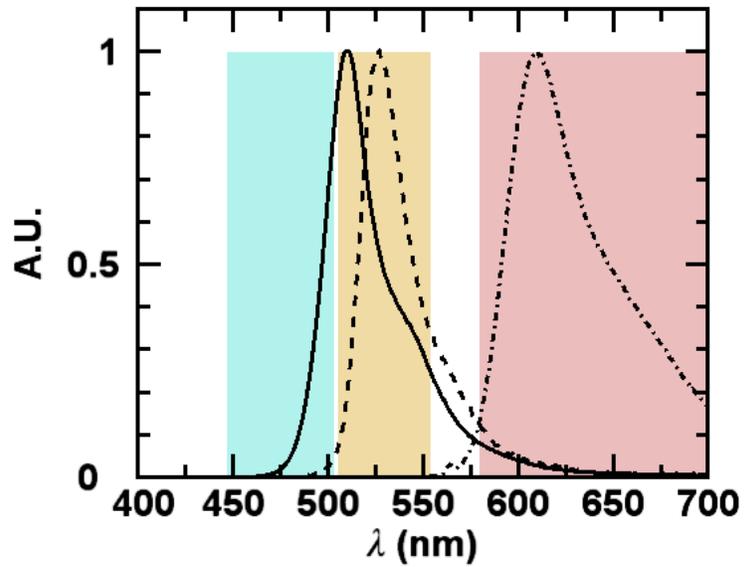
Figure 5.1 Transmission bands of detector channels and emission spectra of labels.

The emission spectra of EGFP (solid line), EYFP (dashed line), and mCherry (dot-dashed line) are shown together with the transmission bands (shaded rectangles) of the three detection channels.

length receptors RARα and RXRα are abbreviated as A and X, respectively. The CoA TIF2 is denoted by T. The G-AL, Y-XL, R-XL, G-AL-R, G-T and R-X vectors were described previously [4,6,28,35]. Mouse RAR (mRARα) was cloned into pEGFP-C1 and pEYFP-C1 at *Xho*I and *Eco*RI sites to construct G-A and Y-A vectors, while mouse RXR (mRXRα) was inserted into pEYFP-C1 at *Xho*I and *Eco*RI sites to construct Y-X.

DMEM (Gibco, Auckland, NZ) and 10% charcoal/dextran stripped fetal bovine serum (JR Scientific Inc., Woodland, CA) was used to maintain U2OS cells (American Type Culture Collection, Manassas, VA). The cells were transfected using GenJet reagent (SignaGen Laboratories, Rockville, MD) according to the manufacturer's instructions 24 hours before measurement. All cells were subcultured into eight-well coverglass chamber slides (Nalge Nunc, International, Rochester, NY) with the media replaced by DPBS (Mediatech, Inc., Manassas, VA) immediately before measurement.

5.2.3 Data analysis

The FFS data from the three detection channels is represented by a sequence of photon count tuples (k_1, k_2, k_3) taken with a sampling time T_S of 50 μs. The subscript of k_i identifies the detection channel of the photon count. Rebinning of the photon counts to other sampling times T was done in software as previously described [75]. Trivariate factorial cumulants $\hat{\kappa}_{[i,j,k]}(T)$ were calculated from the photon counts as described further down. A fit of the experimental $\hat{\kappa}_{[i,j,k]}(T)$ to modeled cumulants $\kappa_{[i,j,k]}(T)$ by non-linear least χ^2 fitting determined the fit parameters. The reduced χ^2 is calculated by

$$\chi^2 = \frac{1}{K-p-1}\sum_q \sum_{i,j,k} \frac{\left(\hat{\kappa}_{[i,j,k]}\left(T_q\right) - \kappa_{[i,j,k]}\left(T_q\right)\right)^2}{Var\left[\hat{\kappa}_{[i,j,k]}\left(T_q\right)\right]}, \tag{5.1}$$

where K is the total number of factorial cumulants used in the fit, q is the number of rebinned sampling times, and p is the number of free fit parameters. Modeling of the factorial cumulants is based on time-integrated fluorescence cumulant analysis (TIFCA) as explained in more detail later. Dead-time and afterpulsing probability of each detector were determined experimentally [80,84] and accounted for in the modeling of the data.

The two-photon excitation volume was modeled by a Gaussian-Lorentzian (GL) point spread function (PSF) with a second-order gamma factor of $\gamma_{2,\infty}$ =3/16 [62]. The third-order gamma factor was experimentally calibrated as previously described [75,80]. The PSF volume V_{PSF} was calibrated from a measurement of a dye solution of known concentration [6,75]. The r-th order binning function $B_r(T, \tau_D)$ was introduced in previous studies [74,75] and was used for TIFCA analysis of the tricolor FFS data. All of these analytical procedures were carried out with programs written in IDL 8.3 (Research Systems, Boulder, CO).

5.3 Results and Discussion

Tricolor FFS relies on three distinctly colored fluorescent proteins (FPs) that serve as labels to uniquely identify protein species by separating the fluorescence emission of each FP into three detection channels. The intensity fluctuations caused by an FP diffusing through the optical observation volume are transmitted to each detection channel and characterized by a brightness vector $\vec{\lambda}_s = \left(\lambda_{1,s}, \lambda_{2,s}, \lambda_{3,s}\right)$, where the label s

identifies the fluorescent protein and $\lambda_{i,s}$ is the brightness value of the i-th detection channel. This tricolor FFS study employed EGFP, EYFP, and mCherry as fluorescent markers, which were denoted by s = G (green), Y (yellow), and R (red), respectively. We started with an experimental characterization of the brightness of the three FPs. While there are several methods for computing brightness from experimental data, time-integrated fluorescence cumulant analysis (TIFCA) seemed most suited for this study, because the closed analytical expressions of TIFCA theory provide a rigorous framework for modeling mixtures of brightness species [28,75].

TIFCA is based on the factorial cumulants of the photon counts and has been described in detail for single- and dual-color FFS [74,75,80]. Because tricolor FFS involves three detection channels, the theory requires trivariate cumulants $\kappa^{(s)}_{[i,j,k]}$ of photon counts. For a single brightness species s the cumulants are modeled by

$$\kappa^{(s)}_{[i,j,k]}(T) = \gamma_{i+j+k}\lambda^i_{1,s}\lambda^j_{2,s}\lambda^k_{3,s}B_{i+j+k}(T,\tau_{D,s})N_s, \tag{5.2}$$

where N_s is the average number of particles within the observation volume and γ_{i+j+k} specifies the $(i+j+k)$-th gamma factor of the PSF [75]. N_s can also be interpreted as a number density or concentration $n_s = N_s/V_{PSF}$ of the particles per PSF volume. Molar concentrations are derived by $c_s = n_s/N_A$ using Avogadro's number N_A. The binning function $B_{i+j+k}(T,\tau_{D,s})$ of order $i+j+k$ accounts for the influence of sampling time T and diffusion time $\tau_{D,s}$ on the fluctuations in the photon counts [75]. Eq. 5.2 is a direct

extension of the corresponding single- and dual-color TIFCA expressions given by

$$\kappa_{[i]}^{(s)}(T) = \gamma_i \lambda_{1,s}^i B_i(T, \tau_{D,s}) N_s \text{ and } \kappa_{[i,j]}^{(s)}(T) = \gamma_{i+j} \lambda_{1,s}^i \lambda_{2,s}^j B_{i+j}(T, \tau_{D,s}) N_s \text{ [74,75].}$$

Our analysis method only uses cumulants $\kappa_{[i,j,k]}$ up to order two ($1 \leq i + j + k \leq 2$), which leads to a set of nine cumulants, $\widetilde{K} = \{ \kappa_{[1,0,0]}, \kappa_{[0,1,0]}, \kappa_{[0,0,1]}, \kappa_{[2,0,0]}, \kappa_{[0,2,0]}, \kappa_{[0,0,2]}, \kappa_{[1,1,0]}, \kappa_{[1,0,1]}, \kappa_{[0,1,1]} \}$. The first six factorial cumulants are described by single-color TIFCA [75] as they only involve the photon count signal from a single detector. The last three cumulants represent the signal correlation between two detectors, which is described by dual-color TIFCA [80]. For example, $\kappa_{[1,1,0]}$ is equivalent to the second-order cumulant $\kappa_{[1,1]}$ of the first and second detector. Thus, since all nine cumulants reduce to single- and dual-color expressions, we can fit $\kappa_{[i,j,k]}(T)$ with existing TIFCA models, which provide error analysis [74,75,80] and take dead-time and after-pulsing of the detector into account [74,75].

We characterized the brightness of EGFP, EYFP, and mCherry with tricolor TIFCA to establish a foundation for future experiments where these FPs serve as protein labels. Experimental calculation of factorial cumulants relies on expressions that relate cumulants to moments of the photon counts [75]. The average $\langle k_i \rangle$ of photon counts of channel i per sampling time, the variance $\langle \Delta k_i^2 \rangle$, and the covariance $\langle \Delta k_i \Delta k_j \rangle$ between channels i and j are sufficient to express all cumulants up to order two: $\kappa_{[1,0,0]} = \langle k_1 \rangle$,

$\kappa_{[0,1,0]} = \langle k_2 \rangle$, $\kappa_{[0,0,1]} = \langle k_3 \rangle$, $\kappa_{[2,0,0]} = \langle \Delta k_1^2 \rangle - \langle k_1 \rangle$, $\kappa_{[0,2,0]} = \langle \Delta k_2^2 \rangle - \langle k_2 \rangle$, $\kappa_{[0,0,2]} = \langle \Delta k_3^2 \rangle$

$- \langle k_3 \rangle$, $\kappa_{[1,1,0]} = \langle \Delta k_1 \Delta k_2 \rangle$, $\kappa_{[1,0,1]} = \langle \Delta k_1 \Delta k_3 \rangle$ and $\kappa_{[0,1,1]} = \langle \Delta k_2 \Delta k_3 \rangle$.

We performed a tricolor FFS measurement on a U2OS cell expressing EGFP and calculated the cumulants $\kappa_{[i,j,k]}^{(G)}(T)$ up to second order for a range of sampling times. A fit of the cumulants to the single-species model of Eq. 5.2 resulted in a reduced $\chi_\nu^2 = 0.86$ with a brightness vector $\vec{\lambda}_G = (0.560, 2.12, 0.278)$ kilocounts per second (kcps), diffusion time $\tau_D = 770$ µs, and average number of particles $N_G = 105$ (Figure 5.2). This procedure was repeated on many cells expressing EGFP at different concentrations. The recovered brightness from each channel is shown as a function of protein concentration (Figure 5.3A) together with a box plot (Figure 5.3B) summarizing the experimental scatter in the brightness data. Based on these experiments the mean and standard deviation of the brightness vector of EGFP is given by $\langle \vec{\lambda}_G \rangle = (0.561 \pm 0.028, 2.21 \pm 0.138, 0.309 \pm 0.021)$ kcps.

We repeated the same experiments on cells expressing EYFP and mCherry. Brightness versus concentration and box plots of the brightness for EYFP are shown in Figures 5.3C and D. The corresponding graphs for mCherry are depicted in Figures 5.3E and F. The mean and standard deviation of the brightness vectors for EYFP and mCherry are $\langle \vec{\lambda}_Y \rangle = (0.010 \pm 0.002, 1.85 \pm 0.108, 0.482 \pm 0.035)$ kcps and $\langle \vec{\lambda}_R \rangle = (0.001 \pm 0.001, 0.003 \pm 0.002, 0.860 \pm 0.109)$ kcps, respectively.

The data demonstrate that tricolor TIFCA robustly determines the brightness of the FPs over a wide concentration range. In addition, the EYFP brightness $\lambda_{1,Y}$ in the first channel is approximately zero, which is consistent with the emission spectrum of EYFP and the optical filter setup (Figure 5.1). Similarly, the optical setup of the three-channel
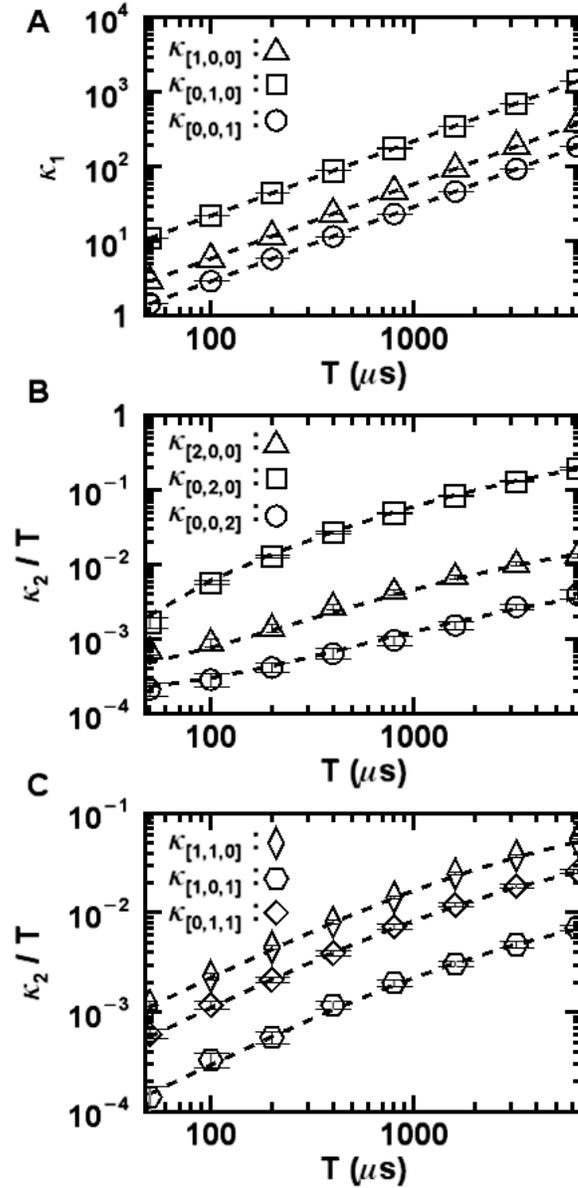
Figure 5.2 Tricolor TIFCA analysis of EGFP data.

The first nine factorial cumulants were fitted by tricolor TIFCA with a reduced χ_2 =0.86 and fit parameters $\vec{\lambda}$ = (0.560, 2.12, 0.278) kcps, τ_D =770 μs and N =105. (A) Experimental cumulants $\kappa_{[1,0,0]}(T)$, $\kappa_{[0,1,0]}(T)$, and $\kappa_{[0,0,1]}(T)$ with fitted curves (dashed lines). (B) Experimental cumulants $\kappa_{[2,0,0]}(T)$, $\kappa_{[0,2,0]}(T)$, and $\kappa_{[0,0,2]}(T)$ divided by the sampling time T together with fit curves (dashed lines). (C) Experimental cumulants $\kappa_{[1,1,0]}(T)$, $\kappa_{[1,0,1]}(T)$, and $\kappa_{[0,1,1]}(T)$ divided by the sampling time T together with fit curves (dashed lines).
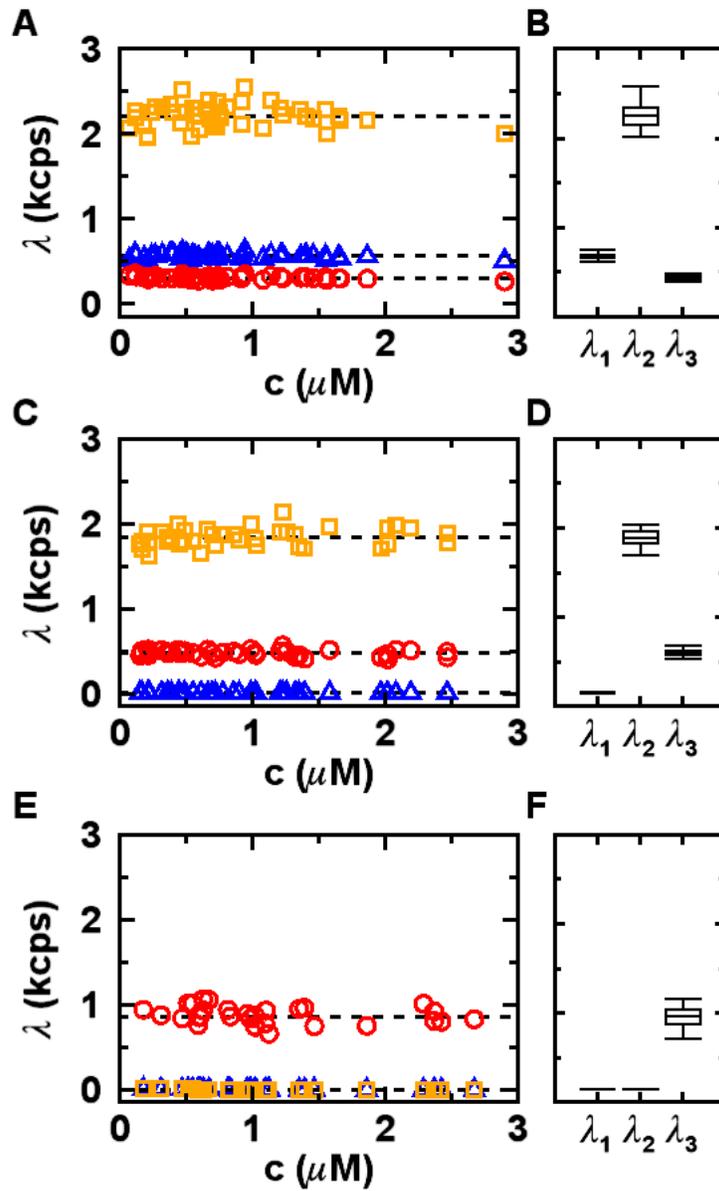
Figure 5.3 Tricolor brightness of EGFP, EYFP and mCherry.

Brightness λ_1 (triangles), λ_2 (squares) and λ_3 (circles) of EGFP (A), EYFP (C), and mCherry (E) are graphed versus concentration. The brightnesses are also represented as box plots for EGFP (B), EYFP (D), and mCherry (F).

detection system ensures that the brightness of mCherry in the first two channels is zero (Figure 5.1), which agrees with the experimental data. Thus, the brightness vectors of the three labels can be written as

$$\vec{\lambda}_G = \left(\lambda_{1,G}, \lambda_{2,G}, \lambda_{3,G}\right)$$
$$\vec{\lambda}_Y = \left(\quad 0, \lambda_{2,Y}, \lambda_{3,Y}\right),$$
$$\vec{\lambda}_R = \left(\quad 0, \quad 0, \lambda_{3,R}\right)$$

(5.3)

which will be used throughout the rest of this study. The six brightness values of Eq. 5.3 are determined by the mean brightness values measured above.

Protein complexes formed by interacting fluorescently-labeled proteins are denoted as $G_p Y_n R_m$, where p, n, and m are the stoichiometry coefficients for EGFP, EYFP, and mCherry. We assume in this study that FRET between the labels is negligible, which is discussed in more detail below. The trivariate cumulant for species $s = G_p Y_n R_m$ is given by Eq. 5.2 with brightness vector

$$\vec{\lambda}_{G_p Y_n R_m} = p\vec{\lambda}_G + n\vec{\lambda}_Y + m\vec{\lambda}_R,$$

(5.4)

where $p, n, m \in \{0, 1, 2, \ldots\}$.

The presence of a single brightness species inside cells is a very unusual situation. The finite binding free energy between proteins generally leads to a mixture of associated and dissociated complexes with a composition that depends on the binding reaction and the concentration of the labeled proteins. The factorial cumulants of such a mixed system are described by a summation of the corresponding cumulants of each species,

$$\kappa_{[i,j,k]}(T) = \sum_s \kappa_{[i,j,k]}^{(s)}(T).$$ (5.5)

While it is in principle possible to deduce the composition of the mixture from Eq. 5.5 by measuring a sufficient number of cumulants $\kappa_{[i,j,k]}$, the limited signal-to-noise ratio of FFS experiments in cells makes this approach unworkable.

To avoid this dead end, we developed tricolor heterospecies partition (HSP) analysis, which is an extension of dual-color HSP [6]. Instead of trying to identify all individual components of the mixture, tricolor HSP seeks to describe the experimental cumulants $\kappa_{[i,j,k]}$ as the sum of cumulants of three apparent brightness species, $\overline{G}$, $\overline{Y}$, and $\overline{R}$,

$$\kappa_{[i,j,k]} = \sum_{s \in \{\overline{G}, \overline{Y}, \overline{R}\}} \kappa_{[i,j,k]}^{(s)} = \sum_{s \in \{\overline{G}, \overline{Y}, \overline{R}\}} \gamma_{i+j+k} \lambda_{1,s}^i \lambda_{2,s}^j \lambda_{3,s}^k B_{i+j+k}(T, \tau_{D,s}) N_s.$$ (5.6)

with brightness vectors

$$\vec{\lambda}_{\overline{G}} = \left(\lambda_{1,\overline{G}}, \lambda_{2,\overline{G}}, \lambda_{3,\overline{G}}\right), \quad \vec{\lambda}_{\overline{Y}} = \left(0, \lambda_{2,\overline{Y}}, \lambda_{3,\overline{Y}}\right), \quad \vec{\lambda}_{\overline{R}} = \left(0, 0, \lambda_{3,\overline{R}}\right)$$ (5.7)

and number of particles of $N_{\overline{G}}$, $N_{\overline{Y}}$, and $N_{\overline{R}}$. The nine cumulants $\kappa_{[i,j,k]}$ up to second order ($1 \le i+j+k \le 2$) provide enough information to identify the nine FFS parameters of the three apparent species $\left(\lambda_{1,\overline{G}}, \lambda_{2,\overline{G}}, \lambda_{3,\overline{G}}, \lambda_{2,\overline{Y}}, \lambda_{3,\overline{Y}}, \lambda_{3,\overline{R}}, N_{\overline{G}}, N_{\overline{Y}}, N_{\overline{R}}\right)$ from a fit to Eq. 5.6. This separation into apparent species is crucial, because it provides the key to distinguish between protein complexes with and without EGFP as illustrated in Figure 3A. The apparent species $\overline{G}$, which we refer to as green heterospecies, describes protein

complexes that contain EGFP, while the other apparent species $\overline{Y}$ and $\overline{R}$ characterize complexes without EGFP label. Because we focused on the characterization of protein species that carry EGFP ($G_pY_nR_m$ with $p > 0$), only the apparent species $\overline{G}$ was examined in this study. It is convenient to express the brightness vector of the green heterospecies in terms of the basis vectors $\left(\vec{\lambda}_G, \vec{\lambda}_Y, \vec{\lambda}_R \right)$

$$\vec{\lambda}_{\overline{G}} = b_G \vec{\lambda}_G + b_Y \vec{\lambda}_Y + b_R \vec{\lambda}_R . \tag{5.8}$$

Thus, the heterospecies $\overline{G}$ is characterized by the coefficients b_G, b_Y, and b_R, which are referred to as normalized brightnesses [105], and the number of particles $N_{\overline{G}}$ in the observation volume. This study only considered protein complexes with a single EGFP protein (GY_nR_m). In this case, the interpretation of the normalized brightness is straightforward,

$$b_G = 1, \quad b_Y = \frac{\sum\limits_{n,m} n N_{GYnRm}}{\sum\limits_{n,m} N_{GYnRm}}, \quad b_R = \frac{\sum\limits_{n,m} m N_{GYnRm}}{\sum\limits_{n,m} N_{GYnRm}} . \tag{5.9}$$

The value of $N_{\overline{G}} = \sum\limits_{n,m} N_{GYnRm}$ represents the total number of molecules carrying an EGFP molecule. The above relations for the green heterospecies $\overline{G}$ are derived by setting Eqs. 5.5 and 5.6 equal and applying linear algebra. The normalized brightness values of the green heterospecies describes an average over all stoichiometries of GY_nR_m weighted by their concentration in the sample, which can be formally expressed by a stoichiometric formula $\overline{G} = GY_{b_Y}R_{b_R}$ with non-integer coefficients, where b_Y and b_R
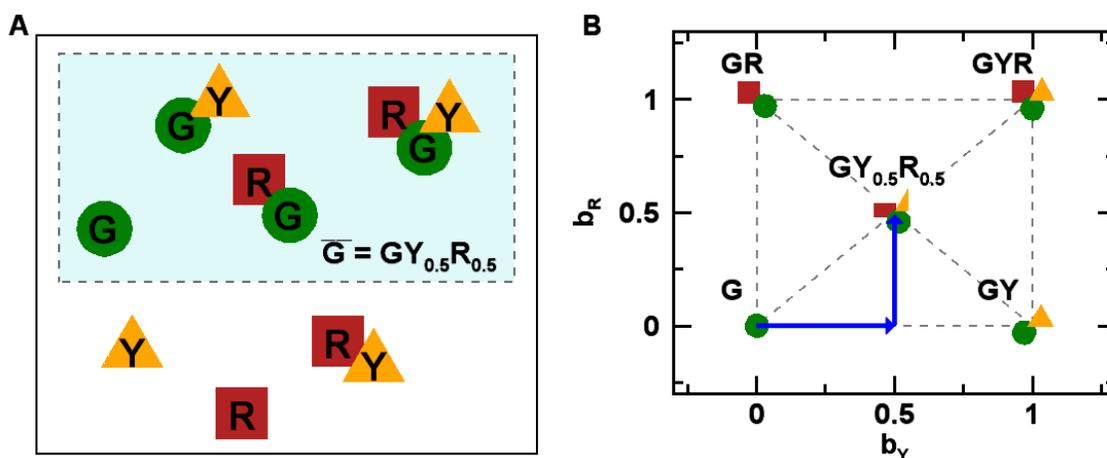
Figure 5.4 Conceptual illustration of tricolor HSP analysis and (b_Y, b_R)-plot.

(A) A mixture of brightness species G, GY, GR, GYR, Y, YR, and R is separated into groups by HSP. The heterospecies $\overline{G}$ contains all brightness species containing G, which in this example applies to species G, GY, GR, and GYR. The average stoichiometry of $\overline{G}$ is $\overline{G}=GY_{0.5}R_{0.5}$. (B) The average stoichiometry of the heterospecies $\overline{G}=GY_{b_Y}R_{b_R}$ is graphically displayed by plotting the normalized brightness b_Y versus b_R. For example, $\overline{G}=GY_{0.5}R_{0.5}$ is represented by point (0.5, 0.5) in the (b_Y, b_R)-plot.

represent the average number of EYFP and mCherry labels associated with a single copy of EGFP ($b_G = 1$) as illustrated in Figure 5.4A.

Since we assume a b_G of one in this study, only two coefficients, b_Y and b_R, are needed to specify the average oligomeric state of the green heterospecies. Plotting a point with coordinates (b_Y, b_R) provides a graphical representation of the composition of the green heterospecies $GY_{b_Y}R_{b_R}$. The points $(0, 0)$, $(1, 0)$, $(0, 1)$ and $(1, 1)$ in Figure 5.4B represent the states G , GY , GR and GYR , respectively. The values of b_Y and b_R specify the degree of binding of Y and R to G. Thus, the point $(0.5, 0.5)$ identifies a heterospecies $\overline{G}$ where each G is on average associated with half a Y and R protein (Figure 5.4).

We evaluated tricolor FFS experimentally using well-characterized proteins. For the first control experiment we performed tricolor FFS on cells expressing EGFP, EYFP and mCherry. Since these proteins are non-interacting, tricolor HSP analysis is expected to identify EGFP as the green heterospecies $(\overline{G}=G)$ with $\vec{\lambda}_{\overline{G}} = 1 \cdot \vec{\lambda}_G + 0 \cdot \vec{\lambda}_Y + 0 \cdot \vec{\lambda}_z$, which corresponds to location $(0, 0)$ on the (b_Y, b_R)-plot. The experimentally determined normalized brightnesses b_Y and b_R localized near $(0, 0)$ (Figure 5.5A), confirming that tricolor-HSP identifies non-interacting EGFP. The normalized brightness b_G of the sample (box plot, Figure 5.6) is close to one, which confirms the presence of monomeric EGFP.

The next experiments measured the ligand binding domain (LBD) of the nuclear receptors RAR and RXR, which were abbreviated as AL and XL, respectively. The
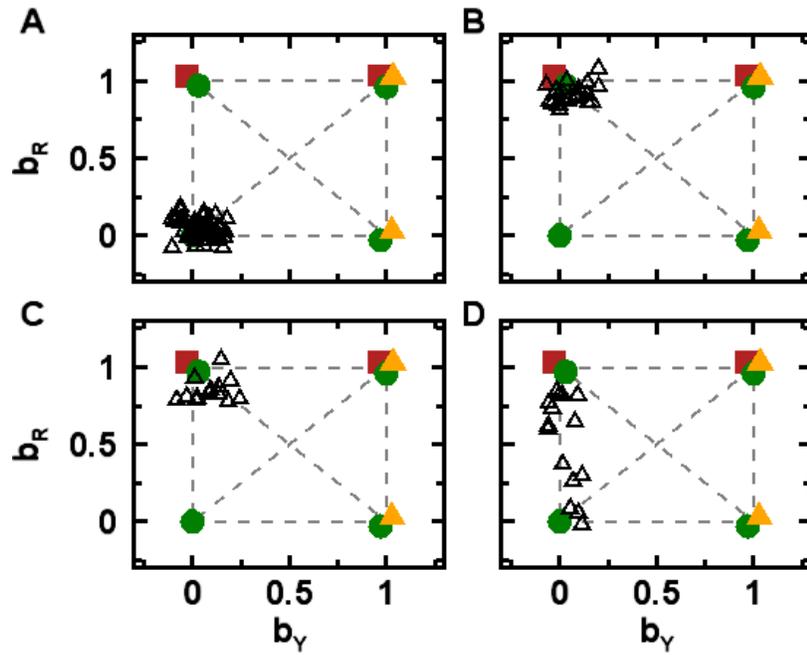
Figure 5.5 (b_Y, b_R)**-plot of non-interacting systems.**

The (b_Y, b_R) values of heterospecies $\overline{G}$ are graphed for cells expressing (A) EGFP + EYFP + mCherry, (B) G-AL-R, (C) G-AL-R + EYFP + mCherry, and (D) G-AL + G-AL-R.
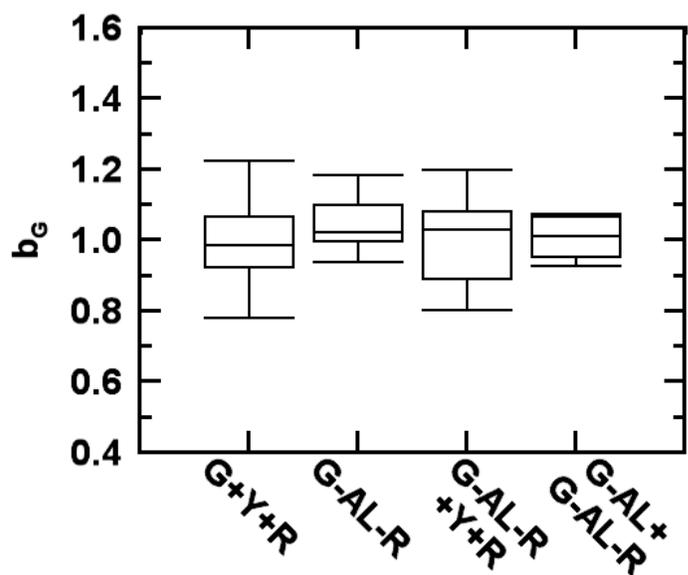
Figure 5.6 Box plot of the normalized brightness b_G.

Each of the four samples of Figure 5.5 are graphed. The median of b_G is in all cases close to one, which confirms a monomeric stoichiometry of EGFP for the heterospecies $\overline{G}$.

fluorescent tag is identified by adding the label G, Y, or R before or after the protein name. For example, G-AL-R represents AL labeled with EGFP at the N-terminus and with mCherry at the C-terminus. AL serves as a spacer to significantly reduce FRET between both labels [28]. We used this construct as a model to mimic the heterodimeric state GR . Tricolor HSP analysis of cells expressing G-AL-R identified (b_Y, b_R) -data clustered around (0, 1), while b_G was close to 1 (Figures 5.5B and 5.6). Thus, the analysis successfully recovered the heterospecies state $\overline{G}$=GR .

We repeated this experiment on cells that expressed EYFP and mCherry in addition to G-AL-R. Since neither EYFP nor mCherry interacts with G-AL-R, HSP analysis should exclude their contributions from the heterospecies $\overline{G}$. Thus, we expected once again to observe the state GR , as was confirmed by experiment (Figures 5.5C and 5.6). For the next experiment, we expressed G-AL-R together with G-AL. These two proteins do not interact [4] and therefore mimic a mixture of monomers G and heterodimers GR. We expected normalized brightness values of $b_G = 1$ and $b_Y = 0$, while the value of b_R would depend on the concentration fraction of heterodimers, $b_R = N_{GR}/(N_G + N_{GR})$. In other words, b_R = 0, if the cell expresses only the monomer G, while b_R = 1, if the cell produces only the heterodimer GR. Any mixture of the two species results in a b_R between these two extrema. The experimental (b_Y, b_R) -data from tricolor HSP analysis (Figures 5.5D and 5.6) were distributed along a line connecting the points (0, 0) and (0, 1), which confirms the presence of a mixture of G and GR. The cell-to-cell variation in the protein expression ratio [105] changes the relative composition of the two states and gives rise to the scatter in the brightness data..

A box plot of b_G for each experiment discussed so far is shown in Figure 5.6. Box plots of b_G for all additional experiments of this study are found in Figure 5.7 for reference. The data demonstrate that the normalized brightness b_G was approximately one for all experiments discussed in this work. Thus, we no longer mention b_G from now on and solely focus on the behavior of the normalized brightness pair (b_Y, b_R).

We next investigated two interacting proteins, starting with AL and XL, which form a very tight heterodimer [35]. FFS data from cells expressing G-AL and Y-XL were analyzed by tricolor HSP. The (b_Y, b_R)-data are located along the line connecting $(0, 0)$ and $(1, 0)$, indicating the presence of G-AL monomers and G-AL / Y-XL dimers (Figure 5.8A). We reasoned that the monomeric populations reflect an excess of G-AL over Y-XL in the measured cell, so that not all G-AL have a binding partner. To test this idea we needed to filter the data based on the expressed protein concentration.

We used the following procedure to identify the total expressed concentration of each labeled protein. The symbols $N_{G,tot}$, $N_{Y,tot}$, and $N_{R,tot}$ were chosen to represent the total number concentration of EGFP-, EYFP-, and mCherry-labeled proteins, respectively. The average fluorescence intensity in each channel is the matrix product of the brightness values of Eq. 5.3 with the total labeled protein concentrations,

$$\begin{pmatrix} \langle F_1 \rangle \\ \langle F_2 \rangle \\ \langle F_3 \rangle \end{pmatrix} = \begin{pmatrix} \lambda_{1,G} & 0 & 0 \\ \lambda_{2,G} & \lambda_{2,Y} & 0 \\ \lambda_{3,G} & \lambda_{3,Y} & \lambda_{3,R} \end{pmatrix} \begin{pmatrix} N_{G,tot} \\ N_{Y,tot} \\ N_{R,tot} \end{pmatrix}. \tag{5.10}$$
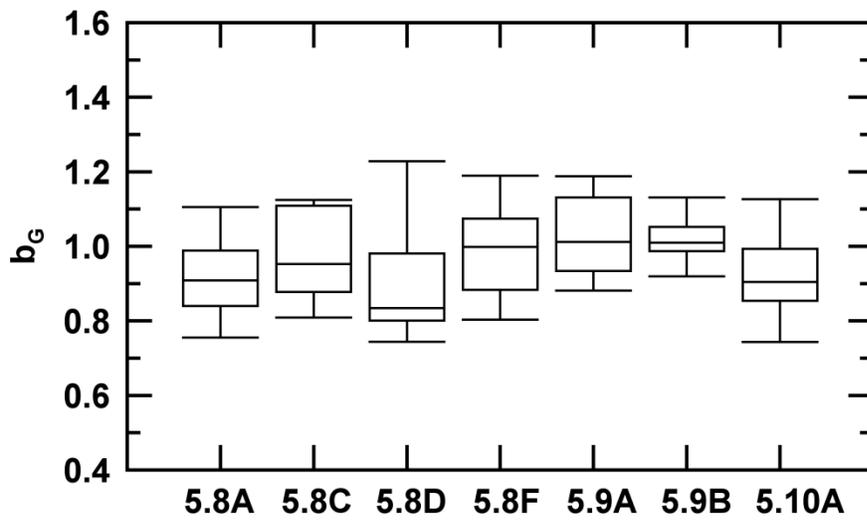
Figure 5.7 Normalized brightness b_G **of samples shown in Figures 5.8, 5.9 and 5.10.**

Box plot of b_G for Figures 5.8A, 5.8C, 5.8D, 5.8F, 5.9A, 5.9, and 5.10A.
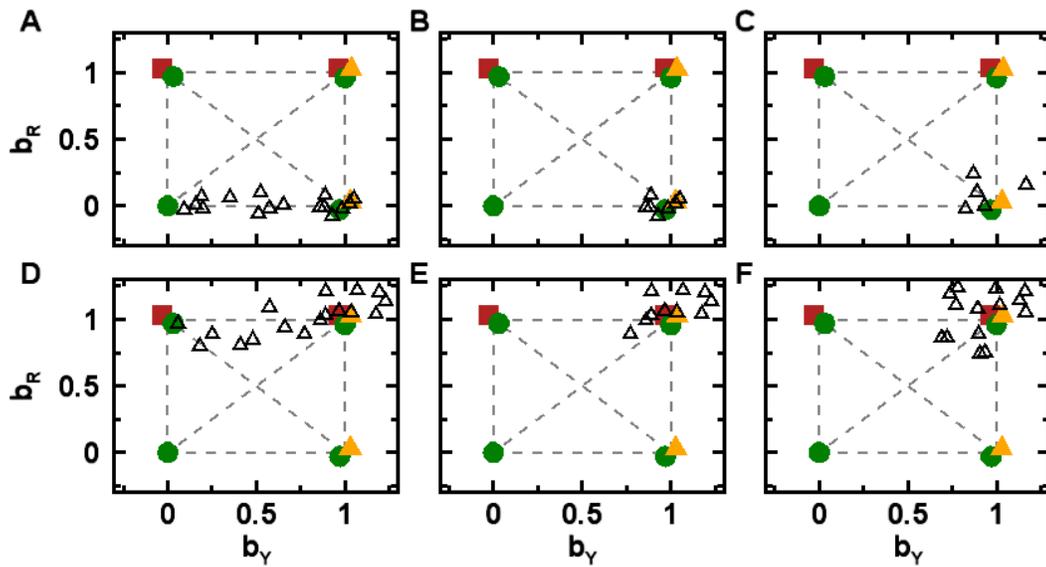
Figure 5.8 (b_Y, b_R)**-plot of heterospecies** $\overline{\text{G}}$ **for interacting systems.**

The (b_Y, b_R) values of the green heterospecies are graphed for different samples. (A) Cells expressing G-AL + Y-XL. (B) Subset of the data in A satisfying the condition $N_{G,tot} < N_{Y,tot}$. (C) Cells expressing G-AL + Y-XL + R satisfying the condition $N_{G,tot} < N_{Y,tot}$. (D) Cells expressing G-AL-R + Y-XL. (E) Subset of the data in D satisfying the condition $N_{G,tot} < N_{Y,tot}$. (F) Cells expressing G-AL-R + Y-XL + and R that satisfy the condition $N_{G,tot} < N_{Y,tot}$.

We applied Eq. 5.10 to determine the values of $\vec{N}_{tot} = \left(N_{G,tot}, N_{Y,tot}, N_{R,tot}\right)$ from the experimentally measured intensities of each channel, $\overrightarrow{\langle F \rangle} = \left(\langle F_1 \rangle, \langle F_2 \rangle, \langle F_3 \rangle\right)$.

We applied the above procedure to the data and selected cells with an excess of the Y-XL binding partner over G-AL, $N_{G,tot} < N_{Y,tot}$. After applying this data cut, the $\left(b_Y, b_R\right)$-data (Figure 5.8B) were firmly in the vicinity of (1, 0), which is consistent with the presence of hetero-dimeric G-AL / Y-XL. We repeated the same exercise on cells expressing G-AL, Y-XL, and mCherry. Because mCherry is not interacting with the G-labeled proteins, we expected to recover a hetero-species $\overline{G}$ with $\left(b_Y, b_R\right)$ values close to (1, 0) as confirmed by the experiment (Figure 5.8C).

The experiments shown in Figures 5.8A, B and C were repeated but with G-A-R instead of G-A. The $\left(b_Y, b_R\right)$ values are shifted up by one unit (Figures 5.8D, E and F) as compared to the previous data, which reflects the additional presence of the R label in the heterospecies. The measured cells contain G-AL-R and G-AL-R / Y-XL, which mimics a mixture of a heterodimer GR and a heterotrimer GYR (Figure 5.8D). To generate a pure sample of G-AL-R / Y-XL complexes we chose cells with an excess of Y-XL over G-AL-R ($N_{G,tot} < N_{Y,tot}$). This selection (Figures 5.8E and F) resulted in $\left(b_Y, b_R\right)$ values close to (1, 1), which is consistent with the expected brightness state of G-AL-R / Y-XL complexes.

The next set of experiments focused on the expression of three proteins. We first measured a mixture of G-AL, Y-XL, and R-XL in cells. In order to ensure full binding to G-AL we only considered cells with an excess of the XL binding partner,

169

$N_{G,tot} < N_{Y,tot} + N_{R,tot}$. Because Y-XL and R-XL compete for binding to the single binding site of G-AL, there are only two forms of complexes possible, G-AL / Y-XL and G-AL / R-XL. These two complexes correspond to HSP brightness pairs (b_Y, b_R) of (1, 0) and (0, 1), which identify the heterodimers GY and GR. The experimental HSP brightness values fall along a line connecting these two points (Figure 5.9A). The position along the line reflects the composition of the two protein complexes and depends on the relative expression ratio of Y-XL and R-XL.

The same type of experiment was performed with the full-length receptors RAR and RXR, which were denoted as A and X, respectively. While the ligand binding domain of the receptors formed a very tight hetero-dimer, the full-length receptors interact more weakly and provide an opportunity to illustrate the effect of the binding reaction $A + X \rightleftarrows AX$ on tricolor FFS. Cells expressing G-A, Y-X, and R-X were studied by tricolor HSP to determine the normalized brightness of the heterospecies $\overline{G}$. The (b_Y, b_R)-data from all measured cells (Figure 5.9B) are distributed within a triangle with vertexes (0, 0), (1, 0), and (0, 1). These vertexes correspond to the states G-A, G-A / Y-X, and G-A / R-X. Because the heterospecies $\overline{G}$ must be composed of a mixture of these states, the normalized brightness (b_Y, b_R) cannot lie outside the triangle.

We selected measurements with an excess of X over A ($N_{G,tot} < N_{Y,tot} + N_{R,tot}$) to provide a binding partner for each G-A protein. This step removes the (b_Y, b_R)-values around (0, 0) and the remaining data clearly indicate interaction between both proteins (Figure 5.9C). However, unlike for the case of the ligand-binding domain (Figure 5.9A),
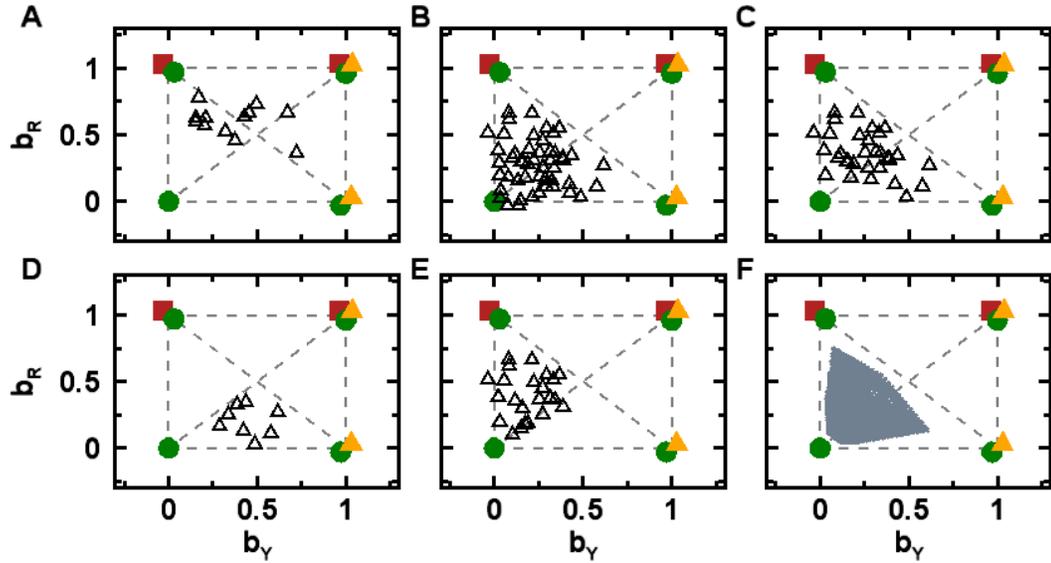
Figure 5.9 (b_Y, b_R) **-plot representing competitive binding of a ternary protein system.**

The (b_Y, b_R) values of the heterospecies $\overline{G}$ are shown for a system with binding interactions $G+Y \leftrightarrow GY$ and $G+R \leftrightarrow GR$. (A) Data from cells expressing G-AL + Y-XL + R-XL. (B) Data from cells expressing G-A + Y-X + R-X. (C) Subset of the data in B satisfying the condition $N_{G,tot} < N_{Y,tot} + N_{R,tot}$. (D) Subset of the data in C satisfying the condition $N_{Y,tot} > N_{R,tot}$. (E) Subset of the data in C satisfying the condition $N_{Y,tot} < N_{R,tot}$. (F) Modeling of (b_Y, b_R) for the reactions $G+Y \leftrightarrow GY$ and $G+R \leftrightarrow GR$ with a $K_D = 1.3$ μM covering the same concentration range as the data shown in panel C.

the brightness data do not align along the line connecting (1, 0) and (0, 1), which reflects the weaker binding affinity of the full-length receptors. We applied a further data cut by selecting measurements with an excess of Y-X over R-X ($N_{Y,tot} > N_{R,tot}$), which is shown in Figure 5.9D. The opposite selection $N_{Y,tot} < N_{R,tot}$ is shown in Figure 7E. This selection cuts the data symmetrically along the symmetry axis of the triangle and demonstrates that binding of Y-X and R-X to G-A has approximately the same affinity. This is of course expected, since Y-X and R-X are identical with the exception of the fluorescent label.

Finally, we performed a simple modeling of the reactions G-A+Y-X $\rightleftarrows$ G-A/Y-X and G-A+R-X $\rightleftarrows$ G-A/R-X using the same dissociation coefficient K_D for both reactions. The range of measured concentrations $N_{Y,tot}$ and $N_{R,tot}$ was used to calculate the concentration of each of the brightness species for the above reactions, which were subsequently converted into cumulants (Eqs. 5.2 and 5.5) and analyzed by HSP to determine the normalized brightness values of the heterospecies $\overline{G}$. All calculations were performed using an excess of the binding partner over G-A to mimic the conditions of Figure 5.9C. We adjusted the K_D until the spread of the modeled (b_Y, b_R)-values matched the experimental measured scatter of Figure 5.9C. The result of the modeling for a K_D of 1.3 μM is shown in Figure 5.9F.

Nuclear receptor binding to coactivator represents the first step in the regulation of transcription [7]. The next experiment probed the interaction of the coactivator TIF2, which we denote as T, with the nuclear receptor pair RAR and RXR. Cells expressing G-

T, Y-A, and R-X were measured and analyzed by tricolor HSP to determine the normalized brightness associated with the heterospecies $\overline{G}$. We selected data with $N_{G,tot} < N_{Y,tot}$ and $N_{G,tot} < N_{R,tot}$ to guarantee that a Y-A and R-X binding partner is available for each G-T protein. The (b_Y, b_R)-data clearly indicate interaction between the nuclear receptors and the coactivator (Figure 5.10A). There is no binding competition of the nuclear receptors to the coactivator, because the data are not confined to a triangle with vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$. Furthermore, the data are asymmetrically distributed, because the triangle with vertices (0, 0), (0, 1), and (1, 1) is essentially not populated (Figure 5.10A). Since the protein expression level of Y-A and R-X covered by the experiments were very similar, the absence of data points in the upper triangle indicates that the interaction between RXR and TIF2 has to be weak compared to other interactions.

To more readily visualize the binding of RAR to TIF2 we selected data points with an excess of Y-A over R-X ($N_{Y,tot} > 2N_{R,tot}$, Figure 5.10B), which results in (b_Y, b_R)-data that are relatively close to the line connecting the points (0, 0) and (1, 0), which is consistent with a heterospecies $\overline{G}$ that consists predominantly of a mixture of free coactivator G-T and coactivator bound to RAR (G-T / Y-A). On the other hand, when selecting cell with an excess of R-X over Y-A ($N_{Y,tot} < N_{R,tot}$, Figure 5.10C), the brightness data are closely aligned with the diagonal connecting (0, 0) and (1, 1). This result indicates that the heterospecies $\overline{G}$ consists of a mixture of G-T and the complex G-T / Y-A / R-X, or if we omit the labels for clarity, the mixture consists of free coactivator T and the complex TAX. In other words, if X is not a limiting factor, we observe TAX
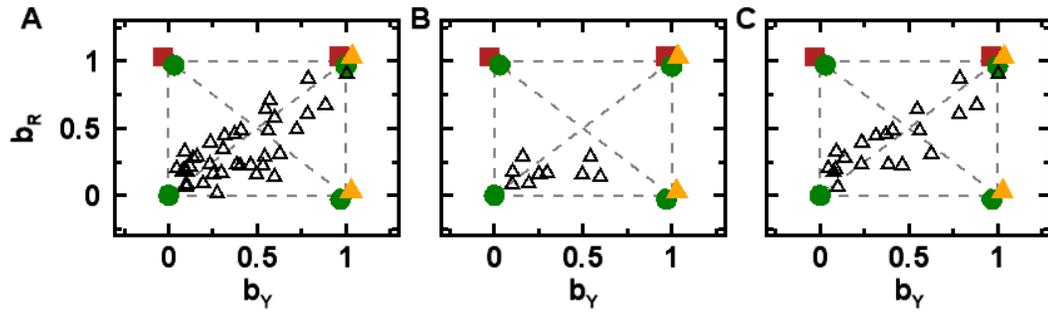
Figure 5.10 (b_Y, b_R)**-plot for the CoA TIF-2 interacting with the NRs RXR and RAR.**

Cells expressing G-T, Y-A and R-X were analyzed by tricolor HSP to extract the heterospecies $\overline{G}$. (A) (b_Y, b_R) of cells that fulfill the conditions $N_{G,tot} < N_{Y,tot}$ and $N_{G,tot} < N_{R,tot}$. (B) Data subset from A that meets the condition $N_{Y,tot} > 2N_{R,tot}$. The brightness points are close to the line connecting states G and GY. (C) Data subset from A that meets the condition $N_{Y,tot} < N_{R,tot}$. The brightness values fall along the diagonal connecting the states G and GYR.

complexes rather than a mix of TA and TAX. This implies that the binding affinity of the NR dimer AX to the coactivator T is much stronger than binding of A to T. This agrees with the generally accepted model in which the coactivator is recruited to the NR dimer [8]. Our finding suggests that the system has two binding reactions, T+A $\rightleftarrows$ TA and T+AX $\rightleftarrows$ TAX, where the latter is the dominant and biologically relevant interaction, because transcription activation requires the coactivator / NR dimer complex [8]. In this study we focused on the interactions between CoA and NRs in the absence of ligands. It is well known that the binding between CoA and NRs is modulated by ligands [7,8], Thus, it would be interesting to directly observe the influence of ligands on the formation of the ternary protein complex in future tricolor HSP experiments.

The distribution of normalized brightness values (b_Y, b_R) provides insight into the interactions of a ternary protein system as demonstrated in this work. Clustering of all (b_Y, b_R) around one location indicates the presence of one dominant brightness species in the sample. If the data form a line connecting two brightness states, the heterospecies is a mix of these two states. Figures 5.5 and 5.8 provide examples for both cases. If the data spread out over an area, then more than two states are involved. The distribution provides clues about the nature of the protein interactions. For example, binding competition for the same binding site confines the brightness data to a specific region of the (b_Y, b_R) plot (Figure 5.9B). Applying data cuts based on the expression level of the three proteins is another helpful tool to decipher the interactions that are taking place. By selecting data with excess of one protein over the others, we were able to gauge the relative strength of different interactions (Figures 5.9D&E and 5.10B&C). While we have not yet examined

the effect of ligands that are known to modulate CoA and NR binding [7,8], it would be interesting to directly observe their influence on complex formation in future experiments.

We only had to track b_Y and b_R of the normalized brightness values $\left(b_G, b_Y, b_R\right)$, because of the absence of homointeractions among the EGFP-labeled proteins as evidenced by $b_G \approx 1$ (Figures 5.6 and 5.7). In the presence of homointeractions of the EGFP-labeled protein, the b_G value increases. In this case, HSP modeling needs to be updated and all three normalized brightness values are required to characterize the averaged composition of the heterospecies $\overline{G}$. Another requirement for HSP is the use of fluorescent proteins that are suitable for quantitative brightness experiments. EGFP, EYFP, and mCherry meet this requirement and have been extensively characterized by us [4,28]. These three proteins also have the advantage of being coexcited at 990 nm, which simplifies the optical setup of the experiment. The models formulated in this work assume that the fluorescent proteins exist to good approximation in a single brightness state. While this is not strictly fulfilled for mCherry [28], the deviation from the model is negligible as long as the heterospecies $\overline{G}$ contains up to one mCherry-labeled protein. If $\overline{G}$ involves multiple mCherry-labeled proteins, a more sophisticated modeling of the cumulants (Eq. 5.2) has to be applied [6,28].

We ignored FRET in our analysis, because its influence on brightness is negligible for the protein systems studied. Previous work demonstrated that an average FRET efficiency of $E\%$ leads to a brightness decrease of the donor by $E\%$, while the brightness of the acceptor typically increases by a similar amount [28,35]. Since the estimated uncertainty of normalized HSP brightness values from Figures 5.5, 5.6, and 5.8

is ~15%, the FRET efficiency would have to significantly exceed this value to be clearly noticeable in the graphical representation of the data.

The median of the normalized brightness b_G of the measured samples (Figures 5.6 and 5.7) ranges from ~1 to ~0.9, which implies that the FRET efficiency E of the donor EGFP is less than 15%. Estimating the FRET efficiency between EYFP and mCherry is less straightforward. The only protein complex we have to consider here is the ternary complex of CoA and the NR dimer of RAR and RXR labeled with EYFP and mCherry. To investigate the presence of FRET between RAR and RXR we want to maximize the fraction of heterodimer in the sample, because this produces the highest average FRET efficiency. The LBD domains of these receptors have a much higher binding affinity than the full length proteins, which should facilitate the detection of a FRET signal. A measurement of the LBD domains using EGFP and EYFP as labels (Figure 5.8C) resulted in a normalized brightness b_G close to one (Figure 5.7), which indicates a FRET efficiency of less than 15%. Thus, we conclude that the influence of FRET on tricolor HSP is negligible in this study. We expect that the analysis strategy presented in this work is more widely applicable, because FRET efficiencies of tagged protein complexes are frequently less than ~15%.

There are relatively few reports of fluorescence-based measurements of ternary protein interactions in the literature. While most of these studies have been carried out in vitro, a few studies in cells using FRET and FRET-based BiFC exist [133,134]. FRET analysis between three different chromophores is challenging and probably limited to cellular complexes that can achieve relatively high FRET efficiencies. Combining BiFC

with FRET is a clever approach and the data analysis is relatively straightforward. However, the covalent linking of proteins by BiFC is a disadvantage as it prevents dissociation of the complex and thus the study of binding equilibria. Tricolor HSP introduces an alternative approach for studying ternary protein interactions in cells. Because HSP uses the intensity fluctuations from complexes passing through the optical observation volume to deduce the make-up of the interacting heteroprotein assembly, it can be applied to systems that have little or no FRET.

We introduced tricolor HSP as a novel method for identifying interactions of ternary protein systems. While not included in the current form of HSP, it should be possible to account for FRET in a more sophisticated model of tricolor HSP. Also, this study focused on qualitative interpretation of HSP data. A simple modeling attempt was shown in Figure 5.9F, which was used to estimate the dissociation coefficient. It would be interesting to explore quantitative modeling of the interactions by accounting for protein concentration and the normalized brightness data in future studies. The goal of this work was to establish the foundation of tricolor HSP. We developed the theory, demonstrated experimental feasibility, and successfully identified the formation of a heterotrimeric complex of one CoA with two NRs. These results indicate the significant potential of tricolor HSP for the characterization of ternary protein systems in the living cell.

6. Summary

FFS has become a powerful tool for studying biomolecular association inside living cells. The technique measures protein-protein interactions quantitatively, non-invasively and in real time. Because protein assemblies play a central role in the regulation of biological processes in the cell, the range of potential applications for FFS is vast. However, applying FFS inside cells is surprisingly challenging. The cell environment influences intensity fluctuations in a number of subtle ways that are not widely recognized. Because conventional FFS theory was developed assuming an aqueous solution as the sample, the existing theory does not account for the cell environment. Thus, there is a considerable risk that the interpretation of any FFS experiment in cells may be flawed. Our group has made a significant effort to transform FFS in cells into a robust method. Part of my thesis was devoted to this ongoing effort by investigating the effect of photodepletion on FFS.

The small volume of yeast cells leads to a modest but systematic depletion of fluorescent proteins. Although the decrease in fluorescence amplitude is generally very small, the calculated brightness can be significantly distorted as observed in Chapter 3. This failure of conventional FFS analysis was traced to the nonstationary nature of the fluorescence signal, while the theory implicitly assumes a stationary process. This problem was solved by dividing the stream of collected photon counts into individual segments. The segment length was chosen so that the nonstationary effect on a single segment was negligible. This approach allowed us to analyze individual segments by applying conventional FFS analysis, while the effect of photodepletion was modeled separately to account for the change in brightness from segment to segment.

The work of Chapter 3, while an important first step, did not yet achieve the inclusion of nonstationary signals into FFS theory. In fact, the method introduced in that chapter breaks down when performing FFS in very small sample compartments as demonstrated by experiments on E. coli bacteria. To overcome this challenge, Chapter 4 was devoted to formulating a first theory that extends FFS to nonstationary signals, which led to the development of MSQ analysis. We derived an analytical MSQ model that incorporated photodepletion, and applied it successfully to determine the oligomeric state of proteins in E. coli cells. The last step also required proper accounting for the size and shape of the E. coli cell, which was accomplished with z-scan FFS. This study provides the necessary framework for applying FFS in small sample compartments of the living cell, such as mitochondria and nucleoli.

Another shortcoming of FFS was its limitation to single- and dual-color experiments. While this is adequate for studying interactions involving one or two proteins, another color needs to be added to probe the interactions of ternary protein systems. We addressed this issue in Chapter 5 by introducing tricolor HSP analysis, which we applied to study the interactions between a coactivator and two nuclear receptors.

This thesis provides the foundation for extending FFS studies to ternary protein systems and to new cell compartments. Chapter 5 used graphs for qualitative assessment of interactions in ternary protein systems. We are currently working on a manuscript that changes tricolor FFS into a quantitative analysis tool. Such an extension promises to provide deeper insight into the regulation of ternary interactions. For example, we are

able to quantify the influence of ligands on the formation of the ternary complex involving RXR, RAR, and TIF2.

We have also begun to apply FFS to the nuclear envelope which separates the nucleus from the cytoplasm and consists of a double membrane separated by a ~40 nm fluid layer. The nuclear envelope is a very challenging environment for FFS studies. In fact, we were not able to interpret the FFS data until MSQ analysis was applied and revealed the presence of an additional process that obscured the interpretation of the data. We also realized that fluctuation imaging techniques are particularly susceptible to the estimator bias described in Chapter 4. Thus, we expect that MSQ will prove important for robust scanning and imaging FFS studies. In addition, we recently discovered that MSQ has surprising advantages over the autocorrelation function in resolving a mixture of fast and slowly diffusing molecules. While all these studies are still in progress, the preliminary results indicate that MSQ analysis could have a much larger impact on FFS than originally anticipated in Chapter 4.

References

1. Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell. 1998;92: 291–294.

2. García-Sáez AJ. The secrets of the Bcl-2 family. Cell Death Differ. 2012;19: 1733–1740.

3. Raboni S, Bettati S, Mozzarelli A. Tryptophan synthase: a mine for enzymologists. Cell Mol Life Sci CMLS. 2009;66: 2391–2403.

4. Chen Y, Wei L-N, Müller JD. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. Proc Natl Acad Sci U S A. 2003;100: 15492–15497.

5. Chen Y, Müller JD. Determining the stoichiometry of protein heterocomplexes in living cells with fluorescence fluctuation spectroscopy. Proc Natl Acad Sci. 2007;104: 3147–3152.

6. Wu B, Chen Y, Müller JD. Heterospecies partition analysis reveals binding curve and stoichiometry of protein interactions in living cells. Proc Natl Acad Sci. 2010;107: 4117–4122.

7. Gronemeyer H, Gustafsson J-A, Laudet V. Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov. 2004;3: 950–964.

8. de Lera AR, Bourguet W, Altucci L, Gronemeyer H. Design of selective nuclear receptor modulators: RAR and RXR as a case study. Nat Rev Drug Discov. 2007;6: 811–820.

9. Tsien RY. The green fluorescent protein. Annu Rev Biochem. 1998;67: 509–544.

10. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. Nat Methods. 2005;2: 905–909.

11. Chudakov DM, Lukyanov S, Lukyanov KA. Fluorescent proteins as a toolkit for in vivo imaging. Trends Biotechnol. 2005;23: 605–613.

12. Truong K, Ikura M. The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. Curr Opin Struct Biol. 2001;11: 573–578.

13. Kenworthy AK. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. Methods San Diego Calif. 2001;24: 289–296.

14. Kerppola TK. Complementary methods for studies of protein interactions in living cells. Nat Methods. 2006;3: 969–971.

15. Kerppola TK. Bimolecular fluorescence complementation: visualization of molecular interactions in living cells. Methods Cell Biol. 2008;85: 431–470.

16. Shyu YJ, Hu C-D. Fluorescence complementation: an emerging tool for biological research. Trends Biotechnol. 2008;26: 622–630.

17. Magde D, Elson E, Webb WW. Thermodynamic Fluctuations in a Reacting System—Measurement by Fluorescence Correlation Spectroscopy. Phys Rev Lett. 1972;29: 705–708. doi:10.1103/PhysRevLett.29.705

18. Elson EL, Magde D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers. 1974;13: 1–27.

19. Thompson NL. Fluorescence Correlation Spectroscopy. In: Lakowicz JR, editor. Topics in Fluorescence Spectroscopy. Springer US; 2002. pp. 337–378.

20. Krichevsky O, Bonnet G. Fluorescence correlation spectroscopy: the technique and its applications. Rep Prog Phys. 2002;65: 251.

21. Thompson NL, Lieto AM, Allen NW. Recent advances in fluorescence correlation spectroscopy. Curr Opin Struct Biol. 2002;12: 634–641.

22. Ries J, Schwille P. Fluorescence correlation spectroscopy. BioEssays News Rev Mol Cell Dev Biol. 2012;34: 361–368.

23. Meseth U, Wohland T, Rigler R, Vogel H. Resolution of fluorescence correlation measurements. Biophys J. 1999;76: 1619–1631.

24. Chen Y, Müller JD, So PT, Gratton E. The photon counting histogram in fluorescence fluctuation spectroscopy. Biophys J. 1999;77: 553–567.

25. Digman MA, Wiseman PW, Choi C, Horwitz AR, Gratton E. Stoichiometry of molecular complexes at adhesions in living cells. Proc Natl Acad Sci U S A. 2009;106: 2170–2175.

26. Saffarian S, Li Y, Elson EL, Pike LJ. Oligomerization of the EGF Receptor Investigated by Live Cell Fluorescence Intensity Distribution Analysis. Biophys J. 2007;93: 1021–1031.

27. Sergeev M, Costantino S, Wiseman PW. Measurement of monomer-oligomer distributions via fluorescence moment image analysis. Biophys J. 2006;91: 3884–3896.

28. Wu B, Chen Y, Müller JD. Fluorescence Fluctuation Spectroscopy of mCherry in Living Cells. Biophys J. 2009;96: 2391–2404.

29. Chen Y, Müller JD, Ruan Q, Gratton E. Molecular Brightness Characterization of EGFP In Vivo by Fluorescence Fluctuation Spectroscopy. Biophys J. 2002;82: 133–144.

30. Andersen PS, Schuck P, Sundberg EJ, Geisler C, Karjalainen K, Mariuzza RA. Quantifying the energetics of cooperativity in a ternary protein complex. Biochemistry (Mosc). 2002;41: 5177–5184.

31. Houtman JCD, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, et al. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. Protein Sci. 2007;16: 30–42.

32. O'Connell CB, Khodjakov AL. Cooperative mechanisms of mitotic spindle formation. J Cell Sci. 2007;120: 1717–1722.

33. Agnati LF, Guidolin D, Leo G, Fuxe K. A boolean network modelling of receptor mosaics relevance of topology and cooperativity. J Neural Transm. 2006;114: 77–92.

34. Courey AJ. Cooperativity in transcriptional control. Curr Biol CB. 2001;11: R250–252.

35. Chen Y, Wei L-N, Müller JD. Unraveling protein-protein interactions in living cells with fluorescence fluctuation brightness analysis. Biophys J. 2005;88: 4366–4377.

36. Bacia K, Kim SA, Schwille P. Fluorescence cross-correlation spectroscopy in living cells. Nat Methods. 2006;3: 83–89.

37. Lonard DM, O'malley BW. Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. Mol Cell. 2007;27: 691–700.

38. McKenna NJ, O'Malley BW. Combinatorial control of gene expression by nuclear receptors and coregulators. Cell. 2002;108: 465–474.

39. Nettles KW, Greene GL. Ligand control of coregulator recruitment to nuclear receptors. Annu Rev Physiol. 2005;67: 309–333.

40. Rosenfeld MG, Lunyak VV, Glass CK. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes Dev. 2006;20: 1405–1428.

41. Hu X, Lazar MA. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. Nature. 1999;402: 93–96.

42. Lin BC, Hong SH, Krig S, Yoh SM, Privalsky ML. A conformational switch in nuclear hormone receptors is involved in coupling hormone binding to corepressor release. Mol Cell Biol. 1997;17: 6131–6138.

43. Nagy L, Kao HY, Love JD, Li C, Banayo E, Gooch JT, et al. Mechanism of corepressor binding and release from nuclear hormone receptors. Genes Dev. 1999;13: 3209–3216.

44. Fletcher TM, Ryu BW, Baumann CT, Warren BS, Fragoso G, John S, et al. Structure and dynamic properties of a glucocorticoid receptor-induced chromatin transition. Mol Cell Biol. 2000;20: 6466–6475.

45. Hager GL, Fletcher TM, Xiao N, Baumann CT, Müller WG, McNally JG. Dynamics of gene targeting and chromatin remodelling by nuclear receptors. Biochem Soc Trans. 2000;28: 405–410.

46. Fletcher TM, Xiao N, Mautino G, Baumann CT, Wolford R, Warren BS, et al. ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. Mol Cell Biol. 2002;22: 3255–3263.

47. Nagaich AK, Hager GL. UV Laser Cross-linking: A Real-Time Assay to Study Dynamic Protein/DNA Interactions During Chromatin Remodeling. Sci Signal. 2004;2004: pl13–pl13.

48. Martinez ED, Dull AB, Beutler JA, Hager GL. High-content fluorescence-based screening for epigenetic modulators. Methods Enzymol. 2006;414: 21–36.

49. Foryst-Ludwig A, Clemenz M, Hohmann S, Hartge M, Sprang C, Frost N, et al. Metabolic Actions of Estrogen Receptor Beta (ERβ) are Mediated by a Negative Cross-Talk with PPARγ. PLoS Genet. 2008;4.

50. Dudek P, Picard D. Genomics of signaling crosstalk of estrogen receptor alpha in breast cancer cells. PloS One. 2008;3: e1859.

51. Ogawa S, Lozach J, Benner C, Pascual G, Tangirala RK, Westin S, et al. Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. Cell. 2005;122: 707–721.

52. De Bosscher K, Haegeman G. Minireview: latest perspectives on antiinflammatory actions of glucocorticoids. Mol Endocrinol Baltim Md. 2009;23: 281–291.

53. Paumelle R, Staels B. Cross-talk between statins and PPARalpha in cardiovascular diseases: clinical evidence and basic mechanisms. Trends Cardiovasc Med. 2008;18: 73–78.

54. Kassel O, Herrlich P. Crosstalk between the glucocorticoid receptor and other transcription factors: molecular aspects. Mol Cell Endocrinol. 2007;275: 13–29.

55. Sun G, Zhu J, Chen L, Raikhel AS. Synergistic action of E74B and ecdysteroid receptor in activating a 20-hydroxyecdysone effector gene. Proc Natl Acad Sci U S A. 2005;102: 15506–15511.

56. Li H, Kim JH, Koh SS, Stallcup MR. Synergistic effects of coactivators GRIP1 and beta-catenin on gene activation: cross-talk between androgen receptor and Wnt signaling pathways. J Biol Chem. 2004;279: 4212–4220.

57. Wang J, Ellwood K, Lehman A, Carey MF, She ZS. A mathematical model for synergistic eukaryotic gene activation. J Mol Biol. 1999;286: 315–325.

58. Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, et al. The nuclear receptor superfamily: the second decade. Cell. 1995;83: 835–839.

59. Chambon P. A decade of molecular biology of retinoic acid receptors. FASEB J Off Publ Fed Am Soc Exp Biol. 1996;10: 940–954.

60. Kersten S, Kelleher D, Chambon P, Gronemeyer H, Noy N. Retinoid X receptor alpha forms tetramers in solution. Proc Natl Acad Sci U S A. 1995;92: 8645–8649.

61. Lin BC, Wong CW, Chen HW, Privalsky ML. Plasticity of tetramer formation by retinoid X receptors. An alternative paradigm for DNA recognition. J Biol Chem. 1997;272: 9860–9867.

62. Skinner JP, Chen Y, Müller JD. Fluorescence fluctuation spectroscopy in the presence of immobile fluorophores. Biophys J. 2008;94: 2349–2360.

63. Macdonald PJ, Chen Y, Wang X, Chen Y, Mueller JD. Brightness analysis by Z-scan fluorescence fluctuation spectroscopy for the study of protein interactions within living cells. Biophys J. 2010;99: 979–988.

64. Hur K-H, Macdonald PJ, Berk S, Angert CI, Chen Y, Mueller JD. Quantitative Measurement of Brightness from Living Cells in the Presence of Photodepletion. PLoS ONE. 2014;9: e97440.

65. Hur K-H, Mueller JD. Quantitative Brightness Analysis of Fluorescence Intensity Fluctuations in E. Coli. PLoS ONE. 2015;10: e0130063.

66. Zipfel WR, Williams RM, Webb WW. Nonlinear magic: multiphoton microscopy in the biosciences. Nat Biotechnol. 2003;21: 1369–1377.

67. Gu M. Advanced Optical Imaging Theory. Springer Berlin Heidelberg; 2000.

68. Berland KM, So PT, Gratton E. Two-photon fluorescence correlation spectroscopy: method and application to the intracellular environment. Biophys J. 1995;68: 694–701.

69.   Hess ST, Webb WW. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. Biophys J. 2002;83: 2300–2317.

70.   Kubo R, Toda M, Hashitsume N. Statistical physics. II, volume 31 of Springer Series in Solid-State Sciences. Springer-Verlag, Berlin,; 1991.

71.   Terry BR, Matthews EK, Haseloff J. Molecular characterisation of recombinant green fluorescent protein by fluorescence correlation microscopy. Biochem Biophys Res Commun. 1995;217: 21–27.

72.   Hink MA, Griep RA, Borst JW, van Hoek A, Eppink MH, Schots A, et al. Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. J Biol Chem. 2000;275: 17556–17560.

73.   Chen Y. Analysis and applications of fluorescence fluctuation spectroscopy. Ph.D., University of Illinois at Urbana-Champaign. 1999.

74.   Müller JD. Cumulant analysis in fluorescence fluctuation spectroscopy. Biophys J. 2004;86: 3981–3992.

75.   Wu B, Müller JD. Time-integrated fluorescence cumulant analysis in fluorescence fluctuation spectroscopy. Biophys J. 2005;89: 2721–2735.

76.   Qian H, Elson EL. On the analysis of high order moments of fluorescence fluctuations. Biophys J. 1990;57: 375–380.

77.   Qian H. On the statistics of fluorescence correlation spectroscopy. Biophys Chem. 1990;38: 49–57.

78.   Mandel L, Sudarshan ECG, Wolf E. Theory of photoelectric detection of light fluctuations. Proc Phys Soc. 1964;84: 435.

79.   Saleh B. Photoelectron statistics: with applications to spectroscopy and optical communication. Springer-Verlag; 1978.

80.   Wu B, Chen Y, Muller JD. Dual-Color Time-Integrated Fluorescence Cumulant Analysis. Biophys J. 2006;91: 2687–2698.

81.   Chen Y, Tekmen M, Hillesheim L, Skinner J, Wu B, Müller JD. Dual-Color Photon-Counting Histogram. Biophys J. 2005;88: 2177–2192.

82.   Kask P, Palo K, Ullmann D, Gall K. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. Proc Natl Acad Sci U S A. 1999;96: 13756–13761.

83.   Slaughter BD, Schwartz JW, Li R. Mapping dynamic protein interactions in MAP kinase signaling using live-cell fluorescence fluctuation spectroscopy and imaging. Proc Natl Acad Sci. 2007;104: 20320–20325.

84. Hillesheim LN, Müller JD. The photon counting histogram in fluorescence fluctuation spectroscopy with non-ideal photodetectors. Biophys J. 2003;85: 1948–1958.

85. Zacharias DA, Violin JD, Newton AC, Tsien RY. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. Science. 2002;296: 913–916.

86. Qian H, Elson EL. Distribution of molecular aggregation by analysis of fluctuation moments. Proc Natl Acad Sci. 1990;87: 5479–5483.

87. Ha T, Tinnefeld P. Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging. Annu Rev Phys Chem. 2012;63: 595–617.

88. Peterman EJG, Brasselet S, Moerner WE. The Fluorescence Dynamics of Single Molecules of Green Fluorescent Protein. J Phys Chem A. 1999;103: 10553–10560.

89. Bogdanov AM, Mishin AS, Yampolsky IV, Belousov VV, Chudakov DM, Subach FV, et al. Green fluorescent proteins are light-induced electron donors. Nat Chem Biol. 2009;5: 459–461.

90. Saha R, Verma PK, Rakshit S, Saha S, Mayor S, Pal SK. Light driven ultrafast electron transfer in oxidative redding of Green Fluorescent Proteins. Sci Rep. 2013;3.

91. Slaughter BD, Huff JM, Wiegraebe W, Schwartz JW, Li R. SAM domain-based protein oligomerization observed by live-cell fluorescence fluctuation spectroscopy. PloS One. 2008;3: e1931.

92. Hillesheim LN, Chen Y, Müller JD. Dual-Color Photon Counting Histogram Analysis of mRFP1 and EGFP in Living Cells. Biophys J. 2006;91: 4273–4284.

93. Macdonald PJ, Chen Y, Mueller JD. Chromophore maturation and fluorescence fluctuation spectroscopy of fluorescent proteins in a cell-free expression system. Anal Biochem. 2012;421: 291–298.

94. Patterson GH, Knobel SM, Sharif WD, Kain SR, Piston DW. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. Biophys J. 1997;73: 2782–2790.

95. Drummond DR, Carter N, Cross RA. Multiphoton versus confocal high resolution z-sectioning of enhanced green fluorescent microtubules: increased multiphoton photobleaching within the focal plane can be compensated using a Pockels cell and dual widefield detectors. J Microsc. 2002;206: 161–169.

96. Didier P, Guidoni L, Bardou F. Infinite average lifetime of an unstable bright state in the green fluorescent protein. Phys Rev Lett. 2005;95: 090602.

97. Ulbrich MH, Isacoff EY. Subunit counting in membrane-bound proteins. Nat Methods. 2007;4: 319–321.

98. Delon A, Usson Y, Derouard J, Biben T, Souchier C. Photobleaching, mobility, and compartmentalisation: inferences in fluorescence correlation spectroscopy. J Fluoresc. 2004;14: 255–267.

99. Caccia M, Camozzi E, Collini M, Zaccolo M, Chirico G. Photon moment analysis in cells in the presence of photo-bleaching. Appl Spectrosc. 2005;59: 227–236.

100. Jain A, Liu R, Ramani B, Arauz E, Ishitsuka Y, Ragunathan K, et al. Probing cellular protein complexes using single-molecule pull-down. Nature. 2011;473: 484–488.

101. Smith EM, Macdonald PJ, Chen Y, Mueller JD. Quantifying protein-protein interactions of peripheral membrane proteins by fluorescence brightness analysis. Biophys J. 2014;107: 66–75.

102. Chaillan-Huntington C, Butler PJ, Huntington JA, Akin D, Feldherr C, Stewart M. NTF2 monomer-dimer equilibrium. J Mol Biol. 2001;314: 465–477.

103. Sanchez-Andres A, Chen Y, Müller JD. Molecular brightness determined from a generalized form of Mandel's Q-parameter. Biophys J. 2005;89: 3531–3547.

104. Fogarty KH, Chen Y, Grigsby IF, Macdonald PJ, Smith EM, Johnson JL, et al. Characterization of cytoplasmic Gag-gag interactions by dual-color z-scan fluorescence fluctuation spectroscopy. Biophys J. 2011;100: 1587–1595.

105. Smith EM, Mueller JD. The statistics of protein expression ratios for cellular fluorescence studies. Eur Biophys J EBJ. 2012;41: 341–352.

106. Saffarian S, Elson EL. Statistical Analysis of Fluorescence Correlation Spectroscopy: The Standard Deviation and Bias. Biophys J. 2003;84: 2030–2042.

107. van Kempen GM, van Vliet LJ. Mean and variance of ratio estimators used in fluorescence ratio imaging. Cytometry. 2000;39: 300–305.

108. Humpolíčková J, Gielen E, Benda A, Fagulova V, Vercammen J, vandeVen M, et al. Probing Diffusion Laws within Cellular Membranes by Z-Scan Fluorescence Correlation Spectroscopy. Biophys J. 2006;91: L23–L25.

109. Cluzel P, Surette M, Leibler S. An Ultrasensitive Bacterial Motor Revealed by Monitoring Signaling Proteins in Single Cells. Science. 2000;287: 1652–1655.

110. Le TT, Harlepp S, Guet CC, Dittmar K, Emonet T, Pan T, et al. Real-time RNA profiling within a single bacterium. Proc Natl Acad Sci U S A. 2005;102: 9160–9164.

111.   Ferguson ML, Le Coq D, Jules M, Aymerich S, Declerck N, Royer CA. Absolute quantification of gene expression in individual bacterial cells using two-photon fluctuation microscopy. Anal Biochem. 2011;419: 250–259.

112.   Coffman VC, Wu J-Q. Counting protein molecules using quantitative fluorescence microscopy. Trends Biochem Sci. 2012;37: 499–506.

113.   Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006;440: 358–362.

114.   Rosenfeld N, Perkins TJ, Alon U, Elowitz MB, Swain PS. A fluctuation method to quantify in vivo fluorescence data. Biophys J. 2006;91: 759–766.

115.   Teng S-W, Wang Y, Tu KC, Long T, Mehta P, Wingreen NS, et al. Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. Biophys J. 2010;98: 2024–2031.

116.   Ferguson ML, Le Coq D, Jules M, Aymerich S, Radulescu O, Declerck N, et al. Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states. Proc Natl Acad Sci U S A. 2012;109: 155–160.

117.   Nayak CR, Rutenberg AD. Quantification of Fluorophore Copy Number from Intrinsic Fluctuations during Fluorescence Photobleaching. Biophys J. 2011;101: 2284–2293.

118.   Gennerich A, Schild D. Fluorescence correlation spectroscopy in small cytosolic compartments depends critically on the diffusion model used. Biophys J. 2000;79: 3294–3306.

119.   Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S. Protein mobility in the cytoplasm of Escherichia coli. J Bacteriol. 1999;181: 197–203.

120.   Meacci G, Ries J, Fischer-Friedrich E, Kahya N, Schwille P, Kruse K. Mobility of Min-proteins in Escherichia coli measured by fluorescence correlation spectroscopy. Phys Biol. 2006;3: 255–263.

121.   Lill Y, Kaserer WA, Newton SM, Lill M, Klebba PE, Ritchie K. Single-molecule study of molecular mobility in the cytoplasm of Escherichia coli. Phys Rev E. 2012;86: 021907.

122.   Digman MA, Brown CM, Sengupta P, Wiseman PW, Horwitz AR, Gratton E. Measuring fast dynamics in solutions and cells with a laser scanning microscope. Biophys J. 2005;89: 1317–1327.

123.   Digman MA, Dalal R, Horwitz AF, Gratton E. Mapping the number of molecules and brightness in the laser scanning microscope. Biophys J. 2008;94: 2320–2332.

124. Ries J, Chiantia S, Schwille P. Accurate Determination of Membrane Dynamics with Line-Scan FCS. Biophys J. 2009;96: 1999–2008.

125. Skinner JP, Chen Y, Müller JD. Position-sensitive scanning fluorescence correlation spectroscopy. Biophys J. 2005;89: 1288–1301.

126. Petersen KE, Petersen K. Ergodic Theory. Cambridge University Press; 1989.

127. Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002;415: 141–147.

128. Gavin A-C, Superti-Furga G. Protein complexes and proteome organization from yeast to man. Curr Opin Chem Biol. 2003;7: 21–27.

129. Wu B, Piatkevich KD, Lionnet T, Singer RH, Verkhusha VV. Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. Curr Opin Cell Biol. 2011;23: 310–317.

130. Zhang J, Campbell RE, Ting AY, Tsien RY. Creating new fluorescent probes for cell biology. Nat Rev Mol Cell Biol. 2002;3: 906–918.

131. Wallrabe H, Periasamy A. Imaging protein molecules using FRET and FLIM microscopy. Curr Opin Biotechnol. 2005;16: 19–27.

132. Kerppola TK. Visualization of molecular interactions by fluorescence complementation. Nat Rev Mol Cell Biol. 2006;7: 449–456. 133. Galperin E, Verkhusha VV, Sorkin A. Three-chromophore FRET microscopy to analyze multiprotein interactions in living cells. Nat Methods. 2004;1: 209–217.

134. Shyu YJ, Suarez CD, Hu C-D. Visualization of AP-1–NF-κB ternary complexes in living cells by using a BiFC-based FRET. Proc Natl Acad Sci. 2008;105: 151–156.

Appendices

Appendix A. Detection optics

This section provides additional information about the detection optics of the tri-color FFS instrument. The optics consists of two dichroic mirrors (DC1 and 2), two band pass filters (BP1 and 2), three lenses (L1, 2 and 3) and one mirror (M) as illustrated in Figure 2.2. The dichroic mirrors DC1 and DC2 are FF506-Di03 (Semrock, Rochester, NY) and 580dcxr (Chroma Technology, Rockingham, VT), respectively. Their transmission spectrum is shown in Figure A.1. The transmission of BP1 (FF02-475/50, Semrock) and BP2 (FF01-531/40, Semrock) are found in Figure A.2. Three lenses of f=75.0 mm (AC254-075-A-ML, Thorlabs, Newton, NJ) are added after DC1. The distance between L1 and L2 is 2f, which is also the distance between L1 and L3. Each of these two lens combinations acts as a 1:1 afocal relay. A silver mirror M (MRA25-P01, Thorlabs) is added between DC2 and L3 to fold the beam path.
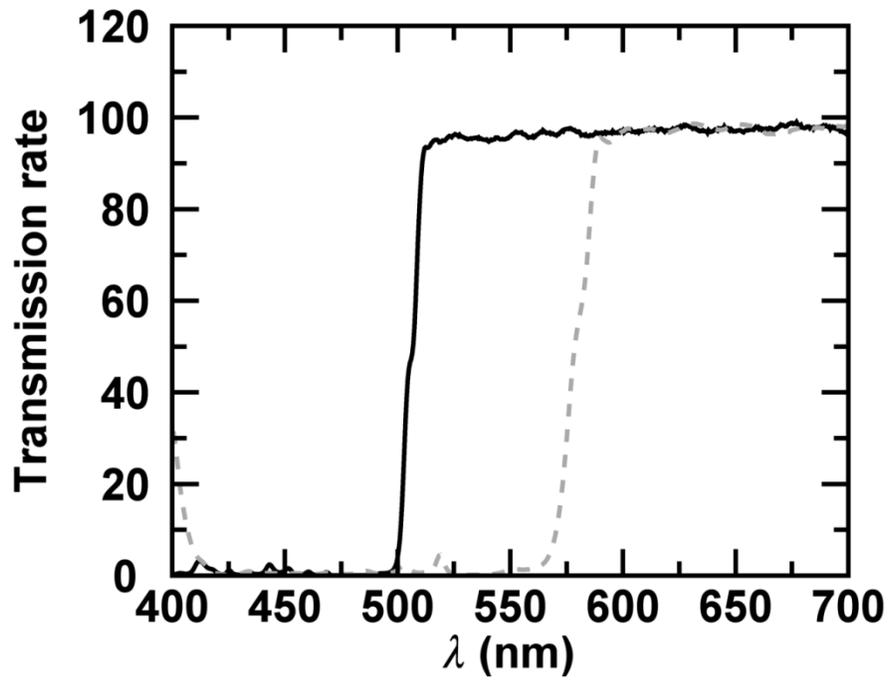
**Figure A.1 Transmission spectrum of dichroic mirrors DC1 (solid line) and DC2 (dashed line)**
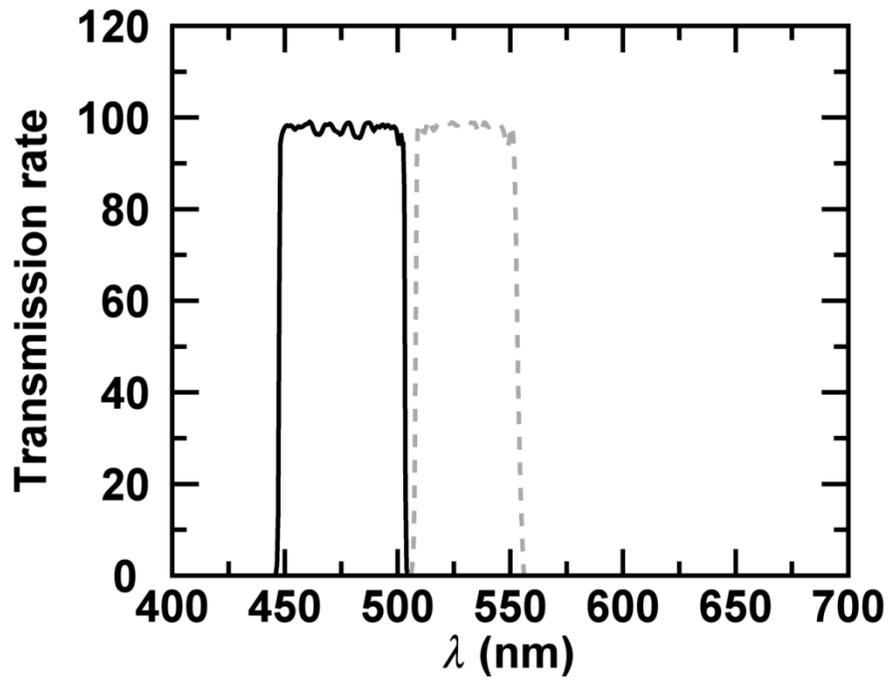
Figure A.2 Transmission spectrum of band-pass filters BP1 (solid line) and BP2 (dashed line)

Appendix B Spherical model

We consider z-scan FFS applied within a small spherical compartment, such as the nucleolus of mammalian cells. Following the notations in section 2.8, the spherical geometry of radius R is described by

$$V_{sph}(R) = \left\{ (\rho', \phi', z') \mid \rho' \in (0, \sigma), \phi' \in (0, 2\pi), z' \in (-R, R) \right\}, \tag{5.1}$$

where $\sigma = \sqrt{R^2 - z'^2}$ determines the boundary along the radial direction at z'.

We first discuss FFS parameters with the PSF's origin located at the center of the sphere (Figure B.1). For the mGL PSF (Eq. 2.4), the r-th order PSF volume (Eq. 2.43) is

$$V_{sph}^{(r)}(R) = \frac{\pi z_0 \omega_0^2}{4r} \int_{-\hat{R}}^{\hat{R}} \left(1 + \hat{z}'^2 \right)^{-r(1+\eta)+1} \left\{ 1 - Exp(-(\varphi_r(\hat{z}'))^2) \right\} d\hat{z}', \tag{5.2}$$

with $\hat{R} = R/z_0$, $\hat{z}' = z'/z_0$ and $\varphi_r(\hat{z}') = 2\sqrt{r\alpha^2 \left(\hat{R}^2 - \hat{z}'^2 \right) \big/ \left(1 + \hat{z}'^2 \right)}$. Therefore, the effective volume and the second gamma-factor of the spherical sample are given by $V_{sph}^{(1)}(R)$ and $\gamma_{2,\,sph}(R) = V_{sph}^{(2)}(R) \big/ V_{sph}^{(1)}(R)$, respectively.

Next, we discuss a scan along the z-axis through the center of the sphere. The r-th order PSF volume depends on the axial position z and is given by

$$V_{sph}^{(r)}(z, R) = \frac{\pi z_0 \omega_0^2}{4r} \int_{-\hat{R}}^{\hat{R}} \left(1 + \left(\hat{z}' - \hat{z} \right)^2 \right)^{-r(1+\eta)+1} \left\{ 1 - Exp(-(\varphi_r(\hat{z}' - \hat{z}))^2) \right\} d\hat{z}', \tag{5.3}$$

with $\hat{z} = z/z_0$. The scan position z is measured with respect to the center of the sphere. The average fluorescence intensity and its variance along the scan path are described by

$$\begin{aligned} \langle F(z) \rangle &= \lambda c V_{sph}^{(1)}(z, R) \\ \langle \Delta F^2(z) \rangle &= \gamma_{2,\,sph}(z, R) \lambda^2 c V_{sph}^{(1)}(z, R) \end{aligned} \tag{5.4}$$
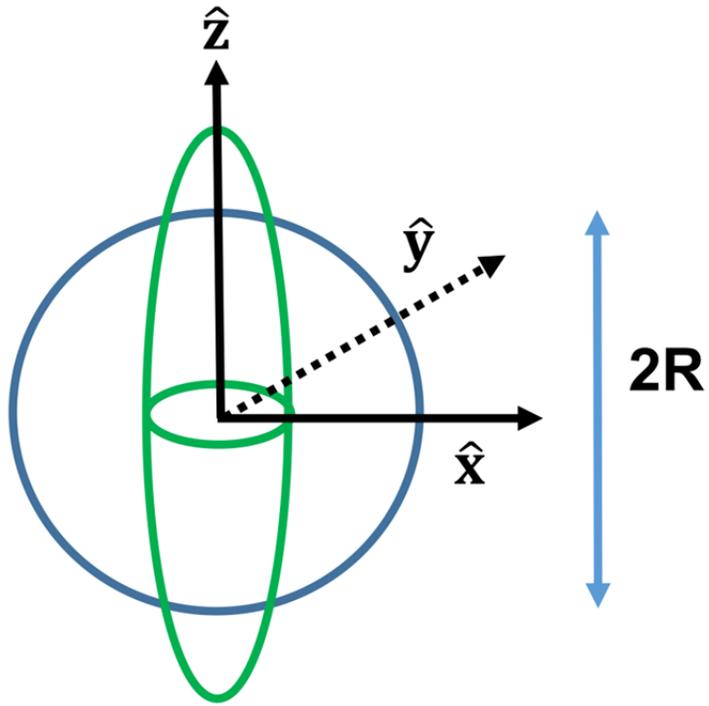
Figure B.1 Sphere with radius R and PSF.

Spherical sample (Blue line) together with PSF (green line) located at center of sphere.